TECHNISCHE
UNIVERSITÄT
DARMSTADT

Chemical reaction networks in random environments

Vom Fachbereich Elektrotechnik und Informationstechnik der
Technischen Universität Darmstadt

zur Erlangung des akademischen Grades eines
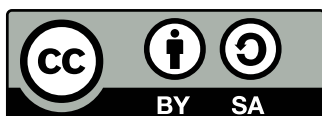Doctor rerum naturalium (Dr. rer. nat.)
genehmigte Dissertation

von

Mark Sinzger-D'Angelo, M.Sc.

Erstgutachter*in: Prof. Dr. techn. Heinz Köppl
Zweitgutachter*in: Prof. Dr. Wilhelm Stannat

Darmstadt 2024

Erklärungen laut Promotionsordnung

§ 8 Abs. 1 lit. d PromO

Ich versichere hiermit, dass von mir zu keinem vorherigen Zeitpunkt bereits ein Promotionsversuch unternommen wurde. Andernfalls versichere ich, dass der promotionsführende Fachbereich über Zeitpunkt, Hochschule, Dissertationsthema und Ergebnis dieses Versuchs informiert ist.

§ 9 Abs. 1 PromO

Ich versichere hiermit, dass die vorliegende Dissertation, abgesehen von den in ihr ausdrücklich genannten Hilfsmittlen, selbstständig und nur unter der Verwendung der angegebenen Quellen verfasst wurde. Weiterhin versichere ich, dass die "Grundsätze zur Sicherung guter wissenschaftlicher Praxis an der Technischen Universität Darmstadt" sowie die "Leitlinien zum Umgang mit digitalen Forschungsdaten an der TU Darmstadt" in den jeweils aktuellen Versionen bei der Verfassung der Dissertation beachtet wurden.

§ 9 Abs. 2 PromO

Ich versichere hiermit, dass die vorliegende Dissertation bisher noch nicht zu Prüfungszwecken gedient hat.

Kurzfassung

Zelluläre Prozesse finden nicht isoliert sondern vielmehr in einem Zusammenhang statt. Um diese Heterogenität zu berücksichtigen, betrachten wir die Modellklasse der chemischen Reaktionsnetzwerke mit äußerem Rauschen (der Umgebung), das sich als zusätzliche Stochastizität in den Reaktionsratenkonstanten äußert. Eine gemeinsame Beschreibung von Teilnetzwerk und Umgebung ist rechenaufwendig. Hier versuchen wir das Teilnetzwerk marginal zu beschreiben, als wäre es noch eingebettet. Das bringt den Vorteil mit sich, stochastische Eigenschaften des Teilnetzwerks den Eigenschaften der Umgebung zuzuschreiben. Die marginale Beschreibung ermöglicht darüber hinaus die Berechnung von Informationsgrößen. Um auszutüfteln wie die Zelle in der Anwesenheit von Umgebungsrauschen zuverlässige Entscheidungen trifft, zielen wir rechnerisch auf die Berechnung der Pfadinformation zwischen Umgebung und Teilnetzwerk ab. Dazu betrachten wir den Poissonkanal, der durch die Reaktionszähler, die das CRN hinreichend beschreiben, motiviert ist. Für geschlossene Ausdrücke und zu Berechnungszwecken brauchen wir vereinfachende Annahmen. Folglich liegt ein besonderer Fokus der Dissertation auf linear Teilnetzwerken in diskretwertigen Markov-Umgebungen und auf allgemeinen Teilnetzwerken in linearer Umgebung.

Wir tragen auf mehreren Ebenen der stochastischen Beschreibung bei: (i) Auf der Ebene der Momente verallgemeinern wir Ergebnisse der Warteschlangentheorie zur exakten Berechnung des stationären Mittelwerts von linear Teilnetzwerken in diskretwertigen Markov-Umgebungen. Unsere Formel hängt von den Reaktionsratenkonstanten des linearen Teilnetzwerks ab, sowie vom Generator und der stationären Verteilung der Markov-Umgebung. Wir erweitern die Spektralzerlegungen in intrinsisches und extrinsischen Rauschen auf den Fall von korrelierten Umgebungskomponenten. (ii) Wir präsentieren Liouville-Mastergleichungen mit Randbedingungen für den Wahrscheinlichkeitsverlauf einer marginalen CRN-Beschreibung mittels Hilfsstatistiken. Unsere Methode der Rückwärtsrekurrenzzeit-Parametrisierung (BReT-P) für stückweise-deterministische Markov-Prozesse (PDMP) führt eine Standardform für die marginale CRN-Beschreibung mittels approximativer Filter ein. Wir leiten verallgemeinerte Mastergleichungen für Beispiele mit wenig Molekülarten her. (iii) Auf der Prozessebene, formalisieren wir eine PDMP-Unterklasse, die Dirac-Maße als Verteilung bei Sprüngen hat, welche wir Dirac-PDMP nennen. Wir führen einen approximativen Marginal-Simulationsalgo-rithmus an, der auf optimalem linearen Filtern basiert.

Für CRN in einer linear stationären Umgebung, d.h. mit Exponential-Auto/Kreuz-Kovarianzfunktion, geben wir einen approximativen Filter basierend auf Snyders optimalem linearen Filtern für Zählprozesse. Indem wir chemische Reaktionen als Ereignisse betrachten, stellen wir eine Verbindung zwischen CRN in einer linearen Umgebung und Hawkesprozessen her, einer Klasse selbstverstärkender Zählprozesse, die zur Analyse von Ereignissen gebräuchlich ist. Wir zeigen, dass man die Hawkes-Annäherung auf äquivalente Weise über ein Momentenschließungsverfahren und die optimale lineare Approximation bezüglich des mittleren quadratischen Fehlers erhält. Darüber hinaus nutzen wir Martingaltechniken um Ergebnisse zu präsentieren zur Übereinstimmung von Hawkesprozess und exaktem marginalem Prozess in ihren Statistiken zweiter Ordnung, d.h. Kovarianz und Auto/Kreuzkorrelation. Indem wir das Hawkesmodell als Referenz nehmen, schreiben wir stochastische Eigenschaften des Teilnetzwerks jeweils der linearen oder nicht-lineare Dynamik der Umgebung zu. Wir führen einen approximativen Marginal-Simulationsalgorithmus ein und verdeutlichen ihn in Fallstudien.

Die empirischen Ergebnisse konzentrieren sich auf die Konformationsänderungen von Makromolekülen und auf Exkursionen. Makromoleküle, die in verschiedenen Konformationen vorhanden sind, wurden für Genexpressionsmodelle unter der Benutzung thermodynamischer Ensembles erforscht. Wir hingegen beziehen die Dynamik der Konformationswechsel als Umgebungskomponente ein. Mit unserem Modellredukionsansatz erhöht diese Einbindung nicht die Zahl der Molekülarten. Damit überbrücken wir eine Lücke zwischen Strukturkinetik und Genexpressionsmodellen, was unser Verständnis von genregulatorischen Netzwerken verbessern und das Designen genetischer Schaltkreise erleichtern kann. Für die Modulation der Translation durch die mRNA quantifizieren wir, wie viel die Dynamik der mRNA-Struktur zur translationalen Heterogenität beiträgt. In weiteren Anwendungen konzentrieren wir uns auf das Phänomen von Exkursionen, nahezu linearen Anhäufungen von Molekülen in Phasen mit verschwindender Abbaurate. Wir erforschen wie dieses den stationären Mittelwert erhöht und führen diesen Anstieg auf den Zustand Null zurück. Wir besprechen einen Regler, der diesen Effekt abschwächt und werten die Fähigkeit des Hawkesmodells aus, Exkursionen adäquat abzubilden.

Um Informationsgrößen des Poissonkanals zu berechnen, stellen wir einen analytischen Ansatz vor, der die Monte-Carlo-Schätzung umgeht. Diese simulationslose Schätzmethode wird ermöglicht, indem wir unseren Augenmerk auf die bedingte Intensität von Zählprozessen und ihre asymptotische Verteilung als die zentralen Größen richten. Für den Poissonkanal mit binärem Markov-Eingabesignal, drücken wir die gegenseitige Information als Riemann-Integral aus. Während in den klassischen Resultaten das kapazitätsausschöpfende Eingabesignal den Off-Zustand bevorzugt, führen wir On-präferierende Bereiche für nach unten beschränkte mittlere Verweildauern in den On- und Off-Zuständen an. Zudem führen wir Evidenz dafür an, dass unter den binären Eingabesignalen die exponentiellen Verweildauern nicht optimal sind. Nachdem wir die Eponentialbedingung lockern, indem wir mehrere Off-Zustände zulassen, diskutieren wir den informationstheoretischen Vorteil des periodischen Durchlaufens mehrerer Off-Zustände.

Durchweg nutzen wir bedingte Wahrscheinlichkeiten als Werkzeug zur Modellreduktion. Wir schlagen zwei Wege, das Bedingen auf die Umgebung und auf das Teilnetzwerk, ein. Wir zeigen die Äquivalenz beider Wege in einer vereinenden Betrachtungsweise, die die Turmeigenschaft auf die Kolmogorov-Rückwärts-Gleichung anwendet. Indem wir verschiedene approximative Filter für die konformationsändernde Umgebung vergleichen, beleuchten wir ihre gegensätzlichen Stärken. Wir führen Zweifel an hinsichtlich der Frage ob die dargestellten Modellreduktionsstrategien den Fluch der Dimensionalität von CRN auflösen kann verglichen mit dem Doob-Gillespie-Algorithmus. Zusammenfassend tragen wir zur Modellreduktion, marginalen Simulation, Berechnung von Informationsgrößen und Attributionstheorie für CRN in zufälliger Umgebung bei.

Abstract

Cellular processes operate in context, rather than in isolation. To account for this heterogeneity, we consider the model class of chemical reaction networks (CRNs) with extrinsic noise (the environment), manifesting as additional stochasticity in the reaction rate constants. A joint description of subnetwork and environment is computationally heavy. Here, we attempt to marginally describe the subnetwork as if it was still embedded. This comes with the merit of attributing stochastic properties of the subnetwork to features of the environment. Additionally, the marginal description enables the estimation of information measures. To puzzle out how the cell makes reliable decisions in the presence of the environmental noise, we computationally target the path mutual information between environment and subnetwork. Namely, we consider the minimal Poisson channel, motivated by the reaction counters that sufficiently describe the CRN. For closed-form expressions and computational purposes, we require simplifying assumptions. Hence, a particular focus of this thesis lies on linear subnetworks in discrete state Markov environments and on general subnetworks in linear environments.

We contribute at different levels of the stochastic description: (i) At the level of moments, we generalize results from queuing theory about the exact stationary mean evaluation of linear CRNs in a discrete state Markov environments. Our analytic expression depends on the reaction rate constants of the linear subsystem, as well as the generator and stationary distribution of the Markov environment. We extend spectral decomposition results on intrinsic and extrinsic noise to the case of correlated environment components. (ii) We present Liouville master equations with boundary conditions for the probability evolution of a marginal CRN description via auxiliary statistics. Our method of the backward recurrence time parametrization (BReT-P) for piecewise-deterministic Markov processes (PDMP) introduces a standard form for the marginal description of CRNs via approximate filters. We derive generalized master equations for examples with a low number of species. (iii) At the process level, we formalize a subclass of PDMPs having Dirac measures at jump times, which we call Dirac-PDMPs. We offer an approximate marginal simulation algorithm based on optimal linear filtering.

For CRNs in a linear stationary environment, i.e., with exponential auto/cross-covariance function, we provide an approximate filter that is based on Snyder's optimal linear filtering for counting processes. By regarding the chemical reactions as events, we establish a link between CRNs in a linear environment and Hawkes processes, a class of self-exciting counting processes widely used in event analysis. We show that the Hawkes approximation is equivalently obtained via moment closure scheme or as the optimal linear approximation under the mean-square error. Furthermore, we use martingale techniques to provide results on the agreement of the Hawkes process and the exact marginal process in their second order statistics, i.e., covariance, auto/cross-correlation. Taking the Hawkes model as a reference, we attribute stochastic properties of the subnetwork to the linear or non-linear dynamics of the environment, respectively. We introduce an approximate marginal simulation algorithm and illustrate it in case studies.

The empirical results focus on structure switching of macromolecules and excursions. Macromolecules that are abundant in different conformations have been studied for gene expression models using thermodynamic ensembles. However, we include switching dynamics as the environmental component. With our model reduction approach, this inclusion does not increase the number of species. Thereby, we bridge a gap between structure kinetics and gene expression models, which can further improve our understanding of gene regulatory networks and facilitate genetic circuit design. For the modulation of translation by mRNA, we provide a method to quantify how mRNA structure dynamics contributes to translational heterogeneity. In a further set of applications we focus on the phenomenon of excursions, near-to-linear accumulation of a species in periods of a vanishing decay rate. We examine how this increases the stationary mean, attributing the increase to the zero state. We discuss a controller that mitigates the effect and evaluate the ability of the Hawkes model to capture excursions.

For estimating information-measures of the Poisson channel, we present an analytic approach that circumvents Monte Carlo sampling. This simulation-free estimation method is enabled by establishing the conditional intensity of counting processes and its asymptotic distribution as central quantitites. For the Poisson channel with binary Markov input, we express the mutual information as a Riemann integral. While in the classical result the capacity-achieving input favors the Off state, we in contrast report On-favoring regimes for lower-bounded average soujourn time in the On and Off states. Additionally, we provide evidence that among the binary inputs the exponential sojourn times are not

optimal. Relaxing the exponential constraint of the binary input by allowing for multiple Off states, we discuss the information-theoretic advantage of cycling through several Off states.

Throughout, stochastic conditioning is used as the tool for model reduction. We take the two routes of conditioning on the environment and on the subnetwork. We present the equivalence of both routes in a unifying approach that uses the tower property on the Kolmogorov backward equation. Comparing different approximate filters for the structure switching environment, we shed light on their orthogonal strengths. We present doubts as to whether the presented model reduction strategies can resolve the curse of dimensionality of CRNs compared to the Doob-Gillespie algorithm. Overall, we contribute to the model reduction, marginal simulation, computation of information-measures and attribution theory for CRNs in random environments.

Acknowledgments

I thank my supervisor Heinz Koeppl, my donors, colleagues, project collaborators, university staff, friends and family for scientific discussion and support, funding, emotional support, advice and opportunities. With gratitude I dedicate this work to Philipp D'Angelo, Seph Forslund-Startceva, Max Gehri, Johannes Luff and Suellen Dutra Pereira.

Darmstadt, 05th of May, 2024

Contents

1   Introduction

In the real world, the systems we encounter are rarely isolated, but they interact in complex ways. Thus, when explaining a phenomenon via mathematical modeling, we face the choice of which components to include in the model. In closed systems, all factors relevant to the modeling of the system and their interactions are explicitly described, all components are internal to the model description. On the contrary, in open systems, external components influence the dynamics of the system, i.e., the system is modulated by its environment. The term environment originates from the theory of open quantum systems, in which the term (heat) bath is equivalently used [1]. When conducting experiments, it is almost never feasible to explicitly capture all the factors that affect the observed system. As a consequence, almost any experimental setup in a real world is an open system. The partial observations are complemented by an external environment component that is hidden from the observations. This holds for cellular processes in particular [2, 3, 4]. Gene regulatory processes operate in a heterogeneous context, rather than in isolation [5, 6]. The heterogeneity has its origin in the cell-to-cell variability (static heterogeneity) as well as in temporal fluctuations. As an example for the former, each cell is equipped with a different number of ribosomes. As an example for the latter, the process of transcription undergoes fluctuations in transcription factors. The number of components that are simultaneously observable in live-cell fluorescence microscopy is currently limited to a handful because the spectra of common fluorophores overlap [7, 8, 9], which makes partial observability a relevant modeling criterion. However, under some circumstances, a closed system models the phenomenon of interest sufficiently well, while facilitating a tractable mathematical representation. Techniques such as the quasi-steady state assumption were developed to close a system [10, 11].

Mathematical modeling is used to narrow down the mechanisms that are in accordance with experimental data and to ultimately make predictions [12]. Viewing models as magnifying glasses, in the optimal case, the model magnifies precisely the relevant components to describe a phenomenon. This can be the observed components and a small number of additional, unobserved, quantities. The exterior components are abstracted into an environment. The environment can be an implicit compilation of external influences whose fingerprint shows in the subnetwork or an explicit physical component. To capture this difference, we distinguish between phenomenological and mechanistic models [13], illustrated by the use of the Hill function [14]. The Hill coefficient has an interpretation as an indicator for cooperativity [15]. In some cases (quorum sensing, cooperative binding to enzymes), a physical mechanism can be detected that implements the cooperativity and justifies the Hill function as a mechanistic model. However, the Hill function is often used in a phenomenological sense [16, 17], due to its properties (bounded, S-shaped, with a concentration of maximal steepness), without having identified a mechanism of cooperativity on the molecular level. Another relevant example for gene regulation is the two-state promoter model [18], often linked to the term 'transcriptional bursting' [19], see figure 1A. In bacteria, the random switching between an inactive and an active state of DNA accessibility
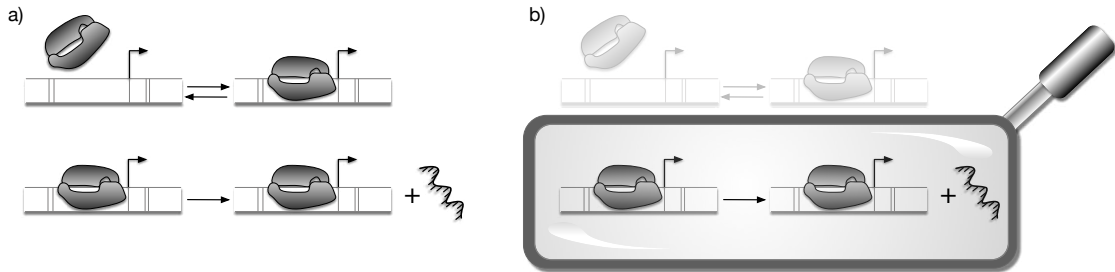
Figure 1: **The standard transcription model with a two-state promoter.** a) A random telegraph promoter switches between its active and its inactive state. Only in the active state, mRNA is synthesized. The promoter stays active, so multiple mRNA strands can be transcribed in one active period. The transcription reaction is of first order with fixed rate. b) Experimentally only the transcription events are observed. The promoter is modeled as a context. An observation model for transcription counts is obtained when we marginalize the joint system over the context. The transcription reaction is of order zero with stochastic rate, i.e., a doubly stochastic Poisson process.

and exponential waiting times could be validated. However, for yeast or mammalian cells, the two-state promoter model was merely used in a phenomenological way because it generates overdispersed mRNA count data [20]. When time-lapse data of the transcriptional process became available, the Markovian assumption was found not to hold for yeast [21] and mammalian cells [22]. Biophysicists suggested mechanistic models that include more than one rate-limiting step [22, 23]. Modelers proposed phenomenological promoter architectures, including cooperativity effects, refractory states and competitive binding [24, 17, 25, 26, 27] as refinements to the two-state promoter model.

In this work, I focus on the particular model class of stochastic chemical reaction networks (CRNs) in a random environment. CRNs are Markov jump processes on the state space $\mathbb{N}^n$ of molecule copy numbers [28]. Transitions are dictated by reactions. Each reaction, as known from textbook chemistry, specifies a number of reactants converting to a number of products. The net change from reactants to products, i.e., the change vector of the reaction, establishes which transitions are possible. The rate with which the reaction occurs is computed from the state, i.e., the copy number of the reactants. It is specified by a functional form, often using the principle of mass-action kinetics. But other functional forms, such as a Hill function, are possible. In the classical case, the functional form includes rate constants. For instance, with the mass-action principle, this is simply a proportionality constant.

Typical quantities of interest for CRNs are the probability distribution, or the generating function, and moments, both in the transient and stationary behavior. While being a pioneer of stochastically described chemical reactions, Bartholomay [29] still emphasized the analogy of the mean equations with the deterministic counterpart. However, the following works [28, 30] demonstrated that bimolecular reactions cause a characteristic deviation of stochastic systems when the system is not in the thermodynamic limit, highlighting the need for other techniques. Early works computed the generating function and mean at stationarity for small bimolecular networks comprising two balanced reactions [31]. The observation that bimolecular reactions result in an unclosed mean equation brought more attention to unclosed (hierarchical) moment equations. Unclosed moment equations have prompted researchers to pursue several strategies. Among them, stochastic simulations for systems with highly abundant components are time-consuming and therefore computationally prohibitive. Equally prohibitive is the use of the master equation, which gave rise to more efficient hybrid methods [32, 33, 34]. Overall, previous studies developed and extensively explored various moment closure schemes [35, 36, 37, 38], identified the limitation of moment closures with regards to bimodal distributions [34] and extinction [39], and pointed out the limited local character of moment closures [40]. More recently, linear programming under positive semi-definite constraints was employed to compute upper and lower moment bounds [41, 42, 43] and approximated the moments in the case of tight bounds.

For the class of CRNs in a random environment, we replace the rate constants by stochastic processes. In this way, we model cell-to-cell heterogeneity, with the random environment manifesting as stochastic reaction rates, and the subsystem embedded into this environment. One way to obtain an embedded subsystem is by partitioning the species of a larger network into subsystem and environment

species. Adhering to the experimental reality of partial observations, one modeling priority is to give a self-contained description of the observed subsystem while accounting for the environment as if the subsystem was still embedded, see figure 1B. A transfer from a joint closed description of the environment and the subsystem to an (approximate) self-contained description of the subsystem is what we call model reduction. We seek for model reductions that succeed to capture a certain phenomenon that the original joint system exhibits. The effect of a random environment on the noise in linear subsystems, i.e., on the variance, is well-studied. Many research groups adopted the decomposition of noise into intrinsic and extrinsic components since its introduction by Swain and colleagues [44], and this concept was expanded upon in later studies [45, 46, 47, 48]. Of equal interest is the question of how the environment shapes the stationary distribution of the subsystem, e.g., bimodality [49]. In the related field of queuing theory, stochasticians studied birth-death process in a random environment. For a Markov environment, O'cinneide and Purdue [50] derived expressions for the stationary factorial moments and reported bimodal stationary distributions. Random-telegraph-modulated service rates were fully characterized in [51] and Falin [52] extended it to simultaneous birth- and death-modulation by a semi-Markov environment.

Several levels of description are possible for stochastic processes [53, 54]: 1) the process evolution, specifying a generative model, such as an SDE or doubly stochastic (multivariate) Poisson process, 2) the probability evolution, specifying the generator or the differential form of the Chapman-Kolmogorov forward equation, 3) the moment evolution, adjoined to the generator. We note that 2) and 3) describe deterministic objects. Usually, we will encounter a model reduction that starts from an open description of the environment and the subsystem, i.e., only first- and second-order moment information on the environment will be used. It is interesting to study failure modes of the quasi-steady state (Q.SS) assumption. Besides the Q.SS-approximated models, we develop a novel reference model from a linearization of the reaction rate based on optimal linear stochastic filtering. The deviations from the reference model reveal which stochastic properties of the subnetwork must be classified as consequences of the non-linear dynamic evolution of the intensity and which persist under the linearized dynamic evolution of the intensity. We are guided by the question, which effects persist under a linearization and can thus be attributed to a linearized version of the environment. Besides attributing subnetwork properties to the environment, a marginal description of the embedded process is of interest for (i) marginal simulations that bypass the co-simulation of the environment, (ii) obtaining new process equations from which moment equations can be derived, (iii) the computation of information-theoretic quantities, and (iv) state estimation of the unobserved environment. A focus of this thesis will be on results for linear systems. We consider linear environments embedding general subnetworks and linear subnetworks embedded in general environments.

Related publications

This thesis is based on the following publications:

[A1] M. Sinzger, M. Gehri, and H. Koeppl. "Poisson channel with binary Markov input and average sojourn time constraints". In: *2020 IEEE International Symposium on Information Theory: Proceedings* (2020). DOI: 10.1109/ISIT44484.2020.9174360.

[A2] Mark Sinzger-D'Angelo and Heinz Koeppl. "Counting Processes with Piecewise-Deterministic Markov Conditional Intensity: Asymptotic Analysis, Implementation and Information-Theoretic Use". In: *IEEE Transactions on Information Theory* (2023), pp. 1–1. DOI: 10.1109/TIT.2023.3293996.

[A3] Mark Sinzger-D'Angelo and Heinz Koeppl. "Hawkes process modelling for chemical reaction networks in a random environment". In: *bioRxiv* (2023), pp. 2023–08.

[A4] Mark Sinzger-D'Angelo, Sofia Startceva, and Heinz Koeppl. "Bye bye, linearity, bye: quantification of the mean for linear CRNs in a random environment". In: *Journal of Mathematical Biology* 87.3 (2023), p. 43.

[A5] Mark Sinzger-D'Angelo et al. "Effects of mRNA conformational switching on translational noise in gene circuits". In: *The Journal of Chemical Physics* 160.13 (2024).

2 CRNs in a random environment

Stochastic CRNs model the synthesis, conversion and decay of molecules under the assumption of spatial homogeneity and dominant effects of low copy numbers [28]. Often, the processes that are modeled are not closed, but are embedded in the cellular context, which we call a random environment [3]. The concept of the random environment is abstract enough to include cell-to-cell heterogeneity, such as a varying polymerase copy number, and fluctuating processes within the cell, such as conformation changes of mRNA. Our way to model the embedding into a random environment is by replacing the constants in the propensity functions by external stochastic process. For instance, the rate of transcription synthesis events is modulated by multiple factors [55]. Recruitment of polymerase and transcription factors as well as unwinding of the DNA strand, all contribute to the activation of a promoter state, where transcription is initiated [56]. In other words, the reaction rate constants are replaced by random (time-varying) quantities. An embedding into a random environment naturally arises when one part of a larger CRN is labeled as the environment and the remaining part is the process of interest. Diffusion processes that remain positive, such as the Cox-Ingersoll-Ross (CIR) process [57] or a positive function of an Ornstein-Uhlenbeck process [58], are other ways to model fluctuating concentrations as environment components. CRNs formalize the stochastic dynamics of molecules of d types as Markov jump processes in the state space $\mathbb{N}^d$ of species copy numbers on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$.

2.1 Multivariate counting processes

By regarding the chemical reactions as events, CRNs are multivariate counting processes. External noise manifests as stochastic reaction rates. To operate mathematically with these models, we provide the martingale calculus for them. The stochastic concept of filtrations is introduced to capture different stochastic conditioning.

Throughout we consider the setting of the subsystem $X(t)$ embedded in a random environment $Z(t)$. The state $X(t) = [X_1(t), \ldots, X_d(t)]^T$ contains the number of molecules of type $\mathrm{X}_1, \ldots, \mathrm{X}_d$ at time t. Transitions of the Markov jump process are dictated by M reactions

$$
\begin{aligned}
\mathrm{R}_1: \quad & \textstyle\sum_{i=1}^d S_{i1}\mathrm{X}_i \quad \xrightarrow{a_1(X(t),Z(t))} \quad \textstyle\sum_{i=1}^d P_{i1}\mathrm{X}_i \\
& \vdots \qquad\qquad\qquad\quad \vdots \\
\mathrm{R}_M: \quad & \textstyle\sum_{i=1}^d S_{iM}\mathrm{X}_i \quad \xrightarrow{a_M(X(t),Z(t))} \quad \textstyle\sum_{i=1}^d P_{iM}\mathrm{X}_i
\end{aligned}
\tag{1}
$$

where $S_{ij}, i = 1, \ldots, d, j = 1, \ldots, M$ are the substrate coefficients and $P_{ij}, i = 1, \ldots, d, j = 1, \ldots, M$ the product coefficients. The superscripts $a_j(X(t), Z(t))$ are the propensity functions. We assume that they depend on the current state $X(t)$ and on an external stochastic process $Z(t)$. For examples of this setting, see section 6. We call $\nu = P - S$ the stoichiometric matrix and its columns $\nu_1, \ldots, \nu_M$ the change vectors. Transitions of the Markov jump process are possible from x to $x + \nu_j$ and occur at a rate given by the propensity function. We make this rigorous in the following counting process description. First, we introduce the multivariate counting process $Y(t) = [Y_1(t), \ldots, Y_M(t)]^T$, where the component $Y_j(t)$ counts how often the Reaction R_j has occurred until time t, so naturally throughout $Y(0) = 0$. Then the current state $X(t)$ can be computed from the initial state and $Y(t)$ as

$$
X(t) = X(0) + \sum_{j=1}^M Y_j(t)\nu_j = X(0) + N \cdot Y(t).
\tag{2}
$$

We use the terms point process and counting process equivalently. Note that we assume a unidirectional coupling, with the environment Z modulating the subnetwork, but not the other way around.

2.1.1 Preliminaries on counting processes

We recall the theory of counting processes, which follows [59, §1 and §2]. The main purpose of this subsection is the definition 2.2 of the conditional intensity of a counting process. Let $(Z_t)_{t \geq 0}$ be a stochastic process and for each $t \geq 0$ denote by $\sigma(Z_s, 0 \leq s \leq t)$ the smallest sigma-algebra for which

all $Z_s, s \le t$ are measurable. A **filtration** is an increasing family $(\mathcal{F}_t)_{t \ge 0}$ of sigma-algebras. The filtration $\mathcal{F}_t^Z := \sigma(Z_s, 0 \le s \le t)$ is said to be **generated by** $Z_t, t \ge 0$ and is called the **canonical filtration** of the process $(Z_t)_{t \ge 0}$. Let $(Z_t)_{t \ge 0}$ be a stochastic process and $(\mathcal{F}_t)_{t \ge 0}$ be a filtration. Then $(Z_t)_{t \ge 0}$ is called **adapted to** $(\mathcal{F}_t)_{t \ge 0}$ if $\mathcal{F}_t^Z \subseteq \mathcal{F}_t$ for all $t \ge 0$. The smallest filtration to which a process is adapted is its canonical filtration.

Definition 2.1 (Counting process, jump times, intensity). Let $(Y_t)_{t \ge 0}$ be a stochastic process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then $(Y_t)_{t \ge 0}$ is called a **counting process** if (i) its paths are right-continuous and piecewise constant, (ii) $Y_0 = 0$ and (iii) $Y_t - Y_{t-} \in \{0, 1\}$ for all $t \ge 0$, where $Y_{t-} = \lim_{s \to t-} Y_s$ denotes the left-sided limit. Denote by $0 = \sigma_0 < \sigma_1 < \sigma_2 < \dots$ the **jump times** of $(Y_t)_{t \ge 0}$, i.e., $\sigma_0 := 0$ and, recursively,

$$\sigma_i := \min\{t > \sigma_{i-1} : Y_t - Y_{t-} = 1\},$$

where $Y_{t-} = \lim_{s \to t-} Y_s$ denotes the left-sided limit. Let $(Y_t)_{t \ge 0}$ be adapted to a filtration $(\mathcal{F}_t)_{t \ge 0}$ and let $(\lambda_t)_{t \ge 0}$ be a non-negative $\mathcal{F}_t$-predictable process. Then $(Y_t)_{t \ge 0}$ admits the $\mathcal{F}_t$-**intensity** λ_t iff
a) for all $t \ge 0$

$$\int_0^t \lambda_s \, \mathrm{d}s < \infty$$

$\mathbb{P}$-almost surely and b) for all non-negative $\mathcal{F}_t$-predictable processes $(C_t)_{t \ge 0}$ it holds

$$\mathbb{E}\left[\int_0^\infty C_s \, \mathrm{d}Y_s\right] = \mathbb{E}\left[\int_0^\infty C_s \lambda_s \, \mathrm{d}s\right].$$

The counting process is **non-explosive**, if $\lim_{i \to \infty} \sigma_i = \infty$.

Throughout, $(Y_t)_{t \ge 0}$ denotes a non-explosive counting process on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and we only consider counting processes $(Y_t)_{t \ge 0}$ that admit a predictable $(\mathcal{F}_t, \mathbb{P})$-intensity λ_t. To highlight its dependency on the filtration $(\mathcal{F}_t)_{t \ge 0}$, we usually include it in the notation. When we compare different probability measures, we denote it explicitly, otherwise the intensity's dependency on $\mathbb{P}$ is ignored.

Definition 2.2 (Conditional intensity). Let $(Y_t)_{t \ge 0}$ be a counting process. We call its $\mathcal{F}_t^Y$-intensity the **conditional intensity** (CI), denoted by $(\hat{\lambda}_t)_{t \ge 0}$.

We follow Brémaud [59] and Daley, Vere-Jones [60] in using the term *conditional intensity* for the $\mathcal{F}_t^Y$-intensity. It is not to be confused with its other usage as a special case of the Papangelou intensity. For an introduction on the CI we refer the reader to [60, §7.2].

The Poisson channel without feedback, introduced by Kabanov [61], is a *doubly stochastic Poisson process* $(Y_t)_{t \ge 0}$ defined via an external measurable process $(Z_t)_{t \ge 0}$ on a state space $\mathcal{Z}$. Informally, $(Y_t)_{t \ge 0}$ is conditionally a Poisson process with non-homogeneous intensity $\lambda(Z_t, t)$ given $(Z_t)_{t \ge 0}$. Formally, we use the following definition 2.3.

Definition 2.3 (Doubly stochastic Poisson process, modulating process). Let $(Y_t)_{t \ge 0}$ be a counting process adapted to $(\mathcal{F}_t)_{t \ge 0}$ with the $\mathcal{F}_t$-intensity λ_t. If there exists a measurable process $(Z_t)_{t \ge 0}$ on a state space $\mathcal{Z}$ and a function $\lambda \colon \mathcal{Z} \times [0, \infty) \to \mathbb{R}_{\ge 0}$, such that $\lambda_t = \lambda(Z_t, t)$ and $\mathcal{F}_0$ contains $\mathcal{F}_\infty^Z$, then $(Y_t)_{t \ge 0}$ is called a **doubly stochastic Poisson process** with $\mathcal{F}_t$-intensity λ_t and $(Z_t)_{t \ge 0}$ is called its **modulating process**.

Remark 2.4. Instead of $(\mathcal{F}_t)_{t \ge 0}$ with $\mathcal{F}_\infty^Z \subseteq \mathcal{F}_0$ we can also consider $\mathcal{F}_t^{Z,Y} := \sigma(Z_s, Y_s : 0 \le s \le t)$. More informally, $(Y_t)_{t \ge 0}$ is conditionally a Poisson process with non-homogeneous intensity $\lambda(Z_t, t)$ given $(Z_t)_{t \ge 0}$. In this work, we consider the setting $\lambda_t = \lambda(Z_t)$. This is precisely the setting of the Poisson channel without feedback in the terminology of Kabanov [61].

Note that we lose the dependency of the intensity on a filtration in the informal definition. This dependency is helpful, because it allows a change of filtration. For this purpose let us recall the definition of the conditional expectation. For a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$ and a sigma-algebra $\mathcal{G}$ the **conditional expectation** $\mathbb{E}[X|\mathcal{G}]$ is the $\mathcal{G}$-measurable random variable $\tilde{X}$ for which

$$\int_B X \, \mathrm{d}P = \int_B \tilde{X} \, \mathrm{d}P$$

for all $B \in \mathcal{G}$. It is unique except for changes on $\mathbb{P}$-null sets.

Theorem 2.5 (Change of filtration, Brémaud 1981)**.** *Let* $(Y_t)_{t \geq 0}$ *be a counting process adapted to* $(\mathcal{F}_t)_{t \geq 0}$ *with the* $\mathcal{F}_t$*-intensity* λ_t *and* $\mathcal{F}_t^Y \subseteq \mathcal{G}_t \subseteq \mathcal{F}_t$*. Assume that a left-continuous* $\mathcal{G}_t$*-adapted version of* $\mathbb{E}[\lambda_t | \mathcal{G}_t]$ *exists. Then* $(Y_t)_{t \geq 0}$ *admits the* $\mathcal{G}_t$*-intensity* $\mathbb{E}[\lambda_t | \mathcal{G}_t]$*. In particular, if a left-continuous* $\mathcal{F}_t^Y$*-adapted version of* $\mathbb{E}[\lambda_t | \mathcal{F}_t^Y]$ *exists, then* $(Y_t)_{t \geq 0}$ *admits the* $\mathcal{F}_t^Y$*-intensity* $\mathbb{E}[\lambda_t | \mathcal{F}_t^Y]$*.*

Proof. The proof is found in [59, §2, theorem T14, p.32-33] and the remark thereafter. $\qquad \square$

Remark 2.6. If a counting process has an $\mathcal{F}_t$-intensity for some filtration $(\mathcal{F}_t)_{t \geq 0}$ and the conditions of theorem 2.5 hold, then it also admits a CI. Note, that the CI is attributed to the canonical filtration.

In this work, we consider the form $\lambda_t = \lambda(Z_t)$, which is the $\mathcal{F}_t^{Z,Y}$-intensity for $\mathcal{F}_t^{Z,Y} := \sigma(Z_s, Y_s : 0 \leq s \leq t)$. The doubly stochastic Poisson process $(Y_t)_{t \geq 0}$ is also known by the term *Cox process* in the literature. We call $(Z_t)_{t \geq 0}$ its *modulating process*. Note that λ_t is not yet the CI of $(Y_t)_{t \geq 0}$. We apply Brémaud's change of filtration theorem 2.5 to obtain that its CI is the conditional expectation $\mathbb{E}[\lambda_t | \mathcal{F}_t^Y]$.

Throughout, we only consider counting processes that admit a CI. In general, we are interested only in the case that $\hat{\lambda}_t$ is not deterministic.

Definition 2.7 (Self-exciting, history-dependent)**.** A counting process $(Y_t)_{t \geq 0}$ is **self-exciting** if for all $t > 0$, the conditional intensity $\hat{\lambda}_t$ is not almost surely a constant. The conditional intensity is **history-dependent** if $\hat{\lambda}_t$ is not $\sigma(Y_t)$-measurable for all $t > 0$.

Note that in [62, p. II.4.1], a self-exciting counting process is more weakly defined as counting process accompanied by its canonical filtration. The CI of a self-exciting counting process in general depends on a measurable function h of the trajectory $Y_{[0,t]}$ and the time t. In some models, it can be written as a function solely of Y_t. If this is not the case, we call it history-dependent.

We now consider the special case that the modulating process $(Z_t)_{t \geq 0}$ of the doubly stochastic Poisson process $(Y_t)_{t \geq 0}$ is a continuous time Markov chain on a finite state space and is time-homogeneous.

Definition 2.8 (Continuous time Markov chain, generator, Markov-modulated Poisson process)**.** Let $\mathcal{Z}$ be a finite set. By a **continuous time Markov chain (CTMC)** $(Z_t)_{t \geq 0}$ on the state space $\mathcal{Z}$ we denote a time-homogeneous Markov process with values in $\mathcal{Z}$. The **generator** $\mathcal{A} \in \mathbb{R}^{|\mathcal{Z}| \times |\mathcal{Z}|}$ of $(Z_t)_{t \geq 0}$ is given, for any $t \geq 0$, by

$$\mathcal{A}_{z,z'} = \lim_{h \to 0} h^{-1} (\mathbb{P}[Z(t+h) = z | Z(t) = z']$$
$$- \mathbb{P}[Z(t) = z | Z(t) = z']). \tag{3}$$

If $(Y_t)_{t \geq 0}$ is a doubly stochastic Poisson process modulated by a CTMC $(Z_t)_{t \geq 0}$, we call $(Y_t)_{t \geq 0}$ a **Markov-modulated Poisson process (MMPP)**.

Remark 2.9. By time-homogeneity, the definition of the generator does not depend on the choice of t.

2.1.2 Standard filtrations for CRNs in a random environment

So far, we have considered counting processes in one dimension. A multivariate counting process $(Y_t)_{t \geq 0} = ([Y_1(t), \ldots, Y_M(t)]^T)_{t \geq 0}$ has the same component-wise properties as in definition 2.1. Additionally, no two components jump simultaneously. A filtration is an increasing family of sigma-algebras, indexed by time, that, intuitively speaking, captures our knowledge at time t. Throughout, as a convention, when s and t are used, then $s \leq t$. Define the filtrations

$$\mathcal{F}_t^{YZ} := \sigma(\{Y(s) : s \leq t\} \cup \{Z(s) : s \leq t\} \cup \{X(0)\})$$
$$\mathcal{F}_t^{Y} := \sigma(\{Y(s) : s \leq t\} \cup \{X(0)\}),$$

where $\sigma(\cdot)$ denotes the smallest sigma-algebra with respect to which all random variables in the argument are measurable. Then we can make the reaction system in Eq. (1) rigorous, by specifying that $Y(t)$ has $\mathcal{F}_t^{YZ}$-intensity $\lambda(t) = [a_1(X(t-), Z(t-)), \ldots, a_M(X(t-), Z(t-))]^T$. By $t-$ we denote the left-sided limit, which makes $\lambda(t)$ a predictable intensity. Equivalently, we also use the filtration

$$\mathcal{F}_t^{YZ\infty} := \sigma(\{Y(s) : s \leq t\} \cup \{Z(s) : s \geq 0\} \cup \{X(0)\}),$$

15

as in definition 2.3. In contrast to $\mathcal{F}_t^{YZ}$, the filtration $\mathcal{F}_t^{YZ\infty}$ has knowledge of the entire trajectory of Z, also the future. Later, we consider different probability measures, which we indicate by $(\mathcal{F}_t^{YZ}, \mathbb{P})$-intensity.

2.1.3 Linear environment and the order of the environmental modulation

General results for classes of environments are difficult to obtain. However, under special conditions, results can be obtained. A well-studied class of environments are the linear stationary ones. When we refer to them, we make simplifying assumptions on the external process $Z(t) \in \mathbb{R}^l$ and the functional form $a(X(t), Z(t))$ of the $\mathcal{F}_t^{YZ}$-intensity $\lambda(t)$. We refer to these as the standard conditions. For the process $Z(t)$, we assume

$$Z(t) \text{ is weakly stationary,} \tag{C1}$$

i.e., its mean $\mathbb{E}[Z(t)]$ is a finite constant over time, and $\mathbb{Cov}[Z(t), Z(s)]$ only depends on $t - s$. Next, we assume that there exist $A, \Sigma \in \mathbb{R}^{l \times l}$ such that all eigenvalues of A have positive real part, Σ is symmetric positive semi-definite, $A\Sigma + \Sigma A^T$ is positive semi-definite and

$$\mathbb{Cov}[Z(t), Z(s)] = e^{-A(t-s)}\Sigma \quad \text{for } t \geq s \geq 0. \tag{C2}$$

For the form of $\lambda(t) = a(X(t), Z(t))$ we assume

$$a(x, z) = \mu(x) + C(x) \cdot z \in \mathbb{R}_{\geq 0}^M \tag{C3}$$

for some $\mu(x) \in \mathbb{R}^M, C(x) \in \mathbb{R}^{M \times l}$. These simplifying assumptions include all linear reaction networks $Z(t)$, i.e., $Z(t)$ is a CRN with only zeroth- and first-order reactions, for which we additionally assume stationarity and ergodicity. For instance, this class covers the random telegraph model, compartmental models with only monomolecular reactions as well as cascades of birth-death processes. Gardiner's regression theorem [63, §3.7.4b, Eq. (3.7.62), p.65] guarantees the matrix exponential form of the time correlation matrix. For processes with continuous state space, the Cox-Ingersoll-Ross process [57] satisfies these conditions. All the examples are Markov processes, see also section 6 for an illustration.

We classify the order of the environmental modulation. The order of a reaction in a CRN is the number of reactants, i.e., the sum $\sum_{i=1}^d S_{ij}$ of substrate coefficients on the left-hand side of the reaction R_j in Eq. (1). In a CRN with a modulating environment, the order of modulation is the order of the subnetwork reaction. If $a_j(x, z)$ does not depend on z, the reaction R_j is unmodulated. By zeroth-order modulation, we denote a form $a_j(x, z) = a^T z + b, a \in \mathbb{R}^l, b \in \mathbb{R}$ and by first-order modulation, we denote a form $a_j(x, z) = x^T A z + x^T b, A \in \mathbb{R}^{d \times l}, b \in \mathbb{R}^d$. To derive results, we consider another condition

$$\mu(x) \equiv \mu, \ \mu_j > 0 \text{ for all } j, \quad C(x) \equiv C, \tag{C4}$$

i.e., the propensities do not depend on the state $X(t)$. This is the case, when only zeroth-order reactions are modulated by the environment and $Z(t)$ is of mean zero. See the section 6 for examples of how the standard conditions are satisfied. For results on linear environments modulating zeroth-order reactions of the subnetwork, see sections 3.2 and 4.1.3.

2.1.4 Martingale calculus for counting processes

Associated with a multivariate counting process $Y(t) \in \mathbb{N}^d$ and its $\mathcal{F}_t$-intensity $\lambda(t)$ for some filtration $\mathcal{F}_t$ that contains $\sigma(Y(s) : s \leq t)$, there is the zero-mean $\mathcal{F}_t$-martingale

$$M(t) = Y(t) - \int_0^t \lambda(u) \, du.$$

In fact, the property that the right-hand side is an $\mathcal{F}_t$-martingale can be taken as the definition of the $\mathcal{F}_t$-intensity for non-explosive $Y(t)$. With this definition, it becomes apparent that the intensity depends on the filtration of the martingale. The increment can be written as $dM(t) = dY(t) - \lambda(t) \, dt$. One benefit of martingale theory is that it provides a rich second-order calculus.

In the following, we use the identities, see for instance [62, proposition II.4.1], for predictable processes $F(t) \in \mathbb{R}^{l_1 \times d}, G(t) \in \mathbb{R}^{l_2 \times d}$

$$\mathbb{E}\left[\int_0^t F(u) \, \mathrm{d}M(u)\right] = 0 \tag{4}$$

$$\mathbb{E}\left[\int_0^t F(u) \, \mathrm{d}M(u) \left(\int_0^s G(u) \, \mathrm{d}M(u)\right)^T\right] = \mathbb{E}\left[\int_0^{\min(s,t)} F(u) \, \mathrm{diag}(\lambda(u)) G(u)^T \, \mathrm{d}u\right]. \tag{5}$$

2.1.5 Piecewise-deterministic Markov processes

The CI of MMPPs is obtained from the so-called Snyder filter, which we identify as a piecewise-deterministic Markov process, see proposition 4.10 below. We first give the formulation of the Snyder filter in proposition 2.10. The motivating example of the pair consisting of the Snyder filter and the MMPP then prompts the investigation of more general pairs consisting of a PDMP, definition 4.7, and a corresponding counting process, definition 4.9. Since this generalization abandons the setting of the Poisson channel with Markov input, the computation of the relative entropy, Eq. (122) and Eq. (123), is then our main motivation.

In the expression of the CI for MMPPs the conditional probabilities given $\mathcal{F}_t^Y$ enter. This is to be expected by Brémaud's change of filtration theorem, see theorem 2.5. The task of finding the *posterior probabilities*, $\Pi_t(z) := \mathbb{P}[Z_t = z | \mathcal{F}_t^Y]$ for all $z \in \mathcal{Z}$, is referred to as filtering. Taken together, we call $\Pi_t(z)_{z \in \mathcal{Z}}$ the *filtering distribution*. The filtering distribution is a multi-variate stochastic process with values in $\Delta := \{\pi \in [0,1]^{|\mathcal{Z}|} : \sum_{z \in \mathcal{Z}} \pi(z) = 1\}$ and the Snyder filter provides its temporal evolution.

Proposition 2.10 (Snyder filter, conditional intensity of MMPP)**.** *In the setting and with the notation of definition 2.8 the filtering distribution $(\Pi_t(z))_{z \in \mathcal{Z}}$ evolves as follows for all $z \in \mathcal{Z}$, according to Snyder [64, Eq.(7.153), p. 396],*

$$\frac{\mathrm{d}}{\mathrm{d}t}\Pi_t(z) = (\mathcal{A}\Pi_t)(z) - (\lambda(z) - \hat{\lambda}_t)\Pi_t(z) \tag{6}$$

between jumps and is updated to [64, Eq.(7.154), p. 396]

$$\Pi_t(z) = \frac{\lambda(z)\Pi_{t-}(z)}{\hat{\lambda}_t} \tag{7}$$

if $Y_t = Y_{t-} + 1$, i.e., for $t = \sigma_i, i \in \mathbb{N}_0$. Taken together, the Snyder filter reads

$$\mathrm{d}\Pi_t(z) = (\mathcal{A}\Pi_t)(z) \, \mathrm{d}t + \frac{(\lambda(z) - \hat{\lambda}_t)\Pi_{t-}(z)}{\hat{\lambda}_t}\{ \, \mathrm{d}Y_t - \hat{\lambda}_t \, \mathrm{d}t\}. \tag{8}$$

By means of the filtering distribution, the conditional intensity is computed as

$$\hat{\lambda}_t = \sum_{z \in \mathcal{Z}} \lambda(z)\Pi_{t-}(z).$$

Proof. The proof is found in [64, theorem 7.4.2, p.383-387, example 7.4.8, p.396]. □

A structural result on the convergence for the Snyder filter is presented in theorem 4.32 and a positive lower bound for the CI is derived in theorem 4.34. The notation in Eq. (8) will be more generally introduced and explained in Eq. (83) below. Loosely speaking, the Snyder filter follows a flow in periods between jumps and is discontinuously updated upon jumps of the counting process $(Y_t)_{t \geq 0}$. The hazard with which jumps occur is the CI, which is evaluated from the filtering distribution at any time via a deterministic function. Finally, the update at jumps is also determined by the state of the filtering distribution. For the definition of the PDMP, which follows the construction of Davis [65, §24, p.57-59 & (24.8) standard conditions, p.62], see Appendix A. The three so-called local characteristics of a PDMP are

- the flow $F \colon \vartheta \to \mathbb{R}^{n_0}$ locally Lipschitz continuous,

- the hazard $l\colon \vartheta \to \mathbb{R}_{\geq 0}$ measurable satisfying an integrability condition, specified in Eq. (290), Appendix A,

- the jump update $Q\colon \mathcal{B}(\vartheta)\times\vartheta \to [0,1]$, such that $Q(B,\cdot)\colon \vartheta \to [0,1]$ is measurable for all $B \in \mathcal{B}(\vartheta)$, where $\mathcal{B}$ denotes the Borel σ-algebra, and $Q(\cdot,\theta)$ is a probability measure for all $\theta \in \vartheta$.

2.1.6 Comment on filtration-dependent intensities

From a modeling perspective, one might wonder how can $(Y_t)_{t\geq 0}$ have two intensities ($\hat{\lambda}_t$ and λ_t) and if one simulates trajectories with either of the two, how can they be on a par in terms of the same statistical properties for $(Y_t)_{t\geq 0}$? This puzzle is solved if one incorporates the $\mathcal{F}_t$-dependency of the intensity as in definition 2.1. In the above case of the Cox process $\mathcal{F}_t = \mathcal{F}_t^Y$ or $\mathcal{F}_t = \mathcal{F}_t^{Z,Y}$. This dependency can be expressed in at least three ways: (i) explicitly but not of operational value for mathematical proofs

$$\lambda_t \ \mathrm{d}t + o(\ \mathrm{d}t) = \mathbb{E}[\ \mathrm{d}Y_t|\mathcal{F}_t],$$

(ii) implicitly, and beneficial for mathematical rigidity and due to the available martingale calculus [66] by requesting

$$Y_t - \int_0^t \lambda_s \ \mathrm{d}s$$

to be an $\mathcal{F}_t$-martingale, (iii) as in the definition 2.1 by means of dual objects, which we chose to avoid the introduction of martingales. The dependency implies that a general counting process $(Y_t)_{t\geq 0}$ can admit several intensities depending on the respective sigma-algebra and for $\mathcal{G}_t \subseteq \mathcal{F}_t$, they can be linked by Brémaud's innovation theorem 2.5.

From the perspective of state estimation, $\hat{\lambda}_t$ is an estimator. The term estimation has the connotation of a deficit. However, far from being an estimator with deficits, the CI $(\hat{\lambda}_t)_{t\geq 0}$ has a standing as corresponding to the canonical filtration, i.e., it is the intensity with respect to the minimal filtration to which $(Y_t)_{t\geq 0}$ is adapted, compare remark 2.6. This is why it can be seen as the canonical one among the $\mathcal{F}_t^Y$- and $\mathcal{F}_t^{Z,Y}$-intensities. The $\mathcal{F}_t^Y$-intensity can be defined for any counting process $(Y_t)_{t\geq 0}$ admitting some intensity, with no need to specify an external process, making it a universal object for counting processes that admit an intensity.

The estimator perspective regards $\mathbb{E}[\lambda_t|\mathcal{F}_t^Y]$ as a function of the history $\{Y_s\colon 0 \leq s \leq t\}$. This complicated functional dependence makes it hard to understand the limit object $\hat{\lambda}_\infty$ of the convergence $\hat{\lambda}_t \to \hat{\lambda}_\infty$ in distribution. In contrast, the conditional intensity view can offer a self-contained description of $\hat{\lambda}_t$ via the Snyder filter, Eq. (8): all the knowledge on $(Y_t)_{t\geq 0}$ needed for understanding $\hat{\lambda}_\infty$ is captured in the jump times of $(\hat{\lambda}_t)_{t\geq 0}$ which are shared with $(Y_t)_{t\geq 0}$. In this way, $(\hat{\lambda}_t)_{t\geq 0}$ emancipates from $(Y_t)_{t\geq 0}$.

For this emancipation it was crucial to note that we can consider $(Y_t)_{t\geq 0}$ to be jumping with intensity $(\hat{\lambda}_t)_{t\geq 0}$ in Eq. (8). In contrast, from a modeling perspective, $(Y_t)_{t\geq 0}$ is often viewed as simulated from a two-step randomization procedure [59, p.21]. First, trajectories of $(Z_t)_{t\geq 0}$ are sampled and secondly, conditional on the trajectory, a non-homogeneous Poisson process trajectory $(Y_t)_{t\geq 0}$ is sampled with intensity $(\lambda(Z_t))_{t\geq 0}$. Then Eq. (8) can be evaluated trajectory-wise. This perspective, however, dilutes the Markov property of $(\Pi_t)_{t\geq 0}$, for which the hazard (the second local characteristic of the PMDP) must be a function l of the state Π_{t-}. The intensity $\lambda(Z_t)$ can, however, not be evaluated from the state Π_{t-} alone. So the part $\mathrm{d}Y_t$ in Eq. (8) uses external knowledge and the equation is not self-contained. The CI $(\hat{\lambda}_t)_{t\geq 0}$ in place of $(\lambda(Z_t))_{t\geq 0}$ fulfills the requirement for l and establishes the Markov property. In the modeling perspective this history-dependent intensity can be realized by keeping track of the ODE (6) between jumps and using a thinning algorithm for $(Y_t)_{t\geq 0}$ or the inverse sampling method for the sojourn times, as well as updating via Eq. (7) upon the occurrence of jumps.

2.2 Motivation

It is one modeling goal to find a closed, i.e., self-contained, description for a given CRN in a random environment that behaves in its statistical properties as if it was still embedded. We call it the marginal description. The goal are both evolution equations for deterministic quantities, i.e., the probability evolution and evolution of moments, as well as process equations, that can be used for stochastic

simulation. Probability and moment evolution equations describe the average behavior of the system, stochastic simulations generate random samples. The descriptive and generative tasks of the marginal description are complemented by the analytic task of attributing subnetwork features to environment features, as well as estimation of information measures. In summary, the various purposes for studying the CRNs in a random environment via stochastic conditioning are (i) marginal simulation, (ii) model reduction, (iii) the computation of information measures, (iv) state estimation and (v) attribution of effects to properties of the random environment.

2.2.1 Marginal simulation

The work by Zechner and Koeppl [3] proposed the marginal simulation algorithm that bypasses the co-simulation of environment components, emphasizing its potential for efficient simulation and variance reduction [67] as known from Rao-Blackwellized estimators [68, 69]. The marginal simulation approach replaces the modulating component of a two-component process to obtain a one-component description of the modulated component. The analysis of marginal simulation thus provides the foundation for extending general gene regulatory networks by sources of heterogeneity. Addressing the problems of simulating gene regulatory networks and the modular design of synthetic circuits, we provide an efficient way of including heterogeneity. For the latter purpose, our method may help with *in silico* studies, i.e., simulations, that precede *in vitro* studies. One ultimate goal of understanding each gene expression step in gene regulatory networks is the engineering and synthesis of genetic circuits [70]. To design complex genetic circuits that function despite their embedding into a cellular context, it is essential to account for identified sources of heterogeneity in a modular manner [71, 72]. The typical workflow of circuit design involves the composition of networks or cascades from well-characterized parts [73, 74]. Adhering to the Design-Build-Test-Learn cycle, it is common to simulate the composed function *in silico* before conducting *in vitro* or *in vivo* experiments [75, 76]. Rational design thus requires principled models that besides reliably predicting the behavior are also modular, in the sense that it consists of single modules that can be easily replaced.

2.2.2 Model reduction

Zechner, Bronstein and Koeppl introduced the marginal process framework [77, 78] to reduce Markov jump models. The reduced models specify new process equations and evolution equations for the probability and moments. Recovering the Markov property, new process equations on an augmented state space can be used to derive generalized (auxiliary-variable) [79] or reduced [80] master equations. The reduced model with a marginal description served the purpose of genetic circuit design [81], realizing the auxiliary variables as chemical species.

2.2.3 Information theory

Beyond the marginal description, the filter mean naturally appears in expressions for information measures. Shamai and Lapidoth used the optimal linear causal estimate to obtain upper bounds on the capacity for the spectrally constrained Poisson channel [82]. Duso, Moor and Zechner used the Gamma filter [83, 84] to estimate the mutual information between two species in a CRN. Furthermore, it has been known since the work on counting processes by Brémaud [59, §VI.6] that the relative entropy between two counting processes is a difference functional of the filter means ('detection formula'). Atar and Weissman used it to derive the error of mismatched estimation of the mutual information, when using an approximate filter [85].

2.2.4 State estimation

We emphasize that the original purpose for the study of the filter mean was state estimation under partial observability [86, 59]. In the literature on CRNs, this approach has been recently rediscovered. In their works, Rathinam et al. and Fang et al. introduce particle filters to solve the filtering problem for latent state estimation [87, 88]. State estimation for CRNs via smoothing, i.e., under the (partial) knowledge of the trajectory until the final time point, is addressed in [89]. The causal, anti-causal and non-causal estimator of the intensity, corresponding to forward filtering, backward filtering and smoothing are usually closely related, as in [90] for the Poisson process. Along with state estimation,
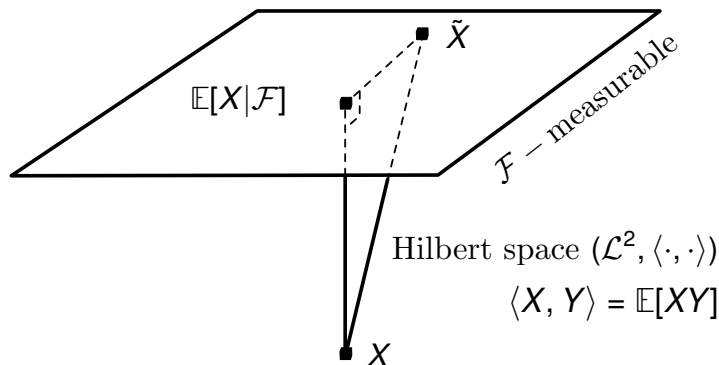
Figure 2: Schematic graph of the conditional expectation with respect to a sigma-algebra $\mathcal{F}$. Projection property and Pythagorean identity is shown in the Hilbert space of square-integrable random variables. The plane indicates the linear subspace of $\mathcal{F}$-measurable random variables.

one of the earliest application was in optimal control [59, §VII] sparked by the initial success of Kalman [91]. A more recent application designed genetic circuits using approximate filters [81].

2.2.5 Attribution to the environment

In addition, approximate and reduced models can be used as baseline models. Deviations from the reference give insight, how subnetwork properties can be attributed to the environment, similar to knock-out experiments. They provide hypotheses about which features of the environment do or do not evoke an observed stochastic property in the subnetwork. More generally, throughout we were guided by the goal of quantifying the effect that the environment has on the moments of the subnetwork.

3 Conditioning on the environment

As a rather general strategy for the aforementioned tasks, we employ stochastic conditioning, i.e., the conditional expectation. Two routes can generally be taken: conditioning on the environment or conditioning on the subnetwork. These approaches differ in that the conditioning on the environment can only generate evolution equations for deterministic quantities, i.e., the probability evolution and evolution of moments. The conditioning on the embedded subnetwork provides process equations in addition, that can be used for marginal simulation.

The conditional expectation is a standard concept from probability theory and we use it here as our main tool to reduce the CRN model, Eq. (1) in a random environment. The conditional expectation of a random variable X with respect to a sigma-algebra $\mathcal{F}$, $\mathbb{E}[X|\mathcal{F}]$ is an $\mathcal{F}$-measurable random variable $\tilde{X}$ with the defining property that

$$\int_A X \, d\mathbb{P} = \int_A \tilde{X} \, d\mathbb{P}$$

for all $A \in \mathcal{F}$. It is unique defined up to changes on a set of measure 0. When $\mathcal{F} = \sigma(Y)$ for another random variable Y, then we write $\mathbb{E}[X|Y]$ instead of $\mathbb{E}[X|\sigma(Y)]$. The conditional expectation satisfies the projection property

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{F}] = \mathbb{E}[X|\mathcal{F}].$$

And more generally the tower property, i.e., for $\mathcal{G} \subseteq \mathcal{F}$

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]|\mathcal{G}] = \mathbb{E}[X|\mathcal{G}],$$

which says that conditioning can be done sequentially starting from the finer, and proceeding to the coarser sigma-algebra. This is especially useful for the particular case of $\mathcal{G} = \{\emptyset, \Omega\}$, telling us

$$\mathbb{E}[\mathbb{E}[X|\mathcal{F}]] = \mathbb{E}[X],$$

i.e., to evaluate an expectation by first conditioning on $\mathcal{F}$ as an intermediary before we take expectations. We can think of $\mathbb{E}[X|Y]$ as a measurable function $g(Y)$ of Y. So we say, that conditioning on Y eliminates the randomness in X, while keeping the randomness in Y. Also it holds, that

$$\mathbb{E}[aX + b\tilde{X}|\mathcal{F}] = a\mathbb{E}[X|\mathcal{F}] + b\mathbb{E}[\tilde{X}|\mathcal{F}],$$

which makes $\mathbb{E}[\cdot|\mathcal{F}]$ a linear projection. If X is square-integrable, then the projection has a stronger geometric meaning in the Hilbert space of square-integrable random variables with the scalar product $\langle X, Y \rangle = \mathbb{E}[XY]$, see figure 2. Namely, $\mathbb{E}[X|\mathcal{F}]$ minimizes $\mathbb{E}[(\tilde{X}-X)^2]$ among all $\mathcal{F}$-measurable random variables $\tilde{X}$. This is seen by proving the Pythagorean identity

$$\mathbb{E}[(\mathbb{E}[X|\mathcal{F}] - \tilde{X})^2] + \mathbb{E}[(\mathbb{E}[X|\mathcal{F}] - X)^2] = \mathbb{E}[(\tilde{X} - X)^2].$$

Hence, $\mathbb{E}[\cdot|\mathcal{F}]$ is an orthogonal linear projection onto the linear subspace of $\mathcal{F}$-measurable random variables.

We now apply this to our case. We started out with the joined system (X_t, Z_t), where X_t is the subsystem state and Z_t is the environment state. When we are interested in the moment of an observable $\mathbb{E}[h(X_t)]$ of X_t, we have two paths that we can take: first eliminating the randomness in X_t or first eliminating the randomness in Z_t. The remaining randomness is subsequently removed in a second step. We cover the elimination of the randomness in X_t here. For the other path, see section 4. In particular, we condition on

$$\mathcal{F}_t^Z := \sigma(\{Z(s) : s \leq t\})$$

or sometimes

$$\mathcal{F}_\infty^Z := \sigma(\{Z(s) : s \geq 0\}).$$

The second variant corresponds to sampling the entire trajectory of Z, such that we can condition on it. Then $\mathbb{E}[h(X_t)]$ is computed as $\mathbb{E}[\mathbb{E}[h(X_t)|\mathcal{F}_t^Z]]$ or $\mathbb{E}[\mathbb{E}[h(X_t)|\mathcal{F}_\infty^Z]]$. Commonly, h can also be chosen as an indicator function to cover probabilities $\mathbb{P}[X_t \in A]$ and $\mathbb{P}[X_t = k]$ within this framework.

3.1 First-order moments

We first investigate the effect of the random environment on the mean of a subsystem species. This effect is visible when first-order reactions are modulated as the following simple example illustrates. Consider a birth-death process with birth rate λ and death rate μ whose mean approaches λ/μ in the equilibrium. Zero-order modulation corresponds to a modulation in the birth rate, i.e., stochastic λ with mean $\bar{\lambda}$. By the linearity of the function $\lambda \mapsto \lambda/\mu$ and the linearity of the mean, the heterogeneous mean $\bar{\lambda}/\mu$ does not reflect the heterogeneity at the mean level. However, if μ is stochastic, the non-linearity of $\mu \mapsto \lambda/\mu$ invalidates this argument. Instead, the excursions in the abundance that emerge during $\mu = 0$ can distort the heterogeneous mean away from $\lambda/\bar{\mu}$ (Fig. 3b). In the experimental context, this means that even bulk data carries the fingerprint of a heterogeneous environment.

The simple birth-death process in a random environment has been studied in queuing theory. It corresponds to an $M/M/\infty$ queue, i.e., a queuing system with infinitely many servers and exponential arrival and serving times. The birth rate is the rate of arrival, while the death rate corresponds to the service rate. For a Markov environment, the stationary distribution, stationary factorial moments, and the transient evolution equation of the moments were derived in [50]. We extend the existing results in queuing theory to more reaction channels than the birth-death channel by considering the general class of linear reaction systems whose reaction rates are modulated by a discrete state Markov environment. As our main contribution, we analytically express the stationary mean for this class of CRNs in terms of the subsystem rate constants, the environment generator matrix and the environment equilibrium distribution. It also provides an exact alternative, at least on the mean level, to various approximations that have been used to understand the effect of extrinsic noise on the mean, variance, power spectrum and distribution. Furthermore, we propose a new method that quantifies the shares that the environment states contribute to the stationary mean of a subsystem species. In section 8.3, we analyze the deviation from the Q.SS behavior in a variety of case studies to illustrate non-linear effects of the random environment.

21
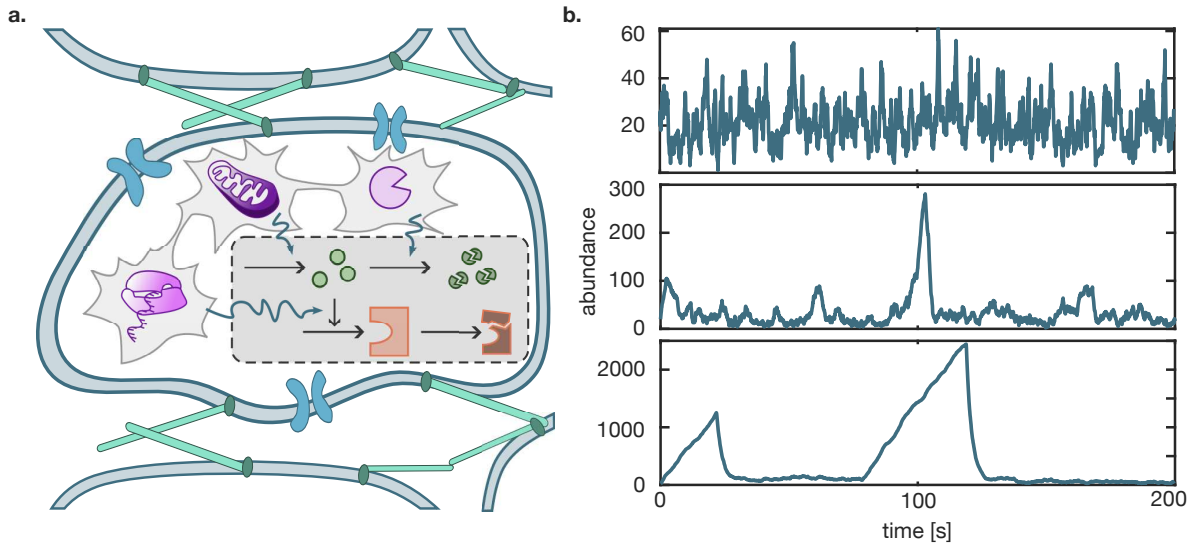
Figure 3: **a.** The cartoon shows an example subsystem of a reaction cascade (dashed box) that is embedded into a cellular environment that includes RNA polymerase activity, ATP availability, and the concentration of a degradation enzyme. Purple environment components from left to right symbolize the RNA polymerase that modulates transcription, mitochondria that supply energy in form of ATP and proteases that actively degrade proteins. More broadly, the embedding of the single cell in a tissue and communication with neighboring cells via the extracellular matrix (light green) or ion channels (blue) can contribute stochasticity. The reaction cascade consists of a sensor molecule (green) and an enzyme (orange) whose expression is mediated by the sensor molecule. **b.** Trajectories of a birth-death-process with heterogeneous death rate μ, modeled as a birth-death process on different timescales. On the fast timescale, the trajectory resembles an unmodulated birth-death process (upper panel). On the intermediate timescale, triangular excursions appear while $\mu = 0$ (middle panel). On the slow timescale, the excursions become more pronounced (lower panel). The mean death rate is the same for all cases.

3.1.1 Setting of embedded linear CRNs

Linear reaction networks are a class of CRNs whose propensities depend linearly on the state of the system, i.e., the vector of species copy numbers. Assuming mass-action kinetics, this restriction corresponds to admitting only first- and zeroth-order reactions. For this class, the first- and second order moment dynamics are well-known both in the transient and stationary phase [92]. In particular, the stationary mean of the stochastic system equals the equilibrium concentration of the corresponding deterministic system [53].

Homogeneous linear CRN

Consider a *linear* CRN with the vector $X_t \in \mathbb{N}^d$ of species counts, namely, a CRN where the vector of propensities for its M reactions is of the form

$$a : \mathbb{N}^d \to \mathbb{R}_{\geq 0}^M, a(x) = \Gamma x + \gamma. \tag{9}$$

If all reactions follow mass-action kinetics, this implies that only zeroth or first order reactions are admissible [92]. Then Γx and γ are the vectors with the propensities of the first-order and the zero-order reactions, respectively. Next, we construct the stoichiometric matrix that stores the change vectors for each reaction:

$$N = \begin{bmatrix} | & & | \\ \nu_1 & \dots & \nu_M \\ | & & | \end{bmatrix} \in \mathbb{Z}^{d \times M}$$

via $N_{ij} = P_{ij} - S_{ij}$ and $\nu_j = P_{\cdot j} - S_{\cdot j}$. This allows to write down the evolution of the mean:

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[X_t] = \mathbb{E}[\sum_{j=1}^M \nu_j a_j(X_t)] = Na(\mathbb{E}[X_t]) = N\Gamma \mathbb{E}[X_t] + N\gamma =: b - A\mathbb{E}[X_t]. \tag{10}$$

The matrix $A = -N\Gamma$ captures the first-order reaction rates, while the vector $b = N\gamma$ accounts for the zero-order reaction rates. Throughout section 3.1, we assume that in the homogeneous case A has only eigenvalues with positive real part. This property is called Hurwitz-stable, and it guarantees the ergodicity of the system, in particular the existence of and convergence to the stationary mean [93].

Linear CRN in a random environment

We have just described the homogeneous case. Next, we embed the linear chemical reaction network in a random environment. We assume that Z is a stationary continuous-time Markov chain (CTMC) on a discrete state space $\mathcal{Z}$. The environment modulates X by Z-dependent propensities which replace Eq. (9) by

$$a(x, z) = \Gamma(z)x + \gamma(z) \tag{11}$$

with corresponding family of matrices $A(z) = -N\Gamma(z)$ and vectors $b(z) = N\gamma(z)$ as defined in Eq. (10). We refer to such a linear CRN in a random environment also as a heterogeneous, or modulated, linear CRN. Throughout section 3.1, we assume that $A(z)$ has only eigenvalues with non-negative real part for each $z \in \mathcal{Z}$. In contrast to the homogeneous case, 0 may be included as an eigenvalue. This assumption is needed for the proof of proposition 3.5 and for the evaluation in Eq. (29) to prove the main theorem 3.7. Both proofs are based on the lemma 3.4 below that uses the assumption.

Most prominently, Z can be a CRN on the state space of (environmental) species counts $\mathcal{Z} \subseteq \mathbb{N}^m$ [94, 3]. In this case, we call X the *subsystem* of the joint reaction network (X, Z). The equation that governs the mean now reads

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[X_t] = \mathbb{E}[b(Z_t)] - \mathbb{E}[A(Z_t)X_t], \tag{12}$$

involving non-linear modulation terms $\mathbb{E}[A(Z_t)X_t]$.

Quasi-steady state reference model

With a Q.SS assumption on the environment, we aim to map a heterogeneous linear CRN to a homogeneous one. To this end, we replace the subsystem propensities $a(x|z)$ by $\bar{a}(x) = \mathbb{E}[\gamma(z)] + \mathbb{E}[\Gamma(z)]x$. Even though the Q.SS method has seen applications with much less coarse approximations in stochastic models, i.e., at the level of the master equation [10], we use the term in the described way at the mean level. The corresponding first-order rate matrix and zero-order rate vector, see Eq.

(10), are $\bar{A} = -N\mathbb{E}[\Gamma(z)], \bar{b} = N\mathbb{E}[\gamma(z)]$. In the case of $\mathcal{Z} \subseteq \mathbb{R}^m_{\geq 0}$ and if Γ or γ depend linearly on z, then $\bar{A} = -N\Gamma(\bar{z})$ or $\bar{b} = N\gamma(\bar{z})$, respectively. The mean Eq. (12) reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t] = \bar{b} - \bar{A}\mathbb{E}[X_t] \tag{13}$$

and converges to the equilibrium

$$\mathbb{E}[X^{Q.SS}_\infty] = \bar{A}^{-1}\bar{b}. \tag{14}$$

We use the Q.SS model as a reference model. The deviation of the heterogeneous from the homogeneous reference model quantifies the effect of the random environment on the subsystem. Quasi-steady state model reduction is typically justified by the assumption that the environment operates on a faster timescale than the subsystem. This particular rather coarse Q.SS model cannot be justified as a model reduction technique, when taking into consideration that more elaborated and superior techniques [10, 11, 95] were developed. However, we justify our Q.SS in the reverse perspective. In the modeling context, when we build a hierarchy of models, we may start with homogeneous reaction rates, i.e., a deterministic constant environment. As we pass to heterogeneous rates, the rate means are kept constant, e.g., if the environment means were obtained from separate measurements.

Then the heterogeneous case deviates from the homogeneous Q.SS case. Is it necessary to include a higher-order moment analysis in order to detect the deviation? We demonstrate how the fingerprint of the heterogeneous rates emerges already in the stationary mean of the subsystem.

As a point of departure, we first portray a base case without deviation at the mean level, i.e., the stationary mean of the Q.SS model coincides with the heterogeneous stationary mean. Suppose, the environment modulates the subsystem only via zero-order reactions, i.e. $A(z) \equiv \bar{A} = A$ is independent of the environment, while $b(z)$ may maintain dependencies. Then Eq. (12) reduces to

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t] = \mathbb{E}[b(z)] - A\mathbb{E}[X_t]$$

with equilibrium state $\mathbb{E}[X_\infty] = A^{-1}\mathbb{E}[b(z)] = \mathbb{E}[X^{Q.SS}_\infty]$. Only in the second (and higher) order moments the exact network deviates from the Q.SS model [48]. The additional term which enters in the variance expressions is commonly interpreted as extrinsic noise, opposed to the intrinsic fluctuations that are attributed to the subsystem alone [47].

While the stationary mean is not affected by the zero-order modulation, this is much different when we allow modulation of first-order reactions. The question how the random environment affects the subsystem even on the mean level guides the sections 3.1.2-3.1.4 and the case study 8.3. Since the analytic expression for the stationary mean of the heterogeneous system that we provide in Eq. (18), (21) and (22) below is complex, we mainly address this question numerically in the case studies, section 8.3. However, in the case of a birth-death process with modulated degradation, an analytical answer can be obtained. We provide it in Eq. (229) below for a two-state environment and more generally in subsection 8.3.4.

3.1.2 Bye bye linearity bye: exact stationary mean evaluation (ESME)

We analytically express the stationary mean for the class of CRNs introduced in the previous section. The expressions only depend on the subsystem rate constants and standard characteristics of the Markov environment that we specify as follows. Denote by $\Lambda(z', z)_{z',z\in\mathcal{Z}}$ its generator and by $\pi(z)$ its stationary distribution, i.e., $\Lambda\pi = 0$. Introduce the notation $\Lambda_0(z) := -\Lambda(z, z) = \sum_{z'\neq z}\Lambda(z', z)$ for the total exit rate. Let $(\tau_n)_{n\in\mathbb{N}}$ be the jump times of Z. These induce the discrete-time Markov chain $(W_n)_n := (Z(\tau_n))_n$ with transition kernel

$$K(z', z) = \begin{cases} \frac{\Lambda(z',z)}{\Lambda_0(z)}, & z' \neq z \\ 0, & z' = z \end{cases}.$$

The process W_n is called embedded discrete-time Markov chain that corresponds to the jump epochs of $Z(t)$ [96].

Proposition 3.1. *Let W be the embedded chain of a Markov jump process on $\mathcal{Z}$ with stationary distribution $\pi(z)$ and total exit rates $\Lambda_0(z)$, $z \in \mathcal{Z}$. Define $\tilde{\pi}(z) := \pi(z)\Lambda_0(z)$. Then $\tilde{\pi}$ satisfies the stationarity condition for the embedded chain, i.e., is an unnormalized stationary distribution of W.*

Proof. We verify the stationarity condition for the embedded chain.

$$\sum_{z \in \mathcal{Z}} K(z', z)\tilde{\pi}(z) = \sum_{z \neq z'} \frac{\Lambda(z', z)}{\Lambda_0(z)} \pi(z)\Lambda(z) = -\Lambda(z', z')\pi(z') = \Lambda_0(z')\pi(z') = \tilde{\pi}(z').$$

$\square$

For the computation of the stationary mean, we use the average values of the process X at the end of intervals $[\tau_n, \tau_{n+1}]$ in environmental state z. The computation of these average values is facilitated by the fact that, conditional on the environment, the subsystem expectation progresses linearly.

Definition 3.2. Define $x(n, z) := \mathbb{E}[X(\tau_{n+1})|Z(\tau_n) = z]$ and $V(t) := \mathbb{E}[X(t)|Z_{[0,t]}]$.

Conditional on the history of Z, the subsystem X is linear and hence V evolves like

$$\frac{\mathrm{d}}{\mathrm{d}t}V(t) = b(Z(\tau_n)) - A(Z(\tau_n))V(t), \quad \tau_n < t < \tau_{n+1} \tag{15}$$

and $t \mapsto V(t)$ is continuous. We note that the value $Z(\tau_n)$ is the only information in the history of the environment that governs the evolution (locally in time). For the value $V(t)$ global in time, the initial jump value $V(\tau_n)$ is needed, which depends in a more complex way on past values of Z. This is solved interval-wise by

$$V(t) = e^{-A(z)(t-t_0)}V(t_0) + \int_{t_0}^{t} e^{-A(z)(s-t_0)} \,\mathrm{d}s \, b(z). \tag{16}$$

By the update characterization of Markov chains, we can find a sequence $(\xi_n)_{n \geq 0}$ of independent random and a deterministic function h, such that ξ_n is independent of $Z(\tau_n)$ and $Z(\tau_{n+1}) = h(Z(\tau_n), \xi_n), n \geq 0$. Furthermore, the $(\xi_n)_{n \geq 0}$ can be chosen independent of $(\tau_{n+1} - \tau_n)_{n \geq 0}$. Choosing $t_0 = \tau_n$ and $t = \tau_{n+1}$ in Eq. (16), we expressed

$$(V(\tau_{n+1}), Z(\tau_{n+1})) = g(V(\tau_n), Z(\tau_n), \tau_{n+1} - \tau_n, \xi_n) \tag{17}$$

for a deterministic update function g. By the update characterization of Markov chains, the independent waiting times $(\tau_{n+1} - \tau_n)_n$ and independence of $\tau_{n+1} - \tau_n$ from $(V(\tau_n), Z(\tau_n))$, we get that the pair $(V(\tau_n), Z(\tau_n))_n$ forms a Markov chain. Furthermore, by the tower property of conditional expectations, we obtain the following link between x and V

$$x(n, z) = \mathbb{E}[V(\tau_{n+1})|Z(\tau_n) = z].$$

At stationarity, $x(n, z)$ does not depend on n anymore, because $(V(\tau_{n+1}), Z(\tau_n))$ has the same distribution as $(V(\tau_n), Z(\tau_{n-1}))$. Hence, we can drop n. More formally:

Definition 3.3. Define $(\tilde{V}_n, \tilde{Z}_n)_n$ as the stationary version of the Markov chain $(V(\tau_n), Z(\tau_n))_n$ and $\tilde{x}(n, z) := \mathbb{E}[\tilde{V}_{n+1}|\tilde{Z}_n = z]$. Define $x(z) := \tilde{x}(1, z)$.

The expressions in Eq. (16) with the evaluation in Eq. (17) suggest calculating the following mean values.

Lemma 3.4. *Let T be exponentially distributed with parameter $\mu > 0$ and suppose that all eigenvalues of $\mathbf{A}$ have non-negative real part. Then $\mu\mathbf{I} + \mathbf{A}$ is invertible and*

(i)

$$\mathbb{E}[e^{-\mathbf{A}T}] = \mu(\mu\mathbf{I} + \mathbf{A})^{-1}$$

(ii)

$$\mathbb{E}[\int_0^T e^{-\mathbf{A}t} \,\mathrm{d}t] = (\mu\mathbf{I} + \mathbf{A})^{-1}$$

(iii)

$$\mathbb{E}[\int_0^T \int_0^t e^{-\mathbf{A}s}\ \mathrm{d}s\ \mathrm{d}t] = \mu^{-1}(\mu\mathbf{I}+\mathbf{A})^{-1}$$

Proof of the lemma.

$$\mathbb{E}[e^{-\mathbf{A}T}] = \int_0^\infty \mu e^{-\mu t}e^{-\mathbf{A}t}\ \mathrm{d}t = \mu(\mu\mathbf{I}+\mathbf{A})^{-1}$$

$$\mathbb{E}[\int_0^T e^{-\mathbf{A}t}\ \mathrm{d}t] = \int_0^\infty \mu e^{-\mu\tau} \int_0^\tau e^{-\mathbf{A}t}\ \mathrm{d}t\ \mathrm{d}\tau = \int_0^\infty e^{-\mathbf{A}t} \int_t^\infty \mu e^{-\mu\tau}\ \mathrm{d}\tau\ \mathrm{d}t$$

$$= \int_0^\infty e^{-\mathbf{A}t}e^{-\mu t}\ \mathrm{d}s = (\mu\mathbf{I}+\mathbf{A})^{-1}$$

$$\mathbb{E}[\int_0^T \int_0^t e^{-\mathbf{A}s}\ \mathrm{d}s\ \mathrm{d}t] = \int_0^\infty \mu e^{-\mu\tau} \int_0^\tau \int_0^t e^{-\mathbf{A}s}\ \mathrm{d}s\ \mathrm{d}t\ \mathrm{d}\tau$$

$$= \int_0^\infty e^{-\mathbf{A}s} \int_s^\infty \int_t^\infty \mu e^{-\mu\tau}\ \mathrm{d}\tau\ \mathrm{d}t\ \mathrm{d}s$$

$$= \int_0^\infty e^{-\mathbf{A}s} \int_s^\infty e^{-\mu t}\ \mathrm{d}t\ \mathrm{d}s = \int_0^\infty e^{-\mathbf{A}s}\mu^{-1}e^{-\mu s}\ \mathrm{d}s$$

$$= \mu^{-1}(\mu\mathbf{I}+\mathbf{A})^{-1}$$

$\square$

Proposition 3.5. *The $x(z)$, defined in definition 3.3, satisfy the linear equations*

$$A(z)x(z) = \sum_{z'\in\mathcal{Z}} \Lambda(z,z')\frac{\pi(z')}{\pi(z)}x(z') + b(z). \tag{18}$$

Suppose that $\pi(z)$ satisfies detailed balance, then

$$A(z)x(z) = (\Lambda^T x)(z) + b(z). \tag{19}$$

Proof.

$$x(z) = \mathbb{E}[V(\tau_{n+1})|Z(\tau_n) = z]$$

$$= \mathbb{E}[e^{-A(z)(\tau_{n+1}-\tau_n)}|Z(\tau_n) = z]\mathbb{E}[V(\tau_n)|Z(\tau_n) = z]$$

$$+ \mathbb{E}[\int_0^{\tau_{n+1}-\tau_n} e^{-A(z)t}\ \mathrm{d}t|Z(\tau_n) = z]b(z)$$

$$= \Lambda_0(z)(\Lambda_0(z)\mathbf{I}+A(z))^{-1} \sum_{z'\neq z} \mathbb{P}[W_{n-1} = z'|W_n = z]x(z') + (\Lambda_0(z)\mathbf{I}$$

$$+ A(z))^{-1}b(z)$$

$$= \Lambda_0(z)(\Lambda_0(z)\mathbf{I}+A(z))^{-1} \sum_{z'\neq z} K(z,z')\frac{\tilde{\pi}(z')}{\tilde{\pi}(z)}x(z') + (\Lambda_0(z)\mathbf{I}+A(z))^{-1}b(z)$$

$$= \Lambda_0(z)(\Lambda_0(z)\mathbf{I}+A(z))^{-1} \sum_{z'\neq z} \Lambda(z,z')\frac{\pi(z')}{\pi(z)\Lambda_0(z)}x(z') + (\Lambda_0(z)\mathbf{I}+A(z))^{-1}b(z)$$

$$= (\Lambda_0(z)\mathbf{I}+A(z))^{-1} \sum_{z'\neq z} \Lambda(z,z')\frac{\pi(z')}{\pi(z)}x(z') + (\Lambda_0(z)\mathbf{I}+A(z))^{-1}b(z)$$

Hence,

$$(\Lambda_0(z)\mathbf{I}+A(z))x(z) = \sum_{z'\neq z} \Lambda(z,z')\frac{\pi(z')}{\pi(z)}x(z') + b(z),$$

or upon adding of $\Lambda(z,z)x(z)$:

$$A(z)x(z) = \sum_{z' \in \mathcal{Z}} \Lambda(z,z') \frac{\pi(z')}{\pi(z)} x(z') + b(z).$$

If detailed balance hold, we substitute

$$\Lambda(z,z') \frac{\pi(z')}{\pi(z)} = \Lambda(z',z),$$

yielding

$$A(z)x(z) = \sum_{z' \in \mathcal{Z}} \Lambda^T(z,z')x(z') + b(z).$$

$\square$

The structure of Eq. (18) can be captured graphically. Let us visualize the recursion by a directed graph on the set of nodes $\mathcal{Z}$. There is an edge from z' to z, if $x(z')$ has a non-zero coefficient for the equation of $x(z)$. Then this graph is precisely the transition graph of the Markov chain Z. In particular the sparsity of Eq. (18) is dictated by the sparsity of Λ.

Remark 3.6. Define $v(z) := \pi(z)x(z)$ and denote by $\mathbf{v}^T = [v(z_0)^T, v(z_1)^T, \ldots]$ the concatenated vector for an enumeration of the environment state. Further define the block matrices

$$\mathbf{A} = \begin{bmatrix} A(z_0) & \mathbf{0} & \cdots \\ \mathbf{0} & A(z_1) & \\ \vdots & & \ddots \end{bmatrix} \in \mathbb{R}^{|\mathcal{Z}|d \times |\mathcal{Z}|d}, \mathbf{B} = \begin{bmatrix} b(z_0) & \mathbf{0} & \cdots \\ \mathbf{0} & b(z_1) & \\ \vdots & & \ddots \end{bmatrix} \in \mathbb{R}^{|\mathcal{Z}|d \times |\mathcal{Z}|}.$$

Then Eq. (18) can be written as

$$(\mathbf{A} - \Lambda \otimes \mathbf{I}_d)\mathbf{v} = \mathbf{B}\pi \tag{20}$$

where $\mathbf{I}_d$ is the $(d \times d)$ identity matrix, $\otimes$ is the Kronecker- or tensor-product and $\pi \in \mathbb{R}^{|\mathcal{Z}| \times 1}$ is the stationary probability vector. Denote by $\mathbf{x}^T = [x(z_0)^T, x(z_1)^T, \ldots]$ the concatenated vector for an enumeration of the environment state. Furthermore consider $\mathbf{b} = \mathbf{B}e$ with $e \in \mathbb{R}^{|\mathcal{Z}| \times 1}$ the vector with ones in all entries. Then Eq. (19) can be written as

$$(\mathbf{A} - \Lambda^T \otimes \mathbf{I}_d)\mathbf{x} = \mathbf{b}.$$

We computed the average values of the process X at the end of intervals $[\tau_n, \tau_{n+1}]$ with environmental 'label' z. With these auxiliary quantities, we obtain the following main result for the stationary mean.

Theorem 3.7 (ESME). *Let X be a linear CRN in a random environment as described in section 3.1.1. With x defined as in definition (3.3) and π the stationary distribution of Z, it holds*

$$\mathbb{E}[X_\infty] = \sum_{z \in \mathcal{Z}} \pi(z)x(z) \tag{21}$$

Proof. See section 3.1.3 below. $\square$

Remark 3.8. If we invest equation (20), the resulting equation (21) for ESME is rewritten in matrix-vector notation

$$\mathbb{E}[X_\infty] = (e^T \otimes \mathbf{I}_d)(\mathbf{A} - \Lambda \otimes \mathbf{I}_d)^{-1}\mathbf{B}\pi \tag{22}$$

with $\otimes$ the Kronecker-product, $\mathbf{I}_d$ the $d \times d$-identity matrix and $e \in \mathbb{R}^{|\mathcal{Z}| \times 1}$ the vector with ones in all entries. For $d = 1$, the expression reduces to

$$e^T(\mathbf{A} - \Lambda)^{-1}\mathbf{B}\pi$$

for diagonal matrices $\mathbf{A}, \mathbf{B}$. This is in agreement with the expression obtained in [50, theorem 3.1]. If $\pi(z)$ satisfies detailed balance, then it also holds

$$\mathbb{E}[X_\infty] = (\pi^T \otimes \mathbf{I}_d)(\mathbf{A} - \Lambda^T \otimes \mathbf{I}_d)^{-1}\mathbf{b}. \tag{23}$$

3.1.3 Proof of the ESME expression

In this section, we provide the proof for the main theorem 3.7 (ESME). This equips the reader with the intuition on quantifying the shares that environmental states contribute to the stationary mean. We define these in the next section. Note that our proof deviates from those in queuing theory in that it does not pursue a generating function approach.

By the ergodic theorem, we can move from the ensemble mean to the temporal mean

$$\mathbb{E}[X_\infty] = \lim_{N\to\infty} \frac{1}{\tau_N} \int_0^{\tau_N} X(t) \, dt. \tag{24}$$

Next, we partition the time axis at the jump times τ_n both in the numerator and denominator to obtain

$$\mathbb{E}[X_\infty] = \lim_{N\to\infty} \frac{1}{\sum_{n=0}^{N-1}(\tau_{n+1}-\tau_n)} \sum_{n=0}^{N-1} \int_{\tau_n}^{\tau_{n+1}} X(t) \, dt. \tag{25}$$

The summands (in both sums) are ordered by their appearance in time. The idea is now to sort them by the values of $Z(\tau_n)$. This is achieved by multiplying each summand with unity of the form $\sum_{z\in\mathcal{Z}} \mathbb{1}(Z(\tau_n) = z)$ and changing the order of summation. The outer sum is now indexed by $z \in \mathcal{Z}$. The next idea is to normalize the numerator and denominator by N to get empirical averages. As $N \to \infty$, the ergodic theorem allows us to move back to expectations. In the denominator, this reveals the average waiting time as a mixture of inverse exit rates:

$$\sum_{z\in\mathcal{Z}} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}(Z(\tau_n) = z)(\tau_{n+1} - \tau_n)$$
$$= \sum_{z\in\mathcal{Z}} \mathbb{P}[Z(\tau_n) = z]\mathbb{E}[\tau_{n+1} - \tau_n | Z(\tau_n) = z] \tag{26}$$
$$= \sum_{z\in\mathcal{Z}} \mathbb{P}[Z(\tau_n) = z]\Lambda_0(z)^{-1}.$$

In the numerator, the contributions of $Z(\tau_n) = z$ to the sum is handled analogously by moving from the empirical mean to the expectation

$$\sum_{z\in\mathcal{Z}} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbb{1}(Z(\tau_n) = z) \int_{\tau_n}^{\tau_{n+1}} X(t) \, dt$$
$$= \sum_{z\in\mathcal{Z}} \mathbb{E}\left[\mathbb{1}(Z(\tau_n) = z) \int_{\tau_n}^{\tau_{n+1}} X(t) \, dt\right] \tag{27}$$
$$= \sum_{z\in\mathcal{Z}} \mathbb{P}[Z(\tau_n) = z]\mathbb{E}\left[\int_{\tau_n}^{\tau_{n+1}} X(t) \, dt | Z(\tau_n) = z\right]. \tag{28}$$

The integrals in Eq. (28) evaluate to

$$\mathbb{E}\left[\int_{\tau_n}^{\tau_{n+1}} X(t) \, dt | Z(\tau_n) = z\right] = \frac{x(z)}{\Lambda_0(z)}, \tag{29}$$

in detail, we have

$$\mathbb{E}\left[\int_{\tau_n}^{\tau_{n+1}} X(t)\ \mathrm{d}t | Z(\tau_n) = z\right]$$

$$=\mathbb{E}[\int_{\tau_n}^{\tau_{n+1}} V(t)\ \mathrm{d}t | Z(\tau_n) = z]$$

$$=\mathbb{E}[\int_0^{\tau_{n+1}-\tau_n} e^{-A(z)(t)}\ \mathrm{d}t | Z(\tau_n) = z]\mathbb{E}[V(\tau_n) | Z(\tau_n) = z]$$

$$+\ \mathbb{E}[\int_0^{\tau_{n+1}-\tau_n} \int_0^t e^{-A(z)s}\ \mathrm{d}s\ \mathrm{d}t | Z(\tau_n) = z]b(z)$$

$$=(\Lambda_0(z)\,\mathbf{I}+A(z))^{-1}\sum_{z'\neq z}\mathbb{P}[W_{n-1}=z'|W_n=z]x(z')$$

$$+\ \Lambda_0(z)^{-1}(\Lambda_0(z)\,\mathbf{I}+A(z))^{-1}b(z)$$

$$=\frac{x(z)}{\Lambda_0(z)}.$$

The consolidating technical calculations read

$$\mathbb{E}[X_\infty] = \lim_{N\to\infty}\frac{\int_0^{\tau_N} X(t)\ \mathrm{d}t}{\tau_N}$$

$$=\frac{\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N}\int_{\tau_n}^{\tau_{n+1}} X(t)\ \mathrm{d}t\sum_{z\in\mathcal{Z}}\mathbb{1}(Z(\tau_n)=z)}{\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N}(\tau_{n+1}-\tau_n)\sum_{z\in\mathcal{Z}}\mathbb{1}(Z(\tau_n)=z)}$$

$$=\frac{\sum_{z\in\mathcal{Z}}\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N}\mathbb{1}(Z(\tau_n)=z)\int_{\tau_n}^{\tau_{n+1}} X(t)\ \mathrm{d}t}{\sum_{z\in\mathcal{Z}}\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N}\mathbb{1}(Z(\tau_n)=z)(\tau_{n+1}-\tau_n)}$$

$$=\frac{\sum_{z\in\mathcal{Z}}\mathbb{P}[Z(\tau_n)=z]\mathbb{E}[\int_{\tau_n}^{\tau_{n+1}} X(t)|Z(\tau_n)=z]}{\sum_{z\in\mathcal{Z}}\mathbb{P}[Z(\tau_n)=z]\mathbb{E}[\tau_{n+1}-\tau_n|Z(\tau_n)=z]}$$

$$=\frac{\sum_{z\in\mathcal{Z}}\tilde{\pi}(z)\frac{x(z)}{\Lambda_0(z)}}{\sum_{z\in\mathcal{Z}}\tilde{\pi}(z)\Lambda_0(z)^{-1}}$$

$$=\frac{\sum_{z\in\mathcal{Z}}\pi(z)x(z)}{\sum_{z\in\mathcal{Z}}\pi(z)}.$$

Remark 3.9. The expression (21) has a rather astonishing interpretation that can be understood as a consequence of the waiting time paradox [97]. Suppose the system operates in stationarity. For simplicity, we assume X is one-dimensional. If we choose a random time t, then with probability $\pi(z)$ we hit an interval $\tau_n \leq t < \tau_{n+1}$ with a 'label' $Z(t) = z$. Call this event I_z. The time point adds $\mathbb{E}[X(t)|I_z]$ to the mean if we think of the stationary mean as $\sum_{z\in\mathcal{Z}}\pi(z)\int_0^\infty \mathbb{E}[X(s)|I_z]\ \mathrm{d}F_z(s)$, and $F_z(s)$ is the distribution of $s = t - \tau_n$ when in $Z(t) = z$. Looking at the structure of expression (21), the integral evaluates to $x(z) = \mathbb{E}[X(\tau_{n+1})|I_z]$. Why is this paradox? We might ad hoc assume that t lands, on average, at some centered location within the interval $[\tau_n, \tau_{n+1}]$, in particular it is, on average, smaller than τ_{n+1}. The progression of $s \mapsto \mathbb{E}[X(s)|I_z] = V(s)$, see Eq. (15), is strictly monotone. Imagine it is increasing (decreasing). Then $t < \tau_{n+1}$ implies $\mathbb{E}[X(t)|I_z] < (>)\mathbb{E}[X(\tau_{n+1})|I_z]$. By the monotonicity of the expectation, this implies $\int_0^\infty \mathbb{E}[X(s)|I_z]\ \mathrm{d}F_z(s) < (>)x(z)$. In contrast, theorem 3.7 informs us that equality holds. This can be understood as a consequence of the waiting time paradox. A uniformly random time point t satisfies that $\tau_{n+1} - t$ is exponentially distributed with parameter $\Lambda_0(z)$ by the memory-less property of the exponential distribution. However, due to symmetry reasons in the uniform choice of t, the distance $t - \tau_n$ of the last Z-jump in backwards time is also exponentially distributed with parameter $\Lambda_0(z)$. To deviate from the main line of thought, the interval that t lands in has, on average, twice the expected length. This paradox is resolved by the size bias effect: longer intervals have a higher chance to be hit by the point t. The explicit size-biased distribution of $\tau_{n+1} - \tau_n$ is provided in [97]. Returning to the main line of thought, a randomly chosen time point is actually an average endpoint, regarded from the perspective of a randomly chosen interval. The difference lies in

the random choice of a time point versus the random choice of an interval. The waiting time paradox permits the slim formulation of theorem 3.7, once the recursion in proposition 3.5 is solved.

Davis [65, p.36] commented on non-Markovian models, which the process $X(t)$ is an example of. "For some non-Markovian models, in renewal theory for instance, special techniques have been developed, but these tend not to extend beyond the very special assumptions on which the models are based, so that generalizations will require a completely fresh approach." Our derivation resembles those common in renewal theory with Eq. (26) computing average sojourn times. And indeed, for the derivation of ESME to work it was crucial to assume the conditionally linear form of the propensities in Eq. (11). From this assumption we obtained equations (15) that are closed in the conditional first moment $V(t)$. As a further consequence of the linear form the update function g in Eq. (17) is linear in $V(\tau_n)$, see Eq. (16), from which we obtain the linear recursions in Eq. (18). If the linear assumption in Eq. (11) was dropped we would not obtain closed equations as in Eq. (15) to begin with. Depending on the functional form of the propensities the equation would involve higher order conditional moments, e.g., for second-order mass-action reactions, or the conditional expectation of rational functions, e.g., for Hill propensities. The analogue to the variable $V(t)$ would need to summarize all the dependencies and would be generally of infinite dimension. A conditional moment closure or other projection methods would generally be required to reduce it to finite dimensions. This may result in non-linear g in Eq. (17) and in non-linear recursions. One way to obtain a linear equation as in Eq. (15) would be the use of conditional probabilities $\mathbb{P}[X(t) = x|Z_{[0,t]}]$, indexed over all $x \in \mathbb{N}^d$, together with an appropriate closure scheme. We anticipate that this can become computationally prohibitive. However, it might be feasible for systems that only access finitely many states, e.g., due to conservation relations. For instance, this is the case if every reaction is a conversion reactions, i.e., $\sum_{i=1}^d S_{ij} = \sum_{i=1}^d P_{ij}$ for all j. While the functional form was required to be linear in the state x, we emphasize that the dependence in the environment component z can be of arbitrary functional form.

3.1.4 Quantification of environmental shares

The stationary mean is a composite result of the subsystem existing in different environmental states. Thus, we aim to quantify the share that each environmental state contributes to the value of the stationary mean. First, let us fix a subsystem species $1 \leq i \leq d$ and consider its stationary mean $\mathbb{E}[X_{i,\infty}]$. When we computed the stationary mean following Eq. (25), we sorted the summands by the environmental states $Z(\tau_n) = z$. Inspired by Eq. (27), we define the *environmental share* that environmental state z contributes to the stationary mean of species i, as

$$\alpha_i(z) := \zeta_i^{-1} \mathbb{E}\left[\mathbb{1}(Z(\tau_n) = z) \int_{\tau_n}^{\tau_{n+1}} X_i(t) \, dt\right] \tag{30}$$

with the normalization

$$\zeta_i := \sum_{z \in \mathcal{Z}} \mathbb{E}\left[\mathbb{1}(Z(\tau_n) = z) \int_{\tau_n}^{\tau_{n+1}} X_i(t) \, dt\right].$$

We note that, by definition, $\sum_{z \in \mathcal{Z}} \alpha_i(z) = 1$. It holds

$$\alpha_i(z) = \frac{\pi(z) x_i(z)}{\mathbb{E}[X_{i,\infty}]}, \tag{31}$$

because $\mathbb{E}[\mathbb{1}(Z(\tau_n) = z) \int_{\tau_n}^{\tau_{n+1}} X_i(t) \, dt] \propto \pi(z) x_i(z)$ by Eq. (29) and proposition 3.1.

This novel technique allows to identify whether one environment state dominates the contribution to the stationary mean. By this tool we pursue our goal to attribute the behavior of the subnetwork to features of the environment. We apply the method in the case studies, section 8.3 below.

3.2 First- and second-order moments of counting processes in a linear random environment

The first- and second-order moment characterization of Hurwitz-stable linear CRNs was provided by [92] and the equivalence with the linear noise approximation was shown. We extend these for the non-stable modulated birth process, i.e., doubly stochastic Poisson process, with a linear stationary modulator. For the derivation we use point process theory [98].

Theorem 3.10. *Let $Y(t)$ be a multivariate counting process for which the $\mathcal{F}_t^{YZ}$-intensity and the external process $Z(t)$ satisfy the standard conditions (C1)-(C4), assuming A in Eq. (C2) is invertible. Firstly, for all $t \geq 0$, we have $\mathbb{E}[Y(t)] = \mu t$. Secondly, for all $t, s \geq 0$*

$$\mathbb{C}\text{ov}[Y(t), Y(s)] = \left[\text{diag}(\mu) + C(A^{-1}\Sigma + \Sigma A^{-T})C^T\right]s$$
$$+ C\left[\Sigma A^{-2T}(\exp(-A^T s) - \mathbf{I}) + (\exp(-At) - \exp(-A(t-s)))A^{-2}\Sigma\right]C^T.$$

Proof. For the first moment, we have

$$\mathbb{E}[Y(t)] = \int \mu + C\mathbb{E}[Z(u)] \,\mathrm{d}u = \mu t. \tag{32}$$

using the zero-mean $\mathcal{F}_t^{YZ}$-martingale $Y(u) - \int_0^t \lambda(u) \,\mathrm{d}u$.

Turning towards the covariance, we employ the covariance decomposition, the multivariate analogue of [98, §3.3, Eq. (3.39)], suppressing the subscript $\mathbb{P}$,

$$\mathbb{C}\text{ov}[Y(t), Y(s)] = \mathbb{E}[\mathbb{C}\text{ov}[Y(t), Y(s)|\mathcal{F}_t^{YZ}]] + \mathbb{C}\text{ov}[\mathbb{E}[Y(t)|\mathcal{F}_t^{YZ}], \mathbb{E}[Y(s)|\mathcal{F}_t^{YZ}]]$$
$$= \text{diag}(\mu)s + C\int_0^t \int_0^s \mathbb{E}[Z(u)Z(\tau)^T] \,\mathrm{d}\tau \,\mathrm{d}u\, C^T$$
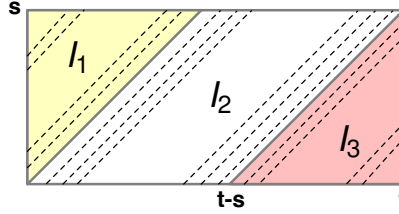
Figure 4: Integration domain

Now we split the integration domain $[0, t] \times [0, s]$ into three parts according to figure 4. Additionally, we change the direction of integration to the 45° diagonal (indicated by dashed lines) on which the auto/cross-covariance $\mathbb{E}[Z(u)Z(\tau)^T]$ is constant. The reason for this is the constant time lag $h = |\tau - u|$ and the assumed stationarity of the auto/cross-covariance. We get, using integration by parts,

$$I_1 = \int_0^s (s-h)\mathbb{E}[Z(u)Z(u+h)^T] \,\mathrm{d}h$$
$$= s\int_0^s \Sigma \exp(-A^T h) \,\mathrm{d}h - \int_0^s h\Sigma \exp(-A^T h) \,\mathrm{d}h = \Sigma A^{-T}s + \Sigma A^{-2T}(\exp(-A^T s) - \mathbf{I})$$
$$I_2 = \int_0^{t-s} s\mathbb{E}[Z(u)Z(u-h)^T] \,\mathrm{d}h = s\int_0^{t-s} \exp(-Ah)\Sigma \,\mathrm{d}h$$
$$= s(\exp(-A(t-s)) - \mathbf{I})A^{-1}\Sigma,$$
$$I_3 = \int_0^s (s-h)\mathbb{E}[Z(u)Z(u-(t-s+h))^T] \,\mathrm{d}h$$
$$= s\int_0^s \exp(-A(t-s+h))\Sigma \,\mathrm{d}h - \int_0^s h\exp(-A(t-s+h))\Sigma \,\mathrm{d}h$$
$$= s\exp(-A(t-s))A^{-1}\Sigma + (\exp(-At) - \exp(-A(t-s)))A^{-2}\Sigma.$$

In total,

$$\mathbb{C}\text{ov}[Y(t), Y(s)] = \text{diag}(\mu)s + C(I_1 + I_2 + I_3)C^T = \left[\text{diag}(\mu) + C(A^{-1}\Sigma + \Sigma A^{-T})C^T\right]s$$
$$+ C\left[\Sigma A^{-2T}(\exp(-A^T s) - \mathbf{I}) + (\exp(-At) - \exp(-A(t-s)))A^{-2}\Sigma\right]C^T.$$

$\square$
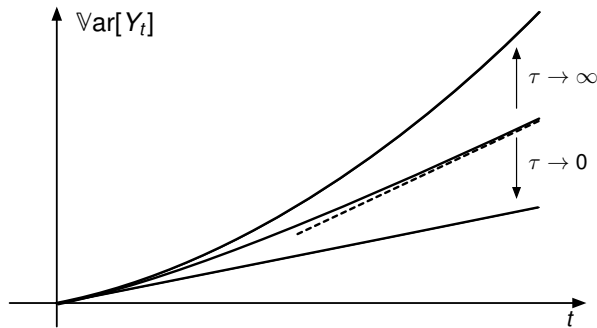
Figure 5: Schematic variance evolution of a Markov-modulated Poisson process. The lowest graph illustrates the Poisson base case which coincides with uncorrelated modulator. The larger the correlation time, the steeper is the asymptotic variance slope.

Relaxing the condition of unidirectional modulation, non-linear reaction systems were embedded in a linear environment while assuming linear subsystem-environment interaction [99], i.e., zeroth-order modulation (in both directions). First- and second order moment characterizations were derived, using the projection operator formalism.

3.3 Fano factor decomposition

For the process class of MMPPs with linear modulator, we have just presented the analytical mean and variance expressions. Asymptotically ($t \to \infty$), the covariance matrix behaves like

$$\lim_{t \to \infty} t^{-1}\mathbb{C}\mathrm{ov}[Y_t] = \mathrm{diag}(\mu_0 + C\mu) + C\left(\Sigma A^{-T} + A^{-1}\Sigma\right)C^T. \tag{33}$$

For processes in one dimension, the inverse autocorrelation decay $\tau = A^{-1} \in \mathbb{R}_{>0}$ is interpreted as the correlation time of the process. Figure 5 shows how the variance initially increases quadratically in time (tangential to the Poisson base case), and enters asymptotic linearity, the slope increasing with τ. For the extreme case $\tau \to 0$, the Poisson base case is recovered, while for $\tau \to \infty$, the asymptotic slope increases unboundedly. This matches the case of static, but random X ($\tau = \infty$) with quadratic (and thus super-linear) asymptotic increase.

The Fano factor, i.e., the ratio of variance and mean, is a noise-quantifying device, suited in the scenario of counting processes, because it assigns 1 to the base Poisson case. The Fano factor F of any super-Poissonian random variable is naturally decomposed into $F = 1 + \Delta F$ to extract its deviation ΔF from the Poisson base case. For counting process models, we evaluate the asymptotic Fano factor $\lim_{t \to \infty} \mathbb{V}\mathrm{ar}[Y(t)]/\mathbb{E}[Y(t)]$. See section 8.1.3 for a decomposition of translational noise. And see section 4.4 for additional tools to compute $\lim_{t \to \infty} t^{-1}\mathbb{C}\mathrm{ov}[Y_t]$ for self-exciting counting processes.

3.4 Second-order moments: spectral formula

Several research groups have contributed to our good understanding of linear subsystems that are modulated in their zeroth-order reactions. Raj et al. [100] calculated the variance for gene expression models, using spectral theory. Warren et al. [92] proved the equivalence with the linear noise approximation for the second-order expressions of Hurwitz-stable linear CRNs. Gupta and Khammash [48] characterized the second-order moments for linear subsystems with zeroth-order modulation by uncorrelated environment components. Zeroth-order modulation by a linear environment is included in the theory on linear systems, because the joint subsystem-environment network is linear. We take a more general approach, providing spectral theory for linear subsystems in a correlated environment, not necessarily linear. We also relax the Hurwitz condition for the subsystem.

Building on the previous work, we provide the second-order result in the frequency domain for a linear CRN X in a random weakly stationary environment Z. Let $X(t) = [X_1(t), \ldots, X_d(t)]^T$ be the state vector of a linear CRN in a random environment, i.e., there are only first- and zeroth-order reactions. We make the assumption that the external process only modulates the zeroth-order

reactions. To formalize this, we assume that the reactions $R_1, \ldots, R_m$ are modulated by $Z(t)$, i.e., the propensity vector is $\lambda(t) = [a_1(Z(t)), \ldots, a_m(Z(t))] = a(Z(t))$ with

$$a(z) = \mu + C \cdot z \in \mathbb{R}_{\geq 0}^M, \quad \mu \in \mathbb{R}^M, C \in \mathbb{R}^{M \times l}, \tag{34}$$

and without loss of generality $\mathbb{E}[Z(t)] = 0$. The reactions $R_{m+1}, \ldots, R_M$ are not modulated. Correspondingly, the stoichiometric matrix decomposes into

$$N = [N_1 \ N_2], \quad N_1 \in \mathbb{Z}^{d \times m}, N_2 \in \mathbb{Z}^{d \times (M-m)}. \tag{35}$$

Let $Y(t) \in \mathbb{N}^m$ count the occurrence of reactions $R_1, \ldots, R_m$ and $\tilde{Y}(t) \in \mathbb{N}^{M-m}$ count the occurrence of reactions $R_{m+1}, \ldots, R_M$. Then the state is obtained as

$$X(t) = X(0) + N_1 Y(t) + N_2 \tilde{Y}(t), \tag{36}$$

compare Eq. (2). Define the filtrations

$$\mathcal{F}_t^{YZ} := \sigma(\{Y(s) : s \leq t\} \cup \{\tilde{Y}(s) : s \leq t\} \cup \{Z(s) : s \leq t\} \cup \{X(0)\})$$
$$\mathcal{F}_t^{Y} := \sigma(\{Y(s) : s \leq t\} \cup \{\tilde{Y}(s) : s \leq t\} \cup \{X(0)\}).$$

The $\mathcal{F}_t^Y$-intensity of the components $\tilde{Y}(t)$ is set to be

$$\tilde{\lambda}(t) = \Lambda X(t-) + b, \quad \Lambda \in \mathbb{R}^{(M-m) \times d}, b \in \mathbb{R}^{M-m}. \tag{37}$$

We assume a general auto/cross-spectrum $S_Z(\omega) \in \mathbb{R}^{m \times m}$,

$$S_{Z,jk}(\omega) = \int_{-\infty}^{\infty} \mathbb{E}[\Delta Z_j(t) \Delta Z_k(t - \tau)] e^{-i\tau\omega} \, d\tau,$$

where $\Delta Z_j(t) := Z_j(t) - \mathbb{E}_{\mathbb{P}}[Z_j(t)]$. How is $S_Z(\omega)$ transferred to $S_X(\omega)$ under $\mathbb{P}$, when assuming weak stationarity also for X? The mean evolution equation of $\mathbb{E}_{\mathbb{P}}[X(t)]$ is

$$\frac{d}{dt}\mathbb{E}_{\mathbb{P}}[X(t)] = \Gamma \mathbb{E}_{\mathbb{P}}[X(t)] + N\tilde{b}, \quad \Gamma = N\tilde{\Lambda} = N_2\Lambda, \quad \tilde{\Lambda} = \begin{bmatrix} 0 \\ \Lambda \end{bmatrix} \in \mathbb{R}^{M \times d}, \tilde{b} = \begin{bmatrix} \mu \\ b \end{bmatrix} \in \mathbb{R}^M. \tag{38}$$

Then all eigenvalues of Γ must have non-positive real part, otherwise the stationary mean of $X(t)$ does not exist. We distinguish the cases in which all eigenvalues of Γ are strictly negative, also called Hurwitz-stability, and the remaining cases, in which Γ is degenerate. We only consider the following mild form of degeneracy.

Definition 3.11. We call $X(t)$ mildly degenerate, when

- eigenvalues of Γ with vanishing real part are 0,

- $f\Gamma = 0$ implies $fN = 0$ for all $f \in \mathbb{R}^d$ and

- in the Jordan canonical form of Γ the Jordan block of the eigenvalue 0 is the zero matrix.

Clearly, finitely many $f_1, \ldots, f_r$ in (ii) can be found that span the left eigenspace of eigenvalue 0. The $f_1, \ldots, f_r$ correspond to conservation laws on the trajectories of $X(t)$, since Eq. (72) implies

$$df_i X(t) = f_i N_1 \, dY(t) + f_i N_2 \, d\tilde{Y}(t) = f_i N A_1 \, dY(t) + f_i N A_2 \, d\tilde{Y}(t) = 0, \tag{39}$$

$$A_1 = \begin{bmatrix} \mathbf{I} \\ 0 \end{bmatrix} \in \mathbb{R}^{M \times m}, A_2 = \begin{bmatrix} 0 \\ \mathbf{I} \end{bmatrix} \in \mathbb{R}^{M \times (M-m)}.$$

Define the probability measure $\bar{\mathbb{P}}$, under which the intensities of $Y(t)$ are replaced by their averages μ, i.e., $Y(t)$ is a homogeneous Poisson process. Then under $\bar{\mathbb{P}}$, the reaction system for $X(t)$ has propensities

$$\bar{\lambda}(t) = \tilde{\Lambda} X(t-) + \tilde{b}, . \tag{40}$$

33

The mean evolution equation of $\mathbb{E}_{\bar{\mathbb{P}}}[X(t)]$ is the same as for $\mathbb{E}_{\mathbb{P}}[X(t)]$ in Eq. (38). When Γ is Hurwitz-stable, then the stationary mean is $\mathbb{E}_{\bar{\mathbb{P}}}(X) = \mathbb{E}_{\mathbb{P}}(X) = -\Gamma^{-1}N\tilde{b}$. Or in the mildly degenerate case, we introduce the notations $\mathbb{P}[\cdot|c]$ and $\bar{\mathbb{P}}[\cdot|c]$ for the measures conditional on $f_i X(0) = c_i, i = 1, \ldots, r$. Then $\mathbb{E}_{\mathbb{P}}(X|c)$ solves

$$\Gamma\mathbb{E}_{\mathbb{P}}(X|c) + N\tilde{b} = 0,$$
$$f_i\mathbb{E}_{\mathbb{P}}(X|c) = c_i, i = 1, \ldots, r.$$

and so does $\mathbb{E}_{\bar{\mathbb{P}}}(X|c)$. Denote by $\mathbb{Cov}_{\bar{\mathbb{P}}}(X)$ the stationary covariance matrix of the system $X(t)$ under $\bar{\mathbb{P}}$ in the Hurwitz-stable case and $\mathbb{Cov}_{\bar{\mathbb{P}}}(X|c)$ the conditional covariance matrix in the degenerate case. Setting up the covariance evolution equation it can be shown that $\mathbb{Cov}_{\bar{\mathbb{P}}}(X)$ satisfies

$$\Gamma\mathbb{Cov}_{\bar{\mathbb{P}}}(X) + \mathbb{Cov}_{\bar{\mathbb{P}}}(X)\Gamma^T + N\operatorname{diag}(\tilde{\Lambda}\mathbb{E}_{\bar{\mathbb{P}}}(X) + \tilde{b})N^T = 0. \tag{41}$$

and analogously for $\mathbb{Cov}_{\bar{\mathbb{P}}}(X|c)$ with $\mathbb{E}_{\bar{\mathbb{P}}}(X)$ replaced by $\mathbb{E}_{\bar{\mathbb{P}}}(X|c)$.

Proposition 3.12. *Assuming that $Z(t)$ is weakly stochastic with auto/cross-spectrum $S_Z(\omega)$, while $X(t)$ is assumed to be weakly stationary with Hurwitz-stable or mildly degenerate Γ and Eq. (35)-(37) hold. Then the auto/cross-spectrum of $X(t)$ is*

$$S_X(\omega) = (i\omega - \Gamma)^{-1}B(-i\omega - \Gamma^T)^{-1} + (i\omega - \Gamma)^{-1}N_1 C S_Z(w) C^T N_1^T (-i\omega - \Gamma^T)^{-1},$$

with $B = -\Gamma\mathbb{Cov}_{\bar{\mathbb{P}}}(X) - \mathbb{Cov}_{\bar{\mathbb{P}}}(X)\Gamma^T$, where $\mathbb{Cov}_{\bar{\mathbb{P}}}(X)$ is evaluated as $\sum_c \bar{\mathbb{P}}[X(0) = c]\mathbb{Cov}_{\bar{\mathbb{P}}}(X|c)$ in the mildly degenerate case.

Proof. We present the proof for Hurwitz-stable Γ, which implies $\lim_{t\to\infty} e^{\Gamma t} = 0$. The changes for mildly degenerate Γ are addressed in the end of the proof. We first expand the term Eq. (75) to the following expression by using the martingale $M(t) = Y(t) - \int_0^t CZ(u) + \mu\,du$ with respect to the filtration $\mathcal{F}_t^{YZ\infty} := \sigma(\{Y(s) : s \le t\} \cup \{Z(s) : s \ge 0\} \cup \{X(0)\})$. In contrast to $\mathcal{F}_t^{YZ}$, the filtration $\mathcal{F}_t^{YZ\infty}$ has knowledge of the entire trajectory of Z, also the future. Then, due to vanishing cross terms,

$$\mathbb{E}\left[\int_0^t e^{\Gamma(t-u)}N_1(CZ(u)\,du + dM(u))\left(\int_0^s e^{\Gamma(s-u)}N_1(CZ(u)\,du + dM(u))\right)^T\right]$$

$$= \int_0^s e^{\Gamma(t-u)}N_1\operatorname{diag}(\mu)N_1 e^{\Gamma^T(s-u)}\,du \tag{42}$$

$$+ \int_0^t\int_0^s e^{\Gamma(t-u)}N_1 C\mathbb{E}[Z(u)Z(\sigma)]C^T N_1^T e^{\Gamma^T(s-u)}\,d\sigma\,du \tag{43}$$

using Eq. (4) and Eq. (5). For convenience denote the term Eq. (43) by $L(t, s)$. We add the terms Eq. (42) and Eq. (76) to yield

$$\int_0^s e^{\Gamma(t-u)}N\operatorname{diag}(\tilde{\Lambda}\mathbb{E}_{\bar{\mathbb{P}}}(X) + \tilde{b})N^T e^{\Gamma^T(s-u)}\,du.$$

By Eq. (41) and using the product rule, this term evaluates to

$$e^{\Gamma(t-s)}\mathbb{Cov}_{\bar{\mathbb{P}}}(X) - e^{\Gamma t}\mathbb{Cov}_{\bar{\mathbb{P}}}(X)e^{\Gamma s}. \tag{44}$$

The covariance $\mathbb{Cov}_{\mathbb{P}}[X(t), X(s)]$ is now the sum of the terms in Eq. (74), Eq. (43) and Eq. (44). Letting $t = s$ and $t \to \infty$, we may decompose

$$\mathbb{Cov}_{\mathbb{P}}[X(0)] = \lim_{t\to\infty} e^{\Gamma t}\mathbb{Cov}_{\mathbb{P}}[X(0)]e^{\Gamma^T t} + \lim_{t\to\infty}\mathbb{Cov}_{\bar{\mathbb{P}}}(X) - e^{\Gamma t}\mathbb{Cov}_{\bar{\mathbb{P}}}(X)e^{\Gamma t} + \lim_{t\to\infty} L(t, t)$$

$$= 0 + \mathbb{Cov}_{\bar{\mathbb{P}}}(X) + \bar{L}, \quad \bar{L} := \lim_{t\to\infty} L(t, t). \tag{45}$$

Inserting this decomposition into Eq. (74), we get

$$\mathbb{Cov}_{\mathbb{P}}[X(t), X(s)] = e^{\Gamma t}\mathbb{Cov}_{\bar{\mathbb{P}}}(X)e^{\Gamma^T s} + e^{\Gamma t}\bar{L}e^{\Gamma^T s} + e^{\Gamma(t-s)}\mathbb{Cov}_{\bar{\mathbb{P}}}(X) - e^{\Gamma t}\mathbb{Cov}_{\bar{\mathbb{P}}}(X)e^{\Gamma s} + L(t, s)$$

$$= e^{\Gamma(t-s)}\mathbb{Cov}_{\bar{\mathbb{P}}}(X) + (L(t, s) + e^{\Gamma t}\bar{L}e^{\Gamma^T s}) \tag{46}$$

Now, in this decomposition the first term is precisely $\mathbb{Cov}_{\bar{\mathbb{P}}}[X(t), X(s)]$. We further recognize the bracket term as the auto/cross-covariance $\mathbb{Cov}_{\mathbb{P}}[\tilde{X}(t), \tilde{X}(s)]$ of the system

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{X}(t) = \Gamma\tilde{X}(t) + N_1 C Z(t),$$

which is initialized such that it satisfies weak stationarity. By the linearity of the Fourier transform, the decomposition Eq. (46) transfers to the spectrum as $S_X(\omega) = S_1(\omega) + S_2(\omega)$. The first term yields

$$S_1(\omega) = (i\omega - \Gamma)^{-1} B (-i\omega - \Gamma^T)^{-1}$$

by the corresponding result on linear CRNs, see [92, Appendix 3] and [63, §4.5.6]. For the second term, we employ the transformation rule for spectra of linearly filtered weakly stationary stochastic processes in continuous time, i.e., the continuous version of [101, §2.5 Eq. (2.94)] or the multivariate version of [102, §4.12 Eq. (4-162b)], see, e.g., [103, §7.5.1 Eq. (7.163)]. With the impulse response function $G(\tau) = \mathbb{1}_{[0,\infty)}(\tau) e^{\Gamma\tau} N_1 C$ and its Fourier transform

$$\hat{G}(\omega) := \int_0^\infty G(\tau) e^{-i\omega\tau} \, \mathrm{d}\tau = (i\omega - \Gamma)^{-1} N_1 C,$$

we derive the following expression

$$S_2(\omega) = \hat{G}(\omega) S_Z(\omega) \overline{\hat{G}(\omega)}^T = (i\omega - \Gamma)^{-1} N_1 C S_Z(w) C^T N_1^T (-i\omega - \Gamma^T)^{-1},$$

which concludes the derivation.

For mildly degenerate Γ, the proof is the same with $\mathbb{P}$ and $\bar{\mathbb{P}}$ replaced by $\mathbb{P}[\cdot|c]$ and $\bar{\mathbb{P}}[\cdot|c]$, respectively. But in order to see that Eq. (45) in the adapted form

$$\mathbb{Cov}_{\mathbb{P}}[X(0)|c] = \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) + \bar{L}$$

still holds, we use the following argument that replaces $\lim_{t\to\infty} e^{\Gamma t} = 0$. Note, that $\bar{L}$ needs no conditioning on c since $\mathbb{E}[Z(u)Z(\sigma)]$ is independent of $X(0)$. It is enough to show

$$v_i \mathbb{Cov}_{\mathbb{P}}[X(0)|c] = v_i \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) + v_i \bar{L} \tag{47}$$

for a basis $v_1, \ldots, v_d$ of $\mathbb{R}^n$. We choose a basis such that Γ is in the Jordan canonical form. Let $v_1, \ldots, v_l$ correspond to the Jordan block of eigenvalue 0. For $i = l+1, \ldots, d$ we have

$$\lim_{t\to\infty} v_i e^{\Gamma t} = v_i \lim_{t\to\infty} e^{\lambda t} q(Nt),$$

where λ has negative real part by property (i) in definition 3.11, q is a polynomial and N a nilpotent matrix, such that the limit is 0. This establishes Eq. (47). Furthermore, by the property (iii) in definition 3.11, $v_1, \ldots, v_l$ can be written as linear combinations of $f_1, \ldots, f_r$ and we can show the equality for f_i, replaced by v_i. With

$$f_i e^{\Gamma t} = f_i \cdot \mathbf{I} + f_i \Gamma t + \cdots = f_i \cdot \mathbf{I} + 0 = f_i$$

we obtain

$$f_i \mathbb{Cov}_{\mathbb{P}}[X(0)|c] = \lim_{t\to\infty} f_i e^{\Gamma t} \mathbb{Cov}_{\mathbb{P}}[X(0)|c] e^{\Gamma^T t} + f_i \lim_{t\to\infty} \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) - f_i e^{\Gamma t} \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) e^{\Gamma^T t} + f_i \bar{L}$$

$$= \lim_{t\to\infty} f_i \mathbb{Cov}_{\mathbb{P}}[X(0)|c] e^{\Gamma^T t} + f_i \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) - \lim_{t\to\infty} f_i \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) e^{\Gamma^T t} + f_i \bar{L}$$

$$= \lim_{t\to\infty} \mathbb{E}_{\mathbb{P}}[(f_i X(0) - \mathbb{E}[f_i X(0)])(X(0) - \mathbb{E}[X(0)])|c] e^{\Gamma^T t} + f_i \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c)$$

$$- \lim_{t\to\infty} \mathbb{E}_{\bar{\mathbb{P}}}[(f_i X(t) - \mathbb{E}[f_i X(t)])(X(t) - \mathbb{E}[X(t)])|c] e^{\Gamma^T t} + f_i \bar{L}$$

$$= \lim_{t\to\infty} \mathbb{E}_{\mathbb{P}}[(c_i - c_i)(X(0) - \mathbb{E}[X(0)])|c] e^{\Gamma^T t} + \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c)$$

$$- \lim_{t\to\infty} \mathbb{E}_{\bar{\mathbb{P}}}[(c_i - c_i)(X(t) - \mathbb{E}[X(t)])|c] e^{\Gamma^T t} + f_i \bar{L}$$

$$= 0 + f_i \mathbb{Cov}_{\bar{\mathbb{P}}}(X|c) - 0 + f_i \bar{L}$$

The second to last equality holds because of what property (ii) in definition 3.11 implies for the trajectories, see Eq. (39). □

This proposition generalizes the corresponding result by Gupta and Khammash [48, theorem 2.1], who assigned the environment components to separate reactions and assumed them to be stochastically independent. Furthermore, the theorem generalizes the result in queuing theory, which corresponds to the birth-death with a Hawkes model [104] for the birth events. The expressions that we provide shed light on the structure of the auto/cross-covariance, e.g., [105, Eq.(64)&(65)].

3.5 Probability evolution

So far, we have described the moments of the subsystem. Now, we present the probability evolution equation of the subsystem in the form of a cumulant expansion as derived by Bronstein [106]. We translate it into the language of the conditional expectation with its projection interpretation. Examples are provided in sections 6.1.1 and 8.3.6. Again, we use the tower property to conditioning on the environment in a first step and resolve the remaining randomness in a second step to arrive at a deterministic equation.

3.5.1 Model assumptions

We assume a multiplicative form for the propensities, such that they factorize into a Z dependent and an X dependent factor. More precisely, we assume that conditionally on the (multidimensional) environment trajectory, the differential form of the Chapman-Kolmogorov equation reads

$$\partial_t p(t, x \mid Z_{[0,t]}) = \sum_{j=1}^{M} Z_j(t)(\Lambda_j p)(t, x | Z_{[0,t]}). \tag{48}$$

Here, Λ_j can be the jump operator of the chemical master equation, the differential operator of the Fokker-Planck equation or the drift operator of the Liouville equation. They are constant. Furthermore, we assume stationarity for the environment. In particular, the mean $\mu_j := \mathbb{E}[Z_j(0)] \equiv \mathbb{E}[Z_j(t)]$ is time-independent.

3.5.2 Generalized master equation: Cumulant expansion

We introduce the collection of marginal probability $p_t := p(t, x)_{x \in \mathcal{X}}$, which is deterministic. Analogously, we introduce the notation $\tilde{p}_t$ for the collection of random variables $p(t, x \mid Z_{[0,t]})_{x \in \mathcal{X}}$. The probability p_t is obtained from $\tilde{p}_t$ by the projection operator $\mathcal{P} : \mathcal{L}_Z^2 \to \mathcal{L}_Z^2$ with $\mathcal{P}\tilde{p}_t := \mathbb{E}\tilde{p}_t = p_t$. Here, $\mathcal{L}_Z^2$ denotes the space of Hilbert space of square-integrable random functions, that are measurable with respect to $Z_{[0,t]}$ for some t.

$$\left\{ f : \Omega \to \mathbb{R}^{\mathcal{X}} : \mathbb{E}\left[\int_{\mathcal{X}} f(x)^2 \, d\nu(x) \right] < 0, \text{there exists } t \geq 0, \text{ s.t. } f \text{ is } Z_{[0,t]} - \text{measurable} \right\}$$

where ν is the Lebesgue measure or counting measure. In order to apply the Mori-Zwanzig formalism, we introduce the orthogonal operator $\mathcal{Q} := \mathbf{I} - \mathbb{E}$, such that $\mathcal{Q}\tilde{p}_t = \tilde{p}_t - p_t =: \Delta p_t$ denotes the deviation from the mean. According to the Mori-Zwanzig formalism, we can decompose the evolution equation (48) to obtain coupled evolution equations for $p_t = \mathcal{P}\tilde{p}_t$ and $\Delta p_t = \mathcal{Q}\tilde{p}_t$. Given the decomposition $\tilde{p}_t = \mathcal{P}\tilde{p}_t + \mathcal{Q}\tilde{p}_t$ and the evolution equation $\partial_t \tilde{p}_t = \mathcal{L}_t \tilde{p}_t$, we obtain the coupled evolution equations

$$\frac{d}{dt}\mathcal{P}\tilde{p}_t = \mathcal{P}\mathcal{L}_t\mathcal{P}\tilde{p}_t + \mathcal{P}\mathcal{L}_t\mathcal{Q}\tilde{p}_t$$

$$\frac{d}{dt}\mathcal{Q}\tilde{p}_t = \mathcal{Q}\mathcal{L}_t\mathcal{P}\tilde{p}_t + \mathcal{Q}\mathcal{L}_t\mathcal{Q}\tilde{p}_t.$$

To this end, we also decompose the environment into $Z_j(t) = \mathbb{E}[Z_j(t)] + \Delta Z_j(t)$, which induces a decomposition of the evolution operator

$$\sum_{j=1}^{M} Z_j(t)\Lambda_j = \overline{\Lambda} + \Delta\Lambda(Z_t),$$

using the deterministic operator $\overline{\Lambda} := \sum_{j=1}^{M} \mu_j \Lambda_j$ and the random operator $\Delta\Lambda(Z_t) := \sum_{j=1}^{M} \Delta Z_j(t)\Lambda_j$. For each of the terms f_i in the re-written Eq. (48)

$$\partial_t(p_t + \Delta p) = (\overline{\Lambda} + \Delta\Lambda(Z_t))(p_t + \Delta p) = \overline{\Lambda}p_t + \overline{\Lambda}\Delta p + \Delta\Lambda(Z_t)p_t + \Delta\Lambda(Z_t)\Delta p =: \sum_{i=1}^{4} f_i,$$

we check whether we are in one of the following two situations $\mathbb{E}[f_i] = f_i$ or $\mathbb{E}[f_i] = 0$. In the first case, we can attribute it entirely to the evolution of p_t, whereas, in the second case, we attribute it entirely to the evolution of Δp_t. If neither is the case, we receive contributions of the term to the evolution of both parts. This procedure corresponds to the decomposition $\mathcal{L}_t f_i = \mathcal{P}\mathcal{L}_t f_i + \mathcal{Q}\mathcal{L}_t f_i = \mathcal{P}\mathcal{L}_t f_i + 0$ and $\mathcal{L}_t f_i = \mathcal{P}\mathcal{L}_t f_i + \mathcal{Q}\mathcal{L} f_i = 0 + \mathcal{Q}\mathcal{L}_t f_i$ in the Mori-Zwanzig formalism for the first and second case, respectively. Adhering to this procedure, we obtain

$$\partial_t p_t = \overline{\Lambda}p_t + \mathbb{E}[\Delta\Lambda(Z_t)\Delta p_t] \tag{49}$$

$$\partial_t \Delta p_t = \Delta\Lambda(Z_t)p_t + \left\{\overline{\Lambda} + (\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_t)\right\}\Delta p_t. \tag{50}$$

In order to eliminate the deterministic constant part $\overline{\Lambda}$ in the evolution equation of Δp_t, we consider the change of variables, known as interaction picture, $\Delta p_t^0 := e^{-t\overline{\Lambda}}\Delta p_t$ and obtain

$$\partial_t \Delta p_t^0 = e^{-t\overline{\Lambda}}\Delta\Lambda(Z_t)p_t + e^{-t\overline{\Lambda}}(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_t)e^{t\overline{\Lambda}}\Delta p_t^0. \tag{51}$$

Its solution is expressed by the time-ordered exponential, i.e., the propagator for time-dependent generators,

$$\Delta p_t^0 = \int_0^t \overleftarrow{\mathcal{T}} \exp\left\{\int_{t'}^t e^{-\tau\overline{\Lambda}}\{(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_\tau)\}e^{\tau\overline{\Lambda}} \, d\tau\right\} e^{-t'\overline{\Lambda}}\Delta\Lambda(Z_{t'})p_{t'} \, dt'$$

$$= \int_0^t \sum_{n=0}^{\infty} \int_{\Delta^n[t',t]} \left(\prod_{k=0}^{n-1} e^{-t_k\overline{\Lambda}}\{(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_{t_k})\}e^{t_k\overline{\Lambda}}\right) d(t_0, \ldots, t_{n-1})e^{-t'\overline{\Lambda}}\Delta\Lambda(Z_{t'})p_{t'} \, dt'$$

$$= \sum_{n=0}^{\infty} \int_{\Delta^{n+1}[0,t]} \left(\prod_{k=1}^{n} e^{-t_k\overline{\Lambda}}\{(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_{t_k})\}e^{t_k\overline{\Lambda}}\right) e^{-t_0\overline{\Lambda}}(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_{t_0})p_{t_0} \, d(t_0, \ldots, t_n)$$

where $\Delta^{n+1}[s,t] := \{(t_0, \ldots, t_n) \in [s,t]^{n+1} \colon t_0 < t_1 < \cdots < t_n\}$ and $\prod_{k=1}^n A_k := A_n \circ \cdots \circ A_1$. We use the notation $\Delta t_k := t_{k+1} - t_k, k = 1, \ldots, n$ with $t_{n+1} := t$ denoting the lags and the set $[M] := \{1, \ldots, M\}$. Upon returning to the variable Δp_t,

$$\Delta p_t = \sum_{n=0}^{\infty} \int_{\Delta^{n+1}[0,t]} \left(\prod_{k=0}^{n} e^{\Delta t_k \overline{\Lambda}}\{(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_{t_k})\}\right) p_{t_0} \, d(t_0, \ldots, t_n),$$

where $\Delta t_k := t_{k+1} - t_k$ with $t_{n+1} := t$ and $t_0 = t'$. When we plug this into Eq. (49), we obtain

$$\partial_t p_t = \overline{\Lambda}p_t + \sum_{n=0}^{\infty} \int_{\Delta^{n+1}[0,t]} \mathbb{E}\left[\Delta\Lambda(Z_t)\left(\prod_{k=0}^{n} e^{\Delta t_k \overline{\Lambda}}\{(\mathbf{I} - \mathbb{E}) \circ \Delta\Lambda(Z_{t_k})\}\right)\right] p_{t_0} \, d(t_0, \ldots, t_n). \tag{52}$$

We insert the definition of $\Delta\Lambda(Z_{t_k})$ to obtain for the expectation

$$\sum_{\iota \in \{1, \ldots, M\}^{n+2}} \sum_{\phi \colon \{0, \ldots, n\} \to \{0,1\}} (-1)^{|\phi|}\mathbb{E}\left[\Delta Z_{\iota(n+2)}(t) \prod_{k=0}^{n} a_{\phi(k)} \circ \Delta Z_{\iota(k+1)}(t_k)\right] \Lambda_{\iota(n+2)} \prod_{k=0}^{n} e^{\Delta t_k \overline{\Lambda}}\Lambda_{\iota(k+1)},$$

where $a_0 := \mathbf{I}, a_1 := \mathbb{E}$ and $|\phi| := \sum_{k=0}^{n} \phi(k)$. Define the term

$$\mathcal{C}_\iota(t_0, \ldots, t_n, t) := \sum_{\phi \colon \{0, \ldots, n\} \to \{0,1\}} (-1)^{|\phi|}\mathbb{E}\left[\Delta Z_{\iota(n+2)}(t) \prod_{k=0}^{n} a_{\phi(k)} \circ \Delta Z_{\iota(k+1)}(t_k)\right],$$

which we refer to as the cumulant term corresponding to ι. We observe, that the contribution of any ϕ that has $\phi(k) = 1 = \phi(k+1)$ for some $k = 0, \ldots, n-1$ vanishes. The same holds true when $\phi(0) = 1$ or $\phi(n) = 1$. Defining the set

$$\Phi := \{\phi\colon \{0, \ldots, n\} \to \{0,1\}, \phi(0) = 0 = \phi(n), \phi(k)\phi(k+1) = 0 \text{ for all } k = 1, \ldots, n-2\},$$

we have

$$\mathcal{C}_\iota(t_0, \ldots, t_n, t) = \sum_{\phi \in \Phi} (-1)^{|\phi|} \mathbb{E}\left[\Delta Z_{\iota(n+1)}(t) \prod_{k=0}^{n} a_{\phi(k)} \circ \Delta Z_{\iota(k)}(t_k)\right].$$

If we define the higher-order autocovariance function

$$\gamma_\iota(t_0, \ldots, t_n, t) := \mathbb{E}\left[\prod_{k=0}^{n+1} \Delta Z_{\iota(k)}\right],$$

then the following recursion holds for the cumulants

$$\mathcal{C}_\iota(t_0, \ldots, t_n, t) = \gamma_\iota(t_0, \ldots, t_n, t) - \sum_{i=1}^{n-1} \gamma_{\iota-}(t_0, t_1, \ldots, t_i)\mathcal{C}_{\iota+}(t_{i+1}, \ldots, t_n, t),$$

where $\iota- = \iota|_{\{0,\ldots,i\}}$ and $\iota+ = \iota|_{\{i+1,\ldots,n\}}$ the restrictions of ι to the smaller sets $\{0, \ldots, i\}$ and $\{i+1, \ldots, n\}$, respectively.

Using the cumulant terms, the Eq. (52) can be written as

$$\partial_t p_t = \overline{\Lambda} p_t + \sum_{n=0}^{\infty} \int_0^t \sum_{\iota \in [M]^{n+2}} \int_{\Delta^n[t',t]} \mathcal{C}_\iota(t', t_1, \ldots, t_n, t)\Lambda_{\iota(n+2)} \prod_{k=0}^{n} e^{\Delta t_k \overline{\Lambda}} \Lambda_{\iota(k+1)} \, \mathrm{d}(t_1, \ldots, t_n) p_{t'} \, \mathrm{d}t'$$

(53)

$$= \overline{\Lambda} p_t + \sum_{n=1}^{\infty} \int_0^t \mathcal{K}_n(t', t) p_{t'} \, \mathrm{d}t',$$

(54)

for the kernels

$$\mathcal{K}_{n+1}(t', t) := \sum_{\iota \in \{1,\ldots,M\}^{n+2}} \int_{\Delta^n[t',t]} \mathcal{C}_\iota(t', t_1, \ldots, t_n, t)\Lambda_{\iota(n+1)} \prod_{k=0}^{n} e^{\Delta t_k \overline{\Lambda}} \Lambda_{\iota(k)} \, \mathrm{d}(t_1, \ldots, t_n).$$

We refer to Eq. (53) as the cumulant expansion of the generalized master equation for a CRN in a random environment. Since we only used that conditionally linear structure of Eq. (48), the same derivation holds for a conditionally linear mean equation

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[X_t \mid Z_{[0,t]}] = \sum_{j=1}^{M} Z_j(t) A_j \mathbb{E}[X_t \mid Z_{[0,t]}].$$

(55)

Remark 3.13. The structure of the (53) can be interpreted as follows. First, we pretend that the environment marginals $Z(t_1), \ldots, Z(t_n)$ at any set of (distinct) time points $t_1, \ldots, t_n \in [0, t]$ are stochastically independent of each other. This provides the first term, which is the averaged CME dynamics. However, pretending this, we made an error, because there is a correlation between time point marginals in the environment. We first correct this error, by pretending, that the current time point marginal is correlated with exactly one other time point marginal, where we allow the second time point to vary across the interval $[0, t]$. For any third time point marginal, we pretend that independence holds. This correction gives us the first cumulant term. Again, we make an error, when we pretend this. We mitigate the error by correcting for the contribution we neglected, when a third time-point marginal is correlated with the other two, which provides the second cumulant term, and so on. We illustrate this in more detail at the end of example 8.3.6.

4 Conditioning on the subnetwork

In the previous section we computed $\mathbb{E}[h(X_t)]$ by first eliminating the randomness in X_t, while conditioning on Z_t. We now take the path via the conditional expectation with respect to X_t. Unlike the procedure in the previous $\mathbb{E}[h(X_t)|\mathcal{F}]$, where $\sigma(X_t) \subseteq \mathcal{F}$, would only result in $\mathbb{E}[h(X_t)|\mathcal{F}] = h(X_t)$, which does not eliminate randomness. Instead, the trick is that we use the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t^X]$ for $\mathcal{F}_t^X = \sigma(\{X_s : s \leq t\})$ on the intensities. We can do so because of the change of filtration theorem by Brémaud, theorem 2.5. In section 4.3.2, we also use the conditional expectation $\mathbb{E}[\cdot|\mathcal{F}_t^X]$ on the right-hand side of the moment dynamics. For instance, if the propensity of a first-order reaction $X_j \to \emptyset$ is Z_t, this involves moments $\mathbb{E}[\mathbb{E}[X_j Z_t | \mathcal{F}_t^X]]$ on the right-hand side of the moment dynamics.

To derive the process equation, previous works established an intimate link between marginal descriptions and stochastic filtering [107, 108, 77, 78]. The property of self-excitation, that is characteristic for the marginal description, emerges here from projecting high-dimensional systems to low-dimensional ones, offering a mechanistic rather than phenomenological perspective on self-excitation. The term stochastic filtering is a synonym for dynamic Bayesian estimation of the random environment components from the observed part of the network [59]. It has been shown that the posterior mean (or filter mean) of the random propensity conditional on the full history of observed reactions is the central quantity that is needed for the marginal description. However, computing it is often impossible. For this reason, we look for approximations of the filter mean. Approximate filters are typically obtained from projections, variational methods, moments closures and assumed density approaches [3, 81, 78].

Note, that, by Eq. (2), $a_j(X(t-), Z(t-))$ is indeed $\mathcal{F}_t^{YZ}$-measurable. In short, we write

$$\mathrm{R}_j : \quad \emptyset \xrightarrow{\ a_j(X(t), Z(t))\ } Y_j, \tag{56}$$

abusing the notation of birth processes for the reaction counters $Y_1, \ldots, Y_M$. By Brémaud's theorem on the change of filtration [59, §II, theorem T14], we obtain that the predictable $\mathcal{F}_t^Y$-intensity is $\hat{\lambda}(t) = \mathbb{E}[\lambda(t)|\mathcal{F}_t^Y]$, provided that a left-continuous version of the conditional expectation exists. This is always the case in our study, as we will see. We then write

$$\mathrm{R}_j : \quad \emptyset \xrightarrow{\ \mathbb{E}[a_j(X(t), Z(t))|\mathcal{F}_t^Y]\ } Y_j. \tag{57}$$

Note that the great advantage of conditioning on X_t is the process level description that we obtain. On the contrary, when conditioning on Z_t, we were only able to target deterministic quantities, i.e., the probability distribution and the moments.

4.1 Process level

Taking the route via the intensity $\hat{\lambda}_t = \mathbb{E}[a_j(X(t), Z(t))|\mathcal{F}_t^Y]$ for the changed filtration $\mathcal{F}_t^Y$, let us first observe that, in general, the $\mathcal{F}_t^Y$-intensity is some measurable function of the trajectory $Y_{[0,t]} := \{Y(t)\}_{0 \leq s \leq t}$ and $X(0)$. As a first example, we consider the Hawkes counting process model in one dimension with exponential kernel. Its positive parameters are the base level μ, the jump height β and the decay α. In differential form, the $\mathcal{F}_t^Y$-intensity $\hat{\lambda}(t) = \hat{Z}(t-)$ reads

$$\mathrm{d}\hat{Z}(t) = -\alpha(\hat{Z}(t) - \mu_0)\,\mathrm{d}t + \beta\,\mathrm{d}Y(t). \tag{58}$$

When a jump of Y occurs at time t, then $\mathrm{d}Y(t) = 1$ and hence $\hat{Z}(t)$ increases by β. In periods of no jumps, the intensity evolves according to the ODE specified in the first part of the evolution equation. In integral form it reads

$$\hat{Z}(t) = \mu_0 + \int_0^t e^{-\alpha(t-u)}\beta\,\mathrm{d}Y(u). \tag{59}$$

Under the assumption $\alpha^{-1}\beta < 1$, there exists a stationary version of $Y(t)$ [109].

In multiple dimensions, the Hawkes process with exponential kernel is more delicate. Hawkes [104] introduced the mutually exciting model for which the $\mathcal{F}_t^Y$-intensity $\hat{\lambda}(t) = \hat{Z}(t-)$ generalizes Eq. (59) and reads in coordinate integral form

$$\hat{Z}_i(t) = \mu_i + \sum_{k=1}^d \int_0^t \beta_{ik} e^{-\alpha_{ik}(t-s)}\,\mathrm{d}Y_k(s). \tag{60}$$

When a jump of Y_k occurs at time t, then $\mathrm{d}Y_k(t) = 1$ and hence $\hat{Z}_i(t)$ increases by β_{ik}. In particular, a jump of a component can influence the intensity of another component, which is usually referred to as mutually exciting. However, this is not the model that we considered in this study. We, in contrast to Eq. (60), read Eq. (58) or Eq. (59) as matrix equations. Accordingly, the matrix analogue to Eq. (58) in coordinate form reads

$$\mathrm{d}\hat{Z}_i(t) = -\sum_{k=1}^{d} \alpha_{ik}(\hat{Z}_k(t) - \mu_k)\,\mathrm{d}t + \sum_{k=1}^{d} \beta_{ik}\,\mathrm{d}Y_k(t). \tag{61}$$

and the matrix analogue to Eq. (59) reads

$$\hat{Z}(t) = \mu_0 + \int_0^t e^{-A(t-u)} B\,\mathrm{d}Y(u),\ A = (\alpha_{ik})_{1 \le i,k \le M}, B = (\beta_{ik})_{1 \le i,k \le M}, \mu_0 \in \mathbb{R}_{>0}^M. \tag{62}$$

Eqs. (61) and (62) are equivalent, but differ from Eq. (60), because in Eq. (62) the exponential operation is a matrix exponential, while in Eq. (60) the exponential is scalar, applied entry-wise to the matrix. Nonetheless, we call Eqs. (61) and (62) also Hawkes model. Eq. (61) generally requires taking the positive part of the intensity components $\hat{\lambda}(t) = \max(0, \hat{Z}(t-))$, except in special cases where positivity can be guaranteed. This makes $Y(t)$ a non-linear Hawkes model. The condition for the existence of a stationary version translates into α being invertible and $\alpha^{-1}\beta$ having a spectral radius strictly less than 1. Given the process description Eq. (57), we turn now to the state estimation of $Z(t)$, or more generally $\hat{\lambda}(t)$, from observation $Y(t)$. This requires the estimation of the posterior mean, which is a Bayesian estimation problem. Filtering theory is concerned with the solution of the estimation problem in the time-evolving setting. Assuming that $Z(t)$ is a Markov process, let the prior probability $p_t(z)$ evolve according to the operator $\mathcal{A}$, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} p_t(z) = (\mathcal{A}p_t)(z),$$

which can be the master equation, if $Z(t)$ is itself a continuous-time Markov chain, or the Fokker-Planck equation, if $Z(t)$ is a diffusion process given by a stochastic differential equation. Then, extending proposition 2.10 to multivariate counting processes, the Snyder filter specifies the evolution of $\hat{\lambda}(t)$ [64]. It can be formulated in terms of the posterior probabilities $\Pi_t(z) := \mathbb{P}[Z(t) = z | \mathcal{F}_t^Y]$ for $Z(t)$ on a discrete state space or $\Pi_t(z)$ the respective posterior probability density on a continuous state space. Extending [64, Eq. (7.138), p.392 & Eq. (7.151), p.396], the posterior probabilities evolve as

$$\mathrm{d}\Pi_t(z) = (\mathcal{A}\Pi_t)(z)\,\mathrm{d}t + \sum_{j=1}^{M} \mathbb{1}(\hat{\lambda}_j(t) > 0) \frac{(a_j(X(t-), z) - \hat{\lambda}_j(t))\Pi_{t-}(z)}{\hat{\lambda}_j(t)} \{\mathrm{d}Y_j(t) - \hat{\lambda}_j(t)\,\mathrm{d}t\}$$

and the posterior moment is $\hat{\lambda}_j(t) = \mathbb{E}_{\Pi_{t-}}[a_j(X(t-), \xi)]$, where we mean that ξ is distributed according to Π_{t-}. The term after the prior dynamics is called the innovation term. It is the $\mathcal{F}_t^Y$-martingale increment scaled by a predictable factor, the innovation gain. This structure ensures that the expectation of the posterior probability $\mathbb{E}[\Pi_t]$ is governed by the prior dynamics, i.e.,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[\Pi_t] = (\mathcal{A}\mathbb{E}[\Pi_t])(z).$$

Assume additionally, that (C3) holds, and denote by F and G the prior dynamics of the mean and the covariance of $Z(t)$, for which we assume they are functions of the first, second and third prior moments

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[Z(t)] = F(\mathbb{E}[Z(t)], \mathbb{C}\mathrm{ov}[Z(t)])$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{C}\mathrm{ov}[Z(t)] = G(\mathbb{E}[Z(t)], \mathbb{C}\mathrm{ov}[Z(t)], (\Phi_{ijk}(t))_{1 \le i,j,k \le l}) \tag{63}$$

$$\Phi_{ijk}(t) := \mathbb{E}[\delta Z_i(t)\delta Z_j(t)\delta Z_k(t)],\ \delta Z_l(t) := Z_l(t) - \mathbb{E}[Z_l(t)].$$

Then the evolution of the posterior mean $\Theta(t) := \mathbb{E}[Z(t)|\mathcal{F}_t^Y]$ can be expressed with the auxiliary processes, which are the second- and third-order posterior moments

$$V_{ik}(t) := \mathbb{E}[\Delta Z_i(t)\Delta Z_k(t)|\mathcal{F}_t^Y]$$
$$S_{ikl}(t) := \mathbb{E}[\Delta Z_i(t)\Delta Z_k(t)\Delta Z_l(t)|\mathcal{F}_t^Y]$$
$$\Delta Z_j(t) := Z_j(t) - \mathbb{E}[Z_j(t)|\mathcal{F}_t^Y].$$

It reads

$$
\begin{aligned}
\mathrm{d}\Theta(t) =& F(\Theta(t), V(t))\,\mathrm{d}t + V(t-)C^T(X(t-))\operatorname{diag}\{\kappa(\hat{\lambda}(t))\}\{\mathrm{d}Y(t) - \hat{\lambda}(t)\,\mathrm{d}t\} \\
\mathrm{d}V_{ik}(t) =& G_{ik}(\Theta(t), V(t), (S_{lmn}(t))_{1\le l,m,n\le l})\,\mathrm{d}t \\
&+ \sum_{j=1}^{M} S_{ik*}C_{*j}^T(X(t-))\kappa_j(\hat{\lambda}(t))\{\mathrm{d}Y_j(t) - \hat{\lambda}_j(t)\,\mathrm{d}t\} \\
&- \left(V(t-)C^T(X(t-))\operatorname{diag}^2\{\kappa(\hat{\lambda}(t))\}\operatorname{diag}\{\mathrm{d}Y(t)\}C(X(t-))V(t-)\right)_{ik}
\end{aligned}
\tag{64}
$$

where $\kappa_j(\lambda) := \mathbb{1}(\lambda_j > 0)\lambda_j^{-1}$ and $\hat{\lambda}(t) = \mu(X(t-)) + C(X(t-))\cdot\Theta(t-)$.

The posterior mean $\Theta(t)$ is also called the state estimate of $Z(t)$. Indeed, it minimizes the quadratic criterion among all $\mathcal{F}_t^Y$-measurable random variables by the property of the conditional expectation. Note that other criteria are also minimized by the state estimate, famously the natural loss criterion [85]. For more details on filtering of point processes, see [59, §IV] and for the more explicit setting of CRNs, we refer the reader to [3] and [78]. In general, the posterior moment equations are not closed and the evolution of the posterior probabilities can become prohibitively high-dimensional.

4.1.1 Approximate filters

Propagating the posterior distribution or equivalently the conditional moments can pose a serious challenge to achieving the goal of fast marginal simulation [3] due to state space explosion and unclosed conditional moment equations. This problem is addressed by approximate filters obtained from conditional moment closure, assumed density filtering, variational inference, entropic matching or projection [81, 78, 64]. We call the marginal simulation with an approximate filter *approximate marginal simulation*.

The construction of approximate filters for the tasks has been guided by the principle of moment closure and projection onto a tractable process class. Zechner used assumed density filtering with the Poisson and the Gamma assumption on the filtering distribution [81]. Eden and Brown as well as Harel and Opper used a Gaussian assumption [58, 110] for the state estimation from neural spike trains. Bronstein derived the method of entropic matching, which projects the filtering distribution along the Kullback-Leibler divergence time-point wise [78]. It was demonstrated for the product-Bernoulli and the product-Poisson ansatz. Alternatively, the minimization of the Kullback-Leibler divergence for the path likelihoods was recently proposed as a guiding principle for the model reduction [80]. This principle had been successfully applied for variational inference before [111].

4.1.2 Optimal linear filtering

While a declared goal of marginal simulation is simulation efficiency, this is only a secondary goal here. Rather, we concentrate on

(a) marginal simulation under limited knowledge about the environment and

(b) attribution of effects to properties of the environment.

For (a) we adhere to the experimental setting, in which only the first- and second-order moments may be estimated robustly. For (b) we are guided by the question, what effects persist under a linearization, and can thus be attributed to a linearized version of the environment. In the situation of (a) we introduce an approximate intensity of the form $\hat{\lambda}^{\mathrm{OL}}(t) = a(X(t-), \hat{Z}(t-))$ with a suboptimal

$\mathcal{F}_t^Y$-measurable estimate $\hat{Z}(t)$ of $Z(t)$. This approximate intensity also serves as a baseline to capture (b) in the case studies 8.1.1, 8.2.2 and 8.4.

In the previous section, we have discussed the problem of unclosed posterior moment equations. Here, we consider a particular projection method to obtain an approximate filter, the optimal linear filter $\hat{Z}(t)$. It was originally introduced for state estimation of the external signal $Z(t)$ that modulates a doubly stochastic Poisson process [64, Eq. (7.238), p.429]. By optimal it is meant that it minimizes the quadratic criterion, i.e., it is the solution

$$\hat{Z}(t) = \underset{\tilde{Z}(t)}{\arg\min}\, \mathbb{E}[(\tilde{Z}(t) - Z(t))^2],$$

among all estimators

$$\tilde{Z}(t) = c(t) + \int_0^t h(t, u)\, \mathrm{d}Y(u),$$

i.e., affine-linear in the trajectory $Y_{[0,t]}$. The doubly stochastic Poisson process, for which stochastic filtering was originally considered by Snyder and Brémaud, precisely corresponds to a process X with zeroth-order modulation by Z in the language of CRNs. In case of higher order modulation we generalize the optimal linear estimator which results in $c(t)$ and $h(t, u)$ with a dependency on the history $X_{[0,t)}$. Assuming the standard conditions (C1)-(C3), the optimal $c(t)$ and $h(t, u)$ can be conveniently formulated in differential vector form with an auxiliary Riccati equation [64, Eq. (7.238), p.429]. By shifting $Z(t)$ to $Z(t) - \mathbb{E}[Z(t)]$, we can assume, without loss of generality, $\mathbb{E}[Z(t)] \equiv 0$ to obtain

$$\mathrm{d}\hat{Z}(t) = -A\hat{Z}(t)\, \mathrm{d}t + B(t)C^T(X(t-))\,\mathrm{diag}\{\kappa(\mu)\}\{\,\mathrm{d}Y(t) - a(X(t-), \hat{Z}(t-))\, \mathrm{d}t\} \qquad (65)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}B(t) = A\Sigma + \Sigma A^T - AB(t) - B(t)A^T - B(t)C^T(X(t))\,\mathrm{diag}\{\kappa(\mu)\}C(X(t))B(t), \qquad (66)$$

initialized in $\hat{Z}(0) = 0, B(0) = \Sigma$. In the examples, 8.1.7, 8.2.5, 8.3.8 and 8.4, these equations are applied. We suppressed the dependency $\mu = \mu(X(t-))$ for readability. Note, that $B(t)$ is deterministic under (C4), i.e., in the case of zeroth-order modulation. The formulation has been proposed by Snyder [64, p.370-371] and resembles the Kalman-Bucy filter. It was motivated by recognizing the optimal $h(t, u)$ from the theory of Gaussian channel. In detail, the optimal linear estimator, which is to say the optimal $c(t)$ and $h(t, u)$, only depends on the first- and second-order information of the modulating process. Grandell [112] specified this dependence by an integral equation for $h(t, u)$ and choosing $c(t)$ such that the estimator is unbiased. However, the very same integral equation appears in filtering with additive noise observations instead of point process observations [113]. Hence, the optimal $h(t, u)$ can be "borrowed" from the Kalman-Bucy filter when assuming a mock Ornstein-Uhlenbeck process with mean 0, satisfying (C1)-(C2) and the observation model

$$R(t) = \int_0^t \lambda(s)\, \mathrm{d}s + \sqrt{\mu}W(t)$$

for $\lambda(t)$ satisfying (C3). This is the continuous-time additive Gaussian white noise channel $r(t) = \lambda(t) + \sqrt{\mu}N(t)$ with signal-to-noise ratio μ^{-1}. The Kalman-Bucy filter specifies the kernel $h(t, u)$ in a differential form, which translates precisely to the equations Eq. (65)-(66) in the case of point process observations, see [114], e.g., for application to neural spike trains. To summarize, in the case of approximate state estimation, $Y(t)$ in these equations is doubly stochastic with the $(\mathcal{F}_t^{YZ}, \mathbb{P})$-intensity $\lambda(t)$.

Building on this result, we repurpose the existing Eqs. (65) - (66) for an approximate marginal description in the following, which we regard as the core idea of this section. Namely, we define the probability measure $\mathbb{Q}$, such that $Y(t)$ is self-exciting with $(\mathcal{F}_t^Y, \mathbb{Q})$-intensity

$$\hat{\lambda}_j^{\mathrm{OL}}(t) := \begin{cases} a_j(X(t-), \hat{Z}(t-)), & a_j(X(t-), \hat{Z}(t-)) > 0 \\ 0, & \text{else.} \end{cases} \qquad (67)$$

Here, $X(t)$ is as in Eq. (2) and $\hat{Z}(t)$ is as in Eq. (65). The idea is that, firstly, sampling $Y(t)$ from $\mathbb{Q}$ only requires the first- and second-order statistics of the environment. In experiments, these might

be the quantities that can be robustly estimated. Secondly, the comparison between $\mathbb{P}$ and $\mathbb{Q}$ allows a classification of stochastic properties as resulting from the linear dynamic evolution of the intensity. Finally, when the environment is high-dimensional, continuous or requiring the simulation of many reaction events, sampling from $\mathbb{Q}$ can be simpler than from $\mathbb{P}$, while $\mathbb{P}$ and $\mathbb{Q}$ are, hopefully, sufficiently close. Since $\mathbb{Q}$ only takes into account the first- and second-order moments of $Z(t)$, we anticipate that statements about $\mathbb{P}$ and $\mathbb{Q}$ being close are confined to the first- and second-order moments of $Y(t)$ as well. We provide structural results for the first- and second-order comparison, section 4.1.3, and empirical findings in sections 8.1, 8.2 and 8.4. The empirical findings focus on the deviation of $\mathbb{P}$ and $\mathbb{Q}$. They reveal what stochastic properties of $Y(t)$ must be classified as consequences of the non-linear dynamic evolution of the intensity and which persist under the linearized dynamic evolution of the intensity.

On a technical note, we chose here to compare two counting processes by introducing one measure for each. The reader might be more familiar with introducing a second process, denoting it by $Y^{\mathrm{OL}}(t)$ say, and comparing $Y(t)$ and $Y^{\mathrm{OL}}(t)$ under $\mathbb{P}$. Those two perspectives are equivalent. After defining $\mathbb{Q}$ we emphasize the difference between approximate state estimation and approximate marginal simulation. State estimation considers $\hat{Z}(z)$ in Eq. (65) under $\mathbb{P}$, while marginal simulation considers $\hat{Z}(t)$ under $\mathbb{Q}$.

Before we present the results, we assume (C4) to establish a link to the (multivariate) Hawkes process with exponential kernel, introduced in the subsection 4.1. At stationarity under $\mathbb{Q}$, Eq. (66) has equilibrated and the time-dependent $B(t)$ can be replaced by the constant asymptotic value $\bar{B} = B(\infty)$. Then in integral form

$$\tilde{Z}(t) = \int_0^t e^{-(A + \bar{B}C^T DC)(t-u)} \bar{B} C^T D \; \mathrm{d}Y(u) - \int_0^t e^{-(A + \bar{B}C^T DC)(t-u)} \bar{B} C^T D \mu \; \mathrm{d}u,$$

with $D := \mathrm{diag}^{-1}(\mu)$. In case $(A + \bar{B}C^T DC)$ is invertible, the second term equilibrates to $-(A + \bar{B}C^T DC)^{-1} \bar{B} C^T D\mu$. Let $\hat{Z}^{\mathrm{H}}(t)$ be the linear filter with both $B(t)$ and the second term equilibrated, then $\hat{\lambda}^{\mathrm{H}}(t) = \mu + C\hat{Z}^{\mathrm{H}}(t)$ is the $(\mathcal{F}_t^Y, \mathbb{Q}^{\mathrm{H}})$-intensity of a (multivariate) Hawkes process $Y(t)$ with

$$\hat{\lambda}^{\mathrm{H}}(t) = \bar{\lambda} + \int_0^t C e^{-(A + \bar{B}C^T DC)(t-u)} \bar{B} C^T D \; \mathrm{d}Y(u) \tag{68}$$

where (using the Sherman–Morrison–Woodbury formula)

$$\begin{aligned} \bar{\lambda} &= (\mathbf{I} - C(A + \bar{B}C^T DC)^{-1} \bar{B} C^T D)\mu = (\mathbf{I} - \Gamma D + \Gamma(\mathrm{diag}(\mu) + \Gamma)^{-1} \Gamma D)\mu \\ &= (\mathbf{I} + \Gamma D)^{-1}\mu, \; \Gamma = CA^{-1}\bar{B}C^T. \end{aligned}$$

The equilibrated scenario can be made rigorous by assuming that the process is defined on the entire real axis and is at stationarity for all time t, provided that $C(A + \bar{B}C^T DC)^{-1} \bar{B} C^T D = \Gamma(\mathrm{diag}(\mu) + \Gamma)^{-1}$ has a spectral radius strictly less than 1. In this case, the lower integral bound in Eq. (68) is replaced by $-\infty$. We refer the reader to the examples in sections 8.1.1, 8.1.7, 8.2 and 8.2.3, where the usage of Eq. (68) is illustrated.

In the multivariate case, the main difference to the classical Hawkes process is the use of a matrix exponential instead of a weighted sum of one-dimensional exponential functions. The drawback of the matrix exponential is that we cannot guarantee a priori that the intensity remains positive. However, for the special case of $M = 1$ we derive a sufficient condition for a positive intensity at all times, see proposition 6.4 below. Overall, in the considered case studies, we did not encounter negative intensities. We now proceed with the time-dependent specification of $Y(t)$ with the $(\mathcal{F}_t^Y, \mathbb{Q})$-intensity as defined in Eq. (65)-(67), which we call the *tilted* version of the Hawkes process.

4.1.3 Structural results for the approximate linear marginal simulation

We have defined two different probability measures $\mathbb{P}$ and $\mathbb{Q}$. The measure $\mathbb{P}$ models the CRN in a random environment as a doubly stochastic process by means of an external process. The measure $\mathbb{Q}$ provides an approximate marginal description of the same CRN. We prove that the measure $\mathbb{Q}$ preserves the second-order properties in case the external process only modulates the zeroth-order reactions.

For the formulation of the results, we assume that

$$\hat{\lambda}_j^{\text{OL}}(t) > 0 \quad \mathbb{Q}\text{-a.s. for all } j = 1, \dots, M, t \geq 0. \tag{C5}$$

This condition implies that $\hat{\lambda}_j^{\text{OL}}(t) = a_j(X(t-), \hat{Z}(t-))$ $\mathbb{Q}$-almost surely for all $j = 1, \dots, M$ and $t \geq 0$. For convenience, we assume again $\mathbb{E}[Z(t)] \equiv 0$ for the external process.

Theorem 4.1. *Let $Y(t)$ be a multivariate counting process for which the $(\mathcal{F}_t^{YZ}, \mathbb{P})$-intensity and the external process $Z(t)$ satisfy the standard conditions (C1)-(C4), assuming A in Eq. (C2) is invertible. Furthermore, let the $(\mathcal{F}_t^Y, \mathbb{Q})$-intensity be given by Eq. (65)-(67) and (C4)-(C5). Firstly, for all $t \geq 0$, we have $\mathbb{E}_{\mathbb{P}}[Y(t)] = \mathbb{E}_{\mathbb{Q}}[Y(t)] = \mu t$. Secondly, for all $t, s \geq 0$*

$$\mathbb{C}\text{ov}_{\mathbb{P}}[Y(t), Y(s)] = \mathbb{C}\text{ov}_{\mathbb{Q}}[Y(t), Y(s)]$$

and, with $*$ indicating $\mathbb{P}$ or $\mathbb{Q}$, the expression is given in theorem 3.10.

Proof. For the first moment, we have, using (C4) and (C5) in Eq. (65), that

$$\mathbb{E}_{\mathbb{Q}}[\hat{Z}(t)] = \mathbb{E}_{\mathbb{Q}}\left[\int \exp(-A(t-u))B(u)C^T \operatorname{diag}\{\mu^{-1}\}\left\{\,\mathrm{d}Y(u) - \hat{\lambda}_j^{\text{OL}}(u)\,\mathrm{d}u\right\}\right] = 0$$

because a deterministic function is integrated with respect to the zero-mean $(\mathcal{F}_t^Y, \mathbb{Q})$-martingale $M(t) = Y(u) - \int_0^t \hat{\lambda}^{\text{OL}}(u)\,\mathrm{d}u$. The conditions (C3)-(C5) imply

$$Y(t) = M(t) + \int_0^t \hat{\lambda}^{\text{OL}}(u)\,\mathrm{d}u = M(t) + \int_0^t \mu + C\hat{Z}(u)\,\mathrm{d}u, \tag{69}$$

from which we obtain

$$\mathbb{E}_{\mathbb{Q}}[Y(t)] = \int_0^t \mu + C\mathbb{E}_{\mathbb{Q}}[\hat{Z}(u)]\,\mathrm{d}u = \mu t = \mathbb{E}_{\mathbb{P}}[Y(t)]. \tag{70}$$

Turning towards the covariance, the left-hand side was derived in theorem 3.10. We compute the right-hand side with martingale techniques, again dropping the subscript $\mathbb{Q}$ and see that the expressions agree. The martingale form of the optimal linear filter is

$$\hat{Z}(t) = \int_0^t \exp(-A(t-u))B(u)C^T \operatorname{diag}^{-1}(\mu)\,\mathrm{d}M(u)$$

First note that the cumulative centered intensity reads

$$\int_0^t \hat{Z}(u)\,\mathrm{d}u = \int_0^t \int_0^s \exp(-A(s-u))B(u)C^T \operatorname{diag}^{-1}(\mu)\,\mathrm{d}M(u)\,\mathrm{d}s$$

$$= \int_0^t \left(\int_u^t \exp(-A(s-u))\,\mathrm{d}s\right) B(u)C^T \operatorname{diag}^{-1}(\mu)\,\mathrm{d}M(u)$$

$$= A^{-1}\int_0^t (\mathbf{I} - \exp(-A(t-u)))B(u)C^T \operatorname{diag}^{-1}(\mu)\,\mathrm{d}M(u). \tag{71}$$

We insert Eq. (69)-(71) into the right-hand side as follows,

$$\mathbb{E}[(Y(t) - \mathbb{E}[Y(t)])(Y(s) - \mathbb{E}[Y(s)])^T]$$

$$=\mathbb{E}\left[\left(M(t) + C\int_0^t \hat{Z}(u)\ \mathrm{d}u\right)\left(M(s)^T + \int_0^s \hat{Z}(u)^T\ \mathrm{d}u\ C^T\right)\right]$$

$$=\mathbb{E}\left[\left(\int_0^t \mathbf{I} + CA^{-1}(\mathbf{I} - \exp(-A(t-u)))B(u)C^T \operatorname{diag}^{-1}(\mu)\ \mathrm{d}M(u)\right) \times \right.$$

$$\left.\left(\int_0^s \mathbf{I} + CA^{-1}(\mathbf{I} - \exp(-A(s-u)))B(u)C^T \operatorname{diag}^{-1}(\mu)\ \mathrm{d}M(u)\right)^T\right]$$

$$=\int_0^s \left\{\mathbf{I} + CA^{-1}(\mathbf{I} - \exp(-A(t-u)))B(u)C^T \operatorname{diag}^{-1}(\mu)\right\} \operatorname{diag}(\mu) \times$$

$$\left\{\operatorname{diag}^{-1}(\mu)CB(u)(\mathbf{I} - \exp(-A^T(s-u)))A^{-T}C^T + \mathbf{I}\right\}\ \mathrm{d}u$$

$$=\int_0^s \mathbf{I}\operatorname{diag}(\mu)\mathbf{I}\ \mathrm{d}u$$

$$+ C\int_0^s A^{-1}(\mathbf{I} - \exp(-A(t-u)))B(u) + B(u)(\mathbf{I} - \exp(-A^T(s-u)))A^{-T}\ \mathrm{d}uC^T$$

$$+ CA^{-1}\int_0^s (\mathbf{I} - \exp(-A(t-u)))B(u)C^T \operatorname{diag}^{-1}(\mu)CB(u) \times$$

$$(\mathbf{I} - \exp(-A^T(s-u)))\ \mathrm{d}uA^{-T}C^T =: \operatorname{diag}(\mu)s + J_1 + J_2.$$

We used Eq. (5) in the third equality and continue with the main trick of the proof, namely recognizing the ODE Eq. (66) in the integrand,

$$J_2 = CA^{-1}\int_0^s (\mathbf{I} - \exp(-A(t-u)))\left\{A\Sigma + \Sigma A^T - AB(u) - B(u)A^T - \frac{\mathrm{d}}{\mathrm{d}u}B(u)\right\} \times$$

$$(\mathbf{I} - \exp(-A^T(s-u)))\ \mathrm{d}uA^{-T}C^T$$

$$= - J_1 + CA^{-1}\left\{\int_0^s \frac{\mathrm{d}}{\mathrm{d}u}\left[(\mathbf{I} - \exp(-A(t-u)))(\Sigma - B(u))(\mathbf{I} - \exp(-A^T(s-u)))\right]\right.$$

$$+ (A\Sigma + \Sigma A^T) - A\Sigma\exp(-A^T(s-u)) - \exp(-A(t-u))\Sigma A^T\ \mathrm{d}u\Big\}A^{-T}C^T$$

$$= - J_1 + C(A^{-1}\Sigma + \Sigma A^{-T})C^T s$$

$$+ C\left\{\Sigma(\exp(-A^T s) - \mathbf{I})A^{-2T} + A^{-2}(\exp(-At) - \exp(-A(t-s)))\Sigma\right\}C^T$$

$$\square$$

We next generalize the result to linear reaction networks in random environment, where the environment modulates the zeroth-order reactions only. To this end, we first extend the theorem 4.1 in the following corollary.

Corollary 4.2. *Let* $[0, t], u \mapsto f(u)$ *and* $[0, s], u \mapsto g(u)$ *be continuous and deterministic functions, then*

$$\mathbb{C}\mathrm{ov}_{\mathbb{P}}\left[\int_0^t f(u)\ \mathrm{d}Y(u), \int_0^s g(u)\ \mathrm{d}Y(u)\right] = \mathbb{C}\mathrm{ov}_{\mathbb{Q}}\left[\int_0^t f(u)\ \mathrm{d}Y(u), \int_0^s g(u)\ \mathrm{d}Y(u)\right]$$

Proof. Define the (matrix-valued and signed) measure $\mathcal{C}_{\mathbb{P}}$ on $\mathbb{R}_{\geq 0}^2$ by

$$\mathcal{C}_{\mathbb{P}}([t_1, t_2] \times [s_1, s_2]) = \mathbb{C}\mathrm{ov}_{\mathbb{P}}[Y(t_2) - Y(t_1), Y(s_2) - Y(s_1)]$$

and analogously $\mathcal{C}_{\mathbb{Q}}$. Then from theorem 4.1 it follows by expanding the term on the right-hand side that $\mathcal{C}_{\mathbb{P}} = \mathcal{C}_{\mathbb{Q}}$ on the rectangles $[t_1, t_2] \times [s_1, s_2]$. Since those form a Π-system generating the Borel-sigma-algebra on $\mathbb{R}_{\geq 0}^2$, the measures $\mathcal{C}_{\mathbb{P}}$ and $\mathcal{C}_{\mathbb{Q}}$ agree on $\mathbb{R}_{\geq 0}^2$. Then the claimed equality holds because the terms can be expressed as

$$\int \mathbb{1}_{[0,t]}f(\tau)\mathbb{1}_{[0,s]}g(\sigma)^T\ \mathrm{d}\mathcal{C}_*(\tau, \sigma),$$

with the $*$ indicating $\mathbb{P}$ or $\mathbb{Q}$, respectively. □

Remark 4.3. The measure $\mathcal{C}_*$ employed in the proof goes by the name of covariance measure in the literature, see [115, §9.5, Eq. (9.5.12)] for a principled introduction. The covariance density and complete covariance density for point processes had been introduced already at an informal basis [116] before Brémaud's rigorous definition of point processes with stochastic intensity. The measure $\mathcal{C}_*$ has the covariance density (with respect to the Lebesgue-measure) $Ce^{A(t-s)}\Sigma C^T$, and the complete covariance density $Ce^{A(t-s)}\Sigma C^T + \delta(t-s)\mu$. The theorem 4.1 and corollary 4.2 can thus be summarized by the observation that the doubly stochastic (multivariate) Poisson process and the (multivariate) Hawkes process have the same complete covariance density. This link was already stated in the original paper by Hawkes [104] for the univariate and the diagonalizable (orthogonal) multivariate case with exponential kernel and constant jumps at stationarity. To the best of our knowledge, the tilted version with time-dependent jumps that corresponds to the stationary doubly stochastic case, but starting at $t = 0$, is new. This offers a solution to the problem of simulating a stationary Hawkes process even when the past is not available to avoid the transient burn-in phase.

Let now $X(t) = [X_1(t), \ldots, X_d(t)]^T$ be as in section 3.4, Eq. (34)-(37). We consider two probability measures $\mathbb{P}$ and $\mathbb{Q}$ for the process $X(t)$. Let $Z(t)$ be an external process satisfying (C1) and (C2) with respect to $\mathbb{P}$. Define the filtrations

$$\mathcal{F}_t^{YZ} := \sigma(\{Y(s) : s \leq t\} \cup \{\tilde{Y}(s) : s \leq t\} \cup \{Z(s) : s \leq t\} \cup \{X(0)\})$$
$$\mathcal{F}_t^Y := \sigma(\{Y(s) : s \leq t\} \cup \{\tilde{Y}(s) : s \leq t\} \cup \{X(0)\}).$$

The $(\mathcal{F}_t^{YZ}, \mathbb{P})$-intensity of the components $Y(t)$ of the joint counting process $[Y(t)^T, \tilde{Y}(t)^T]^T$ are assumed as in (C3) and (C4). The $(\mathcal{F}_t^Y, \mathbb{Q})$-intensity of the components $Y(t)$ are assumed as in Eq. (65)-(67) with (C4). Furthermore, assume (C5) holds for all $j = 1, \ldots, m$ $\mathbb{Q}$-almost surely. We recall that the $\mathcal{F}_t^Y$-intensity of the components $\tilde{Y}(t)$ is set to be

$$\tilde{\lambda}(t) = \Lambda X(t-) + b, \quad \Lambda \in \mathbb{R}^{(M-m) \times d}, b \in \mathbb{R}^{M-m},$$

which is assumed to hold for both measures $\mathbb{P}$ and $\mathbb{Q}$.

Theorem 4.4. *With the probability measures $\mathbb{P}$ and $\mathbb{Q}$ as just defined and assuming $X(0)$ has the same distribution for $\mathbb{P}$ and $\mathbb{Q}$, it holds that*

$$\mathbb{C}\text{ov}_\mathbb{P}[X(t), X(s)] = \mathbb{C}\text{ov}_\mathbb{Q}[X(t), X(s)].$$

Proof. Without loss of generality, we assume that $\mathbb{E}_\mathbb{P}[Z(t)] \equiv 0$, such that $\mathbb{E}_\mathbb{P}[dY(t)] = \mathbb{E}_\mathbb{Q}[dY(t)] = \mu$. By $\mathbb{E}_*$, we indicate that either $\mathbb{P}$ or $\mathbb{Q}$ makes the statement correct. We decompose the differential form of Eq. (36) via the martingale increment $d\tilde{M}(t) = d\tilde{Y}(t) - \tilde{\lambda}(t) dt$

$$dX(t) = N_1 dY(t) + N_2 d\tilde{Y}(t) = N_1 dY(t) + N_2\Lambda X(t) dt + N_2 b dt + N_2 d\tilde{M}(t) \tag{72}$$

In the remainder of the proof, for technical reasons, we consider Eq. (72) with the $\mathcal{F}_t^{Y\infty}$-martingale $\tilde{M}$, where
$$\mathcal{F}_t^{Y\infty} := \sigma(\{Y(s) : s \geq 0\} \cup \{\tilde{Y}(s) : s \leq t\} \cup \{X(0)\}).$$

The reader can think of this filtration as corresponding to first sampling the entire trajectory $Y(t), t \geq 0$, and conditional on it sampling $\tilde{Y}(t)$. Since $\mathcal{F}_t^{Y\infty} \supseteq \mathcal{F}_t^Y$, the $\mathcal{F}_t^{Y\infty}$-intensity is also $\tilde{\lambda}(t)$. From Eq. (72), in integral form it holds with $\Gamma := N_2\Lambda$

$$X(t) = e^{\Gamma t} X(0) + \int_0^t e^{\Gamma(t-u)} N_1 dY(u) + \int_0^t e^{\Gamma(t-u)} N_2(b \, du + d\tilde{M}(u))$$

from which we conclude

$$\mathbb{E}_\mathbb{P}[X(t)] = e^{\Gamma t} \mathbb{E}_*[X(0)] + \int_0^t e^{\Gamma(t-u)} N_1\mu \, du + \int_0^t e^{\Gamma(t-u)} N_2 b \, du = \mathbb{E}_\mathbb{Q}[X(t)]. \tag{73}$$

46

Then for the deviation of the mean $\Delta X(t) := X(t) - \mathbb{E}_*[X(t)]$ both with respect to $\mathbb{P}$ and $\mathbb{Q}$

$$\Delta X(t) = e^{\Gamma t} \Delta X(0) + \int_0^t e^{\Gamma(t-u)} N_1(\ \mathrm{d}Y(u) - \mu \ \mathrm{d}u) + \int_0^t e^{\Gamma(t-u)} N_2 \ \mathrm{d}\tilde{M}(u).$$

In order to evaluate $\mathbb{E}_*[\Delta X(t)\Delta X(s)^T]$, we first reassure that cross terms vanish. Cross terms with $\tilde{M}(u)$ are integrals of $\mathcal{F}_t^{Y\infty}$-predictable integrands with respect to the zero-mean martingale $\tilde{M}(u)$, which vanish by Eq. (4). For this reason, we defined the filtration of $\tilde{M}(u)$ to be $\mathcal{F}_t^{Y\infty}$. The cross term with $\Delta X(0)$ and $Y(u)$ vanish because the increments of $Y(t)$ are independent of the $\Delta X(0)$ by assumption (C4). In total, we obtain

$$\mathbb{C}\mathrm{ov}_*[X(t), X(s)] = \mathbb{E}_*[\Delta X(t)\Delta X(s)^T] = e^{\Gamma t} \mathbb{C}\mathrm{ov}_*[X(0)]e^{\Gamma^T s} \tag{74}$$

$$+ \mathbb{C}\mathrm{ov}_* \left[\int_0^t e^{\Gamma(t-u)} N_1 \ \mathrm{d}Y(u), \int_0^s e^{\Gamma(s-u)} N_1 \ \mathrm{d}Y(u) \right] \tag{75}$$

$$+ \int_0^s e^{\Gamma(t-u)} N_2 \ \mathrm{diag}(\Lambda \mathbb{E}_*[X(t)] + b) N_2^T e^{\Gamma^T(s-u)} \ \mathrm{d}u \tag{76}$$

by the application of Eq. (4) and Eq. (5). All three terms agree for the measures $\mathbb{P}$ and $\mathbb{Q}$, the first one by assumption, the second one by the corollary 4.2 and the third one by Eq. (73) □

In section 4.1.2 we reviewed the optimal linear filter obtained from a projection method. We now present a result that characterizes it as a moment closure.

Theorem 4.5. *Suppose under a probability measure $\mathbb{P}$, the $(\mathcal{F}_t^Y, \mathbb{P})$-intensity $\hat{\lambda}(t)$ of $Y(t)$, strictly positive in all components for all $t \geq 0$, is given by means of a predictable process $V(t) \in \mathbb{R}^{l \times l}$*

$$\mathrm{d}\hat{Z}(t) = - A\hat{Z}(t) \ \mathrm{d}t + V(t-)C^T \ \mathrm{diag}^{-1}\{\hat{\lambda}(t-)\}\{ \ \mathrm{d}Y(t) - \hat{\lambda}(t) \ \mathrm{d}t\}, \tag{77}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}_\mathbb{P}[V(t)] = A\Sigma + \Sigma A^T - A\mathbb{E}_\mathbb{P}[V(t)] - \mathbb{E}_\mathbb{P}[V(t)]A^T - \mathbb{E}_\mathbb{P}\left[V(t)C^T \ \mathrm{diag}^{-1}\{\hat{\lambda}(t)\}CV(t)\right], \tag{78}$$

$$\hat{\lambda}(t) = \mu + C\hat{Z}(t-), \tag{79}$$

initialized in $\hat{Z}(0) = 0, V(t) = \Sigma$. Then the following statements are equivalent.

(i) The process $\left(V(t)C^T \ \mathrm{diag}^{-1}\{\hat{\lambda}(t)\}\right)_t$ is deterministic.

(ii) The moment closure holds for all $t \geq 0$

$$\mathbb{E}_\mathbb{P}\left[V(t)C^T \ \mathrm{diag}^{-1}\{\hat{\lambda}(t)\}CV(t)\right] = \mathbb{E}_\mathbb{P}[V(t)]C^T \ \mathrm{diag}^{-1}\{\mu\}C\mathbb{E}_\mathbb{P}[V(t)]. \tag{80}$$

Furthermore either one implies that $(\hat{Z}(t), \mathbb{E}_\mathbb{P}[V(t)])$ are equal to $(\hat{Z}(t), B(t))$ in Eq. (65)-(66).

Remark: Note that Snyder's exact filter, Eq. (64), has the form for the appropriate F and G, Eq. (63), when using that in the evolution equation for $V_{ik}(t)$, the expectation of the second term is zero by Eq. (4). For the linear reaction networks as well as the CIR process, F and G are in the appropriate forms to match.

For the proof, we need the following generalized Cauchy-Schwarz inequality.

Lemma 4.6. *Let $X_i, i = 1, \ldots, m$ be strictly positive random variables and $Y = (Y_1, \ldots, Y_m)$ be a random vector. Then*

$$\mathbb{E}[Y \ \mathrm{diag}(X)]\mathbb{E}[\mathrm{diag}^2(X)]^{-1}\mathbb{E}[\mathrm{diag}(X)Y^T] \leq \mathbb{E}[YY^T]. \tag{81}$$

Furthermore, equality holds if and only if for every $i = 1, \ldots, m$ there exists a deterministic scalar α_i, such that $Y_i = \alpha_i X_i$.

Proof.

$$\mathbb{E}[Y \operatorname{diag}(X)]\mathbb{E}[\operatorname{diag}^2(X)]^{-1}\mathbb{E}[\operatorname{diag}(X)Y^T] = \sum_{i=1}^{m} \frac{\mathbb{E}[Y_i X_i]^2}{\mathbb{E}[X_i^2]} \leq \sum_{i=1}^{m} \mathbb{E}[Y_i^2] = \mathbb{E}[YY^T]$$

by the term-wise application of the classical Cauchy-Schwarz inequality. The additional statement holds because the sums are equal if and only if term-wise equality holds. □

Proof (theorem 4.5). $(i) \Rightarrow (ii)$. Denote $\beta(t) := V(t)C^T \operatorname{diag}^{-1}(\hat{\lambda}(t))$. Then we compute

$$\mathbb{E}[V(t)]C^T = \mathbb{E}[V(t)C^T] = \mathbb{E}[\beta(t)\operatorname{diag}(\hat{\lambda}(t))] = \beta(t)\operatorname{diag}(\mu) \tag{82}$$

and we obtain

$$\mathbb{E}[V(t)C^T \operatorname{diag}^{-1}(\hat{\lambda}(t))CV(t)] = \beta(t)C\mathbb{E}[V(t)] = \mathbb{E}[V(t)]C^T \operatorname{diag}^{-1}(\mu)C\mathbb{E}[V(t)].$$

$(ii) \Rightarrow (i)$. For each $k = 1, \ldots, n$ apply the lemma for the choice $Y = (V(t)C^T \operatorname{diag}^{-\frac{1}{2}}(\hat{\lambda}(t)))_{k*}$ denoting the k-th row, and $X_i = \sqrt{\hat{\lambda}_i(t)}$. Then the equality of the k-th diagonal entry in Eq. (80) reads as Eq. (81). Consequently, we obtain deterministic scalars $\alpha_{ki}(t)$ from the lemma that satisfy

$$(V(t)C^T \operatorname{diag}^{-\frac{1}{2}}(\hat{\lambda}(t)))_{ki} = \alpha_{ki}(t)\sqrt{\hat{\lambda}_i(t)}.$$

We recognize the right-hand side as the ki-entry of $\alpha(t)\operatorname{diag}^{\frac{1}{2}}(\hat{\lambda}(t))$. Then in matrix notation,

$$V(t)C^T \operatorname{diag}^{-\frac{1}{2}}(\hat{\lambda}(t)) = \alpha(t)\operatorname{diag}^{\frac{1}{2}}(\hat{\lambda}(t))).$$

Upon multiplication with $\operatorname{diag}^{-\frac{1}{2}}(\hat{\lambda}(t))$ (ii) follows.
Let us assume that (i) and (ii) hold. The Eq. (82) implies $\beta(t) = \mathbb{E}[V(t)]C^T \operatorname{diag}^{-1}(\mu)$ and together with Eq. (80), this yields Eq. (65)-(66) for $B(t) = \mathbb{E}[V(t)]$. □

In the special case of $l = 1$, the property (i) more simply states that $V(t)/\hat{\lambda}_j(t)$ is deterministic for $j = 1, \ldots, M$. For the modeler who asks for which environment $Z(t)$ to use the Hawkes approximation the property (i) can be a more intuitive approach than the process Eqs (65)-(66). Furthermore, note that the property (ii) is not a closure scheme on the posterior moments, which are random processes. It is a closure scheme on the moments of posterior moments. Therefore, it is remarkable that this closure scheme already enforces the Hawkes approximate filter. Finally, this characterization sheds light on the relation between different approximate filters. Two closure schemes were discussed in [3] for the CIR modulated birth process, the second-order posterior moment closure expressing the third-order posterior moment as $S(t) = 2V(t)/\Theta(t)$ on the one hand and the one-dimensional Eq. (80), $\mathbb{E}[V(t)^2/\Theta(t)] = \mathbb{E}[V(t)]^2/\mathbb{E}[\Theta(t)]$, on the other hand. Both were conjectured to induce the Gamma filter [3], an assumed density filter, characterized by a two-dimensional evolution equation of mimics of the first- and second-order posterior moments. However, this statement reveals that both are different. The first one induces the Gamma filter and the second one the Hawkes filter.

4.1.4 Auxiliary concepts: Dirac-PDMPs and the backward recurrence time process

Model reduction via conditional expectations has different goals. So far, in the path via conditioning on X that uses the conditional intensity, Eq. (57), we have focused on marginal simulation and attribution of properties in the subnetwork to properties of the environment. But the model reduction via conditioning also aims at new master equations and at the computation of information measures. The goal of this section is to prepare the joint master equation, see Eqs. (104) - (105), by using PDMPs. We recall that PDMPs are defined by their three characteristics, the flow, the hazard and the jump update. In the Snyder filter, proposition 2.10, the jump update is deterministic, which motivates us to confine $Q(\cdot, \theta)$ to a Dirac measure for each $\theta \in \vartheta$. Also in the optimal linear mimic, Eq. (65)-(66), we faced a deterministic jump update. We introduce the following non-standard notion of Dirac-PDMPs for a systematic approach.

Definition 4.7 (Dirac-PDMP)**.** Call a piecewise-deterministic Markov process $(\Theta_t)_{t \geq 0}$ a **Dirac-PDMP** on the state space ϑ if its third local characteristic is $Q(\cdot, \theta) = \delta_{f(\theta)}(\cdot)$ for a measurable function $f \colon \vartheta \to \vartheta$. For the Dirac-PDMP we specify the local characteristics by the triple (F, l, f).

We present the probability evolution equation for the Dirac-PDMP in proposition 4.20 below. Analogously to stochastic differential equations, the process equations of the Dirac-PDMP can be written as

$$\mathrm{d}\Theta_t = F(\Theta_t)\,\mathrm{d}t + (f(\Theta_{t-}) - \Theta_{t-})\,\mathrm{d}Y_t, \tag{83}$$

where $(Y_t)_{t \geq 0}$ is a counting process with the CI $\hat{\lambda}_t = l(\Theta_{t-})$. The counting process $(Y_t)_{t \geq 0}$ is precisely the *embedded counting process* that counts the jumps of the PDMP. For the Snyder filter, we have already used Eq. (83) in Eq. (8). We refer to the first term on the right-hand side as the drift. Readers interested in the corresponding equation for PDMPs may consult [1, §1.5.4.2, p.42-45].

For later reference in the examples and the case studies we add a transformation rule for Dirac-PDMPs, which is the analogue to the Ito transformation rule of stochastic differential equations.

Proposition 4.8 (Ito rule for Dirac-PDMP)**.** *Let $(\Theta_t)_{t \geq 0}$ be a Dirac-PDMP on $\vartheta \subseteq \mathbb{R}^{n_0}$ with local characteristics (F, l, f) and $\Phi \colon \vartheta \to \mathbb{R}^m$ be a differentiable function. Then $(\Phi(\Theta_t))_{t \geq 0}$ is a Dirac-PDMP that follows the process equation*

$$\mathrm{d}\Phi(\Theta_t) = D\Phi(\Theta_t)F(\Theta_t)\,\mathrm{d}t + (\Phi(f(\Theta_{t-})) - \Phi(\Theta_{t-}))\,\mathrm{d}Y_t.$$

Proof. The equation follows from the chain rule of differentiation for the first term and insertion of $\Theta_t = f(\Theta_{t-})$ into $\Phi(\Theta_t) - \Phi(\Theta_{t-})$ if $(Y_t)_{t \geq 0}$ jumps at t for the second term. $\qquad\square$

We emphasize that our focus lies on the object $(Y_t)_{t \geq 0}$ in Eq. (83) and we study it via its CI $l(\Theta_{t-})$. Hence, let us confine the class of counting processes that we investigate in this work.

Definition 4.9 ((F, l, f)-counting process)**.** Let $(Y_t)_{t \geq 0}$ be a counting process. We call it an (F, l, f)-**counting process**, if its finite-dimensional marginal distributions agree with the embedded counting process of a Dirac-PDMP with local characteristics (F, l, f).

Proposition 4.10. *Markov-modulated Poisson processes as defined in definition 2.8 belong to the class of (F, l, f)-counting processes.*

Proof. Let $(Z_t)_{t \geq 0}$ be a CTMC with generator $\mathcal{A}$ and $(Y_t)_{t \geq 0}$ be a Markov-modulated Poisson process, modulated by $(Z_t)_{t \geq 0}$ via $\lambda(Z_t)$. Then according to proposition 2.10 the filtering distribution is a Dirac-PDMP on $\vartheta = \Delta = \{\pi \in [0,1]^{|\mathcal{Z}|} : \sum_{z \in \mathcal{Z}} \pi(z) = 1\}$ with

- $F \colon \vartheta \to \mathbb{R}^{|\mathcal{Z}|}, \pi \mapsto \mathcal{A}\pi - (\mathrm{diag}(\lambda(z)_{z \in \mathcal{Z}}) - l(\pi)\,\mathbf{I}_{|\mathcal{Z}|})\pi$,

- $l \colon \vartheta \to \mathbb{R}_{\geq 0}, l(\pi) := \sum_{z \in \mathcal{Z}} \lambda(z)\pi(z)$,

- $f \colon \vartheta \to \vartheta, \pi \mapsto \mathrm{diag}(\lambda(z)_{z \in \mathcal{Z}})\pi/l(\pi)$.

Since l is bounded by the maximum of λ on $\mathcal{Z}$, the integrability condition Eq. (290), Appendix A, is satisfied. The jump times of $(\Pi_t)_{t \geq 0}$ coincide with the jump times of $(Y_t)_{t \geq 0}$, hence $(Y_t)_{t \geq 0}$ is the embedded counting process of the Dirac-PDMP $(\Pi_t)_{t \geq 0}$. $\qquad\square$

The fact that the Snyder filter is a PDMP has been observed before, e.g., in [115, Example 10.3(e), p.101-103] and [117]. For the particle interpretation of this fact we refer to [118]. The analogous results for the filter with Gaussian instead of Poissonian observations has received much more attention [79, 119, 120].

Proposition 4.10 shows that the class of (F, l, f)-counting processes contains the MMPPs as an interesting subclass. The typical alternative way to obtain an (F, l, f)-counting process is to specify its CI directly via the process equation (83). We sometimes refer to counting processes obtained this way as self-exciting (SE) counting process, even though MMPPs are also self-exciting. The reader may recall the two examples for the class SE that we have seen already. One example is the Hawkes process, 4.1. The second one is the approximate marginal simulation via the optimal linear filter. More generally, the class of Dirac-PDMPs contains the approximate filters as a subclass. The approximate

filters attempt to mimic the structure of the Snyder filter, proposition 2.10 and Eq. (64). As opposed to the Snyder filter with its unclosed hierarchy of conditional moments, the approximate filters use a description by finitely many sufficient state variables. For more approximate filters, see sections 8.1.4, 8.2 and 8.2.3 below.

In both cases (MMPP and SE) we refer to the particular corresponding process $(\Theta_t)_{t\geq 0}$ as the *sufficient state variables of joint Markovian progression*. For the MMPPs, the posterior probabilities are the sufficient state variables of joint Markovian progression. We constrain our analysis to the (F, l, f)-counting processes, which leaves us with studying $l(\Theta_{t-})$ for the Dirac-PDMP $(\Theta_t)_{t\geq 0}$.

We now introduce a standard form for the Dirac-PDMP, which has an especially simple F. It is needed for the formulation and proof of the corollary 4.23 and theorems 4.24 - 4.28, below. Readers mainly interested in the applications can skip this rather technical paragraph. We provide the relevant link to the definition 4.7 and the notation needed for subsequent paragraphs in proposition 4.13 and Eq. (84) - (85). The standard form comes with a notion of dimension for the (F, l, f)-counting process. The name backward recurrence time process (BReT-P) of the standard form originates from its first component, which has the natural interpretation as the time since the last jump.

Definition 4.11 (Backward recurrence time, BReT-P)**.** Let $(Y_t)_{t\geq 0}$ be a counting process with jump times $0 = \sigma_0 < \sigma_1 < \sigma_2 < \ldots$, then the process $(t - \sigma_{Y_t})_{t\geq 0}$ is the **backward recurrence time** [121, §2.1, p.27] of $(Y_t)_{t\geq 0}$.

Let $(\tau(t), \theta(t))_{t\geq 0}$ be a Dirac-PDMP with local characteristics (F, m, f) on the state space $\vartheta = [0, \infty) \times \mathcal{E}, \mathcal{E} \subseteq \mathbb{R}^n$. We call $(\tau(t), \theta(t))_{t\geq 0}$ a **backward recurrence time process (BReT-P)** if the first local characteristic F has the form $F(\tau, \theta) = [1, \mathbf{0}], \mathbf{0} \in \mathbb{R}^n$ and the third characteristic is $f(\tau, \theta) = [0, g(\tau, \theta)]$ for a measurable function $g\colon \vartheta \to \mathcal{E}$. The pair (m, g) summarizes the local characteristics of the BReT-P.

Proposition 4.12. *Let $(\tau(t), \theta(t))_{t\geq 0}$ be a BReT-P and $(Y_t)_{t\geq 0}$ its embedded counting process, then $(\tau(t))_{t\geq 0}$ is the backward recurrence time of $(Y_t)_{t\geq 0}$.*

Proof. Let $0 = \sigma_0 < \sigma_1 < \sigma_2 < \ldots$ be the jump times of $(Y_t)_{t\geq 0}$. By $\dot{\tau}(t) = 1$ for $t \in (\sigma_{i-1}, \sigma_i)$ and $\tau(\sigma_i) = 0$ as specified in definition 4.11, the process $\tau(t)$ is fixed to be $\tau(t) = t - \sigma_{Y_t}$, which is exactly the backward recurrence time of $(Y_t)_{t\geq 0}$. $\qquad\square$

The probability evolution equation for the BReT-P is derived in corollary 4.23 and the BReT-P is used in the proofs of theorems 4.24-4.28 to derive the fixed point equation for embedded Markov chain of the Dirac-PDMP.

Proposition 4.13. *Let $(\Theta_t)_{t\geq 0}$ be a Dirac-PDMP. Then $(\Theta_t)_{t\geq 0}$ induces a BReT-P with the same embedded counting process.*

Proof. Consider a Dirac-PDMP $(\Theta_t)_{t\geq 0}$ with local characteristics (F, l, f). Denote by $(Y_t)_{t\geq 0}$ its embedded counting process with jump times $0 = \sigma_0 < \sigma_1 < \sigma_2 < \ldots$ as in definition 2.1. Denote by

$$u\colon [0, \infty) \times \vartheta \to \vartheta, \tag{84}$$

the function, such that $\tau \mapsto u(\tau, \theta_0)$ is the solution of the initial value problem $\dot{\Theta}_t = F(\Theta_t), \Theta_0 = \theta_0$. Then define

$$\begin{aligned} \tau(t) &:= t - \sigma_{Y_t}, \theta(t) := \Theta_{\xi(t)}, \\ \xi(t) &:= \sigma_{Y_t} = t - \tau(t), m := l \circ u, g := f \circ u. \end{aligned} \tag{85}$$

Then for $t \in (\sigma_{i-1}, \sigma_i) : \dot{\tau}(t) = 1$ and $\dot{\theta}(t) = \mathbf{0}$. And $\tau(\sigma_i) = \sigma_i - \sigma_i = 0$. Hence, $(\tau(t), \theta(t))_{t\geq 0}$ is a BReT-P with local characteristics (m, g). Since the jump times remain unchanged, $(\Theta_t)_{t\geq 0}$ and $(\tau(t), \theta(t))_{t\geq 0}$ have the same embedded counting process. $\qquad\square$

Definition 4.14 (BReT-P counting process)**.** Let $(Y_t)_{t\geq 0}$ be a counting process. We call it **backward recurrence time parametrized counting process (BReT-P counting process)**, if its finite-dimensional marginal distributions agree with the embedded counting process of a BReT-P $(\tau(t), \theta(t))_{t\geq 0}$, which is then called a **corresponding** backward recurrence time processes.

Corollary 4.15. *Every (F, l, f)-counting process is a BReT-P counting process.*

Hence, a study of BReT-P counting processes will be equivalent to our considered class of counting processes in definition 4.9. To summarize the definitions 4.11 and 4.14, we consider counting processes $(Y_t)_{t \geq 0}$ whose CI $(\hat{\lambda}_t)_{t \geq 0}$ is parametrized by a process $(\tau(t), \theta(t))_{t \geq 0}$ in the following form. The first (scalar) component τ is the backward recurrence time. The second component $\theta \in \mathcal{E} \subseteq \mathbb{R}^n$ is a (possibly multi-variate) auxiliary state variable, that possesses three properties

(A1) it is constant between jumps, i.e. $\theta(t) = \theta(t - \tau(t))$,

(A2) there is a measurable $m \colon [0, \infty) \times \mathcal{E} \to \mathbb{R}_{\geq 0}$, satisfying $\hat{\lambda}_t = m(\tau(t-), \theta(t-))$ and for each $\theta^0 \in \mathcal{E}$ there is an $\varepsilon(\theta^0) > 0$ such that

$$\int_0^{\varepsilon(\theta^0)} m(\tau, \theta^0) \, \mathrm{d}\tau < \infty,$$

compare Eq. (290), Appendix A,

(A3) there is a measurable $g \colon [0, \infty) \times \mathcal{E} \to \mathcal{E}$, satisfying $g(\tau(\sigma_i -), \theta(\sigma_i -)) = \theta(\sigma_i)$ at jump times σ_i of $(Y_t)_{t \geq 0}$.

For a BReT-P counting process $(Y_t)_{t \geq 0}$ we denote by $\mathcal{BP}$ the set of corresponding backward recurrence time processes.

Remark 4.16. The set $\mathcal{BP}$ has more than one element, because a new BReT-P $(\tau(t), \tilde{\theta}(t))_{t \geq 0}$ can be defined via $\tilde{\theta}(t) := [\theta(t), 0] \in \mathcal{E} \times \{0\}$ with $\tilde{m} = m \circ \Gamma$ and $\tilde{g} = \Sigma \circ g \circ \Gamma$ for the truncation $\Gamma \colon [0, \infty) \times \mathcal{E} \times \{0\} \to [0, \infty) \times \mathcal{E}, (\tau, \theta, 0) \mapsto (\tau, \theta)$ and the extension $\Sigma \colon \mathcal{E} \to \mathcal{E} \times \{0\}, \theta \mapsto (\theta, 0)$. This example extends the state space artificially. But in other examples the state space can be reduced. To illustrate this, consider a Dirac-PDMP $(\Theta_t)_{t \geq 0}$ in the coordinate form $\Theta_t = (V_1(t), \ldots, V_{n_0}(t)) \in \vartheta \subseteq \mathbb{R}^{n_0}$ with local characteristics (F, l, f). In the proof of proposition 4.13, this induced a BReT-P on $[0, \infty) \times \mathcal{E}, \mathcal{E} \subseteq \mathbb{R}^n$ with $n = n_0$. If components of the third local characteristic f are constant functions, then the dimension n of $\theta(t)$ can be reduced as we demonstrate now. Let us review the proof of proposition 4.13. Suppose that $n < n_0$ and $f_{n+1} \equiv v_{n+1}^0, \ldots, f_{n_0} \equiv v_{n_0}^0$ are constant, i.e., the values $V_{n+1}, \ldots, V_{n_0}$ are reset to the same values $v_{n+1}^0, \ldots, v_{n_0}^0$ at any jump. We now construct the process $(\tau(t), \theta(t))_{t \geq 0}$. For this purpose define the extension and truncation similarly as above via $\Sigma \colon \mathcal{E} \to \mathbb{R}^{n_0}, \Sigma(\theta) := (\theta, v_{n+1}^0, \ldots, v_{n_0}^0)$ and $\Gamma \colon \mathbb{R}^{n_0} \to \mathbb{R}^n, \Gamma(v_1, \ldots, v_n, v_{n+1}, \ldots, v_{n_0}) = (v_1, \ldots, v_n)$. Furthermore, suppose that $\Theta_0 = (\theta(0), v_{n+1}^0, \ldots, v_{n_0}^0)$. Define $\tau(t) := t - \sigma_{Y_t}, \theta(t) := \Gamma(\Theta_{\xi(t)}) \in \mathbb{R}^n, \xi(t) := \sigma_{Y_t} = t - \tau(t)$ as well as $m(\tau, \theta) = l \circ u(\tau, \Sigma(\theta)))$ and $g(\tau, \theta) = \Gamma \circ f(u(\tau, \Sigma(\theta)))$. Let $(\tilde{\tau}(t), \tilde{\theta}(t))_{t \geq 0}$ be the BReT-P, constructed in the proof of proposition 4.13, then $\tilde{\tau}(t) = \tau(t)$ and $\tilde{\theta}(t) = \Sigma(\theta(t))$ ω-wise, in particular, both have the same embedded counting process.

This remark prompts the notion of a dimension for the (F, l, f)-counting processes. It is, loosely speaking, the minimum dimension of $\theta(t)$ that we can reduce the BReT-P to.

Definition 4.17 (Dimension of (F, l, f)-counting process). Let $(Y_t)_{t \geq 0}$ be an (F, l, f)-counting process. Consider the set $\mathcal{D} \subseteq \mathbb{N}_{\geq 0}$ of natural numbers n, for which a BReT-P on $[0, \infty) \times \mathcal{E}, \mathcal{E} \subseteq \mathbb{R}^n$ is in $\mathcal{BP}$. Then the **dimension** of $(Y_t)_{t \geq 0}$, denoted by $\dim(Y_t)_{t \geq 0}$, is the minimum of $\mathcal{D}$.

Remark 4.18. For the MMPPs, the conservation of probability mass and zero-states, defined below, reduce the dimension.

(i) In Eq. (8) the evolution equation of the $\Pi_t(z)$ for one z can be replaced by the trivial evolution of $\sum_z \Pi_t(z)$. The value of this sum is constantly 1. Hence, the number n_0 of sufficient state variables of joint Markovian progression can be reduced to $|\mathcal{Z}| - 1$. A reparametrization may further decrease the number.

(ii) Call any $z \in \mathcal{Z}$ with $\lambda(z) = 0$ a *zero-state*. For zero-states, the reset value of the corresponding $\Pi_t(z)$ in Eq. (7) is 0, i.e., f is constant for these components. Hence, for the MMPP $(Y_t)_{t \geq 0}$ it holds

$$\dim(Y_t)_{t \geq 0} \leq |\{\mathcal{Z} \colon \lambda(z) > 0\}| - 1.$$

The remark 4.16 and the definition of the dimension might require to modify the initial state of the PDMP, or BReT-P. In the remark 4.16, the process was initialized with the constant reset values $\Theta_0 = (\theta(0), v_{n+1}^0, \ldots, v_{n_0}^0)$. Alternatively it is enough to initialize it in a state that can be reached from the above state by the flow. This is not restrictive if the focus lies on the asymptotic analysis and the process is ergodic, as discussed at the end of section 4.2.3 below.

4.2   Joint and marginal master equation using the conditional intensity

For the defined class of (F, l, f)-counting processes the rich Markov theory is now exploited which provides the stationary analysis through the ergodic property and the invariance of measures. We define a notion of stationarity that is tailored for our purpose. For more details on the stationarity of point processes we refer the reader to [115, §12.5, p.222-236].

Definition 4.19 (Asymptotic stationarity, asymptotic conditional intensity distribution). Let $(Y_t)_{t \geq 0}$ be a counting process with a conditional intensity. If its conditional intensity converges in distribution, then we call the counting process **asymptotically stationary**. We call the limiting distribution its **asymptotic conditional intensity distribution (ACID)**.

By stationary analysis, we mean the description of the asymptotic distribution via a stationarity condition for the invariant measure. By ergodicity we mean the property to converge in distribution to the unique invariant measure, independent of the initial distribution. The stationary analysis then provides access to information measures, see section 5.1, via the CI $\hat{\lambda}_t = l(\Theta_{t-}) = m(\tau(t-), \theta(t-))$ and its limiting distribution, the ACID.

4.2.1   Probability evolution equation and stationarity condition

Assuming the absolute continuity of the distribution, the stationary analysis is usually addressed via the evolution equation of the density, which is known as the master equation or differential form of the Chapman-Kolmogorov forward equation for Markov jump processes and as the Fokker-Planck equation for diffusion processes. For PDMPs, we analogously obtain an equation that combines the drift term of the Fokker-Planck equation with the jump terms of the master equation. It goes by different names in the literature. We adopt the names *differential Chapman-Kolmogorov equation* [63, §3.4] and *Liouville master equation* [1, §1.5].

Proposition 4.20 (Differential Chapman-Kolmogorov equation).

(i) *Let $(\Theta_t)_{t \geq 0}$ be a PDMP on the state space $\vartheta \subseteq \mathbb{R}^{n_0}$ with local characteristics (F, l, Q). Suppose that $Q(\cdot, \theta)$ is absolutely continuous and admits a density $\theta' \mapsto q(\theta', \theta)$ for every $\theta \in \vartheta$ and suppose that Θ_t is absolutely continuously distributed with density $\theta \mapsto p(t, \theta)$ for $t > 0$. Suppose that the partial derivatives of F exist. Then for θ in the interior of ϑ, it holds*

$$\partial_t p(t, \theta) = -\sum_{i=1}^{n_0} \partial_i(F_i(\theta)p(t, \theta)) + \int_\vartheta l(\theta')q(\theta, \theta')p(t, \theta') - l(\theta)q(\theta', \theta)p(t, \theta) \, \mathrm{d}\theta'. \qquad (86)$$

(ii) *Let $(\Theta_t)_{t \geq 0}$ be a Dirac-PDMP on the state space $\vartheta \subseteq \mathbb{R}^{n_0}$ with local characteristics (F, l, f). With $\mathring{\vartheta}$ denoting the interior of ϑ define $R := \operatorname{im} f \cap \mathring{\vartheta} \subseteq \vartheta$ and suppose that the restriction $f|_{f^{-1}(R)} \colon f^{-1}(R) \to R$ is a diffeomorphism with inverse $f_- \colon R \to f^{-1}(R)$. Suppose that Θ_t is absolutely continuously distributed with density $\theta \mapsto p(t, \theta)$ for $t > 0$. Suppose that the partial derivatives of F exist. Then for θ in the interior of ϑ, it holds*

$$\partial_t p(t, \theta) = -\sum_{i=1}^{n_0} \partial_i(F_i(\theta)p(t, \theta)) - l(\theta)p(t, \theta) + \mathbb{1}_R(\theta)|\det Df_-(\theta)|l(f_-(\theta))p(t, f_-(\theta)). \qquad (87)$$

Proof. For Eq. (87), according to [65, §26, Eq.(26.15), p.70 & §22, p.53], the generator of the Dirac-PDMP reads

$$\mathcal{A}^*\varphi(\theta) = \sum_{i=1}^{n_0} \partial_i\varphi(\theta)F_i(\theta) + l(\theta)(\varphi(f(\theta)) - \varphi(\theta)). \qquad (88)$$

Then for smooth test functions $\varphi\colon \vartheta \to \mathbb{R}$ that vanish at the boundary of ϑ, we have

$$
\partial_t \int_\vartheta \varphi(\theta) p(t, \theta) \, \mathrm{d}\theta = \partial_t \mathbb{E}[\varphi(\Theta_t)]
$$

$$
= \int_\vartheta \mathcal{A}^* \varphi(\theta) p(t, \theta) \, \mathrm{d}\theta
$$

$$
= \int_\vartheta \sum_{i=1}^{n_0} \partial_i \varphi(\theta) F_i(\theta) p(t, \theta) \, \mathrm{d}\theta - \int_\vartheta l(\theta) \varphi(\theta) p(t, \theta) \, \mathrm{d}\theta + \int_\vartheta l(\theta) \varphi(f(\theta)) p(t, \theta) \, \mathrm{d}\theta.
$$

For the first term

$$
\int_\vartheta \sum_{i=1}^{n_0} \partial_i \varphi(\theta) F_i(\theta) \, \mathrm{d}\theta = - \int_\vartheta \sum_{i=1}^{n_0} \varphi(\theta) \partial_i (F_i(\theta) p(t, \theta)) \, \mathrm{d}\theta
$$

with integration by parts and the assumption that φ vanishes on the boundary. By the assumptions we get that for all $\theta \in \vartheta \backslash f^{-1}(R)$, $f(z)$ lies on the boundary of ϑ and hence $\varphi(f(z)) = 0$. Consequently, for the third term the change of variables formula yields

$$
\int_\vartheta l(\theta) \varphi(f(\theta)) p(t, \theta) \, \mathrm{d}\theta
$$

$$
= \int_{f^{-1}(R)} l(\theta) \varphi(f(\theta)) p(t, \theta) \, \mathrm{d}\theta
$$

$$
= \int_R |\det Df_-(\theta)| l(f_-(\theta)) \varphi(\theta) p(t, f_-(\theta)) \, \mathrm{d}\theta.
$$

Then in summary,

$$
\int_\vartheta \varphi(\theta) \partial_t p(t, \theta) \, \mathrm{d}\theta = \int_\vartheta \varphi(\theta) \left[-\sum_{i=1}^{n_0} \partial_i (F_i(\theta) p(t, \theta)) \right.
$$

$$
\left. - l(\theta) p(t, \theta) + \mathbb{1}_R(\theta) |\det Df_-(\theta)| l(f_-(\theta)) p(t, f_-(\theta)) \right] \, \mathrm{d}\theta,
$$

which implies the claimed equation to hold for θ in the interior of ϑ. For Eq. (86), the proof works analogously starting with

$$
\mathcal{A}^* \varphi(\theta) = \sum_{i=1}^{n_0} \partial_i \varphi(\theta) F_i(\theta) + l(\theta) \int_\vartheta (\varphi(\theta') - \varphi(\theta)) q(\theta', \theta) \, \mathrm{d}\theta'
$$

in place of Eq. (88). We refer to [63, Eq. (3.4.22), p.50], where the Eq. (86) is stated in a more general form. We arrive at Eq. (86) by noting that in Eq. (3.4.1), p. 47 of [63] the function $W(x|z,t)$ factorize into $l(z)q(x,z)$ and in Eq. (3.4.3), p.47 the function B_{ij} vanishes. $\qquad \square$

We follow Petruccione and Breuer [1, §1.5, Eq.(1.150), p.33] in referring to the equations (86) and (87) as Liouville master equations. The Liouville equation describes the probability evolution of an ensemble of particles under a deterministic flow, referring here to the first term of the equation.

Remark 4.21. The equation (86) is linked to Eq. (87) informally via the delta distribution $q(\theta, \theta') = \delta(\theta - f(\theta'))$ and the composition rule of the delta distribution for the substitution $\theta' = f_-(u)$ which yields for the second term in Eq. (86)

$$
\int_{f^{-1}(R)} l(\theta') q(\theta, \theta') p(t, \theta) \, \mathrm{d}\theta'
$$

$$
= \int_{f_-(R)} l(\theta') \delta(\theta - f(\theta')) p(t, \theta) \, \mathrm{d}\theta'
$$

$$
= \int_R l(f_-(u)) \delta(\theta - f(f_-(u))) p(t, f_-(u)) |\det Df_-(\theta)| \, \mathrm{d}u
$$

$$
= \int_R l(f_-(u)) \delta(\theta - u) p(t, f_-(u)) |\det Df_-(\theta)| \, \mathrm{d}u
$$

$$
= \mathbb{1}_R(\theta) l(f_-(\theta)) p(t, f_-(\theta)) |\det Df_-(\theta)| \, \mathrm{d}u.
$$

In the one-dimensional case and with monotone f the proposition 4.20 (ii) can be reformulated.

Corollary 4.22. *Let $(\Theta_t)_{t \geq 0}$ be a Dirac-PDMP on the state space $\vartheta = (a, b]$ or $\vartheta = (a, \infty)$. For the second case, denote $b = \infty$. Let $(\Theta_t)_{t \geq 0}$ follow the stochastic evolution equation*

$$\mathrm{d}\Theta_t = F(\Theta_t)\,\mathrm{d}t + [f(\Theta_{t-}) - \Theta_{t-}]\,\mathrm{d}Y_t$$

and jumps of $(Y_t)_{t \geq 0}$ occur with intensity $t \mapsto l(\Theta_{t-})$ for $l: \vartheta \to \mathbb{R}_{>0}$, i.e., the local characteristics are (F, l, f). We assume there exists $\lambda_1 > 0$, such that $l \geq \lambda_1$. Suppose that $F: \vartheta \to \mathbb{R}$ is differentiable and strictly negative and that $f: \vartheta \to \vartheta$ is differentiable and strictly monotonically increasing. Let $f_-: \vartheta \to \vartheta \cup \{a\}$ be the inverse of f on $\mathrm{im}\, f$ and $f_-(\theta) = a$ for $\theta \notin \mathrm{im}\, f$. Then for a differentiable initial condition $p(0, \theta)$, the probability density evolves according to the PDE for all $t > 0, \theta \in \vartheta$

$$\partial_t p(t, \theta) = -\partial_\theta(F(\theta)p(t,\theta)) - l(\theta)p(t,\theta) + f'_-(\theta)l(f_-(\theta))p(t, f_-(\theta)), \tag{89}$$

where $f'_-(\theta)$ denotes the left-sided derivative of f_- in θ.

Proof. Eq. (89) follows by an application of the Poisson-Liouville equation, proposition 4.20, Eq. (87) by using $f'_-(\theta) = 0$ for all $\theta \notin \mathrm{im}\, f$. Eq. (135) follows from (134) by Leibniz differentiation under the integral sign. $\qquad\square$

A more informal, but instructive, derivation of Eq. (89) considers jump probabilities in small time intervals $[t, t + \Delta t]$. For the derivation, it holds that

$$
\begin{aligned}
&\mathbb{P}[Z(t + \Delta t) \in (-\infty, z]] \\
&= \int_{-\infty}^{f_-(z)} \mathbb{P}[\text{jump in } [t, t + \Delta t] | Z(t) = z'] p(z', t)\,\mathrm{d}z' + o(\Delta t) \\
&\quad + \int_{-\infty}^{z - F(z)\Delta t + o(\Delta t)} \mathbb{P}[\text{no jump in } [t, t + \Delta t] | Z(t) = z'] p(z', t)\,\mathrm{d}z' \\
&= \int_{-\infty}^{f_-(z)} \lambda(z')\Delta t\, p(z', t)\,\mathrm{d}z' + o(\Delta t) + \int_{-\infty}^{z - F(z)\Delta t + o(\Delta t)} (1 - \lambda(z')\Delta t)p(z', t)\,\mathrm{d}z'.
\end{aligned}
$$

Now we take the derivative with respect to z. This yields

$$
\begin{aligned}
&p(z, t + \Delta t) \\
&= f'_-(z)\lambda(f_-(z))p(f_-(z), t)\Delta t \\
&\quad + (1 - \lambda(z)\Delta t)(1 - F'(z)\Delta t)p(z - F(z)\Delta t, t) + o(\Delta t) \\
&= f'_-(z)\lambda(f_-(z))p(f_-(z), t)\Delta t \\
&\quad + (1 - \lambda(z)\Delta t)(1 - F'(z)\Delta t)(p(z, t) - F(z)\Delta t p_z(z, t)) + o(\Delta t) \\
&= f'_-(z)\lambda(f_-(z))p(f_-(z), t)\Delta t + p(z, t) \\
&\quad - \Delta t[\lambda(z)p(z, t) + F'(z)p(z, t) + F(z)p_z(z, t)] + o(\Delta t).
\end{aligned}
$$

Then

$$\lim_{\Delta t \to 0} \frac{p(z, t + \Delta t) - p(z, t)}{\Delta t} = -\partial_z(F(z)p(z, t)) - \lambda(z)p(z, t) + f'_-(z)\lambda(f_-(z))p(f_-(z), t).$$

We now show that the master equation is invariant w.r.t. continuously differentiable monotone transforms. For this purpose let $W_t = h(Z_t)$ for h C^1 and monotone. Let g be the inverse of h. The Ito formula provides the process equation of W

$$dW_t = (h' \circ g)(W_t) \cdot (A \circ g)(W_t)dt + [(h \circ f \circ g)(W_{t-}) - W_{t-}]dY_t$$

and jumps occur with intensity $\lambda_W(w) = \lambda_Z(g(w)) = (\lambda_Z \circ g)(w)$. The process W jumps from w to $\tilde{f} = (h \circ f \circ g)(w)$ and jumps that enter at w, jumped from $\tilde{f}^{-1}(w) = (h \circ f_- \circ g)(w)$. For the first
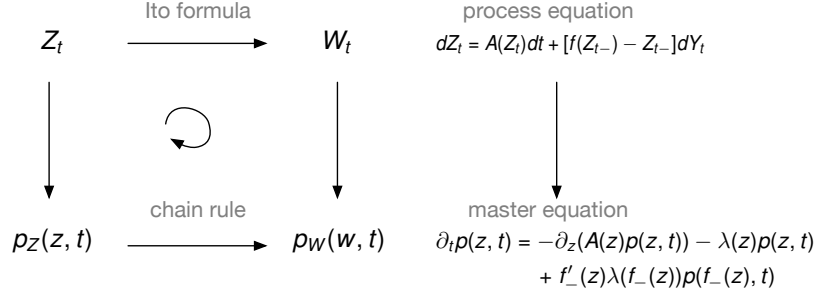
Figure 6: Process equation and probability evolution equation under continuously differentiable monotone transforms $W_t = h(Z_t)$. The translation from process equation to master equation as in corollary 4.22 makes the diagram commute. The transformed process equation is obtained by the Ito formula. The probability densities transform via the chain rule.

term of the process equation note that $h'(g(w))g'(w) = (h \circ g)'(w) = 1$. We omit the t-dependence of p_W in the following derivation of the probability evolution equation for readability.

$$\partial_t p_W(w)$$
$$= - \partial_w \left(\frac{(A \circ g)(w)}{g'(w)} p_W(w) \right) - (\lambda_Z \circ g)(w) p_W(w)$$
$$+ (h \circ f_- \circ g)'(w)(\lambda_Z \circ g \circ h \circ f_- \circ g)(w) p_W((h \circ f_- \circ g)(w))$$
$$= \frac{g''(w)}{(g'(w))^2} F(g(w)) p_W(w) - F'(g(w)) p_W(w) - \frac{F(g(w))}{g'(w)} \partial_w p_W(w) - \lambda_Z(g(w)) p_W(w)$$
$$+ h'(f_-(g(w))) f_-'(g(w)) g'(w) \lambda_Z(f_-(g(w))) p_W((h \circ f_- \circ g)(w))$$

By the transformation rule of densities

$$p_W(w) = g'(w) p_Z(g(w))$$

and

$$\partial_w p_W(w) = g''(w) p_Z(g(w)) + (g'(w))^2 \partial_z p_Z(g(w))$$

we get

$$g'(w) \partial_t p_Z(g(w))$$
$$= g'(w) \left[-F'(g(w)) p_Z(g(w)) - F(g(w)) \partial_z p_Z(g(w)) \right.$$
$$\left. - \lambda_Z(g(w)) p_Z(g(w)) + \lambda_Z(f_-(g(w))) f_-'(g(w)) p_Z(f_-(g(w))) \right]$$
$$= - F'(g(w)) p_W(w) - g'(w) F(g(w)) \frac{\partial_w p_W(w) - g''(w) p_Z(g(w))}{(g'(w))^2}$$
$$- \lambda_Z(g(w)) p_W(w) + \lambda_Z(f_-(g(w))) f_-'(g(w)) p_Z(g((h \circ f_- \circ g)(w)))$$
$$= - F'(g(w)) p_W(w) - \frac{F(g(w)) \partial_w p_W(w)}{g'(w)} + \frac{g''(w) F(g(w))}{(g'(w))^2} [g'(w) p_Z(g(w))]$$
$$- \lambda_Z(g(w)) p_W(w) + \frac{\lambda_Z(f_-(g(w))) f_-'(g(w)) g'(w) p_W(w)}{(g' \circ h)((f_- \circ g)(w))}$$
$$= - F'(g(w)) p_W(w) - \frac{F(g(w)) \partial_w p_W(w)}{(g'(w))} + \frac{g''(w) F(g(w))}{(g'(w))^2} p_W(w)$$
$$- \lambda_Z(g(w)) p_W(w) + \lambda_Z(f_-(g(w))) f_-'(g(w)) g'(w) h'((f_- \circ g)(w)) p_W(w).$$

Hence the diagram in figure 6 commutes. Next, we formulate the proposition 4.20 (ii) for the special case of the BReT-P $(\tau(t), \theta(t))$.

Corollary 4.23. *Let $(\tau(t), \theta(t))_{t \geq 0}$ be a BReT-P on $[0, \infty) \times \mathcal{E}, \mathcal{E} \subseteq \mathbb{R}^n$ with local characteristics (m, g) and let the distribution of $(\tau(t), \theta(t))$ be absolutely continuous with density $(\tau, \theta) \mapsto p(t, \tau, \theta)$. Then it holds for all $\tau > 0$ and θ in the interior of $\mathcal{E}$*

$$\partial_t p(t, \tau, \theta) = -\partial_\tau p(t, \tau, \theta) - m(\tau, \theta) p(t, \tau, \theta). \tag{90}$$

Proof. By definition the BReT-P is a Dirac-PDMP with third local characteristic $f(\tau, \theta) = [0, g(\tau, \theta)]$. Hence $\operatorname{im} f$ is contained in the boundary of $[0, \infty) \times \mathcal{E}$ which implies $R = \emptyset$ and the third term of Eq. (87) vanishes. The first local characteristic $F \equiv [1, \mathbf{0}]$ implies

$$\partial_\tau (F_1(\tau, \theta) \cdot p(t, \tau, \theta)) + \sum_{i=1}^n \partial_{\theta_i} (F_{i+1}(\tau, \theta) p(t, \tau, \theta))$$

$$= \partial_\tau (1 \cdot p(t, \tau, \theta)) + \sum_{i=1}^n \partial_{\theta_i} (0 \cdot p(t, \tau, \theta)) = \partial_\tau p(t, \tau, \theta).$$

$\square$

In section 4.2.4, we provide the master equation for the joint system of subnetwork state and BReT-P. We now aim for the asymptotic distribution of $(\tau(t), \theta(t))_{t \geq 0}$. It is in general not easy to give precise conditions under which an invariant measure for a PDMP exists. We discuss more on this in section 4.2.3 below. For the remainder of this thesis, we assume that a unique invariant measure exists and that $(\tau(t), \theta(t))_{t \geq 0}$ converges in distribution to it. We assume that it is absolutely continuous admitting a density $(\tau, \theta) \mapsto p(\tau, \theta)$. With vanishing left-hand side in Eq. (90), it follows

$$\partial_\tau p(\tau, \theta) = -m(\tau, \theta) p(\tau, \theta). \tag{91}$$

Let ρ solve this equation for $\rho(0, \theta) = 1$. Since Eq. (91) is linear, the density of the invariant measure then satisfies

$$p(\tau, \theta) = p(0, \theta) \rho(\tau, \theta). \tag{92}$$

4.2.2 Boundary condition

The function $\theta \mapsto p_0(\theta) := p(0, \theta)$ is the density of an unnormalized probability measure on $\mathcal{E}$. It has to satisfy a stationarity condition which serves as an integral boundary condition for the family of differential equations (91). We state two versions of the condition. At the end of this subsection, we comment on the interpretation of the following theorem as being the stationarity condition for the embedded Markov chain $(\theta(\sigma_i))_{i \in \mathbb{N}}$, which is known for PDMPs. The novelty is the formulation for Dirac-PDMPs, which to the best of our knowledge is new. The linearity of the equation in the unknown p_0 makes the solution accessible numerically in order to allow for the numerical evaluation of the ACID and the information-theoretic quantities, Eq. (121) and Eq. (123), see the section 5.2 below.

Theorem 4.24. *Let the conditions of corollary 4.23 hold. Then for any $B \in \mathcal{B}(\mathcal{E})$ it holds*

$$\int_B p_0(\theta) \, \mathrm{d}\theta = \int_{g(\tau, \theta) \in B} m(\tau, \theta) \rho(\tau, \theta) p_0(\theta) \, \mathrm{d}\theta \, \mathrm{d}\tau. \tag{93}$$

Proof. Let $t > h > 0$. The probability $\mathbb{P}[\tau(t) \in [0, h), \theta(t) \in B]$ can be written in two ways (up to order $o(h)$), first

$$\int_0^h \int_B p(\tau, \theta, t) \, \mathrm{d}\theta \, \mathrm{d}\tau$$

and second, since we know a jump must have occurred in $(t - h, t]$

$$\int_{t-h}^t \int p(\text{jump at } h' | \theta(h') = \theta', \tau(h') = \tau') p(\tau, \theta, h') \mathbb{1}(g(\tau', \theta') \in B) \, \mathrm{d}\theta' \, \mathrm{d}\tau' \, \mathrm{d}h'$$

$$= \int_{t-h}^t \int m(\tau', \theta') p(\tau', \theta', h') \mathbb{1}(g(\tau', \theta') \in B) \, \mathrm{d}\theta' \, \mathrm{d}\tau' \, \mathrm{d}h'.$$

56

Then dividing by h and letting $h \to 0$ gives the equality

$$\int_B p(0, \theta, t) \, \mathrm{d}\theta = \int_{g(\tau, \theta) \in B} m(\tau, \theta) p(\tau, \theta, t) \, \mathrm{d}\theta \, \mathrm{d}\tau.$$

If we drop the t because of stationarity and use Eq. (92) on the right-hand side, we get the result. $\qquad\square$

The differential version of this for the special case $n = 1$ is the following result.

Theorem 4.25. *For each θ, θ' let $\tau_i(\theta, \theta')$ for $i = 1, \ldots, N(\theta, \theta')$ be an enumeration of the solutions to $g(\tau, \theta') = \theta$, where we allow $N(\theta, \theta') = \infty$. Assume that g is continuously differentiable and $\partial_\tau g(\tau_i(\theta, \theta'), \theta') \neq 0$ for all i. Here, $N(\theta, \theta')$ is the number of such solutions. Then it holds*

$$p_0(\theta) = \int \sum_{i=1}^{N(\theta, \theta')} \frac{m(\tau, \theta') \rho(\tau, \theta')}{|\frac{\partial}{\partial \tau} g(\tau, \theta')|} p_0(\theta') \Bigg|_{\tau = \tau_i(\theta, \theta')} \mathrm{d}\theta', \tag{94}$$

given that the right-hand side is finite.

Proof. Split $\{1, \ldots, N(\theta, \theta')\}$ into $\mathcal{C}_+$ and $\mathcal{C}_-$ depending on whether $\tau \mapsto g(\tau, \theta') - \theta$ has a sign change from $-$ to $+$ or from $+$ to $-$ at $\tau_i(\theta, \theta')$. We choose $B = (-\infty, \theta]$ in Eq. (93) and take the derivative ∂_θ on both sides. Then on the left-hand side we get $p_0(\theta)$. For the right-hand side define $I(\tau, \theta) := m(\tau, \theta) \rho(\tau, \theta) p_0(\theta)$. There exists some choice $\mathcal{C}_+ = \{j_1, \ldots, j_{n_+(\theta, \theta')}\}$ or $\mathcal{C}_+ = \{j_1, \ldots, j_{n_+(\theta, \theta')-1}\}, \infty = j_{n_+(\theta, \theta')}$ and $\mathcal{C}_- = \{i_1, \ldots, i_{n_+(\theta, \theta')}\}$ or $\mathcal{C}_- = \{i_2, \ldots, i_{n_+(\theta, \theta')}\}, i_1 = 0$ for which the right-hand side can be computed as

$$\partial_\theta \int \sum_{k=1}^{n_+(\theta, \theta')} \int_{\tau_{i_k}(\theta, \theta')}^{\tau_{j_k}(\theta, \theta')} I(\tau, \theta') \, \mathrm{d}\tau \, \mathrm{d}\theta'$$

$$= \int \sum_{k=1}^{n_+(\theta, \theta')} -\partial_\theta \tau_{i_k}(\theta, \theta') I(\tau_{i_k}(\theta, \theta'), \theta') + \partial_\theta \tau_{j_k}(\theta, \theta') I(\tau_{j_k}(\theta, \theta'), \theta') \, \mathrm{d}\theta'$$

$$= \int \sum_{i \in \mathcal{C}_+ \cup \mathcal{C}_-} \mathrm{sgn}\{\partial_\tau g(\tau_i(\theta, \theta'), \theta')\} \partial_\theta \tau_i(\theta, \theta') I(\tau_i(\theta, \theta'), \theta') \, \mathrm{d}\theta',$$

where we used, that the ∂_θ-derivatives vanish for the lower and upper limits $\tau_{i_1}(\theta, \theta') = 0$ and $\tau_{j_{n_+(\theta, \theta')}}(\theta, \theta') = \infty$. With

$$1 = \partial_\theta g(\tau_i(\theta, \theta'), \theta')$$
$$= \partial_\tau g(\tau_i(\theta, \theta'), \theta') \cdot \partial_\theta \tau_i(\theta, \theta')$$
$$= |\partial_\tau g(\tau_i(\theta, \theta'), \theta')| \cdot \mathrm{sgn}\{\partial_\tau g(\tau_i(\theta, \theta'), \theta')\} \partial_\theta \tau_i(\theta, \theta')$$

the result follows. $\qquad\square$

Remark 4.26. Note that the integral equation might become singular [122] if there exist $\tau_0, \theta_0, \theta_0'$, such that $g(\tau_0, \theta_0') = \theta_0$, but $\partial_\tau g(\tau_0, \theta_0') = 0$. This may occur both at τ_0 with sign changes of $\tau \mapsto g(\tau_0, \theta_0') - \theta_0$ and without sign changes. In the first case, the equation (94) ought to be interpreted as having an arbitrary value under the integral for this (θ_0, θ_0')-pair. In the second case, the value under the integral in Eq. (94) is defined for every (θ, θ'). However, for θ' close to θ_0' the derivative $\partial_\tau g(\tau_i(\theta_0, \theta'), \theta')$ will approach 0 and hence a singularity will appear nonetheless at (θ_0, θ_0'). The existence of a solution for this singular integral equation needs to be carefully checked.

Remark 4.27. Theorem 4.25 can be generalized to more than one dimension. Let $\gamma_i = (\gamma_{i,1}, \gamma_{i,2})$: $(a_i, b_i) \times \Omega_i \to [0, \infty) \times \Omega, i = 1, \ldots, \eta$ satisfy $g(\gamma_i(t, \theta)) = \theta$ with each γ_i injective and differentiable and $[0, \infty) \times \Omega = \sqcup_{i=1}^\eta \gamma_i((a_i, b_i) \times \Omega_i)$ up to a set of Lebesgue measure zero. Then it holds that

$$p_0(\theta) = \sum_{i: \, \theta \in \Omega_i} \int_{a_i}^{b_i} |\det D\gamma_i(t, \theta)| m(\gamma_i(t, \theta)) \rho(\gamma_i(t, \theta)) p_0(\gamma_{i,2}(t, \theta)) \, \mathrm{d}t.$$

Theorem 4.28. *The normalization of p_0 is related to the mean conditional intensity, i.e.,*

$$\int_{\mathcal{E}} p_0(\theta) \, \mathrm{d}\theta = \mathbb{E}[\hat{\lambda}_\infty]. \tag{95}$$

Proof. Using $\lim_{\tau \to \infty} p(\tau, \theta) = 0$ and Eq. (91), we compute

$$\int p_0(\theta) \, \mathrm{d}\theta = \int p(0, \theta) - \lim_{\tau \to \infty} p(\tau, \theta) \, \mathrm{d}\theta$$
$$= \int \int_0^\infty -\partial_\tau p(\tau, \theta) \, \mathrm{d}\tau \, \mathrm{d}\theta$$
$$= \int \int_0^\infty m(\tau, \theta) p(\tau, \theta) \, \mathrm{d}\tau \, \mathrm{d}\theta$$

and recognize the right-hand side as $\mathbb{E}[m(\tau(\infty), \theta(\infty))] = \mathbb{E}[l(u(\tau(\infty), \theta(\infty)))] = \mathbb{E}[l(\Theta_\infty)] = \mathbb{E}[\hat{\lambda}_\infty]$. $\square$

Remark 4.29. We emphasize that the strong advantage of the BReT-P standard form of the Dirac-PDMP lies in getting rid of the third term in Eq. (87). Figuratively speaking, all probability influx terms align at $\tau = 0$, i.e., the boundary of the domain $[0, \infty) \times \mathcal{E} \ni (\tau, \theta)$. These are processed in the boundary condition (93) instead of the master equation. If the Dirac-PDMP $(\Theta_t)_{t \geq 0}$ in non-standard form is chosen as the state variable, then the stationarity equation (91) grows wider by the third term in Eq. (87), which is needed for any values $\Theta_t = \theta$ in the interior of ϑ that jumps can anticipate, i.e., for $\theta \in R$. Figuratively speaking, these are the probability influx terms. The equation then assumes a difference-differential form, e.g., [60, 7.2.5 (iii)] or the corollary 4.22. The BReT-P circumvents this difference-differential formulation for which solution techniques, like the method of steps, are needed in general. A direct solution technique for a one-dimensional $(\Theta_t)_{t \geq 0}$ can handle the difference-differential formulation. It uses a fixed point method, similar to Eq. (94). We refer the reader to paragraph 5.2.5 below and continue with the BReT-P method here. To summarize, the appealing simplicity of Eq. (91) - being autonomous linear ODEs - comes at the cost of an integral boundary condition ($\tau = 0$) which is more involved.

A further advantage of the BReT-P standard form lies in dimension reduction. Consider again the Dirac-PDMP $\Theta_t = (V_1(t), \ldots, V_{n_0}(t)) \in \vartheta \subseteq \mathbb{R}^{n_0}$ as in remark 4.16. The pair (τ, θ) uniquely informs all state variables $V_{n+1}(t), \ldots, V_{n_0}(t)$ with constant reset values, i.e., via $(\tau, \theta) \mapsto u(\tau, \Sigma(\theta))$. As discussed in remark 4.16, we may consequently dismiss $V_{n+1}(t), \ldots, V_{n_0}(t)$ in the sufficient statistic $\theta(t)$. In case we apply the Markov theory, Eq. (87), to $(\Theta_t)_{t \geq 0}$ instead of $(\tau(t), \theta(t))_{t \geq 0}$ to obtain the analogue of Eq. (91), can we dismiss the state variables with constant reset value as well? Generally not, because the mapping $(V_1(t), \ldots, V_{n_0}(t)) \mapsto (V_1(t), \ldots, V_n(t))$ can be non-injective. In that case there exists no unique mapping $(V_1(t), \ldots, V_n(t)) \mapsto (V_1(t), \ldots, V_{n_0}(t))$. As an example, see fig. 42 in the case study 9.4.2 below with $V_1 = U, V_2 = A$. The trajectory starting at $U(0) = 0.6$ intersects $u = 0.5$ twice.

The embedded Markov chain (EMC) $(\theta(\sigma_i))_{i \in \mathbb{N}}$ of the Dirac-PDMP $(\Theta_t)_{t \geq 0}$ has the transition kernel

$$K(\theta', B) = Q(\theta', B, \infty) = \int_{g(\tau, \theta') \in B} m(\tau, \theta') \rho(\tau, \theta') \, \mathrm{d}\tau. \tag{96}$$

The probability density of the EMC is proportional to p_0 because it equals the probability density of being in a state θ at a jump time. Consequently, Eq. (93) can be interpreted as the stationarity condition for the EMC. When we approached the stationary distribution of the PDMP $(\Theta_t)_{t \geq 0}$ via the BReT-P standard form to derive Eq. (91) and (93), the Eq. (96) shows that we took the well-known strategy via the EMC whose stationary distribution is then lifted to the PDMP via the hazard in Eq. (91). This link was first described in [123]. We refer the reader to the remark 4.33 below. For readers that are more familiar with semi-Markov processes than with PDMPs we provide the semi-Markov perspective on this approach following [124, §2, §3.1, §3.4]. This offers an additional interpretation of the corollary 4.23 and theorems 4.24-4.28 and can thus also leverage improvements on the computational or theoretic side.

Definition 4.30 (Semi-Markov process)**.** A **semi-Markov kernel** is a function $Q\colon \mathcal{E} \times \mathcal{B}(\mathcal{E}) \times [0, \infty) \to \mathbb{R}_{\geq 0}$, such that for each θ, t: $Q(\theta, \cdot, t)$ is a measure with $Q(\theta, \mathcal{E}, t) \leq 1$ and $Q(\theta, \cdot, \infty)$ is a probability measure, for each t, B: $Q(\cdot, B, t)$ is measurable, for each θ, B: $Q(\theta, B, \cdot)$ is a non-decreasing right-continuous function with $Q(\theta, B, 0) = 0$. Let $(X_n, \sigma_n)_{n \in \mathbb{N}_0}$ be a Markov chain on $\mathcal{E} \times [0, \infty)$ with transition kernel $P((x, s), B \times [0, t]) = Q(x, B, t - s)$. Let $(Y_t)_{t \geq 0}$ be a counting process with jump times $(\sigma_n)_{n \in \mathbb{N}_0}$, then $\theta(t) := Z_{Y_t}$ is called a **semi-Markov process**. The process $\tau(t) = t - \sigma_{Y_t}$ is the backward recurrence time and $(\tau(t), \theta(t))_{t \geq 0}$ the associated Markov process. The function $S_x(\tau) = 1 - Q(x, E, \tau)$ is the **survival function** and if it is differentiable $h_x(\tau) = -\frac{\mathrm{d}}{\mathrm{d}\tau} \ln S_x(\tau)$ is the **hazard**. The process $(Z_n)_{n \in \mathbb{N}}$ is the **embedded Markov chain**.

Remark 4.31. Let $H(\theta, [s, t]) := \rho(s, \theta) - \rho(t, \theta)$. For the semi-Markov kernel

$$Q(\theta', B, \tau) = \int_0^\tau \mathbb{1}_B(g(\tau', \theta')) H(\theta', \mathrm{d}\tau') = \int_0^\tau \mathbb{1}_B(g(\tau', \theta')) m(\tau, \theta') \rho(\tau, \theta') \, \mathrm{d}\tau'$$

the BReT-P $(\tau(t), \theta(t))$ can be seen as the associated Markov process of the semi-Markov process $\theta(t)$ with $\tau(t)$ being its backward recurrence time. Conditioned on being in state θ, the function $m(\tau, \theta)$ is the hazard and $\rho(\tau, \theta)$ has an interpretation as the survival function $\mathbb{P}[T(\theta) > \tau]$ of the sojourn time $T(\theta)$. The embedded Markov chain (EMC) $(\theta(\sigma_i))_{i \in \mathbb{N}}$ has the transition kernel

$$K(\theta', B) = Q(\theta', B, \infty) = \int_{g(\tau, \theta') \in B} m(\tau, \theta') \rho(\tau, \theta') \, \mathrm{d}\tau. \tag{97}$$

The probability density of the EMC is proportional to p_0 because it equals the probability density of being in a state θ at a jump time. Consequently, Eq. (93) can be interpreted as the stationarity condition for the EMC. Finally, the relation $p(\tau, \theta) = p_0(\theta) \rho(\tau, \theta)$ reflects the fact that the EMC's stationary distribution and the sojourn time factorize asymptotically [125].

4.2.3 Existence and uniqueness of the stationary distribution

The questions of existence and uniqueness of a stationary distribution for PDMPs are in general not easy to answer [65, p.127]. Research on these questions as well as on stability, i.e., the ergodic property, has seen multiple approaches [126, 127] and is ongoing [128, 129]. By the ergodic property we mean the convergence to the invariant measure independent of the initial condition. However, in the case of the Snyder filter for MMPPs as described in proposition 2.10, the questions can be answered in the affirmative for both existence and uniqueness. We transfer the analogous results [79] on Gaussian observations to the counting process observations giving a step-by-step repetition of the proof.

Theorem 4.32. *Let $(Z_t)_{t \geq 0}$ be a stationary ergodic CTMC on a finite state space $\mathcal{Z}$ with stationary distribution $\mu = (\mu_x)_{z \in \mathcal{Z}}$. The PDMP given by the Snyder filter in proposition 2.10 has a unique invariant distribution and Π_t converges to it in distribution for the initialization $\Pi_0 = \mu$.*

Proof. The existence follows from a compactness argument. As before let $\Delta := \{\pi \in \mathbb{R}_{\geq 0}^{|\mathcal{Z}|}; \sum_{z \in \mathcal{Z}} \pi(z) = 1\}$ denote the set of probability measures on $\mathcal{Z}$. Denote by $\mathcal{M}(\Delta)$ the probability measures on Δ, equipped with the weak topology, i.e., the topology induced by weak convergence of measures. Since Δ is compact, $\mathcal{M}(\Delta)$ is also compact in the weak topology. Denote by $(\Pi_t^\pi)_{t \geq 0}$ the process that evolves according to Eq. (8) with initial condition $\Pi_0^\pi = \pi$ and if the superscript is suppressed we mean $\Pi_0 = \mu$. Define $\nu_n \in \mathcal{M}(\Delta)$ for all $\Gamma \in \mathcal{B}(\Delta)$ by

$$\nu_n(\Gamma) := \frac{1}{n} \int_0^n \mathbb{P}[\Pi_t \in \Gamma] \, \mathrm{d}t.$$

By compactness there exists a subsequence $(\nu_{n_k})_{k \geq 0}$ that converges weakly to $\nu \in \mathcal{M}(\Delta)$. Then for

any $f \colon \Delta \to \mathbb{R}$ continuous

$$
\begin{aligned}
\int_\Delta \mathbb{E}[f(\Pi_t^\pi)] \, \mathrm{d}\nu(\pi) &= \lim_{k\to\infty} \frac{1}{n_k} \int_0^{n_k} \mathbb{E}[f(\Pi_{t+s})] \, \mathrm{d}s \\
&= \lim_{k\to\infty} \frac{1}{n_k} \int_t^{t+n_k} \mathbb{E}[f(\Pi_s)] \, \mathrm{d}s \\
&= \lim_{k\to\infty} \frac{1}{n_k} \int_0^{n_k} \mathbb{E}[f(\Pi_s)] \, \mathrm{d}s \\
&= \int_\Delta f \, \mathrm{d}\nu.
\end{aligned}
$$

This proves that ν is invariant. A measure $\Phi \in \mathcal{M}(\Delta)$ has barycenter μ if $b_\Phi := \int_\Delta \pi \, \mathrm{d}\nu(\pi) = \mu$. For the uniqueness, the idea of the proof is to introduce an order relation on $M_\mu := \{\Phi \in \mathcal{M}(\Delta) : b_\Phi = \mu\}$ and show that the maximal and the minimal invariant measure with respect to this order relation coincide. The maximal and minimal invariant measure are constructed as the limiting distribution of $(\Pi_t^\pi)_{t\geq 0}$ with π distributed according to the maximal and minimal element of M_μ, respectively.

Denote by $\mathcal{C}(\Delta)$ the continuous functions and by $\mathcal{CC}(\Delta)$ the continuous convex functions. By interpreting $\mathcal{M}(\Delta)$ as a subset of the dual space of $\mathcal{C}(\Delta)$, we write for $\Phi \in \mathcal{M}(\Delta), F \in \mathcal{C}(\Delta)$ the scalar product $\langle \Phi, F \rangle$ to mean $\int_\Delta F \, \mathrm{d}\Phi$. Let us thus define the order relation $\leq$ on M_μ via

$$
\Phi \leq \Psi :\Leftrightarrow \forall F \in \mathcal{CC}(\Delta) : \langle \Phi, F \rangle \leq \langle \Psi, F \rangle.
$$

When we refer to maximal and minimal in the following we mean with respect to this order relation. The maximal element in M_μ is $\overline{\Phi} := \sum_{z\in\mathcal{Z}} \mu_z \delta(\pi^z)$, where $\delta(a)$ is the Dirac measure $\delta(a)[B] = \mathbb{1}_B(a)$ and π^z is the vertex of Δ with $\pi^z(z') = \delta_{zz'}$ for all $z' \in \mathcal{Z}$. The minimal element in M_μ is $\underline{\Phi} := \delta(\mu)$. Define the function

$$
h \colon \Delta \times [0,\infty) \times \mathcal{C}(\Delta) \to \mathbb{R}, (\pi, t, F) \mapsto h(\pi, t, F) := \mathbb{E}[F(\Pi_t^\pi)].
$$

Then $h(\cdot, t, F) \in \mathcal{C}(\Delta)$ We now show the following

- For each $F \in \mathcal{CC}(\Delta)$ and $t \geq 0$: $h(\cdot, t, F) \in \mathcal{CC}(\Delta)$, corresponding to lemma 3.2. of [79].

- For each $F \in \mathcal{CC}(\Delta) : t \mapsto \langle \underline{\Phi}, h(\cdot, t, F) \rangle$ is increasing and $t \mapsto \langle \overline{\Phi}, h(\cdot, t, F) \rangle$ is decreasing, corresponding to lemma 3.3. of [79].

The strategy is to condition on appropriate sigma-algebras and use Jensen's inequality. For the first claim, let $\pi_1, \pi_2 \in \Delta, \lambda \in (0,1)$ and $\pi_0 := \lambda\pi_1 + (1-\lambda)\pi_2$. Define $\Phi_\lambda := \lambda\delta(\pi_1) + (1-\lambda)\delta(\pi_2)$. Then for π distributed according to Φ_λ by Jensen's inequality

$$
\begin{aligned}
&h(\lambda\pi_1 + (1-\lambda)\pi_2, t, F) \\
&= h(\pi_0, t, F) = \mathbb{E}[F(\Pi_t^{\pi_0})] \\
&= \mathbb{E}[F(\mathbb{E}[\Pi_t^\pi | \mathcal{F}_t^Y])] \\
&= \mathbb{E}[F(\mathbb{E}[\mathbb{E}[\Pi_t^\pi | \mathcal{F}_t^Y \vee \sigma(\Pi_0^\pi)] | \mathcal{F}_t^Y])] \\
&\leq \mathbb{E}[\mathbb{E}[F(\mathbb{E}[\Pi_t^\pi | \mathcal{F}_t^Y \vee \sigma(\Pi_0^\pi)]) | \mathcal{F}_t^Y]] \\
&= \mathbb{E}[\mathbb{1}(\Pi_0^\pi = \pi_1)F(\Pi_t^{\pi_1})] + \mathbb{E}[\mathbb{1}(\Pi_0^\pi = \pi_2)F(\Pi_t^{\pi_2})] \\
&= \mathbb{P}[\Pi_0^\pi = \pi_1]\mathbb{E}[F(\Pi_t^{\pi_1})] + \mathbb{P}[\Pi_0^\pi = \pi_2]\mathbb{E}[F(\Pi_t^{\pi_2})] \\
&= \lambda h(\pi_1, t, F) + (1-\lambda)h(\pi_2, t, F).
\end{aligned}
$$

For the second claim, let $(Z_t)_{t\geq 0}$ be stationary and consider $\mathcal{F}_s^Z := \sigma(Z_u, 0 \leq u \leq s)$. Then for any $F \in \mathcal{C}(\Delta)$:

$$
\mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_t^Y \vee \mathcal{F}_s^Z])_{z\in\mathcal{Z}})] = \sum_{z'\in\mathcal{Z}} \mu_{z'} \mathbb{E}[F(\Pi_{t-s}^{\pi^{z'}})]. \tag{98}
$$

In particular this holds for $s = 0$. By noting that $\mathcal{F}_t^Y \vee \mathcal{F}_0^Z \subseteq \mathcal{F}_t^Y \vee \mathcal{F}_s^Z$, we obtain with Jensen's inequality for any $F \in \mathcal{CC}(\Delta)$

$$
\begin{aligned}
\langle \overline{\Phi}, h(\cdot, t - s, F) \rangle \\
&= \sum_{z' \in \mathcal{Z}} \mu_{z'} \mathbb{E}[F(\Pi_{t-s}^{\pi^{z'}})] \\
&= \mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_t^Y \vee \mathcal{F}_s^Z])_{z \in \mathcal{Z}})] \\
&= \mathbb{E}[\mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_t^Y \vee \mathcal{F}_s^Z])_{z \in \mathcal{Z}}) | \mathcal{F}_t^Y \vee \mathcal{F}_0^Z]] \\
&\geq \mathbb{E}[F(\mathbb{E}[(\mathbb{P}[Z_t = z | \mathcal{F}_t^Y \vee \mathcal{F}_s^Z])_{z \in \mathcal{Z}} | \mathcal{F}_t^Y \vee \mathcal{F}_0^Z])] \\
&= \mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_t^Y \vee \mathcal{F}_0^Z])_{z \in \mathcal{Z}})] \\
&= \langle \overline{\Phi}, h(\cdot, t, F) \rangle.
\end{aligned}
$$

Note that the larger sigma-algebra corresponded to the shorter time interval. For $\underline{\Phi}$, in contrast, the larger sigma-algebra corresponds to the longer time interval, so we can expect the opposite inequality to hold. For this purpose define $\mathcal{F}_{s,t}^Y := \sigma(Y_u - Y_s, s \leq u \leq t)$. Then $\mathcal{F}_{s,t}^Y \subseteq \mathcal{F}_t^Y$ and for any $F \in \mathcal{CC}(\Delta)$

$$
\begin{aligned}
\langle \underline{\Phi}, h(\cdot, t, F) \rangle &= \mathbb{E}[F(\Pi_t^\mu)] \\
&= \mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_t^Y])_{z \in \mathcal{Z}})] \\
&= \mathbb{E}[\mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_t^Y])_{z \in \mathcal{Z}}) | \mathcal{F}_{s,t}^Y]] \\
&\geq \mathbb{E}[F(\mathbb{E}[(\mathbb{P}[Z_t = z | \mathcal{F}_t^Y])_{z \in \mathcal{Z}} | \mathcal{F}_{s,t}^Y])] \\
&= \mathbb{E}[F((\mathbb{P}[Z_t = z | \mathcal{F}_{s,t}^Y])_{z \in \mathcal{Z}} |)] \\
&= \mathbb{E}[F((\mathbb{P}[Z_{t-s} = z | \mathcal{F}_{t-s}^Y])_{z \in \mathcal{Z}} |)] \\
&= \langle \underline{\Phi}, h(\cdot, t - s, F) \rangle.
\end{aligned}
$$

This proves the second claim. Since $t \mapsto \langle \underline{\Phi}, h(\cdot, t, F) \rangle$ and $t \mapsto \langle \overline{\Phi}, h(\cdot, t, F) \rangle$ are bounded by compactness of Δ and continuity of F, the limits for $t \to \infty$ exist, and induce invariant measures Φ_0 and Φ_1, by

$$
\langle \Phi_0, F \rangle = \lim_{t \to \infty} \langle \underline{\Phi}, h(\cdot, t, F) \rangle
$$

and

$$
\langle \Phi_1, F \rangle = \lim_{t \to \infty} \langle \overline{\Phi}, h(\cdot, t, F) \rangle
$$

for all $F \in \mathcal{CC}(\Delta)$. By Eq. (98), and since $\mathcal{C}(\Delta)$ is the closure of linear combination of elements in $\mathcal{CC}(\Delta)$, this implies the convergence in distribution for $t \to \infty$

$$
\left(\mathbb{P}[Z_t = z | \mathcal{F}_t^Y \vee \mathcal{F}_0^Z] \right)_{z \in \mathcal{Z}} \to \Phi_1
$$

and analogously

$$
\left(\mathbb{P}[Z_t = z | \mathcal{F}_t^Y] \right)_{z \in \mathcal{Z}} \to \Phi_0.
$$

This is exactly the claimed convergence of $(\Pi_t)_{t \geq 0}$ for $\Pi_0 = \mu$, if we prove that Φ_0 is the unique invariant measure. We first prove that Φ_0 is the minimal invariant measure and Φ_1 is the maximal one. Let Φ be an arbitrary invariant measure, then for any $t \geq 0$ because of the convexity of $h(\cdot, t, F)$ and the definition of $\underline{\Phi}$ and $\overline{\Phi}$ as minimal and maximal, respectively

$$
\langle \underline{\Phi}, h(\cdot, t, F) \rangle \leq \langle \Phi, h(\cdot, t, F) \rangle \leq \langle \overline{\Phi}, h(\cdot, t, F) \rangle.
$$

Now, $\langle \Phi, h(\cdot, t, F) \rangle = \langle \Phi, h(\cdot, 0, F) \rangle = \langle \Phi, F \rangle$ constant for all $t \geq 0$ and hence when $t \to \infty$

$$
\langle \Phi_0, F \rangle \leq \langle \Phi, F \rangle \leq \langle \Phi_1, F \rangle.
$$

This proves, that Φ_0 and Φ_1 define the minimal and maximal invariant measure.

In the derivations above we used that the PDMP is time-homogeneous. Since we were only interested in distributions, this allowed - by time shifts of the Markov semigroup - to fix the random

variable at a particular t. Then we conditioned on different sigma-algebra to make use of Jensen's inequality, instead of varying both the random variable and the sigma-algebra with respect to time. This strategy will now again be used to prove that the minimal and maximal invariant measure coincide. To this end, let $(\tilde{Z}_t)_{t \in \mathbb{R}}$ be the stationary CTMC on the full real axis. And let $(\tilde{Y}_t)_{t \in \mathbb{R}}$ be the corresponding integer-valued doubly stochastic Poisson process, i.e., there are sigma-algebras $\mathcal{F}_{s,t}^{\tilde{Z}} := \sigma(\tilde{Z}_u : s \le u \le t)$, $\mathcal{F}_{s,t}^{\tilde{Y}} := \sigma(\tilde{Y}_v - \tilde{Y}_u : s \le u \le v \le t)$ and $\mathcal{F}_{s,t}^{\tilde{Z},\tilde{Y}} := \mathcal{F}_{s,t}^{\tilde{Y}} \vee \mathcal{F}_{-\infty,\infty}^{\tilde{Z}}$, such that

$$\mathbb{E}[\int_s^t C_u \, \mathrm{d}\tilde{Y}_u] = \mathbb{E}[\int_s^t C_u \lambda(\tilde{Z}_u) \, \mathrm{d}u]$$

for all $(C_u)_{s \le u \le t}$ that are $\mathcal{F}_{s,u}^{\tilde{Z},\tilde{Y}}$-predictable. Then

$$(\mathbb{P}[\tilde{Z}_0 = z | \mathcal{F}_{-\infty,0}^{\tilde{Y}} \vee \mathcal{F}_{-\infty,-n}])_{z \in \mathcal{Z}} \tag{99}$$

agrees in distribution with

$$(\mathbb{P}[Z_n = z | \mathcal{F}_n^Y \vee \mathcal{F}_0^Z])_{z \in \mathcal{Z}},$$

hence the limit of expression (99) for $n \to \infty$ is distributed as Φ_1. Analogously,

$$(\mathbb{P}[\tilde{Z}_0 = z | \mathcal{F}_{-n,0}^{\tilde{Y}}])_{z \in \mathcal{Z}} \tag{100}$$

agrees in distribution with

$$(\mathbb{P}[Z_n = z | \mathcal{F}_n^Y])_{z \in \mathcal{Z}},$$

hence the limit of expression (100) for $n \to \infty$ is distributed as Φ_0. By Levy's downwards theorem the expression (99) converges almost surely to

$$(\mathbb{P}[\tilde{Z}_0 = z | \mathcal{F}_{-\infty,0}^{\tilde{Y}} \vee \bigcap_{n \ge 0} \mathcal{F}_{-\infty,-n}])_{z \in \mathcal{Z}},$$

and by Levy's upwards theorem the expression (100) converges almost surely to

$$(\mathbb{P}[\tilde{Z}_0 = z | \mathcal{F}_{-\infty,0}^{\tilde{Y}}])_{z \in \mathcal{Z}}.$$

By the ergodicity of the CTMC $(Z_t)_{t \ge 0}$ the sigma-algebra $\bigcap_{n \ge 0} \mathcal{F}_{-\infty,-n}$ is trivial and consequently, the limits agree and in particular Φ_0 and Φ_1 agree. This concludes the proof. $\square$

The existence and uniqueness of an invariant measure for a PDMP, as well as ergodicity, are often established via properties of the EMC, see the remark 4.31. The existence of an invariant measure is linked to recurrence, or weaker, irreducibility combined with a compactness property, see [130, theorem 4.2]. Since the state space of the EMC is continuous, concepts such as φ-irreducibility, φ-recurrence or Harris recurrence are needed. We refer the interested reader to Asmussen [131, §VI.1, p.168-171 & §VII.3, p.198-206] for a collection of results and the link to regenerative processes and to Meyn and Tweedie [132] for a compendium of results for Markov chains on general state spaces. If existence, uniqueness or ergodicity could be shown for the embedded chain, it remains to transfer the property to the PDMP. Conditions under which properties are transferable have been derived in the equivalence results on ergodicity, recurrence and stability for the PDMP and its embedded chain [127].

Remark 4.33. The link between the set of stationary distributions of the EMC and the stationary distributions of the PDMP was established for the first time in [123]. A simple sufficient condition that guarantees the one-to-one-correspondence between both is the existence of constants $\lambda_1, \lambda_2 > 0$, such that for all $t \ge 0$

$$\lambda_1 \le \hat{\lambda}_t \le \lambda_2. \tag{101}$$

We state and prove in the following that the condition is satisfied by the Snyder filter.

Theorem 4.34. Let $(Z_t)_{t \ge 0}$ be an irreducible CTMC on the finite state space $\mathcal{Z}$. Then the piecewise-deterministic Markov process given by the Snyder filter in proposition 2.10 satisfies the condition (101).

We first introduce the proof idea and the notation required for the proof. Let $\mathcal{Z}_0 := \{z \in \mathcal{Z}; \lambda(z) = 0\}$ the set of zero-states. As above denote the state space of the PDMP by $\Delta = \{\pi \in \mathbb{R}_{\geq 0}^{|\mathcal{Z}|}; \sum_{z \in \mathcal{Z}} \pi(z) = 1\}$, which is compact in the hyperplane $\mathcal{H} := \{\pi \in \mathbb{R}^{|\mathcal{Z}|}; \sum_{z \in \mathcal{Z}} \pi(z) = 1\}$. We construct a subset $\tilde{\Delta} \subseteq \Delta$ which (i) is positively invariant under the flow F of the filtering equation (positively indicates forward in time), namely under

$$F(\pi) := \mathcal{A}\pi - (\operatorname{diag}(\lambda(z)_{z \in \mathcal{Z}}) - l(\pi) \mathbf{I}_{|\mathcal{Z}|})\pi,$$

(ii) satisfies $l(\pi) \geq \lambda_1$ for all $\pi \in \tilde{\Delta}$ for a global $\lambda_1 > 0$ and (iii) contains the initial condition Π_0 as well as all Π_{σ_i} for jump times σ_i. The construction of $\tilde{\Delta}$ is recursive in $k = |S|$ with $S \subseteq \mathcal{Z}_0$. In each recursion step we construct a convex polytope $\Delta^k \subseteq \Delta$ in the ambient space $\mathcal{H}$ that is positively invariant, starting from $\Delta^0 := \Delta$. Intuitively, $\tilde{\Delta}$ is constructed by cutting off the vertices of $\mathcal{Z}_0$ first, the edges of $\mathcal{Z}_0$ next and continuing with the higher dimensional faces until the convex hull of $\mathcal{Z}_0$ is entirely removed, thereby always cutting parallel to the faces.

For the proof we require the following notions from the theory on convex polytopes. A convex polytope is the convex hull of finitely many points or, equivalently, the intersection of closed half spaces. A closed half-space H is given by a normal vector n_H and a vector π_H via $H = \{\pi \in \mathcal{H} : \langle \pi - \pi_H, n_H \rangle \geq 0\}$. Since we work in the ambient space $\mathcal{H}$, the normal vector n_H is contained in the tangent vector space $T_{\mathcal{H}} = \{\pi \in \mathbb{R}^{|\mathcal{Z}|}; \sum_{z \in \mathcal{Z}} \pi(z) = 0\}$ of $\mathcal{H}$ and $\pi_H \in \mathcal{H}$. The normal vector n_H is pointing inward. To each half space H, a supporting hyperplane $h := \{\pi \in \mathcal{H} : \langle \pi - \pi_H, n_H \rangle = 0\}$ is assigned. A convex polytope, given by half spaces $H_1, \cdots, H_n$, is positively invariant under F, if for all boundary points

$$\pi : \pi \in h_i \Rightarrow \langle F(\pi), n_{H_i} \rangle \geq 0. \tag{102}$$

In particular, if the boundary point lies in the intersection $h_i \cap h_j$, it must satisfy the condition for both.

Define the vertices of Δ as π^z with $\pi^z(z') = \delta_{zz'}$ for all $z' \in \mathcal{Z}$ and $\pi^S = 1/|S| \sum_{z \in S} \pi^z$. For each $S \subseteq \mathcal{Z}$ define the corresponding face

$$B_S := \operatorname{conv}\{\pi^z : z \in S\} = \{\pi \in \Delta : \pi(z) = 0 \text{ for all } z \in S^C\}.$$

For simplicity we introduce $B := B_{\mathcal{Z}_0}$. Clearly, $S \subseteq S' \Rightarrow B_S \subseteq B_{S'}$. Denote by $\mathring{B}_S$ the interior of B_S in the subspace topology on B_S, i.e.,

$$\mathring{B}_S = B_S \backslash \bigcup_{S' : S' \subsetneq S} B_{S'} = \{\pi \in \Delta : \pi(z) > 0 \Leftrightarrow z \in S\}.$$

Now, for each $S \subseteq \mathcal{Z}$, we define the vector

$$n_S := \sum_{z \in S^C} \alpha \pi^z - \sum_{z \in S} \beta \pi^z,$$

where $\alpha, \beta > 0$ depend on S and are chosen such that $\sum_{z \in \mathcal{Z}} n_S(z) = 0$ and $\|n_S\| = 1$.

Remark 4.35. The vector n_S is a normal vector for B_S and points inward of Δ. More precisely, the set $h_S := \{\pi \in \mathcal{H} : \langle \pi - \pi^S, n_S \rangle = 0\}$ is a supporting hyperplane of Δ, i.e., $\Delta \subseteq \{\pi \in \mathcal{H} : \langle \pi - \pi^S, n_S \rangle \geq 0\}$. Furthermore, it holds $h_S \cap \Delta = B_S$.

Proof. For all $\pi \in \Delta$

$$\begin{aligned}
\langle \pi - \pi^S, n_S \rangle &= \sum_{z \notin S} \alpha \pi(z) \pi^z(z) - \sum_{z \in S} \beta \pi(z) \pi^z(z) + \sum_{z \in S} \beta \cdot \frac{1}{|S|} \pi^z(z) \\
&= \sum_{z \notin S} \alpha \pi(z) - \sum_{z \in S} \beta \pi(z) + \beta \\
&= \sum_{z \notin S} (\alpha + \beta) \pi(z) \geq 0.
\end{aligned}$$

Then

$$B_S = \{\pi \in \Delta; \forall z \in S^C : \pi(z) = 0\} = \{\pi \in \Delta; \sum_{z \notin S} \pi(z) = 0\} = \{\pi \in \Delta; \langle \pi - \pi^S, n_S \rangle = 0\}.$$

$\square$

Lemma 4.36. *Let $S \subseteq \mathcal{Z}_0$. Under the conditions of the theorem 4.34 for each $\pi \in \mathring{B}_S$ it holds $\langle F(\pi), n_S \rangle > 0$.*

Proof. Let $\pi \in \mathring{B}_S \subseteq B$. Since $\pi(z) = 0$ for all $z \notin \mathcal{Z}_0$, we obtain $l(\pi) = 0$ and $\mathrm{diag}(\lambda(z)_{z \in \mathcal{Z}})\pi = 0$ and hence $F(\pi) = \mathcal{A}\pi$. Then by using Eq. (3)

$$
\begin{aligned}
\langle F(\pi), n_S \rangle =& \alpha \sum_{z \notin S} \langle \mathcal{A}\pi, \pi^z \rangle - \beta \sum_{z \in S} \langle \mathcal{A}\pi, \pi^z \rangle \\
=& \lim_{h \to 0} h^{-1} \alpha \sum_{z' \in S} \mathbb{P}[Z_{t+h} \notin S | Z_t = z']\pi(z') - \lim_{h \to 0} h^{-1}\beta \sum_{z' \in S}(\mathbb{P}[Z_{t+h} \in S | Z_t = z'] - 1)\pi(z') \\
=& \lim_{h \to 0} h^{-1}(\alpha + \beta) \sum_{z' \in S} \mathbb{P}[Z_{t+h} \notin S | Z_t = z']\pi(z').
\end{aligned}
$$

Each summand is non-negative. Since $\mathcal{A}$ is irreducible and $S \neq \mathcal{Z}$, we obtain $\lim_{h \to 0} h^{-1}\mathbb{P}[Z_{t+h} \notin S | Z_t = z'] > 0$ for at least one $z' \in S$. Hence the sum is strictly positive. $\square$

Lemma 4.37. *Let $D \subseteq \Delta$ be a compact set with $D \cap B_S \subseteq \mathring{B}_S$. Then there exists an $\varepsilon_S > 0$ such that for all $\pi \in D$ with $\langle \pi - \pi^S, n_S \rangle = \varepsilon_S$ it holds $\langle F(\pi), n_S \rangle \geq 0$.*

Proof of lemma 4.37. The proof uses the remark 4.35 and lemma 4.36. The property of lemma 4.36 holds for all π in the compact set $D \cap B_S$. Since $\pi \mapsto \langle F(\pi), n_S \rangle$ is continuous, its minimum M_S is assumed on $D \cap B_S$ and is strictly positive. Again by continuity, there is a set $U_S \subseteq D$, which is open in D, contains $D \cap B_S$ and on which $\langle F(\pi), n_S \rangle > M_S/2$. The complement $D \backslash U_S$ is closed and hence compact as a subset of the compact set D. Now we write, by using $B_S = h_S \cap \Delta$ from remark 4.35,

$$
D \cap B_S = \{\pi \in D : \langle \pi - \pi^S, n_S \rangle = 0\}.
$$

By the choice of $U_S \supseteq D \cap B_S$, every $\pi \in D \backslash U_S$ is in the complement of $D \cap B_S$, and thus satisfies $\langle \pi - \pi^S, n_S \rangle > 0$. Hence, the continuous function $\pi \mapsto \langle \pi - \pi^S, n_S \rangle$ assumes its strictly positive minimum m_S on $D \backslash U_S$. For $\varepsilon_S := m_S/2$ it then holds that all $\pi \in D \backslash U_S$ satisfy $\langle \pi - \pi^S, n_S \rangle > \varepsilon_S$, thus

$$
\begin{aligned}
\pi \in D, \langle \pi - \pi^S, n_S \rangle \leq \varepsilon_S &\Rightarrow \pi \in U_S \\
&\Rightarrow \langle F(\pi), n_S \rangle > M_S/2 \geq 0.
\end{aligned}
$$

$\square$

Proof of theorem 4.34. Clearly, $\hat{\lambda}_t \leq \max\{\lambda(z) : z \in \mathcal{Z}\} < \infty$. It remains to prove the lower bound. We claim that for each $k = 0, \ldots, |\mathcal{Z}_0|$ we can define a convex polytope Δ^k that is positively invariant and for which $\Delta^k \cap B_S = \emptyset$ for all $S \subseteq \mathcal{Z}_0$ with $|S| = k$. Start the recursive construction by $\Delta^0 := \Delta$. Clearly, Δ^0 is positively invariant under F. Now assume, that Δ^{k-1} with the above properties has been constructed, then $\Delta^{k-1} \cap B_S \subseteq \mathring{B}_S$ for all $|S| = k$. Define

$$
\Delta^k := \Delta^{k-1} \cap \bigcap_{S \subseteq \mathcal{Z}_0, |S|=k} \{\pi \in \mathcal{H} : \langle \pi - \pi^S, n_S \rangle \geq \varepsilon_S\} \tag{103}
$$

with ε_S as in lemma 4.37 for $D = \Delta^{k-1}$. Then

- Δ^k is a polytope. To see this, let π_H satisfy $\langle \pi_H - \pi^S, n_S \rangle = \varepsilon_S$. Then

$$
\langle \pi - \pi^S, n_S \rangle = \langle \pi - \pi_H + \pi_H - \pi^S, n_S \rangle = \langle \pi - \pi_H, n_S \rangle + \varepsilon_S
$$

  hence $\langle \pi - \pi^S, n_S \rangle \geq \varepsilon_S \Leftrightarrow \langle \pi - \pi_H, n_S \rangle \geq 0$. Consequently, each condition $\langle \pi - \pi^S, n_S \rangle \geq \varepsilon_S$ defines a half space.

- For all $\pi \in B_S$: $\langle \pi - \pi_S, n_S \rangle = 0 < \varepsilon_S$ and hence $\Delta^k \cap B_S = \emptyset$ for all $S \subseteq \mathcal{Z}_0$ with $|S| = k$.

- Δ^k is positively invariant, because in the defining Eq. (103), we may replace Δ^{k-1} by the intersection of halfspaces $\mathcal{H}_1, \ldots, \mathcal{H}_n$. Then each boundary point of Δ^k that lies in a corresponding hyperplane h_i satisfies the condition (102) by the induction assumption. Each boundary point of Δ^k that lies in the hyperplane $\{\pi \in \mathcal{H} : \langle \pi - \pi^S, n_S \rangle = \varepsilon_S\}$ satisfies the condition (102) by lemma 4.37.

Hence, the induction step is completed. For the positively invariant $\tilde{\Delta} := \Delta^{|\mathcal{Z}_0|}$ and $\varepsilon := \varepsilon_{\mathcal{Z}_0}$, we have by construction in Eq. (103) for all $\pi \in \tilde{\Delta}$: $\langle \pi - \pi^{\mathcal{Z}_0}, n_{\mathcal{Z}_0} \rangle \geq \varepsilon$ and hence

$$
\begin{aligned}
l(\pi) = \sum_{z \in \mathcal{Z}} \pi(z)\lambda(z) &= \sum_{z \notin \mathcal{Z}_0} \pi(z)\lambda(z) \\
&\geq \min\{\lambda(z) : z \notin \mathcal{Z}_0\} \sum_{z \notin \mathcal{Z}_0} \pi(z) \\
&\geq \min\{\lambda(z) : z \notin \mathcal{Z}_0\} \cdot \frac{\varepsilon}{\alpha + \beta} =: \lambda_1 > 0.
\end{aligned}
$$

Finally, we have that $\Pi_{\sigma_i} \in \{\pi \in \Delta : \sum_{z \in \mathcal{Z}_0} \pi(z) = 0\}$, which is compact and has empty intersection with B. Hence it has a positive distance from B. And $\Pi_0 \notin B$ because it is the stationary distribution of Z_t which satisfies $\Pi_0(z) > 0$ because $\mathcal{A}$ is irreducible. Then by choosing $\varepsilon_S, S \subseteq \mathcal{Z}_0$ small enough, the uniformly positive distance from B guarantees that Π_0 and $\Pi_{\sigma_i}, i = 1, 2, \ldots$ are in Δ^k for every $k = 1, \ldots, |\mathcal{Z}_0|$, in particular in $\tilde{\Delta}$.

□

It is beyond the scope of this work to investigate the exact conditions on the local characteristics (m, g) of a BReT-P, for which (i) a unique invariant measure exists and (ii) the BReT-P is ergodic, meaning that it converges in distribution to $(\tau_\infty, \theta_\infty)$ independent of the initial distribution. An example of a non-ergodic process $(\tau(t), \theta(t))$ was investigated in [133]. We assume for the remainder of this thesis that (m, g) is such that (i) and (ii) hold. Furthermore, we assume that $m \colon [0, \infty) \times \mathcal{E} \to \mathbb{R}_{>0}$ is continuous. Then $\hat{\lambda}_t = m(\tau_t, \theta_t)$ converges in distribution to the ACID. In case it admits a density, we denote it $m \mapsto p_\lambda(m)$. Furthermore, we assume that the families $\{\phi(\hat{\lambda}_t) : t \geq 0\}$ and $\{\ell(\hat{\lambda}_t, \hat{\mu}_t) : t \geq 0\}$ are uniformly integrable. Then the convergence of the means (not to be confused with L^1-convergence) is guaranteed [134, Appendix, Prop 2.3, p.494] in theorem 5.1. The condition (101) is sufficient for the uniform integrability, since ϕ and ℓ restricted to the domains $[\lambda_1, \lambda_2]$ and $[\lambda_1, \lambda_2] \times [\mu_1, \mu_2]$ are bounded.

Ergodicity allows us to replace the initial distribution by an absolutely continuous one, which makes the evolution equation (90) and the stationarity equation (91) valid. Also, since we are merely interested in the asymptotic properties, we may assume that the Dirac-PDMP or the BReT-P is initialized in the stationary distribution. This assumption implicitly underlies the remark 4.16 and the definition 4.17 of the dimension. This guarantees that it is possible to reduce the dimension of the state variables of joint Markovian progression already before the first jump in remark 4.16. For instance, the Snyder filter, proposition 2.10, is originally initialized in the stationary distribution of the signal process $(Z_t)_{t \geq 0}$, i.e., in a fixed state of the state space Δ. This is because $\mathcal{F}_0^Y = \{\emptyset, \Omega\}$ and by definition $\Pi_0(z) = \mathbb{P}[Z_0 = z | \mathcal{F}_0^Y] = \mathbb{P}[Z_0 = z]$. However, a stationary version can be interpreted as induced by a doubly stochastic Poisson process $(Y_t)_{t \in \mathbb{R}}$ on the whole real axis as in the proof of theorem 4.32. In one example below, section 9.2, we additionally treat the transient behavior, i.e., capturing explicitly the behavior before the occurrence of the first jump.

4.2.4 Joint master equation

Building on the master equation (90) for a one-dimensional counting process, we next provide the corresponding equations for a CRN in a random environment. These are hybrid master equations for a general set of reaction channels and finite dimensional sufficient statistics. As before, integral boundary conditions are needed for a complete description. For the BReT-P with m_j and g_j as in

65

definition 4.11

$$\partial_t p(t, x, \tau, \theta) = -\sum_{j=1}^{M} m_j(x, \tau, \theta) p(t, x, \tau, \theta) - \sum_{j=1}^{M} \partial_\tau p(t, x, \tau, \theta) \tag{104}$$

$$\int_B p(t, x, 0, \theta) \, \mathrm{d}\theta = \sum_{j=1}^{M} \int_{g_j(x-\nu_j, \tau, \theta) \in B} m_j(x - \nu_j, \tau, \theta) p(t, x - \nu_j, \tau, \theta) \, \mathrm{d}\theta \, \mathrm{d}\tau \tag{105}$$

for any $B \in \mathcal{B}(\mathcal{E})$. The characteristic feature of the probability evolution equation is the partial derivative with respect to the sufficient statistic.

For the marginal evolution equation of $p(t, x)$, we depart from the Kolmogorov backward equation, and use the tower property on the filtration $\mathcal{F}_t$ generated by the sufficient statistics

$$\begin{aligned}
\partial_t p(t, x) =\ & \partial_t \mathbb{E}[\mathbb{1}(X_t = x)] \\
=\ & \sum_{j=1}^{M} \mathbb{E}[a_j(Z_t, x - \nu_j) \mathbb{1}(X_t = x - \nu_j)] - \sum_{j=1}^{M} \mathbb{E}[a_j(Z_t, x) \mathbb{1}(X_t = x)] \\
=\ & \sum_{j=1}^{M} \mathbb{E}[\mathbb{E}[a_j(Z_t, x - \nu_j) \mid \mathcal{F}_t] \mathbb{1}(X_t = x - \nu_j)] - \sum_{j=1}^{M} \mathbb{E}[\mathbb{E}[a_j(Z_t, x) \mid \mathcal{F}_t] \mathbb{1}(X_t = x)] \\
=\ & \sum_{j=1}^{M} \mathbb{E}[m_j(x - \nu_j, \tau_t, \theta_t) \mathbb{1}(X_t = x - \nu_j)] - \sum_{j=1}^{M} \mathbb{E}[m_j(x, \tau_t, \theta_t) \mathbb{1}(X_t = x)] \\
=\ & \sum_{j=1}^{M} \mathbb{E}[m_j(x - \nu_j, \tau_t, \theta_t) \mid X_t = x - \nu_j] p(t, x - \nu_j) - \sum_{j=1}^{M} \mathbb{E}[m_j(x, \tau_t, \theta_t) \mid X_t = x] p(t, x).
\end{aligned}$$

Alternatively, marginalizing out sufficient statistics from Eqs. (104)-(105), we obtained

$$\begin{aligned}
\int_0^\infty \int_{\mathcal{E}} \partial_\tau p(t, x, \tau, \theta) \, \mathrm{d}\theta \, \mathrm{d}\tau =\ & 0 - \int_{\mathcal{E}} p(t, x, 0, \theta) \, \mathrm{d}\theta \\
=\ & -\sum_{j=1}^{M} \int_{g_j(x-\nu_j, \tau, \theta) \in \mathcal{E}} m_j(x - \nu_j, \tau, \theta) p(t, x - \nu_j, \tau, \theta) \, \mathrm{d}\theta \, \mathrm{d}\tau \\
=\ & -\sum_{j=1}^{M} \int_0^\infty \int_{\mathcal{E}} m_j(x - \nu_j, \tau, \theta) p(t, x - \nu_j, \tau, \theta) \, \mathrm{d}\theta \, \mathrm{d}\tau \\
=\ & -\sum_{j=1}^{M} \mathbb{E}[m_j(x - \nu_j, \tau_{t-}, \theta_{t-}) \mid X_t = x - \nu_j] p(t, x - \nu_j)
\end{aligned}$$

and hence

$$\partial_t p(t, x) = \sum_{j=1}^{M} \mathbb{E}[m_j(x - \nu_j, \tau_t, \theta_t) \mid X_t = x - \nu_j] p(t, x - \nu_j) - \sum_{j=1}^{M} \mathbb{E}[m_j(x, \tau_t, \theta_t) \mid X_t = x] p(t, x). \tag{106}$$

We recover the familiar structure of probability inflow and outflow terms in the marginal master equation. The derivation illustrated how the inflow term corresponds to terms in the fine-grained joint master equation with sufficient statistics, namely the partial derivative with respect to the sufficient statistic, Eq. (104) in combination with the boundary condition, Eq. (105).

4.3 Moment evolution equation

In the previous section we assumed, there are finitely many sufficient state variables θ of joint Markovian progression that describe the filter distribution or an approximation thereof. In general, however, the filter distribution is a measure-valued Markov process. We describe its evolution equation for moments of test functions. First, introduce the following operator notation for the filter equation.

4.3.1 Operator notation for the measure-valued filter equation

The measure-valued random variable Π_t evolves as

$$d\Pi_t(z) = (\mathcal{A}\Pi_t)(z)dt + \sum_j \frac{(h_j(z) - \langle h_j, \Pi_{t-}\rangle)\Pi_{t-}(z)}{\langle h_j, \Pi_{t-}\rangle}(dY_j(t) - c_j(x)\langle h_j, \Pi_{t-}\rangle dt).$$

Or with the pointwise multiplication operator $(M_f\pi)(z) := f(z)\pi(z)$

$$d\Pi_t = (\mathcal{A}\Pi_t)dt + \sum_j \frac{(M_{h_j} - \langle h_j, \Pi_{t-}\rangle)\Pi_{t-}}{\langle h_j, \Pi_{t-}\rangle}(dY_j(t) - c_j(x)\langle h_j, \Pi_{t-}\rangle dt).$$

Then the joint process (X_t, Π_t) is a piecewise-deterministic Markov process.

4.3.2 Moment equation

Let us now consider test functions in the multiplicative form $\varphi(x, \pi) = c(x) \cdot \langle h, \pi\rangle$, where $\langle h, \cdot\rangle$ is a linear functional. This covers all conditional expectations $\mathbb{E}_\pi[\varphi(Z)]$ via

$$h(\pi) = h_\varphi(\pi) = \int \varphi(z)\pi(z)\, dz.$$

According to the Ito transformation rule for piecewise-deterministic Markov process we obtain

$$d\varphi(X_t, \Pi_t) = \frac{\partial\varphi}{\partial\pi}\mathcal{A}\Pi_{t-}dt + \sum_j \frac{\partial\varphi}{\partial\pi}c_j(X_t)(\langle h_j, \Pi_t\rangle - M_{h_j})\Pi_t dt$$

$$+ \sum_j \left\{ \varphi\left(X_t + \nu_j, \frac{M_{h_j}\Pi_{t-}}{\langle h_j, \Pi_{t-}\rangle}\right) - \varphi(X_{t-}, \Pi_{t-})\right\} dY_j(t)$$

$$= c(X_t)\langle \mathcal{A}^\dagger h, \Pi_t\rangle$$

$$+ c(X_t)\sum_j c_j(X_t)(\langle h, \Pi_t\rangle\langle h_j, \Pi_t\rangle - \langle h, M_{h_j}\Pi_t\rangle)dt$$

$$+ \sum_j \left\{ c(X_t + \nu_j)\frac{\langle h, M_{h_j}\Pi_{t-}\rangle}{\langle h_j, \Pi_{t-}\rangle} - c(X_t)\langle h, \Pi_{t-}\rangle\right\} dY_j(t).$$

For $h = h_\varphi$ this reads

$$d\varphi(X_t, \Pi_t) = c(X_t)\langle \mathcal{A}^\dagger\varphi, \Pi_t\rangle$$

$$+ c(X_t)\sum_j c_j(X_t)(\langle\varphi, \Pi_t\rangle\langle h_j, \Pi_t\rangle - \langle\varphi h_j, \Pi_t\rangle)dt$$

$$+ \sum_j \left\{ c(X_t + \nu_j)\frac{\langle\varphi h_j, \Pi_{t-}\rangle}{\langle h_j, \Pi_{t-}\rangle} - c(X_t)\langle\varphi, \Pi_{t-}\rangle\right\} dY_j(t)$$

$$= c(X_t)\mathbb{E}_{\Pi_t}[\mathcal{A}^\dagger\varphi]$$

$$+ c(X_t)\sum_j c_j(X_t)(\mathbb{E}_{\Pi_t}[\varphi(Z_t)]\mathbb{E}_{\Pi_t}[h_j(Z_t)] - \mathbb{E}_{\Pi_t}[\varphi(Z_t)h_j(Z_t)])dt$$

$$+ \sum_j \left\{ c(X_t + \nu_j)\frac{\mathbb{E}_{\Pi_{t-}}[\varphi(Z_{t-})h_j(Z_{t-})])}{\mathbb{E}_{\Pi_{t-}}[h_j(Z_{t-})]} - c(X_t)\mathbb{E}_{\Pi_{t-}}[\varphi(Z_{t-})]\right\} dY_j(t).$$

The left hand side is $c(X_t)\mathbb{E}_{\Pi_t}[\varphi(Z_t)]$. Upon using the martingale increments $dY_j(t) - c_j(X_t)\langle h_j, \Pi_{t-}\rangle dt$ we arrive at

$$\frac{d}{dt}\mathbb{E}[c(X_t)\langle\varphi, \Pi_t\rangle] = \mathbb{E}[c(X_t)\langle \mathcal{A}^\dagger\varphi, \Pi_t\rangle] + \sum_j \mathbb{E}[\{c(X_t + \nu_j) - c(X_t)\}c_j(X_t)\langle\varphi h_j, \Pi_t\rangle] \qquad (107)$$

The left hand side is simply

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[c(X_t)\langle\varphi,\Pi_t\rangle] = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[c(X_t)\mathbb{E}[\varphi(Z_t)|X_{[0,t]}]] = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[c(X_t)\varphi(Z_t)],$$

whereas the right hand side is composed of the prior dynamics of the environment and the change due to jumps in X

$$\mathbb{E}[\{c(X_t+\nu_j)-c(X_t)\}c_j(X_t)\langle\varphi h_j,\Pi_t\rangle] = \mathbb{E}[\{c(X_t+\nu_j)-c(X_t)\}c_j(X_t)\mathbb{E}[\varphi(Z_t)h_j(Z_t)|X_{[0,t]}]]$$
$$= \mathbb{E}[\{c(X_t+\nu_j)-c(X_t)\}\varphi(Z_t)\lambda_j(X_t,Z_t)].$$

This establishes the equivalence with the Kolmogorov backward equation for the joint system for multiplicative moments $c(X_t)\varphi(Z_t)$. The linking identity is the tower property of conditional expectations. For non-linear functionals of Π_t, we expect differing evolution equations. The conditional variance is such a non-linear functional. In the paragraph 4.5 below we elaborate more on the equivalence of master equations via the tower property.

4.4 Variance of Markov-modulated and self-exciting counting process

In subsection 3.3 we introduced the asymptotic variance slope for Markov-modulated Poisson processes. Here, we cover the analogous results for a large class of self-exciting counting processes, which includes the marginal counting processes obtained by the multinomial filters. Namely, we assume that the (multivariate) counting process $Y \in \mathbb{N}^d$ is driven by the intensity vector $\lambda_t = C\Theta_{t-} + \alpha, C \in \mathbb{R}^{d\times l}$ with the global parameter vector $\Theta_t \in \mathbb{R}^l$ governed by the stochastic evolution equation

$$\mathrm{d}\Theta_t = -A(\Theta_t - \bar{\Theta})\,\mathrm{d}t + V(t)\mathrm{diag}^{-1}(\lambda_t)(\,\mathrm{d}Y_t - \lambda_t\,\mathrm{d}t), \tag{108}$$

where $A \in \mathbb{R}^{l\times l}$ and $V(t) \in \mathbb{R}^{l\times d}$ is a (predictable) process mimicking the conditional covariance. Then we obtain the central result analogous to Eq. (33)

$$\lim_{t\to\infty}\frac{\mathbb{C}\mathrm{ov}[Y_t]}{t} = \mathrm{diag}(\mathbb{E}\lambda_\infty) + C(A^{-1}\mathbb{E}[V(\infty)] + \mathbb{E}[V(\infty)^T]A^{-T} + A^{-1}\mathbb{C}\mathrm{ov}[\Theta_\infty] + \mathbb{C}\mathrm{ov}[\Theta_\infty]A^{-T})C^T. \tag{109}$$

The proof uses martingale techniques. By the Doob-Meyer decomposition for counting processes, decompose $Y_t = Q_t + \int_0^t \lambda_u\,\mathrm{d}u$, where Q_t is a (multivariate) mean-zero martingale. Then we compute

$$\mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_t - \mathbb{E}[Y_t])^T]$$
$$=\mathbb{E}\left[\left(Q_t + C\int_0^t \Theta(u) - \bar{\Theta}\,\mathrm{d}u\right)\left(Q_t^T + \int_0^t \Theta(u)^T - \bar{\Theta}^T\,\mathrm{d}u\,C^T\right)\right]$$
$$=\mathbb{E}\left[\int_0^t \mathbf{I} + CA^{-1}(\mathbf{I} - \exp(-A(t-u)))V(u)\,\mathrm{diag}^{-1}(\lambda_u)\,\mathrm{d}Q_u\times\right.$$
$$\left.\left(\int_0^t \mathbf{I} + CA^{-1}(\mathbf{I} - \exp(-A(t-u)))V(u)\,\mathrm{diag}^{-1}(\lambda_u)\,\mathrm{d}Q_u\right)^T\right]$$
$$=\int_0^t \mathbb{E}\left\{\mathbf{I} + CA^{-1}(\mathbf{I} - \exp(A(u-t)))V(u)\,\mathrm{diag}^{-1}(\lambda_u)\right\}\mathrm{diag}(\lambda_u)\times$$
$$\left\{\mathrm{diag}^{-1}(\lambda_u)V(u)^T(\mathbf{I} - \exp(A^T(u-t)))A^{-T}C^T + \mathbf{I}\right\}\,\mathrm{d}u$$
$$=\int_0^t \mathbb{E}[\mathrm{diag}(\lambda_u)]\,\mathrm{d}u$$
$$+\int_0^t CA^{-1}(\mathbf{I} - \exp(-A(t-u)))\mathbb{E}[V(u)] + \mathbb{E}[V(u)]^T(\mathbf{I} - \exp(-A^T(t-u)))A^{-T}C^T\,\mathrm{d}u$$
$$+CA^{-1}\int_0^t (\mathbf{I} - \exp(-A(t-u)))\mathbb{E}\left[V(u)\,\mathrm{diag}^{-1}(\lambda_u)V(u)^T\right](\mathbf{I} - \exp(-A^T(t-u)))\,\mathrm{d}u\,A^{-T}C^T. \tag{110}$$

68

Furthermore, we apply the Ito formula for counting processes to derive the evolution equation

$$
\begin{aligned}
d\Theta_t \Theta_t^T = & - A(\Theta_t - \bar{\Theta})\Theta_t^T - \Theta_t(\Theta_t - \bar{\Theta})^T A^T + V(t)\operatorname{diag}^{-1}(\lambda_t)V(t)^T \ dt \\
& + \Theta_t \ dQ_t^T \operatorname{diag}^{-1}(\lambda_t)V(t)^T + V(t)^T \operatorname{diag}^{-1}(\lambda_t) \ dQ_t\Theta_t^T \\
& + V(t)\operatorname{diag}^{-1}(\lambda_t)\operatorname{diag}(\ dQ_t)\operatorname{diag}^{-1}(\lambda_t)V(t)^T.
\end{aligned}
$$

Taking expectations on both sides we obtain

$$
\frac{d}{dt}\mathbb{E}\left[\Theta_t\Theta_t^T\right] = -A\mathbb{Cov}[\Theta_t] - \mathbb{Cov}[\Theta_t]A^T + \mathbb{E}\left[V(t)\operatorname{diag}^{-1}(\lambda_t)V(t)^T\right], \tag{111}
$$

which at stationarity equals zero. Then from Eq. (110), with only the terms multiplied by $\mathbf{I}$ surviving and using Eq. (111) equal to zero, we obtain the asymptotic behavior

$$
\begin{aligned}
& \lim_{t\to\infty} t^{-1}\mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_t - \mathbb{E}[Y_t])^T] \\
& = \operatorname{diag}(\mathbb{E}\lambda_\infty) + CA^{-1}\mathbb{E}[V(\infty)] + \mathbb{E}[V(\infty)^T]A^{-T}C^T \\
& \quad + C(A^{-1}\mathbb{Cov}[\Theta(\infty)] + \mathbb{Cov}[\Theta(\infty)]A^{-T})C^T.
\end{aligned}
$$

The variance result applies to the scenario of approximate marginal simulation via approximate filters, see subsection 8.1.5.

Corollary 4.38. *For $d = l = 1$, the asymptotic variance reads*

$$
\lim_{t\to\infty} t^{-1}\mathbb{Var}[Y_t] = \mathbb{E}\lambda_\infty + \frac{2(C\mathbb{E}[V(\infty)] + C^2\mathbb{Var}[\Theta_\infty])}{A}. \tag{112}
$$

and the asymptotic Fano factor, correspondingly,

$$
\lim_{t\to\infty} \mathrm{F}[Y_t] = 1 + \frac{2(C\mathbb{E}[V(\infty)] + C^2\mathbb{Var}[\Theta_\infty])}{\mathbb{E}\lambda_\infty A}. \tag{113}
$$

If $\lambda_t = \Theta_t, C = 1$ in the Snyder filter, then

$$
\lim_{t\to\infty} t^{-1}\mathbb{Var}[Y_t] = \mathbb{E}\Theta_\infty + \frac{2(\mathbb{E}[V(\infty)] + \mathbb{Var}[\Theta_\infty])}{A}. \tag{114}
$$

For the Fano factor, correspondingly,

$$
\lim_{t\to\infty} \mathrm{F}[Y_t] = 1 + \frac{2(\mathbb{E}[V(\infty)] + \mathbb{Var}[\Theta_\infty])}{\mathbb{E}\Theta_\infty A}. \tag{115}
$$

4.5 Unifying approach

We have taken the routes via stochastic conditioning on either the environment or the subnetwork to arrive at generalized master equations. While the evolution equations may differ, the function that solves it is the same. However, the different routes can be taken as starting points for different approximation. As a summary, we present the unified version of the generalized master equations, which highlights the equivalence of approaches. The unifying class of generalized master equations, with the bracket $\langle \cdot \rangle$ denoting the expectation, reads

$$
\begin{aligned}
\partial_t p(t, x) = & \sum_{j=1}^{M}\langle\mathbb{E}[a_j(Z_t, x - \nu_j)\mathbb{1}(X_t = x - \nu_j) \mid \mathcal{F}_t]\rangle - \sum_{j=1}^{M}\langle\mathbb{E}[a_j(Z_t, x)\mathbb{1}(X_t = x) \mid \mathcal{F}_t]\rangle \\
= & \begin{cases} \sum_{j=1}^{M}\langle Z_j(t)\lambda_j(x - \nu_j)p(t, x - \nu_j \mid \mathcal{F}_t)\rangle - \langle Z_j(t)\lambda_j(x)p(t, x \mid \mathcal{F}_t)\rangle, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \mathcal{F}_t \text{ contains } \sigma(Z_t) \\ \\ \sum_{j=1}^{M}\langle\mathbb{E}[Z_j(t) \mid \mathcal{F}_t]\lambda_j(x - \nu_j)\mathbb{1}(X_t = x - \nu_j)\rangle - \langle\mathbb{E}[Z_j(t) \mid \mathcal{F}_t]\lambda_j(x)\mathbb{1}(X_t = x)\rangle, \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad \text{if } \mathcal{F}_t \text{ contains } \sigma(X_t) \end{cases}
\end{aligned}
$$
$$\tag{116}$$

and in both cases we assumed the form $a_j(z, x) = z_j \lambda_j(x)$.

As a characteristic feature each generalized master equation for CRNs in random environment requires the expected value of its propensities at jump times. To evaluate this expected value, the unifying approach suggests to use the tower property of conditional expectations. The choice of the sigma-algebra that we condition on in the intermediate step determines which approach is taken. Section 3.5 corresponded to $\mathcal{F}_t = \sigma(Z_s; s \in [0, t])$ and section 4.2 to $\mathcal{F}_t = \sigma(X_s; s \in [0, t])$. The method of conditional moments [34] is obtained, when choosing the j- and x-dependent $\mathcal{F}_t^{(j,x)} = \sigma(\{X_t = x - \nu_j\})$ in the first sum and $\mathcal{F}_t^{(x)} = \sigma(\{X_t = x\})$ in the second sum. In that case the summands in the last line simplify to $\mathbb{E}[Z_j(t) \mid X_t = x - \nu_j] \lambda_j(x - \nu_j) p(t, x - \nu_j) - \mathbb{E}[Z_j(t) \mid X_t = x] \lambda_j(x) p(t, x)$. Öcal et al. [80] chose $\mathcal{F}_t = \sigma(X_t)$ along with a Markovian projection for the KL divergence as an information theoretic loss.

5   Information-theoretic applications

The results obtained in the previous section contribute to the problem of estimating information measures for CRNs. From the perspective of information transmission in general CRNs, the discrete nature of the reaction events, i.e., Poissonian observations, as opposed to continuous (Gaussian) observations, is of interest. When we model the characteristics of the low copy number regime (as opposed to approximating via a Gaussian diffusion), this discreteness can bring to attention the bottleneck character of sensor molecules. Namely, the signal they sense may be a concentration of a high copy number transmitter, but their synthesis events are restricted to the discrete regime [135, 136]. Furthermore, Poissonian channels are suited for the inclusion of temporal effects because they are capable of modeling time-varying inputs and outputs. However, computing the path MI over Poissonian channels remains difficult. Many works of research resort to basing the MI on single time-point marginals ignoring any encoding in the temporal profile [137, 138, 139]. Other approaches include: (i) Gaussian approximations of the input to make use of analytic results [135], (ii) Monte Carlo estimators [140, 141, 83, 142]. (iii) Often, the intractability decoyed researchers to resort to other channels such as the Gaussian channel [143, 144].

This Poisson channel naturally appears in the setting of CRNs in a random environment as we defined it, because its stochastic rates are the input to the channel and the reaction counters are its output. Namely, the Eq. (56) defines an information-theoretic channel, that transmits information from Z_t and X_t to Y_t. Assuming the condition (C4), the channel is the Poisson channel with input Z_t and output Y_t. Much is already known about the Poisson channel. In particular, its capacity under amplitude constraint $\lambda(Z_t) \in [0, c]$ was determined by Kabanov [61] shortly after its introduction. However, the capacity-achieving input distribution is found to switch infinitely fast between 0 and c [61, 145]. This flaw in terms of physical interpretability was addressed by adding constraints on the bandwidth [146, 147, 82]. While reducing the class of input processes this necessitates to review the task of computing the mutual information (MI) between trajectories.

The definition of the MI depends on the Radon-Nikodym-derivative between probability measures over the joint trajectories of input and output process. More precisely, if $\mu_T^{Z,Y}$ is the probability measure of joint paths $\{(Z_t, Y_t), 0 \le t \le T\}$ as specified by the doubly stochastic Poisson process, whereas μ_T^Z and μ_T^Y are the measures for marginals $\{Z_t, 0 \le t \le T\}$ and $\{Y_t, 0 \le t \le T\}$ with the product measure $\mu_T^Z \times \mu_T^Y$ of joint paths, then the MI and the mutual information rate (MIR) are

$$\mathbb{I}(Z_{[0,T]}, Y_{[0,T]}) := \mathbb{E}\left[\ln \frac{\mathrm{d}\mu_T^{Z,Y}}{\mathrm{d}(\mu_T^Z \times \mu_T^Y)}\right], \quad \bar{\mathbb{I}}(Z, Y) = \lim_{T \to \infty} \frac{1}{T}\mathbb{I}(Z_{[0,T]}, Y_{[0,T]}) \tag{117}$$

The definition, although complicated, is sometimes taken as a departure for computations [90]. Alternatively, at least two strategies of computing or estimating the MI for the Poisson channel are usually considered. The first one represents the MI as the difference between the entropy of the output and the conditional entropy of the output given the input [141, 142]. The second one uses the expression obtained by Liptser [148].

Besides the path MI, we are motivated to look at the relative entropy and the relative entropy rate

(RER) between path measures $\mathbb{P}$ and $\mathbb{Q}$ for the counting process $(Y_t)_{t\geq 0}$

$$D(\mathbb{P}^T\|\mathbb{Q}^T) := \mathbb{E}_\mathbb{P}\left[\ln \frac{\mathrm{d}\mathbb{P}^T}{\mathrm{d}\mathbb{Q}^T}\right], \quad \overline{D}(\mathbb{P}\|\mathbb{Q}) := \lim_{T\to\infty}\frac{1}{T}D(\mathbb{P}^T\|\mathbb{Q}^T). \tag{118}$$

Here, for $T > 0$ denote by $\mathbb{P}^T$ ($\mathbb{Q}^T$) the corresponding path measure of $(Y_t)_{0\leq t\leq T}$, i.e., the restriction of $\mathbb{P}$ ($\mathbb{Q}$) to $\mathcal{F}_T^Y$. When we compare stochastic processes, the relative entropy, also known as Kullback-Leibler (KL) divergence in the literature, serves as an information measure to assess their similarity. As an application in statistics, due to its definition as the expectation of the logarithm of the likelihood ratio, the relative entropy is also closely related to sequential probability ratio tests [149]. For similar applications in optimal control and finance see [59, §VII] for intensity control and [150, 151] for dynamic pricing in finance. Statistical estimators for the RER in case of piecewise constant intensities are given in the form of Monte Carlo estimates in [152], which uses the RER for sensitivity analysis with respect to system parameters.

5.1 Information measures and the conditional intensity

The main purpose of this section is the link between the ACID, definition 4.19, and information-theoretic quantities related to counting processes. Since we opt for asymptotic information-theoretic quantities that complement transient ones and are simpler to compute, we assume stationarity as defined in the this definition. A simulation-free way to compute the ACID is presented in section 5.2.4. We demonstrate that the CI and the ACID appear in expressions for Eq. (117) and (118), two information-theoretic quantities related to counting processes. Both can be conveniently expressed by means of the so-called **natural loss function** [85], defined as $\ell\colon \mathbb{R}_{\geq 0}\times\mathbb{R}_{>0}\to\mathbb{R}_{\geq 0}, \ell(x,\hat{x}) = x\ln(x/\hat{x}) - x + \hat{x}$.

Theorem 5.1. *Let $\phi\colon \mathbb{R}_{\geq 0}\to\mathbb{R}, \phi(z) := z\ln z$ if $z > 0$ and $\phi(0) = 0$.*

1. *Let $(Y_t)_{t\geq 0}$ be a doubly stochastic Poisson process with the $\mathcal{F}_t^{Z,Y}$-intensity $\lambda_t = \lambda(Z_t)$ and $(Z_t)_{t\geq 0}$ its modulating process. Let $(\hat{\lambda}_t)_{t\geq 0}$ be the conditional intensity of $(Y_t)_{t\geq 0}$. Then for the path mutual information it holds*

$$\mathbb{I}(Z_{[0,T]}, Y_{[0,T]}) = \int_0^T \mathbb{E}[\phi(\lambda_t)] - \mathbb{E}[\phi(\hat{\lambda}_t)] \,\mathrm{d}t \tag{119}$$

$$= \int_0^T \mathbb{E}[\ell(\lambda_t, \hat{\lambda}_t)] \,\mathrm{d}t. \tag{120}$$

2. *Suppose, additionally, that $(\lambda_t)_{t\geq 0}$ is a stationary process. Furthermore, suppose $(Y_t)_{t\geq 0}$ is asymptotically stationary, and $\{\phi(\lambda_t) : 0 \leq t < \infty\}$ as well as $\{\phi(\hat{\lambda}_t) : 0 \leq t < \infty\}$ are uniformly integrable, then*

$$\bar{\mathbb{I}}(Z,Y) = \mathbb{E}[\phi(\lambda_\infty)] - \mathbb{E}[\phi(\hat{\lambda}_\infty)], \tag{121}$$

where $\hat{\lambda}_\infty$ is distributed according to the ACID and λ_∞ is distributed according to the stationary distribution of $(\lambda_t)_{t\geq 0}$.

3. *Let $(Y_t)_{t\geq 0}$ be a counting processes on $(\Omega,\mathcal{F})$ with the conditional $\mathbb{P}$-intensity $(\hat{\lambda}_t)_{t\geq 0}$ and the conditional $\mathbb{Q}$-intensity $(\hat{\mu}_t)_{t\geq 0}$, respectively. For $T > 0$ denote by $\mathbb{P}^T$ ($\mathbb{Q}^T$) the corresponding path measure of $(Y_t)_{0\leq t\leq T}$, i.e., the restriction of $\mathbb{P}$ ($\mathbb{Q}$) to $\mathcal{F}_T^Y$. Then the relative entropy between path measures is*

$$D(\mathbb{P}^T\|\mathbb{Q}^T) = \mathbb{E}\left[\int_0^T \ell(\hat{\lambda}_t, \hat{\mu}_t) \,\mathrm{d}t\right], \tag{122}$$

where the expectation is taken over the path measure $\mathbb{P}^T$.

4. *If the joint process $(\hat{\lambda}_t, \hat{\mu}_t)_{t\geq 0}$ converges in distribution on the probability space $(\Omega,\mathcal{F},\mathbb{P})$ to $(\hat{\lambda}_\infty, \hat{\mu}_\infty)$ and $\{\ell(\hat{\lambda}_t, \hat{\mu}_t) : 0 \leq t < \infty\}$ is uniformly integrable, then*

$$\overline{D}(\mathbb{P}\|\mathbb{Q}) := \lim_{T\to\infty}\frac{1}{T}D(\mathbb{P}^T\|\mathbb{Q}^T) = \mathbb{E}_P\left[\ell(\hat{\lambda}_\infty, \hat{\mu}_\infty)\right]. \tag{123}$$

71

Proof.   1. According to Liptser [148, §19.5, Corollary, Eq. (19.140) and after, p.347-348]

$$\mathbb{I}(Z_{[0,T]}, Y_{[0,T]}) = \int_0^T \mathbb{E}[\lambda_t \ln \lambda_t - \hat{\lambda}_t \ln \hat{\lambda}_t] \; \mathrm{d}t$$

$$= \int_0^T \mathbb{E}[\lambda_t \ln \frac{\lambda_t}{\hat{\lambda}_t} + (\lambda_t - \hat{\lambda}_t) \ln \hat{\lambda}_t - \lambda_t + \hat{\lambda}_t] \; \mathrm{d}t$$

$$= \int_0^T \mathbb{E}[\lambda_t \ln \frac{\lambda_t}{\hat{\lambda}_t} + \mathbb{E}[\lambda_t - \hat{\lambda}_t | \mathcal{F}_t^Y] \ln \hat{\lambda}_t - \lambda_t + \hat{\lambda}_t] \; \mathrm{d}t$$

$$= \int_0^T \mathbb{E}[\ell(\lambda_t, \hat{\lambda}_t)] \; \mathrm{d}t.$$

2. Since ϕ is continuous $\phi(\lambda_t)_{t\geq 0}$ and $\phi(\hat{\lambda}_t)_{t\geq 0}$ also converge in distribution. Then by [134, Appendix, Prop 2.3, p.494], the means $\mathbb{E}[\phi(\lambda_t)]$ and $\mathbb{E}[\phi(\hat{\lambda}_t)]$ converge to $\mathbb{E}[\phi(\lambda_\infty)]$ and $\mathbb{E}[\phi(\hat{\lambda}_\infty)]$ and the l'Hospital rule proves the claim.

3. By [59, §VI.6., remark after theorem T12, p.188] the Radon-Nikodym derivative is

$$\frac{\mathrm{d}\mathbb{P}^T}{\mathrm{d}\mathbb{Q}^T} = \prod_{n \geq 1} \frac{\hat{\lambda}_{\sigma_n}}{\hat{\mu}_{\sigma_n}} \mathbb{1}(\sigma_n \leq T) \exp\left(\int_0^T \hat{\mu}_t - \hat{\lambda}_t \; \mathrm{d}t \right).$$

Then for the relative entropy

$$\mathbb{E}_{\mathbb{P}^T}\left[\ln \frac{\mathrm{d}\mathbb{P}^T}{\mathrm{d}\mathbb{Q}^T} \right] = \mathbb{E}_{\mathbb{P}^T}\left[\sum_{n \geq 1} \ln \frac{\hat{\lambda}_{\sigma_n}}{\hat{\mu}_{\sigma_n}} \mathbb{1}(\sigma_n \leq T) + \int_0^T \hat{\mu}_t - \hat{\lambda}_t \; \mathrm{d}t \right]$$

$$= \mathbb{E}_{\mathbb{P}^T}\left[\int_0^T \ln \frac{\hat{\lambda}_t}{\hat{\mu}_t} \; \mathrm{d}Y_t + \int_0^T \hat{\mu}_t - \hat{\lambda}_t \; \mathrm{d}t \right]$$

$$= \mathbb{E}_{\mathbb{P}^T}\left[\int_0^T \ln \frac{\hat{\lambda}_t}{\hat{\mu}_t} \cdot \hat{\lambda}_t + \hat{\mu}_t - \hat{\lambda}_t \; \mathrm{d}t \right]$$

$$= \mathbb{E}_{\mathbb{P}^T}\left[\int_0^T \ell(\hat{\lambda}_t, \hat{\mu}_t) \; \mathrm{d}t \right].$$

4. Since ℓ is continuous $\ell(\hat{\lambda}_t, \hat{\mu}_t)_{t \geq 0}$ also converges in distribution. Then by [134, Appendix, Prop 2.3, p.494], the mean $\mathbb{E}_P\left[\ell(\hat{\lambda}_t, \hat{\mu}_t) \right]$ converges to $\mathbb{E}_P\left[\ell(\hat{\lambda}_\infty, \hat{\mu}_\infty) \right]$ and the l'Hospital rule proves the claim.

$\square$

Remark 5.2. In Eq. (123) the projection $\hat{\lambda}_\infty$ on the first component of the joint limit $(\hat{\lambda}_\infty, \hat{\mu}_\infty)$ is $\mathbb{P}$-distributed according to the ACID of the conditional $\mathbb{P}$-intensity $\hat{\lambda}_t$. However, due to the asymmetry of the relative entropy the projection $\hat{\mu}_\infty$ on the second component is neither necessarily $\mathbb{Q}$-distributed nor $\mathbb{P}$-distributed according to the ACID of the conditional $\mathbb{Q}$-intensity $\hat{\mu}_t$.

5.2   Simulation-free computation

The difficulty in computing the MI via Eq. (119) clearly lies in the computation of $\mathbb{E}[\phi(\hat{\lambda}_t)]$. Different approaches address this. (i) In case of a stationary input $(\lambda_t)_{t\geq 0}$, Eq. (121) can simplify the computation. The processes $(\lambda_t)_{t\geq 0}$ and $(\hat{\lambda}_t)_{t\geq 0}$ can be confined to the moments $\mathbb{E}[\phi(\cdot)]$ of their limiting distributions. (ii) Properties of the function ϕ can be used. For instance ϕ is Lipschitz continuous on bounded intervals, which was used for a reduction to second order moments of $\hat{\lambda}_\infty$ for bounded λ [61]. The non-negative third derivative of ϕ was used in a similar way [82]. (iii) Kabanov and Davis used martingale theory with its rich second order analysis tools. (iv) The conditioning on a coarser

sigma-field [147] or the use of a suboptimal estimator [82] provided upper bounds. (v) Alternatively, one needs to have the knack for choosing tractable input process classes. For example, piecewise-constant input trajectories, whose amplitudes follow a Markov chain were considered [153]. (vi) Being an expectation the MI is often estimated via Monte Carlo sampling [140, 83].

Using our results from sections 4.2.1 and 4.2.2, we combined the approaches (i) and (v). In pursuit of (v), we restrict ourselves to stationary continuous-time Markov chains $(Z_t)_{t \geq 0}$ with a low number of input states. For this case, we present a simulation-free method for the computation of the asymptotic information-theoretic quantities, the MIR, Eq. (121), and the RER, Eq. (123).

This technique is applied in the case studies 9.4.1, 9.4.3 and 8.2.1. In section 9.2 below, we portray an example that permits an analytic solution for the mutual information, see theorems 9.2 and 9.3 below. Generally, it not possible to compute information-theoretic quantities for the Poisson channel in analytically closed form. We propose a simulation-free estimation method using the above filtering and Markov theory to avoid estimation from Monte Carlo simulations. This section provides an implementation to obtain the ACID, definition 4.19, and the asymptotic information-theoretic quantities, Eq. (121) and Eq. (123) for the counting processes specified in definition 4.7 and 4.9.

5.2.1 Discretization

We assume that Y_t is a BReT-P counting process as in definition 4.14 and we discretize the integral boundary condition (93). To this end, we assume that $p_0(\theta)$ is supported on $\Omega \subseteq \mathbb{R}^n$. We choose a partition $(\Omega_i)_{i=1,\dots,N}$ with equal volume $\text{vol}(\Omega_i) = \text{vol}(\Omega_j) =: \nu$. For $n = 1$, this is an equidistant partition. We discretize $p_0(\theta)$ as

$$p_0(\theta) = \sum_{i=1}^{N} a_i \mathbb{1}_{\Omega_i}(\theta)$$

with unknowns a_i and discretize $m(\tau, \theta)$, respectively $\rho(\tau, \theta)$, as

$$m(\tau, \theta) = \sum_{i=1}^{N} m(\tau, \theta_i) \mathbb{1}_{\Omega_i}(\theta), \quad \rho(\tau, \theta) = \sum_{i=1}^{N} \rho(\tau, \theta_i) \mathbb{1}_{\Omega_i}(\theta) \tag{124}$$

for a choice of representatives $\theta_i \in \Omega_i$, e.g. the center of Ω_i. Define the border crossing time points $\tau_0(\theta') := 0$ and recursively $\tau_k(\theta') := \min\{\tau > \tau_{k-1} : g(\tau, \theta') \in \partial\Omega_i \text{ for some } i\}$. Define $I(k, \theta') := i$ if $g(\tau, \theta') \in \Omega_i$ for $\tau_{k-1}(\theta') < \tau < \tau_k(\theta')$. Then for $B = \Omega_i$ the equation (93) reads

$$a_i = \sum_{j=1}^{N} \left(\sum_{k:I(k,\theta'_j)=i} \int_{\tau_{k-1}(\theta'_j)}^{\tau_k(\theta'_j)} m(\tau, \theta'_j) \rho(\tau, \theta'_j) \, \mathrm{d}\tau \right) a_j$$

$$= \sum_{j=1}^{N} \left(\sum_{k:I(k,\theta'_j)=i} \rho(\tau_{k-1}(\theta'_j), \theta'_j) - \rho(\tau_k(\theta'_j), \theta'_j) \right) a_j. \tag{125}$$

Defining the bracket term as $A_{i,j}$, the equation can be written $\vec{a} = A\vec{a}$ in matrix form. Observe that by a telescope sum argument and $\rho(0, \theta'_j) = 1$ the matrix A is left stochastic. If A is quasi-positive, the fixed point equation has a unique non-zero solution by the Perron-Frobenius theorem. Then $\vec{a}$ can be approximated by taking a column of a large enough power A^{2^L} and multiplying by the product of weight ν and normalization constant $\mathbb{E}[\hat{\lambda}_\infty]$.

5.2.2 The mutual information rate

We assume that $(Y_t)_{t \geq 0}$ is an asymptotically stationary MMPP as in definition 2.8 and 4.19 with the local characteristics $(\bar{F}, l, f)$ of a corresponding Dirac-PDMP and (m, g) the local characteristics of the corresponding BReT-P as in Eq. (84) - (85). Let $\lambda_t \to \lambda_\infty$ in distribution. The MIR is

$$\bar{\mathbb{I}}(Z, Y) = \mathbb{E}[\phi(\lambda_\infty)] - \int \int_0^\infty \phi(m(\tau, \theta)) p(\tau, \theta) \, \mathrm{d}\tau \, \mathrm{d}\theta. \tag{126}$$

The first term is a finite sum. The outer integral can be approximated by

$$\sum_{i=1}^{N} p_0(\theta_i)\nu \int_0^{\infty} \phi(m(\tau,\theta_i))\rho(\tau,\theta_i)\, \mathrm{d}\tau, \tag{127}$$

where $p_0(\theta_i)$ is found as in section 5.2.1. Define the partial integral $J(t,\theta')$ by

$$J(t,\theta') := \int_0^t \phi(m(\tau,\theta'))\rho(\tau,\theta')\, \mathrm{d}\tau.$$

Then for each i, the integral in Eq. (127) is efficiently solved by the joint ODE system, the dot denoting the τ-derivative,

$$\dot{\rho}(\tau,\theta_i) = -l(\Theta_\tau)\rho(\tau,\theta_i) \tag{128}$$

$$\dot{\Theta}_\tau = F(\Theta_\tau) \tag{129}$$

$$\dot{J}(\tau,\theta_i) = \phi(l(\Theta_\tau))\rho(\tau,\theta_i). \tag{130}$$

The initial conditions are $\rho(0,\theta_i) = 1$, $\Theta_0 = (\theta_i, v_{n+1}^0, \dots, v_{n_0}^0)$ and $J(0,\theta_i) = 0$.

Remark 5.3 (J converges to the MIR exponentially fast). If $m(\tau,\theta) \geq m_0 > 0$ uniformly over τ,θ, then $\rho(\tau,\theta) \leq e^{-m_0\tau}$. Let furthermore $m(\tau,\theta) \leq R(\theta)$. Then

$$|J(\infty,\theta') - J(t,\theta')| \leq \int_t^{\infty} |\phi(m(\tau,\theta'))|e^{-m_0\tau}\, \mathrm{d}\tau = \frac{\max\{\phi(R(\theta')), \frac{1}{e}\}}{m_0}e^{-m_0 t}.$$

In order to obtain exponential convergence of $\int p_0(\theta')J(t,\theta')\, \mathrm{d}\theta'$ to the integral in Eq. (126) for $t \to \infty$, it needs to holds that

$$\int p_0(\theta') \max\{\phi(R(\theta')), \frac{1}{e}\}\, \mathrm{d}\theta' < \infty.$$

Sufficient conditions are, for instance, if R is bounded or if $\theta \mapsto R(\theta)$ is linear in θ and the second moment of p_0 exists.

As a summary, a numerical approximation of the MIR can be obtained by solving for $J(\infty,\theta_i)$ via Eq. (128) - (130) and $p_0(\theta_i)$ via Eq. (125). The weights in Eq. (125) can be determined on the fly while solving Eq. (128) - (130).

5.2.3 The relative entropy rate

Let $(Y_t)_{t\geq 0}$ be an (F,l,f)-counting process with the corresponding BReT-P counting process characteristics (m,g), see Eq. (84) and (85). Let $(\tilde{F},\tilde{l},\tilde{f})$ and $(\tilde{m},\tilde{g})$ be the local characteristics of another counting process. Trajectory-wise for trajectories of $(Y_t)_{t\geq 0}$ the CI can be evaluated via Eq. (83) with $\hat{\lambda}_t = l(\Theta_{t-})$ and

$$\mathrm{d}\tilde{\Theta}_t = \tilde{F}(\tilde{\Theta}_t)\, \mathrm{d}t + (\tilde{f}(\tilde{\Theta}_{t-}) - \tilde{\Theta}_{t-})\, \mathrm{d}Y_t,$$

with $\hat{\mu}_t = \tilde{l}(\tilde{\Theta}_{t-})$. For the computation of the RER the measure of the (F,l,f)-counting process $(Y_t)_{t\geq 0}$ is used, i.e., for both trajectory-wise evaluations $(Y_t)_{t\geq 0}$ is a counting process with the CI $(\hat{\lambda}_t)_{t\geq 0}$. Then define the concatenated Dirac-PDMP $(\Theta_t, \tilde{\Theta}_t)_{t\geq 0}$ on the state space $\vartheta \times \tilde{\vartheta}$. For this purpose let $\Gamma : \vartheta \times \tilde{\vartheta} \to \vartheta$ and $\tilde{\Gamma} : \vartheta \times \tilde{\vartheta} \to \tilde{\vartheta}$ be the projections. Then the local characteristics of the concatenated process are $(F \circ \Gamma, \tilde{F} \circ \tilde{\Gamma}), l \circ \Gamma, (f \circ \Gamma, \tilde{f} \circ \tilde{\Gamma})$. Note the asymmetry in the second local characteristic due to the asymmetry of the relative entropy. The corresponding BReT-P characteristics for $(\tau(t), \theta(t), \tilde{\theta}(t))$ on $[0,\infty) \times \mathcal{E} \times \tilde{\mathcal{E}}$ and projections $\gamma : [0,\infty) \times \mathcal{E} \times \tilde{\mathcal{E}} \to [0,\infty) \times \mathcal{E}, \tilde{\gamma} : [0,\infty) \times \mathcal{E} \times \tilde{\mathcal{E}} \to [0,\infty) \times \tilde{\mathcal{E}}$ are $m \circ \gamma, (g \circ \gamma, \tilde{g} \circ \tilde{\gamma})$. If the stationary distribution $p(\tau,\theta,\tilde{\theta})$ is found, then the RER can be deduced by

$$\overline{D}(P\|\mathbb{Q}) = \int_{\tilde{\mathcal{E}}} \int_{\mathcal{E}} \int_0^{\infty} \ell(m(\tau,\theta), \tilde{m}(\tau,\tilde{\theta}))p(\tau,\theta,\tilde{\theta})\, \mathrm{d}\tau\, \mathrm{d}\theta\, \mathrm{d}\tilde{\theta}.$$

The outer double integral can be approximated by

$$\sum_{i=1}^{N} p_0(\theta_i, \tilde{\theta}_i)\nu \int_0^{\infty} \ell(m(\tau, \theta_i), \tilde{m}(\tau, \tilde{\theta}_i))\rho(\tau, \theta_i, \tilde{\theta}_i) \, \mathrm{d}\tau, \tag{131}$$

where $p_0(\theta_i, \tilde{\theta}_i)$ is found as in section 5.2.1. Define the partial integral $L(t, \theta', \tilde{\theta}')$ by

$$L(t, \theta', \tilde{\theta}') := \int_0^t \ell(m(\tau, \theta'), \tilde{m}(\tau, \tilde{\theta}'))\rho(\tau, \theta', \tilde{\theta}') \, \mathrm{d}\tau.$$

Then for each i, the integral in Eq. (131) is efficiently solved by the joint ODE system, the dot denoting the τ-derivative,

$$\dot{\rho}(\tau, \theta_i, \tilde{\theta}_i) = -l(\Theta_\tau)\rho(\tau, \theta_i, \tilde{\theta}_i)$$
$$\dot{\Theta}_\tau = F(\Theta_\tau)$$
$$\dot{\tilde{\Theta}}_\tau = \tilde{F}(\tilde{\Theta}_\tau)$$
$$\dot{L}(\tau, \theta_i, \tilde{\theta}_i) = \ell(l(\Theta_\tau), \tilde{l}(\tilde{\Theta}_\tau))\rho(\tau, \theta_i, \tilde{\theta}_i).$$

The initial conditions are $\rho(0, \theta_i, \tilde{\theta}_i) = 1$, $\Theta_0 = (\theta_i, v_{n+1}^0, \dots, v_{n_0}^0)$, $\tilde{\Theta}_0 = (\tilde{\theta}_i, \tilde{v}_{\tilde{n}+1}^0, \dots, \tilde{v}_{\tilde{n}_0}^0)$ and $L(0, \theta_i, \tilde{\theta}_i) = 0$.

5.2.4 Numerical approximation of the ACID

The ACID is obtained as the distribution p_λ of $m(\tau, \theta)$, where (τ, θ) is distributed according to the density $p(\tau, \theta)$, see Eq. (91) and (92). We discretize $(0, \infty)$ with mesh size Δm and compute the weights

$$p_i := \mathbb{P}[m_{i-1} \leq m(\tau, \theta) \leq m_i]$$
$$= \int \int_0^{\infty} \mathbb{1}_{(m_{i-1}, m_i]}(m(\tau, \theta))\rho(\tau, \theta)p_0(\theta) \, \mathrm{d}\tau \, \mathrm{d}\theta.$$

If we define $\tau_k^{(m)}(\theta')$ similarly as $\tau_k(\theta')$, namely by $\tau_0^{(m)}(\theta') = 0$ and recursively $\tau_k^{(m)}(\theta') := \min\{\tau > \tau_{k-1}^{(m)}(\theta') : m(\tau, \theta') = m_i \text{ for some } i\}$, then

$$p_i \approx \nu \cdot \sum_{j=1}^{N} p_0(\theta_j') \sum_{k: I'(k, \theta_j')=i} \int_{\tau_{k-1}^{(m)}(\theta_j')}^{\tau_k^{(m)}(\theta_j')} \rho(\tau, \theta_j') \, \mathrm{d}\tau, \tag{132}$$

where $I'(k, \theta') := i$ if $g(\tau, \theta') \in \Omega_i$ for $\tau_{k-1}(\theta') < \tau < \tau_k(\theta')$. We further have

$$\Delta m \cdot p_\lambda(m) \approx \sum_{i=1}^{\infty} p_i \mathbb{1}_{(m_{i-1}, m_i]}(m).$$

The ACID can be used for an empirical assessment of the similarity between counting processes, e.g., by visualizing the histogram or by computing similarity measures between densities, such as the Wasserstein metric.

5.2.5 Direct method

In special cases, the fixed point method 5.2.1 can be applied directly to the ACID $p_\lambda(m)$ without the need of the BReT-P, i.e., we derived a linear fixed point equation

$$p_\lambda(m) = \int K(m, m')p_\lambda(m') \, \mathrm{d}m' \tag{133}$$

for the stationary solution of equation (89) in corollary 4.22.

Theorem 5.4. *Assume the conditions of corollary 4.22. Suppose that the function $G\colon \vartheta \to \mathbb{R}$ evolves as*

$$G'(\theta) = -\frac{l(\theta)}{A(\theta)}G(\theta)$$

on ϑ with $G(c) = 1$ for some $c \in \vartheta$. If there is $p\colon \vartheta \to \mathbb{R}_{\geq 0}$ that fulfills

$$p(\theta) = -\int_{f_-(\theta)}^{b} \frac{l(z')G(\theta)}{A(\theta)G(f(z'))}p(z') \; \mathrm{d}z', \tag{134}$$

for all $\theta \in \vartheta$, then the stationarity condition is satisfied for all $\theta \in \vartheta$:

$$0 = -\partial_\theta(A(\theta)p(\theta)) - l(\theta)p(\theta) + f'_-(\theta)l(f_-(\theta))p(f_-(\theta)). \tag{135}$$

Proof. Eq. (135) follows from (134) by Leibniz differentiation under the integral sign. $\square$

The direct method applies for the case studies 9.4.1 and 8.2 below.

6 Examples of environments

Linear environments offer convenient choices as discussed in the introduction, section 2.1.3. Here we introduce common examples, ranging from discrete state spaces with few states, and countably many states to the positive reals. Along with their properties known from the literature, we also present novel results that our case studies refer to. For the examples 6.1 and 6.2, we use that CRNs with only zeroth- and first-order reactions are linear.

6.1 Conversion process

The conversion process describes metastable switching of a fixed total number N of molecules, that can be in k different states, structures or conformations. This is an ubiquitous principle in cell biology. Promoter states determine transcription efficiency, mRNA structures allocate different metastable states to facilitate or hamper translation, enzymes change conformation upon activation to catalyze reactions. We illustrate the model for promoter switching, but the promoter is exchangeable with other macromolecules. Suppose there are N transcription sites with the same promoter architecture. At each site, the promoter can be in one of k states. The switching between the states occurs according to a Markov jump process with transition rates $c_{i,j}$ for the i-th to the j-th state. When being in the i-th state, the transcription rate at which mRNA are produced is r_i. We assume the transcription sites have the same stochastic properties, i.e., the same $c_{i,j}$ and r_i and are stochastically independent. More formally, we define the conversion model $\mathcal{M}_S$ to consist of the k different promoters and their $k(k-1)$ inter-conversion reactions

$$R_{j,i}, R_{i,j}\colon \quad \mathrm{Z}_j \xrightleftharpoons[c_{i,j}]{c_{j,i}} \mathrm{Z}_i, \tag{136}$$

where $j, i \in \{1, \dots, k\}$ and $j \neq i$, with the coupling matrix

$$\begin{pmatrix} 0 & c_{1,2} & \cdots & c_{1,k} \\ c_{2,1} & 0 & \cdots & c_{2,k} \\ \vdots & & & \\ c_{k,1} & c_{k,2} & \cdots & 0 \end{pmatrix}.$$

The conversion process is $Z(t) = [Z_1(t), \dots, Z_k(t)]$ where $Z_i(t)$ counts the number of transcription sites that are in state i at time t. the conversion process is a closed system with only monomolecular (conversion) reactions, in particular it is linear and hence (C2) is satisfied. Its stationary distribution

is the multinomial distribution with parameters $p_1, \ldots, p_k$ and N. The system parameters are derived to be

$$A_{ij} = -c_{j,i}, \ c_{i,i} = \sum_{\substack{j=1 \\ j \neq i}}^{k} c_{i,j}, \tag{137}$$

$$p = [p_1, \ldots, p_k]^T, \ \text{s.t.} \ Ap = 0, \sum_{i=1}^{k} p_i = 1 \tag{138}$$

$$\Sigma_{ij} = N(\delta_{ij} p_i - p_i p_j). \tag{139}$$

The multinomial stationary distribution implied the form of the covariance matrix, Eq. (139). According to Jahnke and Huisinga [154] p is uniquely determined by the condition in Eq. (138) if A is irreducible. We assume (C1) to hold.

In combination with state-dependent transcription rates $[r_1, \ldots, r_k]$, such that the transcription reaction is zero-order modulated as

$$R: \quad \emptyset \xrightarrow{\ \sum_{i=1}^{k} r_i \ Z_i\ } \text{Prot}, \tag{140}$$

the conditions (C3) and (C4) are satisfied with

$$C = r = [r_1, \ldots, r_k] \, , \mu = N \langle r, p \rangle. \tag{141}$$

The special case $N = 1, k = 2$ is the random telegraph process. For convenience, we call Z_1 the On state and Z_2 the Off state and abbreviate $c_1 := c_{2,1}, c_2 := c_{1,2}$. Often in its standard form it satisfies $r_1 = 1, r_2 = 0$ but in general we redefine $\lambda_1 := r_1, \lambda_0 := r_2$ with $\Delta\lambda := \lambda_1 - \lambda_0$.

6.1.1 The cumulants of the random telegraph process

We derived the cumulants of the random telegraph process $Z(t)$ based on the higher-order autocovariance functions [155]. The $(n+1)$-th cumulant term for $t_1 < \cdots < t_n < t$, defined in section 3.5, reads

$$\mathcal{C}(t_1, \ldots, t_n, t) = \frac{(c_2 - c_1)^{n-1}}{(c_1 + c_2)^{n-1}} \gamma(0, t - t_1) = \frac{(c_2 - c_1)^{n-1}}{(c_1 + c_2)^{n-1}} \cdot \frac{c_1 c_2}{(c_1 + c_2)^2} e^{-(c_1 + c_2)(t - t_1)}, \tag{142}$$

where $\gamma(t_1, \ldots, t_n) = \mathbb{E}[(Z(t_1) - \mathbb{E}Z(t_1)) \ldots (Z(t_n) - \mathbb{E}Z(t_n))]$ is the higher-order autocovariance function. Our derivation of Eq. (142) departs from the following recursive definition of the cumulant terms

$$\mathcal{C}(t_1, \ldots, t_n, t) = \gamma(0, \Delta t_1, \ldots, \Delta t_n) - \sum_{i=2}^{n-1} \gamma(0, \Delta t_1, \ldots, \Delta t_{i-1}) \mathcal{C}(t_{i+1}, \ldots, t_n, t).$$

We proceed by induction on n. Clearly, for $n = 1$ and $n = 2$, the identity holds, because the sum is empty. Assuming that the conjecture (142) holds up to and including $n - 1$. Then we compute with

$\sigma^2 = \frac{c_1 c_2}{(c_1+c_2)^2}$, $\theta = \frac{c_2-c_1}{c_1+c_2}$, $\varphi = c_1 + c_2$ and $f_j = e^{\varphi \Delta t_j}$ as in [155] for $n = 2k - 1$

$$\mathcal{C}(t_1,\ldots,t_n,t) = \sigma^2 e^{-\varphi(t-t_1)}\theta^{n-1} + e^{-\varphi(t-t_1)}\sum_{r=1}^{k-1}\theta^{2k-2-2r}(\sigma^2)^{r+1}A^{(r)}(2k-2)$$

$$-\sum_{j=1}^{k-1}\gamma(0,\Delta t_1,\ldots,\Delta t_{2j-1})\sigma^2 e^{-\varphi(t-t_{2j+1})}\theta^{n-(2j+1)}$$

$$-\sum_{j=1}^{k-2}\gamma(0,\Delta t_1,\ldots,\Delta t_{2j})\sigma^2 e^{-\varphi(t-t_{2j+2})}\theta^{n-(2j+2)}$$

$$=\sigma^2 e^{-\varphi(t-t_1)}\theta^{n-1} + e^{-\varphi(t-t_1)}\sum_{r=1}^{k-1}\theta^{n-1-2r}(\sigma^2)^{r+1}A^{(r)}(2k-2)$$

$$-\sum_{j=1}^{k-1}\sigma^2 e^{-\varphi(t_{2j}-t_1)}\sum_{r=1}^{j}\theta^{2j-2-2(r-1)}(\sigma^2)^{r-1}A^{(r-1)}(2j-2)\sigma^2 e^{-\varphi(t-t_{2j+1})}\theta^{n-(2j+1)}$$

$$-\sum_{j=1}^{k-2}\sigma^2 e^{-\varphi(t_{2j+1}-t_1)}\sum_{r=1}^{j}\theta^{2j-1-2(r-1)}(\sigma^2)^{r-1}A^{(r-1)}(2j-1)\sigma^2 e^{-\varphi(t-t_{2j+2})}\theta^{n-(2j+2)}$$

$$=\sigma^2 e^{-\varphi(t-t_1)}\theta^{n-1} + e^{-\varphi(t-t_1)}\sum_{r=1}^{k-1}\theta^{n-1-2r}(\sigma^2)^{r+1}A^{(r)}(2k-2)$$

$$-\sum_{j=1}^{k-1}e^{-\varphi(t-t_1)}\sum_{r=1}^{j}\theta^{n-1-2r}(\sigma^2)^{r+1}A^{(r-1)}(2j-2)f_{2j}$$

$$-\sum_{j=1}^{k-2}e^{-\varphi(t-t_1)}\sum_{r=1}^{j}\theta^{n-1-2r}(\sigma^2)^{r+1}A^{(r-1)}(2j-1)f_{2j+1}$$

$$=\sigma^2 e^{-\varphi(t-t_1)}\theta^{n-1} + e^{-\varphi(t-t_1)}\sum_{r=1}^{k-1}\theta^{n-1-2r}(\sigma^2)^{r+1}\times$$

$$\left(A^{(r)}(2k-2) - \sum_{j=r}^{k-2}A^{(r-1)}(2j-1)f_{2j+1} - \sum_{j=r}^{k-1}A^{(r-1)}(2j-2)f_{2j}\right)$$

$$=\sigma^2 e^{-\varphi(t-t_1)}\theta^{n-1}$$

The last equality holds because by the definition of the $A^{(r)}(m)$ the bracket term is 0 for each $r = 1,\ldots,k-1$ by the following lemma. The derivation for $n = 2k$ works analogously.

Lemma 6.1. *Let $r \in \mathbb{N}$. For any $k \geq r+1$ it holds*

$$A^{(r)}(2k-2) = \sum_{j=r}^{k-2}A^{(r-1)}(2j-1)f_{2j+1} + \sum_{j=r}^{k-1}A^{(r-1)}(2j-2)f_{2j}$$

Proof. By induction on k, for the induction start $k = r+1$, as in [155],

$$f_{2r}A^{(r-1)}(2r-2) = f_{2r}(f_2 \cdots f_{2r-2}) = A^{(r)(2r)}.$$

And the induction step $k \to k+1$

$$A^{(r)}(2k)$$

$$= \text{(all terms in } A^{(r)}(2k) \text{ excluding } f_{2k}) + \text{(all terms in } A^{(r)}(2k) \text{ including } f_{2k})$$

$$= A^{(r)}(2k-1) + f_{2k}A^{(r-1)}(2k-2)$$

$$= A^{(r)}(2k-2) + f_{2k-1}A^{(r-1)}(2k-3) + f_{2k}A^{(r-1)}(2k-2)$$

$$= \sum_{j=r}^{k-2} A^{(r-1)}(2j-1)f_{2j+1} + \sum_{j=r}^{k-1} A^{(r-1)}(2j-2)f_{2j} + f_{2k-1}A^{(r-1)}(2k-3) + f_{2k}A^{(r-1)}(2k-2)$$

$$= \sum_{j=r}^{k-1} A^{(r-1)}(2j-1)f_{2j+1} + \sum_{j=r}^{k} A^{(r-1)}(2j-2)f_{2j}.$$

$\square$

6.1.2 Stochastic filtering for the random telegraph model

Suppose $(Z_t)_{t\geq 0}$ is the random telegraph model, i.e., two-state Markov on $\mathcal{Z} = \{0,1\}$ with the generator

$$\mathcal{A} = \begin{pmatrix} -c_1 & c_2 \\ c_1 & -c_2 \end{pmatrix}.$$

The intensity function $\lambda \colon \mathcal{Z} \to \mathbb{R}_{\geq 0}$ maps the input to $\lambda_0 := \lambda(0) \geq 0$, which is called the dark current, and $\lambda_1 := \lambda(1) > \lambda(0)$. Define the amplitude $\Delta\lambda := \lambda_1 - \lambda_0$.

Proposition 6.2. *The doubly stochastic counting process $(Y_t)_{t\geq 0}$ with intensity λ_t is an (F, l, f)-counting process with $\dim(Y_t)_{t\geq 0} = 1$ for $\lambda_0 > 0$ and $\dim(Y_t)_{t\geq 0} = 0$ for $\lambda_0 = 0$.*

The proof proceeds by establishing the local characteristics of the corresponding Dirac-PDMPs.

Proof. According to proposition 4.10, the MMPP $(Y_t)_{t\geq 0}$ is an (F, l, f)-counting process. The remark 4.18 establishes 1, when $\lambda_0 > 0$, and 0, when $\lambda_0 = 0$, as upper bounds for $\dim(Y_t)_{t\geq 0}$. Applying the Snyder filter (proposition 2.10), the pair $(\Pi_t(0), \Pi_t(1))$ of posterior probabilities evolves according to (8)

$$d\Pi_t(0) = (c_2\Pi_t(1) - c_1\Pi_t(0))\, dt$$
$$+ \frac{(\lambda_0 - \hat{\lambda}_t)\Pi_{t-}(0)}{\hat{\lambda}_t}\{\, dY_t - \hat{\lambda}_t\, dt\} \tag{143}$$

$$d\Pi_t(1) = (c_1\Pi_t(0) - c_2\Pi_t(1))\, dt$$
$$+ \frac{(\lambda_1 - \hat{\lambda}_t)\Pi_{t-}(1)}{\hat{\lambda}_t}\{\, dY_t - \hat{\lambda}_t\, dt\} \tag{144}$$

and $(Y_t)_{t\geq 0}$ jumps with hazard $\hat{\lambda}_t = \Pi_{t-}(0)\lambda_0 + \Pi_{t-}(1)\lambda_1$. This provides the canonical two-dimensional Dirac-PDMP which demonstrates that $(Y_t)_{t\geq 0}$ is an (F, l, f)-counting process. As in remark 4.18(i), we can replace $\Pi_t(0) = 1 - \Pi_t(1)$ due to the conservation of probability mass for the conditional probabilities, which yields the one-dimensional sufficient state variable $(\Pi_t(1))_{t\geq 0}$ of Markovian progression. Then $\tilde{\lambda}_t = \lambda_0 + \Delta\lambda\Pi_t(1)$, the right-continuous version of $\hat{\lambda}_t$, is an affine-linear transform and hence in bijection with $\Pi_t(1)$. Consequently, $(\tilde{\lambda}_t)_{t\geq 0}$ is an equivalent choice for a one-dimensional sufficient state variable of Markovian progression. By abuse of the notation, we now replace $\tilde{\lambda}_t$ by the state variable $\hat{\lambda}_t$ with the same symbol as the CI, even though $\hat{\lambda}_t$ is the left-continuous version of $\tilde{\lambda}_t$. Using the Ito-rule for Dirac-PDMPs (proposition 4.8) on Eq. (144), we obtain the evolution Eq. (220).

This equation establishes F and f as to match Eq. (83), i.e., $f(\hat{\lambda}) = \hat{\lambda}^{-1}(\hat{\lambda} - \lambda_0)(\lambda_1 - \hat{\lambda}) + \hat{\lambda}$, and the hazard is $l = \mathrm{id}$ of the Dirac-PDMP on the state space $\vartheta = [\lambda_0, \lambda_1]$. By $l(\hat{\lambda}_{t-}) = \hat{\lambda}_{t-}$, it is clear that the CI is meant to be the left-continuous version of the state variable. The function

$u\colon [0,\infty) \times \vartheta \to \vartheta$, Eq. (84), links the Dirac-PDMP to the BReT-P standard form via Eq. (85). To derive u define $\tilde{\omega}_1 < \tilde{\omega}_2$ as the roots of the quadratic equilibrium equation

$$0 = c_1 \Delta\lambda - (c_2 + c_1 + \Delta\lambda)\omega + \omega^2$$

and $\omega_i := \tilde{\omega}_i + \lambda_0, i = 1, 2$. With $\Delta\omega := \omega_2 - \omega_1$ we obtain

$$u(\tau,\theta) = \omega_2 - \frac{\Delta\omega}{1 + \frac{\theta - \omega_1}{\omega_2 - \theta} e^{-\Delta\omega\tau}}.$$

For the local characteristics m and g of the BReT-P it holds $m = \mathrm{id} \circ u = u, g = f \circ u$. Taking a look at the jump update f reveals that for $\lambda_0 = 0$ we have

$$f(\hat{\lambda}) = \frac{\hat{\lambda}(\lambda_1 - \hat{\lambda})}{\hat{\lambda}} + \hat{\lambda} = \lambda_1,$$

hence the upper interval bound of the state space $\vartheta = [\lambda_0, \lambda_1]$ is targeted independent of the state θ. This is equivalent to $\Pi_t(0) = \lambda_0 \Pi_{t-}(0)/\hat{\lambda}_t = 0$ for $t = \sigma_i$ as discussed in remark 4.18(ii). It reduces the dimension of the BReT-P to 0, i.e., $(\tau(t))_{t\geq 0}$ is a Markov process on its own with $m(\tau) = u(\tau, \lambda_1)$. Then $\dim(Y_t)_{t\geq 0} = 0$ by definition 4.17, when $\lambda_0 = 0$. For $\lambda_0 > 0$, Eq. (220) shows that it is impossible to parametrize the CI only with the backward recurrence time, because the image of f is not a singleton. $\qquad\square$

Hence, we showed the claim and established the local characteristics (F, l, f) and (m, g) of the MMPP to prepare the computation of the MIR in section 9.2.1.

6.1.3 Optimal linear filter for structure switching

We consider the equilibrated optimal linear (Hawkes) approximation of the conversion system in Eqs. (136) - (141), i.e., $\hat{\lambda}^{\mathrm{H}}(t)$ as in Eq. (68). For this system, we provide the equilibrium value of the ODE for $\hat{\lambda}^{\mathrm{H}}(t)$ between jumps, from which we compute $\hat{Z}^{\mathrm{H}}(t)$ that is shown to be positive. For the derivations, we require $C \in \mathbb{R}^{1 \times l}$ and A to be invertible. For the conversion process, this is not the case, since $\mathbf{1}A = 0$, but by truncation to $l = k - 1$ we can enforce it, see the Appendix B.

The equilibrium $\bar{z}$ of the linear dynamics

$$\dot{z} = -(A + \bar{B}C^T \mu^{-1} C)z - \bar{B}C^T \tag{145}$$

between jumps is evaluated by the Sherman-Morrison formula as

$$\bar{z} = -(A + \bar{B}C^T C \mu^{-1})^{-1}\bar{B}C^T$$
$$= -A^{-1}\bar{B}C^T + \frac{A^{-1}\bar{B}C^T \mu^{-1} C A^{-1}\bar{B}C^T}{1 + CA^{-1}\bar{B}C^T \mu^{-1}} = \frac{-A^{-1}\bar{B}C^T}{1 + CA^{-1}\bar{B}C^T \mu^{-1}}.$$

We assume that $(A + \bar{B}C^T \mu^{-1} C)$ has only eigenvalues with positive real part. Then the equilibrium is stable and the equilibrium value $\bar{\lambda}$ of $\hat{\lambda}^{\mathrm{H}}(t)$ between jumps is

$$\bar{\lambda} = \mu + C\bar{z} = \mu - \frac{\gamma}{1 + \gamma\mu^{-1}} = \frac{\mu}{1 + \gamma\mu^{-1}}$$
$$\gamma = CA^{-1}\bar{B}C^T.$$

Furthermore, jumps of $z(t)$ by $\bar{B}C^T \mu^{-1}$ increase $\hat{\lambda}^{\mathrm{H}}(t)$ by

$$\Delta\hat{\lambda}^{\mathrm{H}} = C\bar{B}C^T \mu^{-1} \geq 0.$$

Proposition 6.3. *The term* $\gamma = CA^{-1}\bar{B}C^T$ *is the positive root of*

$$0 = 2CA^{-1}\Sigma C - 2y - y^2 \mu^{-1}.$$

Proof. First, we note that $CA^{-1}\Sigma C^T \geq 0$, since

$$2CA^{-1}\Sigma C^T = CA^{-1}(A\Sigma + \Sigma A^T)A^{-T}C^T$$

and by condition (C2) $A\Sigma + \Sigma A^T$ is positive semidefinite. Since $B(0) = \Sigma$, we have

$$B(0) = -\Sigma C^T \mu^{-1} C\Sigma = -\Sigma C^T \sqrt{\mu^{-1}}(\Sigma C^T \sqrt{\mu^{-1}})^T \leq 0.$$

Then by [156, Corollary 3] $B(t)$ is monotonically non-increasing. By the property of complete observability, $B(t)$ converges monotonically to the equilibrium $\bar{B} \geq 0$. Define $x(t) := CA^{-1}B(t)A^{-T}C^T$. Then by the monotonicity of $B(t)$, it holds that $\frac{d}{dt}x(t) < 0$. With minor algebraic manipulation, we express

$$\frac{d}{dt}x(t) = 2CA^{-1}\Sigma C^T - 2y(t) - y(t)^2\mu^{-1} =: f(y(t))$$

in terms of $y(t) = CA^{-1}B(t)C^T$. Then necessarily, at equilibrium of $x(t)$ and $y(t)$, it holds $0 = 2CA^{-1}\Sigma C^T - 2\gamma - \gamma^2\mu^{-1}$. The positive and the negative solution γ_1 and γ_2 split the real line into three intervals $(-\infty, \gamma_2), (\gamma_2, \gamma_1), (\gamma_1, \infty)$ on which $f(y)$ is negative, positive, negative. The initial value $y(0) = CA^{-1}\Sigma C^T$ is positive and $f(y(0)) < 0$, hence $y(0) \in (\gamma_1, \infty)$. But since $f(y(t)) = \frac{d}{dt}x(t) < 0$ for all $t \geq 0$, the solution $y(t) \in (\gamma_1, \infty)$ for all $t \geq 0$. Hence, γ is the positive root γ_1. $\square$

Then immediately we obtain

$$\bar{\lambda} = \frac{\mu}{\sqrt{1 + 2CA^{-1}\Sigma C^T\mu^{-1}}} \in (0, \mu). \tag{146}$$

For the condition (C5) to hold we provide the following sufficient criterion which guarantees that $\hat{\lambda}^{\mathrm{H}}(t)$ stays positive at all times.

Proposition 6.4. *Suppose that $z_1, \ldots, z_l$ form a basis of real eigenvectors of $(A + \bar{B}C^T\mu^{-1}C)$ for the positive eigenvalues $\nu_1, \ldots, \nu_l$, such that $Cz_i > 0$ for all $i = 1, \ldots, l$. Let L be the matrix whose columns are the eigenvectors $z_1, \ldots, z_l$ and assume that*

$$L^{-1}\bar{B}C^T \text{ has only positive entries.} \tag{D1}$$

Then $\hat{\lambda}^{\mathrm{H}}(t) > 0$ for all $t \geq 0$.

Proof. We consider the half-space $\Lambda_+ := \{z; Cz + \mu > 0\}$ with boundary $\{z; Cz + \mu = 0\}$ as well as the inward normal vector C, and the cone $U := \{\bar{z} + \sum_i \alpha_i z_i; \alpha_i > 0\}$. By Eq. (146), $\bar{z} \in \Lambda_+$ and hence $Cz_i > 0$ implies $U \subseteq \Lambda_+$. By assumption, there exist positive $\beta_i, i = 1, \ldots, l$ such that $\bar{B}C^T\mu^{-1} = \sum_i \beta_i z_i$. Consequently, $f(U) \subseteq U$ for the map $f(u) := u + \bar{B}C^T\mu^{-1}$, which means that jumps of the trajectory $\hat{Z}^{\mathrm{H}}(t)$ by $\bar{B}C^T\mu^{-1}$ remain in U. The initial value $\hat{Z}^{\mathrm{H}}(0) = 0$ is in U because

$$L^{-1}(0 - \bar{z}) = L^{-1}(A + \bar{B}C^T\mu^{-1}C)^{-1}\bar{B}C^T = \mathrm{diag}\{\nu\}^{-1}L^{-1}\bar{B}C^T \tag{147}$$

has only positive entries. Finally, we verify that the dynamics in Eq. (145) leaves U invariant. Let z be any boundary point of U, i.e. $z - \bar{z} = \sum_i \beta_i z_i, \beta_1, \ldots, \beta_l \geq 0$ and there exists i such that $\beta_i = 0$. Let η be any corresponding inward normal vector, in particular, $\langle z_i, \eta \rangle = 0$ for all $i = 1, \ldots, l$. Then

$$\langle \dot{z}, \eta \rangle = \langle -(A + \bar{B}C^T\mu^{-1}C)(z - \bar{z}), \eta \rangle = \langle -\sum_i \beta_i \nu_i z_i, \eta \rangle = -\sum_i \beta_i \nu_i \langle z_i, \eta \rangle = 0$$

$$\square$$

As a remark, we note that the condition D1 is equivalent to

$$-L^{-1}\bar{z} \text{ has only positive entries,} \tag{D2}$$

by Eq. (147). Geometrically, the condition corresponds to requiring that the ray starting from $\bar{z}$ and going through 0 lies within the cone U, defined in the proof.

6.2 Birth-death process

The birth-death process is a one-dimensional linear reaction network with zeroth-order birth (or synthesis) reaction and first-order death (or decay) reaction

$$R_1, R_2: \quad \emptyset \underset{\mu}{\overset{\lambda}{\rightleftharpoons}} Z, \tag{148}$$

Its stationary distribution is the Poisson with mean $\frac{\lambda}{\mu}$. Its mean evolution reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Z(t)] = \lambda - \mu\mathbb{E}[Z(t)]$$

and its autocovariance function in the weak stationary case is $\mathrm{Cov}[Z(t), Z(s)] = \frac{\lambda}{\mu}e^{-\mu|t-s|}$. More generally, its transient probability distribution remains invariant in the class of Poisson distributions for time-varying birth and death rate parameters.

Proposition 6.5. *For a birth-death process $X(t)$ with time-dependent birth rate $\lambda(t)$ and death rate $\mu(t)$ and Poisson(Λ_0)-distributed $X(0)$, the marginals $X(t)$ are Poisson-distributed, i.e.*

$$P[X(t) = k] = \frac{\Lambda(t)^k}{k!}e^{-\Lambda(t)}$$

where Λ is such, that is solves

$$\dot{\Lambda}(t) = \lambda(t) - \mu(t)\Lambda(t), \Lambda(0) = \Lambda_0.$$

By the variation of constant formula

$$\Lambda(t) = \Lambda_0 \exp\left(-\int_0^t \mu(\tau) \, \mathrm{d}\tau\right) + \int_0^t \lambda(s) \exp\left(-\int_s^t \mu(\tau) \, \mathrm{d}\tau\right) \, \mathrm{d}s.$$

If $\Lambda_0 = 0$, then $P[X(0) = 0] = 1$.

Proof. By means of generating functions: Define $p(t, n) = P[X(t) = n]$

$$g(t, s) = \sum_{n=0}^{\infty} s^n p(t, n)$$

Then $g(t, s)$ has to solve the partial differential equation

$$\partial_t g(t, s) = (s - 1)(\lambda(t)g(t, s) - \mu(t)\partial_s g(t, s))$$

with initial condition

$$g(0, s) = \exp((s - 1)\Lambda_0).$$

The function $g(t, s) = \exp((s - 1)\Lambda(t))$ satisfies

$$\partial_s g(t, s) = \Lambda(t)g(t, s)$$

and

$$\partial_t g(t, s) = \dot{\Lambda}(t)(s - 1)g(t, s)$$

and thus solves the PDE. But $[0, 1] \ni s \mapsto g(t, s)$ is the generating function of Poisson$(\Lambda(t))$. $\qquad\square$

Remark 6.6. This result can also be obtained by just checking, that

$$p(t, n) = \frac{\Lambda(t)^n}{n!}e^{-\Lambda(t)}$$

satisfies

$$\partial_t p(t, n) = \lambda(t)(p(t, n - 1) - p(t, n)) - \mu(t)((n + 1)p(t, n + 1) - np(t, n))$$

The Poisson distribution is a limit distribution of the (N, p)-binomial distribution for $N \to \infty, p \to 0, Np \to \frac{\lambda}{\mu}$. More generally, the birth-death process can be seen as a limit of the conversion process with two structures, under the limit $N \to \infty, c_{1,2} \to 0, Nc_{1,2} \to \lambda$.

6.3 CIR process

The CIR process is a Markov process on a continuous state space, that evolves according to the SDE

$$
\mathrm{d}Z(t) = \gamma(\mu - Z(t)) \; \mathrm{d}t + \sqrt{\frac{2\gamma\sigma^2 Z(t)}{\mu}} \; \mathrm{d}W(t) \tag{149}
$$

with velocity of adjustment γ, saturation level μ, diffusion parameter σ and a standard one-dimensional Brownian motion. It was first introduced to model interest rates [57]. The condition $\mu \geq \sigma^2$ is sufficient for the process to remain positive. The corresponding Fokker-Planck operator reads

$$
(\mathcal{A}p)(z,t) = \gamma p(z,t) + \gamma \left(\frac{2\sigma^2}{\mu} - (\mu - z) \right) \partial_z p(z,t) + \frac{\gamma\sigma^2}{\mu} z \partial_{zz} p(z,t)
$$

with the Gamma distribution $\mathcal{G}(\mu^2/\sigma^2, \mu/\sigma^2)$ as its stationary distribution. By $Z \sim \mathrm{CIR}(\mu, \gamma, \sigma^2)$, we abbreviate a CIR process with respective parameter triple, initialized in its stationary distribution. Mean and auto-covariance for a CIR process, initialized in a distribution with mean μ_0 and variance σ_0^2 read

$$
\mathbb{E}[Z(t)] = \mu + e^{-\gamma t}(\mu_0 - \mu) \tag{150}
$$

and

$$
\mathrm{Cov}[Z(s), Z(t)] = \sigma^2 e^{-\gamma(t-s)} + \sigma^2 e^{-\gamma(s+t)} \left(1 - \frac{2\mu_0}{\mu} + \frac{\sigma_0^2}{\sigma^2} \right) + 2\sigma^2 e^{-\gamma t} \left(\frac{\mu_0}{\mu} - 1 \right). \tag{151}
$$

The derivation uses martingale techniques as follows. Analogously to [63, example 4.5.4], derive an SDE for the transformed process

$$
Y(t) = (Z(t) - \mu)e^{\gamma t}.
$$

The intuition behind choosing this transformation is the following:

$$
z(t) = \mu - e^{-\gamma t}
$$

solves the deterministic equation $\mathrm{d}Z(t) = \gamma(\mu - Z(t)) \; \mathrm{d}t$, which by the transformation is transformed to constant -1. Hence figuratively speaking, only the fluctuating part survives. Applying Ito's formula to $Y(t)$:

$$
\begin{aligned}
\mathrm{d}Y(t) &= \; \mathrm{d}(Z(t) - \mu) \; \mathrm{d}(e^{\gamma t}) + \; \mathrm{d}(Z(t) - \mu)e^{\gamma t} + (Z(t) - \mu) \; \mathrm{d}(e^{\gamma t}) \\
&= 0 + (\gamma(\mu - Z(t)) \; \mathrm{d}t + \sigma\sqrt{\frac{2\gamma}{\mu} Z(t)} \; \mathrm{d}W(t))e^{\gamma t} + \gamma(Z(t) - \mu)e^{\gamma t} \; \mathrm{d}t \\
&= \sigma e^{\gamma t} \sqrt{\frac{2\gamma}{\mu} Z(t)} \; \mathrm{d}W(t).
\end{aligned}
$$

Upon resubstituting, we obtain

$$
Z(t) = \mu + e^{-\gamma t}(Z(0) - \mu) + e^{-\gamma t}\sigma \int_0^t e^{\gamma s} \sqrt{\frac{2\gamma}{\mu} Z(s)} \; \mathrm{d}W(s).
$$

Then we calculate the mean

$$
\mathbb{E}[Z(t)] = \mu + e^{-\gamma t}(\mathbb{E}[Z(0)] - \mu)
$$

and, by using that for the martingale

$$
M(t) = \int_0^t e^{\gamma s} \sqrt{\frac{2\gamma}{\mu} Z(s)} \; \mathrm{d}W(s)
$$

it holds $\mathbb{E}[M(s)M(t)] = \mathbb{E}[M(s)^2]$, see below, the covariance for $s \leq t$

$$\mathbb{E}[Z(t)Z(s)] - \mathbb{E}[Z(s)]\mathbb{E}[Z(t)]$$

$$= e^{-\gamma(s+t)}\mathbb{V}\mathrm{ar}[Z(0)] + 0 + 0$$

$$+ e^{-\gamma(s+t)}\sigma^2 \mathbb{E}\left[\left(\int_0^t e^{\gamma s'}\sqrt{\frac{2\gamma}{\mu}Z(s')}\,\mathrm{d}W(s')\right)\left(\int_0^s e^{\gamma s'}\sqrt{\frac{2\gamma}{\mu}Z(s')}\,\mathrm{d}W(s')\right)\right]$$

$$= e^{-\gamma(s+t)}\mathbb{V}\mathrm{ar}[Z(0)] + e^{-\gamma(s+t)}\sigma^2 \mathbb{E}\left[\left(\int_0^s e^{\gamma s'}\sqrt{\frac{2\gamma}{\mu}Z(s')}\,\mathrm{d}W(s')\right)^2\right]$$

$$= e^{-\gamma(s+t)}\mathbb{V}\mathrm{ar}[Z(0)] + e^{-\gamma(s+t)}\sigma^2 \int_0^s \mathbb{E}\left[\left(e^{\gamma s'}\sqrt{\frac{2\gamma}{\mu}Z(s')}\right)^2\right]\,\mathrm{d}s'$$

$$= e^{-\gamma(s+t)}\mathbb{V}\mathrm{ar}[Z(0)] + e^{-\gamma(s+t)}\sigma^2 \int_0^s e^{2\gamma s'}\frac{2\gamma}{\mu}E[Z(s')]\,\mathrm{d}s'$$

$$= \sigma^2 e^{-\gamma(t-s)} + \sigma^2 e^{-\gamma(s+t)}\left(1 - \frac{2\mathbb{E}[Z(0)]}{\mu} + \frac{\mathbb{V}\mathrm{ar}[Z(0)]}{\sigma^2}\right)$$

$$+ 2\sigma^2 e^{-\gamma t}\left(\frac{\mathbb{E}[Z(0)]}{\mu} - 1\right)$$

For martingales $\{X(t)\}_{t \geq 0}$ and $0 \leq s \leq t$, it holds $\mathbb{E}[X(s)X(t)] = \mathbb{E}[X(s)^2]$, because

$$\mathbb{E}[X(s)X(t)] = \mathbb{E}[X(s)^2] + \mathbb{E}[X(s)(X(t) - X(s))] = \mathbb{E}[X(s)^2] + \mathbb{E}[X(s)\mathbb{E}[X(t) - X(s)|\mathcal{F}_s]] = \mathbb{E}[X(s)^2],$$

where the second equality follows from the $\mathcal{F}_s$-measurability of the factor $X(s)$ combined with the tower property $\mathbb{E}[\mathbb{E}[Y|\mathcal{G}]] = \mathbb{E}[Y]$ and the third equality is implied by the martingale property.

The following expression for the variance

$$\mathbb{V}\mathrm{ar}[Z(t)] = e^{-2\gamma t}\mathbb{V}\mathrm{ar}[Z(0)] + \sigma^2 \int_0^s e^{-2\gamma(t-s)}\frac{2\gamma}{\mu}\mathbb{E}[Z(s)]\,\mathrm{d}s$$

implies that the linear differential equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{V}\mathrm{ar}[Z(t)] = -2\gamma\left(\mathbb{V}\mathrm{ar}[Z(t)] - \frac{\sigma^2}{\mu}\mathbb{E}[Z(t)]\right), \quad \mathbb{V}\mathrm{ar}[Z(0)] = \sigma_0^2$$

is satisfied. In summary, mean and variance solve the closed differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[Z(t)] = -\gamma(\mathbb{E}[Z(t)] - \mu), \qquad\qquad \mathbb{E}[Z(0)] = \mu_0$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{V}\mathrm{ar}[Z(t)] = -2\gamma\left(\mathbb{V}\mathrm{ar}[Z(t)] - \frac{\sigma^2}{\mu}\mathbb{E}[Z(t)]\right), \qquad\qquad \mathbb{V}\mathrm{ar}[Z(0)] = \sigma_0^2.$$

In case of a CIR started in $\mu_0 = \mu, \sigma_0^2 = \sigma^2$, the expressions simplify and read

$$\mathbb{E}[Z(t)] = \mu, \quad \mathrm{Cov}[Z(s), Z(t)] = \sigma^2 e^{-\gamma|t-s|}.$$

In Eq. (63), for a CIR process $Z(t)$, we have

$$F(\Theta, V) = -\gamma(\Theta - \mu), \quad G(\Theta, V) = -2\gamma\left(V - \frac{\sigma^2}{\mu}\Theta\right). \tag{152}$$

Consequently, the first and second conditional moment equations in Eq. (64) for a CIR process $Z(t)$ read

$$\mathrm{d}\Theta(t) = (-\gamma(\Theta(t) - \mu) - cV(t))\,\mathrm{d}t + \frac{V(t-)}{\Theta(t-)}\,\mathrm{d}Y_j(t) \tag{153}$$

$$\mathrm{d}V(t) = \left[-2\gamma\left(V(t) - \frac{\sigma^2}{\mu}\Theta(t)\right) - cS(t)\right]\,\mathrm{d}t + \left[\frac{S(t-)}{\Theta(t-)} - \frac{V^2(t-)}{\Theta^2(t-)}\right]\,\mathrm{d}Y_j(t) \tag{154}$$

The CIR-modulated reactions have motivated the approximate Gamma filter, see section 8.2.3 below.

6.4 Static Gamma heterogeneity

For a static environment random variable, the Gamma distribution is a convenient choice. Unlike the normal distribution it has a positive support. Unlike the exponential distribution it has two parameters such that its mean and variance can be independently chosen. The Gamma distribution has the probability density function

$$p(z) \propto z^{\alpha-1} e^{-z\beta}, \quad z > 0$$

where α denotes the shape parameter and β the rate parameter. The normalization constant is

$$\int p(z) \, \mathrm{d}z = \frac{\Gamma(\alpha)}{\beta^\alpha}.$$

The mean and variance of a $\mathcal{G}(\alpha, \beta)$-distributed Z are $\mathbb{E}[Z] = \frac{\alpha}{\beta}$ and $\mathbb{V}\mathrm{ar}[Z] = \frac{\alpha}{\beta^2}$. For $\mathcal{G}(\alpha, \beta)$-distributed Z and $Y|Z = z$ Poisson distributed with mean z, the conditional distribution $Z|Y = k$ is $\mathcal{G}(\alpha + k, \beta + 1)$. This property is called the Gamma distribution's conjugacy with the Poisson distribution. It is a convenient property of the Gamma distribution for modulated counting processes, compared to the lognormal distribution.

The hybrid master equation, Eq. (87), is available for the static Gamma heterogeneity. Setting $Z_j, j = 1, \ldots, M$ are independent Gamma-distributed random variables with shape α_j and rate β_j. The heterogeneous reaction rate of the j-th reaction is $Z_j a_j(n(t))$, where $n(t)$ is the vector of species copy numbers at time t and a_j is the state-dependent propensity function. Introduce the sufficient statistics $W(t) = (W_1(t), \ldots, W_M(t))$ with $W_j(t) = \int_0^t a_j(n(t')) \, \mathrm{d}t'$. Then $\dot{W}_j(t) = a_j(n(t))$ and hence, we apply Eq. (87) to obtain

$$
\begin{aligned}
\partial_t p(t, x, w) = \sum_{j=1}^M m_j(x - \nu_j, w) p(t, x - \nu_j, w) \\
- m_j(x, w) p(t, x, w) - \sum_{j=1}^M \lambda_j(x) \partial_{\omega_j} p(t, x, w).
\end{aligned}
\tag{155}
$$

as in [106, p.96]. When the state $X(t)$ describes the reaction count vector, and the initial species copy number vector $n(0)$ is fixed, then $\nu_j = e_j$, and, building on the previous work, $m_j(x, w) = \frac{\alpha_j + x_j}{\beta_j + w_j} \lambda_j(x)$ and $\lambda_j(x) = a_j(n(0) + \sum_{j=1}^M x_j \nu_j)$ with the state-dependent propensity function $a_j(n)$. The evolution equation of $p(t, x)$ obtained from marginalization over w. Marginalize out w as in Eq. (106) to obtain

$$
\begin{aligned}
\partial_t p(t, x) &= \sum_{j=1}^M \mathbb{E}\left[\frac{\alpha_j + x_j - 1}{\beta_j + W_j(t)} \lambda_j(x - e_j) \mathbb{1}(X(t) = x - e_j)\right] - \mathbb{E}\left[\frac{\alpha_j + x_j}{\beta_j + W_j(t)} \lambda_j(x) \mathbb{1}(X(t) = x)\right] \\
&= \sum_{j=1}^M \left\langle \frac{\alpha_j + x_j - 1}{\beta_j + W_j(t)} \Big| x - e_j, t \right\rangle \lambda_j(x - e_j) p(t, x - e_j) - \left\langle \frac{\alpha_j + x_j}{\beta_j + W_j(t)} \Big| x, t \right\rangle \lambda_j(x) p(t, x).
\end{aligned}
$$

The equation is not closed, but depends on the expected value of the propensities. We solve the expected values for the particular example, section 8.5.1, below.

7 Marginal simulation

The joint simulation of the external process $Z(t)$ and the reaction system $X(t)$ provides samples for $Z(t)$ and $X(t)$. Disregarding $Z(t)$, as this is mostly not of interest, means a marginal simulation of $X(t)$ that behaves in its statistical properties as if it was still embedded. It can be obtained using the $\mathcal{F}_t^Y$-intensity. Simulation with an approximate instead of Snyder's exact filter yields an approximate marginal simulation.

Approximate marginal simulation using the aforedescribed filters requires algorithm that can handle time- and history-dependent propensities. Here, we use the modified next reaction algorithm by

Anderson [157]. It proceeds by assigning each reaction counter an internal time that progresses as a unit Poisson process. Then the internal time is transformed to the global time via the inverse of the cumulative intensity. We now apply this algorithm for the simulation of a reaction system $X(t)$ that has probability $\mathbb{Q}$, i.e., with $\mathcal{F}_t^Y$-intensity $\hat{\lambda}^{\mathrm{OL}}(t)$ given by Eq. (65)-(67). To this end, we need to propagate the approximate state estimate $\hat{Z}(t)$ and posterior covariance mock $B(t)$ jointly between jumps via

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{Z}(t) = -A\hat{Z}(t) - B(t)C^T \operatorname{diag}\{\kappa(\mu)\}(\mu + C\hat{Z}(t)) \tag{156}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}B(t) = A\Sigma + \Sigma A^T - AB(t) - B(t)A^T - B(t)C^T \operatorname{diag}\{\kappa(\mu)\}CB(t). \tag{157}$$

The coefficients μ_j and C_{jk} can be state-dependent. Additionally, each reaction R_j gets assigned another state variable $\Lambda_j(t)$, its cumulative intensity, which evolves as

$$\dot{\Lambda}_j(t) = \mu_j + \langle C_{j*}, \hat{Z}(t)\rangle. \tag{158}$$

We collect them in the vector $\Lambda(t) = [\Lambda_1(t), \ldots, \Lambda_M(t)]$. For determining which reaction occurs, we need to check whether the cumulative intensity has reached the internal time difference. For this we define events, i.e., time points

$$t_j : \Lambda_j(t_j) = \Delta T_j. \tag{159}$$

Finally, $\hat{Z}(t)$ needs to be updated when the reaction R_j occurs, i.e.,

$$\hat{Z}(t+) = \hat{Z}(t-) + B(t)C^T \operatorname{diag}\{\kappa(\mu)\}e_j, \tag{160}$$

where e_j is the j-th unit vector. The complete algorithm reads as follows.

Algorithm 1 Approximate marginal simulation with the optimal linear filter

---

1: **Input**: System parameters A, Σ, state-dependent functions $\mu(x), C(x)$, the stoichiometric matrix N, the initial value X_0 and length T of the simulated time interval.
2: Initialize internal times $T_j = 0$, jump times $P_j \sim \operatorname{expo}(1)$ and $\Delta T_j = P_j - T_j, j = 1, \ldots, M$.
3: Initialize global time $t_0 = 0$.
4: Initialize coefficients $\mu = \mu(X_0), C = C(X_0)$.
5: Initialize $\hat{Z}(t_0) = 0$, $B(t_0) = \Sigma$, $\Lambda(t_0) = 0$.
6: Save $t = t_0$. Save $X = X_0$.
7: **while** $t_0 < T$ **do**
8:     Set events as in Eq. (159), $j = 1, \ldots, M$
9:     Evolve $\hat{Z}, B, \Lambda$ jointly according to Eq. (156)-(158) on $[t_0, \tau]$ until $\tau = \min_j\{t_j\}$.
10:     Assign $J = \arg\min_j\{t_j\}$.
11:     Update $t_0 \leftarrow \tau$ and $\hat{Z}(t_0) \leftarrow \hat{Z}(\tau)$, $B(t_0) \leftarrow B(\tau)$, $\Lambda(t_0) \leftarrow \Lambda(\tau)$. Save $t \leftarrow [t, t_0]$.
12:     Update state $X_0 \leftarrow X_0 + \nu_J$. Save $X \leftarrow [X, X_0]$.
13:     Update internal times $T_j \leftarrow T_j + \Lambda_j(t_0), j = 1, \ldots, M$. Reset $\Lambda(t_0) \leftarrow 0$.
14:     Draw $\Delta P \sim \operatorname{expo}(1)$. Update jump time $P_J \leftarrow P_J + \Delta P$. Update $\Delta T_j \leftarrow P_j - T_j, j = 1, \ldots, M$.
15:     Update $\hat{Z}(t_0)$ according to Eq. (160) with $j = J$.
16:     Update coefficients $\mu = \mu(X_0), C = C(X_0)$.
17: **end while**
18: **return** Jump times t and states X.

We formulated the algorithm for Anderson's modified next reaction method, but other algorithms that can handle time-dependent propensities [158, 159] can replace it, i.e., lines 3, 8, 10, 13, 14 must be modified.

7.1 Independent environment components

When the environment has independent components, the line 9 of the algorithm can be simplified. Assume a diagonal structure for $A = \operatorname{diag}\{\alpha\}, \Sigma = \operatorname{diag}\{\sigma^2\}$ and $C = \operatorname{diag}\{c\}$, i.e., the environment

components $Z_1(t), \ldots, Z_M(t)$ are stochastically independent and $Z_j(t)$ modulates R_j via $\mu_j(X(t)) + c_j(X(t))Z_j(t)$. In this case $B(t)$ is also diagonal, say $\mathrm{diag}\{\beta(t)\}$. We assume that $\mu_j(X(t)) = 0$ implies $c_j(X(t)) = 0$ to avoid negative rates. We analytically solve the joint ODE system of $\hat{Z}_j(t), \beta_j(t), \Lambda_j(t)$ for $j = 1, \ldots, M$ separately. For the derivation, we first solved the one-dimensional Riccati equation for $\beta_j(t)$ with standard techniques. Upon plugging $\beta_j(t)$ into the evolution of $\hat{Z}_j(t)$, the variation of constants formula for linear systems with time-dependent coefficients yielded $\hat{Z}_j(t)$, where we took advantage of the logarithmic derivative $\frac{\mathrm{d}}{\mathrm{d}t} \ln \xi_j(t)$ in $\beta_j(t)$. Note that the derivative and first integral of $\xi_j(t)$ are analytically expressed as linear combinations of $a_j(t)$ and $b_j(t)$, together yielding the expression for $\hat{\lambda}_j^{\mathrm{OL}}(t)$. Finally, for $\Lambda_j(t)$, the integral was solved by the technique of partial fractions and the substitution $u = e^{\rho_j(t - t_0)}$. Altogether, if $c_j \neq 0$,

$$\hat{Z}_j(t) = \hat{Z}_j(t_0) \frac{\xi_j(t_0)}{\xi_j(t)} + \frac{\mu_j}{c_j \xi_j(t)} \left(\int_{t_0}^t \alpha_j \xi_j(\tau) - \xi_j'(\tau) \, \mathrm{d}\tau \right)$$

$$\beta_j(t) = \frac{\mu_j}{c_j^2} \left(\frac{\xi_j'(t)}{\xi_j(t)} - \alpha_j \right)$$

$$\xi_j(t) = \rho_j a_j(t) + \left(\alpha_j + \frac{c_j^2}{\mu_j} \beta_j(t_0) \right) b_j(t)$$

$$\rho_j = \sqrt{\alpha_j^2 + \frac{2\alpha_j \sigma_j^2 c_j^2}{\mu_j}}$$

$$a_j(t) = \cosh(\rho_j \cdot (t - t_0)), \quad b_j(t) = \sinh(\rho_j \cdot (t - t_0))$$

$$\hat{\lambda}_j^{\mathrm{OL}}(t) = \mu_j + c_j \hat{Z}_j(t) = A_j + \frac{-2A_j + B_j e^{\rho_j \cdot (t - t_0)}}{C_j e^{2\rho_j \cdot (t - t_0)} + 1}$$

$$\Lambda_j(t) = \int_{t_0}^t \hat{\lambda}_j^{\mathrm{OL}}(\tau) \, \mathrm{d}\tau$$

$$= \frac{A_j}{\rho_j} \ln \left(\frac{C_j e^{\rho_j(t - t_0)} + e^{-\rho_j(t - t_0)}}{C_j + 1} \right)$$

$$+ \begin{cases} \left[\frac{B_j}{\sqrt{C_j} \rho_j} \arctan \left(\sqrt{C_j} e^{\rho_j(t - t_0)} \right) \right]_{t_0}^t, & C_j > 0 \\[2ex] \left[\frac{B_j}{\sqrt{|C_j|} \rho_j} \operatorname{artanh} \left(\frac{e^{-\rho_j(t - t_0)}}{\sqrt{|C_j|}} \right) \right]_{t_0}^t, & C_j < 0, B_j < 0 \end{cases}$$

$$C_j = \frac{\rho_j + \alpha_j + \frac{c_j^2}{\mu_j} \beta_j(t_0)}{\rho_j - \alpha_j - \frac{c_j^2}{\mu_j} \beta_j(t_0)}, \quad A_j = \frac{\mu_j \alpha_j}{\rho_j},$$

$$B_j = 2 \frac{(c_j \hat{Z}_j(t_0) + \mu_j)\rho_j - \frac{\mu_j \alpha_j (\alpha_j + \frac{c_j^2}{\mu_j} \beta_j(t_0))}{\rho_j}}{\rho_j - \alpha_j - \frac{c_j^2}{\mu_j} \beta_j(t_0)}.$$

If $c_j = 0$,

$$\hat{Z}_j(t) = e^{-\alpha_j(t - t_0)} \hat{Z}_j(t_0)$$

$$\beta_j(t) = \sigma_j^2 + e^{-2\alpha_j(t - t_0)} (\beta_j(t_0) - \sigma_j^2)$$

$$\Lambda_j(t) = \mu_j(t - t_0).$$

Note, that for the case $c_j = 0$ the mimics of the state estimate $\hat{Z}_j(t)$ and the posterior variance $\hat{Z}_j(t)$ follow the prior dynamics of mean and variance. We demonstrate the use of this filter in the case study 8.3.8 when we investigate a two-stage gene expression in uncorrelated random environment.

7.2 Sinzger's min-thin algorithm

For the time-dependent intensities $\lambda(t)$ in the marginal simulation algorithm the modeler requires a first reaction method, which comes with a disadvantage over Doob-Gillespie's direct next reaction method. Namely, proposed waiting times are rejected, when choosing the first reaction, and the time-dependence can make the computation of waiting times inefficient. This disadvantage is mitigated if a large amount of reaction events in the environment complicates the Doob-Gillespie method. Independent of the question, at which threshold of reaction events the disadvantages are on a par, we present an optimized first reaction method for our purposes. The key idea is to sample waiting times with the inverse method from finitely many auxiliary intensities $f_1(t), \ldots, f_k(t)$ whose sum dominates $\lambda(t)$. The auxiliary $f_1(t), \ldots, f_k(t)$ are chosen such that the inverse method is applicable in closed form for each. Additionally, we decompose $\lambda(t) = \lambda_1(t) + \cdots + \lambda_k(t)$, each part dominated by the corresponding $f_i(t)$. A thinning step is applied that accepts the time with $\lambda_i(t)/f_i(t)$. With this method, we combine the advantages of both the thinning method [158] and the Anderson MNRM [157]. Namely, on the one hand, the Anderson method is only strong, if the integrated intensity can be inverted, on the other hand, the thinning method is only strong if the acceptance rate is large enough.

To explain why the proposed method generates the correct law of sampled trajectories, we review how the following two operations on intensities induce operations on simulated trajectories: (i) addition and (ii) multiplication with a number between zero and one. Let us assume that $\{Y_1(t)\}_{t\in[0,T]}$ is a counting process with intensity $\lambda_1(t)$ and $\{Y_2(t)\}_{t\in[0,T]}$ is a counting process with intensity $\lambda_2(t)$ which is stochastically independent of $\{Y_1(t)\}_{t\in[0,T]}$. Then $Y(t) := Y_1(t) + Y_2(t)$ is a counting process with intensity $\lambda(t) = \lambda_1(t) + \lambda_2(t)$. In the point process picture we can visualize this. If the jumps of $Y_1(t)$ and $Y_2(t)$ were marked with different colors on a common real axis, then $Y(t)$ amounts to "forgetting" their color. If $\{Y_1(t)\}_{t\in[0,T]}$ is a counting process with intensity $\lambda_1(t)$ and for each of the jump times $(\tau_n)_{n\geq 0}$ we throw a coin, i.e. $U_n \sim \text{bernoulli}(p(\tau_n))$, then $Y^p(t) := \sum_{n=1}^{\infty} \mathbb{1}\{\tau_n \leq t, U_n = 1\}$ is a counting process with intensity $\lambda^p(t) = p(t)\lambda_1(t)$. Graphically we keep each point τ_n with a probability $p(\tau_n)$. The latter procedure, called thinning, is the reverse of the addition. Each point that is kept is marked with one color, each thinned point is marked with another color. Then $Y^p(t)$ and $Y^{1-p}(t) := Y_1(t) - Y^p(t)$ are stochastically independent counting processes with intensities $\lambda^p(t)$ and $\lambda^{1-p}(t) = (1 - p(t))\lambda_1(t)$. The decomposition $\lambda_1(t) = \lambda^p(t) + \lambda^{1-p}(t)$ indicates that "forgetting" the colors as before reverses the procedure. We call this decomposing operation (iii) assignment of marks.

The origin of why the three operations on intensities result in the correct law of the counting processes lies in the corresponding operations of exponentially distributed random variables. Let the distributions for two independent random waiting times T_1, T_2 (here, not to be confused with the internal times of algorithm 1) be given by their survival functions

$$\mathbb{P}[T_i > t] = \exp\left(-\int_0^t \lambda_i(s) \, ds\right).$$

Then

$$\begin{aligned}
\mathbb{P}[\min(T_1, T_2) > t] &= \mathbb{P}[T_1 > t, T_2 > t] \\
&= \mathbb{P}[T_1 > t]\mathbb{P}[T_2 > t] \\
&= \exp\left(-\int_0^t \lambda_1(s) + \lambda_2(s) \, ds\right).
\end{aligned}$$

Let $S_1, S_2, \ldots$ be random jump times from a point process with intensity $\lambda(t)$ and conditional on $S_1 = t_1, S_2 = t_2, \ldots$ let $U_1, U_2, \ldots$ be independent Bernoulli random variables with $p_i = \frac{\tilde{\lambda}(t_i)}{\lambda(t_i)}$ where $\tilde{\lambda}(t) \leq \lambda(t)$ for all $t \geq 0$. Let $I := \min(i : U_i = 1)$ and $\tilde{T} := S_I$, then for $\Delta_n(t) := \{\vec{t} \in [0,t]^n : t_1 \leq$

$\cdots \leq t_n\}$ we obtain

$$
\begin{aligned}
\mathbb{P}[\tilde{T} > t] &= \sum_{n=0}^{\infty} \int_{\Delta_n(t)} \mathbb{P}[\forall i = 1, \ldots, n : U_i = 0 | S_1 = t_1, \ldots, S_n = t_n, S_{n+1} > t] \\
&\quad \times p[S_1 = t_1, \ldots, S_n = t_n, S_{n+1} > t] \, \mathrm{d}\vec{t} \\
&= \sum_{n=0}^{\infty} \int_{\Delta_n(t)} \prod_{i=1}^{n} \left\{ \frac{\lambda(t_i) - \tilde{\lambda}(t_i)}{\lambda(t_i)} \cdot \lambda(t_i) \right\} \exp\left(-\int_0^t \lambda(s) \, \mathrm{d}s \right) \, \mathrm{d}\vec{t} \\
&= \exp\left(-\int_0^t \tilde{\lambda}(s) - \lambda(s) \, \mathrm{d}s \right) \exp\left(-\int_0^t \lambda(s) \, \mathrm{d}s \right) \\
&= \exp\left(-\int_0^t \tilde{\lambda}(s) \, \mathrm{d}s \right)
\end{aligned}
$$

Building on this, the idea of Sinzger's min-thin simulation algorithm is to combine two algorithms, the modified next reaction algorithm by Anderson and the thinning algorithm by Lewis, to harvest the advantages of both for an efficient simulation. To illustrate the conjunction of the algorithms by Lewis and Anderson assume exemplary that the intensity of a one-dimensional counting process is

$$
\lambda(t) = a_n + b_n \exp(-c_n t) \tag{161}
$$

between $(t(n), t(n+1))$, where $t(0) < t(1) < t(2) < \ldots$ are the jump times Then Anderson's modified next reaction algorithm requires to invert the cumulative intensity

$$
\Lambda(t) = \int_0^t \lambda(s + t(n)) \, \mathrm{d}s = t a_n + b_n / c_n (1 - \exp(-c_n t)).
$$

However, this function is not analytically invertible, which slows down the algorithm. On the contrary, if we take the cumulative intensities of the intensities

$$
\lambda_1(t) = a_n, \quad \lambda_2(t) = b_n \exp(-c_n t)
$$

separately, we obtain invertible cumulative intensities for both. This corresponds to simulating two counting processes $(Y_1(t))_{t \geq 0}$ and $(Y_2(t))_{t \geq 0}$ with intensities $\lambda_1(t)$ and $\lambda_2(t)$, respectively and then adding both $Y(t) := Y_1(t) + Y_2(t)$.

Similarly, if we have

$$
\lambda(t) = \alpha_n(t) + b_n \exp(-c_n t)
$$

for which $\int_0^t \alpha_n(s + t(n)) \, \mathrm{d}s$ is not analytically invertible, but we know $0 \leq \alpha_n(t) \leq a_n$. Then we can simulate the two counting processes $(Y_1(t))_{t \geq 0}$ and $(Y_2(t))_{t \geq 0}$ as above and we thin the jumps of counting process $Y_1(t)$ with probability $\alpha_n(t)/a_n$ to yield the counting process $\tilde{Y}_1(t)$. Then $Y(t) := \tilde{Y}_1(t) + Y_2(t)$

This motivates the following general decomposition. Suppose that

$$
\lambda(t) = \sum_{i=1}^{N} \lambda_i(t) \tag{162}
$$

for non-negative $\lambda_i(t)$, such that there exist $f_i^{(n)}(t) \geq \lambda_i(t), t \in [t(n), t(n+1)]$ for which

$$
\Lambda_{i,n}(t) := \int_0^t f_i^{(n)}(s + t(n)) \, \mathrm{d}s \tag{163}
$$

is analytically invertible in t. Let $\tilde{Y}_i(t), i = 1, \ldots N$ be counting processes with intensity $f_i^{(n)}(t)$, and $Y_i(t)$ the counting processes obtained from $\tilde{Y}_i(t)$ by thinning with $\lambda_i(t)/f_i^{(n)}(t)$. Then

$$
Y(t) = \sum_{i=1}^{N} Y_i(t) \tag{164}
$$

89

Algorithm 2 Sinzger's min-thin algorithm

1: **Input**: Intensities $\lambda_i(t), i = 1, \ldots, N$, dominating intensities $f_i^{(n)}(t)$, length T of the simulated time interval.

2: Initialize internal times $T_i = 0$, jump times $P_i \sim \text{expo}(1)$ and $\Delta T_i = P_i - T_i, i = 1, \ldots, N$.

3: Initialize global time $t_0 = 0$ and event counter $n = 0$.

4: Save $t(0) \leftarrow t_0$.

5: Evaluate $t_i \leftarrow \Lambda_{i,n}^{-1}(\Delta T_i), i = 1, \ldots, N$.

6: **while** $t_0 < T$ **do**

7:     Assign $I = \arg\min_i\{t_i\}$. Assign $\tau \leftarrow t_I$. Update $t_0 \leftarrow t_0 + \tau$. Update $t_i \leftarrow t_i - \tau, i \neq I$.

8:     Sample $u \sim \text{unif}[0,1]$.

9:     **if** $u < \lambda_I(t_0)/f_I^{(n)}(t_0)$ **then**

10:         For $i \neq I$: Update internal times $T_i \leftarrow T_i + \Lambda_{i,n}(t_0 - t(n))$, update $\Delta T_i \leftarrow P_i - T_i$, evaluate $t_i \leftarrow \Lambda_{i,n}^{-1}(\Delta T_i)$.

11:         Update $n \leftarrow n + 1$. Save $t(n) \leftarrow t_0$.

12:     **end if**

13:     Update internal time $T_I \leftarrow P_I$. Draw $\Delta P \sim \text{expo}(1)$. Update jump time $P_I \leftarrow P_I + \Delta P$. Update $\Delta T_I \leftarrow P_I - T_I$. Update $t_I \leftarrow \Lambda_{I,n}^{-1}(\Delta T_I)$.

14: **end while**

15: **return** Jump times $t(0), \ldots, t(n)$.

---

has the desired law. This translates into the algorithm, where $\Lambda_{i,n}^{-1}(\cdot)$ denotes the inverse of $t \mapsto \Lambda_{i,n}(t)$ in lines 5, 10 and 13. If the minimum of the proposed times is thinned, i.e., if the condition in line 9 is false, a new time must be proposed line 13 to compete with the remaining proposed times in line 7.

The challenge of the decomposition Eq. (162) lies in the condition that the $\lambda_i(t)$ must be non-negative, which becomes apparent when b_n in Eq. (161) is negative. The case that for one component there are almost surely only finitely many jumps on $(0, \infty)$ is included in the procedure and does not cause a conflict, since the component then suggests ∞ as a jump time and is not chosen. This is the case for the above example $Y_2(t)$ with intensity $\lambda_2(t) = b_n \exp(-c_n t)$.

Our presented algorithm 2 portrays the case of a univariate counting process $Y(t)$. It can be combined with Anderson's modified next reaction method to obtain the multivariate case $(Y_1(t), \ldots, Y_M(t))$, for the reaction channels $j = 1, \ldots, M$. Each single channel $Y_j(t)$ is decomposed as in Eq. (164), i.e., $Y_j(t) = \sum_{i=1}^{N(j)} Y_{j,i}(t)$, where the number of components $N(j)$ can depend on j. Correspondingly, in the algorithm 2, we use double indices $T_{i,j}, i = 1, \ldots, N(j), j = 1, \ldots, M$.

We now apply Sinzger's min-thin trick to obtain a fast simulation algorithm for a chemical reaction network with independent environment components. The above difficulty of a non-negative decomposition results in a distinction of various cases for the intensity $\hat{\lambda}^{\text{OL}}(t) = \mu + c\hat{Z}(t)$. For convenience we suppress the subscript j. Define $\bar{\beta} := \lim_{t \to \infty} \beta(t) = \frac{\mu}{c^2}(\rho - \alpha), \bar{\lambda} = \lim_{t \to \infty} \hat{\lambda}^{\text{OL}}(t) = \frac{\mu\alpha}{\rho}$ and $\lambda_0 := \hat{\lambda}^{\text{OL}}(t_0)$. First of all we consider the case

- $\underline{\beta(t_0) = \bar{\beta}}$. Then $\hat{\lambda}^{\text{OL}}(t) = \bar{\lambda} + (\lambda_0 - \bar{\lambda})e^{-\rho(t-t_0)}$.

  - $\underline{\lambda_0 \geq \bar{\lambda}}$ is the case considered above and we choose

$$f_1(t) = \lambda_1(t) = \bar{\lambda} \tag{165}$$

$$f_2(t) = \lambda_2(t) = (\lambda_0 - \bar{\lambda})e^{-\rho(t-t_0)}, \tag{166}$$

  whereas for

  - $\underline{\lambda_0 < \bar{\lambda}}$

$$\lambda_1(t) = (\bar{\lambda} - \lambda_0)(1 - e^{-\rho(t-t_0)}) \tag{167}$$

$$f_1(t) = (\bar{\lambda} - \lambda_0) \cdot \min(1, \rho(t - t_0)) \tag{168}$$

$$f_2(t) = \lambda_2(t) = \lambda_0, \tag{169}$$

The next case we consider is

- $\underline{\beta(t_0) < \bar{\beta}}$. Then

$$\hat{\lambda}^{\mathrm{OL}}(t) = \bar{\lambda}\frac{C - e^{-2\rho(t-t_0)}}{C + e^{-2\rho(t-t_0)}} + \frac{Be^{-\rho(t-t_0)}}{C + e^{-2\rho(t-t_0)}} \tag{170}$$

$$C = \frac{\beta(t_0) + \bar{\beta} + \frac{2\alpha\mu}{c^2}}{\bar{\beta} - \beta(t_0)} \tag{171}$$

$$B = 2\left(\bar{\lambda} - \frac{(\lambda_0 - \bar{\lambda})\rho\mu}{c^2(\beta(t_0) - \bar{\beta})}\right) \tag{172}$$

The first summand in Eq. (170) is positive, because $\rho > 0$ and $C > 1$. For the second summand, it depends on the sign of B from which we obtain the case

- $\frac{\lambda_0 - \bar{\lambda}}{\beta(t_0) - \bar{\beta}} \leq \frac{c^2\alpha}{\rho^2}$. In this case $B \geq 0$ and we choose

$$f_1(t) = \lambda_1(t) = \bar{\lambda} \cdot \frac{C - e^{-2\rho(t-t_0)}}{C + e^{-2\rho(t-t_0)}} \tag{173}$$

$$f_2(t) = \lambda_2(t) = \frac{Be^{-\rho(t-t_0)}}{C + e^{-2\rho(t-t_0)}}, \tag{174}$$

- $\frac{\lambda_0 - \bar{\lambda}}{\beta(t_0) - \bar{\beta}} > \frac{c^2\alpha}{\rho^2}$. In this case $B < 0$ and we choose

$$\lambda_1(t) = \hat{\lambda}^{\mathrm{OL}}(t) \tag{175}$$

$$f_1(t) = \bar{\lambda} \cdot \frac{C - e^{-2\rho(t-t_0)}}{C + e^{-2\rho(t-t_0)}} \tag{176}$$

$$f_2(t) = \lambda_2(t) = 0. \tag{177}$$

Finally, we consider the case

- $\underline{\beta(t_0) \geq \bar{\beta}}$. Then again Eq. (170) holds and the first summand is positive because $C < -1$ and $\rho > 0$. The denominator of the second term is negative. Again the sign of the second term depends on the sign of B. We distinguish again the cases

- $\frac{\lambda_0 - \bar{\lambda}}{\beta(t_0) - \bar{\beta}} \leq \frac{c^2\alpha}{\rho^2}$. In this case $B \geq 0$ which makes the second term negative. Then we choose Eq. (175) - Eq. (177) as above, whereas for

- $\frac{\lambda_0 - \bar{\lambda}}{\beta(t_0) - \bar{\beta}} > \frac{c^2\alpha}{\rho^2}$ we have we use Eq. (173) - Eq. (174). For later reference we write again

$$f_2(t) = \lambda_2(t) = \frac{Be^{-\rho(t-t_0)}}{C + e^{-2\rho(t-t_0)}}, \quad B < 0, C < -1, \tag{178}$$

Note that the dynamics of $\hat{\lambda}^{\mathrm{OL}}(t)$ is dominated as follows

$$\dot{\hat{\lambda}}^{\mathrm{OL}}(t) = -\rho(\hat{\lambda}^{\mathrm{OL}}(t) - \bar{\lambda}) - (\beta(t) - \bar{\beta})\hat{\lambda}^{\mathrm{OL}}(t) \geq -\rho(\hat{\lambda}^{\mathrm{OL}}(t) - \bar{\lambda}),$$

because $\beta(t) > \bar{\beta}$ for all t and $\hat{\lambda}^{\mathrm{OL}}(t) > 0$. Consequently, $\hat{\lambda}^{\mathrm{OL}}(t) \leq \bar{\lambda} + (\lambda_0 - \bar{\lambda})e^{-\rho(t-t_0)}$ Then we distinguish cases by the sign of $\lambda_0 - \bar{\lambda}$.

* $\underline{\lambda_0 - \bar{\lambda} > 0}$

The cumulative intensities and their inverses are for Eq. (165)

$$\Lambda(t) = \bar{\lambda}t \tag{179}$$

$$\Lambda^{-1}(T) = \frac{T}{\bar{\lambda}} \tag{180}$$

and analogously for Eq. (169), for Eq. (166)

$$\Lambda(t) = \frac{\lambda_0 - \bar{\lambda}}{\rho}(1 - e^{-\rho t}) \tag{181}$$

$$\Lambda^{-1}(T) = -\rho^{-1}\ln\left(1 - \frac{T\rho}{\lambda_0 - \bar{\lambda}}\right), 0 \leq T < \frac{\lambda_0 - \bar{\lambda}}{\rho} \tag{182}$$

for Eq. (168)

$$\Lambda(t) = \begin{cases} (\bar{\lambda} - \lambda_0)\min\left(\frac{1}{2}\rho t^2\right), & t \leq \rho^{-1} \\ (\bar{\lambda} - \lambda_0)\min\left(t - \frac{1}{2\rho}\right), & \text{else} \end{cases} \tag{183}$$

$$\Lambda^{-1}(T) = \begin{cases} \sqrt{\frac{2T}{(\bar{\lambda} - \lambda_0)\rho}}, & T \leq \frac{\bar{\lambda} - \lambda_0}{2\rho} \\ \frac{T}{\bar{\lambda} - \lambda_0} + \frac{1}{2\rho}, & \text{else} \end{cases}, \tag{184}$$

as well as

$$\Lambda(t) = \bar{\lambda}t + \frac{\bar{\lambda}}{\rho}\ln\left(\frac{2\rho + \frac{c^2}{\mu}(\beta(t_0) - \bar{\beta})(1 - e^{-2\rho t})}{2\rho}\right) \tag{185}$$

$$\Lambda^{-1}(T) = \frac{T}{\bar{\lambda}} - \rho^{-1}\ln(2) + \rho^{-1}\ln\left(1 + C^{-1} + \sqrt{(1 + C^{-1})^2 - 4C^{-1}e^{-2T\rho\bar{\lambda}^{-1}}}\right) \tag{186}$$

for Eq. (173) and Eq. (176). Finally, for Eq. (174), we obtain

$$\Lambda(t) = \frac{B}{\sqrt{C}\rho}\left(\arctan(\sqrt{C}e^{\rho t}) - \arctan(\sqrt{C})\right) \tag{187}$$

$$\Lambda^{-1}(T) = \rho^{-1}\ln\left(\tan\left(\frac{T\sqrt{C}\rho}{B} + \arctan(\sqrt{C})\right)\sqrt{C^{-1}}\right), 0 \leq T < (\frac{\pi}{2} - \arctan(\sqrt{C}))\frac{B}{\sqrt{C}\rho} \tag{188}$$

whereas Eq. (178) yields

$$\Lambda(t) = \frac{-B}{\sqrt{-C}\rho}\left(\text{artanh}(\sqrt{-C^{-1}}) - \text{artanh}(\sqrt{-C^{-1}}e^{-\rho t})\right) \tag{189}$$

$$\Lambda^{-1}(T) = -\rho^{-1}\ln\left(\tanh\left(\frac{T\sqrt{-C}\rho}{B} + \text{artanh}(\sqrt{-C^{-1}})\right)\sqrt{-C}\right), 0 \leq T < \text{artanh}(\sqrt{-C^{-1}}) \cdot \frac{-B}{\sqrt{-C}\rho} \tag{190}$$

Here, we have portrayed Sinzger's min-thin algorithm 2 for our tilted Hawkes model. However, the algorithm can be more generally applied for multivariate counting processes with piecewise-deterministic intensities.

8 Case studies

We illustrate the derived methods and results in a variety of case studies. The different environments of section 6 are employed. We derive probability evolution equations using the stochastic conditioning. A focus of the case studies is on the comparison of approximate filters. The stationary mean is evaluated for the birth-death process in different environments that modulate the death rate and the contribution of environmental shares is analyzed. The examples cover zeroth- and first-order modulation. Preparing for the channel interpretation of zeroth-order modulation, the table 8 summarizes the models that were analyzed with the sufficient variables of joint Markovian progression, see section 4.1.4.

8.1 Modular embedding of structure conversion

At all three stages of gene expression the allosteric principle postulates that regulation occurs via the switching between different states or conformations. The promoter region of the DNA strand can be

Model (*Type)	Θ_t	n_0	dim	l	u	F	f
Random telegraph	$\Pi_t(1)$	1	0	$c\pi_1$	✓	Eq. (275)	1
Dark current	$\hat{\lambda}_t$	1	1	id	✓	Eq. (220)	$\frac{(\hat{\lambda}-\lambda_0)(\lambda_1-\hat{\lambda})}{\hat{\lambda}} + \hat{\lambda}$
Single On N Off	$(\Pi_t(z_1),\ldots,\Pi_t(z_{|\mathcal{Z}|-1}))$	$|\mathcal{Z}|-1$	0	$c\pi_1$	-	Eq. (6)	$(1,0,\ldots,0)$
Double On - N Off	$(U_t,A_t,\Pi_t(z_3),\ldots,\Pi_t(z_{|\mathcal{Z}|-1}))$	$|\mathcal{Z}|-1$	1	ca	-	Ito from Eq. (6)	$(u,1,0,\ldots,0)$
- Single Off	(U_t,A_t)	2	1	ca	-	(287)-(288)	$(u,1)$
Hawkes*	$\hat{\lambda}_t$	1	1	id	✓	Eq. (58)	$\hat{\lambda}+\beta$
Hawkes conversion*	$(z_1(t),\ldots,z_k(t))$	k	$k-1$	$\mu+Cz$	✓	Eq. (145)	$z+BC^T\mu^{-1}$
Multinomial*	$(\theta_1(t),\ldots,\theta_{k-1}(t))$	$k-1$	$k-1$	Eq. (202)	-	Eq. (200)	Eq. (201)
Binomial*	$\theta(t)$	1	1	Eq. (203)	✓	Eq. (200)	Eq. (201)
Gamma filter*	(M_t,S_t)	2	2	cm	-	Eq. (222)	$(m,s)+(\frac{s}{m},\frac{s^2}{m^2})$

Table 1: Snyder filters and approximate filters for zeroth-order modulated reactions. The type indicates whether the model is a Snyder filter (Markov-modulated Poisson process, no *) or an approximate filter (self-exciting counting process, *). The minus $(-)$ in the column u indicates that no analytic solution is available for $\tau \mapsto u(\tau, v^0)$. For the Hawkes conversion process, see also the Appendix B

accessible for transcription or shielded. The mRNA can fold in different secondary structures and transition among these via metastable switching. Proteins undergo conformational changes to fulfill their function as enzymes, or to regulate as transcription factors via activation or repression. Both the composition of the thermodynamic ensemble, i.e., the stationary distribution of structures, and the switching dynamics between those, play a role. Structure kinetics and gene expression is often studied separately. Here, we present approaches that allow quantitative information on structure switching from biophysical experiments or coarse-grain molecular dynamics simulations to be included in gene regulatory chemical reaction network models without an increase in the number of species. Addressing the problems of simulating gene regulatory networks and the modular design of synthetic circuits, we provide an efficient way of including heterogeneity due to structure switching. For the latter purpose, our method may help with *in silico* studies, i.e., simulations, that precede *in vitro* studies. The ultimate goal is the incorporation of the approaches in larger genetic circuits and gene regulatory networks.

8.1.1 Promoter-mediated transcription

The transcription of mRNA is a process that is shaped by the cellular context. The number of polymerases, fluctuating transcription factor concentrations, noncognate binding and chromatin remodeling are some of the factors that determine the rate of transcription. Often these influences are abstracted into a low number of discrete promoter states that determine the transcription rate. The effect of switching between an active and an inactive promoter state is often equated with transcriptional bursting. We continue the example of section 6.1 where $X(t)$ denoted the total number of mRNA from multiple transcription sites. Assume further that mRNA decays with a rate γ. Then $X(t)$ is a birth-death process whose birth rate is modulated by a conversion process $Z(t) = [Z_1(t), \ldots, Z_k(t)]$ as described in section 6.1. The switching rates are $c_{i,j}$ and the transcription rates were denoted by r_i.

We first considered a promoter model which unidirectionally transitions through an active, an inactive and a refractory state, see figure 7A. We chose equal transition rates, hence the promoter uniformly spends time in either state. Assuming leaky transcription, we set the transcription rate for the inactive state to a small leakage value, the rate for the active state to the highest and for the refractory state to an intermediate value. Furthermore, we considered four independent copy of the promoter. We numerically verified the condition (D1) for the considered choice of parameters. We looked at the stationary distribution of the mRNA, figure 7B to compare the exact system with the equilibrated optimal linear (Hawkes) model, Eq. (68). From theorem 4.4 we know that the variances coincide. However, also the shape is essentially captured. When looking at the true and approximated state estimate (fig 7C) for a mRNA trajectory (fig 7D), both also agree to a large extent. Even though the equilibrated optimal linear (Hawkes) model does not capture the base level, the state estimate seems to allocate the regime where the linearization fits well for a large amount of time.
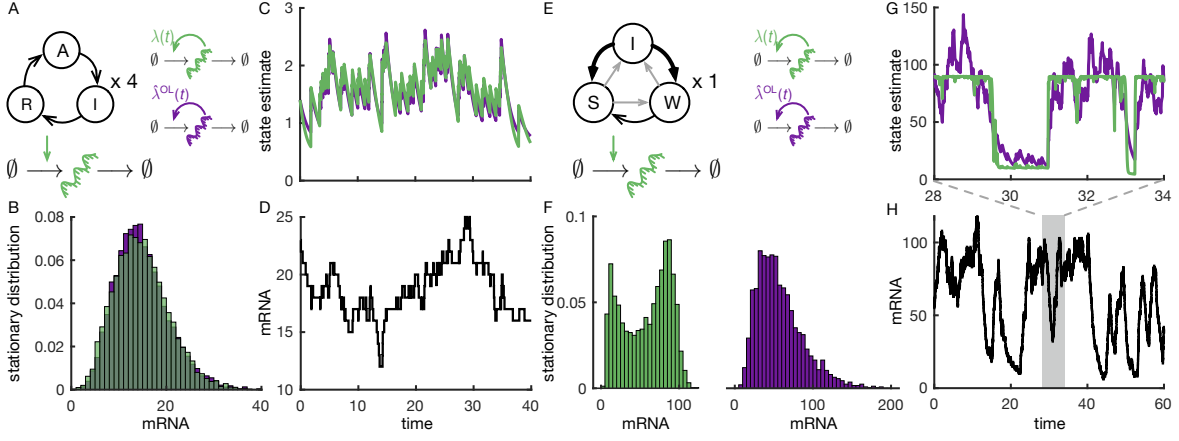
Figure 7: Promoter-mediated transcription. A. Transcription model with promoter states modeled by a discrete Markov process. The promoter model transitions through an active, an inactive and a refractory state with all three transition rates being equal to 0.2. Each state summarizes at what rate transcription is initiated. Assuming a leakage of 0.01, we chose exemplarily for the transcription rates 1.01 (A), 0.01 (I), 0.06 (R). Furthermore, we considered four independent copies corresponding to different transcription sites in the cell. We compared the true system (green) and the equilibrated optimal linear (Hawkes) approximation (purple). B. Histogram of the stationary distribution for the exact simulations, using Gillespie (green) and the approximate marginal simulations, using Anderson's modified next reaction method (purple). C. State estimate of the effective (posterior mean) transcription rate obtained as an average over the posterior probabilities of being in either promoter state. Green uses the true posterior probabilities, purple the equilibrated optimal linear (Hawkes) model, Eq. (68). D. The mRNA trajectory for which the state estimate was computed. It was simulated, using the Gillespie algorithm. E. A different promoter model is used compared to A with an inactive, a weak and a strong state. The arrow thickness indicates the transition strength. Transition and transcription rates were $c_{WI} = c_{SW} = c_{SI} = 0.2, c_{WS} = 0.4, c_{IS} = c_{IW} = 2, \lambda_S = 90, \lambda_W = 10, \lambda_I = 0$ (borrowed from [24]). We considered only a single transcription site. F. Histogram as in B. The exact bimodal distribution is not captured by the equilibrated optimal linear (Hawkes) model, even though the variances agree. G. State estimates as in C for the time interval $[28, 34]$. H. Sample trajectory corresponding to G, simulated using Gillespie. The degradation rate of mRNA was chosen 0.1 in model A and E.
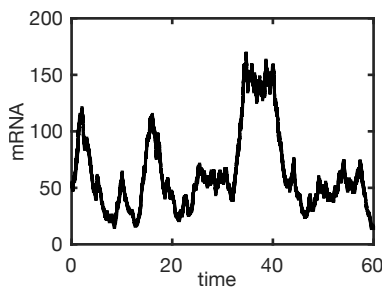
Figure 8: A sample trajectory of the mRNA counts for the approximate marginal simulation, i.e., the analog to fig 7H
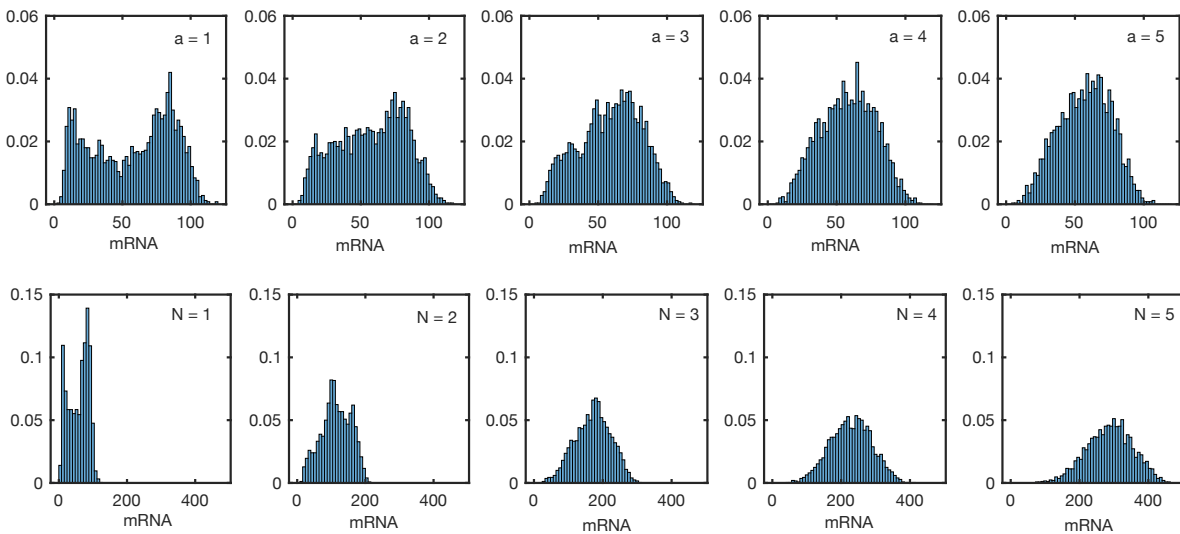
Figure 9: **Details on promoter-mediated transcription.** Histograms of the stationary mRNA, simulated from the exact system using the Doob-Gillespie algorithm. Final time point was $t = 60$. For the upper row, the transition rates were multiplied by $a = 1, \ldots, 5$. For the lower row N was increased from 1 to 5. Otherwise, the parameters were as in 7E. The number of trajectories was 2500.

We now turn towards a promoter model, which has more extreme properties than the previous one. The new model allows transitions in all directions, where two states, a strong and a weak one, are dominantly allocated, see figure 7E. The major changes are: (i) an increase of the transcription rates compared to the transition rates, (ii) only one transcription site as opposed to four copies. Again, we numerically verified the condition (D1). Then we see that the true bimodal distribution is not captured by the equilibrated optimal linear (Hawkes) model, even though the variance is captured correctly (fig 7F). Looking at the exact and the approximate state estimates (fig 7G and H) we see that the very sharp transitions are not well captured by the equilibrated optimal linear (Hawkes) model. These non-linearities are out of reach for the linearization. The sample trajectory (fig 7H) was simulated with the exact system (Doob-Gillespie). An mRNA trajectory simulated by the approximate marginal simulation is shown in the figure 8. It differs from the exact system in underestimating the time spent in the regime of 80 to 100 mRNA copies, compare also the histograms in figure 7F.

We hypothesized that both a relaxation of (i) and (ii) could lead to a better agreement with the linearization. As a necessary condition, the bimodal distribution must transform to a unimodal one. To relax (i) we multiplied the transition rates by a factor $a > 1$, and to relax (ii) we varied N. In figure 9 we depict the corresponding histograms for $a = 1, \ldots, 5$ and $N = 1, \ldots, 5$, which provides evidence that the necessary condition is met.

If in the case of relaxing (i) the timescale of the promoter approaches the timescale of the transcription then the knowledge about the state is more vague, hence the domain of state estimate values that are allocated narrows. On this domain, the linearization can have a good accuracy for a larger propor-

95

tion. The extreme case would be a much faster promoter timescale than the transcription timescale, for which the state estimate distribution would narrow to a peak at the mean value. If in the case of relaxing (ii) we increase the number of copies, the prior distribution of intensity values becomes unimodal, in the sense that the distribution of $\frac{1}{\sqrt{N}}(\sum_{i=1}^{N} \lambda(Z_i) - N\mathbb{E}[\lambda(Z_1)]$ approaches a Gaussian for the i.i.d. promoter copies Z_i distributed as the stationary prior distribution of one promoter. Let us assume the extreme case of perfect state estimation, with the distribution of state estimate values equal to the prior distribution of intensity values. This distribution would resemble more the distribution of intensity value a Hawkes model can achieve, see section 8.2 below. With imperfect state estimation, the linearization is even improved by the above argument. In summary, we conjecture that both with a lower frequency of point observations and with a more Gaussian shape of the prior distribution of intensity values, the accuracy of the equilibrated optimal linear (Hawkes) model can increase. We continue this case study in sections 8.2.2 and 8.3.6

8.1.2 Conformational mRNA switching dynamics mediates translation

With the conversion process environment, not only the promoter switching can be captured, but also the conformational changes of macromolecules, such as mRNA or enzymes. A growing body of experimental work indicates a negative correlation between translation efficiency and mRNA secondary structure [160]. Structure switching thus contributes to translational heterogeneity [161, 4], which is not covered in the classical approach via CRN models. While the use of statistical physics has illuminated how the ensemble of mRNA secondary structures is composed, it is still an open question in what way and how much the dynamics between these structures contributes to translational heterogeneity [4]. A recent study hints at the importance of the unfolding kinetics near the ribosome binding site [162]. Although the unfolding and refolding of mRNA can happen on its own, it is more commonly mediated by ligands, small RNAs or proteins [163, 161, 164]. Riboswitches are included in this class of translational regulators [165].

The dynamics of RNA structure changes occurs at different timescales [166, 167] (figure 10E). At the fastest timescale of about femto- to nanoseconds, the structure changes as a result of intramolecular thermal motion [161], e.g., exhibiting base wobbling or rotational diffusion. These fast changes of the states, summarized by all atom positions, are called equilibrium fluctuations. A conformation, or metastable state, of a macromolecule, such as mRNA, denotes a subset of atom configurations that preserves large-scale geometric properties [168]. At a medium timescale, the switching is due to conformational changes between metastable states. The dynamics remains in the metastable state for a medium timescale with a probability of almost one until entering another metastable state [169]. Due to energy barriers, these state transitions of functional importance happen only rarely [161]. Metastable states are often linked to energy minima of the free energy landscape of mRNA folding [167]. Note the role of co-transcriptional folding [170] that determines in which suboptimal local minimum the structure might be trapped. At the slowest timescale of seconds or minutes, in contrast to equilibrium fluctuations such as base wobbling, we consider switching due to changes in the energy landscape that are induced by cellular cues, e.g., a change in ligand concentrations [171]. In particular, such altered energy landscape comes with changes in the canonical ensemble, the conformations and the conformational dynamics, i.e., the transition rates between metastable states [172]. While the state at the fast and medium timescales is the physical position of atoms and the abstract metastable state, the state at the slowest timescale is the energy landscape itself, i.e., the full RNA ensemble and transition dynamics. A transition between such states is also called redistribution [167]. Note that Dethoff et al. use the term conformational transition exclusively in the sense of altering the energy landscape. Assuming that there is a dominant conformation for each altered energy landscape, a discrete state Markov model might still be sufficiently accurate at the slow timescale. The dynamics of the cellular effector then determines the transition between dominant mRNA structures.

We interpret Eq. (136) as mRNA structure switching and (140) as translation reaction. Different mRNA conformations are associated with different translation rates. Denote the translation model consisting of Eqs. (140) and (136) as $\mathcal{M}_{\mathrm{S}}$, where S stands for structure.
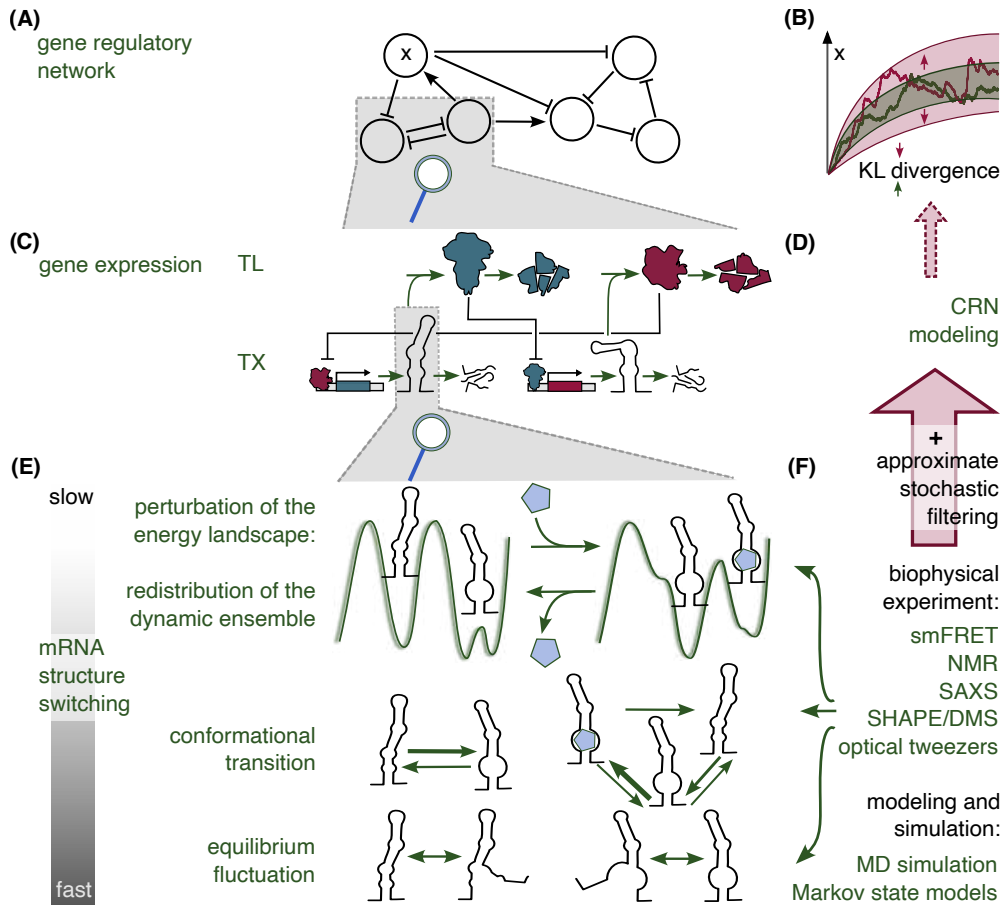
Figure 10: The inclusion of mRNA structure switching in the CRN modeling of gene expression in gene regulatory networks. (A) A representative of a gene regulatory network. Nodes are genes, arrows indicate activation and repression. (B) Statistical properties of protein trajectories, such as the variance, the Fano factor. The figure shows trajectories without (green) and with (red) sources of heterogeneity and indicates how the standard deviation increases. The Fano factor increase or the KL divergence between trajectories can indicate the contribution of the heterogeneity. (C) The gene regulatory motifs (activation and repression) are modeled in a classical two-stage gene expression model, separating transcription (TX) and translation (TL). (D) Classical CRN models are extended to reflect translational heterogeneity. Approximate stochastic filtering links biophysical knowledge about the reaction rate constants to CRN models. (E) The translation reaction is refined to allow for mRNA conformations with different translation rates. The dynamics of mRNA conformational switching occurs on different timescales. At the slowest timescale, the conformation or energy landscape is altered by cellular effectors. Exemplarily, we depicted changed energy landscapes for a low (left) and high (right) ligand concentration. The dynamics of the cellular effector that modulates the conformational landscape dictates the timescale. At an intermediate timescale, transitions occur between conformations, i.e., sets of atom configurations that preserve large-scale geometric shapes. At the fastest timescale, equilibrium fluctuations due to intramolecular thermal motion occur. When accumulated, those lead to the more rarely occurring conformational transitions. (F) The rates of transition and the translation rates are deduced from biophysical experiments and (MD) algorithms, such as oxRNA, in combination with hidden Markov state models.

8.1.3   Contribution of mRNA structure switching to translational noise

Experimental evidence hints at the contribution of mRNA structure to the translational noise. In order to demonstrate this effect in the modeling context, we consider the standard transcription-translation

model and enhance it with a two-structure model. The transcription-translation base model

$$\mathcal{M}_{\text{TX}} : \begin{array}{ll} R_1, R_2 & : \emptyset \xrightarrow{\mu} Z \xrightarrow{\delta} \emptyset \\ R_3 & : Z \xrightarrow{r} Z + \text{Prot} \end{array}$$
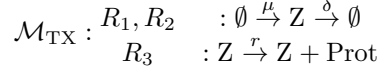
is extended by accounting for structures to

$$\mathcal{M}_{\text{TX,S}} : \begin{array}{ll} R_{1,2} & : Z_1 \xrightarrow{c_{1,2}} Z_2 \\ R_{2,1} & : Z_2 \xrightarrow{c_{2,1}} Z_1 \\ R_1 & : Z_1 \xrightarrow{r_1} Z_1 + \text{Prot} \\ R_2 & : Z_2 \xrightarrow{r_2} Z_2 + \text{Prot} \\ R_3, R_4 & : \emptyset \xrightarrow{\alpha \cdot c_{2,1}} Z_1 \xrightarrow{\delta} \emptyset \\ R_5, R_6 & : \emptyset \xrightarrow{\alpha \cdot c_{1,2}} Z_2 \xrightarrow{\delta} \emptyset. \end{array}$$

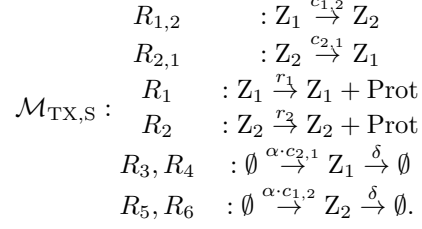When the structure Z_i is synthesized, the probability to be in Z_1 is chosen as $f := c_{2,1}/(c_{1,2} + c_{2,1})$ to match the stationary ratio of structures due to switching between structures. This explains the choice of the birth rates $\alpha \cdot c_{2,1}$ and $\alpha \cdot c_{1,2}$. The models are matched in that (i) the constant transcription rate of the model $\mathcal{M}_{\text{TX}}$ equals $\mu = \alpha(c_{1,2} + c_{2,1})$, the total transcription rate of $\mathcal{M}_{\text{TX,S}}$, and (ii) its constant translation rate equals the mean translation rate of $\mathcal{M}_{\text{TX,S}}$, i.e., $r = fr_1 + (1-f)r_2$. Without loss of generality we assume the structure Z_1 to possess the higher translation rate, i.e., $r_1 > r_2$, and introduce the fold change $\rho := r_1/r_2$.

Both models are instances of Markov-modulated Poisson processes where the modulator is a linear CRN. How to choose the parameters of the models $\mathcal{M}_{\text{TX}}, \mathcal{M}_{\text{TX,S}}, \mathcal{M}_{\text{S}}$ in Eq. (33) is demonstrated below. For example, $\tau_{\text{TX}} = \delta^{-1}$ and $\tau_{\text{S}} = (c_{1,2} + c_{2,1})^{-1}$ are the timescales of transcription $R_3 - R_6$ and structure switching $R_{1,2}, R_{2,1}$ in $\mathcal{M}_{\text{TX,S}}$, respectively.
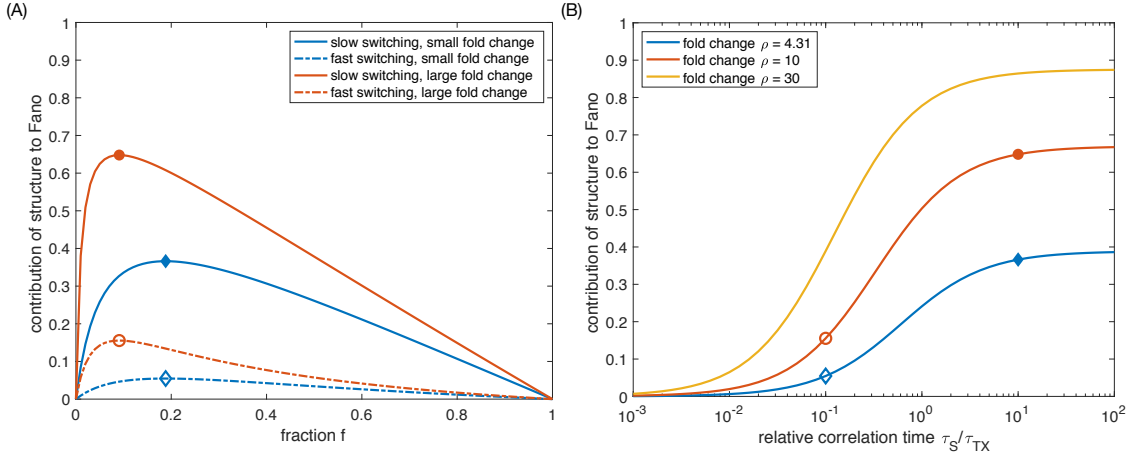
Figure 11: Translational noise. Contribution of the structure compared to the transcript copy number as in the expression Eq. (197). (A) Noise share as a function of the fraction of structure Z_1. Local maxima occur at $f = (\rho + 1)^{-1}$, i.e., when $(1 - f)/f = \rho$. Different curves show different relative correlation times $\tau_{\text{S}}/\tau_{\text{TX}}$ (solid is high, dashed is low) and different fold changes ρ (blue is low, orange is high). (B) The dependence of the noise share on the relative correlation time for different fold changes ρ. The fraction f was set to $(\rho+1)^{-1}$ to yield the maximal value. Curves saturate at $(\rho-1)^2/(\rho+1)^2$. Diamonds and circles belong to corresponding parameter choices in the subfigures.

Let $Y(t)$ count the number of translation synthesis events. We assess the total translational noise by means of the asymptotic Fano factor $F_{\text{TX,S}}$ of the protein synthesis, i.e. $\lim_{t \to \infty} \mathbb{V}\text{ar}[Y(t)]/\mathbb{E}[Y(t)]$. In order to disentangle the different noise sources we decompose the total noise into contributions from translational intrinsic noise (Poissonian), translational noise due to transcription and due to switching between structures

$$F_{\text{TX,S}} = 1 + \Delta F_{\text{TX}} + \Delta F_{\text{S|TX}}. \tag{191}$$

We computed $F_{TX,S}$ and F_{TX} as in Eqs. (32) and (33), section 3.3. Then $\Delta F_{TX} = F_{TX} - 1$ yielded $\Delta F_{S|TX}$.

For the model $\mathcal{M}_{TX}$ the first- and second order statistics of the modulating birth-death process N are

$$\mathbb{E}[N] = \frac{\mu}{\delta}, \mathbb{C}\text{ov}[N_{t+h}, N_t] = \frac{\mu}{\delta} e^{-\delta h}.$$

With $C = r$ we obtain for the asymptotic variance slope

$$\dot{\mathbb{V}}\text{ar} = \lim_{t \to \infty} t^{-1} \mathbb{V}\text{ar}[Y_t] = r \cdot \frac{\mu}{\delta} + \frac{2r^2 \frac{\mu}{\delta}}{\delta}$$

and consequently, for the Fano factor:

$$F_{TX} = 1 + 2 \cdot \frac{r}{\delta} = 1 + 2 \cdot \frac{f(r_1 - r_2) + r_2}{\tau_{TX}^{-1}}$$

$$= 1 + 2 \cdot \tau_{TX} \cdot r_2 (f(\frac{r_1}{r_2} - 1) + 1)$$

$$= 1 + 2\frac{\tau_{TX}}{\tau_{TL}}(f(\rho - 1) + 1).$$

For the model $\mathcal{M}_{TX,S}$, the modulator $X = [Z_1, Z_2]$ induces as parameters the matrix

$$A = \begin{bmatrix} c_{1,2} + \delta & -c_{2,1} \\ -c_{1,2} & c_{2,1} + \delta \end{bmatrix},$$

the mean

$$\mathbb{E}[Z_1] = \frac{f\mu}{\delta}, \mathbb{E}[Z_2] = \frac{(1-f)\mu}{\delta}$$

while the matrix entries of the stationary covariance matrix Σ must satisfy

$$0 = \alpha c_{2,1} + \frac{\alpha c_{1,2} c_{2,1}}{\delta} - (\delta + c_{1,2})\Sigma_{11} + c_{2,1}\Sigma_{12}$$

$$0 = \alpha c_{1,2} + \frac{\alpha c_{1,2} c_{2,1}}{\delta} - (\delta + c_{2,1})\Sigma_{22} + c_{1,2}\Sigma_{12}$$

$$0 = -(2\delta + c_{1,2} + c_{2,1})\Sigma_{12} - \frac{2\alpha c_{1,2} c_{2,1}}{\delta} + c_{1,2}\Sigma_{11} + c_{2,1}\Sigma_{22},$$

which is solved by $\Sigma_{11} = \frac{f\mu}{\delta}, \Sigma_{12} = 0, \Sigma_{22} = \frac{(1-f)\mu}{\delta}$. The stationary distribution can be shown to be a joint of independent Poisson random variables. But through the matrix A, temporal cross-covariances enter, such that the variance slope of Y_t reads

$$\dot{\mathbb{V}}\text{ar} = \frac{r\mu}{\delta} + [r_1, r_2] \left(\Sigma A^{-T} + A^{-1}\Sigma \right) [r_1, r_2]^T$$

$$= \frac{r\mu}{\delta} + \frac{2\mu}{\delta(c_{1,2} + c_{2,1} + \delta)}(r^2(c_{1,2} + c_{2,1}) + \delta(r_1^2 f + r_2^2(1 - f))).$$

With transformations analogous to the ones above

$$F_{TX,S} = 1 + \frac{2r}{\delta} + \frac{2(r_1 - r_2)^2 f(1 - f)}{(1 + \frac{\delta}{c_{1,2} + c_{2,1}})(c_{1,2} + c_{2,1})r} \tag{192}$$

$$= F_{TX} + \frac{2(\rho - 1)^2 f(1 - f)}{(\frac{\tau_{TL}}{\tau_S} + \frac{\tau_{TL}}{\tau_{TX}})(f(\rho - 1) + 1)}. \tag{193}$$

For the model $\mathcal{M}_S$ with fixed number N of mRNAs we represent the modulator as one-dimensional process $Z = Z_1$ and use the conservation relation $Z_1 + Z_2 = N$. Let $Y_j(t), j = 1, 2$ count the number of translation reactions via the structure Z_j, and let $Y(t) = [Y_1(t), Y_2(t)]$, whereas $\overline{Y}(t) := Y_1(t) + Y_2(t)$. Then $\Sigma = \frac{c_{2,1} c_{1,2} N}{(c_{1,2} + c_{2,1})^2}, C = [r_1, -r_2]^T, \mu_0 = [0, Nr_2], \mu = \frac{c_{2,1} N}{c_{1,2} + c_{2,1}}, A = c_{1,2} + c_{2,1}$ and hence

$$\lim_{t \to \infty} \frac{1}{t} \text{Cov}(Y(t)) = \frac{N}{c_{1,2} + c_{2,1}} \text{diag}([r_1 c_{2,1}, r_2 c_{1,2}]) + \begin{bmatrix} r_1 \\ -r_2 \end{bmatrix} \frac{2 c_{2,1} c_{1,2} N}{(c_{1,2} + c_{2,1})^3} [r_1, -r_2].$$

Consequently,

$$\lim_{t\to\infty} \frac{1}{t}\mathbb{V}\mathrm{ar}[\overline{Y}(t)] = \lim_{t\to\infty}\frac{1}{t}\mathbb{V}\mathrm{ar}[Y_1(t) + Y_2(t)]$$

$$= [1,1]\lim_{t\to\infty}\frac{1}{t}\mathrm{Cov}(Y_t)\begin{bmatrix}1\\1\end{bmatrix}$$

$$= \frac{(r_1 c_{2,1} + r_2 c_{1,2})N}{c_{1,2} + c_{2,1}} + (r_1 - r_2)^2\frac{2c_{2,1}c_{1,2}N}{(c_{1,2} + c_{2,1})^3}$$

$$= rN + 2\tau_S r_2^2(\frac{r_1}{r_2} - 1)^2 f(1 - f)N$$

$$= rN + 2\tau_S r_2^2(\rho - 1)^2 f(1 - f)N.$$

Hence, for the Fano factor

$$\mathrm{F_S} = 1 + \frac{2\tau_S r_2^2(\frac{r_1}{r_2} - 1)^2 f(1 - f)}{r_2(f(\frac{r_1}{r_2} - 1) + 1)}$$

$$= 1 + \frac{2\tau_S(\rho - 1)^2 f(1 - f)}{\tau_{TL}(f(\rho - 1) + 1)}$$

$$= 1 + \frac{1 + \frac{\tau_S}{\tau_{TX}}}{\frac{\tau_{TL}}{\tau_S} + \frac{\tau_{TL}}{\tau_{TX}}} \cdot \frac{2(\rho - 1)^2 f(1 - f)}{(f(\rho - 1) + 1)}$$

$$= 1 + \left(1 + \frac{\tau_S}{\tau_{TX}}\right)\Delta\mathrm{F_{S|TX}}.$$

Overall, for the noise contributions we obtained the terms

$$\Delta\mathrm{F_{TX}} = 2\frac{\tau_{TX}}{\tau_{TL}}(f(\rho - 1) + 1), \tag{194}$$

$$\Delta\mathrm{F_{S|TX}} = \frac{2(\rho - 1)^2 f(1 - f)}{(\frac{\tau_{TL}}{\tau_S} + \frac{\tau_{TL}}{\tau_{TX}})(f(\rho - 1) + 1)} \tag{195}$$

with (i) the fold change ρ, (ii) the share f of Z_1, (iii) the relative timescales τ_S/τ_{TL} and (iv) τ_{TX}/τ_{TL}, where $\tau_{TL} = r_2^{-1}$.

Both models $\mathcal{M}_{TX}$ and $\mathcal{M}_{TX,S}$ account for the transcription. Like $\mathcal{M}_{TX,S}$, the above translation model $\mathcal{M}_S$ (Eq. (136)-(140)) with $k = 2$ accounts for two mRNA structures. However, it neglects transcription. Instead, the number of transcripts in $\mathcal{M}_S$ is conserved. We set this constant number of transcripts to $N = \mu\delta^{-1}$, the mean of the combined model $\mathcal{M}_{TX,S}$, thereby matching the two models. Then $\Delta\mathrm{F_{S|TX}}$ is related to the Fano factor $\mathrm{F_S}$ of $\mathcal{M}_S$ by

$$\Delta\mathrm{F_S} = (1 + \frac{\tau_S}{\tau_{TX}})\Delta\mathrm{F_{S|TX}}. \tag{196}$$

Note that the correction factor in Eq. (196) only depends on the timescale difference between transcription and switching. And note further that $\Delta\mathrm{F_{S|TX}}$ does not depend on N, so the matching can be neglected.

In order to set the translational noise due to structure in relation to the translational noise due to transcription, we considered

$$\frac{\Delta\mathrm{F_{S|TX}}}{\Delta\mathrm{F_{TX,S}}} = \frac{\Delta\mathrm{F_{S|TX}}}{\Delta\mathrm{F_{S|TX}} + \Delta\mathrm{F_{TX}}} = \frac{1}{1 + (1 + \frac{\tau_{TX}}{\tau_S})h(\rho, f)}, \tag{197}$$

with $h(\rho, f) = ((\rho - 1)^{-1} + f)^2/f(1 - f)$, $\rho > 1, f \in (0,1)$. In this model, the contribution of the structure depends only on (i) the fold change ρ, (ii) the share f of Z_1 and (iii) the relative timescale τ_S/τ_{TX}. The dependence is visualized in figure 11. For a fixed ρ, the function h is minimal if $f = (\rho + 1)^{-1}$ which maximizes the contribution of the structure to the noise.

The contribution of the structure to the translational noise compared to the noise due to transcript copy number fluctuation increases with the fold change and with the relative timescale, when fixing the

remaining parameters (figure 11A). The dependence on the fraction f shows a local maximum (figure 11A) which, independently of the timescale, is located at $f = (\rho+1)^{-1}$. This means that the structure maximally impacts the translation noise when $f r_1 = (1-f)r_2$, i.e., when the odds of allocating the low-translated structure equals the fold change. Or in other words, the larger the separation between the high and low translation rate the less does the maximally noisy switching allocate the highly-translated structure. The odds of its abundance decrease in exactly the way that the fold change increases. We now consider the fraction at its maximizing value $f = (\rho+1)^{-1}$ (figure 11B). Not surprisingly, slow switching structures have a higher impact on the translational noise when ρ is fixed. However, this effect is limited, where the limit is dictated by the fold change of the translation rates (horizontal asymptotes in figure 11B at $(\rho-1)^2/(\rho+1)^2$). When fixing the relative timescale, however, we found that the contribution of the structure can be brought arbitrarily close to 1 for large enough ρ.

In more detail, the relative noise contribution of the structure, Eq. (197), is

$$\nu(f, \rho, \tau_{\mathrm{S}}/\tau_{\mathrm{TX}}) = \frac{(\rho-1)^2 f(1-f)}{(\rho-1)^2 f(1-f) + \left(\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}+1\right)(f(\rho-1)+1)^2}.$$

For fixed ρ, it can be derived analytically that $\nu(f, \rho, \tau_{\mathrm{S}}/\tau_{\mathrm{TX}})$ is maximized at $f^* = (\rho+1)^{-1}$. Then

$$\nu(f^*, \rho, \tau_{\mathrm{S}}/\tau_{\mathrm{TX}}) = \frac{(\rho-1)^2}{(\rho+1)^2 + 4\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}} \cdot \rho}$$

and

$$\lim_{\tau_{\mathrm{S}}/\tau_{\mathrm{TX}} \to \infty} \nu(f^*, \rho, \tau_{\mathrm{S}}/\tau_{\mathrm{TX}}) = \left(\frac{\rho-1}{\rho+1}\right)^2.$$

Furthermore, $\lim_{\rho \to \infty} \nu(f^*, \rho, \tau_{\mathrm{S}}/\tau_{\mathrm{TX}}) = 1$ for every $\tau_{\mathrm{S}}/\tau_{\mathrm{TX}} > 0$. Finally, $\nu(f^*, \rho, \tau_{\mathrm{S}}/\tau_{\mathrm{TX}}) = \alpha \in (0, 1)$ is solved by

$$\rho_\alpha = \frac{\alpha^{-1}+1+2\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}}{\alpha^{-1}-1} + \sqrt{\left(\frac{\alpha^{-1}+1+2\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}}{\alpha^{-1}-1}\right)^2 - 1},$$

which for $\alpha = 1/2$ yields

$$\rho_{1/2} = 3 + 2\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}} + \sqrt{\left(3 + 2\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}\right)^2 - 1} \approx 6 + 4\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}},$$

whereas $\alpha = 0.05$ yields

$$\rho_{0.05} = \frac{21}{19} + \frac{2}{19}\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}} + \sqrt{\left(\frac{21}{19} + \frac{2}{19}\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}\right)^2 - 1}.$$

The value of $\rho_{0.05}$ approaches $(1 + \sqrt{0.05})/(1 - \sqrt{0.05}) = 1.58$ for large $\tau_{\mathrm{S}}/\tau_{\mathrm{TX}}$ and is approximately $2.21 + 0.21\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}$ for $\tau_{\mathrm{S}}/\tau_{\mathrm{TX}} < 0.1$.

In summary, the contribution of the structure makes up half when $\rho \approx 6 + 4\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}$ and potentially becomes negligible (contribution at most 0.05) when $\rho \leq 2.21 + 0.21\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}}$ in the fast switching regime ($\frac{\tau_{\mathrm{S}}}{\tau_{\mathrm{TX}}} < 0.1$) and $\rho \leq 1.58$ for slow switching ($\frac{\tau_{\mathrm{S}}}{\tau_{\mathrm{TX}}} > 10$). Considering fast switching compared to the mRNA lifetime, fold changes of $\rho \geq 4.31$ yield potentially non-negligible contributions by this argumentation.

Next, we compared the contribution of structure switching to other sources of noise in the protein abundance. Promoter switching is a well established noise source. When the promoter alternates between the On and Off state, transcription occurs in bursts, and this heterogeneity transfers to the protein counts during the translation step. We used the standard three-state gene expression model with a two-state Markov model for the promoter. To include promoter switching we replaced the reactions R_3 and R_5 in the model $\mathcal{M}_{\mathrm{TX,S}}$ by

$$
\begin{aligned}
R_7 \quad &: \mathrm{G}_{\mathrm{off}} \xrightarrow{k_{\mathrm{on}}} \mathrm{G}_{\mathrm{on}} \\
R_8 \quad &: \mathrm{G}_{\mathrm{on}} \xrightarrow{k_{\mathrm{off}}} \mathrm{G}_{\mathrm{off}} \\
R_{3a} \quad &: \mathrm{G}_{\mathrm{on}} \xrightarrow{\alpha \cdot c_{2,1}} \mathrm{G}_{\mathrm{on}} + \mathrm{Z}_1 \\
R_{5a} \quad &: \mathrm{G}_{\mathrm{on}} \xrightarrow{\alpha \cdot c_{1,2}} \mathrm{G}_{\mathrm{on}} + \mathrm{Z}_2.
\end{aligned}
$$
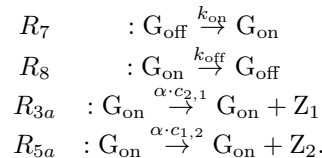
to obtain the model $\mathcal{M}_{\mathrm{TX,G,S}}$. Denote by $|\mathrm{G}|$ the fixed sum of $\mathrm{G_{on}}$ and $\mathrm{G_{off}}$, which typically is 1. Define $\mathbb{V}\mathrm{ar_G} = |\mathrm{G}|\mathrm{P_{off}P_{on}}$, where $\mathrm{P_{on}} = 1 - \mathrm{P_{off}} = \frac{k_{\mathrm{on}}}{k_{\mathrm{on}}+k_{\mathrm{off}}}$, $\overline{f} = 1 - f, k = k_{\mathrm{on}} + k_{\mathrm{off}}$ and $\mathbb{E}_{\mathrm{X}} = |\mathrm{G}|\mathrm{P_{on}}\frac{\mu}{\delta}$. With

$$A = \begin{bmatrix} k_{\mathrm{on}} + k_{\mathrm{off}} & 0 & 0 \\ -f\mu & c_{1,2} + \delta & -c_{2,1} \\ -(1-f)\mu & -c_{1,2} & c_{2,1} + \delta \end{bmatrix},$$

$$\Sigma = \begin{bmatrix} \mathbb{V}\mathrm{ar_G} & \frac{f\mu\mathbb{V}\mathrm{ar_G}}{k+\delta} & \frac{\overline{f}\mu\mathbb{V}\mathrm{ar_G}}{k+\delta} \\ \frac{f\mu\mathbb{V}\mathrm{ar_G}}{k+\delta} & \frac{f^2\mu^2\mathbb{V}\mathrm{ar_G}}{\delta(k+\delta)} + f\mathbb{E}_{\mathrm{X}} & \frac{f\overline{f}\mu^2\mathbb{V}\mathrm{ar_G}}{\delta(k+\delta)} \\ \frac{\overline{f}\mu\mathbb{V}\mathrm{ar_G}}{k+\delta} & \frac{f\overline{f}\mu^2\mathbb{V}\mathrm{ar_G}}{\delta(k+\delta)} & \frac{\overline{f}^2\mu^2\mathbb{V}\mathrm{ar_G}}{\delta(k+\delta)} + \overline{f}\mathbb{E}_{\mathrm{X}} \end{bmatrix},$$
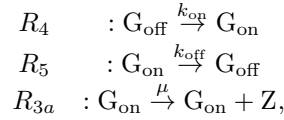
we obtained from Eq. (33)

$$\mathrm{F_{TX,G,S}} = 1 + 2 \cdot \frac{r}{\delta} + 2 \cdot \frac{\mathrm{P_{off}}\mu r}{(k_{\mathrm{on}} + k_{\mathrm{off}})\delta} + 2 \cdot \frac{(r_1 - r)(r - r_2)}{(c_{2,1} + c_{1,2} + \delta)r}.$$

The expression analogously to Eq. (191) reads

$$\mathrm{F_{TX,G,S}} = 1 + \Delta\mathrm{F_{TX}} + \Delta\mathrm{F_{G|TX}} + \Delta\mathrm{F_{S|TX,G}}$$
$$\Delta\mathrm{F_{G|TX}} = 2\frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{TL}}}\mathrm{P_{off}^2}\overline{\beta}(f(\rho - 1) + 1)$$

with $\Delta\mathrm{F_{TX}}$ as in Eq. (194) and $\Delta\mathrm{F_{S|TX,G}} = \Delta\mathrm{F_{S|TX}}$ as in Eq. (195). Here, $\overline{\beta} = \mu/k_{\mathrm{off}}$ denotes the mean number of mRNA per burst and $\mathrm{P_{off}}$ the frequency of being in the promoter Off state. In the corresponding reference model $\mathcal{M}_{\mathrm{TX,G}}$, that we obtained from $\mathcal{M}_{\mathrm{TX}}$ by replacing the reaction R_3 by

$$R_4 \quad : \mathrm{G_{off}} \xrightarrow{k_{\mathrm{on}}} \mathrm{G_{on}}$$
$$R_5 \quad : \mathrm{G_{on}} \xrightarrow{k_{\mathrm{off}}} \mathrm{G_{off}}$$
$$R_{3a} \quad : \mathrm{G_{on}} \xrightarrow{\mu} \mathrm{G_{on}} + \mathrm{Z},$$

we obtained

$$A = \begin{bmatrix} k_{\mathrm{on}} + k_{\mathrm{off}} & 0 \\ -\mu & \delta \end{bmatrix}, \Sigma = \begin{bmatrix} \mathbb{V}\mathrm{ar_G} & \frac{\mu\mathbb{V}\mathrm{ar_G}}{k_{\mathrm{on}}+k_{\mathrm{off}}+\delta} \\ \frac{\mu\mathbb{V}\mathrm{ar_G}}{k_{\mathrm{on}}+k_{\mathrm{off}}+\delta} & \frac{\mu^2\mathbb{V}\mathrm{ar_G}}{\delta(k_{\mathrm{on}}+k_{\mathrm{off}}+\delta)} + \mathbb{E}_{\mathrm{X}} \end{bmatrix},$$

and thus from Eq. (33)

$$\mathrm{F_{TX,G}} = 1 + 2 \cdot \frac{r}{\delta} + 2 \cdot \frac{\mathrm{P_{off}}\mu r}{(k_{\mathrm{on}} + k_{\mathrm{off}})\delta} = \mathrm{F_{TX}} + 2 \cdot \frac{\mathrm{P_{off}^2}\mu\delta^{-1}(f(\rho - 1) + 1)}{k_{\mathrm{off}}r_2^{-1}}.$$

This implies for the contribution of the structure switching

$$\frac{\Delta\mathrm{F_{S|TX,G}}}{\Delta\mathrm{F_{TX,G,S}}} = \frac{1}{1 + (1 + \frac{\tau_{\mathrm{TX}}}{\tau_{\mathrm{S}}})(1 + \mathrm{P_{off}^2}\overline{\beta})h(\rho, f)}, \tag{198}$$

We observed that, in this extended model, the contribution of the structure additionally depends on $\mathrm{P_{off}}$ and $\overline{\beta}$. The contribution is larger, when the promoter is more frequently active and when the mean number of mRNA per burst is smaller. The qualitative dependence on the parameters ρ, f and $\tau_{\mathrm{S}}/\tau_{\mathrm{TX}}$ is as before. Secondly, we asked by how much the effect of structure switching is diluted by the protein decay. We extended the model by a protein decay reaction. In detail, $\mathcal{M}_{\mathrm{TX,G,S,P}}$ and $\mathcal{M}_{\mathrm{TX,G,P}}$ are extensions of $\mathcal{M}_{\mathrm{TX,G,S}}$ and $\mathcal{M}_{\mathrm{TX,G}}$, respectively, by the reaction $\mathrm{Prot} \to \emptyset$ with rate κ. Then the theory of linear CRNs provides the stationary variances, from which the Fano factors are computed as

$$\Delta\mathrm{F_{TX,G,P}} = \frac{\mathrm{P_{off}}r\mu(\delta + k + \kappa)}{(\delta + \kappa)(k + \kappa)(k + \delta)} + \frac{r}{\delta + \kappa}$$
$$\Delta\mathrm{F_{TX,G,S,P}} = \Delta\mathrm{F_{TX,G,P}} + \frac{(r_1 - r)(r - r_2)}{r(c_{2,1} + c_{1,2} + \delta + \kappa)}$$

This yields

$$\frac{\Delta F_{S|TX,G,P}}{\Delta F_{TX,G,S,P}} = \frac{1}{1 + \left(1 + \frac{c_{2,1}+c_{1,2}}{\delta+\kappa}\right)\left(1 + P_{off}^2 \overline{\beta}\frac{1+\frac{\kappa}{k+\delta}}{1+\frac{\kappa}{k}}\right)h(\rho,f)}$$

$$= \frac{1}{1 + (1 + (\frac{\tau_S}{\tau_{TX}} + \frac{\tau_S}{\tau_P})^{-1})\left(1 + P_{off}^2\overline{\beta} \cdot \frac{1+(\frac{\tau_P}{\tau_G}+\frac{\tau_P}{\tau_{TX}})^{-1}}{1+\frac{\tau_G}{\tau_P}}\right)h(\rho,f)}, \tag{199}$$

where $\tau_P = \kappa^{-1}$ is the inverse protein degradation rate and $\tau_G = k^{-1}$ is the inverse sum of the On and Off switching rates of the promoter. For $\tau_P \gg \tau_{TX}$, the expression amounts to Eq. (198). This expresses what the structure switching contributes to the Fano factor compared to other external protein noise sources, i.e., the intrinsic transcription noise and the promoter switching, while excluding the intrinsic protein noise from the consideration. As an additional observation, if we include the intrinsic protein noise in the comparison, and again assume $\tau_P \gg \tau_{TX}$, we obtain

$$\frac{\Delta F_{S|TX,G}}{1 + \Delta F_{TX} + \Delta F_{G|TX} + \Delta F_{S|TX,G}} = \frac{1}{1 + \left((1 + \frac{\tau_{TX}}{\tau_S})(1 + P_{off}^2\overline{\beta}) + \pi^{-1}\right)h(\rho,f)},$$

where $\pi = r/\delta$ is the average number of protein translated by one mRNA.

The findings in this section justify incorporating noise due to mRNA structure into CRN modeling. Since with increasing number of structures, direct stochastic simulations of the conversion reactions are expected to become inefficient, we aim for a model reduction that incorporates effects of the switching. This motivates the remainder of our work. While we chose a model reduction approach via approximate stochastic filtering, the projection operator formalism has been used as an alternative in the literature [173, 174, 175, 99, 176]. We opted for an approximate stochastic simulation approach that does not presume specific copy number regimes and switching timescales. The projection operator formalism provides process equations only for the diffusion approximations. On a discrete state space, it yields probability evolution equations with memory terms, from which no immediate process equation can be derived for simulations. Stochastic filtering is more commonly applied for state estimation in partially observed reaction networks [58, 89, 87], whereas for model reduction purposes it was proposed by Zechner and Koeppl [3] and rediscovered by Öcal et al [80].

8.1.4 Approximate filters for mRNA structure switching

The Hawkes filter can be compared to the multinomial filter, which is obtained as a multinomial closure of the posterior mean equation (64). The stationary prior distribution of the heterogeneity network is a multinomial $M(N, (p_1, \ldots, p_k))$. This suggests a multinomial moment closure, assuming $M(N, (\theta_1(t), \ldots, \theta_k(t)))$ for the posterior distribution. We use $\theta_i(t) = N^{-1}M_i(t)$ and express the second order terms in the posterior mean equation (64) as

$$V_{ji} = \begin{cases} N\theta_i(t)\theta_j(t), & i \neq j \\ N\theta_i(t)(1 - \theta_i(t)), & i = j. \end{cases}$$

This leads to the multinomial parameter evolution equation

$$\frac{d}{dt}\theta_i = \sum_{\substack{j=1 \\ j\neq i}}^{k}(-c_{i,j}\theta_i + c_{j,i}\theta_j) - r_i\theta_i + \sum_{j=1}^{k}r_j\theta_j\theta_i \tag{200}$$

in-between jumps of the translation counting process. When $Y(t)$ jumps at t, we get

$$\theta_i(t) = \begin{cases} \theta_i(t-)\left(1 - \frac{1}{N} + \frac{r_i}{N\sum_{j=1}^{k-1}(r_j-r_k)\theta_j(t-)+r_k}\right) & i \neq k \\ 1 - \sum_{i=1}^{k-1}\theta_i(t-) & i = k. \end{cases} \tag{201}$$

We call

$$R: \quad \emptyset \xrightarrow{\;N\sum_{i=1}^{k} r_i\; \theta_i(t)\;} \text{Prot} \tag{202}$$

the model $\mathcal{M}_{\mathrm{M}}^P$, where M stands for multinomial. It is shown in [A5] that the multinomial moment closure coincides with an entropic matching evolution equation. For $k = 2$, we call the model $\mathcal{M}_{\mathrm{B}}^P$, where B denotes binomial. The intensity of $\mathcal{M}_{\mathrm{B}}^P$ is equivalently expressed by

$$\lambda_t^{\mathrm{B}} = Nr_1\theta(t) + Nr_2(1 - \theta(t)). \tag{203}$$

The model $\mathcal{M}_{\mathrm{B}}^Y$ differentiates between translation reactions R_1, counted by $Y_1(t)$, and R_2, counted by $Y_2(t)$. The evolution equation between jumps is equal to Eq. (200). However, the available information about which reaction fired modifies the jump update equation Eq. (201) to

$$\theta(t) = \begin{cases} \frac{1}{N} + \frac{N-1}{N}\theta(t-), & \text{if } Y_1(t) = Y_1(t-) + 1 \\ \frac{N-1}{N}\theta(t-), & \text{if } Y_2(t) = Y_2(t-) + 1 \end{cases}.$$

Complementary, we considered the Hawkes approximate filter as in section 8.1.1. The mRNA switching replaces promoter switching and protein production events replace transcription events

$$R: \quad \emptyset \xrightarrow{\;\hat{Z}^{\mathrm{H}}(t)\;} \text{Prot}. \tag{204}$$

We indicate this reduced model by $\mathcal{M}_{\mathrm{L}}^P$ with L denoting linear, For $k = 2$, it simplifies to a one-dimensional evolution equation

$$\mathrm{d}\hat{Z}^{\mathrm{H}}(t) = -A(\hat{Z}^{\mathrm{H}}(t) - \mu)\,\mathrm{d}t + \beta(t)(\,\mathrm{d}Y(t) - \hat{Z}^{\mathrm{H}}(t)\,\mathrm{d}t), \tag{205}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\beta(t) = -2A\beta(t) + 2A\Sigma - \beta(t)^2, \tag{206}$$

$$\hat{Z}^{\mathrm{H}}(0) = \mu, \quad \beta(0) = \frac{\Sigma}{\mu}, \tag{207}$$

with

$$A = c_{1,2} + c_{2,1}, \Sigma = N(r_1 - r_2)^2 \frac{c_{1,2}c_{2,1}}{(c_{1,2} + c_{2,1})^2}, \mu = N\frac{r_1 c_{2,1} + r_2 c_{1,2}}{c_{1,2} + c_{2,1}}.$$

When we are interested in the asymptotic behavior, we may assume that $\beta(t)$ has equilibrated and replace it by the constant $\beta = \sqrt{A^2 + \frac{2A\Sigma}{\mu}} - A$. Then the model reads in differential form

$$\mathrm{d}\hat{Z}^{\mathrm{H}}(t) = -\alpha(\hat{Z}^{\mathrm{H}}(t) - \mu_0)\,\mathrm{d}t + \beta\,\mathrm{d}Y(t), \quad \hat{Z}^{\mathrm{H}}(0) = \mu,$$
$$\alpha = A + \beta, \quad \mu_0 = \frac{\mu A}{A + \beta}, \tag{208}$$

and in integral form

$$\hat{Z}^{\mathrm{H}}(t) = \mu_0 + \int_0^t \beta e^{-\alpha(t-u)}\,\mathrm{d}Y(u), \tag{209}$$

upon replacement of $\hat{Z}^{\mathrm{H}}(0) = \mu_0$ for convenience, which is justified for the asymptotic analysis. In particular, under the equilibrium assumption the model $\mathcal{M}_{\mathrm{L}}^P$ for $k = 2$ yields a Hawkes process with standard exponential kernel. We indicate whether we used the non-equilibrated or the equilibrated case.

8.1.5 Comparison of Fano factors for approximate filters

In this section, we apply the result of section 4.4 and investigate the Fano factor in the two-structure-model with constant mRNA number. While comparing the contributions of noise due to structure and intrinsic transcriptional noise in figure 11, here we investigated the relevance of fast switching

timescales compared to the slowest translation rate, i.e., the dependence on $\tau_S/\tau_{TL} = \frac{r_2}{c_{1,2}+c_{2,1}}$. As discussed in subsection 8.1.3, the Fano contribution ΔF_S due to structure is

$$\Delta F_S = \frac{2\tau_S(\rho-1)^2 f(1-f)}{\tau_{TL}(f(\rho-1)+1)}. \tag{210}$$

For the Fano factor F_B of the binomial filter, Eqs. (200)-(201) for $k = 2$, we use that it is of the form in Eq. (108) with one-dimensional global parameter and $V(t) = (r_1 - r_2)\theta(t-)(1 - \theta(t-))$, see the section 8.1.6 below. Using Eq. (109), we find that $\Delta F_B = \Delta F_S + \Delta F_{B|S}$ with

$$\Delta F_{B|S} = \frac{2\tau_S(\rho-1)^2(N-1)\mathbb{V}\mathrm{ar}[\theta_\infty]}{\tau_{TL}(f(\rho-1)+1)}. \tag{211}$$

This shows that the binomial filter systematically overestimates the exact Fano factor, because $\Delta F_{B|S} > 0$ if $N \geq 2$. We explain this as follows. The binomial filter estimates the filter mean with only one degree of freedom, since the filter parameter θ is a scalar. The binomial ansatz, however, relies on the implicit assumption that the events for finding either of the N mRNA molecules in structure Z_1 are independent. The success probability is then calibrated, such that the mean matches, because there is only one degree of freedom. But the events are not independent, since knowledge about the recent history of the translation process adds implicit negative correlations. There is one exception. In the case, when no event was observed for a long time, the filtering distribution converges to a binomial distribution, implying that the independence assumption is met. For the derivation of the convergence, suppose the $\theta_1, \ldots, \theta_k$ satisfy

$$0 = \sum_{\substack{i=1 \\ i\neq j}}^{k} c_{ij}\theta_i - c_{ji}\theta_j - r_j\theta_j + \sum_{i=1}^{k} r_i\theta_i\theta_j$$

for all $j = 1, \ldots, \theta_k$. Let $x_1, \ldots, x_k \in \mathbb{N}_0$. Then upon multiplication with x_j/θ_j and summation we get

$$0 = \sum_{\substack{i,j=1 \\ i\neq j}}^{k} c_{ij}x_j\frac{\theta_i}{\theta_j} - c_{ji}x_j - \sum_{j=1}^{k} r_jx_j + \sum_{i=1}^{k} r_i\theta_i(\sum_{j=1}^{k} x_j)$$

$$= \sum_{\substack{i,j=1 \\ i\neq j}}^{k} c_{ij}x_j\frac{\theta_i}{\theta_j} - c_{ji}x_j - \sum_{j=1}^{k} r_j(x_j - \theta_j\sum_{i=1}^{k} x_i).$$

Now, we employ the multinomial ansatz

$$\pi(x_1, \ldots, x_k) = \theta_1^{x_1} \ldots \theta_k^{x_k}\frac{N!}{x_1! \ldots x_k!}$$

for any $x_1, \ldots, x_k \in \mathbb{N}_0$ that add up to N.

Plugging this ansatz into the dynamics (right-hand side) of the exact filtering equation, we obtain

$$\frac{d}{dt}\pi(x_1, \ldots, x_k)$$

$$= \sum_{\substack{i,j=1 \\ i\neq j}}^{k} c_{ij}(x_i + 1)\pi(\ldots, x_i + 1, \ldots, x_j - 1 \ldots)$$

$$- c_{ji}x_j\pi(x_1, \ldots, x_k) - \sum_{j=1}^{k} r_j(x_j - \mathbb{E}_\pi[X_j])\pi(x_1, \ldots, x_k)$$

$$= \pi(x_1, \ldots, x_k)\left\{\sum_{\substack{i,j=1 \\ i\neq j}}^{k} c_{ij}x_j\frac{\theta_i}{\theta_j} - c_{ji}x_j - \sum_{j=1}^{k} r_j(x_j - N\theta_j)\right\}$$

$$= 0$$

by applying $N = \sum_{i=1}^{k} x_i$. Hence, the multinomial distribution is the equilibrium distribution of the filtering equation between jumps.

In order to evaluate the term $\mathbb{V}\mathrm{ar}[\theta_\infty]$ in Eq. (211), we made use of that computes the asymptotic conditional intensity distribution (ACID) in a simulation-free way. The technique makes use of the convenience that θ_t is a piecewise-deterministic Markov process.

The reduced model $\mathcal{M}_B^Y$ yields the analogous result, Eq. (211), with the respective θ_t. In this case, the asymptotic moment $\mathbb{V}\mathrm{ar}[\theta_\infty]$ can be obtained by a tight moment bound approach [42, 41]. Applying the Ito formula for counting processes to the powers $\theta_t^n, n = 2, \ldots, N$ yields a system of $N - 1$ linear equations in the moments $\mathbb{E}[\theta_\infty^2], \ldots \mathbb{E}[\theta_\infty^{N+1}]$. Then $\mathbb{E}[\theta_\infty^2]$ can be maximized and minimized under the additional inequality constraints

$$\mathbb{E}[\theta_\infty] \geq \mathbb{E}[\theta_\infty^2] \geq \mathbb{E}[\theta_\infty^3] \geq \cdots \geq \mathbb{E}[\theta_\infty^{N+1}] \geq 0.$$

For large enough N the obtained upper and lower bound are tight. Details on the moment equations and the linear program are found in [A5].

The reduced model $\mathcal{M}_L^P$ agrees with the exact model in terms of the Fano factor, compare 4.1 for the non-equilibrated case. For the equilibrated case, Eq. (208), see section 8.1.6 below. Apparently, the model uses its degrees of freedom μ_0, α, β to match the three parameters (i) mean, (ii) variance and (iii) correlation time of the modulating process Z_1 or, equivalently, $f, \rho, \tau_S/\tau_{TL}$, that determine the asymptotic Fano factor.

8.1.6 Derivation of the Fano factors for the model reductions $\mathcal{M}_B^P$, $\mathcal{M}_B^Y$ and $\mathcal{M}_L^P$

For the model $\mathcal{M}_B^P$, i.e., $\mathcal{M}_M^P$ for $k = 2$

$$R: \quad \emptyset \xrightarrow{\ (r_1 - r_2)N\theta(t-) + Nr_2\ } \mathrm{Prot}, \tag{212}$$

compare Eqs. (200)-(201), the intensity is expressed affine-linearly in $\theta(t-)$ of the form in Eq. (108) with $\Theta_t = \theta(t) \in [0,1]$, $d = l = 1$, $\bar{\Theta} = c_{2,1}(c_{1,2} + c_{2,1})^{-1}$, $V(t) = (r_1 - r_2)\theta(t-)(1 - \theta(t-))$, $A = c_{1,2} + c_{2,1}, C = N(r_1 - r_2), \alpha = Nr_2$.

Then Eq. (109) yields

$$\lim_{t\to\infty} t^{-1}\mathbb{V}\mathrm{ar}[Y_t] = \mathbb{E}\lambda_\infty + 2CA^{-1}(\mathbb{E}[V(\infty)] + C\mathbb{V}\mathrm{ar}[\theta_\infty])$$

$$= (r_1 - r_2)N\frac{c_{2,1}}{c_{1,2} + c_{2,1}} + Nr_2 + \frac{2(r_1 - r_2)^2 N}{c_{1,2} + c_{2,1}}\left(\frac{c_{2,1}}{c_{1,2} + c_{2,1}} - \frac{Nc_{2,1}^2}{(c_{1,2} + c_{2,1})^2} + (N-1)\mathbb{E}[\theta_\infty^2]\right)$$

$$= (r_1 - r_2)N\frac{c_{2,1}}{c_{1,2} + c_{2,1}} + Nr_2 + \frac{2(r_1 - r_2)^2 N}{c_{1,2} + c_{2,1}}\left(\frac{c_{1,2}c_{2,1}}{(c_{1,2} + c_{2,1})^2} + (N-1)\mathbb{V}\mathrm{ar}[\theta_\infty^2]\right)$$

$$= \lim_{t\to\infty} t^{-1}\mathbb{V}\mathrm{ar}[\tilde{Y}_t] + \frac{2(r_1 - r_2)^2 N(N-1)}{c_{1,2} + c_{2,1}}\mathbb{V}\mathrm{ar}[\theta_\infty],$$

where $\tilde{Y}_t$ denotes the counting process of model $\mathcal{M}_S$. The asymptotic Fano factor is hence given by

$$F_B = F_S + \frac{2r_2^2(\frac{r_1}{r_2} - 1)^2(N-1)\mathbb{V}\mathrm{ar}[\theta_\infty]}{\tau_S^{-1}r_2(f(\frac{r_1}{r_2} - 1) + 1)}$$

$$= F_S + \frac{2\tau_S(\rho - 1)^2(N-1)\mathbb{V}\mathrm{ar}[\theta_\infty]}{\tau_{TL}(f(\rho - 1) + 1)}.$$

For the model $\mathcal{M}_B^Y$ we have a bivariate counting process $Y_t = [Y_1(t), Y_2(t)]$ with $\Theta_t = \theta(t) \in [0,1]$, $d = 2, l = 1$, $V(t) = [V_1(t), V_2(t)]$ for $V_1(t) = r_1\theta(t-)(1 - \theta(t-)), V_2(t) = -r_2\theta(t-)(1 - \theta(t-))$, $A = c_{1,2} + c_{2,1}, C_1 = Nr_1, C_2 = -Nr_2, \alpha_1 = 0, \alpha_2 = Nr_2$. Denote $\overline{Y}(t) := Y_1(t) + Y_2(t)$. Then the

asymptotic variance Y_t is given via the asymptotic covariance of Y_t, i.e.,

$$\lim_{T \to \infty} \frac{\mathbb{V}\mathrm{ar}[\overline{Y}(T)]}{T} = [1,1] \lim_{t \to \infty} \frac{1}{t} \mathrm{Cov}(Y_t) \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

$$= (r_1 - r_2)N \frac{c_{2,1}}{c_{1,2} + c_{2,1}} + Nr_2 + \frac{2(r_1 - r_2)^2 N}{c_{1,2} + c_{2,1}} \left(\frac{c_{1,2}c_{2,1}}{(c_{1,2} + c_{2,1})^2} + (N-1)\mathbb{V}\mathrm{ar}[\theta_\infty^2] \right)$$

$$= \lim_{t \to \infty} t^{-1}\mathbb{V}\mathrm{ar}[\tilde{Y}_t] + \frac{2(r_1 - r_2)^2 N(N-1)}{c_{1,2} + c_{2,1}}\mathbb{V}\mathrm{ar}[\theta_\infty],$$

where $\tilde{Y}_t$ denotes the counting process of model $\mathcal{M}_\mathrm{S}$. The asymptotic Fano factor is hence given by

$$F_\mathrm{B} = F_\mathrm{S} + \frac{2\tau_\mathrm{S}(\rho - 1)^2(N-1)\mathbb{V}\mathrm{ar}[\theta_\infty]}{\tau_\mathrm{TL}(f(\rho - 1) + 1)}.$$

For $\mathcal{M}_\mathrm{L}^P$, we use $\Theta(t) = Z_0(t)$ and $V(t) = \bar{B}C^T \hat{Z}^\mathrm{H}(t)\mu^{-1}$. Then at equilibrium in Eq. (111) we obtain

$$A\mathbb{C}\mathrm{ov}[\Theta_\infty] - \mathbb{C}\mathrm{ov}[\Theta_\infty]A^T$$
$$= \mathbb{E}\left[V(\infty) \mathrm{diag}^{-1}(\lambda_\infty)V(\infty)^T \right]$$
$$= \bar{B}C^T \mu^{-1}C\bar{B}$$

and, assuming that A is invertible, see appendix B, in Eq. (109)

$$\lim_{t \to \infty} t^{-1}\mathbb{E}[(Y_t - \mathbb{E}[Y_t])(Y_t - \mathbb{E}[Y_t])^T]$$
$$= \mathbb{E}\lambda_\infty + CA^{-1}(\bar{B}A^T + A\bar{B})A^{-T}C^T + CA^{-1}(\bar{B}C^T\mu^{-1}C\bar{B})A^{-T}C^T$$
$$= \mathbb{E}\lambda_\infty + CA^{-1}(A\Sigma + \Sigma A^T)A^{-T}C^T$$
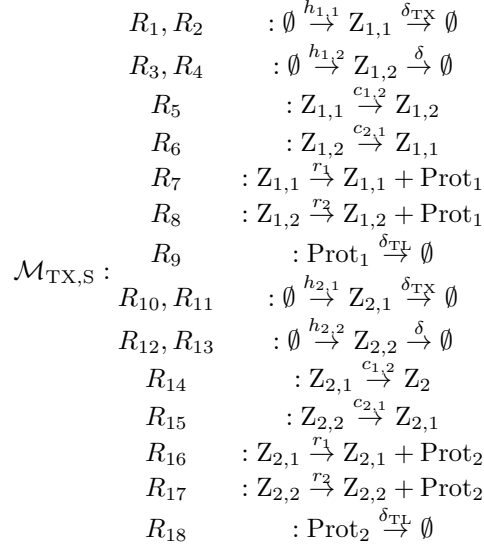$$= \mathbb{E}\lambda_\infty + C(\Sigma A^{-T} + A^{-1}\Sigma)C^T.$$

8.1.7 Demonstration of modular embedding of mRNA structure switching

We demonstrate how the binomial and the Hawkes filter can be used to incorporate intragenic translational heterogeneity due to mRNA structure into existing gene networks in a modular way. For this purpose, we need to generalize the filters for time-varying N. We assume that each mRNA structure degrades with the same rate and that new mRNA molecules are synthesized with probability given by the stationary distribution of the prior process, i.e., a categorical distribution $[p_1, \ldots, p_k]$ as in Eq. (138). Let us first consider the case $N(t) = N(t-) - 1$. Since the $N(t-)$ mRNA molecules are non-distinguishable, by symmetry each of them is in conformation Z_i with probability $\mathbb{E}_{\tilde{\pi}(t-)}[Z_i]/N(t-)$, where $\tilde{\pi}(t)$ denotes the exact or approximate filtering distribution. After any of the non-distinguishable molecules degraded, again by symmetry, the remaining ones are still in the conformation Z_i with probability $\mathbb{E}_{\tilde{\pi}(t-)}[Z_i]/N(t-)$, which implies $\mathbb{E}_{\tilde{\pi}(t)}[Z_i] = \frac{N(t)}{N(t)+1}E_{\tilde{\pi}(t-)}[Z_i]$. For the multinomial filter, this means $\theta_i(t) = \theta_i(t-)$. Let us next consider $N(t) = N(t-) + 1$. The newly synthesized mRNA molecule is in conformation Z_i with probability p_i, whereas the $N(t-)$ are of type Z_i with probability $\mathbb{E}_{\tilde{\pi}(t-)}[Z_i]/N(t-)$. Hence, the expected mRNA molecule of type Z_i at time t are $\mathbb{E}_{\tilde{\pi}(t)}[Z_i] = \mathbb{E}_{\tilde{\pi}(t-)}[Z_i] + p_i$. For the multinomial filter, this reads $\theta_i(t) = \frac{N(t)-1}{N(t)}\theta_i(t-) + \frac{p_i}{N(t)}$. For the Hawkes model, let us first consider the case $N(t) = N(t-) - 1$, then $Z_0(t) = \frac{N(t)}{N(t-)}Z_0(t-)$ because it scales linearly in N, and furthermore $\beta(t) = \beta(t-)$, because the ratio of the variance and the mean does not scale with N. Next, we consider $N(t) = N(t-) + 1$. Then $Z_0(t) = Z_0(t-)$, because a stationary contribution to the mean of the new mRNA does not contribute to $Z_0(t)$, which is linearly shifted to have mean zero. The contribution of the new mRNA to the variance is independent of the conditional variance of the present mRNAs, which means it is added. For the ratio of the variance and the mean this implies

$$\beta(t) = \frac{N(t-)}{N(t-) + 1}\beta(t-) + \frac{1}{N(t-) + 1}\frac{(r_1 - r_2)^2 f(1-f)}{(r_1 f + r_2(1-f))}.$$
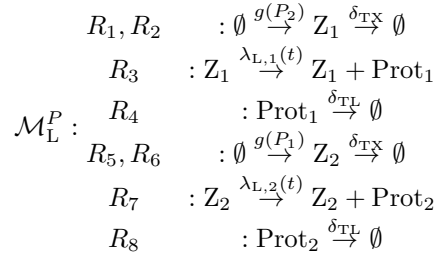
As a case study, we considered a stochastic toggle switch [177] which exhibits bistability caused by mutual repression of two gene products. The first output protein represses the mRNA of the second

output protein via a decreasing Hill function and vice versa. In the model, we included the effect of mRNA structure, such that the full symmetric reaction system that we considered contains the following reactions

$$\mathcal{M}_{\text{TX,S}}: \quad
\begin{aligned}
R_1, R_2 \quad & : \emptyset \overset{h_{1,1}}{\rightarrow} Z_{1,1} \overset{\delta_{\text{TX}}}{\rightarrow} \emptyset \\
R_3, R_4 \quad & : \emptyset \overset{h_{1,2}}{\rightarrow} Z_{1,2} \overset{\delta}{\rightarrow} \emptyset \\
R_5 \quad & : Z_{1,1} \overset{c_{1,2}}{\rightarrow} Z_{1,2} \\
R_6 \quad & : Z_{1,2} \overset{c_{2,1}}{\rightarrow} Z_{1,1} \\
R_7 \quad & : Z_{1,1} \overset{r_1}{\rightarrow} Z_{1,1} + \text{Prot}_1 \\
R_8 \quad & : Z_{1,2} \overset{r_2}{\rightarrow} Z_{1,2} + \text{Prot}_1 \\
R_9 \quad & : \text{Prot}_1 \overset{\delta_{\text{TL}}}{\rightarrow} \emptyset \\
R_{10}, R_{11} \quad & : \emptyset \overset{h_{2,1}}{\rightarrow} Z_{2,1} \overset{\delta_{\text{TX}}}{\rightarrow} \emptyset \\
R_{12}, R_{13} \quad & : \emptyset \overset{h_{2,2}}{\rightarrow} Z_{2,2} \overset{\delta}{\rightarrow} \emptyset \\
R_{14} \quad & : Z_{2,1} \overset{c_{1,2}}{\rightarrow} Z_2 \\
R_{15} \quad & : Z_{2,2} \overset{c_{2,1}}{\rightarrow} Z_{2,1} \\
R_{16} \quad & : Z_{2,1} \overset{r_1}{\rightarrow} Z_{2,1} + \text{Prot}_2 \\
R_{17} \quad & : Z_{2,2} \overset{r_2}{\rightarrow} Z_{2,2} + \text{Prot}_2 \\
R_{18} \quad & : \text{Prot}_2 \overset{\delta_{\text{TL}}}{\rightarrow} \emptyset
\end{aligned}$$

with $h_{1,1} = g(P_2)f, h_{1,2} = g(P_2)(1-f), h_{2,1} = g(P_1)f, h_{2,2} = g(P_1)(1-f), g(x) = \frac{\alpha}{1+(x/p)^H}$, where p is the half-max concentration, α is the full induction and H the Hill coefficient.

We also considered the reduced models, that consist only of the TX, TL and degradation reactions. To this end, we replaced the translation rate constants by one of the following time- and history-dependent filters, the Hawkes and binomial filter. The Hawkes model reads

$$\mathcal{M}_{\text{L}}^P: \quad
\begin{aligned}
R_1, R_2 \quad & : \emptyset \overset{g(P_2)}{\rightarrow} Z_1 \overset{\delta_{\text{TX}}}{\rightarrow} \emptyset \\
R_3 \quad & : Z_1 \overset{\lambda_{\text{L},1}(t)}{\rightarrow} Z_1 + \text{Prot}_1 \\
R_4 \quad & : \text{Prot}_1 \overset{\delta_{\text{TL}}}{\rightarrow} \emptyset \\
R_5, R_6 \quad & : \emptyset \overset{g(P_1)}{\rightarrow} Z_2 \overset{\delta_{\text{TX}}}{\rightarrow} \emptyset \\
R_7 \quad & : Z_2 \overset{\lambda_{\text{L},2}(t)}{\rightarrow} Z_2 + \text{Prot}_2 \\
R_8 \quad & : \text{Prot}_2 \overset{\delta_{\text{TL}}}{\rightarrow} \emptyset
\end{aligned}$$

where $\lambda_{\text{L},i}(t) = \hat{Z}^{\text{H}}$, as in Eq. (205) with updates of $N(t)$ as in the first paragraph of section 8.1.7. For the Hawkes filter we used the non-equilibrated version, since with the repeated modification of N we cannot expect the filter to be at equilibrium, but assume that it spends a non-negligible amount of time in the transient regime. For the binomial model $\mathcal{M}_{\text{B}}^P$, the reactions R_3, R_7 obtain altered propensities $\lambda_{\text{B},i}(t)$ as in Eq. (203) with updates of $N(t)$ as in the first paragraph of section 8.1.7.

The switching parameters and heterogeneous translation rates are accounted for by the filter. In the case study, the parameters were chosen to match the mean translation rate of the extended model with the one of the initial model as described in section 8.1.3 after the model introduction. We emphasize that the extension does not alter the reaction types, in particular, it does not increase the number of reaction types. Due to the model reduction via filtering, the switching reactions are not explicitly simulated. Their effect enters implicitly via corresponding history-dependent propensities, that rely on tracking a sufficient statistic, as opposed to depending on the current network state. For this reason, a stochastic simulation algorithm that handles time-varying propensities is required, which excludes the convenient Doob-Gillespie algorithm. Accordingly, the results were obtained by our algorithm presented in section 7.1 for the Hawkes model and a modified algorithm 1 for the binomial model.

The stochastic toggle switch can operate in three different regimes. Here, we looked at the two bistable regimes with (i) stable modes and (ii) unstable modes. The third monostable regime, which is characterized by a unimodal joint distribution of the two protein species, is not of interest here. The bistable regime is characterized by a bimodal joint distribution. For (i), transitions between the
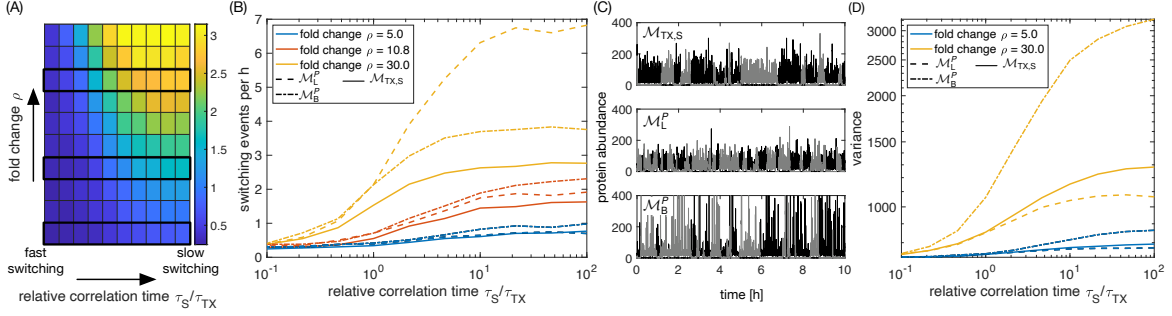
Figure 12: Symmetric toggle switch with conformational mRNA switching. (A) Heatmap showing the average number of switching events per hour for varying fold change and relative correlation time. Fold change increases from bottom ($\rho = 5$) to top ($\rho = 50$), relative correlation time increases from left ($\tau_S/\tau_{TX} = 0.1$) to right ($\tau_S/\tau_{TX} = 100$) with logarithmic axes. Estimated values from simulation of the reaction system $\mathcal{M}_S$. (B) The average number of switching events as a function of the relative correlation time for three different fold changes (boxed by black in A). (C) Sample trajectories for the first (black) and second (gray) protein species, $\rho = 30.0, \tau_S/\tau_{TX} = 100$. Exact and reduced models are indicated. (D) Stationary variance of the protein abundance of the first protein species as a function of the relative correlation time for two different fold changes ($\rho = 10.8$ is not shown to avoid cluttering). In B and D, exact model $\mathcal{M}_S$ (solid line, values as in A) and reduced models $\mathcal{M}_L^P$ (dashed), $\mathcal{M}_B^P$ (dotted) are compared. Parameters were $\alpha = 1s^{-1}, H = 5, p = 10, \delta_{TX} = 0.2s^{-1}, \delta_{TL} = 0.1s^{-1}$. Translation rates r_1, r_2 and transition rates $c_{1,2}, c_{2,1}$ between structures were chosen such that $fr_1 + (1-f)r_2 = 1s^{-1}$ with $f = (\rho + 1)^{-1}$, and according to values of ρ as well as $\tau_S/\tau_{TX} = \delta_{TX}/(c_{1,2} + c_{2,1})$. To estimate the values shown in A, B and D, 100 trajectories were drawn, initialized in $(Z_{1,1}, Z_{1,2}, \text{Prot}_1, Z_{2,1}, Z_{2,2}, \text{Prot}_2) = (5, 5, 0, 5, 5, 0)$ for the exact model and $(Z_1, \text{Prot}_1, Z_2, \text{Prot}_2) = (10, 0, 10, 0)$ for the reduced models. Simulations were terminated after $500, 000$ reactions that changed Prot_1 or Prot_2.

two modes are rare, and the bimodality is a result of the initial transition to either mode. For (ii), we observed switching between the modes in the time course of a trajectory. A symmetric toggle switch, which we study here, allocates both modes with equal probability. We expected that a heterogeneous translation due to mRNA structure switching can alter the regime in which the toggle switch operates. We conjectured that the stable modes of a bistable toggle switch can be destabilized. As we have seen, the protein variance increases upon the introduction of heterogeneous mRNA structures. This broadens the distribution of each mode, which increases the probability of transitions from one mode to the other. Indeed, we found parameter choices for the Hill coefficient, the saturation values and the degradation rates which exhibit a destabilizing behavior. As in section 8.1.3, the fold change between translation rates is the ratio r_1/r_2. We investigated how the frequency of transitions varies for different fold changes and timescales of mRNA conformational switching.

For the estimation of the number of switching events, 100 trajectories were simulated, until $500, 000$ jumps had occurred in reactions that altered Prot_1 or Prot_2. The system state at the jump times yielded a time series. The time series was smoothed by a moving average with window length 100, and the sign changes of the difference between the smoothed protein abundances were counted. Each of the 100 trajectories yielded one total number of sign changes. We added all total numbers and divided by the sum of elapsed time over all 100 trajectories. The stationary variance of Prot_1 was estimated as a time average over the appended trajectories. By the symmetry of the system in Prot_1 and Prot_2, we appended all trajectories of Prot_2, even though both are clearly not independent.

Consistently, we found that the quasi-steady state assumption holds for fast conformational switching. We observed that the switch switches more frequently with increased fold change for slow enough conformational switching (figure 12 A). This behavior is qualitatively captured by the reduced Hawkes and binomial models (figure 12 B). However, quantitatively, the reduced models deviate, overestimating the number of switching events, compare also the sample trajectories (figure 12 C). Because of the non-linear propensities for the repression, realized by decreasing Hill functions, we cannot expect the protein variance of the Hawkes model to agree with the exact model. We observed a lower variance

(figure 12 D). At the same time, the effect of overdispersion for the binomial model persists, compare section 8.1.5. The switching behavior for the Hawkes and binomial models shows an interesting non-monotonic relation (figure 12 B). While for an intermediate fold change $\rho = 10.8$ the Hawkes model switches less than the binomial model, for a high fold change $\rho = 30.0$ the Hawkes model switches more. We explain this by the following observation. Due to the constraints of its linear approximation, the Hawkes model spends an overestimated amount of time in intermediate propensity values, whereas the binomial model leans away from the central propensity values towards the high and low propensity values. For larger fold changes, in particular, this brings the two modes closer together for the Hawkes model, whereas it broadens the modes for the binomial model. This explains not only the underestimated variance for the Hawkes model, but also implies that the Hawkes model has more overlaps between the modes, which makes transitions more likely. For the binomial model, on the contrary, the more likely transitions result from increased allocation of low propensity values. This example shows that quantitative deductions from approximate stochastic filtering, as presented here, need to be treated carefully when a CRN contains non-linear propensities. Not only the monotonic agreement of the approximate filter's propensity with Snyder's filter matter, but also the proportion with which the filter allocates the propensity values.

8.2 Hawkes modeling

For the inclusion of structure switching in the previous case study, we observed cases where the Hawkes model and the exact joint model agree, as well as cases with a discrepancy. Based on these observations, we investigated the Hawkes approximation in more detail. We used the ACID as a discrimination tool and compared the Hawkes approximate filter, Eq. (208), to the Snyder filter for the random telegraph model with a dark current. Formally, the model is equivalent to the promoter-mediated transcription model or mRNA-structure-mediated translation model with $k = 2$ and $N = 1$.

We first provide the technical details of the computational approach for the ACID that uses sections 5.2.1 and 5.2.4. For this purpose, the Eq. (58) provides the one-dimensional sufficient state variable of Markovian progression with linear $F(z) = -\alpha(z - \mu_0)$, $f(z) = z + \beta$ and $l = \mathrm{id}$. The CI $\hat{\lambda}_t$ is the left-continuous version of $\hat{Z}(t)$ in Eq. (58). The link to the BReT-P is established by Eq. (84) and (85)

$$m(\tau, \theta) = u(\tau, \theta) = \mu_0 + e^{-\alpha\tau}(\theta - \mu_0), \; g(\tau, \theta) = m(\tau, \theta) + \beta. \tag{213}$$

The next proposition parallels the proposition 9.6. In the Eq. (215) below we also provide the analogue of Eq. (286), i.e. the kernel for the direct method. The support of $p_0(\theta)$ is $(\mu_0 + \beta, \infty)$ and contained in (μ_0, ∞). We consider the equidistant partition $b_i = \mu_0 + i\Delta\theta, i = 0, \ldots, N$ for $\Delta\theta \cdot N$ large enough to cover most of the probability weight, i.e., for

$$\int_0^{\Delta\theta \cdot N} p_0(\theta + \mu_0 + \beta) \, \mathrm{d}\theta \approx \int_0^\infty p_0(\theta + \mu_0 + \beta) \, \mathrm{d}\theta.$$

Proposition 8.1. *Suppose that we partition the truncated $p_0(\theta)$-support $\Omega = (\mu_0 + \beta, \mu_0 + \beta + \Delta\theta \cdot N]$ into equidistant intervals $(b_{i-1}, b_i]$ with $b_i = \mu_0 + \beta + i \cdot \Delta\theta, i = 1, \ldots, N$ and choose representatives $\theta_i = \frac{b_i + b_{i-1}}{2}$. Then the matrix entries in (125) are given by*

$$A_{i,j} = R(b_i, \theta_j) \wedge 1 - R(b_{i-1}, \theta_j) \wedge 1$$

with

$$R(\theta, \theta') = \left(\frac{\theta - \beta - \mu_0}{\theta' - \mu_0} \right)^{\frac{\mu_0}{\alpha}} e^{-\alpha^{-1}(\theta' - \theta + \beta)}.$$

Proof. Continuing with Eqs. (213), we solve the Eq. (91) to obtain

$$\rho(\tau, \theta) = e^{-\mu_0\tau} \exp\left(\frac{\theta - \mu_0}{\alpha}(e^{-\alpha\tau} - 1) \right).$$

Define the function

$$T(\theta, \theta') := \alpha^{-1} \ln\left(\frac{\theta' - \mu_0}{\theta - \mu_0} \right). \tag{214}$$

The number of solutions of $g(\tau, \theta') = \theta$ is $N(\theta, \theta') \in \{0, 1\}$ with $\tau(\theta, \theta') = T(\theta - \beta, \theta')$. Then the matrix entries in Eq. (125) are given by

$$A_{i,j} = \rho(\tau(b_i, \theta_j) \vee 0, \theta_j) - \rho(\tau(b_{i-1}, \theta_j) \vee 0, \theta_j)$$
$$= R(b_i, \theta_j) \wedge 1 - R(b_{i-1}, \theta_j) \wedge 1.$$

$\square$

The ACID was obtained from a fixed point iteration as described in 5.2.1 and, according to section 5.2.4, we employed $\tau^{(m)}(\theta, \theta') = T(\theta, \theta')$, see Eq. (214), with numerical integration in Eq. (132). Examples are shown in fig. 13. The other way to obtain the ACID via the direct method in Eq. (134) yields for the Hawkes process

$$K(m, m') = \frac{m'(m - \mu_0)^{\frac{\mu_0}{\alpha} - 1}}{\alpha(m' + \beta - \mu_0)^{-\frac{\mu_0}{\alpha}}} e^{\frac{1}{\alpha}(m - m' - \beta)} \tag{215}$$

in Eq. (133). By "iterative solution" [178, p.2] this fixed point iteration could have originally been meant instead of the method of steps. Fig. 13 shows agreement for three Hawkes examples.

The approximate $p_\lambda(m)$ may serve as initial distribution of $\hat{\lambda}_0$ for the stationary Hawkes process. Using martingale theory, the equilibrium variance of $\hat{\lambda}_t$ was derived to equal

$$\mathbb{Var}[\hat{\lambda}_\infty] = \frac{\alpha \mu_0 \beta^2}{2(\alpha - \beta)^2}. \tag{216}$$

For this purpose the process equation for the Hawkes process is rewritten as

$$d\hat{\lambda}_t = -\gamma(\hat{\lambda}_t - c\mu) \, dt + \beta \, dQ_t$$

with the canonical $\mathcal{F}_t^Y$-martingale increment $dQ_t = dY_t - \hat{\lambda}_t \, dt$. The parameters are linked as derived in Eq. (68) with the correspondence $\bar{\lambda} = \mu_0, \bar{B}D = \beta, A = \gamma$, i.e.,

$$\beta = \sqrt{\gamma^2 + \frac{2c\gamma\sigma^2}{\mu}} - \gamma, \ \alpha = \gamma + \beta, \ \alpha\mu_0 = c\mu\gamma, \tag{217}$$

see also Eq. (208). We are interested in the asymptotic behavior, i.e. $\hat{\lambda}_\infty$. It can be brought to finite time t under the shift [60] of the time domain $[0, \infty)$ to $(-\infty, t)$:

$$\hat{\lambda}_t = c\mu + \int_{-\infty}^t e^{-\gamma(t-s)} \beta \, dQ_s. \tag{218}$$

By the Ito isometry for counting processes, Eq. (5), we get Eq. (216) from Eq. (218) via:

$$\mathbb{Var}[\hat{\lambda}_t] = \mathbb{E}[(\hat{\lambda}_t - c\mu)^2] = \mathbb{E}\left[\left(\int_{-\infty}^t e^{-\gamma(t-s)} \beta \, dQ_s\right)^2\right] = \int_{-\infty}^t e^{-2\gamma(t-s)} \beta^2 \mathbb{E}[\hat{\lambda}_s] \, ds$$

$$= \frac{c\beta^2\mu}{2\gamma} = c^2\sigma^2 - c\mu\beta = \mathbb{Var}[\lambda_t] - \mathbb{E}[\lambda_t]\beta. \tag{219}$$

In section 8.2.4 below, we take Eq. (219) as a point of departure to define a relative effective noise that is transferred from the environment to the subnetwork. The parameter sets in fig. 13 were chosen, such that the ACID's first and second order moments are constant, but vary in the exponential decay parameter α of the Hawkes kernel. This makes the shape vary qualitatively. While, for fast decay (α large), the region near the base value μ_0 is frequented more heavily, for slow decay, the CI spends more time in the middle regime around the mean $c\mu$. This illustrates that the ACID analysis goes beyond the mean and variance analysis, i.e., that the ACID is parameterized by more than two parameters.

Let us return to the perspective of the previous subsection, where we saw the Hawkes process as an approximation of the MMPP modulated by $(Z_t)_{t \geq 0}$. We emphasize that the stationary distributions, i.e., the distribution of λ_∞ for common input processes $(Z_t)_{t \geq 0}$, such as the Cox-Ingersoll-Ross process,
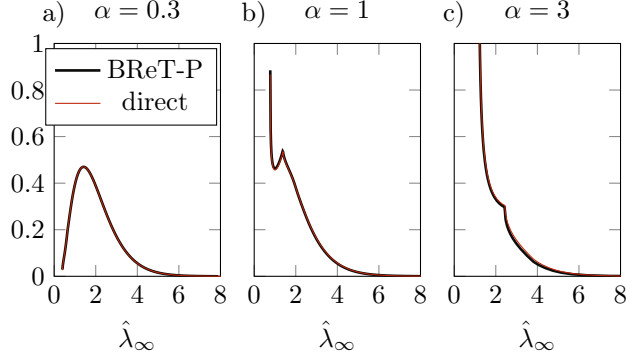
111

Figure 13: **ACID for the Hawkes process.** ACID mean $c\mu = \frac{\alpha\mu_0}{\alpha-\beta} = 2$ and ACID variance $\frac{\alpha\mu_0\beta^2}{2(\alpha-\beta)^2} = 1$ were constant, while $\alpha \in \{0.3, 1, 3\}$ varied. The truncation $\Delta\theta \cdot N$ of the support was chosen to be the 0.999-quantile of the Gamma distribution with mean $c\mu$ and variance $c^2\sigma^2$. Discretization granularity was $N = 200$. Additionally, three equidistant representatives θ_i, see Eq. (124) were chosen in each interval. Their mean function evaluations were used in Eq. (125) for the coefficients of A. Number of iterations was $L = 15$. Comparison of BReT-P and direct method is shown, where the discretization granularity was $N = 1000$ for the direct method.

see case study 8.2.3 below, the birth-death process or random telegraph model, are entirely character-ized by their mean and their variance. The autocovariance decay parameter γ, i.e., A in Eq. (C2), is not captured by λ_∞. In contrast, the ACID $\hat{\lambda}_\infty$ does capture a change in γ, see fig. 13, so it contains temporal information about the input process. Next we go into detail about the comparison of the MMPP and its Hawkes approximation with the ACID and the RER. This exceeds the second-order analysis, which by theorem 4.1 fails to tell them apart.

8.2.1 ACID discriminates Snyder and Hawkes filter

In the promoter-mediated transcription model with three structures, case study 8.1.1, we observed that the Hawkes model does not capture the trajectory-wise non-linear Riccati dynamics of the Snyder filter, see figure 7G. While the trajectories provide an empirical comparison of the dynamically evolving conditional intensities. The information-measure RER, on the contrary, can assess the deviation of the path measures. Alternatively, we use the ACID for an ensemble comparison of assumed conditional intensity values, that summarizes the comparison of trajectories. We compare the Hawkes filter to the exact Snyder filter, i.e.,

$$\mathrm{d}\hat{\lambda}_t = \left\{ c_1\Delta\lambda - (c_2 + c_1 + \Delta\lambda)(\hat{\lambda}_t - \lambda_0) + (\hat{\lambda}_t - \lambda_0)^2 \right\} \, \mathrm{d}t + \frac{(\hat{\lambda}_{t-} - \lambda_0)(\lambda_1 - \hat{\lambda}_{t-})}{\hat{\lambda}_{t-}} \, \mathrm{d}Y_t, \tag{220}$$

see also section 6.1.2 and information-theoretic applications 9.2 and 9.4.1 below. We fixed the dynamic range $\Delta\lambda = 1$ and varied the intensity λ_0 for the Off state, i.e., the leakage. The RER and the empirical comparison of the ACID is employed to detect the parameter regimes where the approximate marginal simulation deviates from the exact marginal simulation (fig. 14b, c). Let $\mathbb{P}$ denote the path measure of the random telegraph modulated counting process with leakage, i.e., with the CI given by the Snyder filter and $\mathbb{Q}$ the Hawkes process. Following section 5.2.3, the RER $\overline{D}(\mathbb{P}\|\mathbb{Q})$ was computed. Figure 14e) shows that for fixed switching rates, gain and amplitude, the deviation gets more severe for smaller leakage. We also computed other similarity measures, depicted in fig. 14f). The Wasserstein metric compares the ACIDs, while the metric

$$\lim_{T \to \infty} \frac{1}{T} \int_0^T |\hat{\lambda}_t - \hat{\lambda}_t^H| \, \mathrm{d}t. \tag{221}$$

implements a path-average L^1-comparison. It depends on the context, i.e., the approximation goal, when to use a comparison of the ACIDs and when to use the RER for assessing an approximation.
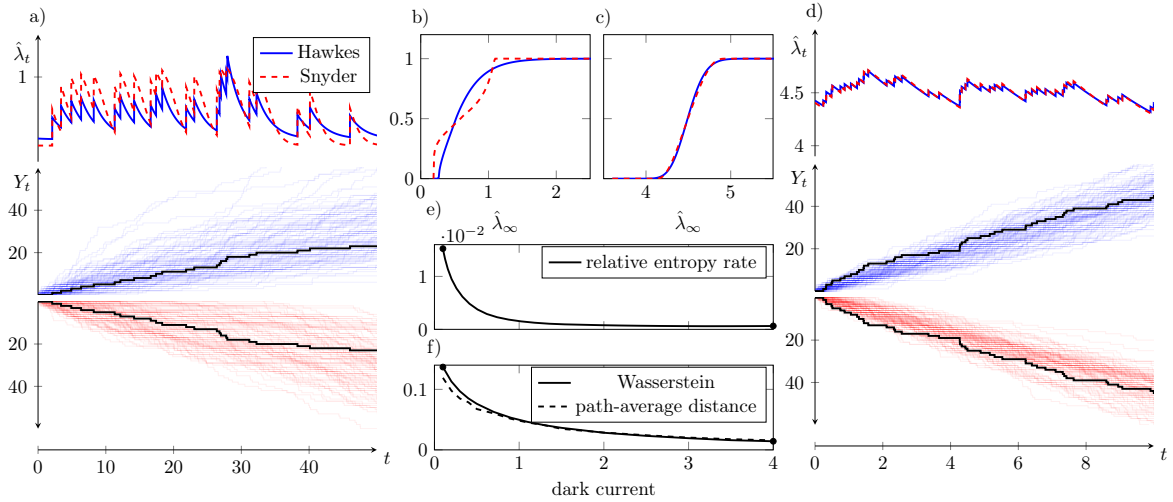
Figure 14: **The random telegraph with dark current modulates a Poisson process.** Comparison of the exact Snyder filter and the optimal linear filter (Hawkes process). a), d) Lower panel shows 100 realizations of the approximate marginal simulation (blue, increasing) and the exact simulation (red, decreasing). The lower y axis was flipped to simplify visual comparison. Upper panel compares $\hat{\lambda}_t$ for the Hawkes (blue) and the exact (red, dashed). The trajectory $t \mapsto Y_t$ for which both where computed, is shown in the lower panel (black). b), c) The cdfs of the ACIDs are compared for Hawkes (blue) and Snyder filter (red, dashed). e) The RER between the counting process obtained by the Snyder filter, i.e., the MMPP with the random telegraph input, and the Hawkes process is depicted for increasing dark current. f) The Wasserstein metric between the Hawkes ACID and the exact ACID for increasing dark current. The dashed line depicts the path metric (221) obtained from a) as the average distance between both paths. (The trajectory $t \mapsto Y_t$ was obtained from the Snyder filter.) All plots used the parameters $\Delta\lambda = 1, c_1 = c_2 = 0.1, c = 1$. Dark current varied in e), while a), b) used $\lambda_0 = 0.1$, and c), d) used $\lambda_0 = 4$. Both values are indicated as dots in e) and f).

113

When approximating the MIR via Eq. (121) with an approximate filter $\hat{\lambda}_t$, a comparison of the ACIDs seems suited. According to [85], the RER precisely quantifies the error of a mismatched computation of the MIR, i.e., if an approximate filter is used in Eq. (8), while $(Y_t)_{t\geq 0}$ is distributed according to the original counting process. Note that this differs from approximating Eq. (121) with $\hat{\lambda}_\infty$ being the ACID of the approximate filter. In the latter case, also the measure with respect to which the expectation is taken, changes (compare remark 5.2) or, in other words, $(Y_t)_{t\geq 0}$ in Eq. (8) is a counting process with the mismatched CI.

8.2.2 Mixing effects

At the end of case study 8.1.1, we observed that a higher copy number in the environment of switching structures and faster switching than synthesis could impose more agreement with the Hawkes model. Moreover, the section 8.2.1 revealed how the KL divergence decreases with faster switching and low fold change even for the extreme case of $N = 1$. In addition, we observed that the binomial model overestimated the variance of translation events with higher N, while being accurate for $N = 1$ (section 8.1.5). With its quadratic evolution equation it is able to capture the case when the intensity values are concentrated more in the upper and lower region of the interval of achievable intensity values, see figures 7F and 14a)-b). Building on these observations, we hypothesized that $\mathcal{M}_{\mathrm{M}}^P$ and $\mathcal{M}_{\mathrm{L}}^P$, see section 8.1.4, are advantageous in different regimes. The model $\mathcal{M}_{\mathrm{L}}^P$ appears to be appropriate when the intensity of the exact model $\mathcal{M}_{\mathrm{S}}^P$ allocates the intermediate regime most of the time. We conjectured that different mixing behaviors imply this condition, i.e., (i) large enough N, (ii) smaller ρ, (iii) larger k and (iv) faster τ_{S} compared to τ_{TX}. As for (iii), with one more conformation that allocates an intermediate translation rate, we would expect an additional mixing effect in terms of allocating medium intensity values. We hypothesized that $k = 3$ is disadvantageous over $k = 2$ for the multinomial compared to the Hawkes model.

In order to check this systematically, we considered a cube with the three axes k, N, ρ and the values $k \in \{2, 3\}, N \in \{2, 5\}, \rho \in \{10, 30\}$, compare figure 15A. For each of the eight combinations (the vertices of the cube), we evaluated whether $\mathcal{M}_{\mathrm{L}}^P$ or $\mathcal{M}_{\mathrm{M}}^P(\mathcal{M}_{\mathrm{B}}^P)$ was more advantageous in terms of a lower KL divergence rate. If our hypothesis of the mixing behavior was true, we would expect to find a separating hyperplane, s.t., the vertices on one side of the hyperplane would be advantageous for $\mathcal{M}_{\mathrm{L}}^P$ and the vertices on the other side would be advantageous for $\mathcal{M}_{\mathrm{M}}^P$. As a fourth parameter we varied $\tau_{\mathrm{S}}/\tau_{\mathrm{TL}}$. To elaborate our hypothesis, we included intermediate values in our study to visualize the transition between mixing and non-mixing conditions. In figure 15B, the KL divergence rates for the faces $k = 2$ and 3 spanned by (N, ρ) are shown, while in the case $k = 3$ we iterated over three network topologies. Results are shown for the fixed relative timescale $\tau_{\mathrm{S}}/\tau_{\mathrm{TL}} = 1$ and for the three-node-chain topology, see figure 16. For a faster timescale of $\tau_{\mathrm{S}}/\tau_{\mathrm{TL}} = 0.1$ and for the approximately balanced and flux network the results are shown in the figure 16.

Our analysis revealed a higher complexity than what we conjectured in the cube hypothesis. The main reasons are influences that were not captured in the abstraction with three parameters, such as the network topology in the case $k = 3$ and the timescale of conformation switching. But also among the three parameters, we found deviations from the hypothesis. While for N there was a clear opposite trend for $\mathcal{M}_{\mathrm{M}}^P$ and $\mathcal{M}_{\mathrm{L}}^P$ as hypothesized, for ρ homogeneous trends can be observed and for k effects were heterogeneous for $\mathcal{M}_{\mathrm{M}}^P$, depending on the network structure and the timescale. Thus, in the following, we elaborate on the details of these findings. In order to resolve effects only present in either $\mathcal{M}_{\mathrm{L}}^P$ or $\mathcal{M}_{\mathrm{M}}^P$, we chose to present the KL divergence rates for both, instead of their difference, which would obscure some trends.

Independent of k the multinomial model has the worst accuracy for $\rho = 30, N = 5$, the extreme values among the ones studied. Decreasing N or decreasing ρ from there improved the accuracy. In contrast, the Hawkes model exhibits an opposite trend. Its worst accuracy is found for $\rho = 30, N = 2$ while *increasing* N or decreasing ρ improves the accuracy. As expected, for slower switching the overall accuracy decreases, compare figure 16. Changes with N confirm our hypothesis: For fixed ρ, decreasing N improves the accuracy for $\mathcal{M}_{\mathrm{M}}^P$, while increasing k improves the accuracy for $\mathcal{M}_{\mathrm{L}}^P$. The larger ρ the stronger is this effect. The outlier to this trend, $\mathcal{M}_{\mathrm{L}}^P, k = 2, N = 2, 3, \tau_{\mathrm{S}}/\tau_{\mathrm{TL}} = 0.1$ can likely be attributed to a numerical approximation error.

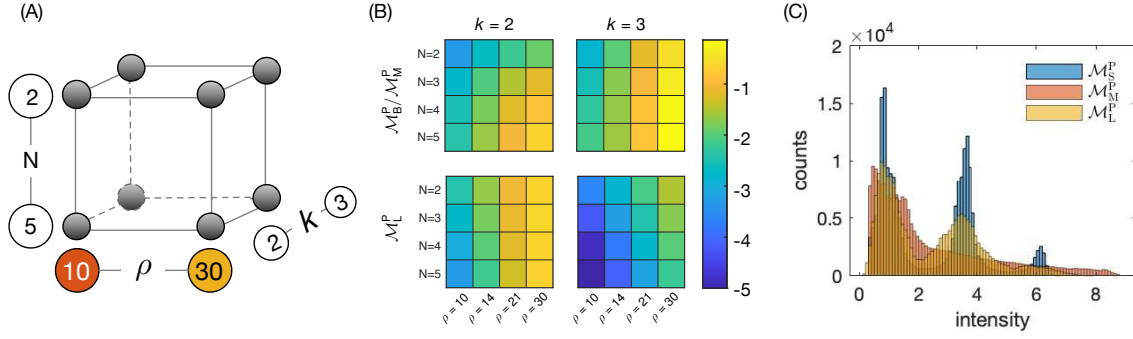For $\mathcal{M}_{\mathrm{L}}^P$, the additional mRNA conformation with intermediate translation rate ($k = 3$ instead

Figure 15: Comparison of models $\mathcal{M}_M^P$ and $\mathcal{M}_L^P$ in the context of different mixing regimes characterized by k, N, and ρ. (A) Cube with vertices corresponding to the minimum and maximum values of the three mixing [...] k [...] (logarithm [...] latter case, the underlying network topology is the three-node-chain from figure 16. In all cases, the relative timescale τ_S/τ_{TL} was set to 1. The full heatmap table for all three $k = 3$ topologies and the additional [...] intensities [...] $r_1 = 0.1, r_2 = 0.3, r_3 = 3$, (thus $\rho = 30$), and switching rates $c_{1,2} = 4 = c_{2,3}, c_{2,1} = 6, c_{3,2} = 8$ scaled so that $\tau_S/\tau_{TL} = 1$. Used with permission of Maleen Hanst.
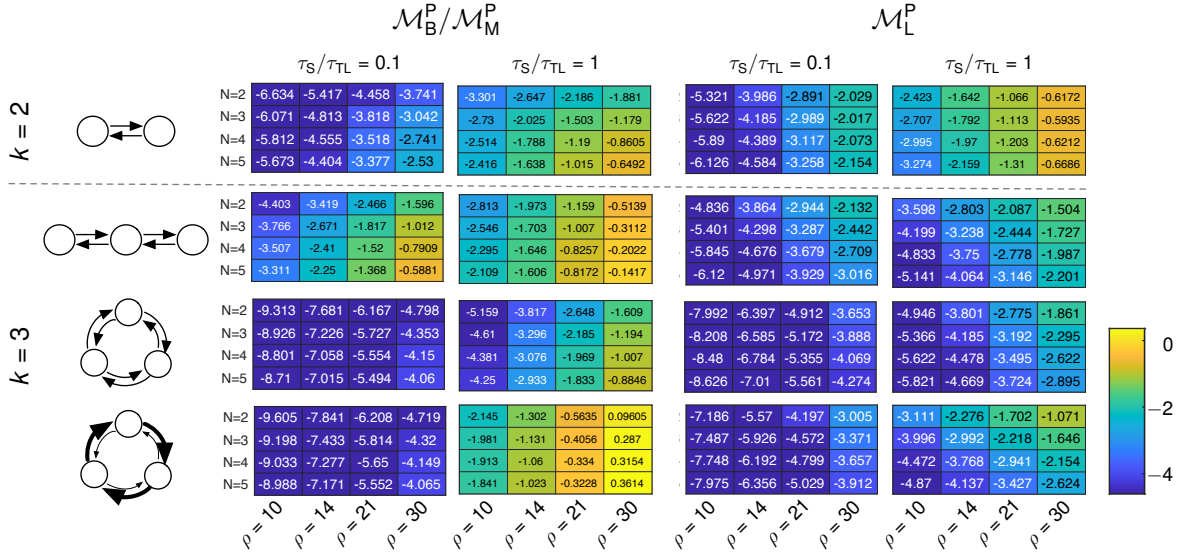

Figure 16: Full table of heatmaps: Kullback-Leibler divergence rate (logarithmic scale) for (i) the multinomial and Hawkes model, (ii) $k = 2$ and 3, and (iii) relative timescales of $\tau_S/\tau_{TL} \in \{0.1, 1\}$. Heatmaps spanned along (N, ρ). Used with permission of Maleen Hanst.

of $k = 2$) causes an the effect claimed in our hypothesis, with higher accuracy throughout the three network topologies for both timescales. In the multinomial case, the behavior is depends on the topologies and timescales. For the depicted chain network, the accuracy decreases strongly for both timescales as our hypothesis claimed. For the approximately balanced and flux network, the result is more heterogeneous and depends on the timescale as well. For fast switching, the accuracy increased for both of the two topologies. For slow switching the accuracy decreased strongly for the flux and slightly for the approximately balanced network. Contrary to our hypothesis, decreasing ρ improves the accuracy for both models. For $\mathcal{M}_L^P, k = 2$ at both timescales, the fold change ρ is the dominant factor determining the accuracy, compared to the number of mRNA molecules N.

In figure 15C, histograms of the intensity values at equidistant time points for the models $\mathcal{M}_S^P$, $\mathcal{M}_L^P$, $\mathcal{M}_B^P$ are shown. We observed three distinguished peaks in $\mathcal{M}_S^P$ corresponding to $0, 1, 2$ mRNAs in structure Z_3. Opposed to the hypothesized limitation of the Hawkes model in non-mixing cases, we see that the peaks are captured qualitatively, even though they are not peaked enough.

We conclude this section with two remarks on what makes either filter advantageous, giving interpretations that are to be tested in further experiments. First, we found that the worst overall accuracy was observed in the multinomial case for the chain and flux networks within the slow switching regime ($\tau_S/\tau_{TL} = 1$). We hypothesize that the reason for the strong deviation is the network structure which violates the independence assumption in-between molecules for the multinomial filter, but this need to be evaluated in more detail in ensuing work. Second, the Hawkes and multinomial filter have different failure modes. The Hawkes filter tends to over-allocate the intermediate regime of achievable intensity values, see figure 15C, which becomes a limiting drawback when there are not enough mixing properties, in particular for very small N. In order to achieve the same variance as the exact model, the intensity at peaks is over-estimated which compensates for over-allocating the intermediate regime, see figure 15C. For the multinomial filter, a general characteristic is that it is too steep in the intermediate regime and tends to over-allocate the high and low achievable values, implying the overestimation of the variance. Similarly as for the Hawkes model, we observed that it overestimated at the intensity peaks.

8.2.3 Hawkes versus Gamma filter

The Gamma filter [3, 81] is an approximate filter obtained from conditional moment closure. It departs from a doubly stochastic Poisson process $(Y_t)_{t \geq 0}$ that is modulated by a Cox-Ingersoll-Ross (CIR) process [57]. As introduced in section 6.3, the CIR process is a Markov process with stationary CIR-mean μ and CIR-autocovariance function $\sigma^2 e^{-\gamma t}$, i.e., $Z_t \sim \text{CIR}(\mu, \sigma^2, \gamma)$ turning $(Y_t)_{t \geq 0}$ into an MMPP for which we assume $\lambda_t = c Z_t$. The Gamma filter is the assumed density filter, where $Z_t | Y_{[0,t]}$ is assumed to be Gamma distributed. With two degrees of freedom for the Gamma distribution, two equations sufficiently describe the filter. One governs the mean, one governs the variance. Expressing the Gamma's third centered moment in terms of mean and variance justifies the replacement $\mathbb{E}[(Z_t - \mu)^3 | \mathcal{F}_t^Y] = \frac{2\mathbb{E}[(Z_t - \mu)^2 | \mathcal{F}_t^Y]^2}{\mathbb{E}[Z_t | \mathcal{F}_t^Y]}$. Consider an approximate marginal simulation of $(Y_t)_{t \geq 0}$ that uses the Gamma filter. This yields a self-exciting process $(Y_t^G)_{t \geq 0}$ with (M_t, S_t) mimicking $(\mathbb{E}[Z_t | \mathcal{F}_t^Y], \mathbb{E}[(Z_t - \mu)^2 | \mathcal{F}_t^Y])$, the CI $\hat{\lambda}_t^G = c M_{t-}$ and

$$dM_t = \{-\gamma(M_t - \mu) - cS_t\} \, dt + \frac{S_{t-}}{M_{t-}} \, dY_t^G$$

$$dS_t = \{-2\gamma(S_t - \frac{\sigma^2}{\mu} M_t) - 2c\frac{S_t^2}{M_t}\} dt + \frac{S_{t-}^2}{M_{t-}^2} \, dY_t^G. \tag{222}$$

It is instructive to contrast the approximate marginal simulation of $(Y_t)_{t \geq 0}$ with the estimation of $(Z_t)_{t \geq 0}$. Both perspectives can make use of the equations (222). In approximate marginal simulation $(Y_t^G)_{t \geq 0}$ is by definition self-exciting with $\hat{\lambda}_t^G = c M_{t-}$, while in estimation Y_t^G in Eq. (222) is replaced by the original $(Y_t)_{t \geq 0}$, which is self-exciting with intractable CI $\hat{\lambda}_t = c\mathbb{E}[Z_t | \mathcal{F}_t^Y]$. We proceed with $(Y_t^G)_{t \geq 0}$, i.e., the approximate marginal simulation. While the CI $\hat{\lambda}_t^G$ alone is not of Markovian progression, the joint (M_t, S_t) is, yielding $\dim(Y_t)_{t \geq 0} = n_0 = 2$. With $\theta(t) = (M_t, S_t)$ at jump times, the sufficient statistic can be defined.

The Gamma filter and the Hawkes filter were obtained from the CIR modulated Poisson process with a different ansatz. The ACID is now employed to discriminate between approximate marginal simulations with either one, indicating its limitations as a discrimination tool for approximate filters. It, however, informs a decision, in which parameter regime to replace the Gamma by the optimal linear filter when approximately computing the MIR. Moreover, the Gamma filter is used to illustrate the method for $\dim(Y_t^G)_{t \geq 0} = 2$.

The Gamma ACID was computed using sections 5.2.1 and 5.2.4. The (m, s)-plane was truncated in a way to respect the minimal value of m in the progression Eq. (222) and to cover $99, 5\%$ of the probability mass of a Gamma distribution with mean $c\mu$ and $c^2\sigma^2$. The bounds of the auxiliary s were dictated by the minimum and maximum in Eq. (222) for the above determined range of m. The

rectangular (m, s)-domain was partitioned into 100×50 congruent rectangles. Denote the boundaries of the rectangles by b_i^m and b_i^s, respectively. Similar to the Hawkes ACID, 3×3 equally spaced representatives θ_i, see Eq. (124), were chosen in each rectangle. Their mean function evaluations were used in Eq. (125) for the coefficients of A. As for the DOnSOff numerical approximation, the times $\tau_k(\theta'_j)$ in (125) that satisfy

$$g_1(\tau, \theta'_j) = b_i^m \quad \text{or} \quad g_2(\tau, \theta'_j) = b_i^s \tag{223}$$

were found by evolving the ODE system (91), (222) and checking for the event (223).

Consider the MMPP $(Y_t)_{t \geq 0}$ whose external signal is a CIR-process, i.e., satisfies Eq. (C1). The Hawkes process seen as optimal linear filter and the Gamma filter both approximate the CI of $(Y_t)_{t \geq 0}$. We compare them for the same parameters μ, σ^2, γ, c. First, we inspect the asymptotic mean slope and variance slope for the Gamma filter.

Proposition 8.2. *For $(Y_t)_{t \geq 0}$ the CIR-modulated Poisson process and $(Y_t^H)_{t \geq 0}$ the corresponding Hawkes process and $(Y_t^G)_{t \geq 0}$ the approximation via the Gamma filter it holds*

$$\lim_{t \to \infty} \frac{1}{t} \mathbb{V}\mathrm{ar}[Y_t^G] = \lim_{t \to \infty} \frac{1}{t} \mathbb{V}\mathrm{ar}[Y_t] = \lim_{t \to \infty} \frac{1}{t} \mathbb{V}\mathrm{ar}[Y_t^H].$$

Proof. By application of Eq. (112) for $\Theta_t = M_t$ and $V_t = S_t$ we obtain that

$$\lim_{t \to \infty} \frac{1}{t} \mathbb{V}\mathrm{ar}[Y_t^G] = c\mu + \frac{2c^2(\mathbb{E}[S_\infty] + \mathbb{V}\mathrm{ar}[M_\infty])}{\gamma}. \tag{224}$$

Furthermore, the Gamma filter satisfies the variance decomposition

$$\mathbb{E}[S_\infty] + \mathbb{V}\mathrm{ar}[M_\infty] = \sigma^2, \tag{225}$$

by the following derivation. The Ito rule, proposition 4.8, yields the evolution equation of M_t^2 from Eq. (108)

$$\mathrm{d}M_t^2 = \{-2\gamma M_t(M_t - \mu) - 2cM_tS_t\} \, \mathrm{d}t + (M_t + \frac{S_t}{M_t})^2 - M_t^2 \, \mathrm{d}Y_t. \tag{226}$$

By applying the $\mathbb{E}$ and $\mathrm{d}Y_t - cM_{t-} \, \mathrm{d}t = \mathrm{d}Q_t$ we get

$$c\mathbb{E}\left[\frac{S_t^2}{M_t}\right] = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[M_t^2] + 2\gamma\mathbb{E}[M^2] - 2\gamma\mu^2 = \frac{\mathrm{d}}{\mathrm{d}t}\mathbb{V}\mathrm{ar}[M_t] + 2\gamma\mathbb{V}\mathrm{ar}[M_t].$$

Taking the expectation in Eq. (226) and Eq. (222) yields

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{E}[M_t^2] - \mu^2) = c\mathbb{E}\left[\frac{S_t^2}{M_t}\right] - 2\gamma(\mathbb{E}[M_t^2] - \mu^2)$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[S_t] = -2\gamma\mathbb{E}[M_t^2] + 2\gamma\sigma^2 - c\mathbb{E}\left[\frac{S_t^2}{M_t}\right].$$

So the derivative of the sum evolves as

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathbb{E}[S_t] + \mathbb{V}\mathrm{ar}[M_t]) = -2\gamma(\mathbb{E}[S_t] + \mathbb{V}\mathrm{ar}[M_t]) + 2\gamma\sigma^2.$$

Since the sum starts in the steady state $\mathbb{E}[S_0] + \mathbb{V}\mathrm{ar}[M_0] = \sigma^2$, it stays constant for all t and in particular in the asymptotic. As a consequence of Eq. (225) plugged into Eq. (224), the Gamma filter agrees with the exact CIR-modulated Poisson process in asymptotic first and second order moment, like the Hawkes process did. $\square$

First and second order analysis cannot tell the counting processes obtained from the Hawkes and the Gamma filter apart. A comparison of their ACIDs shows that they are remarkably similar. The cdfs of their ACIDs approximately agree for a range of values σ^2, γ, while the mean and gain were kept constant $\mu = 2, c = 1$. The Wasserstein metric was computed for the regime $(\sigma^2, \gamma) \in \{0.05, 0.1, \dot{2}\} \times \{0.2, 0.4, \ldots, 4\}$. For the Gamma (Hawkes) filter, the numerical method yielded $93,0\%$ ($98,5\%$) of
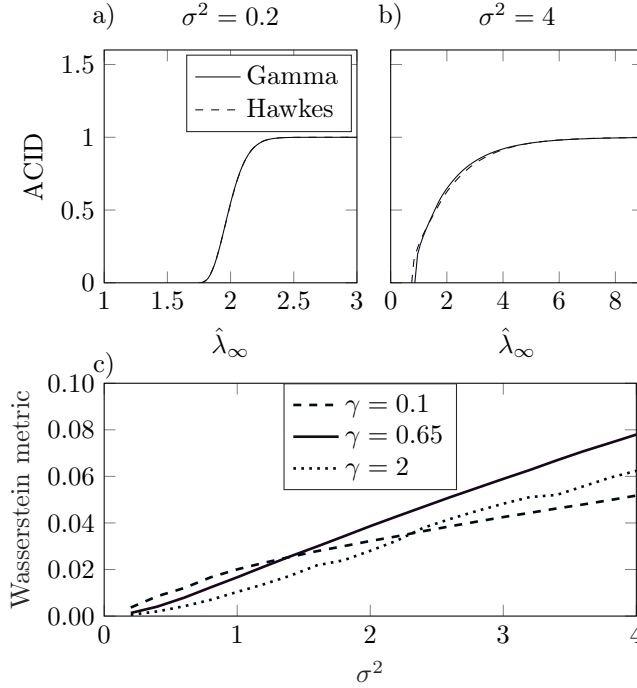
Figure 17: **Comparison of the ACID for Gamma and Hawkes.** a) and b) show the cdf of the ACID for the Gamma filter and Hawkes process for $\mu = 2, \gamma = 0.65$ and different values of σ^2. c) The Wasserstein metric was computed for a range of γ, σ^2, while $\mu = 2, c = 1$ were fixed. In the examined regime γ has little effect, while the difference measure slightly increases with growing σ^2. The largest deviation for $\gamma = 0.65, \sigma^2 = 4$ was depicted in b) still showing agreement of the ACIDs.

ACIDs that had mean value less than 0.01 from the true value μ and $40, 6\%$ $(53, 1\%)$ with a difference less than 0.001. For the Gamma filter, most outliers (deviation > 0.01) were detected for $\gamma = 0.05$ or $\sigma^2 > 4 \cdot \gamma - 0.45$. For the Hawkes filter all outliers were detected at $\gamma = 0.05$. The Wasserstein metric values ranged from 0.0005 to 0.13. When neglecting $\gamma = 0.05$ the largest value was 0.078 for $\sigma^2 = 4, \gamma = 0.65$ and a decreasing trend for decreasing σ^2 was detected, relatively independent of γ. The large values for $\gamma = 0.05$ can be explained by the numerical inaccuracy in the Gamma ACID. Exemplary graphs for $\mu = 2, \gamma = 0.65, c = 1$ and smaller vs. larger variance are depicted in fig. 17. The most prominent dissimilarity is found in base values and it becomes more pronounced for larger variance.

In summary, a comparison of the ACIDs for a range of parameters revealed a slight increase of the Wasserstein metric for increased σ^2. The parameter γ had little effect on the Wasserstein metric. Looking at the extreme case, fig. 17b depicts the example with the largest Wasserstein metric among the considered parameters, revealing that the ACIDs are still very similar. Due to the ACID's limitation as a partial characteristic, we cannot deduce that the path measures are close in some notion of distance. However, we conclude the following. When quantities are computed that only depend on the ACID, the optimal linear filter - appealing with efficient analytic expressions - might replace the Gamma filter. The MIR, Eq. (121), is such a quantity.

The MIR along the Poisson channel was efficiently approximated by Monte Carlo simulation in [83, Case study 1] via the Gamma filter. We replaced the Gamma filter by the even more efficient Hawkes filter. In more detail, for the birth-death input process $(Z_t)_{t \geq 0}$ with birth rate $\gamma \mu$ and death rate γ (i.e., mean μ and autocovariance function $\mu e^{-\gamma t}$) the gamma filter's conditional variance equation is slightly modified:

$$\mathrm{d}S_t = \{-\gamma(2S_t - M_t - \mu) - 2c\frac{S_t^2}{M_t}\}dt + \frac{S_{t-}^2}{M_{t-}^2}\,\mathrm{d}Y_t.$$
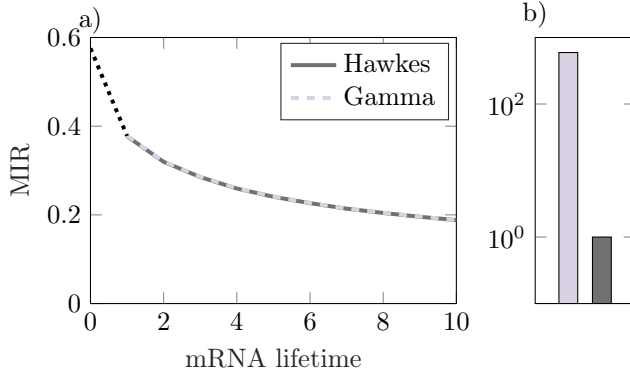
Figure 18: Comparison of the MIR $\bar{\mathbb{I}}(Z, Y)$ for an MMPP $(Y_t)_{t \geq 0}$, approximately computed with the Gamma and Hawkes filter. The input Z was a birth-death process with birth rate $\gamma\mu$ and death rate γ. In the biological context this corresponds to the gene expression model with mRNA counts Z_t and protein translation event counts Y_t. a) The x-axis shows the average mRNA lifetime γ^{-1}. Both the Gamma and Hawkes approximate MIR were computed as Monte Carlo average using [83, Eq. (16)] with hyperparameters $T = 200$, sample size $10,000$. Parameters were $\mu = 10, c = 1$ and $\gamma = 1, \ldots, 10$. The value at 0 was determined analytically via $\mathbb{E}[\phi(Z_\infty)] - \phi(\mathbb{E}[Z_\infty])$, with $Z_\infty \sim \mathrm{Pois}(\mu)$. b) shows the relative simulation time of the Gamma vs Hawkes in logarithmic scale with colors as in a).

In the limit $\gamma \to \infty$ ACID is a delta distribution at $\hat{\lambda}_\infty = \mathbb{E}[\lambda_\infty]$ and hence

$$\lim_{\gamma \to \infty} \bar{\mathbb{I}}(Z, Y) = \mathbb{E}[\phi(\lambda_\infty)] - \phi(\mathbb{E}[\lambda_\infty]).$$

Note, that both the Gamma filter and the Hawkes filter yield only an approximation of the exact MIR. Between them, the Hawkes filter can be preferred in this case with its gain in efficiency and no loss in accuracy relative to the Gamma. In what respect the replacement works for more complicated reaction networks, must be carefully evaluated.

8.2.4 Effective noise conjecture

In [3] a direct relation between the effective noise $\mathbb{Var}[\hat{\lambda}_\infty]$ and the noise of the output Y_t was conjectured. The effective noise conjecture implies that second-order information about the environment is lost in the marginal description of the subnetwork. However, the remark 4.3 shows that both the variance of the environment and the autocovariance decay are recovered by the Hawkes process, and both parameters are disentangled. Hence, at least for the modulated birth process, no second-order information is lost, when the autocovariance function is available. For linear subnetworks with zeroth-order modulation, the proposition 3.12 shows, how the second-order information of the environment is convolved in the second-order information of the subnetwork. Whether the relation can be deconvolved depends on the knowledge of Γ and B as well as the rank of $N_1 C$.

We next suppose that only the variance of the Y_t is available. Let us look at the claim that $\mathbb{Var}[\hat{\lambda}_\infty]$, i.e., the variance of the ACID, can be interpreted as the effective variance transferred to Y. In the expression (114) both the effective noise $\mathbb{Var}[\hat{\lambda}_\infty]$ and suppressed noise $\mathbb{E}[V(\infty)]$ contribute, which again rejects the conjecture. The conjecture suggested that a reduction of the environmental variance yields the variance in the subnetwork. From Eq. (33), the autocorrelation time τ is the factor that reduces the contribution of the environmental variance to the variance in the subnetwork. We asked whether, alternatively, a relative effective noise captures how the environmental variance is reduced in the subnetwork, thus providing a direct relation. If we define the relative effective noise as $\eta := \frac{\mathbb{Var}[\mathrm{ACID}]}{c^2 \sigma^2}$ then the Hawkes Fano factor, Eq. (113) and Eq. (216) allows to be rewritten as

$$\mathrm{F} = 1 + \frac{4\eta}{(1 - \eta)^2}. \tag{227}$$

Does this relation also hold for the Gamma filter?

Theorem 8.3. *Let (M_t, S_t) be the Gamma filter associated with the quadruple $(\mu, \sigma^2, \gamma, c)$ and $\hat{\lambda}_t^{\mathcal{G}} = cM_t$ the conditional intensity. Further let $\hat{\lambda}_t^{\mathcal{H}}$ be the Hawkes intensity corresponding to $(\mu, \sigma^2, \gamma, c)$. Define $\eta^{\mathcal{G}} := \frac{\mathbb{V}\mathrm{ar}[\hat{\lambda}_\infty^{\mathcal{G}}]}{c^2\sigma^2}$ and $\eta^{\mathcal{H}} := \frac{\mathbb{V}\mathrm{ar}[\hat{\lambda}_\infty^{\mathcal{H}}]}{c^2\sigma^2}$. Then*

$$\eta^{\mathcal{G}} > \eta^{\mathcal{H}}.$$

Proof. Considering the rescaling $(\frac{\mu}{c}, \frac{\sigma^2}{c^2}, \gamma, 1)$ we may assume w.l.o.g $c = 1$. From Eq. (222) we obtain

$$0 = -2\gamma(\mathbb{E}[S_\infty] - \sigma^2) - \mathbb{E}\left[\frac{S_\infty^2}{M_\infty}\right].$$

Cauchy-Schwarz applied on $\sqrt{M_\infty}$ and $\frac{S_\infty}{\sqrt{M_\infty}}$ yields

$$\mathbb{E}\left[\frac{S_\infty^2}{M_\infty}\right] \geq \frac{\mathbb{E}[S_\infty]^2}{\mathbb{E}[M_\infty]}.$$

Equality can be excluded: Suppose it would hold, then $\frac{S_\infty}{\sqrt{M_\infty}} = \alpha\sqrt{M_\infty}$ for α deterministic. Clearly $\alpha = \frac{S_\infty}{M_\infty}$ is not deterministic, as seen from applying Ito's formula, proposition 4.8, to Eq. (222). Taken together, we obtain

$$2\gamma(\mathbb{E}[S_\infty] - \sigma^2) + \frac{\mathbb{E}[S_\infty]^2}{\mu} < 0.$$

This implies $\mathbb{E}[S_\infty] < \mu\beta$. By the relations Eq. (225) and (219) for $c = 1$ it follows $\mathbb{V}\mathrm{ar}[\hat{\lambda}_\infty^{\mathcal{G}}] > \mathbb{V}\mathrm{ar}[\hat{\lambda}_\infty^{\mathcal{H}}]$, which implies the claim. $\qquad\square$
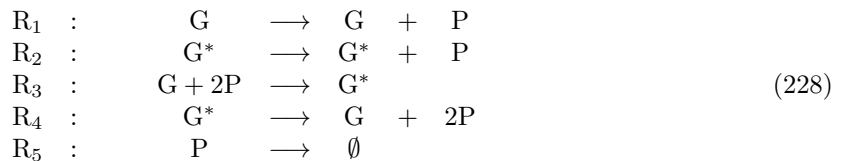
As a consequence of the theorem the relation Eq. (227) is not exact for the Gamma filter. The right hand side is strictly increasing in η. If the relation held for the Gamma filter, the statement of the theorem would contradict proposition 8.2. However, the numerical evaluation suggests that for the evaluated parameter regime the deviation is not severe.

This procedure allows for a generalization: Suppose we have an intensity process $\hat{\lambda}_t = cM_t$ that depends on some auxiliary process as in Eq. (108). The asymptotic Fano factor for such a process is obtained from Eq. (113). Then we can define $\eta := \lim_{t\to\infty} \frac{\mathbb{V}\mathrm{ar}[\hat{\lambda}_t]}{\mathbb{E}[c^2 V_t] + \mathbb{V}\mathrm{ar}[\hat{\lambda}_t]}$, appealing to the equalities (225) and (219). This is an intrinsic definition of relative effective noise that requires no environmental process with parameter σ^2. If V_t can be obtained from a low number of state variables of joint Markovian progression, the method introduced can compute the joint distribution of $(\hat{\lambda}_\infty, V_\infty)$. Both $\mathbb{V}\mathrm{ar}[\hat{\lambda}_\infty]$ and $\mathbb{E}[V_\infty]$ can be obtained from it to check how well η satisfies Eq. (227). So we hypothesize an approximate relation (227) between the relative effective noise and the noise of the output, measured by the asymptotic Fano factor. This relation can be intrinsically checked using the ACID and the corresponding asymptotic distribution of the auxiliary process.

In summary, we reject the hypothesis of the effective noise contribution, that is based on a decomposition of the variance into the contribution by the ACID and a suppressed noise The situation is more complex and depends on an interplay of the variance and autocorrelation time, Eq. (33), or the knowledge of both the variance of the ACID and the asymptotic expected conditional variance in Eq. (114). We suggested a new direct relation that is based on a definition for a relative effective noise. As shown, a relative effective noise can also be intrinsically defined for a self-exciting counting process.

8.2.5 Genetic feedback loop with fluctuating decay

As a next example we consider the following genetic feedback system studied by Holehouse et al. [49]:

$$
\begin{array}{llrcll}
\mathrm{R}_1 & : & \mathrm{G} & \longrightarrow & \mathrm{G} + \mathrm{P} & \\
\mathrm{R}_2 & : & \mathrm{G}^* & \longrightarrow & \mathrm{G}^* + \mathrm{P} & \\
\mathrm{R}_3 & : & \mathrm{G} + 2\mathrm{P} & \longrightarrow & \mathrm{G}^* & \\
\mathrm{R}_4 & : & \mathrm{G}^* & \longrightarrow & \mathrm{G} + 2\mathrm{P} & \\
\mathrm{R}_5 & : & \mathrm{P} & \longrightarrow & \emptyset &
\end{array}
\tag{228}
$$

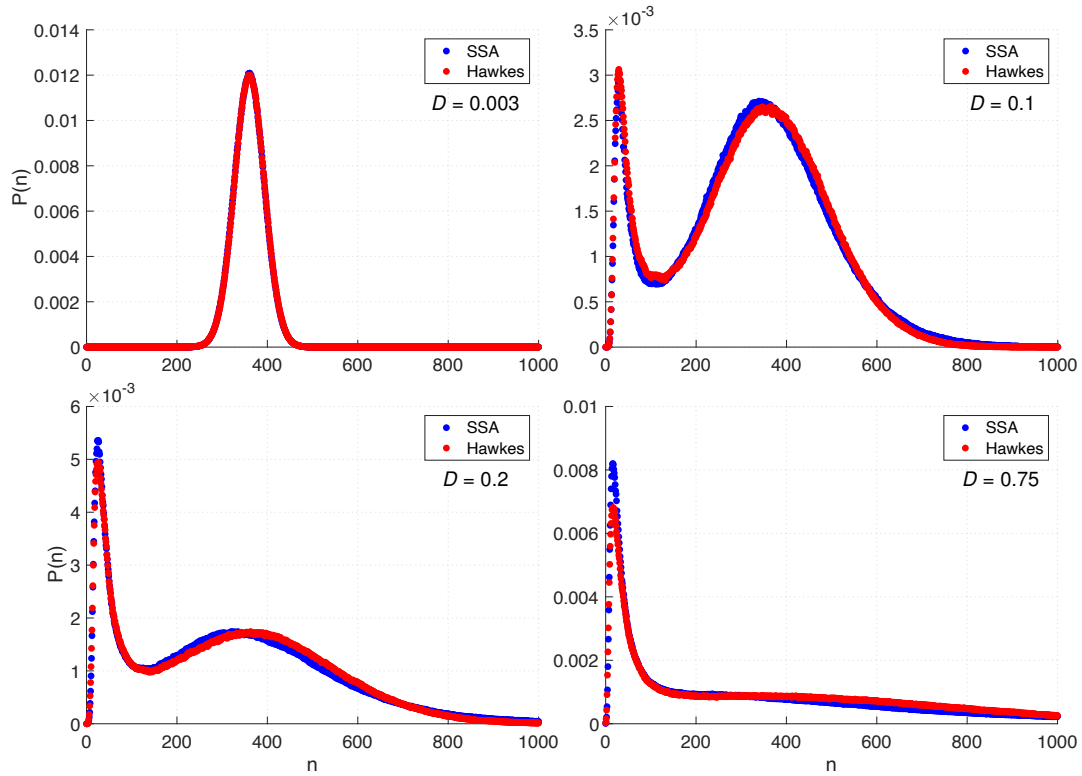with extrinsic noise in the reaction rate parameters. Reaction R_3 introduces the feedback.

Figure 19: Genetic feedback loop with fluctuating decay as specified by the reaction system Eq. (228). Environment was chosen with characteristics $\mu, \sigma^2 = D/\tau, \gamma = 1/\tau$. Parameters were $[24, 464, 1000/\Omega^2, 1000, 1]$ for the reaction rate constants, $\Omega = 200, \mu = 1, \tau = 1$ and D indicated above, as in [49, Figure 3]. Figure shows Monte Carlo estimates from 10^7 equidistant samples from one trajectory that was terminated after $4 \cdot 10^8$ reaction events. The first 10^7 reaction events were discarded to ensure the chain reached the stationary regime.

Holehouse et al. found for this system bimodal distributions in the case of a fluctuating decay rate. We observed that the proposed tilted Hawkes model captures this bimodality, see figure 19. The environment was simulated as a birth-death process, to mimic the reference model in the literature case study. There is a difference with the case study 8.1.1. In case study 8.1.1, Z was bimodal, and bimodality was transferred to X. Here, Z is not bimodal, but bimodality emerges from the topology of X under the presence of noise in the decay rate.

8.3 Birth-death process in a random environment

Being guided by the general question of how the random environment leaves a fingerprint on the ensemble statistics of the subsystem, we now focus on a minimal system to capture the fingerprint on the mean. When a random environment modulates zeroth-order reactions, the mean value is not altered compared to averaging the random environment to a reaction rate constant, see theorem 4.1. As a minimal example consider a birth-death process with birth rate λ and death rate μ whose mean approaches λ/μ in the equilibrium. By the argument for the mean of zeroth-order modulation we skip the case of a stochastic λ. If μ is stochastic, in contrast, we expect effects of the random environment on the mean, because the mean does not commute with the denominator, i.e., $\mathbb{E}[\lambda/\mu] \neq \mathbb{E}[\lambda]/\mathbb{E}[\mu]$. Additionally, the excursions during $\mu = 0$ (Fig. 3b) hint at a deviation from $\mathbb{E}[\lambda]/\mathbb{E}[\mu]$.
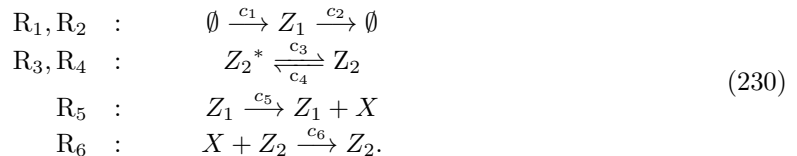
We illustrate this effect of three different stochastic environments E1 - E3 (Fig. 20a-c) on the stationary mean of a birth-death process X. Our focus is on comparing the stationary mean of the Q.SS model and ESME (theorem 3.7) to evaluate the effect of the random environment on the mean of the embedded system. ESME requires the generator and the stationary distribution of the Markov environment, as well as the reaction rate constants of the linear subsystem. The generator and reaction rate constants are specified by the model, whereas the stationary distribution can be obtained numerically, or in special cases analytically, e.g., for the class of monomolecular CRNs [154]. The modulated first-order death reaction can be seen as a bimolecular reaction. Here, we demonstrate the effect of different relative speeds between the environment Z_2 and the subspecies X. The environments E1 - E3 modulate the birth and death rates independently. The birth modulation via a birth-death process Z_1 is the same in all cases, while the complexity of the death rate modulation Z_2 increases.

8.3.1 Death modulation via random telegraph (E1)

First, we considered the scenario where the death rate is modulated by a two-state Markov process (Fig. 20a). A two-state modulation highlights the effect that the Off state Z_2^* has on X. When the instant decay rate is zero, the molecular numbers of X increase unboundedly. We call these phases excursions. We expect the stationary mean to depend on (i) the length and (ii) the frequency of excursions, because (i) the temporal average during one excursion increases with the length of the excursion, and (ii) if excursions occur more frequently, the excursion average value is weighted more strongly. The frequency can be characterized by $P_{\text{Off}} := \mathbb{P}[Z_2 = 0]$, whereas the length of excursions is proportional to the relative correlation time. The autocorrelation function of the random telegraph process, as well as the birth-death process, is of the form $e^{-t/\tau}$, where t is the time lag and τ is the correlation time. Thus, for the random telegraph process Z_2 with On and Off switching rate c_3, c_4, the correlation time is $\tau_Z = (c_3 + c_4)^{-1}$, while for the birth-death process with constant degradation c_6 it is $\tau_X = c_6^{-1}$. With these definitions, the expression of the stationary mean takes the form

$$\mathbb{E}[X_\infty] = \mathbb{E}[X_\infty^{\text{Q.SS}}] \left(1 + P_{\text{Off}} \frac{\tau_Z}{\tau_X}\right). \tag{229}$$

In the derivation, we used the ESME, Eq. (23). The birth rate is a birth-death process and the death rate is a two-state Markov process (random telegraph model), i.e., consider the CRN (see Fig. 20a.)
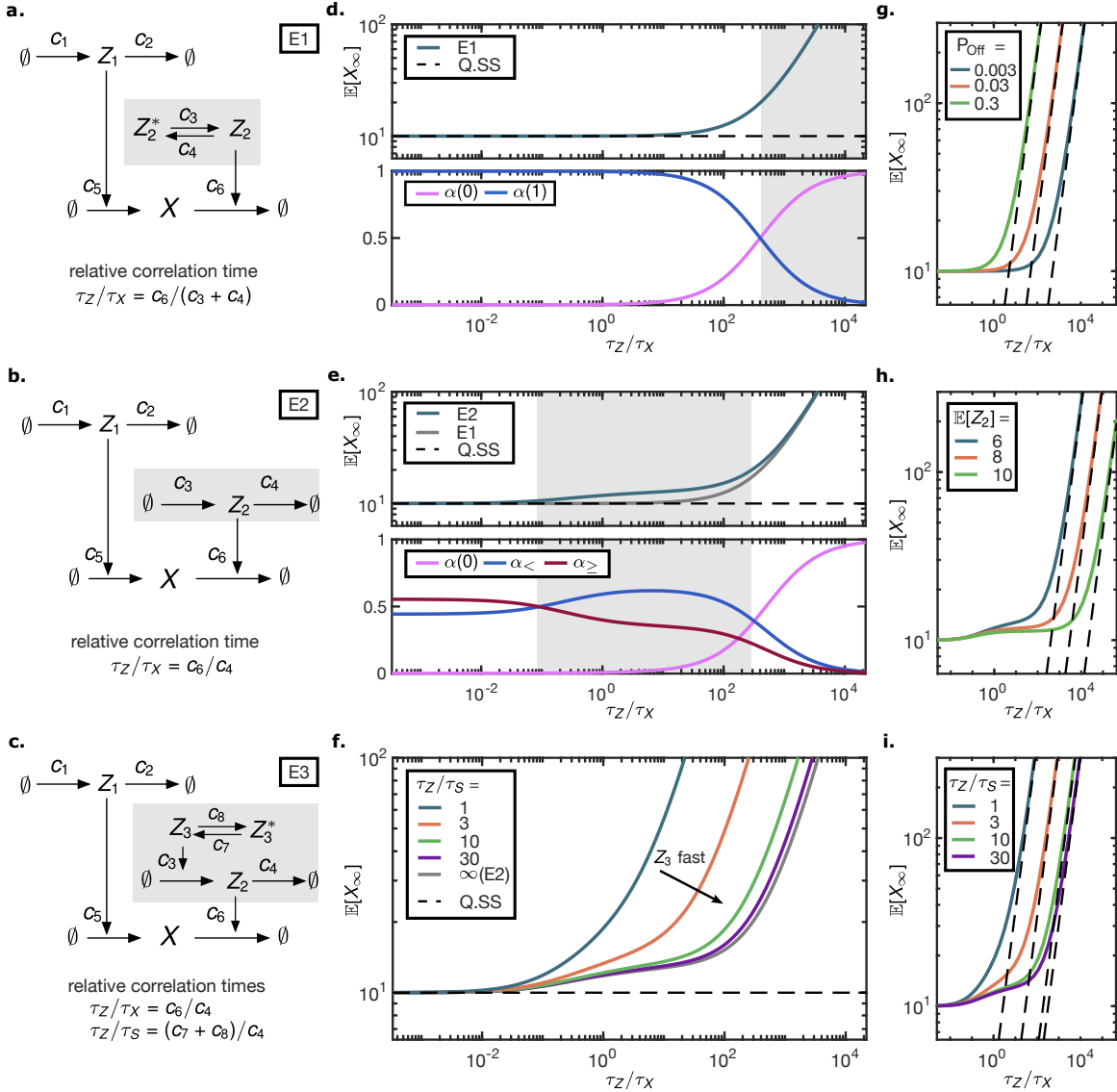
$$
\begin{array}{rcl}
R_1, R_2 & : & \emptyset \xrightarrow{c_1} Z_1 \xrightarrow{c_2} \emptyset \\
R_3, R_4 & : & Z_2^* \underset{c_4}{\overset{c_3}{\rightleftharpoons}} Z_2 \\
R_5 & : & Z_1 \xrightarrow{c_5} Z_1 + X \\
R_6 & : & X + Z_2 \xrightarrow{c_6} Z_2.
\end{array}
\tag{230}
$$

Figure 20: **a,b,c** Environment schemes for the birth-death process X. The two processes Z_1 and Z_2 modulate the birth and the death rate, respectively. **d**, **e**, **f** The stationary mean as a function of the relative correlation time systematically exceeds the Q.SS mean, to which it converges for a fast environment (valid Q.SS assumption). **d** *Lower panel*. The contributions of Z_2 to the stationary mean (E1). For a slow environment the share of $Z_2 = 0$ dominates. **e** *Upper panel*. The asymptotic, the intermediate and the degenerate regime can be distinguished (E2). The matched plot of E1, Eq. (234), shows agreement in the fast and slow regimes and a discrepancy for the intermediate regime. *Lower panel*. The contributions of $Z_2 = z$ for $z \in \{0\}, (0, \mathbb{E}[Z_2]), [\mathbb{E}[Z_2], \infty)$ to the stationary mean indicate the distinct regimes. **f** Mutable Z_2 synthesis (E3): The curves reflect different relative speeds between environmental components Z_2, Z_3. For a fast Z_3 the muting is neglectable and the environmental effect E2 from **e** is recovered. For slow Z_3 the deviation from the Q.SS mean increases. **g,h,i** The asymptotic behavior for E1 - E3 as the relative subsystem speed tends to ∞. The log-log plot demonstrates that $\mathbb{E}[X_\infty]$ increases proportional with the relative correlation time. The dashed line (proportionality constant) was computed according to section 8.3.4. Curves correspond to different values of **g** $\mathbb{P}[Z_2 = 0]$ in E1, **h** environment mean in E2, **i** relative speed of environmental components in E3. The relative correlation times $\tau_Z/\tau_X = c_6/(c_3 + c_4)$ (E1), c_6/c_4 (E2, E3) progress from slow to fast subsystem (hence slow environment) in increasing direction. All parameter choices kept the Q.SS mean $\mathbb{E}[X_\infty^{Q.SS}] = 10$ and the ratio $c_5/c_6 = 1$ fixed. For E2, E3 $\mathbb{E}[Z_2] = 8$ was chosen unless indicated otherwise, while in E1 $\mathbb{E}[Z_2]$ was calibrated to match $\mathbb{P}[Z_2 = 0]$ with E2. For E3 $\mathbb{E}[Z_3] = 0.8$ was held constant when varying $\tau_Z/\tau_S = (c_7 + c_8)/c_4$.

The stationary means of Z_1 and Z_2 are easily identified as

$$\mathbb{E}[Z_1] = \frac{c_1}{c_2}, \qquad \mathbb{E}[Z_2] = \frac{c_3}{c_3 + c_4}.$$

In this case, the dimension of the subsystem is $d = 1$. Then $A(z_1, z_2) = c_6 z_2$ and $b(z_1, z_2) = c_5 z_1$ are scalars. Since Z_1 only enters via the zero-order reactions, i.e., via b, we use the Q.SS assumption and set it to its mean. This reduces the environment to $Z = Z_2$ on the state space $\mathcal{Z} = \{0, 1\}$. The generator of the environment is

$$\Lambda = \begin{bmatrix} -c_3 & c_4 \\ c_3 & -c_4 \end{bmatrix}$$

with stationary distribution $\pi = \text{Bernoulli}(c_3/(c_3 + c_4))$. The expression of the stationary mean is then (Eq. (23))

$$\mathbb{E}[X_\infty] = \frac{c_1 c_5 (c_3 + c_4)}{c_2 c_3 c_6} \left(1 + \frac{c_6 c_4}{(c_3 + c_4)^2} \right). \tag{231}$$

In the queuing literature, [51] already derived an equivalent expression in terms of rate constants, along with other characteristics for the model, e.g., stationary distribution, using a generating function approach. The bracket term is larger than 1, yielding a systematic positive deviation compared to the Q.SS as a reference model (section 3.1.1). Note that this can be expected from the moment equations. The mean equation for X is

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[X(t)] = c_5 \mathbb{E}[Z_1(t)] - c_6 \mathbb{E}[Z_2(t) X(t)] \tag{232}$$

$$= c_5 \mathbb{E}[Z_1(t)] - c_6 \mathbb{E}[Z_2(t)] \mathbb{E}[X(t)] - c_6 \text{Cov}[Z_2(t), X(t)]. \tag{233}$$

The first two terms in Eq. (233) would yield the Q.SS dynamics. Since Z_2 and X are negatively correlated, the stationary mean is larger compared to the Q.SS mean. As Eq. (229) shows, the deviation is proportional to the separation of timescales τ_Z/τ_X and to the probability $\mathbb{P}[Z_2 = 0]$.

Figure 20d (upper panel) portrays the stationary mean $\mathbb{E}[X_\infty]$ as a function of the relative correlation time $\tau_Z/\tau_X = c_6/(c_3 + c_4)$, given that the Q.SS mean and P_{Off} are constant. Increasing the relative correlation time can be achieved by accelerating X or decelerating Z_2. Figure 20d (lower panel) shows how the share of $Z_2 = 0$ increases with the mean waiting time in $Z_2 = 0$ (or decreases with increasing speed of Z_2). In the regime $\tau_Z/\tau_X \to 0$, the stationary mean reaches the Q.SS mean, confirming that the Q.SS assumption is valid for sufficiently fast Z_2 or, equivalently, slow X.

8.3.2 Death modulation via birth-death process (E2)

We next investigated whether the generic expression Eq. (229) still holds when the death modulator Z_2 is itself a birth-death process (Fig. 20b). To this end, we altered $\tau_Z = c_4^{-1}$ and $P_{\text{Off}} = \exp(-\mathbb{E}[Z_2])$ and asked whether

$$f\left(\frac{c_6}{c_4} \right) = \mathbb{E}[X_\infty^{\text{Q.SS}}] \left(1 + \frac{c_6}{c_4} \exp(-\mathbb{E}[Z_2]) \right) \tag{234}$$

approximates $\mathbb{E}[X_\infty]$, computed via ESME. In Eq. (230), replace the two reactions R_3, R_4 by

$$R_3, R_4: \qquad \emptyset \xrightarrow{c_3} Z_2 \xrightarrow{c_4} \emptyset. \tag{235}$$

The full CRN is visualized in figure 20b. While the collection of rates A, b of the X dynamics remains the same, we make adjustments in the stationary mean state space, generator and stationary distribution of Z_2 accordingly:

$$\mathbb{E}[Z_1] = \frac{c_1}{c_2}, \qquad \mathbb{E}[Z_2] = \frac{c_3}{c_4}.$$

The standard birth-death generator on state space $\mathcal{Z} = \mathbb{N}_0$ is given by

$$\Lambda(z, z') = \begin{cases} c_3, & z = z' + 1 \\ c_4 z', & z = z' - 1 \\ -(c_3 + c_4 z'), & z = z' \end{cases} \tag{236}$$
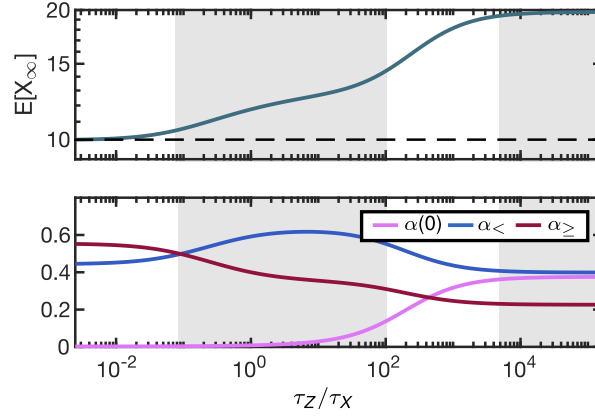
124

Figure 21: **Model E2 with leakage. Upper panel**. Stationary mean as a function of $\ln(c_4)$. The stationary mean systematically exceeds the Q.SS mean $\frac{c_1 c_5}{c_2 c_6 (c_3/c_4 + \lambda_0)} = 9.97$. Four regimes can be distinguished: the asymptotic regime, the intermediate regime, the rising degenerate regime and the saturating degenerate regime. **Lower panel**. The contributions of Z_2 to the stationary mean. Parameters were $\frac{c_1 c_5}{c_2} = 60, \frac{c_3}{c_4} = 6, c_6 = 1, \lambda_0 = 0.02$.

with stationary distribution $\pi = \text{Poisson}(c_3/c_4)$. The expression of the stationary mean was already presented in [52] and [50]. Analogously to model E1, we analyzed the behavior of the stationary mean $\mathbb{E}[X_\infty]$ as a function of the relative correlation time c_6/c_4. For computational purposes, we truncate the state space $\mathcal{Z}$ to $\mathcal{Z}_N = \{0, 1, \ldots, N\}$ with a large enough N. Here, $N = 99$ was used.

Returning to Eq. (234), the approximation is valid for the extreme cases of (i) c_4 large compared to c_6 or (ii) c_4 small compared to c_6 (Fig. 20e, upper panel). Namely, we found $\mathbb{E}[X_\infty] = \mathcal{O}(f(\frac{c_6}{c_4}))$ for (i) $\frac{c_6}{c_4} \to \infty$ and (ii) $\frac{c_6}{c_4} \to 0$. The means $\mathbb{E}[Z_1]$ and $\mathbb{E}[Z_2]$, as well as c_5/c_6, were fixed.

Next, we partitioned the states of Z_2 into three classes: the zero state, the non-zero states below the mean, and the states equal to or above the mean. The relative speed τ_Z/τ_X of the environment defines which of these three classes dominates in terms of the corresponding environmental share, i.e., the effect on the stationary mean. In order to interpret the deviation of E2 from E1, we quantified environmental shares according to the three classes of environment states:

$$\alpha(0), \quad \alpha_{<\bar{z}} := \sum_{k=1}^{\lceil \mathbb{E}[Z_2] - 1 \rceil} \alpha(k), \quad \alpha_{\geq \bar{z}} := \sum_{k=\lceil \mathbb{E}[Z_2] \rceil}^{\infty} \alpha(k).$$

We found that, depending on the relative speed of the environment, the subsystem can be in one of three phases (Fig. 20e, lower panel). For a small c_6/c_4 ratio, the non-zero shares $\alpha_{<\bar{z}}$ and $\alpha_{\geq \bar{z}}$ both contribute significantly, with $\alpha_{\geq \bar{z}}$ showing a slight dominance over $\alpha_{<\bar{z}}$, while the share of $\alpha(0)$ is negligible. For a medium c_6/c_4 ratio, $\alpha_{<\bar{z}}$ takes the lead in dominance, while $\alpha(0)$ is still negligible. Finally, for large c_6/c_4, the share $\alpha(0)$ dominates the contribution to the mean. Returning to the upper panel of the figure 20e, we confirm that the mean $\mathbb{E}[X_\infty]$ as a function of τ_Z/τ_X undergoes the same phase transitions.

The qualitative behavior for large τ_Z/τ_X is driven by unbounded excursions in the state $Z_2 = 0$. However, in biological systems, these are generally prevented by, e.g., a leakage in the death rate. Upon introducing a base death rate λ_0, we expect the stationary mean to saturate at the upper bound as the relative environmental speed approaches zero. To demonstrate this, we generalize the propensity of the death reaction to $a(x|z) = (c_6 z + \lambda_0) x$. This altered A, i.e., $A(z) = c_6 z + \lambda_0$. The figure 21 indicates four qualitatively different regimes of the stationary mean over the relative correlation time. In particular, the three phases of the model E2 analysis (Fig. 20e.) persist, whereas the fourth phase with the largest τ_Z/τ_X reaches saturation.

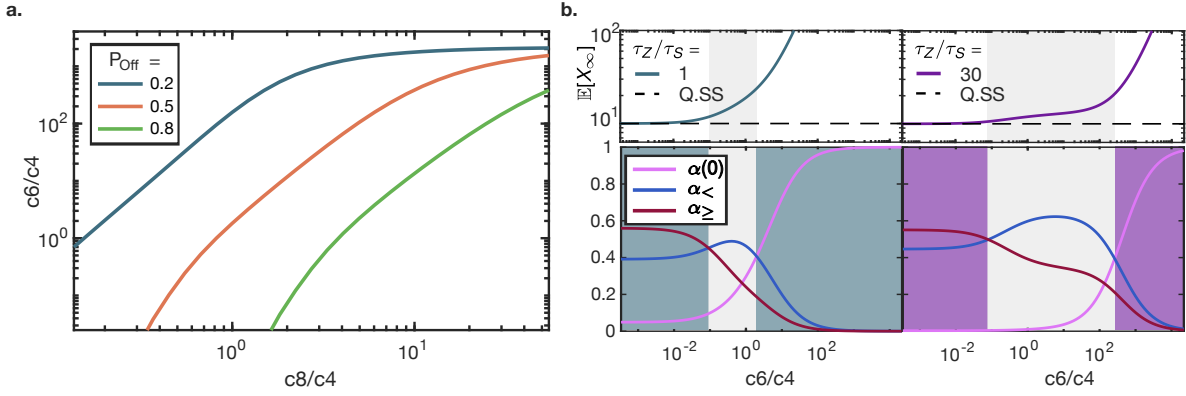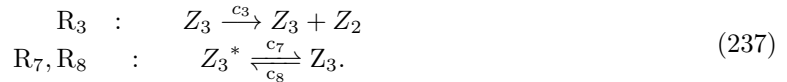Figure 22: **Model E3 Mutable c_3 synthesis. a.** The figure shows the c_4 value at which $\alpha(0)$ and $\alpha_{<\bar{z}}$ intersect as a function of $\frac{c_8}{c_4}$. Dominant $\alpha(0) > \alpha_{<\bar{z}}$ indicates that the stationary mean has entered the degenerate regime. The three curves illustrate different values of $\mathbb{P}[Z_3 = 0]$. For $\mathbb{P}[Z_3 = 0] = 0.5, 0.8$ and slow enough relative speed c_8/c_4 no intersection was found, because $\alpha(0)$ dominated $\alpha_{<\bar{z}}$ for all c_4. The mean $\mathbb{E}[Z_2] = 8$ was fixed and c_3/c_4 adapted accordingly. Parameters were $c_1 = 0.4, c_2 = 0.01, c_5 = 1, c_6 = 0.5$ and $\mathbb{E}[X_\infty^{\mathrm{Q.SS}}] = 10$. **b.** The left upper panel shows the slow switching Z_3, whereas the right panel shows the fast switching Z_3, compare figure 20f. The corresponding shares are depicted in the lower panels. Parameters were as in figure 20f.

8.3.3 Mutable synthesis of the modulator (E3)

We next considered the case where the modulating process Z_2 is itself modulated by another process, Z_3 (Fig. 20c). The modulator $Z_3 \leftrightarrow Z_3^*$ is a two-state Markov process that acts as a switch with On (Z_3) and Off (Z_3^*) states, switching rates c_7, c_8, and correlation time $\tau_S = (c_7 + c_8)^{-1}$. Here, Z_2 can be seen as a regulatory protein produced from a promoter Z_3 that alternates between On and Off states. Then Z_2 is a Markov-modulated birth-death process and, as such, represents an example of a non-Markovian death rate modulator. Since the joint environment (Z_2, Z_3) is Markovian, ESME applies.

In detail, we modified reaction R_3 of model E2, see Eq. (230) and Eq. (235), to

$$
\begin{aligned}
R_3 &: & Z_3 &\xrightarrow{c_3} Z_3 + Z_2 \\
R_7, R_8 &: & Z_3^* &\underset{c_8}{\overset{c_7}{\rightleftharpoons}} Z_3.
\end{aligned}
\tag{237}
$$

Then the two-dimensional environment $Z = (Z_2, Z_3) \in \mathbb{N} \times \{0, 1\}$ has a stationary distribution that is expressed via the confluent hypergeometric function. Expressions for $\mathbb{P}[Z_2 = m, Z_3 = 0]$ and $\mathbb{P}[Z_2 = m, Z_3 = 1]$ were derived analogously to [18] by expanding the generating function given therein.

Lemma 8.4. *Define* $a = \frac{c_7}{c_4}, b = \frac{c_7 + c_8}{c_4}, \mu = \frac{c_3}{c_4}$. *Then it holds*

$$
\mathbb{P}[Z_2 = m, Z_3 = 0] = \frac{\Gamma(b)\Gamma(m + a)(b - a)}{\Gamma(a)\Gamma(m + b + 1)} {}_1F_1(m + a, m + b + 1, -\mu)\frac{\mu^m}{m!}
$$

$$
\mathbb{P}[Z_2 = m, Z_3 = 1] = \frac{\Gamma(b)\Gamma(m + a + 1)}{\Gamma(a)\Gamma(m + b + 1)} {}_1F_1(m + a + 1, m + b + 1, -\mu)\frac{\mu^m}{m!}.
$$

Proof. Define the partial probability mass functions $p_j(m) = \mathbb{P}[Z_2 = m, Z_3 = j], j = 0, 1$. They are determined from Peccoud and Ycart's expression for the corresponding generating functions

$$
g_0(z) = \frac{c_8}{c_7 + c_8} {}_1F_1(a, b + 1, \mu(z - 1)), \quad g_1(z) = {}_1F_1(a, b, \mu(z - 1)) - g_0(z)
$$

In order to get the coefficients of the Taylor series at expansion point 0 the following integral formula for ${}_1F_1$ is used. It is correct because the condition $b > a > 0$ holds.

$$
{}_1F_1(a, b, x) = \frac{\Gamma(b)}{\Gamma(a)\Gamma(b - a)} \int_0^1 e^{xu} u^{a-1}(1 - u)^{b-a-1} \, \mathrm{d}u
\tag{238}
$$

126

Upon $x = \mu(z-1)$ we get

$$
\begin{aligned}
g_0(z) =& \frac{(b-a)\Gamma(b+1)}{b\Gamma(a)\Gamma(b-a+1)} \int_0^1 e^{\mu z u} e^{-\mu u} u^{a-1}(1-u)^{b-a}\, du \\
=& \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \sum_{n=0}^\infty \frac{z^n \mu^n}{n!} \int_0^1 e^{-\mu u} u^{n+a-1}(1-u)^{(n+b+1)-(n+a)-1}\, du \\
=& \frac{\Gamma(b)}{\Gamma(a)\Gamma(b-a)} \sum_{n=0}^\infty \frac{z^n \mu^n}{n!} {}_1F_1(n+a, n+b+1, -\mu) \frac{\Gamma(n+a)\Gamma(b-a+1)}{\Gamma(n+b+1)}
\end{aligned}
$$

Consequently

$$
p_0(m) = \frac{\Gamma(b)\Gamma(m+a)(b-a)}{\Gamma(a)\Gamma(m+b+1)} {}_1F_1(m+a, m+b+1, -\mu) \frac{\mu^m}{m!}
$$

By the analogous representation of ${}_1F_1(a, b, \mu(z-1))$ we get

$$
\begin{aligned}
\frac{m!\Gamma(a)}{\mu^m} p_1(m) =& \frac{\Gamma(b)\Gamma(m+a)}{\Gamma(m+b)} {}_1F_1(m+a, m+b, -\mu) - \frac{m!\Gamma(a)}{\mu^m} p_0(m) \\
=& \frac{\Gamma(b)}{\Gamma(b-a)} \int_0^1 e^{-\mu u} u^{a+m-1}(1-u)^{b-a-1}\, du \\
& - \frac{(b-a)\Gamma(b)}{\Gamma(b-a+1)} \int_0^1 e^{-\mu u} u^{a+m-1}(1-u)^{b-a-1}(1-u)\, du \\
=& \frac{\Gamma(b)}{\Gamma(b-a)} \int_0^1 e^{-\mu u} u^{a+m-1}(1-u)^{b-a-1}(1-(1-u))\, du \\
=& \frac{\Gamma(b)}{\Gamma(b-a)} {}_1F_1(m+a+1, m+b+1, -\mu) \frac{\Gamma(m+a+1)\Gamma(b-a)}{\Gamma(m+b+1)}
\end{aligned}
$$

Hence

$$
p_1(m) = \frac{\Gamma(b)\Gamma(m+a+1)}{\Gamma(a)\Gamma(m+b+1)} {}_1F_1(m+a+1, m+b+1, -\mu) \frac{\mu^m}{m!}.
$$

$\square$

ESME was calculated numerically using Eq. (22) with truncation $N = 100$. Figure 22a shows the entry of the stationary mean into the degenerate regime as a function of the relative speed of the modulator $Z_3 \leftrightarrow Z_3^*$. Figure 22b depicts the environmental shares for the slow and the fast modulator.

We varied the relative correlation time $\tau_Z/\tau_X = \frac{c_6}{c_4}$ while keeping the ratios $\frac{c_7}{c_4}, \frac{c_8}{c_4}$, and $\frac{c_3}{c_4}$ constant. Furthermore, we varied the relative speed of the environmental components by additionally varying the relative correlation time $\tau_Z/\tau_S = \frac{c_7+c_8}{c_4}$ while keeping the fraction of time the modulator is active, $\mathbb{E}[Z_3] = \frac{c_7}{c_7+c_8}$, constant. For a large relative speed $\frac{c_7+c_8}{c_4}$, the original model with a constant birth rate $c_3\mathbb{E}[Z_3]$ is recovered as expected.

Comparing the stationary means at different relative switching speeds of the modulator (Fig. 20f), we found the following. For slower relative speeds, the deviation from the Q.SS mean becomes more pronounced already at smaller correlation times c_4^{-1}, and the intermediate phase vanishes. Meanwhile, figure 22a visualizes how the entry into the degenerate regime depends on τ_Z/τ_S and $\mathbb{E}[Z_3]$. The muting prolongs the excursions of X in the zero or sub-average Z_2 states.

8.3.4 Stationary mean for slow environment

We considered each of the environments E1 - E3 without leakage in the death rate of X and analyzed the behavior for a slow environment and a fast subsystem, i.e., $\tau_Z/\tau_X \to \infty$. We aimed at isolating the effect of the timescale separation. For this purpose, we kept the means and the relative speed of the environment components, as well as c_5/c_6, fixed. As the plots of the environmental shares (Fig. 20d,e, 22b) suggest, the state $Z_2 = 0$ is the only one that contributes in the limit case $\tau_Z/\tau_X \to \infty$.

Consider any of the models E1, E2, E3. Under the fixation of $\mathbb{E}[X_\infty^{\text{Q.SS}}], \frac{c_5}{c_6}, \mathbb{E}[Z_2], \mathbb{E}[Z_3]$ and τ_Z/τ_S the stationary mean only depends on the relative timescale τ_Z/τ_X. Hence in the following we fix

$\tau_X = c_6^{-1} = 1$. Then, $\tau_Z/\tau_X \to \infty$ is equivalent to $c_4 \to 0$. By our main theorem (3.7), we obtain $\mathbb{E}[X_\infty] = \mathcal{O}(\pi(0)x(0))$ for $c_4 \to 0$ in the models 1a, 1b and $\mathbb{E}[X_\infty] = \mathcal{O}(\pi(0,0)x(0,0) + \pi(0,1)x(0,1))$ for model E3. By definition $x(0) = \mathbb{E}[X(\tau_{n+1})|Z_2(\tau_n) = 0]$. The function $t \mapsto \mathbb{E}[X(t)|Z_2(\tau_n) = 0]$ progresses affine-linearly in time. For very slow Z_2 the excursions are long and the base value at τ_n, from which they depart, is small compared to the value they reach at τ_{n+1}. We thus neglect the base value to obtain a linear progression with constant slope $c_5\mathbb{E}[Z_1]$, yielding

$$x(0) = \mathcal{O}(c_5\mathbb{E}[Z_1]\mathbb{E}[\tau_{n+1} - \tau_n|Z_2(\tau_n) = 0]) = \mathcal{O}\left(\frac{c_1 c_5}{c_2 c_3}\right).$$

In total we obtain for the stationary mean of models E1, E2

$$\mathbb{E}[X_\infty] = \mathcal{O}\left(\frac{c_1 c_5}{c_2 c_3}\mathbb{P}[Z_2 = 0]\right).$$

Since $\mathbb{P}[Z_2 = 0]$ only depends on the mean $\mathbb{E}[Z_2]$ which we keep fixed, the parameter c_3 which scaled linearly with c_4 dominates the asymptotic behavior $c_4 \to 0$. More precisely, we obtain

$$\lim_{c_4 \to 0} \frac{\mathbb{E}[X_\infty]}{c_4^{-1}} = \begin{cases} c_5\mathbb{E}[Z_1]\frac{(1-\mathbb{E}[Z_2])^2}{\mathbb{E}[Z_2]}, & Z_2 \text{ random telegraph} \\ c_5\mathbb{E}[Z_1]\frac{\exp(-\mathbb{E}[Z_2])}{\mathbb{E}[Z_2]}, & Z_2 \text{ birth-death} \end{cases}.$$

For model E3 the two states $(Z_2, Z_3) = (0,0)$ and $(0,1)$ contribute for $c_4 \to 0$. After setting up the recursion that couples $x(0,0)$ and $x(0,1)$, see next paragraph, we obtain

$$\lim_{c_4 \to 0} \frac{\mathbb{E}[X_\infty]}{c_4^{-1}} = c_5\mathbb{E}[Z_1](\frac{\mathbb{P}[Z_2 = 0]}{\mathbb{E}[Z_2]} + \frac{c_4}{c_8}\frac{\mathbb{P}[Z_2 = 0, Z_3 = 0]}{1 - \mathbb{E}[Z_3]})$$

Note that $\mathbb{P}[Z_2 = 0]$ and $\mathbb{P}[Z_2 = 0, Z_3 = 0]$ only depend on the relative rates $\frac{c_7}{c_4}, \frac{c_8}{c_4}, \frac{c_3}{c_4}$, which were fixed, i.e., the timescale of the joint environment (Z_2, Z_3) was varied. The inverse proportional dependence on c_4 for models E1 - E3 is visualized in figures 20c.,f.,i.

For $c_4 \to 0$ we compute the stationary mean under assumption that all terms in Eq. (21) vanish except for $z = (0,0),(0,1)$. The recursion equations (18) for the two states $(0,0)$ and $(0,1)$ read

$$0 = -c_7\pi(0,0)x(0,0) + c_8\pi(0,1)x(0,1) + c_4\pi(1,0)x(1,0) + \frac{c_5 c_1}{c_2}\pi(0,0)$$

$$0 = -(c_8 + c_3)\pi(0,1)x(0,1) + c_7\pi(0,0)x(0,0) + c_4\pi(1,1)x(1,1) + \frac{c_5 c_1}{c_2}\pi(0,1)$$

By assumption,

$$\max\{\pi(1,1)x(1,1), \pi(1,0)x(1,0)\} \ll \min\{\pi(0,1)x(0,1), \pi(0,0)x(0,0)\}$$

so we can set $\pi(1,1)x(1,1) = \pi(1,0)x(1,0) = 0$. Then the 2 dimensional linear system has the solution

$$\pi(0,0)x(0,0) = \frac{c_5 c_1}{c_2} \cdot (\frac{c_8 + c_3}{c_3 c_7}\pi(0,0) + \frac{c_8}{c_3 c_7}\pi(0,1))$$

$$\pi(0,1)x(0,1) = \frac{c_5 c_1}{c_2} \cdot \frac{\pi(0,0) + \pi(0,1)}{c_3}$$

which by $\mathbb{E}[X_\infty] = \pi(0,0)x(0,0) + \pi(0,1)x(0,1)$ yields the result for $\lim_{c_4 \to 0} \mathbb{E}[X_\infty]c_4$.

In summary, we derived

$$\mathbb{E}[X_\infty] = \mathcal{O}\left(\frac{\tau_Z}{\tau_X}\right), \tag{239}$$

i.e., the stationary mean grows proportionally to the timescale differences. The parallel asymptotes in the log-log plots with unit slope in the figures 20g,h,i reflect this dependence.

### 8.3.5 Application to a division-dilution model

In a model similar to E1, subsection 8.3.1, Beentjes et al. [179] studied the dilution of protein copy numbers due to cell division. The mean copy number in a single lineage increased in a model with Erlang distributed division time and binomial partitioning at cell division (model III in the reference) compared to the same model with deterministic division times (model II) and to a birth-death model with averaged effective degradation rate (model I). We can attribute this deviation to the random cell division times. To demonstrate this attribution, we elaborate on the analogy to the random duration of excursions in E1. Let the extrinsic noise component (the environment) be the cell cycle duration. To formalize this, we introduce the time points of division, denoted by a sequence of random variables $0 =: \tau_0 < \tau_1 < \tau_2 < \ldots$, such that the $\tau_{i+1} - \tau_i$ are i.i.d. Let $X(t)$ be the number of proteins that increases on average with rate λ between environment jumps, be it by geometrically distributed bursts or by simple birth reactions. Furthermore, let $X(\tau_i)$ follow a Binomial distribution with $p = 1/2$ and $N = X(\tau_i-)$. If the protein production, the binomial partition at division, and the environment jumps are all assumed to be stochastically independent, it is easy to see that $Y(t) := E[X(t)|(\tau_i)_i]$ evolves as $\dot{Y}(t) = \lambda$ between the jumps and is set to $Y(\tau_i) = 1/2 \cdot Y(\tau_i-)$ at the environment jumps. It is furthermore easy to derive $\mathbb{E}[Y(\tau_i)] = \lambda \mathbb{E}[\tau_1]$ for the stationary system. Then, analogous to the derivation in section 3.1.3, we obtain

$$\mathbb{E}[X_\infty] = \lim_{N \to \infty} \frac{\frac{1}{N} \sum_{i=1}^N \int_{\tau_{i-1}}^{\tau_i} Y(t) \, \mathrm{d}t}{\frac{1}{N} \sum_{i=1}^N \tau_i - \tau_{i-1}} = \lambda \mathbb{E}[\tau_1] + \frac{1/2 \lambda \mathbb{E}[\tau_1^2]}{\mathbb{E}[\tau_1]}$$
$$= \frac{\lambda}{2} \left(3\mathbb{E}[\tau_1] + \frac{\mathbb{V}\mathrm{ar}[\tau_1]}{\mathbb{E}[\tau_1]} \right).$$

For $(N, N/y)$-Erlang distributed τ_1, i.e., $\mathbb{E}[\tau_1] = y$ and $\mathbb{V}\mathrm{ar}[\tau_1] = y^2/N$, this is in accordance with Eq. (39) in [179], which was derived using a generating function approach. Compared with our two-state Markov model of alternating linear increase and exponential decay in X, the division-dilution model replaces the periods of exponential decay by instantaneous decays at divisions. In both models, the stochasticity in the duration of excursions (referred to as cell cycle length variability in the reference) causes an increase in the stationary mean. While in the reference paper this is the single cause of the increase compared to the model I, in our case the increase is caused by an interplay of the random duration, the frequency of the excursions, and the timescale separation between Z and X. Note that the reference models (model I and Q.SS, respectively) are constructed differently. The degradation rate of the model I was tuned such that the stationary mean matches with that of the deterministic division-dilution model II. As a consequence, it encodes the timescale of the environment, whereas in the Q.SS reference model, the degradation rate is uncoupled from the environment timescale. This uncoupling provides a degree of freedom which causes a larger relative increase of the stationary mean value in our case study than in Beentjes et al., compared to the respective reference models.

8.3.6 Probability evolution equation for the telegraph-modulated birth-death process

Zeroth-order modulation was seen to not impact the mean of the subsystem. However, the probability evolution equation is affected. As a minimal example, we considered the promoter-mediated transcription with mRNA decay, see case study 8.1.1. To keep the example minimal, we considered two promoter structures, an active and an inactive one, i.e., a random telegraph model

$$Z \xrightleftharpoons[c_1]{c_2} Z^*,$$
$$Z^* \xrightarrow{\lambda} Z^* + X, \quad 1 \leq i \leq k \tag{240}$$
$$X \xrightarrow{\delta} \emptyset.$$
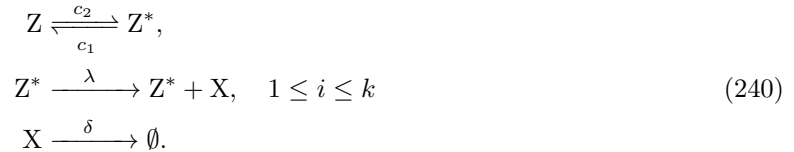
The cumulant approach, section 3.5, provides analytically closed probability evolution equations in this case. The time point marginals of the random telegraph process at stationarity are Bernoulli distributed with success probability $\psi := \frac{c_1}{c_1 + c_2}$, having a variance of $\sigma^2 := \frac{c_1 c_2}{(c_1 + c_2)^2}$. We applied the cumulant expansion (53) for the canonical birth operator $\Lambda_1 = \mathcal{B}$ with birth rate λ, the canonical

death operator $\Lambda_2 = \mathcal{D}$ with death rate δ and the mean operator $\overline{\Lambda} := \psi \mathcal{B} + \mathcal{D}$. Define $\varphi := c_1 + c_2$ is the switching timescale of the random telegraph process and $\theta := \mathbb{E}[(-1)^{Z_0}] = 1 - 2\psi = \frac{c_2 - c_1}{c_1 + c_2}$. For the derivation we use the Baker-Campbell-Hausdorff formula, namely, if $[X, Y] = sX$, then

$$X \exp(Y) = \exp(Y) e^s X. \tag{241}$$

Using Eq. (142), derived in section 6.1.1,

$$
\begin{aligned}
\partial_t p_t &= \overline{\Lambda} p_t + \sum_{n=0}^{\infty} \int_0^t \int_{\Delta^n[t',t]} \mathcal{C}(t', t_1, \ldots, t_n, t) \mathcal{B} \prod_{k=0}^n e^{\Delta t_k \overline{\Lambda}} \mathcal{B} \, \mathrm{d}(t_1, \ldots, t_n) p_{t'} \, \mathrm{d}t' \\
&= \overline{\Lambda} p_t + \sum_{n=0}^{\infty} \int_0^t \int_{\Delta^n[t',t]} \theta^n \sigma^2 e^{-\varphi(t-t')} \mathcal{B} \prod_{k=0}^n e^{\Delta t_k \overline{\Lambda}} \mathcal{B} \, \mathrm{d}(t_1, \ldots, t_n) p_{t'} \, \mathrm{d}t' \\
&= \overline{\Lambda} p_t + \int_0^t \sigma^2 \mathcal{B} e^{-\varphi(t-t')} \sum_{n=0}^{\infty} \int_{\Delta^n[t',t]} e^{t\overline{\Lambda}} \prod_{k=1}^n e^{-t_k \overline{\Lambda}} (\theta \mathcal{B}) e^{t_k \overline{\Lambda}} \, \mathrm{d}(t_1, \ldots, t_n) e^{-t'\overline{\Lambda}} \mathcal{B} p_{t'} \, \mathrm{d}t' \\
&= \overline{\Lambda} p_t + \int_0^t \sigma^2 \mathcal{B} e^{-\varphi(t-t')} e^{(t-t')(\overline{\Lambda} + \theta \mathcal{B})} \mathcal{B} p_{t'} \, \mathrm{d}t' \\
&= \overline{\Lambda} p_t + \mathcal{B} \int_0^t \sigma^2 e^{(t-t')(\overline{\Lambda} + \theta \mathcal{B} - \varphi)} \mathcal{B} p_{t'} \, \mathrm{d}t'.
\end{aligned}
$$

In summary, we resolved all terms in the series to obtain an explicit expression for the kernel without the need of a truncation. Hence, the following generalized master equation governs the probability evolution

$$\partial_t p_t = \overline{\Lambda} p_t + \mathcal{B} \int_0^t \sigma^2 \exp\left[(\overline{\Lambda} + \theta \mathcal{B} - \varphi)(t - t')\right] \mathcal{B} p_{t'} \, \mathrm{d}t'. \tag{242}$$

Upon defining the integral term as q_t one recovers the memory-less evolution equation system

$$
\begin{aligned}
\partial_t p_t &= \overline{\Lambda} p_t + \mathcal{B} q_t \\
\partial_t q_t &= \sigma^2 \mathcal{B} p_t + (\overline{\Lambda} + \theta \mathcal{B} - \varphi) q_t.
\end{aligned}
\tag{243}
$$

This system is a linear transformation of [18, Eq.(5)], which reads, for $p_t^{(z)}(x) := \mathbb{P}[Z_t = z, X_t = x]$,

$$
\begin{aligned}
\partial_t p_t^{(0)} &= -c_1 p_t^{(0)} + c_2 p_t^{(1)} + \mathcal{D} p_t^{(0)} \\
\partial_t p_t^{(1)} &= c_1 p_t^{(0)} - c_2 p_t^{(1)} + \mathcal{B} p_t^{(1)} + \mathcal{D} p_t^{(1)}.
\end{aligned}
$$

Namely, for $p_t(x) = p_t^{(1)}(x) + p_t^{(0)}(x)$ and $q_t(x) = (1 - \psi) p_t^{(1)}(x) - \psi p_t^{(0)}(x)$ we obtain Eq. (243) by a linear transform.

In Eq. (242) as it is written, commuting the birth operator with the operator exponential simply introduces another factor $\exp(-\delta(t - t'))$. More precisely, using Eq. (241) with $X = \mathcal{B}, Y = (t - t')(\overline{\Lambda} + \theta \mathcal{B} - \varphi)$, and $[X, Y] = (t - t')[\mathcal{B}, \mathcal{D}] + (t - t')[\mathcal{B}, (\psi + \theta)\mathcal{B} - \varphi] = -(t - t')\delta \mathcal{B} + 0$, we obtained

$$\partial_t p_t = \overline{\Lambda} p_t + \int_0^t \sigma^2 e^{(t-t')(\overline{\Lambda} + \theta \mathcal{B} - (\varphi + \delta))} \mathcal{B}^2 p_{t'} \, \mathrm{d}t'. \tag{244}$$

We make two remarks about what happens to Eq. (242), when (i) changing the modulated reaction, (ii) altering the state variable p_t to be the generating function. When (i) the random telegraph environment modulates the death reaction instead of the birth reaction, then the same equation holds with $\mathcal{B}$ replaced by $\mathcal{D}$ and with $\overline{\Lambda} = \mathcal{B} + \psi \mathcal{D}$ accordingly. In detail, parallel to the calculation for the modulated birth rate, we obtained

$$
\begin{aligned}
\partial_t p_t &= \overline{\Lambda} p_t + \sum_{n=0}^{\infty} \int_0^t \int_{\Delta^n[t',t]} \mathcal{C}(t', t_1, \ldots, t_n, t) \mathcal{D} \prod_{k=0}^n e^{\Delta t_k \overline{\Lambda}} \mathcal{D} \, \mathrm{d}(t_1, \ldots, t_n) p_{t'} \, \mathrm{d}t' \\
&= \overline{\Lambda} p_t + \mathcal{D} \int_0^t \sigma^2 e^{(t-t')(\overline{\Lambda} + \theta \mathcal{D} - \varphi)} \mathcal{D} p_{t'} \, \mathrm{d}t'.
\end{aligned}
$$

Upon defining the integral term as q_t one recovers the memory-less evolution equation system

$$\partial_t p_t = \overline{\Lambda} p_t + \mathcal{D} q_t$$
$$\partial_t q_t = \sigma^2 \mathcal{D} p_t + (\overline{\Lambda} + \theta \mathcal{D} - \varphi) q_t. \tag{245}$$

Unlike in the above case, the operator $\mathcal{D}$ does not commute with the operator exponential by the simple compensation with a factor. Hence, we do not obtain an equation analogous to Eq. (244).

When we seek (ii) an evolution equation for the generating function $\hat{p}_t \colon [0,1] \to \mathbb{R}_{\geq 0}$ with $\hat{p}_t(s) := \mathbb{E}[s^{X_t}]$ instead of p_t, then the same (242) holds true using the corresponding operators $(\tilde{\mathcal{B}}f)(s) = -\lambda(1-s)f(s), (\mathcal{D}f)(s) = \delta(1-s)\partial_s f(s)$. Note that both $X = \Lambda, Y = \overline{\Lambda} - p_1\Lambda = \mathcal{B}$ and $X = \mathcal{L}, Y = \overline{\mathcal{L}} - p_1\mathcal{L}$ satisfy the same commutator relation $[X, Y] = -\delta p_1 Y$.

To contrast the cumulant expansion approach for system in Eqs. 240 with the filtering approach, we consider as a sufficient statistic the backward recurrence time $\tau(t)$ of the birth reaction counting process, definition 4.11. It denotes the time since the last birth event and evolves according to $\partial_t \tau(t) = 1, t \in (t_{i-1}, t_i), \tau(t_i) = 0$. The hybrid master equation (104)-(105) is found in the section 4.2.4. The marginalized hybrid master Eq. (106) in this example can be simplified by the help of Eq. (105). Namely, from

$$\partial_t p_t(\tau, x) = -\partial_\tau p_t(\tau, x) - \lambda g(\tau) p_t(\tau, x) + \mathcal{D} p_t(\tau, x)$$

and integration over τ on the half axis $[0, \infty)$ by use of $\int_0^\infty \lambda g(\tau) p_t(\tau, x) \, d\tau = p_t(0, x+1)$ we obtain

$$\partial_t p_t(x) = \tilde{q}_t(x+1) - \tilde{q}_t(x) + (\mathcal{D} p_t)(x) \tag{246}$$

with $\tilde{q}_t(x+1) := p_t(0, x+1)$. Moreover, we derive

$$\begin{aligned} p_t(0, x+1) &= \lim_{h \to 0} \frac{\mathbb{P}[\tau(t+h) \in [0, h), X(t+h) = x+1]}{h} \\ &= \lim_{h \to 0} \frac{\mathbb{P}[X(t+h) = x+1, X(t) = x]}{h} \\ &= \lambda \mathbb{P}[Z(t) = 1, X(t) = x] \\ &= \lambda p_t^{(1)}(x). \end{aligned}$$

Then Eq. (246) becomes $\partial_t p_t(x) = \mathcal{B} p_t^{(1)}(x) + \mathcal{D} p_t(x)$, which is equivalent to the evolution of $\partial_t(p_t^{(0)}(x) + p_t^{(1)}(x))$ in Eq. (243). This illustrates the equivalence of the approaches, that condition on either the environment or the subnetwork with the approach without conditioning, see section 4.5. The equivalence is here obtained from linear transformations. We expect that linear transformations cannot in general link the approaches.

8.3.7 Mean evolution for the telegraph-modulated birth-death process

In the previous paragraphs we derived generalized master equations for the birth-death process with a modulated birth reaction. We also commented on how the equations are altered when the death reaction is modulated. For the modulated birth reaction, the generalized mean equation agrees with the Q.SS mean equation (section 3.1.1), i.e., $\partial_t \mathbb{E}[X_t] = \lambda \mathbb{E}[Z_t] - \delta \mathbb{E}[X_t]$. In contrast, for the modulated death reaction we obtained equations that deviate from the naive mean equation, see Eq. (233). The mean equations can be derived directly,

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[X_t] = \bar{A} \mathbb{E}[X_t] + \mathbb{E}[\Delta A(Z_t) \Delta X_t]$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[X_t] = \bar{A} \mathbb{E}[X_t] + \int_0^t \mathcal{C}(t', t) A e^{(t-t')\bar{A}} A \mathbb{E}[X_{t'}] \, \mathrm{d}t' + \dots$$

$$+ \int_0^t \int_{t'}^t \mathcal{C}(t', t_1, t) A e^{(t-t_1)\bar{A}} A e^{(t_1-t')\bar{A}} A \mathbb{E}[X_{t'}] \, \mathrm{d}t'$$

$$+ \int_0^t \int_{t'}^t \mathcal{C}(t', t_1, t_2, t) A e^{(t-t_2)\bar{A}} A e^{(t_2-t_1)\bar{A}} A e^{(t_1-t')\bar{A}} A \mathbb{E}[X_{t'}] \, \mathrm{d}t'.$$

Then

$$\int_0^t \int_{t_1}^t \cdots \int_{t_{n-1}}^t e^{-\frac{c_1}{c_1+c_2}\delta(t-t_1)}(-\mu)^{n+1}\mathcal{C}(t_1,\ldots,t_n,t)\mathbb{E}[X_{t_1}]\,\mathrm{d}t_n\ldots\,\mathrm{d}t_2\,\mathrm{d}t_1$$

$$=\int_0^t \mu^2 e^{-\frac{c_1}{c_1+c_2}\delta(t-t_1)}\mathbb{E}[X_{t_1}]\int_{t_1}^t\cdots\int_{t_{n-1}}^t (-\mu)^{n-1}\mathcal{C}(t_1,\ldots,t_n,t)\,\mathrm{d}t_n\ldots\,\mathrm{d}t_2\,\mathrm{d}t_1$$

$$=\int_0^t \mu^2 e^{-\frac{c_1}{c_1+c_2}\delta(t-t_1)}\mathbb{E}[X_{t_1}]\frac{\left(-\delta\frac{c_2-c_1}{c_1+c_2}\cdot(t-t_1)\right)^{n-1}}{(n-1)!}\gamma(0,t-t_1)\,\mathrm{d}t_1.$$

Summing over n, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t]=\lambda-\frac{\delta c_1}{(c_1+c_2)}\mathbb{E}[X_t]$$
$$+\sum_{n=1}^\infty \int_0^t \int_{t_1}^t \cdots \int_{t_{n-1}}^t e^{-\frac{c_1}{c_1+c_2}\delta(t-t_1)}(-\mu)^{n+1}\mathcal{C}(t_1,\ldots,t_n,t)\mathbb{E}[X_{t_1}]\,\mathrm{d}t_n\ldots\,\mathrm{d}t_2\,\mathrm{d}t_1$$
$$=\lambda-\frac{\delta c_1}{(c_1+c_2)}\mathbb{E}[X_t]+\frac{\mu^2 c_1 c_2}{(c_1+c_2)^2}\int_0^t e^{-(\frac{c_1}{c_1+c_2}\delta+\delta\frac{c_2-c_1}{(c_1+c_2)}+(c_1+c_2))(t-t_1)}\mathbb{E}[X_{t_1}]\,\mathrm{d}t_1$$
$$=\lambda-\frac{\delta c_1}{(c_1+c_2)}\mathbb{E}[X_t]+\frac{\mu^2 c_1 c_2}{(c_1+c_2)^2}\int_0^t e^{-(\frac{\delta c_2}{c_1+c_2}+(c_1+c_2))(t-t_1)}\mathbb{E}[X_{t_1}]\,\mathrm{d}t_1.$$

For the equilibrium, introduce $y(t)$ for the integral term to obtain the two-dimensional linear ODE

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix}\mathbb{E}[X_t]\\y(t)\end{bmatrix}=\begin{bmatrix}\lambda\\0\end{bmatrix}+\begin{bmatrix}-\delta\psi & \delta^2\sigma^2\\1 & -(\delta(1-\psi)+\varphi)\end{bmatrix}.\tag{247}$$

Following the standard stability analysis for a two dimensional ODE, we compute the trace, determinant and discriminant

$$\mathrm{tr}=-(\delta+c_1+c_2)<0$$
$$\det=c_1\delta>0$$
$$\Delta=(\delta+c_1+c_2)^2-4c_1\delta>(\delta-c_1)^2\geq 0.$$

The trace-determinant criterion reveals that the fixed point is stable. In summary, applying the cumulant expansion approach, we derived

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t]=\lambda-\delta\psi\mathbb{E}[X_t]+\delta^2\sigma^2\int_0^t e^{-(\delta(1-\psi)+\varphi)(t-t')}\mathbb{E}[X_{t'}]\,\mathrm{d}t'\tag{248}$$

with the stable equilibrium $\mathbb{E}[X_\infty]=\frac{\lambda}{\delta\psi}\left(1+\frac{(1-\psi)\delta}{\varphi}\right)$. This expression for the stationary mean is in accordance with Eq. (229).

When truncating the cumulant expansion after the first cumulant term, we obtain the same functional form of the evolution equation, reading

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t]=\lambda-\delta\psi\mathbb{E}[X_t]+\delta^2\sigma^2\int_0^t e^{-(\delta\psi+\varphi)(t-t')}\mathbb{E}[X_{t'}]\,\mathrm{d}t'\tag{249}$$

We notice that the exponent of the exponential kernel is different. This also results in mismatched stationary means. As another consequence, if we inferred parameters based on the non-truncated model and the truncated model, the interpretation of the inferred parameters would differ. Finally, the mismatch indicates a non-intuitive property of an optimal projection. Suppose we constrained the evolution equation to a second-order truncated functional form with variable exponent and we wanted to project the exact evolution equation onto the functional form. Then the second-order truncation is not generally the optimal choice of the variable exponent.

To illustrate the correspondence with the alternative route via conditioning on the subnetwork, we apply Eq. (107), for $\varphi = 1$ and $\varphi = \text{id}$ reads

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t] = \lambda - \delta\mathbb{E}[X_t\mathbb{E}[Z_t|X_{[0,t]}]]$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X_t\mathbb{E}[Z_t|X_{[0,t]}]] = \mathbb{E}[X_t(c_1 - (c_1 + c_2)\mathbb{E}[Z_t|X_{[0,t]}])] + \lambda\mathbb{E}[\mathbb{E}[Z_t|X_{[0,t]}]] - \delta\mathbb{E}[X_t\mathbb{E}[Z_t^2|X_{[0,t]}]].$$

In this particular example, we make use of $Z_t^2 = Z_t$ to obtain the closed two-dimensional system

$$\frac{\mathrm{d}}{\mathrm{d}t}\begin{bmatrix} \mathbb{E}[X_t] \\ \mathbb{E}[X_t\Pi_t] \end{bmatrix} = \begin{bmatrix} \lambda \\ \frac{\lambda c_1}{c_1+c_2} \end{bmatrix} + \begin{bmatrix} 0 & \delta \\ c_1 & -(c_1+c_2+\delta) \end{bmatrix}. \tag{250}$$

Note that the same equation also holds for $\mathbb{E}[X_t\Pi_t]$ replaced by $\mathbb{E}[X_tZ_t]$ by the tower property. In fact, $\mathbb{E}[X_t\Pi_t] = \mathbb{E}[X_tZ_t]$. Hence, in this case, the filtering is not necessary to find the solution. For the centered variable $y(t) = \mathbb{E}[X_t\Pi_t] - \mathbb{E}[X_t]\mathbb{E}[\Pi_t]$, the equation (247) holds. We found that the centered variable $y(t) = \mathbb{E}[X_t\Pi_t] - \mathbb{E}[X_t]\mathbb{E}[\Pi_t]$ equals the integral term in (248). This demonstrates that the approaches via conditioning on the environment and via conditioning on the subnetwork lead to equations which are transformations of each other, emphasizing the equivalence of the approaches without using properties of the conditional expectation, section 4.5.

In remark 3.13, we interpreted the terms in the cumulant expansion. To illustrate this interpretation more clearly, we considered the example of a birth-death process $X(t)$ and a stationary process $Z(t)$ with mean μ_2 that modulates the decay reaction. Let c_1 be the birth rate. For convenience, we assumed $X(0) = 0$. Using the conditioning on the environment, the conditional mean evolves as

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X(t)|Z_{[0,t]}] = c_1 - Z(t)\mathbb{E}[X(t)|Z_{[0,t]}],$$

solved by

$$\mathbb{E}[X(t)|Z_{[0,t]}] = \int_0^t c_1 \exp\left(-\int_s^t Z(\sigma)\,\mathrm{d}\sigma\right)\,\mathrm{d}s = c_1\int_0^t e^{-\mu_2(t-s)}\exp\left(-\int_s^t Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\,\mathrm{d}s$$

with mean

$$\mathbb{E}[X(t)] = \mathbb{E}[\mathbb{E}[X(t)|Z_{[0,t]}]] = c_1\int_0^t e^{-\mu_2(t-s)}\mathbb{E}\left[\exp\left(-\int_s^t Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\right]\,\mathrm{d}s.$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X(t)] = c_1 - \mu_2\mathbb{E}[X(t)] - E\left[(Z(t) - \mu_2)\mathbb{E}[X(t)|Z_{[0,t]}]\right] \tag{251}$$

$$= c_1 - \mu_2\mathbb{E}[X(t)] - c_1\int_0^t e^{-\mu_2(t-s)}\mathbb{E}\left[(Z(t) - \mu_2)\exp\left(-\int_s^t Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\right]\,\mathrm{d}s. \tag{252}$$

If we assume the approximation that $Z(\sigma)_{\sigma\geq t'}$ is independent of $Z(\sigma)_{\sigma<t'}$ for any $t' \in [0,t)$, we get

$$\mathbb{E}\left[(Z(t) - \mu_2)\exp\left(-\int_s^t Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\right] \tag{253}$$

$$= \mathbb{E}\left[(Z(t) - \mu_2)\int_s^t (Z(t') - \mu_2)\exp\left(-\int_s^{t'} Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\,\mathrm{d}t'\right] \tag{254}$$

$$= \int_s^t \mathbb{E}\left[(Z(t) - \mu_2)(Z(t') - \mu_2)\exp\left(-\int_s^{t'} Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\right]\,\mathrm{d}t' \tag{255}$$

$$\approx \int_s^t \mathbb{E}[(Z(t) - \mu_2)(Z(t') - \mu_2)]\,\mathbb{E}\left[\exp\left(-\int_s^{t'} Z(\sigma) - \mu_2\,\mathrm{d}\sigma\right)\right]\,\mathrm{d}t'. \tag{256}$$
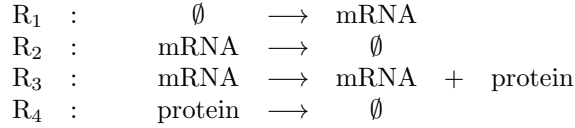
Plugging this approximation into Eq. (252), we obtain

$$
\frac{\mathrm{d}}{\mathrm{d}t}\mathbb{E}[X(t)]
$$

$$
\approx c_1 - \mu_2\mathbb{E}[X(t)]
$$

$$
- c_1 \int_0^t e^{-\mu_2(t-s)} \int_s^t \mathbb{E}\left[(Z(t)-\mu_2)(Z(t')-\mu_2)\right]\mathbb{E}\left[\exp\left(-\int_s^{t'} Z(\sigma)-\mu_2\ \mathrm{d}\sigma\right)\right]\ \mathrm{d}t'\ \mathrm{d}s
$$

$$
= c_1 - \mu_2\mathbb{E}[X(t)]
$$

$$
- c_1 \int_0^t \mathbb{E}[(Z(t)-\mu_2)(Z(t')-\mu_2)]e^{-\mu_2(t-t')} \int_0^{t'} e^{-\mu_2(t'-s)}\mathbb{E}\left[\exp\left(-\int_s^{t'} Z(\sigma)-\mu_2\ \mathrm{d}\sigma\right)\right]\ \mathrm{d}s\ \mathrm{d}t'
$$

$$
= c_1 - \mu_2\mathbb{E}[X(t)] - \int_0^t \mathrm{Cov}[Z(t'),Z(t)]e^{-\mu_2(t-t')}\mathbb{E}[X(t')]\ \mathrm{d}t'.
$$

This recovers exactly the cumulant expansion truncated after the first cumulant term, see [106, Example 4.2.3]. Hence, when we assume that the environment trajectories before and after the intermediate time point t' are independent, we obtain the first cumulant term.

8.3.8   Gene expression in a random environment

Next, we extend the single birth-death process to a cascade of two birth-death processes and place it in a random environment. For this example, we illustrate the effects of the environmental parameter, i.e., the mean, variance and autocorrelation time, using the tilted Hawkes model for independent environmental components, section 7.1. We applied Sinzger's min-thin algorithm for an efficient simulation. Each component modulates one of the four reactions of the following gene expression model, see figure 23A.

$$
\begin{array}{llccll}
\mathrm{R}_1 & : & \emptyset & \longrightarrow & \mathrm{mRNA} & \\
\mathrm{R}_2 & : & \mathrm{mRNA} & \longrightarrow & \emptyset & \\
\mathrm{R}_3 & : & \mathrm{mRNA} & \longrightarrow & \mathrm{mRNA} & + \quad \mathrm{protein} \\
\mathrm{R}_4 & : & \mathrm{protein} & \longrightarrow & \emptyset &
\end{array}
$$

When all environment components are set to their constant mean level, we recover the unembedded gene expression model, i.e. the homogeneous Q.SS reference model. As we incorporate the modulation of the first-order reactions, we do not expect the mean to coincide with the mean of the reference model. As relevant characteristics of the environment, we considered the variance and the autocorrelation time of the individual environment components. We varied these characteristics for the first-order modulators Z_2, Z_3 and Z_4 one at a time, while leaving the remaining ones fixed. Figure 23B depicts the setup of the simulation study. For the variance σ_j^2, we chose μ_j^2 as an upper bound for the considered domain. This bound is motivated by the CIR process, which guarantees the strict positivity of the trajectory only for $\sigma_j^2 \leq \mu_j^2$. For the autocorrelation time $\tau_j = \gamma_j^{-1}$, we distinguished three regions. The region for which the timescale of changes in the environment is of the order of the mRNA timescale or faster (low autocorrelation), the region where the environment timescale is shared with the protein timescale (medium autocorrelation) and the slower-than-protein timescale (high autocorrelation). Additionally, we investigated how the timescale differences between the mRNA level and the protein level in the reference model affects the results. For this purpose, we separately considered a fast mRNA timescale (Z_1 and Z_2 multiplied by $c_1 = c_2 = 1$) and a slow mRNA timescale (Z_1 and Z_2 multiplied by $c_1 = c_2 = 0.1$). This choice keeps the mean of the protein level the same for the homogeneous case, but brings the mRNA and protein timescales closer to each other in the latter case. The gene expression model is a cascade of two birth-death processes. We thus were interested to understand what was different between changing the modulation at the mRNA level (R_2) compared to the protein level (R_3 and R_4). Also, we expected to see differences between the modulation of the birth reaction (R_3) and the death reactions (R_2 and R_4).

We first investigated the regime of low environment correlation time, at most of the order of the mRNA level. We saw no effect in the stationary mean of the protein (figure 23C) for varying the modulation at the protein level (R_3 and R_4). For a variation at the mRNA level (R_2), the stationary
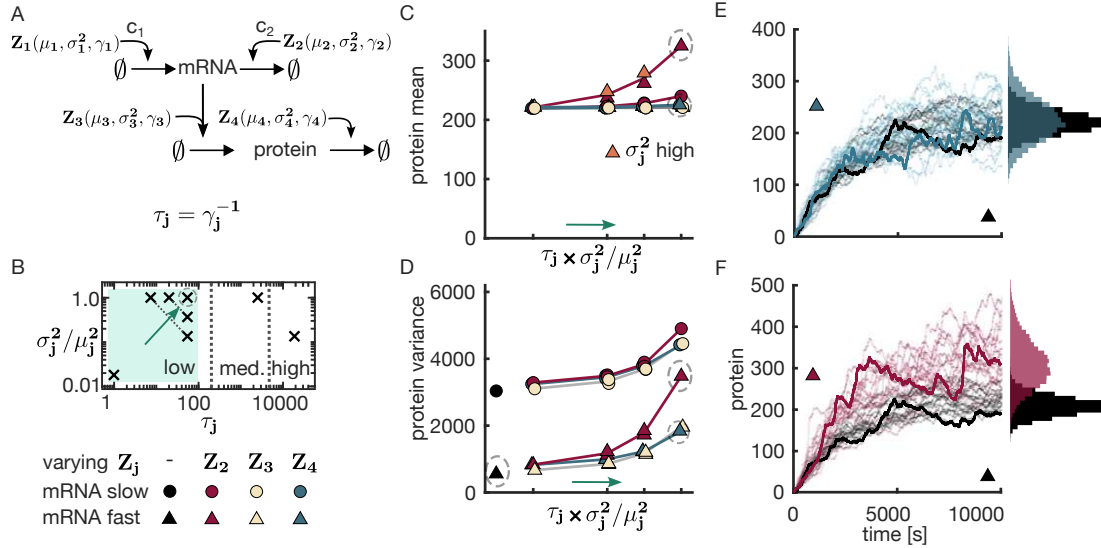
Figure 23: **Gene expression model in a fast random environment.** A. Four independent environment components modulate the gene expression model. Each component Z_j is characterized by its mean μ_j, its variance σ_j^2 and its autocorrelation decay γ_j. The inverse of γ_j is the autocorrelation time τ_j. The birth and death reactions of the mRNA can be multiplied by factors c_1 and c_2. B. The setup of the simulation study - We simulated trajectories of the model in A with the approximate marginal simulation algorithm for the tilted Hawkes approximation. For each component Z_2, Z_3 and Z_4, we varied σ_j^2 and τ_j. The parameter pairs (σ_j^2, τ_j) for which we simulated trajectories with are marked by crosses in the logarithmic (σ_j^2, τ_j)-plane. For the comparison of the three components, we varied σ_j^2 relative to μ_j^2. Along the decreasing diagonals, the product $\tau_j \times \sigma_j^2/\mu_j^2$ is constant. We distinguished the low autocorrelation regime from the medium and high regime. A correlation time of $50 - 500s$ (upper end of low regime to lower end of the medium regime) was the correlation time of the mRNA. We conducted simulations for a slow ($c_1 = c_2 = 0.1$) and fast ($c_1 = c_2 = 1$) mRNA timescale. The black circle and triangle symbolizes the reference model, in which all environment components are replaced by their constant mean values. C and D. The mean and the variance of the protein at time point $t = 10000$, estimated from 10000 sampled trajectories, is shown for the different conditions. The ticks of the x-axis correspond to the four diagonals of the low regime in B. Accordingly, the second and third tick combine two values which overlap, sometimes to an extent that only one is visible. For Z_2 we indicated the pair with the larger σ_2^2 by orange (C). Lines were added and different colors for the same tick were shifted horizontally to improve readability. The stationary variance of the homogeneous model was added for reference in black. E and F. Sample protein trajectories of the reference model were plotted in black to contrast the embedded model with the largest variance and correlation for fast mRNA timescale and Z_4 (E) and Z_2 (F). Histograms of time $t = 10000$ were added. Circled marks in B, C and D indicate that for these parameter choice trajectories are shown in E and F. The components that were not varied were fixed at $((\mu_j, \sigma_j^2, \gamma_j))_{j=1,\dots,4} = ((0.06, 0.001, 0.05), (0.02, 10^{-5}, 0.3), (0.03, 3 \cdot 10^{-4}, 0.05), (4 \cdot 10^{-4}, 4 \cdot 10^{-9}, 0.3))$. In the shaded low regime of B parameters were $\sigma_j^2/\mu_j^2 = e^{-k}, k = 0, 1, 2, 4$ and $\tau_j = e^k, k = 0, 2, 3, 4$.

mean increased when we increased σ_2^2 or τ_2. This effect was stronger for the fast mRNA timescale and for larger σ_2^2 when the product $\sigma_2^2 \times \tau_2$ was fixed. The autocorrelation showed no substantial change in either case, see figure 25. In contrast, we saw a noticeable effect on the stationary variance of the protein in all conditions, see figure 23D. The variance of the reference system was calculated as [92]

$$\frac{\mu_1 \mu_3}{\mu_2 \mu_4} \left(1 + \frac{\mu_3}{c_2 \mu_2 + \mu_4} \right).$$

It is composed of a Poissonian and an additional term. The Poissonian term accounts for the intrinsic variance at the protein level, while the additional term captures the impact of the mRNA level. The latter is dominated by the mRNA timescale $c_2 \mu_2$, which is larger than μ_4. As a consequence, we get a significantly higher variance level for the slow mRNA timescale, which is attributed to the effect of the mRNA level on the protein noise. On top of these two reference values of the variance, the modulating environment further increases the variance. We found that the effect was the strongest for a modulation at the mRNA level by R_2, while R_3 and R_4 showed a similar effect. Figure 23E illustrates the broader distribution compared to the reference model for the modulation by Z_4 with maximal σ_4^2 and τ_4. Can the increase be attributing to either an increased environment autocorrelation time τ_j or environment variance σ_j^2? We obtained Monte Carlo estimates of the variance at time $t = 10000$ for different pairs of autocorrelation time and variance, see figure 23B. Interestingly, the simulation studies seem to indicate that the product of τ_j and σ_j^2 majorly determines the variance level. This observation is in line with the increase of the variance for the zero-order modulation.

An exception to this observation is the condition with a fast mRNA timescale and varying R_2 modulation, which shows a steeper increase in the variance compared to the other conditions (fig 23D). Also for the slow mRNA timescale the R_2 modulation shows a deviation compared to R_3 and R_4 in the condition with the maximal σ_2^2 and the highest τ_2. Note the parallel to the protein mean in figure 23C. Sample trajectories for the modulation of R_2 with the largest values of τ_2 and σ_2^2 and for the fast mRNA timescale are depicted in 23F. They are contrasted with trajectories from the reference system. We observe the increase in both the variance and the mean. In summary, we found that in the regime of low environment autocorrelation (the order of the mRNA at most), there is an effect on the variance for all conditions, which becomes significant only when τ_j approaches the autocorrelation time of the reference mRNA level and σ_j^2 approaches the maximal value. For the modulation of the mRNA level, R_2 there was a higher increase compared to R_3 and R_4, whose increase majorly depended on the product of τ_j and σ_j^2. This additional contribution is likely associated with the increased protein mean for this condition. The trajectories look qualitatively the same as for the reference model.

Secondly, we investigated the regimes of medium and high environment autocorrelation. We were interested whether our finding on the variance for the low regime still holds, or whether factors other than the product of τ_j and σ_j^2 determine the variance. The exception of R_2 modulation hints at other contributing factors. For this purpose, we chose two pairs of autocorrelation time and variance, which have the same product, figure 24A. One pair utilizes the maximal variance combined with a medium environment correlation time which is of the order of the reference protein timescale. The other pair utilizes an autocorrelation larger than the autocorrelation time of the protein, combined with a medium value for the variance. In the protein mean, we saw a different behavior for the modulation of death reactions compared to the birth reaction, figure 24B. For R_2, we observed largely increased values in the means for the first (correlation time, variance)-pair compared to the second. For R_4, we observed an increase comparable with the increase for R_2 in the low regime under fast mRNA timescale. In the variance, we noticed different values of both pairs for all reactions, figure 24C and also the autocorrelation exhibits a noticeable difference 24D. Before we investigate these differences further, we first note, that for the mean, it was to be expected that there is no effect when the birth reaction R_3 is modulated. The mean equations read

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[\mathrm{mRNA}(t)] = \mathbb{E}[Z_1(t)] - \mathbb{E}[Z_2(t)\mathrm{mRNA}(t)]$$

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbb{E}[\mathrm{protein}(t)] = \mathbb{E}[Z_3(t)\mathrm{mRNA}(t)] - \mathbb{E}[Z_4(t)\mathrm{protein}(t)].$$

Since the environment component $Z_3(t)$ is independent of $\mathrm{mRNA}(t)$, the first term in the second line equals $\mu_3 \mathbb{E}[\mathrm{mRNA}(t)]$. Hence, the mean equations do not depend on σ_3^2 and γ_3. Consequently,
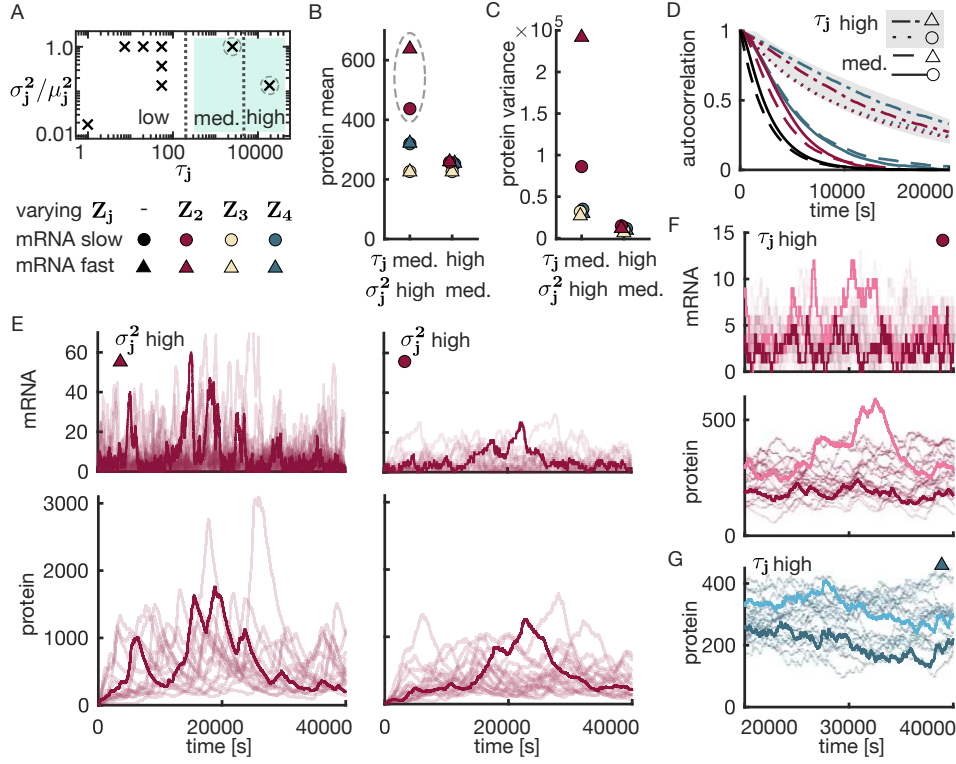
Figure 24: **Gene expression model in a slow random environment.** A. As fig 23B, but with focus on the medium and high correlation time. The mark in the medium regime is the timescale $\tau_j = \mu_4$ of the reference model. We simulated trajectories of the model in fig 23A with the approximate marginal simulation algorithm (tilted Hawkes approximation) for the two marked parameter pairs. Circled marks indicate that E depicts corresponding trajectories. B and C. The protein mean and variance at time $t = 40000$ estimated from 10000 sampled trajectories for the different conditions. Different colors for the same tick were shifted horizontally to improve readability. D. The protein autocorrelation as a function of the lag. Black trajectories indicate the analytically computed autocorrelation of the reference model. Colored curves indicate the correlation coefficient of protein trajectories for the embedded model at different conditions of varying Z_2 and Z_4, estimated from 10000 sampled trajectories. For reference model and medium τ_j, dashed = fast mRNA timescale, solid = slow mRNA timescale. The high correlation regime is shaded by gray (dashed-dotted = fast mRNA timescale, dotted = slow mRNA timescale). E. Sample trajectories of the mRNA and the protein for the maximal σ_2^2 value and medium (protein-like) correlation time τ_2. The left mRNA-protein pair depicts the fast mRNA timescale and the right one depicts the slow mRNA timescale. F. Sample trajectories of the mRNA and the protein for the pair of a high (more-than-protein) correlation time τ_2 and a medium σ_2^2 value with slow mRNA timescale. Trajectories are depicted from $t = 20000$ after stationarity has been reached. G. Sample trajectories of the protein for the pair of a high (more-than-protein) correlation time τ_4 and a medium σ_4^2 value with a fast mRNA timescale, at stationarity as in F. In the shaded medium and high regime of A parameters were $\sigma_j^2/\mu_j^2 = e^{-k}$ and $\tau_j = 2500 \cdot e^k, k = 0, 2$.

the mean protein level is expected to be the same as in the low regime. Now, we turn to the more interesting modulation of the death reactions. For the modulation of R_4, we notice a difference in the protein means. From the sample trajectories, see figure 26A, we conclude that periods of low environment value Z_4 can cause excursions, meaning periods of strong deviations from the mean, in the protein. We further observe that the mean does not depend on the relative timescales of the mRNA and protein. A way to explain this is by arguing that the protein mean depends linearly on the average protein birth rate. The protein birth on average is majorly affected by the mRNA mean. The mRNA however is the same for both the fast and the slow mRNA timescale. Conversely, this means that when we see a difference in the protein mean for the two timescales of the mRNA, this might be attributed to different mean mRNA levels. And indeed, for the modulation of R_2, we see a significant difference between both timescales (figure 24B). We confirmed that this difference can be attributed to a difference in the mean levels of the mRNA, see figure 26B. By looking at sample trajectories in figure 24E, we see excursions at the mRNA level. By the same argument as for the R_4 modulation at the protein level, these can be attributed to phases of low Z_2. To explain the difference between the two mRNA timescales, we see that the excursions are heavier for the fast mRNA timescale (left-hand side of 24E) than for the slow one (right-hand side of 24E). This can be explained by a higher average mRNA birth rate $c_1 \mu_1$ for $c_1 = 1$ compared to $c_1 = 0.1$, while the length of the excursions depends mostly on τ_2, which is the same for both cases. Looking for other quantities that are equal or differ for both cases, we note that the scaled variance compared to the scaled mean squared $c_2^2 \sigma_2^2 / (c_2 \mu_2)^2$ is also the same. However, both cases differ in their values of $c_2^2 \sigma_2^2$, which seem to play a negligible role. The mRNA excursions and their strength directly translate to the protein level (lower panels of 24E). We have thus explained the behavior of the protein mean for the pair of high variance and medium correlation. For the other pair with high correlation and medium variance a similar behavior is not observed, compare figure 24F and G. The lack of this behavior hints at the crucial role of a sufficiently large environment variance in order to expose relevant excursions.

When we now turn towards the variance (fig 24C), we see a very different behavior compared to the regime of low correlation (fig 23D). The variance for the low regime was dominated by the base variance and a contribution by the product of τ_j and σ_j^2, which increased as τ_j approached the mRNA timescale and σ_j^2 its maximal value. For the medium and high regime, we see no significant contribution of the base variance any more. It is still noticeable when we look at the differences between the slow and the fast mRNA timescales for an R_3 and an R_4 modulation. However, what seems to become the dominating factors are the mean level of the protein (largely caused by excursions) and the environment variance. Consequently, the trajectories, figure 24E, look qualitatively different from the reference case.

While for the protein mean and variance, the environment variance σ_j^2 played the essential role, this is not the case when we look at the autocorrelation of protein trajectories, figure 24D. We analytically computed the autocorrelation function for the reference system as in [92], and compared it to the correlation coefficient obtained from 10000 Monte Carlo sample trajectories. We computed the correlation coefficient between the final time point at $t = 40000$ and time points further in the past, increasing the lag from 0 to 20000. To avoid cluttering, we excluded the R_3 modulation. Its behavior is qualitatively the same as R_2 and R_4, see figure 26C. Both the medium and the high regime differed significantly from the reference case, showing a slower autocorrelation decay. When comparing the high regime and the medium regime, the high regime even showed a much slower decay than the medium regime. When we look for differences among the modulation by R_2 and R_4 combined with the slow and fast mRNA timescale, we find that the R_4 modulation in the fast mRNA timescale shows the slowest decay, and the R_2 in the slow regime the least slow one. The sample trajectories, figure 24F and G, indicate that two individual trajectories can remain separate for a rather long time. For the R_4 modulation in the fast mRNA timescale, we see that the trajectories show very little mixing behavior (fig 24G). This illustrates the strong correlation between even far apart time points.

To summarize, we observed an increase in the protein mean, when both the correlation time and the variance of an environment component that modulates a decay reaction, are large enough. We attributed this effect to excursions. We saw a stronger effect for the modulation of the mRNA decay than the protein decay. Interestingly, our findings differ from Keizer et al. [180] who investigated the same gene expression model in a different environment, i.e., using log-normally distributed slow extrinsic noise. For the parameter regime considered in their work, the mean protein levels were the
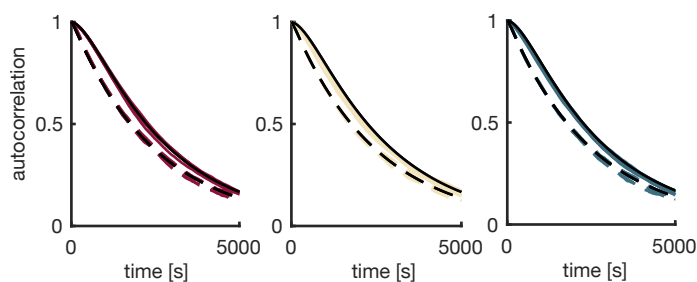
Figure 25: **Details on the gene expression model in the fast environment.** The protein autocorrelation as a function of the lag (dashed = fast mRNA timescale, solid = slow mRNA timescale). Analytical autocorrelation for the reference model (black), shown with the correlation coefficient of protein trajectories for the embedded model (colored) at different conditions (from left to right: varied modulation by Z_2, Z_3, Z_4), estimated from 10000 trajectories. It was computed from the last time point, paired with previous time points.
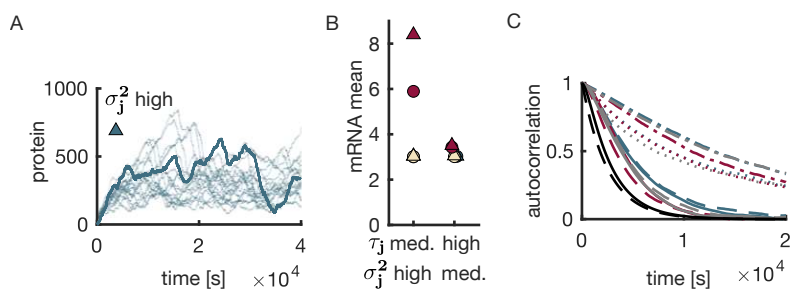

Figure 26: **Details on the gene expression model in the slow environment.** A. Sample trajectories of the protein for the maximal σ_2^2 value and medium (protein-like) correlation time τ_2 with the fast mRNA timescale. B. The mRNA mean at time $t = 40000$ estimated from 10000 sampled trajectories for the different conditions. Different colors for the same tick were shifted horizontally to improve readability. C. The protein autocorrelation as in 24D for all conditions in the medium and high regime.

same regardless of modulating the mRNA decay or the protein decay. In our case study, the increase of the mean contributed to an increase in the variance, which, for the regime of medium to high autocorrelation, is no longer dominated by the base variance and $\sigma_j^2 \times \tau_j$. The protein autocorrelation increases essentially with the autocorrelation of any environment component, with minor differences for the considered conditions.

8.4 Synthetic controller mitigating the heterogeneous degradation rate

One goal of synthetic biology is to design circuits that are robust to environmental changes. The setpoint objective specifies a target copy number or concentration at which the species of interest should be kept robustly, i.e., the concentration is supposed to re-adapt when environmental changes perturb it. Building on the results of the case study 8.3, we considered the setting of a birth-death process X and an environmental birth-death process Z that modulates the degradation of X (Fig. 27a). Additionally, a controller species U senses the environment and acts on the birth rate of X to attenuate the effect and achieve the setpoint objective for X [81, Fig S.9c]. We chose the Q.SS mean as the setpoint, and as the deviation measure we employed the relative deviation

$$\Delta = \frac{\mathbb{E}[X_\infty] - \mathbb{E}[X_\infty^{\mathrm{Q.SS}}]}{\mathbb{E}[X_\infty^{\mathrm{Q.SS}}]}, \tag{257}$$

or the accuracy measure Δ^{-1}, respectively. In the previous sections, we saw that $Z = 0$ can be the main driver of deviations from the Q.SS. The controller U works against this effect. During phases of otherwise unbounded X excursions, the birth rate of X is now down-regulated by U and, in the extreme case, comes to a halt at a plateau.

Compared to environment E2, we replaced the reaction R_1 in Eq. (230) , (235) by

$$R_1: \qquad Z_2 \xrightarrow{c_1} Z_2 + Z_1.$$
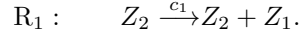
In order to apply our general framework, we regard Z_2 as the one-dimensional environment and (Z_1, X) as the modulated linear CRN. Then $\mathcal{Z} = \mathbb{N}_0$ and Λ is the generator of a birth-death process with birth rate c_3 and death rate c_4. Detailed balance is satisfied by Poisson-distributed π with parameter c_3/c_4. For the dynamics of V we obtain

$$A(z) = \begin{bmatrix} c_2 & 0 \\ -c_5 & c_6 z \end{bmatrix}$$

$$b(z) = \begin{bmatrix} c_1 z \\ 0 \end{bmatrix}.$$

We fix c_6, c_5 and the means $\frac{c_1}{c_2}, \frac{c_3}{c_4}$. For different speeds c_2 of Z_1 we let the speed c_4 of Z_2 vary, see figure 20k. For large c_4 the stationary mean saturates to the Q.SS mean $c_1 c_5/c_2 c_6 = 10$ independent of c_2. The saturation level for small c_4 depends on the speed c_2 of Z_1. For a fast enough Z_1 a local maximum appears. For a slower Z_1 the saturation level increases. The stationary mean curve for small c_2 resembles the leakage case displaying the same four phases.

With stochastically independent birth and death modulation that we considered so far, the stationary mean was not affected by fluctuations in the birth modulation, and we could apply the Q.SS assumption on Z_1. Here, on the contrary, the timescale of the controller species U matters in reacting robustly to the environment. Figure 27b depicts the stationary mean of X for different controller speeds c_2/c_6. The deviation Δ gets more pronounced for a slower controller, whereas a fast controller achieves better accuracy Δ^{-1}. For the effect to become apparent, the environment needs to be sufficiently slow (τ_Z/τ_X large). When the controller operates slowly, it does not have the attenuating effect. In this regime, the target species achieves a base accuracy that depends on a given environment speed and its mean (see Fig. 27c). The base accuracy decreases when the environment gets lower in mean or slower in timescale. In particular, a slow controller achieves worse accuracy in a slow environment compared to a fast one. As a contrary effect, as the controller speeds up, it departs to a better accuracy later in a fast environment than in a slow one. That is why in figure 27c the accuracy curves for the slow and the fast environments intersect, which also explains the local maxima in 27b. The slower environment
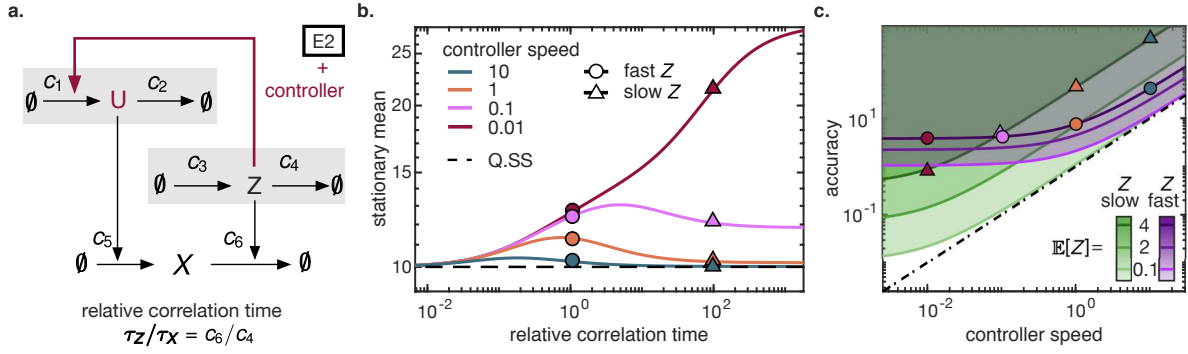
Figure 27: **a.** Reaction scheme for the environment with controller. The birth-death environment E2 is mitigated by a controller that senses the environment and regulates the birth rate of X accordingly. **b.** The stationary mean as a function of the relative correlation time τ_Z/τ_X. Different relative controller speeds c_2/c_6 are plotted. Circles (triangles) indicate fast (slow) environment. **c.** Accuracy Δ^{-1} as a function of controller speed c_2/c_6 for different values of environment mean $\mathbb{E}[Z]$ and speed c_4/c_6. Green (purple) indicates slow (fast) environment. The mean value $\mathbb{E}[Z] = 4$ is highlighted by triangles (slow) or circles (fast) to match **b**. The slow environment achieves lower accuracy than the fast one for inactive controller (controller speed near 0). In order to improve accuracy, less increase in controller speed is needed for slow than for fast environment. As a consequence the accuracy curves for slow and fast environment intersect. The dashed line indicates the the critical controller speed needed to reach a given accuracy level for all environment mean and speed. The slow and low environment exhausts this universal accuracy-controller relation. Parameters were $\mathbb{E}[X_\infty^{Q.SS}] = c_1/c_2 = 10, c_6\mathbb{E}[Z_2] = 4(2, 0.1), c_6/c_4 = 1$ (circles, purple), 100 (triangles, green).
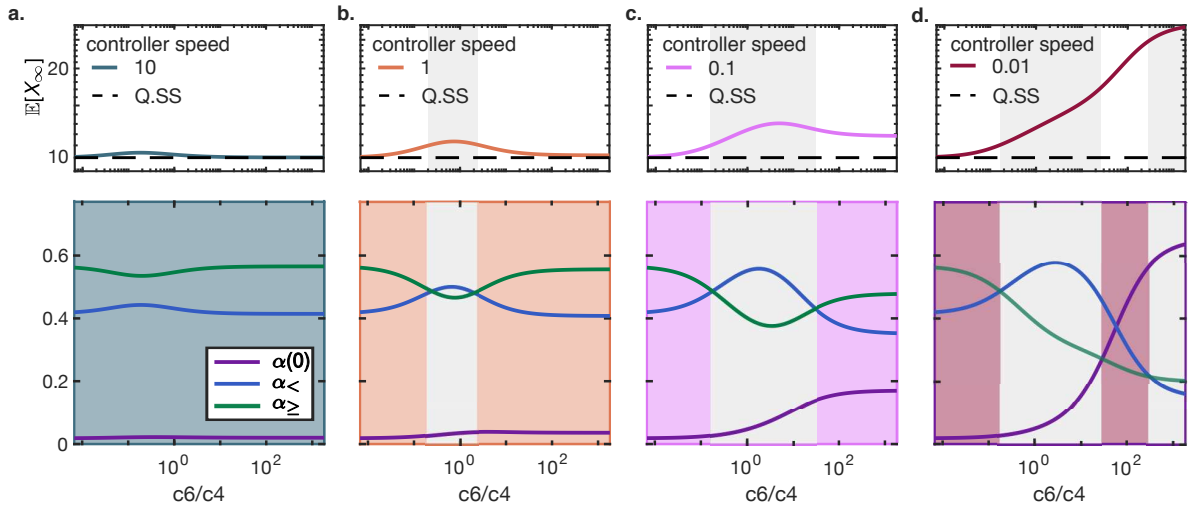

Figure 28: The controller U becomes slower from **a** to **d**. Parameters were $c_5 = 1, c_6 = 1, c_1/c_2 = 10, c_3/c_4 = 4$. For small c_6/c_4 the stationary mean saturates to the Q.SS mean $c_1 c_5/c_2 c_6 = 10$ independent. The saturation level for large c_6/c_4 depends on the speed c_2 of U. For fast enough U a local maximum appears. For slower U the saturation level increases and the contribution $\alpha(0)$ dominates more and more.

- although having a lower base accuracy as a handicap - can be compensated for (in terms of accuracy) already at a slower controller speed. For the details on the shares $\alpha(0), \alpha_{<\bar{z}}, \alpha_{\geq \bar{z}}$, see figure 28.

As a key question, we asked at which timescale the controller must operate to mitigate environmental perturbations with a given accuracy. To formalize this, we request the deviation in Eq. (257) to stay below a critical margin Δ^* or the accuracy to stay above $1/\Delta^*$. Interestingly, we found that for each accuracy margin, a critical controller speed can be chosen that operates universally for all environment speeds and means (see the dashed line in Fig. 27c). For fixed c_5/c_6 and c_1/c_2, the critical relative controller speed is given by

$$c_2^*/c_6 = (\Delta^*)^{-1}. \tag{258}$$

When the environment changes in mean or timescale, the controller holds the accuracy within the tolerated margin. Since the robustness that we analyze is defined via the steady state behavior, our statement is restricted to environmental changes that occur so rarely that the steady state can be reached between changes (see Fig. 29).

Which environment speed and mean exhaust the critical accuracy? It is not the fast and furious (i.e., large mean) environment that causes deviation from Q.SS. For fast environment, the degradation rate of X averages out to the mean, making the Q.SS assumption valid. For large (furious) environment, the main driver of deviation 0 is hardly visited. In addition, the furious environment boils up the controller to act in a regime where it has a higher signal-to-noise ratio. This leaves the slow and low Z to exhaust the critical deviation. During excursions ($Z = 0$), the average dynamics follows

$$\dot{u} = -c_2 u, \quad u(0) = \mathbb{E}[U|Z = 1] = c_1/c_2$$
$$\dot{x} = c_5 u, \quad x(0) = \mathbb{E}[X_\infty^{\text{Q.SS}}].$$

Consequently, the deviation for an infinitely long excursion rises, on average, to the value

$$x(\infty) - \mathbb{E}[X_\infty^{\text{Q.SS}}] = \frac{c_5 c_1}{c_2} \int_0^\infty e^{-c_2 t} \, \mathrm{d}t = \frac{\mathbb{E}[X_\infty^{\text{Q.SS}}] c_6}{c_2}.$$

The excursion becomes the single dominating share as $\mathbb{E}[Z]$ tends to 0, and $\tau_Z \to \infty$ justifies infinitely long excursions. Since this dominating case is linear, the average dynamics is justified and the heuristic derivation of Eq. (258) can be made rigorous.

In summary, we observed that the accuracy margin was exhausted by the regime of slow and low environment, leaving the controller with the simple task to react as fast as to guarantee that a plateau is reached within the tolerated deviation, on average. This finding joins the variations on the theme 'faster sensor molecules achieve higher accuracy'. Note that the setting and the control objective differ from [135]. In the latter, a sensor molecule recorded the progression of the target species, while here it senses the environment. The objective of suppressing fluctuations, i.e., the variance, in the controlled species induced the quartic root law on the sensing event counts. Here, in contrast, we asked for the stationary mean to stay within a tolerance of the setpoint, and this induced an inverse proportional law on the speed with which the controller responds to the environment. To summarize, both findings show that, when under the influence of a random environment, the accuracy can be increased at the cost of a faster sensor molecule. Our finding stands out due to its independence of size and speed of Z and the proportional relation in Eq. (258).

With analytic expressions for the stationary mean at hand, we aim to test whether the impact on the stationary mean is also captured quantitatively by the approximate marginal simulation with the Hawkes approximate filter. Since the joint process (U, Z) is used as the environment for the birth-death process X, this is an example of a birth-death process embedded in a correlated environment. We picked two different controller speeds from the setting of figure 27B. Then the figure 30B depicts the stationary mean as a function of the environment speed for the exact model and the tilted Hawkes model. We recall that the faster controller has the maximal point in the interior. The tilted Hawkes model captures this behavior. In the sample trajectories figure 30C and D of the state estimates, we notice that the fast controller tracks the environment and the slow one only follows with a sincere delay or not at all. Hence, the state estimate mimics capture the behavior of environment and controller qualitatively. We note here that we did not encounter negative values of the state estimates $(\hat{U}, \hat{Z})$.

Returning to the slow controller in figure 30B, we observe a deviation of more than 100% from the set point, which the tilted Hawkes model does not exhibit. To see that the excursions cause the
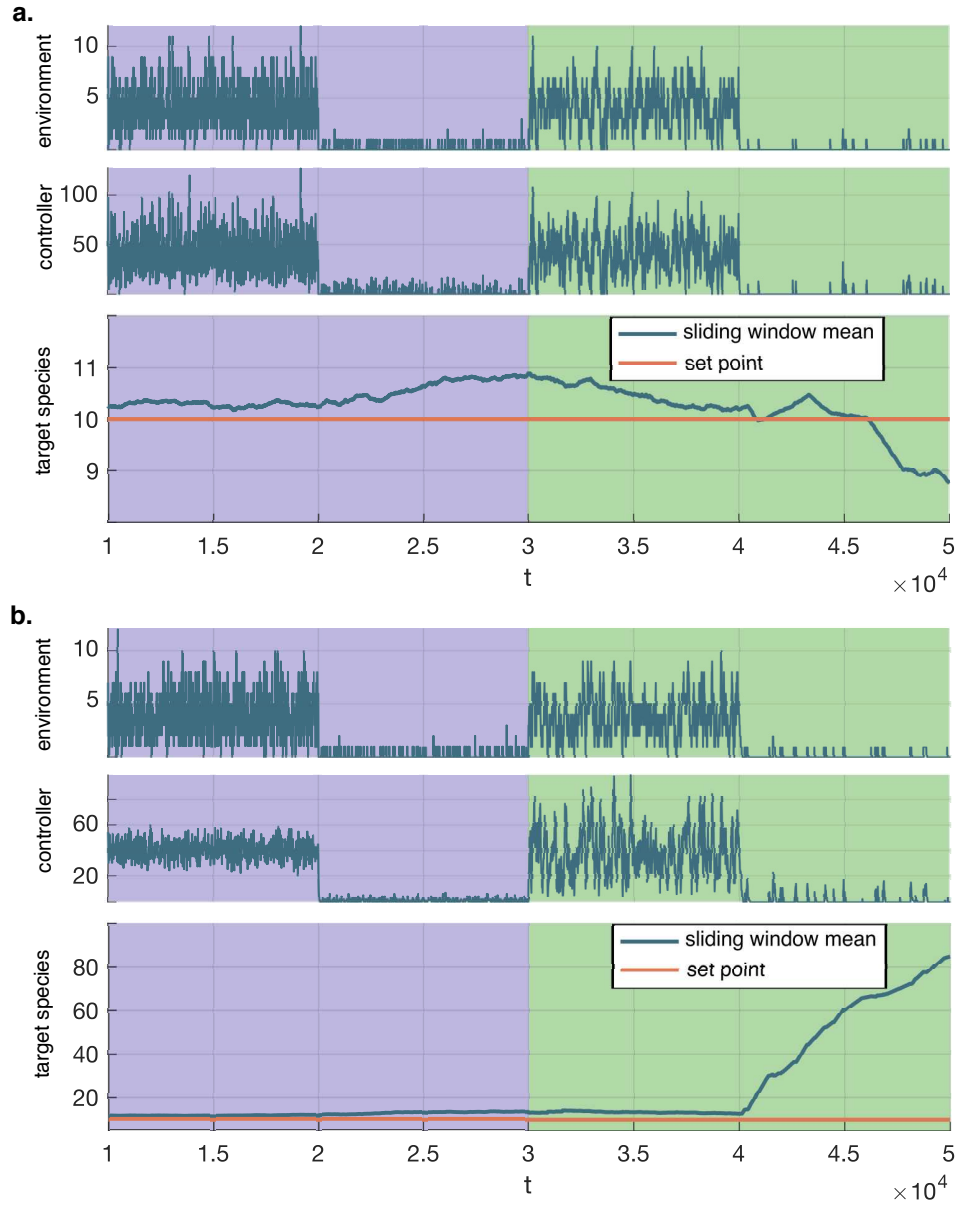
Figure 29: Trajectories for the environment E2 with controller. Environment Z and controller U are shown in plain form, while X is depicted as a moving average with window length $t = 10,000$ to match the statement about the stationary mean. A burn-in time of $t = 10,000$ was used. The environment progresses through four stages of different mean and speed. **a.** A fast controller ($c_2/c_6 = 10$) keeps the target species near the setpoint, on average, as the environment changes. According to the theory the deviation stays within 10% of the setpoint. In the last phase a higher deviation indicates that stationarity is not reached within the window length. The drop below the setpoint is due to a low plateau value in the trajectory that would be counter-balanced by larger plateau values in the long run. **b.** A slow controller ($c_2/c_6 = 0.1$) achieves low accuracy for the critical stage of low and slow environment (last phase). Parameter values were $\mathbb{E}[X_\infty^{Q.SS}] = \mathbb{E}[U] = 10, c_6 = 1$ and in the four phases the four pairs $(\mathbb{E}[Z], c_4) = (4, 3), (0.1, 3), (4, 0.03), (0.1, 0.03)$ were used.
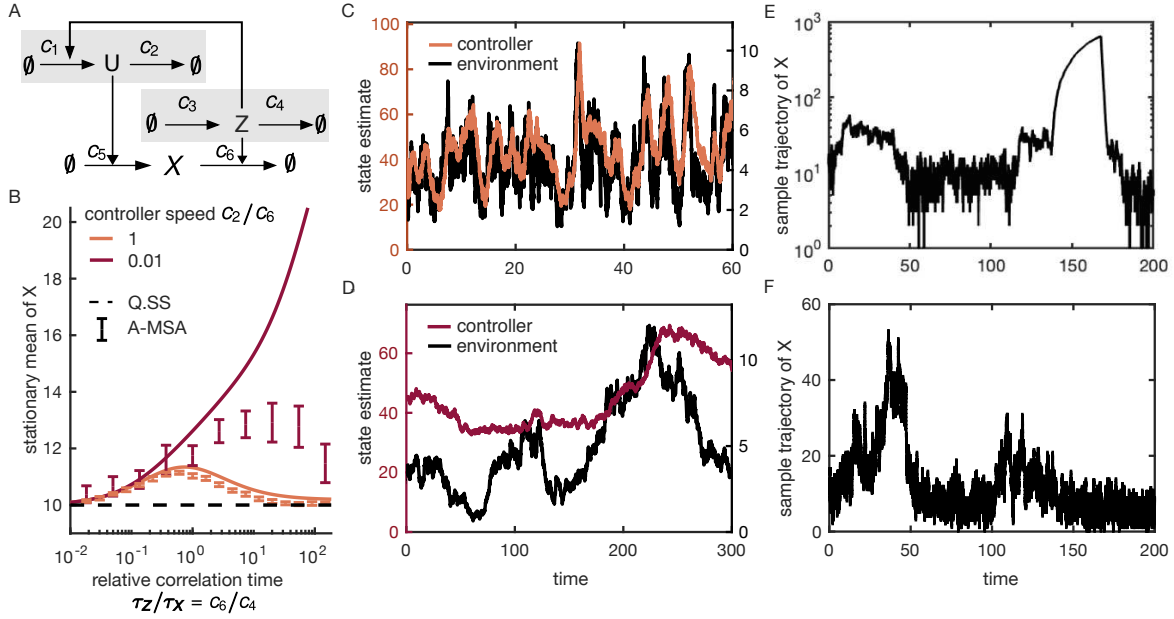
Figure 30: Modulation of first-order reaction. A. A birth-death process X is embedded into a correlated environment (U, Z). The controller U senses the environment component Z which modulates the death rate of X, and it regulates the birth rate of X accordingly to achieve the set point goal of keeping X near $\frac{c_1 c_5}{c_2 c_6}$. B. The stationary mean is plotted for two controller speeds as a function of the environment speed, dictated by the relative correlation time c_6/c_4. The environment becomes slower from left to right. As a reference, the set point (quasi-steady state) is added. The stationary mean serves as a proxy to assess whether the control goal is achieved. The error bars indicate the stationary mean of the tilted Hawkes model (approximate marginal simulation algorithm). For each of the 20 replicates it was computed as the temporal average of a trajectory of $55,000$ $(100,000)$ transitions, truncated by a burn in period of the first 5000 $(50,000)$ transitions for the fast (slow) controller. The 95% bootstrap confidence interval for the mean, using 1000 bootstrap samples of the 20 replicates, is shown. C. Sample trajectory for the approximate state estimates of the controller U and the environment component Z in the system with a slow controller. D. As in C for a slow controller. E. Sample trajectory of X for the slow controller (Doob-Gillespie). F. Sample trajectory of X for the slow controller (approximate marginal simulation). Parameters were $c_1/c_2 = 10, c_3/c_4 = 4, c_5 = c_6 = 1$ and $c_2 = 1, c_6/c_4 = 0.6$ (C), $c_2 = 0.01, c_6/c_4 = 54.6$ (D-F).

deviation from the set point to a large extent, we examined the shares that the different environment states contribute to the stationary mean (fig 28). Also, when we compare sample trajectories of the exact system (fig 30E) and the tilted Hawkes model (fig 30F), we see an excursion occurring for the exact system at around $t = 150$, while the tilted Hawkes model does not show an excursion. We conclude that the tilted Hawkes model underestimates the effect of the excursions on the stationary mean. We did not rule out that for a slower environment, the severe deviation from the set point and the rise to a plateau value for the slow controller can be qualitatively captured. Also, it is not unlikely that the simulation time was too short for the last confidence interval (slowest considered environment, slow controller) in figure 30E to properly capture the mean. When we cross-checked the Monte Carlo estimation procedure by the exact system with the same hyperparameters, the mean was often not captured by the confidence interval. This supports the hypothesis of failed convergence. Beside this possible explanation for the discrepancy between the exact mean and the approximation, we recall that in the case study 8.3.8, we saw the excursions only for sufficiently high values of σ^2. Consulting these findings, we hypothesize that the stationary variance of Z could be too small for excursions to arise. In summary, we see that, on the one hand, the tilted Hawkes model captures important features of the model such as the controller (not) following the environment and the maximal deviation occurring in the interior environment speed. The quantitative deviation attributed to underestimating excursions, on the other hand, poses a clear limitation to the approximation. This comparison also classifies the excursions as imposed by the non-linear dynamics of the environment, as the linearized Hawkes model does not capture them.

8.5 Gamma-modulated reactions

So far, we have considered examples with a dynamic environment. Characteristically, its reaction counter had a variance that is asymptotically linear with respect to time, as discussed in section 3.2. We now assume a static Gamma-distributed environment introduced in section 6.4 causing the variance of Gamma-modulated birth process to increase quadratically with time, see figure 5.

8.5.1 Gamma-modulated decay reaction

We considered a minimal system with only $M = 1$ reaction, a decay $X \to \emptyset$ with rate $Z\delta x$, where Z is Gamma(α, β)-distributed. The scaled random variable δZ is Gamma Gamma$(\alpha, \beta/\delta)$-distributed. Thus, by redefining $\tilde{\beta} = \beta/\delta$, we can assume without loss of generality $\delta = 1$. Then, with $W(t) := \int_0^t X(s)\,\mathrm{d}s$ and $p(t, x) := \mathbb{P}[X(t) = x]$,

$$\partial_t p(t, x) = \mathbb{1}(x < N)\left\langle \frac{\alpha + N - x - 1}{\beta + W(t)}\middle| x + 1, t\right\rangle (x + 1)p(t, x + 1) - \left\langle \frac{\alpha + N - x}{\beta + W(t)}\middle| x, t\right\rangle xp(t, x). \quad (259)$$

We claim with $X(0) = N$,

$$\left\langle \frac{\alpha + N - x}{\beta + W(t)}\middle| x, t\right\rangle = \frac{\alpha \sum_{k=0}^{N-x}(-1)^k \binom{N-x}{k}(\beta + (x + k)t)^{-(\alpha+1)}}{\sum_{k=0}^{N-x}(-1)^k \binom{N-x}{k}(\beta + (x + k)t)^{-\alpha}}. \quad (260)$$

The derivation uses the joint conditional distribution of the decay time points $(T_1, \ldots, T_{N-x})$ conditional on $\{X(t) = x\} = \{T_{N-x} \le t, T_{N-x+1} > t\}$, derived with Bayes rule as follows

$$
\begin{aligned}
&p_{T_1, \ldots, T_{N-x}}(t_1, \ldots, t_{N-x} \mid x, t) \\
&\propto \mathbb{P}(T_{N-x+1} > t \mid T_1 = t_1, \ldots, T_{N-x} = t_{N-x}) p_{T_1, \ldots, T_{N-x}}(t_1, \ldots, t_{N-x}) \\
&= \int_t^\infty \int_0^\infty p_{T_{N-x+1}}(t_{N-x+1} \mid T_1 = t_1, \ldots, T_{N-x} = t_{N-x}, Z = z) \times \\
&\qquad\qquad p_{T_1, \ldots, T_{N-x}}(t_1, \ldots, t_{N-x} \mid Z = z) p_Z(z) \; \mathrm{d}z \; \mathrm{d}t_{N-x+1} \\
&\propto \int_t^\infty \int_0^\infty \prod_{i=0}^{N-x} z(N-i) e^{-z(N-i)(t_{i+1}-t_i)} e^{-z\beta} z^\alpha \; \mathrm{d}z \; \mathrm{d}t_{N-x+1} \\
&\propto \int_t^\infty \int_0^\infty \exp\left(-z\left(\beta + \sum_{i=0}^{N-x}(N-i)(t_{i+1}-t_i)\right)\right) z^{(\alpha+(N-x+1)} \; \mathrm{d}z \; \mathrm{d}t_{N-x+1} \\
&\propto \int_t^\infty \left(\beta + \sum_{i=0}^{N-x}(N-i)(t_{i+1}-t_i)\right)^{-(\alpha+N-x+1)} \mathrm{d}t_{N-x+1} \\
&= \int_t^\infty (\beta + t_1 + \cdots + t_{N-x} + x t_{N-x+1})^{-(\alpha+N-x+1)} \; \mathrm{d}t_{N-x+1} \\
&\propto (\beta + t_1 + \cdots + t_{N-x} + xt)^{-(\alpha+N-x)}.
\end{aligned}
$$

By $\propto$ we mean that there is constant depending on t, x, N, α, β, such that multiplication with this constant turns the relation into an equality. Define the integrals

$$
C_{n,\alpha,\beta}(t_0) := \frac{\Gamma(\alpha)}{\Gamma(\alpha-n)} \int_0^{t_0} \cdots \int_0^{t_{n-1}} (\beta + t_1 + \cdots + t_n)^{-\alpha} \; \mathrm{d}t_n \ldots \; \mathrm{d}t_1, \tag{261}
$$

which by induction can be shown to evaluate to

$$
C_{n,\alpha,\beta}(t_0) = \frac{1}{n!} \sum_{k=0}^n (-1)^k \binom{n}{k} (\beta + k t_0)^{-\alpha+n}.
$$

Then, with $W(t) = xt + T_1 + \cdots + T_{N-x}$,

$$
\begin{aligned}
&\left\langle \frac{\alpha+N-x}{\beta+W(t)} \middle| x, t \right\rangle \\
&= (\alpha+N-x) \cdot \left(\int_0^t \int_0^{t_{N-x-1}} \cdots \int_0^{t_1} p_{T_1, \ldots, T_{N-x}}(t_1, \ldots, t_{N-x} \mid x, t) \; \mathrm{d}t_1 \ldots \; \mathrm{d}t_{N-x} \right)^{-1} \times \\
&\quad \int_0^t \int_0^{t_{N-x-1}} \cdots \int_0^{t_1} (\beta + xt + t_1 + \cdots + t_{N-x})^{-1} p_{T_1, \ldots, T_{N-x}}(t_1, \ldots, t_{N-x} \mid x, t) \; \mathrm{d}t_1 \ldots \; \mathrm{d}t_{N-x} \\
&= \frac{\alpha C_{N-x,\alpha+N-x+1,\beta+xt}(t)}{C_{N-x,\alpha+N-x,\beta+xt}(t)} \\
&= \frac{\alpha \sum_{k=0}^{N-x}(-1)^k \binom{N-x}{k}(\beta+(x+k)t)^{-(\alpha+1)}}{\sum_{k=0}^{N-x}(-1)^k \binom{N-x}{k}(\beta+(x+k)t)^{-\alpha}}.
\end{aligned}
$$

Then, Eq. (259) is solved by

$$
p(t, x) = \binom{N}{x} \sum_{k=0}^{N-x}(-1)^k \binom{N-x}{k}\left(1 + \beta^{-1}(x+k)t\right)^{-\alpha} = \frac{N! \beta^\alpha}{x!} C_{N-x,\alpha+N-x,\beta+xt}(t), \tag{262}
$$

which can be checked by differentiating. We recognize the discrete distribution on $x \in \{0, \ldots, N\}$ with

$$
p(x \mid \alpha, \gamma, N) = \binom{N}{x} \sum_{k=0}^{N-x}(-1)^k \binom{N-x}{k}(1+(x+k)\gamma)^{-\alpha}
$$

as the Gamma-binomial distribution $\text{GammaBin}(\alpha, \gamma, N)$ with the Gamma parameters (α, γ^{-1}) and N the number of trials [181]. This is a binomial mixture distribution, where the success parameter is e^{-Z} with Gamma-distributed Z, i.e., the conditional $p(x|z)$ is the binomial distribution with parameters (e^{-z}, N).

For the mean of a $\text{GammaBin}(\alpha, \gamma, N)$-distributed X, we find

$$
\begin{aligned}
\mathbb{E}[X] &= \sum_{x=1}^{N} x p(x \mid \alpha, \gamma, N) \\
&= \sum_{x=1}^{N} \binom{N}{x} x \sum_{k=0}^{N-x} (-1)^k \binom{N-x}{k} (1 + (x+k)\gamma)^{-\alpha} \\
&= N \sum_{x=1}^{N} \binom{N-1}{x-1} \sum_{k=0}^{N-x} (-1)^k \binom{N-x}{k} (1 + (x+k)\gamma)^{-\alpha} \\
&= N \sum_{x=0}^{N-1} \binom{N-1}{x} \beta^\alpha \sum_{k=0}^{N-(x+1)} (-1)^k \binom{N-(x+1)}{k} (1 + (x+1+k)\gamma)^{-\alpha} \\
&= N \sum_{x=0}^{N-1} \binom{N-1}{x} \beta^\alpha \sum_{k=0}^{N-1+x} (-1)^k \binom{N-1+x}{k} (1 + \gamma + (x+k)\gamma)^{-\alpha} \\
&= N (1+\gamma)^{-\alpha} \sum_{x=0}^{N-1} p(x \mid \alpha, \gamma/(1+\gamma), N-1) \\
&= N (1+\gamma)^{-\alpha}.
\end{aligned}
$$

Thus, with $\gamma = \beta^{-1} t$, the mean $E[X(t)] = N(1 + \beta^{-1}t)^\alpha$ satisfies the evolution equation

$$
\frac{\mathrm{d}}{\mathrm{d}t} E[X(t)] = \frac{-\alpha}{\beta + t} E[X(t)]. \tag{263}
$$

Similarly, for the second falling moment of a $\text{GammaBin}(\alpha, \gamma, N)$-distributed X, we obtain

$$
\mathbb{E}[X(X-1)] = N(N-1)(1+2\gamma)^\alpha, \tag{264}
$$

which implies for the variance

$$
\mathbb{V}\mathrm{ar}[X] = N(N-1)(1+2\gamma)^\alpha + N(1+\gamma)^\alpha - N^2(1+\gamma)^{2\alpha}, \tag{265}
$$

The distribution $\text{GammaBin}(\alpha, \beta^{-1}t, N)$ was to be expected for $p(x, t)$, since conditional on $Z = z$, each of the N molecules degrades independently of the other molecules and the probability of having escaped degradation after time t is e^{-zt}. The transformed random variable tZ is precisely Gamma-distributed with parameters $(\alpha, \beta t^{-1})$.

In the limit for $\alpha \to \infty, \gamma \to 0$ with $\alpha\gamma \to \mu$ the distribution $\text{GammaBin}(\alpha, \gamma, N)$ converges to a binomial distribution with parameters $(e^{-\mu}, N)$ with corresponding mean $Ne^{-\mu}$ and variance $Ne^{-\mu}(1 - e^{-\mu})$.

8.5.2 Gamma-modulated conversion reaction

In the previous subsection, we obtained a Gamma-binomial distribution for the Gamma-modulated decay reaction. Along the same lines of argument, we obtain a binomial mixture distribution for the conversion process

$$
\text{X}^* \underset{\text{Z}}{\overset{\text{r}}{\rightleftharpoons}} \text{X}
$$
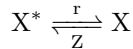
that is initialized in the binomial distribution with parameters (θ_0, N). Again, the Gamma-distributed Z is transformed to $[0, 1]$ to serve as a success probability. However, this time the transform is not $Z \mapsto e^{-tZ}$, but

$$
\text{Z} \mapsto \theta_0 e^{-(r+Z)t} + \frac{r}{r+Z}(1 - e^{-(r+Z)t}),
$$

such that the limit distribution for $t \to \infty$ is a binomial mixture distribution with success probability $\frac{r}{r+Z}$ for Gamma-distributed Z.

8.5.3 Gamma-modulated birth-death reaction

Taking the limit $N \to \infty, r \to 0$ while $rN \to \lambda$, we obtain a birth-death process with birth rate λ and Gamma-modulated decay. Suppose, the initial distribution is a Poisson distribution with mean λ_0, then $p(x,t)$ is a Poisson mixture distribution, where the parameter is

$$\lambda_0 e^{-Zt} + \frac{\lambda}{Z}(1 - e^{-Zt})$$

with Gamma-distributed Z. Its limit distribution for $t \to \infty$ is a Poisson mixture distribution with parameter $\frac{\lambda}{Z}$ for Gamma-distributed Z. This is different from a birth-death process with Gamma-distributed birth rate and decay rate κ, whose distribution $p(x,t)$ is a Poisson mixture distribution with parameter

$$\lambda_0 e^{-\kappa t} + \frac{Z}{\kappa}(1 - e^{-\kappa t})$$

for Gamma-distributed Z. For $\lambda_0 = 0$, this is a Gamma-Poisson distribution, more commonly known as negative binomial distribution.

9 Information-theoretic results

There is mounting evidence that the information encoded in the temporal concentration profiles of biomolecules plays a key role in cellular sensing and decision making [182, 183] and helps to overcome biochemical noise [184]. The computation of mutual information (MI) between time-varying, biomolecular signals is complex, and analytical solutions so far relied on Gaussian [143] or steady-state approximations [185]. Other papers based the MI on single time-point marginals ignoring any encoding in the temporal profile, e.g., [138, 137]. Works, that account for the discrete nature of chemical reactions, commonly assume diffusion approximations as inputs [135, 186] or are based on stochastic simulation [141, 83, 140]. Restrictions to sub-classes of discrete-state input processes permit analytical bounds on the capacity [153], often challenging diffusion based results [136].

9.1 Path mutual information in cell signaling

Our particular interest lies in studying the MI between time-varying signals in the biological cell. It is widely assumed that the optimization of the MI could be an evolutionary strategy to mitigate noise in the cellular signaling via metabolites and in gene regulatory networks [139]. The exact mechanisms how temporal features result in decision making are unclear [187]. A path MI that is sensitive to temporal effects has been recently introduced for a class of chemical reaction networks [83]. The MI is often interpreted as a measure for the amount of input states that can be resolved accurately [188]. From an engineering perspective maximizing the path MI between sensor and actuator could thus define a design principle to construct cellular circuits. Information theory has been used to reveal fundamental physical limits of information transmission guiding our understanding of gene regulatory motifs as information processing units [135, 189].

Why is the Poisson channel used as a communication model in cells? Transcription, i.e., the process of synthesizing mRNA, can be modeled as a counting process that records the number of completed mRNA molecules. In promoter-mediated transcription [24] the accessibility of the promoter determines at which rate molecules are transcribed. The promoter state is then modeled as a finite-state CTMC combined with an intensity function and the transcription as its corresponding doubly stochastic Poisson process. The classical promoter model is the two-state random telegraph process (Fig. 1a), i.e., Markovian switching between a high and low or zero transcription rate [18], but recently more general models were studied [22, 55]. Among them the class with more than two states and a binary intensity function separating the states into On states (high) and Off states (low or zero) was considered [190]. The directed graph that captures possible transitions between the On and Off states is called the promoter architecture. In particular, the model class contains the two-state non-Markov models with phase-type sojourn time distribution. Constituting the smallest building block of gene regulation, the promoter-mediated transcription is hypothesized to be optimized in an information-theoretic sense. In order to find the optimal promoter architecture under appropriate physical constraints an efficient computation of the MI between promoter state and mRNA transcription events is required.

9.2 Poisson channel with binary input and average sojourn time constraints

The Poisson channel was introduced as a model for direct-detection optical communication systems [191]. A random telegraph wave switches a light-source on and off and photons are detected by a photosensitive device. Upon entry of photons to the detector, photoelectrons are generated. The number of detection events are counted by $(Y_t)_{t \geq 0}$ according to a Markov-modulated Poisson process with an intensity that follows the random telegraph wave. Additionally, due to a random generation of electrons, a small electric current flows in the device. We model the input $(Z_t)_{t \geq 0}$ as a random telegraph model on $\mathcal{Z} := \{0, 1\}$ with On and Off rates c_1, c_2.

Regarding the question of how the biological cell operates under uncertainty, we analyzed the minimal gene expression model with a two-state promoter (see e.g. [100] and Fig. 2a in [3]) as an analytically tractable example. Despite our biological interpretation of the mathematics, the following analysis is transferable to other such applications. Switching stochastically between inactivated state z_1 and activated state z_2, the promoter is modeled by a stationary random telegraph process $Z(t)$, i.e., a binary, time-homogeneous, stationary Markov process (BMP). Its state linearly modulates the synthesis rate of messenger RNA (mRNA) molecules $Y(t)$. The decay of mRNA molecules can be ignored from an information theoretic point of view, because birth events are uniquely identified from birth-death trajectories [83]. The joint distribution of (Z, Y), that factorizes in the conditional $Y|Z$ and the input path distribution μ_Z, is equivalent to the Poisson channel, whose class of input processes $Z(t)$ is restricted to BMPs. We distinguish between channels with leakage ($z_1 > 0$) and without leakage (zero dark current case $z_1 = 0$), see Fig. 33a. By $Z_{[0,T]}$ we denote the trajectory $Z(t)_{0 \leq t \leq T}$ of the time-varying input signal with transmission duration T. The sojourn times σ_1 in z_1 and σ_2 in z_2 are exponentially distributed with parameters c_1 and c_2 such that $\mathbb{E}[\sigma_1] = c_1^{-1}$ and $\mathbb{E}[\sigma_2] = c_2^{-1}$. The channel output $Y(t)$ fires at rate $c_3 Z(t)$, where c_3 is the channel gain that dictates the timescale of $Y(t)$. We denote the switching times of Z as s_i with $0 < s_1 < \cdots < T$ and the jump times of Y as t_i with $0 < t_1 < \cdots < T$.

9.2.1 Capacity of the Poisson channel and optimal allocation of binary input

When it comes to computing the capacity of the Poisson under constraints, classically, the input peak and average power are constrained. Then among general inputs the class of BMPs achieves capacity, however at the cost of infinite switching rates c_1, c_2 [61, 145]. This physical implausibility motivates bandwidth-like constraints [146, 82]. By restricting general signals to binary inputs with $|s_i - s_{i+1}| \geq \Delta$ for some $\Delta > 0$, [147] reported a transition from asymmetric to symmetric allocation of the states z_1 and z_2 for the capacity-achieving input, as Δ increases. We consider the Poisson channel with BMPs as input class and investigate the following bandwidth-like constraint (B1) $0 < c_1 \leq r_1, 0 < c_2 \leq r_2$, i.e., lower-bounding the average sojourn times by $\mathbb{E}[\sigma_1] \geq r_1^{-1}, \mathbb{E}[\sigma_2] \geq r_2^{-1}$. A special case is the homogeneous constraint $\max\{c_1, c_2\} \leq r_0$, in analogy to [147]. We consider the path-wise MI $\mathbb{I}(Z_{[0,T]}, Y_{[0,T]})$ [148] and the information rate, defined as $\bar{\mathbb{I}}(Z, Y) := \lim_{T \to \infty} T^{-1} \mathbb{I}(Z_{[0,T]}, Y_{[0,T]})$. Optimizing (i) the MI and (ii) the information rate among all admissible input path distributions μ_Z yields (i) the capacity C_T and (ii) the information rate capacity C. For fixed z_1, z_2, the distribution μ_Z is parametrized by the system parameters $c_1, c_2 > 0$. Fixing c_3, the capacity-achieving distributions are characterized by the following constrained optimization problems

$$\text{(i)} \sup_{c_1, c_2} \frac{1}{T} \mathbb{I}(Z_{[0,T]}, Y_{[0,T]}), \quad \text{(ii)} \sup_{c_1, c_2} \bar{\mathbb{I}}(Z, Y)$$

subject to the constraint (B1). Applying Jensen's inequality to the second integrand term of Eq. (119) and using the identity $\phi(cz) = c\phi(z) + z\phi(c)$ as well as the stationarity of BMPs, the MI is bounded by

$$\frac{\mathbb{I}(Z_{[0,T]}, Y_{[0,T]})}{c_3 T} \leq (1-m)\phi(z_1) + m\phi(z_2) - \phi(z_1 + m\Delta z) =: J(z_1, z_2, m),$$

where $\Delta z := z_2 - z_1$ is the dynamic range and $m := P[Z(t) = z_2] = c_1/(c_1 + c_2)$ coincides with the average power $\mathbb{E}[Z(t)]$ for a normalized signal $z_1 = 0, z_2 = 1$. Optimizing over the general average power constraint $0 < z_1 + \Delta z \cdot m \leq z_1 + \Delta x \cdot p_0$ implies

$$C_T \leq c_3 J(z_1, z_2, \min\{p_0, \bar{p}\}), \tag{266}$$

Figure 31: Poisson channel with Two-State input reduced to binary random variables on small intervals $[t, t + \Delta]$

where $\bar{p} = \left\{ \exp\left(\Delta z^{-1}(\phi(z_2) - \phi(z_1)) - 1 \right) - z_1 \right\} / \Delta z$.

Kabanov and Davis showed that for $c_1, c_2 \to \infty$ the bound (266) is indeed achieved with the asymptotic ratio $m = \min\{p_0, \bar{p}\}$ [61, 145]. The case $z_1 = 0, z_2 = 1$ (no leakage) with $p_0 \geq 1/e$ reduces to $C = C_T = c_3/e$ with $m = 1/e$. Why is the asymmetric allocation optimal for large switching rates?

In the following we caricature the Poisson channel as an input-output-pair of binary random variables. The idea is that the number of jumps of the channel output on small time intervals is essentially a Bernoulli variable, i.e. only 0 or 1. In order to regard the input as a Bernoulli variable as well, we split the time axis in small intervals, on which the input is constant. If the switching rates are large, this corresponds to choosing the inputs on consecutive time intervals independently. This justifies that the following picture captures the path mutual information by merely investigating a single small time interval. Now we turn to a more formal description.

Consider an equidistant discretization of the time axis of length Δ. Suppose on each interval $[t, t + \Delta], [t + \Delta, t + 2\Delta], \dots$ the input Z is constant. Later, we will let $\Delta \to 0$. We write $Z_{t\dots t+\Delta}$ for the value of Z on the interval $[t, t+\Delta]$. Consider additionally the output $Y[t, t+\Delta] := Y(t+\Delta) - Y(t)$. Figure 31 summarizes the (conditional) probabilities of the two-stage experiment: $Z_{t\dots t+\Delta}$ takes the values z_1 and z_2 with probabilities p and $1 - p$. $Y[t, t + \Delta] | Z_{t\dots t+\Delta} = z_i$ takes the values 0 and 1 with probabilities $1 - z_i \Delta + o(\Delta)$ and $z_i \Delta + o(\Delta)$ and all larger values with probabilities $o(\Delta)$.

Now, we aim for the mutual information of the two binary random variables $Z := Z_{t\dots t+\Delta}$ and $Y := Y[t, t + \Delta]$ up to $o(\Delta)$. We will first derive the result for $z_1 = 0, z_2 = 1$, and then state it for general z_i.

We decompose the mutual information $\mathbb{I}(Z, Y) = H(Z) - H(Z|Y)$, which is shown in figure 32. The entropy

$$H(Z) = -p \ln(p) - (1 - p) \ln(1 - p) \tag{267}$$

of the input Z is maximized by $p = \frac{1}{2}$. The evaluation of the conditional entropy $H(Z|Y)$ requires the knowledge of $P[Z = z_i | Y = 0]$ and $P[Z = z_i | Y = 1]$. Using the notation $p(z|y) := P[Z = z | Y = y]$, we have

$$p(1|1) = 1, \quad p(0|1) = 0, \quad p(1|0) = \frac{p(1 - \Delta)}{1 - p\Delta}, \quad p(0|0) = \frac{1 - p}{1 - p\Delta}.$$

Then

$$-H(Z|Y) = \sum_{z \in \{z_1, z_2\}, y \in \{0,1\}} p(z, y) \ln(p(z|y))$$

$$= p\Delta \ln(1) + p(1 - \Delta) \ln\left(\frac{p(1 - \Delta)}{1 - p\Delta} \right) + (1 - p) \ln\left(\frac{1 - p}{1 - p\Delta} \right)$$

$$= p(1 - \Delta) \ln(p) - p\Delta + p^2\Delta + (1 - p) \ln(1 - p) + (1 - p)p\Delta + o(\Delta)$$

$$= p \ln(p) + (1 - p) \ln(p) - \Delta p \ln(p) + o(\Delta)$$

where we used the logarithm laws and the expansion $\ln(1 - h\Delta) = -h\Delta + o(\Delta)$. Adding terms, we get

$$\mathbb{I}(Z, Y) = H(Z) - H(Z|Y) = -\Delta p \ln p + o(\Delta).$$

Hence, the mutual information is of order Δ and the instantaneous gain is maximized for $p = \exp(-1)$ (Fig. (32)). This explains the asymmetric allocation that favors the state $z_1 = 0$.
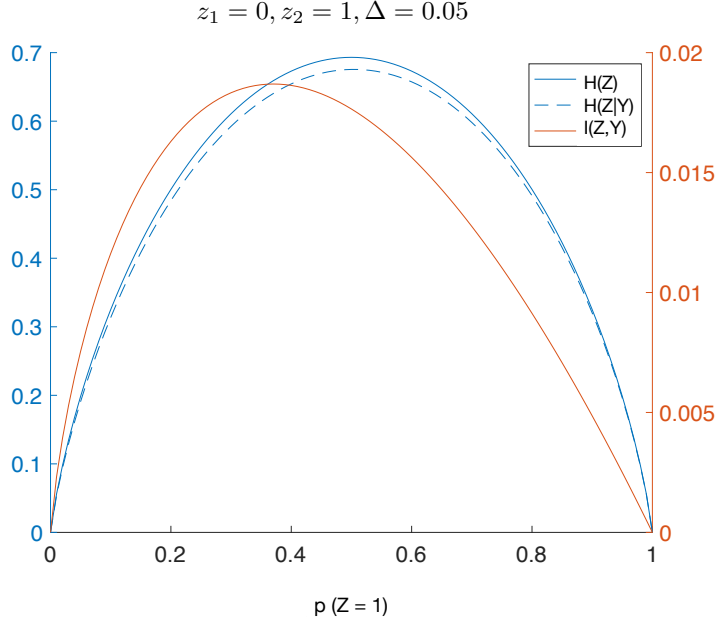
Figure 32: The entropy, conditional entropy and mutual information for the discrete caricature of the Poisson channel with Two-State input. The mutual information $I(Z,Y)$ is maximized for $p(Z=1) = \exp(-1)$.

For general z_i it holds that

$$\mathbb{I}(Z,Y) = \Delta(pz_2 \ln z_2 + (1-p)z_1 \ln z_1 - (p_2 z_2 + (1-p)z_1) \ln(p_2 z_2 + (1-p)z_1)) + o(\Delta)$$

Define the bracket term as $m_1 = p_2 z_2 + (1-p)z_1$, such that for the mean it holds $\mathbb{E}[Y] = \mathbb{P}[Y = 1] = \Delta m_1$. Using the logarithm laws and the expansion $\ln(1 - h\Delta) = -h\Delta + o(\Delta)$ we compute the conditional entropy

$$
\begin{aligned}
-H(Z|Y) &= \sum_{z \in \{z_1, z_2\}, y \in \{0,1\}} p(z,y) \ln(p(z|y)) \\
&= pz_2 \Delta \ln\left(\frac{pz_2 \Delta}{m_1 \Delta}\right) + (1-p)z_1 \Delta \ln\left(\frac{(1-p)z_1 \Delta}{m_1 \Delta}\right) \\
&\quad + p(1 - z_2 \Delta) \ln\left(\frac{p(1 - z_2 \Delta)}{1 - m_1 \Delta}\right) + (1-p)(1 - z_1 \Delta) \ln\left(\frac{(1-p)(1 - z_1 \Delta)}{1 - m_1 \Delta}\right) \\
&= pz_2 \Delta \ln(p) + pz_2 \Delta \ln(z_2) + (1-p)z_1 \Delta \ln(1-p) + (1-p)z_1 \Delta \ln(z_1) \\
&\quad + p(1 - z_2 \Delta) \ln(p) - pz_2 \Delta + (1-p)(1 - z_1 \Delta) \ln(1-p) - (1-p)z_1 \Delta \\
&\quad - ((pz_2 + (1-p)z_1)\Delta \ln(m_1) - pm_1 \Delta - (1-p)m_1 \Delta) + o(\Delta) \\
&= p \ln(p) + (1-p) \ln(1-p) + \Delta(pz_2 \ln z_2 + (1-p)z_1 \ln z_1 - m_1 \ln(m_1)) + o(\Delta) \\
&= -H(Z) + \Delta(pz_2 \ln z_2 + (1-p)z_1 \ln z_1 - m_1 \ln(m_1)) + o(\Delta)
\end{aligned}
$$

By the decomposition $I(Z,Y) = H(Z) - H(Z|Y)$ and definition of m_1 the claim follows.

9.2.2 Average sojourn time constraints

For the Poisson channel with a random telegraph input, a bandwidth-like constraint is imposed to avoid the physically implausible capacity-achieving input with infinitely fast switching rates. We thus analyze how $\bar{\mathbb{I}}(Z,Y)$ depends on the system parameters c_1, c_2, c (as in the Eq. (220)) with the goal to answer questions of optimality under constraints $0 < c_1 \le r_1, 0 < c_2 \le r_2$. Continuing with Eq. (220) for vanishing dark current, $\lambda_0 = 0$, we consider the Dirac-PDMP $(\Pi_t(1))_{t \ge 0}$ on $[0,1]$, which is
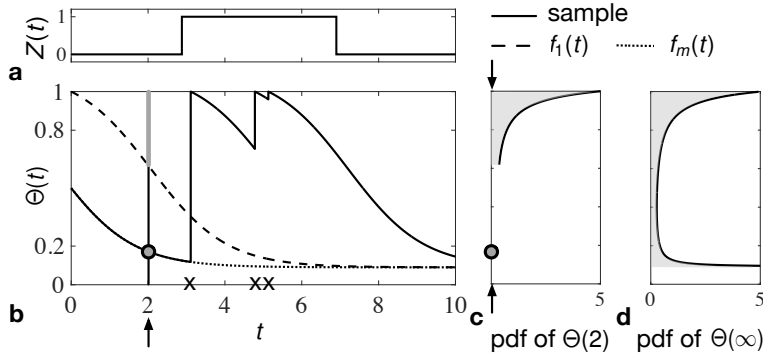
Figure 33: Dynamics of the causal estimator $\Theta(t)$ for $c_1 = c_2 = 0.1, c = 1, z_1 = 0, z_2 = 1$. **a** Sample trajectory of the binary signal input $Z(t)$. **b** The solid line is a sample trajectory of $\Theta(t)$. Crosses indicate the jump times t_1, t_2, t_3 of Y. The dotted line $f_m(t)$ is the common trajectory of $\Theta(t)$ prior to the first jump. The dashed line $f_1(t)$ separates the (t, θ)-plane. Prior to the first jump the trajectories evolve below, afterwards they evolve above. **c** The probability distribution of the causal estimator at time $t = 2$ is composed of a Dirac measure with weight $\kappa(2)$ at $f_m(2)$ and a density supported on $(f_1(2), 1]$. **d** Probability density of the asymptotic causal estimator.

the normalized version of $\hat{\lambda}_t = \lambda_1 \cdot \Pi_t(1)$. Its hazard is $l(\pi) = c\pi$ with $\pi := \pi_1$ and $c := \Delta\lambda = \lambda_1$ for simplicity. Furthermore, we consider the normalized parameters $\tilde{c}_1 := c_1/c, \tilde{c}_2 := c_2/c$ and, for convenience, drop the tilde again. Let $\omega_1 < \omega_2$ be the roots of the quadratic equation at equilibrium

$$0 = c(c_1 - (c_1 + c_2 + 1)\omega + \omega^2)$$

with difference $\Delta\omega := \omega_2 - \omega_1$. We now use the results of sections 4.2.1 and 4.2.2.

9.2.3 Mutual information for the Poisson channel with binary Markovian input

First, we present the transient results. Since all trajectories $\Theta(t)$ start in the stationary mean $\mathbb{E}[\Theta(0)] = \mathbb{E}[Z(0)] = c_1/(c_1 + c_2) = m$, the probability distribution of $Z(0)$ is a point mass. In order to track its progression, we switch to the trajectory-wise perspective (220). Solving the deterministic part of (220), i.e. the Riccati equation $\frac{d}{dt} f_a(t) = A(f_a(t)), f_a(0) = a$, yields

$$f_a(t) = \omega_1 + \Delta\omega(1 + e^{\Delta\omega ct}(\Delta\omega/(a - \omega) - 1))^{-1}$$

$$\Theta(t) = f_m(t)\mathbb{1}_{[0,t_1)}(t) + \sum_i f_1(t - t_i)\mathbb{1}_{[t_i, t_{i+1})}(t)$$

For each t the ensemble of trajectories is partitioned by $f_1(t)$ into the ones that have and those that have not yet jumped:

$$t < t_1 \Leftrightarrow Z(t) \le f_1(t) \Leftrightarrow \Theta(t) = f_m(t). \tag{268}$$

Returning to the distribution perspective, Fig. 33c visualizes the ensemble of trajectories stopped at a fixed t, while Fig. 33d visualizes the asymptotic distribution. The following theorem fully describes the distribution of $\Theta(t)$. Besides preparing the main result, theorem 9.2, it can be interesting in its own right in the related fields of filtering and control theory.

Theorem 9.1. *The probability measure* $\mu_t \colon \mathcal{B}(\omega_1, 1] \to [0, 1], \mu_t(B) = \mathbb{P}[Z(t) \in B]$, *defined for Borel sets* $B \subseteq (\omega_1, 1]$, *is a hybrid measure*

$$\mu_t(B) = \kappa(t)\delta_{f_m(t)}(B) + \nu_t(B), \tag{269}$$

composed of a Dirac measure $\delta_z(B) = \mathbb{1}_B(z), z \in \mathbb{R}$ *at* $f_m(t)$ *with weight*

$$\kappa(t) = e^{-\omega_1 ct} \cdot \left(1 - (\Delta\omega)^{-1}(m - \omega_1)(1 - e^{-\Delta\omega ct})\right)$$

152

and an absolutely continuous measure $\nu_t(\mathrm{d}z) = \pi(a)\mathrm{d}a$ supported on $(f_1(t), 1]$ with time-independent density

$$\pi(a) = \frac{cc_1}{c_2(c_1 + c_2)} \left(\frac{z - \omega_1}{1 - \omega_1}\right)^{\frac{3-(c_2+c_1+1)/\Delta\omega}{2}} \left(\frac{\omega_2 - 1}{\omega_2 - z}\right)^{\frac{3+(c_2+c_1+1)/\Delta\omega}{2}}, \tag{270}$$

where

Proof. We use Eq. (268) and compute

$$\mathbb{P}[\Theta(t) \in B] = \mathbb{P}[\Theta(t) \in B, t_1 > t] + \mathbb{P}[Z(t) \in B, t_1 \le t]$$
$$= \mathbb{P}[t_1 > t]\delta_{f_m(t)}(B) + \mathbb{P}[\Theta(t) \in B \cap (f_1(t), 1]]$$

having the form of Eq. (269). First,

$$\mathbb{P}[t_1 > t] = \exp\left(-c\int_0^t f_m(s)\,\mathrm{d}s\right) = \kappa(t).$$

Second, by the equivalence

$$\Theta(t) \in B \cap (f_1(t), 1] \Leftrightarrow t - \sup\{t_i \mid t_i \le t\} \in f_1^{-1}(B), t_1 \le t$$

the absolute continuity of the t_i implies that $\nu_t(B) = \mathbb{P}[\Theta(t) \in B \cap (f_1(t), 1]]$ is an absolute continuous measure, i.e., $\nu_t(\mathrm{d}a) = p(t, a)\mathrm{d}a$ with some density $p(t, a)$ supported on $(f_1(t), 1]$. Its solution is obtained by the method of characteristics [192], initiated at the boundary condition $p_1(t) := p(t, 1)$ and propagated through the rewritten linear PDE (104)

$$\frac{\partial}{\partial t}p(t, a) + A(a)\frac{\partial}{\partial a}p(t, a) = p(t, a)\left\{-ca - \frac{\mathrm{d}}{\mathrm{d}a}A(a)\right\}.$$

It remains to evaluate $p_1(t)$. We compute

$$\mathbb{E}[cZ(t)] = \mathbb{E}\left[\lim_{h\to 0} \frac{\mathbb{P}[Y(t+h) - Y(t) = 1 \mid Z_{[0,t]}]}{h}\right]$$
$$= \lim_{h\to 0} \frac{\mathbb{P}[Y(t+h) - Y(t) = 1]}{h}$$
$$= \lim_{h\to 0} \frac{\mathbb{P}[\Theta(t+h) \in (f_1(h), 1]]}{h}$$
$$= -f_1'(0)p(t, 1) = -A(1)p_1(t) = c_2 p_1(t).$$

The absolute continuity of ν_t was used in the fourth equality. Stationarity of Z implies that $p_1(t) \equiv cm/c_2$ independent of t. Plugging this in, $p(t, z) = \pi(z)\mathbb{1}_{(f_1(t),1]}(a)$ is obtained. $\square$

Accordingly, for the mutual information, we obtain the following analytic expression.

Theorem 9.2. *Let $z_1 = 0, z_2 = 1$, then*

$$\frac{1}{T}\mathbb{I}(Z_{[0,T]}, Y_{[0,T]}) = -\frac{c}{T}\int_0^T \phi(f_m(t))\kappa(t)\,\mathrm{d}t$$
$$- c\int_{\omega_1}^1 \phi(a)\pi(a)(1 - f_1^{-1}(a)/T)\,\mathrm{d}a.$$

Proof. Using theorems 9.1, the identity $\phi(cz) = c\phi(z) + z\phi(c)$, and Fubini's theorem, we evaluate Eq. (119). $\square$

Next, we state the asymptotic results.

Theorem 9.3. *The process $(\Pi_t(1))_{t \geq 0}$ converges in distribution, where the density of the limiting distribution is*

$$p_Z(z) = \frac{cc_1}{c_2(c_1 + c_2)} \left(\frac{z - \omega_1}{1 - \omega_1} \right)^{\frac{3 - (c_2 + c_1 + 1)/\Delta\omega}{2}} \left(\frac{\omega_2 - 1}{\omega_2 - z} \right)^{\frac{3 + (c_2 + c_1 + 1)/\Delta\omega}{2}}, \tag{271}$$

when we assume Z is distributed according to the limiting distribution. The limiting distribution of the backward recurrence time $(\tau(t))_{t \geq 0}$ is

$$p(\tau) = \frac{cc_1}{c_1 + c_2} \left(e^{-c\omega_1 \tau} \frac{\omega_2 - 1}{\Delta\omega} + e^{-c\omega_2 \tau} \frac{1 - \omega_1}{\Delta\omega} \right). \tag{272}$$

The MIR of the random telegraph model and its Poisson channel output equals

$$\bar{\mathbb{I}}(Z, Y) = -\frac{cc_1}{c_1 + c_2} \int_0^\infty \phi\left(\omega_2 - \frac{\Delta\omega}{1 + \frac{1 - \omega_1}{\omega_2 - 1} e^{-\Delta\omega\tau}} \right) \left[e^{-\omega_1 \tau} \frac{\omega_2 - 1}{\Delta\omega} + e^{-\omega_2 \tau} \frac{1 - \omega_1}{\Delta\omega} \right] \, \mathrm{d}\tau \tag{273}$$

$$= -c \int_{\omega_1}^1 \phi(z) p_Z(z) \, \mathrm{d}z. \tag{274}$$

Proof. According to theorem 4.32, $\Pi_t(1)$ converges in distribution to its unique invariant distribution. The distribution in Eq. (271) satisfies the stationarity condition obtained from (104) by equating the right side to zero. For Eq. (272) compute the solution $u(\tau)$ of

$$\dot{\Pi}_\tau(1) = c(c_1 - (c_1 + c_2 + 1)\Pi_\tau(1) + \Pi_\tau(1)^2), \quad \Pi_0(1) = 1 \tag{275}$$

that yields

$$u(\tau) = \omega_2 - \frac{\Delta\omega}{1 + \frac{1 - \omega_1}{\omega_2 - 1} e^{-c\Delta\omega\tau}}. \tag{276}$$

As in Eq. (85) we have $m(\tau) = cu(\tau)$. Then the unnormalized density $\rho(\tau)$ is obtained from solving Eq. (91)

$$\rho(\tau) = e^{-c\omega_1 \tau} \frac{\omega_2 - 1}{\Delta\omega} + e^{-c\omega_2 \tau} \frac{1 - \omega_1}{\Delta\omega}$$

and $p_0 = \mathbb{E}[\lambda_\infty] = \frac{cc_1}{c_1 + c_2}$. By $p = p_0 \cdot \rho$, the claimed Eq. (272) is proved. Using $\phi(cx) = c\phi(x) + x\phi(c)$ the MIR can be written as

$$\bar{\mathbb{I}}(Z, Y) = -cp_0 \int_0^\infty \phi(u(\tau)) \rho(\tau) \, \mathrm{d}\tau$$

$$= -\frac{cc_1}{c_1 + c_2} \int_0^\infty \phi\left(\omega_2 - \frac{\Delta\omega}{1 + \frac{1 - \omega_1}{\omega_2 - 1} e^{-\Delta\omega\tau}} \right) \left[e^{-\omega_1 \tau} \frac{\omega_2 - 1}{\Delta\omega} + e^{-\omega_2 \tau} \frac{1 - \omega_1}{\Delta\omega} \right] \, \mathrm{d}\tau.$$

The linear time scaling $\tau \mapsto c\tau$ was used in the second equality. Analogously, Eq. (274) is obtained from Eq. (271). We identify Eq. (273) as a reparametrization of Eq. (274) for $\omega = \omega_1$ and $\Pi_t(1) \xrightarrow{d} \Theta$

$$\bar{\mathbb{I}}(Z, Y) = -\mathbb{E}[c\phi(Z)] = -c \int_\omega^1 \phi(z) \pi_Z(z) \, \mathrm{d}z. \tag{277}$$

The link can be obtained via the transformation rule applied on the transformation $u \colon [0, \infty) \to (\omega, 1]$. Let T satisfy $\tau(t) \xrightarrow{d} T$, i.e., T follows the stationary distribution of $\tau(t)$. Then

$$\int_\omega^1 \phi(z) p_\Theta(z) \, \mathrm{d}z = \int_{u^{-1}(\omega)}^{u^{-1}(1)} \phi(u(t)) p_\Theta(u(t)) \cdot u'(t) \, \mathrm{d}t$$

$$= \int_0^\infty \phi(u(t)) \left[-p_\Theta(u(t)) \cdot u'(t) \right] \, \mathrm{d}t$$

$$= \int_0^\infty \phi(u(t)) p_T(t) \, \mathrm{d}t$$

$$= p_0 \int_0^\infty \phi(u(t)) \rho(t) \, \mathrm{d}t.$$

$\square$

Remark 9.4. The reparametrization Eq. (273) has the advantage over Eq. (274) that vertical asymptotes of the integrand at $z = \omega$ can be avoided. Furthermore, the integral bounds do not depend on the system parameters c_1, c_2, c, allowing more uniformly chosen integral bounds in the numerical approximation, compare remark 5.3.

9.2.4 Computation of partial derivatives

Getting rid of the dependence on the parameters in the integral bounds has another advantage. It enables the computation of partial derivatives. Let us pick up the tilde again. In order to answer questions of optimality of $\bar{\mathbb{I}}(Z, Y)$, the partial derivatives of $\bar{\mathbb{I}}(Z, Y) = \bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, c)$ are relevant. Note the relation $\partial_1 \bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, c) := \partial_{\tilde{c}_1} \bar{\mathbb{I}}(Z, Y) = c\partial_{c_1} \bar{\mathbb{I}}(Z, Y)$. In particular the nullclines $[\partial_1 \bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, c) = 0]$ and $[\partial_2 \bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, c) = 0]$ contain information about optimal points. By the implicit function theorem, the $\tilde{c}_1$-nullcline $\tilde{c}_1 \mapsto h(\tilde{c}_1)$ satisfies the ODE

$$h'_c(\tilde{c}_1) = -\frac{\partial_{11}\bar{\mathbb{I}}(\tilde{c}_1, h(\tilde{c}_1), c)}{\partial_{12}\bar{\mathbb{I}}(\tilde{c}_1, h(\tilde{c}_1), c)}. \tag{278}$$

Furthermore, to decide on the convexity of the c_1-nullcline, positivity of

$$h''_c(\tilde{c}_1) = \left[\frac{2\partial_{11}\bar{\mathbb{I}}\partial_{112}\bar{\mathbb{I}}}{\partial_{12}\bar{\mathbb{I}}^2} - \frac{\partial_{111}\bar{\mathbb{I}}}{\partial_{12}\bar{\mathbb{I}}} - \frac{\partial_{11}\bar{\mathbb{I}}^2\partial_{122}\bar{\mathbb{I}}}{\partial_{12}\bar{\mathbb{I}}^3}\right](\tilde{c}_1, h_c(\tilde{c}_1), c) \tag{279}$$

must be checked. This motivates to compute partial derivatives up to the third order. The numerical method is exemplified for $\partial_1 \bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, c)$. In order to appreciate the reparametrization, we observe that the Leibniz rule for differentiation of the parameter integral in Eq. (277) fails when there is an asymptote at ω_1. Then the lower boundary term evaluates to $-\infty$. Thus, let us exploit the reparametrization. Define $u_1(\tau) := \partial_{\tilde{c}_1} u(\tau)$ and $p_1(\tau) := \partial_{\tilde{c}_1} p(\tau)$ as well as

$$J_1(\tau) := \partial_{\tilde{c}_1} J(\tau) = \partial_{\tilde{c}_1} \int_0^\tau -c\phi(u(t))p(t)\ \mathrm{d}t = \partial_{\tilde{c}_1} \int_0^{c\tau} -\phi(u(t/c))p(t/c)\ \mathrm{d}t.$$

For the evolution of $u_1(\tau)$ one takes advantage of $\dot{u}_1(\tau) = \partial_1 \dot{u}(\tau)$. The joint evolution (for the time scaling $\tau \mapsto c\tau$, i.e., $\tilde{f}(\tau) := f(\tau/c)$ for $f = p, u, p_1, u_1, J_1$ and dropping the tilde again) is given by

$$\dot{p}(\tau) = -u(\tau)p(\tau) \tag{280}$$

$$\dot{u}(\tau) = \tilde{c}_1 - (\tilde{c}_2 + \tilde{c}_1 + 1)u(\tau) + u(\tau)^2 \tag{281}$$

$$\dot{p}_1(\tau) = \partial_1 \dot{p}(\tau) = -u_1(\tau)p(\tau) - u(\tau)p_1(\tau) \tag{282}$$

$$\dot{u}_1(\tau) = \partial_1 \dot{u}(\tau) = 1 - u(\tau)$$
$$\qquad - (\tilde{c}_2 + \tilde{c}_1 + 1)u_1(\tau) + 2u(\tau)u_1(\tau) \tag{283}$$

$$\dot{J}_1(\tau) = -\phi'(u(\tau))u_1(\tau)p(\tau) - \phi(u(\tau))p_1(\tau) \tag{284}$$

with initial conditions $(\frac{c\tilde{c}_1}{\tilde{c}_1+\tilde{c}_2}, 1, \frac{c\tilde{c}_2}{(\tilde{c}_1+\tilde{c}_2)^2}, 0, 0)$. The saturation value $\lim_{\tau\to\infty} J_1(\tau)$ is the partial derivative $\partial_1 \bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, c)$.

9.2.5 Mutual information in the phase plane

For the analysis we set $c = 1$, which is justified by the scaling behavior $\bar{\mathbb{I}}(c_1, c_2, c) = c\bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2, 1) =: c\bar{\mathbb{I}}(\tilde{c}_1, \tilde{c}_2)$. For convenience, we drop the tilde. Along the line $(c_1, (e-1)c_1)$, both partial derivatives are numerically found not to switch sign. From this, we conclude that the nullclines do not intersect. This excludes local optima of $\bar{\mathbb{I}}(c_1, c_2)$. For a rectangular constraint $0 < c_1 \le r_1, 0 < c_2 \le r_2$, the maximizing pair (c_1^*, c_2^*) is consequently always located on one boundary $c_1 = r_1$ or $c_2 = r_2$. The (c_1, c_2)-plane is split into regions A, B and C by the nullclines. The pair $[\text{sgn}(\partial_1\mathbb{I}), \text{sgn}(\partial_2\mathbb{I})]$ characterizes the regions: $[1, -1], [1, 1], [-1, 1]$ on A,B,C. Depending on the location of (r_1, r_2), the maximum (c_1^*, c_2^*) shows different behavior. If located in A, it holds $c_1^* = r_1, c_2^* < r_2$, while a location in B enforces $c_1^* = r_1, c_2^* = r_2$, and finally, $c_1^* < r_1, c_2^* = r_2$ for (r_1, r_2) in region C. As a summary,
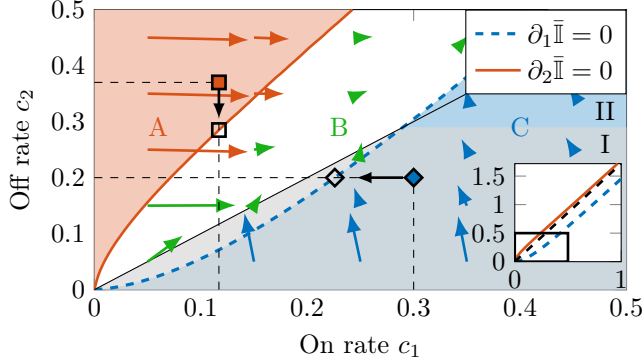
Figure 34: **Phase plane analysis of the MIR** for the Poisson channel with random telegraph input. Nullclines $\partial_1 \bar{\mathbb{I}} = 0$ and $\partial_2 \bar{\mathbb{I}} = 0$ were evaluated, using $T = 1000$. Colored arrows indicate the gradient of the MIR, calculated alike. Optimization domains are rectangular. Depending on the location of the domain's upper right corner (r_1, r_2), the optimum is assumed on the nullclines $\partial_1 \bar{\mathbb{I}} = 0$ [(r_1, r_2) in C] and $\partial_2 \bar{\mathbb{I}} = 0$ [(r_1, r_2) in A], respectively, or in the interior of region B [(r_1, r_2) in B]. Regions I and II contain all (r_1, r_2), whose optima favor the On and Off state, respectively. In the inset the dashed line $c_2 = (e - 1)c_1$ separates the nullclines. The gain was fixed at $c = 1$.

the maximum is always located in region B or its boundary. Numerically, we find that the c_1-nullcline crosses the bisection line (c_1, c_1) at 0.29. An evaluation of $\partial_1 \bar{\mathbb{I}}(c_1, c_2)$ for $c_1 \in (0, 0.29)$ indicates no further intersections. The derivative (278) at $c_1 = 0.29$ is found to be larger than one. Furthermore, Eq. (279) evaluates to positive values in the region $\{(c_1, c_2) \in [0, 0.3]^2 : c_1 - 0.07 \leq c_2 \leq c_1 + 0.01\}$. Hence, the isoclines that transit the region are convex, in particular the c_1-nullcline. The constraint $0 < c_2 \leq r_2$ for $r_2 \in (0, 0.29)$ then returns a maximum (c_1^*, c_2^*) with $c_1^* > c_2^*$. This shows that a bandwidth-like constraint can impose an On-favoring maximum complementing the classical result by Kabanov. The sequence of input processes for the Poisson channel that exhaust its capacity is a random telegraph process with $c_1, c_2 \to \infty$ and $c_2/c_1 \to e - 1$. The stationary input distribution favors the Off state, occupying it $1 - 1/e$ of the time.

Optimizing the allocation of the On and the Off state under constraint (B1) can be intuitively explained as an interplay between different forces that maximize the efficiency and precision of signal transmission. On the one hand, a force, reducing average sojourn times is predominant in region B of Fig. 34. This force aims at increasing the amount of signals transmitted. On the other hand, forces that increase the sojourn time in the On and Off states are predominant in regions A and C, respectively. A larger sojourn time in the On state increases the likelihood of observing the On state at the channel output. A larger sojourn time in the Off state decreases the likelihood of misinterpreting the period between consecutive channel output pulses as an input Off phase. The phase diagram in Fig. 34 explicitly quantifies how the ensemble of forces is balanced.

Given the analyzed model's and the constraints' rather minimal nature it might not account for biophysical reality. Yet assuming that evolutionary strategies aim at achieving capacity [193] the model supplies the hypothesis that system parameters $\tilde{c}_1, \tilde{c}_2$ are located in or near the optimal region B [194]. The prediction made by this interpretation is yet to be verified by experimentalists.

9.3 Optimal promoter architecture

The goal of this section is to provide evidence that the circular motif is information-theoretically optimal for the 3- and 4- state promoter with one active and otherwise inactive states. We also provide evidence that there is an information-theoretically optimal number of steps for the circle.

We considered the Poisson channel

$$R : \emptyset \xrightarrow{\lambda(Z_t)} Y$$

with Markov input $Z_t \in \mathcal{Z}$. The input is discrete and $\lambda(Z_t) \in \{0, 1\}$. For this reason, we call the input *binary*. Moreover, we assume that there is a unique $z \in \mathcal{Z}$, such that $\lambda(z) = 1$. For this reason, we

say that the input has exactly one On state, while the remaining states are Off states. Without loss of generality, $\mathcal{Z} = \{1, \ldots, k\}$ and $\lambda(1) = 1$. Denote by α_{ij} the transition rate from i to j, where $i, j \in \mathcal{Z}$. Then, according to remark 4.18, the CI is of dimension 0 with the backwards recurrence time τ being sufficient without additional state variables. In other words, the process Y_t is a renewal process. The mutual information $\bar{\mathbb{I}}(Z, Y)$ is computed by method 5.2.2, i.e., Eqs. (128)- (130) with auxiliary states $\Theta_\tau = (\Pi_\tau(1), \ldots, \Pi_\tau(k-1))$ and the evolution equation F for $i < k$

$$\dot{\Pi}_\tau(i) = -(\sum_{j, i \neq j} \alpha_{ij})\Pi_\tau(i) + \sum_{i, i \neq j} \alpha_{ji}\Pi_\tau(j) + c\Pi_\tau(1)\Pi_\tau(i) - c\mathbb{1}(i = 1)\Pi_\tau(i),$$

where $\Pi_\tau(k) = 1 - \sum_{j<k} \Pi_\tau(j)$ and $c = 1$. The equation is initialized at $(\Pi_0(1), \ldots, \Pi_0(k-1)) = (1, 0, \ldots, 0)$. In Eq. (128) and (130), the function $l(\Pi) = \Pi(1)$ is used, accordingly.

We considered Markov transition architectures $A \in \mathcal{A}(k)$ for $|\mathcal{Z}| = 3$ and $|\mathcal{Z}| = 4$, Formally we define the set of architectures for $k = |\mathcal{Z}|$ as

$$\mathcal{A}(k) = \{A \in \{0, 1\}^{k \times k}; \forall i, j \in \{1, \ldots, k\} : a_{ii} = 0, (A^k)_{ij} > 0\}/\sim,$$

where we divide by the equivalence relation

$$A \sim B \Leftrightarrow \exists \Phi \in \mathcal{S}(\{1, \ldots, k\}), \Phi(1) = 1 : \forall i, j \in \{1, \ldots, k\} : a_{\Phi(i)\Phi(j)} = b_{ij}.$$

with $\mathcal{S}$ being the symmetric group. More informally, Φ corresponds to a relabeling of the states that keep the active state 1 fixed. The entry $a_{ij} = 1$ means that there is a transition from j to i, whereas $a_{ij} = 0$ means there is no such transition. The condition $(A^k)_{ij} > 0$ guarantees that Z_t is irreducible. For each Markov transition architectures $A \in \mathcal{A}(k)$ we optimized the function $\bar{\mathbb{I}}(Z, Y) = \bar{\mathbb{I}}(\alpha_{ij})$ in the parameter space of positive matrix entries

$$\{(\alpha_{ij}); i \neq j, a_{ij} = 1\} = \mathbb{R}_{>0}^{\chi(A)}, \ \chi(A) = \sum_{i,j} a_{ij}.$$

To capture the constraints on the mean On and Off time, define the set $\mathcal{Q}$ of Q matrices that are transition matrices of irreducible CTMCs, i.e., off diagonal entries are non-negative, row-sums are zero. Call 1 the active state in $\mathcal{Z}$. For Z_t with transition matrix Q, and jump times $\tau_1, \tau_2, \ldots$, the mean Off time is

$$m(Q) = \mathbb{E}[\min\{\tau_k - \tau_{n+1}\}|k > n+1, Z_{\tau_n} = 1 = Z_{\tau_k}].$$

For given $A \in \mathcal{A}(|\mathcal{Z}|)$, given mean Off time μ and mean On time M, we optimized over the set

$$\{Q \in \mathcal{Q}; \max(\mathrm{sgn}(Q), 0) = A, m(Q) = \mu, Q_{11} = -M^{-1}\} \subseteq \mathbb{R}_{>0}^{\chi}(A)$$

where the sgn and max functions are applied entry-wise to map the Q into $\mathcal{A}(k)$.

The results for the three-state and four-state inputs are shown in figures 35, 36 and in the appendix C. Within the constraints of the architecture, a one-directional circling through all states was preferred over branching or smaller circles. Furthermore, Off states were equally allocated. The experimentally obtained optimal input thus corresponds to an Erlang-distributed sojourn time.

Next, after we identified the circular motif as maximizing the mutual information rate, we were interested, whether a higher number of steps in the circle always increases the mutual information. To this end, we fixed the mean On and Off time. Interestingly, we found an information-maximizing number of steps in the circle, see figure 37. This can be explained, when we notice that with an increasing number of steps the Off time distribution becomes less informative. Its variance decreases to 0, the distribution converges to a delta distribution with peak at the fixed mean. As a consequence, information is only transmitted by the On time, while the Off time takes up part of the trajectory without being informative. Since in the mutual information rate we divide by the time, this nearly deterministic Off time causes loss of information transmission efficiency.

9.4 Leakage and non-exponential On time distributions

In the previous sections we restricted our study of the Poisson channel to binary inputs without leakage and with exponentially distributed On sojourn time. Directions that could be taken for leakage and non-exponential On time distributions are considered in the following two examples.

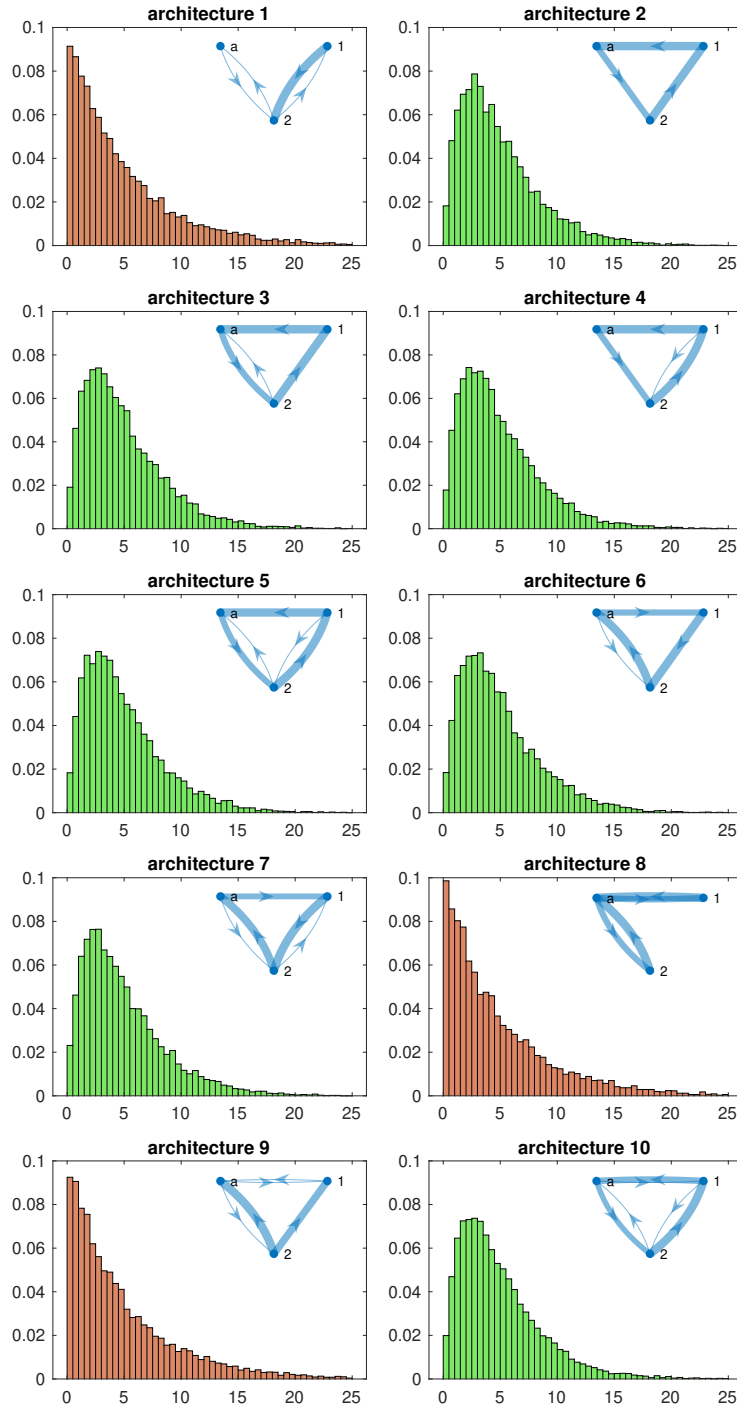OFF-time distribution for 3 state architectures, OFF-mean = 5, ON-mean = 3.3

Figure 35: Results of our experiment of the Poisson channel with three-state input, of which one was active and two were inactive. The 10 different architectures were investigated. The active state is labeled with a. The thickness of the arrow shows the transition rates of the optimal input under the fixed average sojourn time. The coloring from red (low) to green (high) encodes the value of the mutual information rate. The histogram shows the corresponding Off time distribution.
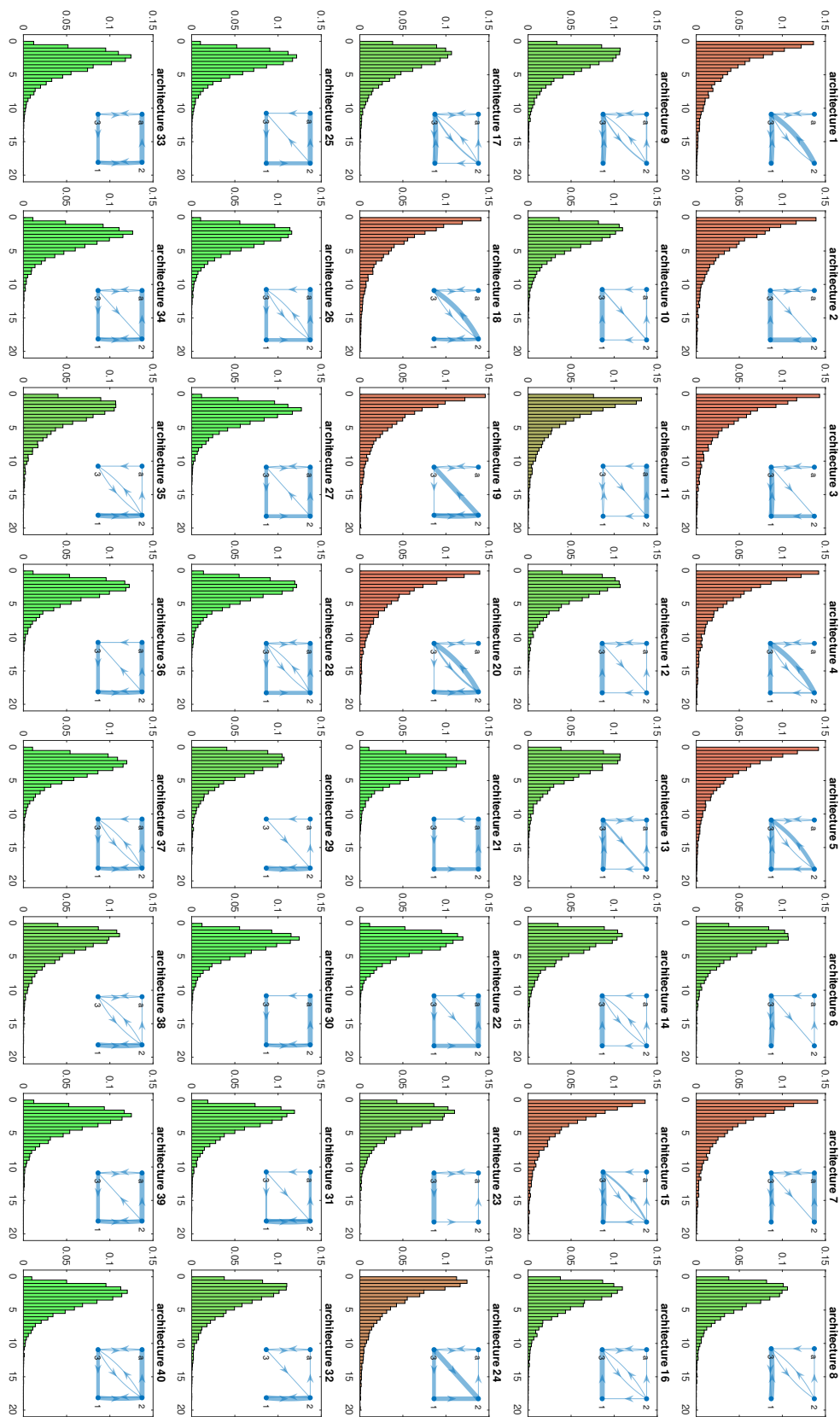
Figure 36: Results of our experiment of the Poisson channel with four-state input, of which one was active and three were inactive. The 287 different architectures were investigated, of which 40 are shown. For the remaining 247, see the appendix C. Labels, colors and arrows as in figure 35.
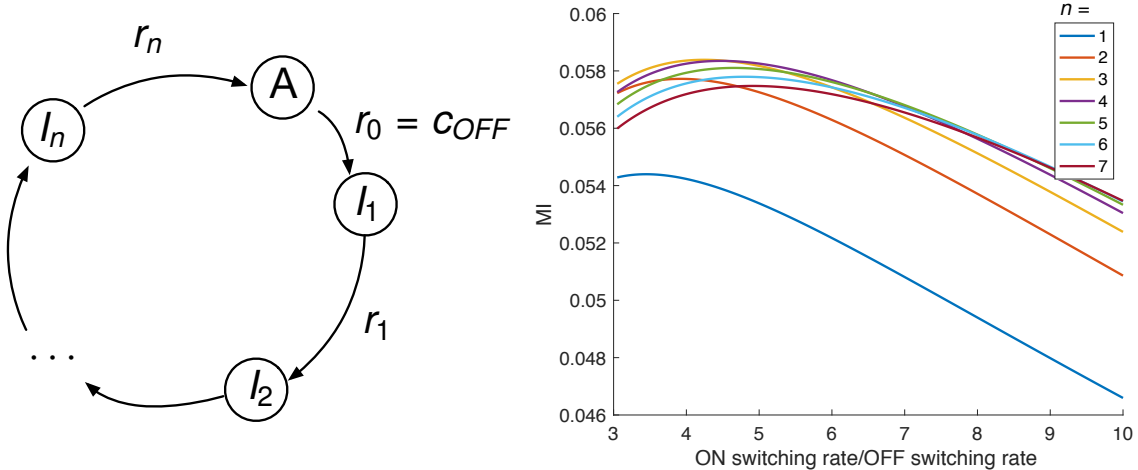
Figure 37: Circular promoter model with n inactive and one active state (left) and the mutual information rate between promoter input and Poisson channel output (right). The maximal mutual information rate is non-monotone in n. We used $r_1 = \cdots = r_n$ and On switching rate r_1/n, the inverse of the mean Off sojourn time n/r_1.

9.4.1 Random telegraph with dark current

We now consider the MMPP $(Y_t)_{t\geq 0}$ with a random telegraph modulator and $\lambda_0 > 0$. As seen in the proposition 6.2, $\dim(Y_t)_{t\geq 0} = 1$ and the integral boundary condition, theorem 4.24, will be in place in contrast to the previous example with $\dim(Y_t)_{t\geq 0} = 0$. We numerically investigate how a non-zero dark current alters $\bar{\bar{\mathbb{I}}}(Z, Y)$. For this purpose, we apply sections 5.2.1 and 5.2.2.

Continuing the example 9.2 with $f, c_1, c_2, \omega_1, \omega_2, \lambda_0, \lambda_1, m$ as defined in there, we first make the following observation to choose the discretization domain.

Remark 9.5. For every θ the trajectory $m(\cdot, \theta)$ is decreasing and $\lim_{\tau\to\infty} m(\tau, \theta) = \omega_1$. Consequently, for any τ, θ

$$g(\tau, \theta) > f(\omega_1) = \frac{(\omega_1 - \lambda_0)(\lambda_1 - \omega_1)}{\omega_1} =: f_\infty.$$

Proposition 9.6. *Suppose that we partition the $p_0(\theta)$-support $\Omega = (f_\infty, \lambda_1]$ into equidistant intervals $(b_{i-1}, b_i]$ with $b_i = f_\infty + i \cdot \frac{1 - f_\infty}{N}, i = 1, \ldots, N$ and choose representatives $\theta_i = \frac{b_i + b_{i-1}}{2}$. Then the matrix entries in Eq. (125) are given by*

$$A_{i,j} = R(b_i, \theta_j) \wedge 1 - R(b_{i-1}, \theta_j) \wedge 1$$

with

$$R(\theta, \theta') = \left(\frac{f^{-1}(\theta) - \omega_1}{\theta' - \omega_1} \right)^{\frac{\omega_1}{\Delta\omega}} \left(\frac{\omega_2 - \theta'}{\omega_2 - f^{-1}(\theta)} \right)^{\frac{\omega_2}{\Delta\omega}},$$

where $f, \omega_1, \omega_2, \Delta\omega$ are as in example 9.2.

Proof. Continuing the example 9.2 and solving Eq. (91), we obtain

$$\rho(\tau, \theta) = e^{-\omega_1 \tau} \frac{\omega_2 - \theta}{\Delta\omega} + e^{-\omega_2 \tau} \frac{\theta - \omega_1}{\Delta\omega}.$$

Define the function

$$T(\theta, \theta') := \Delta\omega^{-1} \left\{ \ln\left(\frac{\omega_2 - \theta}{\theta - \omega_1} \right) + \ln\left(\frac{\theta' - \omega_1}{\omega_2 - \theta'} \right) \right\}. \tag{285}$$

The number of solutions of $g(\tau, \theta') = \theta$ is $N(\theta, \theta') \in \{0, 1\}$ with $\tau(\theta, \theta') = T(f^{-1}(\theta), \theta')$. Then the matrix entries in Eq. (125) are given by

$$A_{i,j} = \rho(\tau(b_i, \theta_j) \vee 0, \theta_j) - \rho(\tau(b_{i-1}, \theta_j) \vee 0, \theta_j)$$
$$= R(b_i, \theta_j) \wedge 1 - R(b_{i-1}, \theta_j) \wedge 1.$$
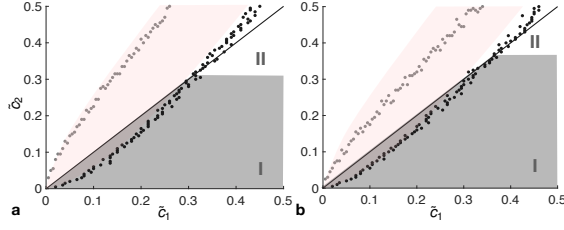
160

Figure 38: Phase planes with leakage. **a** $x_1 = 0.01, x_2 = 1$ **b** $x_1 = 0.1, x_2 = 1$. Black and gray marks indicate the nullclines $\partial_1 \bar{\bar{\mathbb{I}}} = 0$ and $\partial_2 \bar{\bar{\mathbb{I}}} = 0$, respectively, and were obtained from Monte Carlo simulations with sample size $2 \cdot 10^6$. The red shaded area indicates region B of the case $x_1 = 0$.
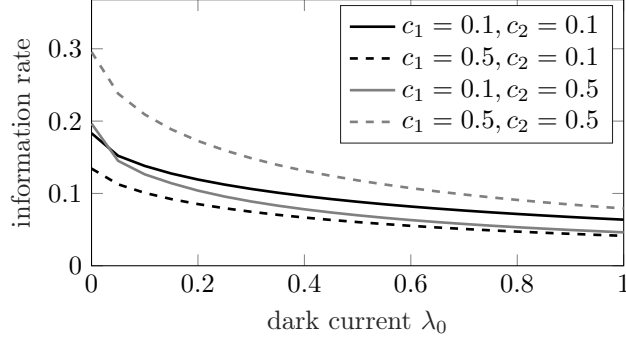

Figure 39: **Random telegraph input with dark current λ_0 and amplitude $\Delta\lambda = 1$.** The MIR $\bar{\bar{\mathbb{I}}}(c_1, c_2)$ is plotted for different values of λ_0, c_1, c_2. The graphs of $c_1 = 0.1, c_2 = 0.1$ and $c_1 = 0.1, c_2 = 0.5$ intersect, showing that dark current can alter the monotonicity properties of $\bar{\bar{\mathbb{I}}}(c_1, c_2)$ in the (c_1, c_2)-plane.

$\square$

The initial condition $p_0(\theta)$ was found by fixed-point iteration A^{2^L} of Eq. (125) with L iterations. The MIR $\bar{\bar{\mathbb{I}}}(Z, Y) = \bar{\bar{\mathbb{I}}}(c_1, c_2)$, Eq. (121), was computed with ODE system (128) - (130). Results for different c_1, c_2 and increasing dark current are shown in figure 39. For any examined pair (c_1, c_2), the MIR decreases with dark current as expected. The figure reveals a notable property. For fixed $c_1 = 0.1$, the plots for $c_2 = 0.1$ and $c_2 = 0.5$ intersect. This means that increasing the dark current increases the MIR $\bar{\bar{\mathbb{I}}}(0.1, 0.1)$ relative to $\bar{\bar{\mathbb{I}}}(0.1, 0.5)$. Consequently, an increased dark current can qualitatively alter the monotonicity and optimality properties in the (c_1, c_2)-phase plane. For example, the On favoring region increases with a dark current, see figure 38.

For the histogram approximation of p_λ, as described in section 5.2.4, we compute $\tau^{(m)}(\theta, \theta') = T(\theta, \theta')$, using Eq. (285) and explicitly obtain in (132)

$$\int_{\tau^{(m)}(m_i, \theta')}^{\tau^{(m)}(m_{i-1}, \theta')} \rho(\tau, \theta') \, \mathrm{d}\tau = \frac{(\omega_2 - \theta')^{\frac{\omega_2}{\Delta\omega}}}{\Delta\omega(\theta' - \omega_1)^{\frac{\omega_1}{\Delta\omega}}} \left[\frac{(m - \omega_1)^{\frac{\omega_1}{\Delta\omega}}}{\omega_1(\omega_2 - m)^{\frac{\omega_1}{\Delta\omega}}} + \frac{(m - \omega_1)^{\frac{\omega_2}{\Delta\omega}}}{\omega_2(\omega_2 - m)^{\frac{\omega_2}{\Delta\omega}}} \right]_{m=m_{i-1}}^{m=m_i}.$$

Fig. 40 shows the agreement of the computation with and without the Monte-Carlo sampling. The ACID $p_\lambda(m)$ has an asymptote at ω_1 and is not differentiable at $f_\infty = f(\omega_1)$.

The linear fixed point equation (134) can be used to return a numerical approximation of $p_\lambda(m)$ directly. However, it can have a singularity at the equilibrium θ characterized by $A(\theta) = 0$. And the ad-hoc discretization into equidistant intervals is not guaranteed to be a stochastic matrix as in Eq. (125).

The equation (134), which is the analogue to theorem 4.25, yields

$$K(m, m') = \frac{m'(m - \omega_1)^{\frac{\omega_1}{\Delta\omega} - 1}(\omega_2 - f_-(m'))^{\frac{\omega_2}{\Delta\omega}}}{(f_-(m') - \omega_1)^{\frac{\omega_1}{\Delta\omega}}(\omega_2 - m)^{\frac{\omega_2}{\Delta\omega} + 1}}. \tag{286}$$
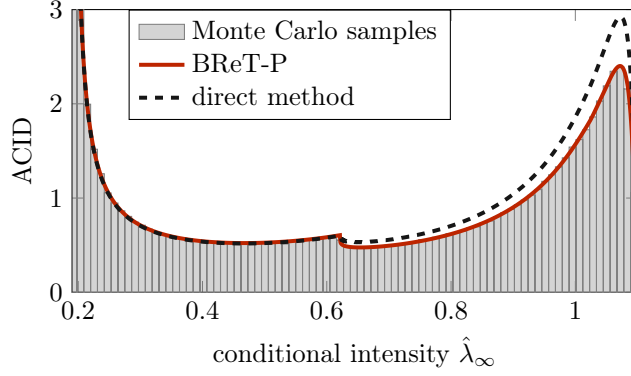
161

Figure 40: **Random telegraph input with dark current.** ACIDs obtained from Monte Carlo samples and simulation-free computation via BReT-P agree. The direct method, computed by fixed point iteration of Eq. (286), shows an inaccuracy. The density has a vertical asymptote at ω_1 and is non-smooth at $f_\infty = f(\omega_1)$. Parameters were $c_1 = c_2 = 0.1, \lambda_0 = 0.1, \Delta\lambda = 1$. Sample size was 10^6 for the Monte Carlo simulation. Grid point number was 2000 for the p_0-support $(f_\infty, \lambda_1]$ and number of fixed point iterations was $L = 12$.
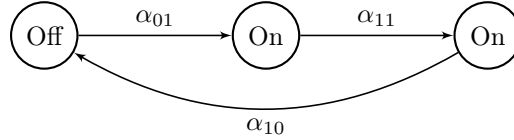
Figure 41: **State diagram for the Double On Single Off (DOnSOff) model.** The three-state model is Markovian. A refractory second active state realizes a non-exponential sojourn time in On. For $\alpha_{11} = \alpha_{10}$ the sojourn time is an Erlang distribution. The Off sojourn time remains exponential.

in the Eq. (133). In fig. 40 the accuracy for the considered random telegraph with dark current suffers, possibly caused by the singularity at the equilibrium ω_1. This is in contrast with fig. 13 that showed good accuracy for the direct method.

9.4.2 Non-Markov On time distribution

In section 9.3 we considered promoter models which are binary and Markovian with precisely one unique On state. Consequently, the On time was exponentially distributed. Now, we relax this condition. Let $(Z_t)_{t\geq 0}$ be an ergodic CTMC with n states, let $\lambda : \mathcal{Z} \to \{0, c\}$ be binary with two active, i.e., non-zero states z_1, z_2 and $(Y_t)_{t\geq 0}$ the corresponding MMPP. We refer to the model as double On (DOn). Then introduce the conditional probability of being in an active state

$$A_t = \Pi_t(z_1) + \Pi_t(z_2)$$

and $U_t := \Pi_t(z_1)/A_t$ the contribution of z_1 to this conditional probability. Then

$$\Theta_t := (U_t, A_t, \Pi_t(z_3), \ldots, \Pi_t(z_{|\mathcal{Z}|-1}))$$

are sufficient variables of joint Markovian progression. The reset value at jumps is $(U_{t-}, 1, 0, \ldots, 0)$. Since all but the first component are set to a constant at jumps, we may choose the sufficient statistic $\theta(t) = U_{t-\tau(t)}$, hence $\dim(Y_t)_{t\geq 0} = 1$, compare remark 4.16. The progression of Θ_t can be found by the chain rule, proposition 4.8, from Eq. (8). We only elaborate on it for the circular CTMC with double On state and single Off state (DOnSOff), depicted in figure 41. Let $\alpha_{01}, \alpha_{11}, \alpha_{10}$ be the transition rates of going from inactive to first active, first to second active and second active back to inactive, then the generator is of the form

$$\mathcal{A} = \begin{pmatrix} -\alpha_{11} & 0 & \alpha_{01} \\ \alpha_{11} & -\alpha_{10} & 0 \\ 0 & \alpha_{10} & -\alpha_{01} \end{pmatrix}.$$
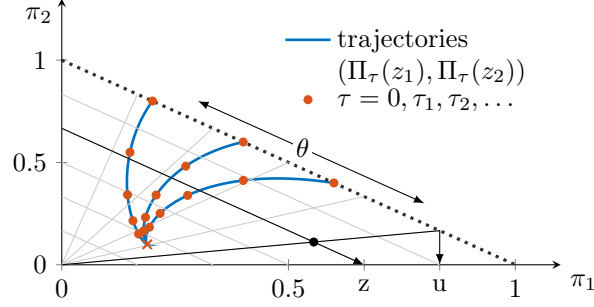
Figure 42: **The Double On Single Off input.** Trajectories $(\Pi_\tau(z_1), \Pi_\tau(z_2))$ evolve in the (π_1, π_2)-plane with $\pi_1 + \pi_2 \leq 1$. The gray grid indicates the change of coordinates $(u, a) = (\pi_1/(\pi_1 + \pi_2), \pi_1 + \pi_2) \in [0, 1]^2$. The initial value is $[u_0, a_0] = [\theta, 1]$. Dots indicate the temporal evolution, i.e. trajectory values at equally spaced time points. The equilibrium is marked by the cross. Parameter values were $\alpha_{01} = 0.4, \alpha_{11} = \alpha_{10} = 0.8, c = 1, \theta = 0.2, 0.4, 0.6$

The evolution of the conditional probabilities $\Pi_t(z_1)$ and $\Pi_t(z_2)$ is, according to the Snyder filter in proposition 2.10,

$$\dot{\Pi}_t(z_1) = \alpha_{01}(1 - A_t) - \alpha_{11}\Pi_t(z_1) - c(1 - A_t)\Pi_t(z_1)$$
$$\dot{\Pi}_t(z_2) = \alpha_{11}\Pi_t(z_1) - \alpha_{10}\Pi_t(z_2) - c(1 - A_t)\Pi_t(z_2).$$

Via the function $\varphi(\pi_1, \pi_2) = (\pi_1/(\pi_1 + \pi_2), \pi_1 + \pi_2) = (u, a)$ and the Ito rule, proposition 4.8, this transforms to, omitting t,

$$\dot{U} = -\alpha_{11}U + \alpha_{01}\frac{(1 - A)(1 - U)}{A} + \alpha_{10}U(1 - U) \tag{287}$$
$$\dot{A} = -c(1 - A)A + \alpha_{01}(1 - A) - \alpha_{10}(1 - U)A. \tag{288}$$

The update function according to Eq. (7) is $f(u, a) = (u, 1)$, i.e., constant in the second component. Figure 42 visualizes how trajectories for different initial values $[U(\tau), A(\tau)]_{\tau=0} = [\theta, 1]$ evolve. The coordinate system shows the plane (π_1, π_2), while the grid indicates the transformed radial-like coordinates (u, a). The DOnSOff model is equivalent to a binary semi-Markov process with exponential sojourn time in the Off state, while the sojourn time in the On state is the convolution of two exponential distributions. It serves as an example of a non-Markovian binary input to the Poisson channel whose $\bar{\bar{\mathbb{I}}}(Z, Y)$ is compared to the Markov case in the following paragraph.

9.4.3 Double On single Off

Binary Markov input processes exhaust the capacity of the Poisson channel when their switching rates tend to infinity. The defining property for optimality in the limit is only the proportion On/Off. With the autocorrelation time going to 0 for an exhausting sequence of binary semi-Markov processes also, the Markov property might as well be relaxed. Consider the following capacity problem: We restrict the input process class to binary semi-Markov processes and impose a lower bound on the average sojourn times in the On and the Off state. Is the Markov case with its exponential sojourn times the capacity-achieving input? Here, we consider the DOnSOff model, which is a semi-Markov processes with exponential sojourn time in the Off and Erlang sojourn time in the On state. For the numerical evaluation, following sections 5.2.1 and 5.2.2, we discretize $[0, 1] \ni \theta$. The state variables $(u(\tau, \theta), a(\tau, \theta))$ evolve according to (287) - (288) with initial conditions $[u(0, \theta), a(0, \theta)] = [\theta, 1]$ and $g(\tau, \theta) = u(\tau, \theta)$, compare Eq. (85). The times $\tau_k(\theta'_j)$ in (125), that satisfy

$$g(\tau, \theta'_j) = b_i, \tag{289}$$

were found by evolving the ODE system (91), (287), (288) and checking for the event (289). The matrix entries (125) were evaluated and $p_0(\theta)$ was found by fixed-point iteration with 2^L iterations. The rates
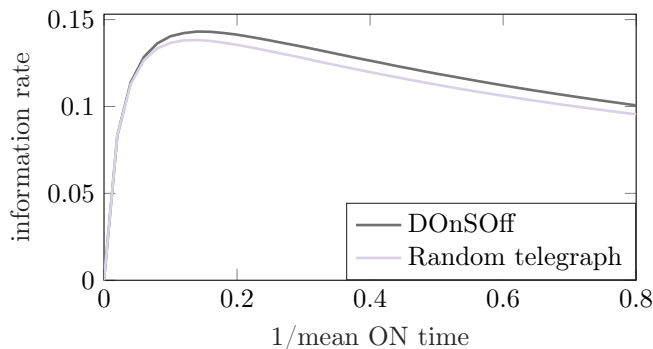
Figure 43: **The Double On Single Off model** increases the MIR compared to the random telegraph model. Parameters were $\alpha_{01} = c_1 = 0.04$ and $\alpha_{11} = \alpha_{10} = 2 \cdot c_2$. This choice guaranteed a model match in terms of mean On and Off times.

$\alpha_{01} = 0.04, \alpha_{11} = \alpha_{10} = 1.6$ exhibit a MIR of $\bar{\bar{\mathbb{I}}}(Z, Y) = 0.101$ compared to $\bar{\bar{\mathbb{I}}}(Z, Y) = 0.096$ for the Markov case. The rates were tuned such that the average sojourn times are the same. This example shows that the Markov input does not generally solve the capacity problem with average sojourn time constraint. It remains an open research question which On and Off sojourn time distributions are capacity-achieving.

10 Summary, Discussion and Outlook

For the analysis of stochastic CRNs in a random environment, we contributed at different levels of the stochastic description. At the process level, we established a link between CRNs in a linear random environment with Hawkes processes. Furthermore, we introduced the class of Dirac-PDMPs and the BReT-P, a standard form of marginal process equations that make use of auxiliary state variables. From these, by proceeding to the probability evolution layer, we derived generalized master equations and their asymptotic analysis. Those enabled the simulation-free estimation of information measures. We presented the ACID, a characterization tool for counting processes. At the level of moments, with our new method ESME, we computed the stationary means of linear CRNs in a Markov environment. For the seemingly simple example of the birth-death process with environmentally modulated decay rate, our case study 8.3 revealed a complex effect of the environment, even on the mean. We could explain the asymptotic cases of a slow and a fast subnetwork. Between the two asymptotic regimes, however, we observed a complex transition phase. At the levels of moments, probabilities and the stochastic process, we compared the Snyder filter and approximate filters by using, respectively, asymptotic Fano factors, the RER and anecdotal case studies.

We would like to point out misconceptions in the marginal process framework as introduced by Zechner [3] and Bronstein [78]. Firstly, we found evidence that the interpretation of the variance decomposition as effective and suppressed noise does not hold. Secondly, the Gamma filter is not exact for the CIR-modulated reactions. Thirdly, the CI was previously seen as an optimal estimator of the environment based on the subnetwork observations. Here, we advocate the perspective of filtration-dependent intensities. With this change of perspective, the CI can emancipate as a state variable that, on an augmented state space, is Markov. Furthermore, the filtration provides another tuning knob that can be flexibly modified when developing model approximations.

10.1 Marginal simulation

We presented a novel approximate marginal simulation scheme for CRNs in random environments based on the Hawkes approximation. The method builds on optimal linear state estimation for doubly stochastic Poisson processes. The simulation scheme studied here provides a convenient way to account for variance in the subsystem caused by the environment, even when only limited information about the environment is available, i.e., the mean, covariance and auto/cross-covariance decay of the environment.

This matches well with the situation of the experimentalist who has limited information about the environment, but would like to include that information in the model.

It would be of interest to investigate whether the linearizing approach with a Gaussian environment that has the appropriate first- and second-order characteristics, generates trajectories equivalent to the Hawkes model. For the linear noise approximated environment, one can argue that the intensity can become negative. In contrast, when considering independent environment components, our approach guarantees positivity by the one-dimensional Hawkes model. For a correlated multi-component environment that modulates a single reaction, we derived a sufficient criterion for positivity. However, in general, we could not guarantee that the filter remains positive. Yet, we anticipate that the possibility of negative rates is less restrictive than for the linear noise approximation, because we use marginal rates. The latter ones have a smaller variance, and thus explore a smaller range of intensity values, making negative intensities less likely.

Independent of the use of approximate filters, we introduced Sinzger's min-thin algorithm, a new simulation technique for time-dependent piecewise deterministic rates. It combines the strengths of the Anderson MNRM and Lewis' thinning method.

Using the established link with the Hawkes process, future research can transfer more results from this well-studied self-exciting counting process to CRNs in a random environment. The Hawkes process can be simulated using an immigration-birth-scheme based on branching processes [195]. We anticipate that marginal simulation with approximate filters can be implemented in analogous schemes.

10.1.1 Evaluation of approximate filters

Although several approximate filter schemes have been proposed in the last decade, there has been limited progress in systematic studies of their structural properties and estimation of their accuracy. For the optimal linear filter, we provided structural results in terms of the second-order moment agreement for zeroth-order modulation. It is to be expected that only statements about the first- and second-order moments are possible, since the approximation ignores other features of the environment.

We presented different similarity measures for counting processes. On the one hand, the RER is an established measure for the comparison of two counting processes based on their path measure. We provided a simulation-free computational method for it. On the other hand, a comparison based on the ACIDs is less established. The ACID may offer an accessible lower-dimensional statistic for comparison that still exceeds the first- and second-order analysis. We exemplified this for the comparison of an exact and an approximate filter or of two approximate filters. The ACID was successful at differentiating between them. While the relative entropy can detect whether path measures are equal, the ACID, in contrast, does not have this ability. By our current state of knowledge, it does not allow a conclusion about the distance of the path measures in case two ACIDs are close.

Our characterization theorem 4.5 sheds light on the moment closure $\mathbb{E}\left[\frac{V(t)^2}{\hat{\lambda}(t)}\right] = \mathbb{E}[V(t)]^2/\mathbb{E}[\hat{\lambda}(t)]$, which was misconceived before. In [3] this moment relation and its supposed equivalence with $S_3 = 2V/\hat{\lambda}$ was erroneously used to derive the Gamma filter from the CIR-modulated reaction channel. This would have justified the Gamma filter as being exact for the CIR-modulated counting model. However, we showed that the moment closure instead imposes the optimal linear filter.

For mRNA structure switching and its effect on translation in gene regulatory networks, we tested several approximate filters. As a limitation of the method we saw the following effect. In cases with a non-negligible deviation from the quasi-steady state, i.e., when the effect of the environment is needed, the approximated models is also more prone to inaccuracies. The non-negligible cases were those when the structure considerably contributed to the Fano factor. In contrast, the quasi-steady state assumption was clearly valid for a fast switching timescale compared to the translation timescale, i.e., a mixing property in time. And also, when enough conditions for a mixing behavior in the states were combined: (i) more than two structures, (ii) $N \geq 20$, and (iii) low fold change in the translation rate among structures. The results raise the question, in parameter regime is at once (i) biologically relevant, (ii) non-negligibly contributing to the Fano factor (iii) well approximated by the reduced models and (iv) less complex than the non-reduced model. The studies conducted here provide evidence that this regime might be small, which questions the capability of the approximate filter approaches.

Namely, we observed complementary coverage in performance by the Hawkes and the multinomial (binomial) model. The performance was measured in terms of a lower deviation of the KL divergence

rate. While the Hawkes model performed better when a mixing behavior in the states is given, i.e., when Snyder's filter allocates the intermediate values, the multinomial performed better when due to lacking mixing properties, Snyder's filter allocates more frequently the upper and lower values of the achievable regime. This is the case in particular with low N. In the case study of a minimal gene regulatory network, the toggle switch, we investigated how the symmetric stable switch is destabilized with heterogeneous translation rates due to conformational switching. The binomial and the Hawkes model captured qualitatively the increasing switching frequency for slower conformational transitions and increased fold change. However, in line with the conclusions from the accuracy assessment, we saw complementary failure modes. While the binomial model severely overestimated the variance, the Hawkes model underestimated the variance only slightly, in return overestimating the switching strongly. Hence, more optimistically formulated, a single filter may be insufficient, but hybrid approaches may be pursued if still in line with (iv).

We have doubts if a marginal simulation algorithm that uses filters can increase simulation efficiency compared to a Doob-Gillespie-algorithm on the joint system. For any such marginal simulation approach, auxiliary variables must be coevolved. In all examples we considered, there were at least as many auxiliary variables as environmental species. Both the Hawkes filter and the multinomial filter had a tie between auxiliary variables and environmental species, whereas the Snyder filter and the Gamma filter had an increase in state variables. Even with a tie in the state space, this replacement comes at the cost of losing the runtime advantages of Doob-Gillespie's next reaction method. Moreover, generally, ODEs must be coevolved. Closed form solutions for the filter, efficiently simulated using Sinzger's min-thin, were an exception rather than a rule. Even in this case, the Doob-Gillespie algorithm is faster. In addition, there is an accuracy loss from approximate filters. It remains unclear, whether there exists a case with many environment species that can be reduced to only a few effective auxiliary variables. Even in this case, we consider it more likely that a modified Doob-Gillespie algorithm can be employed to eliminate some species using quasi-steady state assumptions. In contrast, simulation efficiency is possibly gained in the case with a continuous environment, e.g., a concentration evolving according to an SDE. In this case, the environment might alternatively be realized as a CTMC with matched stochastic properties, to return to the Doob-Gillespie algorithm as the method of choice.

10.1.2 Design of approximate filters

We advocate that, before the design of approximate filters, one should initially define the approximation goal. As seen with the comparison of the Hawkes and multinomial filters, approximations can have complementary benefits and weaknesses. Alternatively, as seen with the Hawkes and Gamma filter, efficiency can be increased without the loss of accuracy.

We anticipate that the optimal linear filter can be used as a base case which can be adjusted and improved. We suggest adjustments in a controlled and hierarchical way to account for structural properties at each level of the adjustment. We imagine that a principled approach for the adjustments can take at least two routes. First, the base case uses a linearization in the trajectory of reaction counts. A canonical adjustment is the incorporation of higher-order terms in increasing order. Since we covered the case of first-order dependence, the next step could be a second-order dependence. A second route could use variational methods to find the optimal functional form of improvements, similar to [196]. Thereby, we would not impose the polynomial order, but account for trajectory features. For instance, this could be hierarchically achieved by an increasing family of sigma-algebras that control for the trajectory features. The structural results revealed that the Hawkes model and the doubly stochastic model coincide in the covariance measure. A matched covariance measure can also guide the design of approximations for non-linear systems. It is the second term when expanding the logarithm of the characteristic functional of the random measure associated with the counting process in terms of cumulants [115, §9.5, proposition 9.5.V]. Higher order terms can successively be matched.

While the construction of the optimal linear filter used the quadratic loss criterion, other loss criteria can be considered. Atar and Weissman showed that the natural loss criterion provides a principled choice that is well-suited for positive random quantities [85]. In figure 44 we illustrate how the optimal Hawkes parameters are altered when the natural loss criterion is used. Projection criteria that differ from the quadratic loss criterion could be applied to rule out the possibility that the filter becomes negative.

The usefulness of the ACID for the design of approximate filters might be strengthened if we find
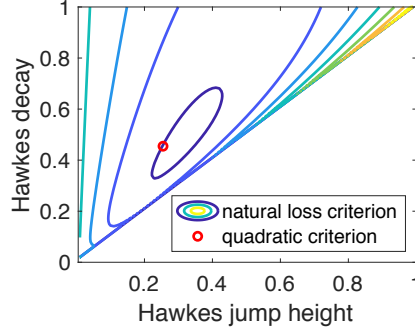
Figure 44: Optimal Hawkes parameters. Exemplarily, we chose a random telegraph modulated counting process model. Consider the task of finding the optimal one-dimensional Hawkes model $\hat{\lambda}^{\mathrm{H}}(t)$ with parameters as in Eq. (58). Level sets indicate the natural loss criterion $\lim_{T \to \infty} \frac{1}{T} \int_0^T \ell(\hat{\lambda}(t), \hat{\lambda}^{\mathrm{H}}(t)) \, \mathrm{d}t$ in the (β, α)-plane. The red circle locates the optimum with respect to the quadratic criterion. The stationary mean $\mu_0 \alpha / \beta$ of the Hawkes intensity is equal for both criteria.

statements that allow conclusions of the following forms. (i) If the distributions of $\lambda_\infty^{(1)}$ and $\lambda_\infty^{(2)}$ are close and the deterministic dynamics $F^{(1)}$ and $F^{(2)}$ are close, then the jump updates $f^{(1)}$ and $f^{(2)}$ are close. (ii) If $\lambda_\infty^{(1)}$ and $\lambda_\infty^{(2)}$ are close and $f^{(1)}$ and $f^{(2)}$ are close, then $F^{(1)}$ and $F^{(2)}$ are close. This is subject for future research that might allow concluding the closeness of the path measures from the closeness of two ACIDs under mild additional conditions.

For the binomial model we observed and proved that it systematically overestimates the exact variance of the Markov-modulated counting process. The binomial ansatz is limited to only one free parameter. Within the theory of approximate filters, this is an interesting observation, since it provides an example where limited free parameters imply that the second-order moment cannot be captured properly. In contrast, both the Gamma filter and the optimal linear filter shared the variance decomposition with the exact model. The finding on the binomial filter suggests a road map for future improvement of the filter ansatz. For assumed density filtering, distributions that are underdispersed compared to the binomial distribution should be preferred to replace the conversion environment. A canonical way to introduce more degrees of freedom would be the use of binomial mixture models. However, these generally increase the variance and are thus not suited. Underdispersion models, such as the underdispersed continuation of the beta-binomial distribution [197], can be promising. Independent of the precise route taken to design approximate filters in future works, this thesis provides a stepping stone toward a systematic study of approximate filter schemes, that can foster better understanding of environment effects and advance the discovery of new schemes tailored to the specific needs in the biological context.

10.2 Model reduction and generalized master equations

The goal of model reduction approaches is to arrive at generalized master equations and process equations. For this purpose, we used stochastic conditioning on either (i) the environment or (ii) the subnetwork. For (i), we targeted only deterministic quantities, while for (ii), we also arrive at process equations via stochastic filtering. This comes at the expense of an augmented state space, i.e., the sufficient statistics of joint Markovian progression. In the case of the Snyder filter, this is the posterior distribution with the corresponding filtering equation.

The Markovian nature of the filtering equation was also paraphrased as a 'mysterious' concept of recursiveness by Brémaud [59, p.84]. Back in the days, more emphasis was on its purpose of saving memory-space. *A priori*, the CI depends on the history of $(Y_t)_{t \geq 0}$. With the dependence only on the current state, there is no need to record the history. In contrast, we use the insight on the Markovian nature to analyze the asymptotic conditional intensity distribution. More generally, Daley and Vere-Jones discuss the representation of counting processes with a CI by a Markov process whose states represent past histories. They emphasize that the usefulness of this observation is limited, although there are cases where the 'joint statistics' [115, Example 10.3(e), p.101-102] can be reduced to a low-

dimensional Euclidean space. The idea of resorting to the governing Markov process to describe the convergence to the equilibrium of point processes was also explored [115, §12.5, p.225-229]. However, this idea was not linked to the theory of PDMPs in particular. We spotlight the subclass of Dirac-PMDPs as a novelty for which we provide the related expressions, such as the differential Chapman-Kolmogorov equation (or Liouville master equation) and the stationarity condition of the embedded Markov chain.

The simulation-free method for computing the limiting object, i.e., the ACID, is modular. The parametrization of the sufficient statistic θ can be varied. The discretization scheme, used here to find p_0 according to Eq. (93) or (94), can be replaced with another technique. The normalization constant can be employed from Eq. (95), or the resulting distribution $p(\tau, \theta)$ can be normalized. Depending on whether the entire ACID or a summary statistic, such as variance or $\mathbb{E}[\phi(\cdot)]$, is of interest, the method can be modified. We expect that there is room for improvement: when substituting single modules in the method, the precision might be increased, computation time decreased and limitations relaxed. While our grid discretization was an ad-hoc approach, we consider our main contribution to lie in the formulation of the BReT-P and the derivation of the integral boundary conditions, i.e., the stationarity condition of the embedded Markov chain for Dirac-PDMPs. For the Hawkes process, to the best of our knowledge, only the difference-differential equation (135) was available so far [60, 7.2.5 (ii)].

For the example 6.1.2 of dimension zero, we entered the territory of renewal processes, as discussed by [115, §10.3, p.95 & example 12.5(a), p.227]. Then Eq. (272) for ρ expresses the known fact that the sojourn time for the marginal distribution is a mixture of two exponential distributions [198]. Renewal processes in general fall within our framework. They are characterized by $\dim(Y_t)_{t \geq 0} = 0$, as the backward recurrence time $\tau(t)$ is sufficient for their description. Their CI is $\hat{\lambda}_t = -\frac{\dot{\rho}(\tau(t))}{\rho(\tau(t))}$ for the survival function $\rho(\tau)$. It is well established that the stationary distribution of τ has the probability density $\frac{\rho(\tau)}{\int \rho(u) \, du}$ [121, §5.1, Eq. (3), p. 61] and the theorem 4.28 reduces to stating that the expected hazard, called the renewal density, is equal to $(\int \rho(u) \, du)^{-1}$ [121, p. 62]. The transformation $\tau \mapsto -\frac{\dot{\rho}(\tau)}{\rho(\tau)}$ provides the ACID, i.e., the hazard.

Faced with the problem of the intractable augmented state space, we require approximations. The study of the approximation via optimal linear filtering revealed that the Hawkes model and the doubly stochastic model coincide in the covariance measure. This hints at a future application for model reduction, as the covariance measure is the second term in a cumulant expansion of the random measure associated with counting processes.

Our Hawkes model is especially well-tailored for linear environment that modulate zeroth-order reactions. Consequently, we expect that it can be used to replace linear leafs in the reaction network graph and linear transit networks. As a word of caution, we would like to recall that, for this kind of model reduction to work, the replaced linear environment must not be bi- or multimodal, as we have seen in the case study 8.1.1. Multimodality also includes discrete-state environment models with few states.

The optimal linear filter for counting processes has seen some application [199, 200]. Alternatively, the Kalman filter has been used after applying a diffusion approximation [201], in the linear noise approximated case [202], or for a continuous model to begin with [203]. Our assessment of the literature suggests that the approach via diffusion approximations has been more common. In contrast, we emphasize that the Kalman-like approach by Snyder, on which we built the Hawkes model, does not require a diffusion approximation. The link between the Kalman filter for a Gaussian setup and the Kalman-like filter for point processes is a shared integral equation for the estimation kernel. With this insight, Snyder transferred the results from the continuous-time additive white Gaussian noise channel to point processes.

The reduction techniques of approximate filtering and truncation of the cumulant expansion can have difficulties in capturing the asymptotic mean correctly, as seen in the case studies 8.3.8, 8.4 and [106, p.70]. The expressions for the stationary mean might assist in detecting the parameter regimes in which approximations succeed. On the one hand, ESME does not require Monte Carlo approximations. On the other hand, our quantification of environmental shares can permit insight into the failure mode of the approximate simulation or model reduction technique. Besides the detection of parameter regimes, the knowledge of the exact stationary mean can tune approximation methods towards capturing the asymptotic mean correctly.

At the same time, our expressions are limited to a particular class of CRNs, i.e., the ones that can be decomposed into a linear subsystem and an environment, where the modulation is only allowed unidirectionally, from the environment to the subsystem. Often, such decomposition does not exist, because the requirements of linearity and unidirectional modulation are limiting. Let us consider the common violations of these requirements. (1) There exists a reaction $A + B \rightarrow C$ where at least one of the three species is in the subsystem, e.g., the MAPK/ERK pathway [204]. If A is an environmental species, it needs to be preserved during the reaction, and A and B cannot be both in the subsystem due to the linearity assumption. (2) Any partition into two sets of species contains bimolecular reactions of the forms $A_i + B_i \rightarrow A_i + *$ with A_i in one set and B_i in the other set for one i and the other way around for another i. Finally, (3) reaction rates that have a Hill or other non-linear dependency violate the linearity constraint. We suggest that, for non-linear propensities, Monte Carlo simulations with the method of conditional moments [34] are used to obtain numerical estimates of the stationary mean. Additionally, our expressions require the stationary distribution of the environment, which further limits our approach.

Previously, it was believed that generalized master equations decrease complexity over master equations of joint Markov models, especially when alternatives, such as quasi-steady state model reductions, are too simple to be exhaustive. In our studies, we could not validate that generalized master equations with this benefit are within reach. Instead, the unifying approach showed that generalized master equations require evaluating conditional expectations of random propensities, given the state of the system. For the telegraph-modulated birth-death process, we derived evolution equations via conditioning on the environment and conditioning on the subnetwork. For this tractable joint system, we found that both the approach via the cumulant expansion and via stochastic filtering arrived at the same equations as the classical CME for the joint system. In the Gamma-mediated decay reactions, complex calculations via stochastic filtering yield the probability distribution of the subsystem. However, the solution is more simply obtained by conditioning on the Gamma random variable and solving the classical CME, i.e., the route of section 3.

The ansatz of the generalized CME via conditioning on the subnetwork, section 4, is the replacement of environmental species by their state estimates, which become the new continuous state variables. In the studied examples, the necessary number of state variables was at least as high as the number of environmental species. More commonly, an augmented state space is needed to recover the Markov property. Moreover, replacing discrete by continuous states implies the need for challenging hybrid approaches. The advantages of this ansatz over joint Markov approaches in terms of reducing the complexity have not become apparent in our studies. Overall, we conclude that (i) the curse of dimensionality is not resolved by the two approaches in sections 3 and 4, and (ii) between both approaches, the route in section 3 might be preferred for analytic approaches, as it parallels the structure of the generative model, which defines the subnetwork conditionally on the environment.

10.3 Computation of information measures

We contributed to information theory in cell biology by presenting a Monte Carlo-free numerical computation of the path mutual information rate of a Poisson channel for Markovian input with a low state number. We aimed at computing the MIR and RER in a simulation-free way, i.e., without Monte Carlo simulations. This involved the evaluation of ODEs on a grid with the dimension given by the dimension of the sufficient statistic. We considered doubly stochastic Poisson processes (signal process along Poisson channel) and self-exciting counting processes. The ACID is accessible by our method in the case of very low number of non-zero signal states. With no limitations on the number of zero states, the method enables us to analyze binary semi-Markov inputs, i.e., have non-exponential phase-type Off-times. We demonstrated how the RER can compare two counting processes.

In section 8.2.3, we compared the Gamma filter and the optimal linear filter in their ability to estimate the mutual information in the two-state model of gene expression. We noted that the optimal linear filter improves the computation time with no decrease in accuracy. The conditions under which either of the two approximate filters is preferable still have to be determined more systematically. The characterization result for the optimal linear filter can be used to guide the intuition.

For the Poisson channel, we considered the input class restricted to binary processes that have a Markov representation with one or two active states. For this class, we were able to contribute a novel

MI optimization method that does not require costly Monte Carlo simulation. Instead, we took an analytic approach based on ODEs and a fixed-point iteration. With this approach, we observed interesting effects, such as optimal promoter cycling and a favored allocation of the On state. Ultimately, it is the goal to add one more stochastic layer in the channel. Namely, in gene regulatory networks, the channel between a transcription factor as an input and mRNA transcription as an output is more interesting. The promoter serves as a mediator. Mathematically, this requires another external stochastic process that modulates the rates of the promoter switching. It is still unclear, whether the same techniques that use the ACID and the BReT-P can be applied.

10.4   Attribution of subnetwork features to the environment

We contributed several decomposition results. The Fano factor decomposition for mRNA structure switching showed its contribution to translational noise. The spectral formula for correlated environment has the common structure of intrinsic and extrinsic noise components. As a further novelty, we provided the method of environmental shares that quantifies the contribution of the environmental states to the stationary mean. The method of the environmental shares allowed to correctly attribute the excursions to the zero state of the environment. Moreover, it showed an interesting decomposition of the parameter regions, in which different environment states dominate. Beyond these positive results, we also provided a line of argument why the decomposition of subnetwork noise into effective and suppressed noise as proposed by [3] seems misconceived.

In the case study 8.3.8, we found that excursions persist under the linearization, and for this reason, we classified them as results of linear environment features, at least qualitatively. The same holds for the local maximum in the case study 8.4, figure 30B. Non-linear features of the environment, such as bimodality, skewness, a non-linear evolution of the intensity, and a discrete state space with few states, are expected to cause deviations in $X(t)$. These deviations suggest using more details than first- and second-order moments of the stochastic description of $Z(t)$. In the modeling context, if the Hawkes model lacks specifics of the observed $X(t)$, this indicates that non-linear environmental features may cause these specifics. The case study 8.2.5 showed that the Hawkes model is not generally incapable of bimodal distributions if the bimodality results from the architecture of the embedded system instead of being caused by the bimodality of the environment.

To benefit future research on state estimation, the partial observations of the system that is being modulated could be used to infer the three characteristics of the environment using the proposed Hawkes model. For this purpose, it would be desirable to derive path likelihoods for the Hawkes-modeled CRNs when it is reasonable to assume that the environment has an exponentially decaying auto/cross-covariance. Overall, among the studied applications, we see the largest benefit of approximate filtering and conditioning on the environment in the areas of information theory and attribution theory. Disentangling intrinsically stochastic effects of the subnetwork and extrinsic effects of different environment components beyond the decomposition of the variance can improve our fundamental understanding of how systems are shaped by being embedded and advance the design of synthetic biology in a specific cellular context.

A Piecewise-deterministic Markov processes

Here, the piecewise-deterministic Markov process is redefined. We follow the construction of Davis [65, §24, p.57-59 & (24.8) standard conditions, p.62] step by step.

Construction A.1 (Piecewise-deterministic process). *Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. And let $U_1, V_1, U_2, V_2, \ldots$ be independent and uniformly distributed random variables $\Omega \to [0, 1]$. Let $\vartheta \subseteq \mathbb{R}^{n_0}$ be given. This will be the state space of the constructed process. Let $\mathcal{B}(\vartheta)$ denote the Borel sets of ϑ and let further three function be given*

- *$F \colon \vartheta \to \mathbb{R}^{n_0}$ locally Lipschitz continuous*

- *$l \colon \vartheta \to \mathbb{R}_{\geq 0}$ measurable satisfying an integrability condition, specified in Eq. (290)*

- *$Q \colon \mathcal{B}(\vartheta) \times \vartheta \to [0, 1]$, such that $Q(B, \cdot) \colon \vartheta \to [0, 1]$ is measurable for all $B \in \mathcal{B}(\vartheta)$, where $\mathcal{B}$ denotes the Borel σ-algebra, and $Q(\cdot, \theta)$ is a probability measure for all $\theta \in \vartheta$.*

Then with these characteristics construct the following piecewise-deterministic process $(\Theta_t)_{t \geq 0}$ starting from the fixed initial value $\Theta_0 := S_0 := \theta \in \vartheta$ at $\sigma_0 := 0$. First, for each $v^0 \in \vartheta$ denote by $[0, \infty) \to \vartheta, \tau \mapsto u(\tau, v^0)$ the solution of the initial value problem

$$\dot{\Theta}_t = F(\Theta_t), \quad \Theta_0 = v^0$$

and assume that there exists an $\varepsilon(v^0) > 0$, such that

$$\int_0^{\varepsilon(v^0)} l(u(\tau, v^0)) \, \mathrm{d}\tau < \infty. \tag{290}$$

Now we construct the process recursively, for $i = 0, 1, \ldots$. Suppose, that we have constructed the process until $t = \sigma_i$ and $\Theta_{\sigma_i} = S_i$. Then by the standard technique of the inverse cumulative distribution function, let $F_{i+1} \colon [0, 1] \to [0, \infty)$ such that $T_{i+1} = F_{i+1}(U_{i+1})$ is distributed as

$$\mathbb{P}[T_{i+1} > t] = \exp\left(-\int_0^t l(u(\tau, S_i)) \, \mathrm{d}\tau\right)$$

and set $\sigma_{i+1} := \sigma_i + T_{i+1}$. Next, set $\Theta_t = u(t - \sigma_i, S_i)$ for $t \in (\sigma_i, \sigma_{i+1})$ and let $\tilde{F}_{i+1} \colon [0, 1] \to \vartheta$ be such that $S_{i+1} = \tilde{F}_{i+1}(V_{i+1})$ is distributed as $Q(\cdot, u(T_{i+1}, S_i))$. The process has then been constructed until $t = \sigma_{i+1}$ and $\Theta_{\sigma_{i+1}} := S_{i+1}$.

According to Davis [65, §25, theorem 25.5, p.64], this process has the strong Markov property.

Definition A.1 (Piecewise-deterministic Markov process, embedded counting process). A Markov process $(\Theta_t)_{t \geq 0}$ on the state space ϑ is called **piecewise-deterministic Markov process** (PDMP), if its finite-dimensional marginal distributions agree with the construction A.1 for some triple (F, l, Q). Then (F, l, Q) are called the local characteristics of the PDMP and the process $Y_t := \sum_{i=1}^{\infty} \mathbb{1}(\sigma_i \leq t)$ with $(\sigma_i)_{i \in \mathbb{N}}$ as in the construction A.1 is called the **embedded counting process**.

Remark A.2. The class of PDMPs as studied by Davis allows an *active boundary* meaning that an entry into the active boundary triggers jumps automatically [65, §24, p.60 and §34, p.116]. In our setting there is no active boundary.

B Invertibility of matrices for promoter-mediated transcription

This section discusses how to make A and $A + \bar{B}C^T\mu^{-1}C$ invertible. The matrix A in Eq. (137) has vanishing column sums, i.e., $\mathbf{1}^T A = \mathbf{0}^T$ for the vector $\mathbf{1} = [1, \ldots, 1]^T$. For this reason A is non-invertible. The matrix Σ^2 has column and row sums zero, i.e., $\mathbf{1}^T \Sigma^2 = \mathbf{0}^T$ and $\Sigma^2 \mathbf{1} = \mathbf{0}$. We derive an equation equivalent to Eq. (65) - Eq. (66) for the reduced state $\tilde{Z}_0(t) \in \mathbb{R}^{k-1}$, dropping the k-th coordinate of $Z_0(t)$. This way, we make A invertible by reducing its dimension.

To this end, we define Γ and decompose the matrices A and Σ^2 as

$$\Gamma := \begin{bmatrix} \mathbf{I}_{k-1} \\ -\mathbf{1}^T \end{bmatrix} \in \mathbb{R}^{k \times (k-1)}, \quad A = \begin{bmatrix} A_0 \\ -\mathbf{1}^T A_0 \end{bmatrix} = \Gamma A_0, \quad \Sigma^2 = \begin{bmatrix} \tilde{\Sigma}^2 & -\tilde{\Sigma}^2 \mathbf{1} \\ -\mathbf{1}^T \tilde{\Sigma}^2 & \mathbf{1}^T \tilde{\Sigma}^2 \mathbf{1} \end{bmatrix} = \Gamma \tilde{\Sigma}^2 \Gamma^T,$$

with $A_0 \in \mathbb{R}^{(k-1) \times k}$, $\tilde{\Sigma}^2 \in \mathbb{R}^{(k-1) \times (k-1)}$. Then we define $\tilde{A} := A_0 \Gamma \in \mathbb{R}^{(k-1) \times (k-1)}$, $\tilde{C} := C\Gamma \in \mathbb{R}^{1 \times (k-1)}$. Let $\tilde{B}(t) \in \mathbb{R}^{(k-1) \times (k-1)}$ solve

$$\frac{\mathrm{d}}{\mathrm{d}t} \tilde{B} = -\tilde{A}\tilde{B} - \tilde{B}\tilde{A}^T + \tilde{A}\tilde{\Sigma}^2 + \tilde{\Sigma}^2 \tilde{A}^T - \tilde{B}\tilde{C}^T \mu^{-1} \tilde{C}\tilde{B}.$$

Then we derive the evolution equation for

$$\beta := \begin{bmatrix} \tilde{B} & -\tilde{B}\mathbf{1} \\ -\mathbf{1}^T \tilde{B} & \mathbf{1}^T \tilde{B}\mathbf{1} \end{bmatrix} = \Gamma \tilde{B}\Gamma^T.$$

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t} \beta = & -\Gamma\tilde{A}\tilde{B}\Gamma^T - \Gamma\tilde{B}\tilde{A}^T\Gamma^T + \Gamma\tilde{A}\tilde{\Sigma}^2\Gamma^T + \Gamma\tilde{\Sigma}^2\tilde{A}^T\Gamma^T - \Gamma\tilde{B}\tilde{C}^T\mu^{-1}\tilde{C}\tilde{B}\Gamma^T \\
= & -\Gamma A_0\Gamma\tilde{B}\Gamma^T - \Gamma\tilde{B}\Gamma^T A_0^T\Gamma^T + \Gamma A_0\Gamma\tilde{\Sigma}^2\Gamma^T + \Gamma\tilde{\Sigma}^2\Gamma^T A_0^T\Gamma^T - \Gamma\tilde{B}\Gamma^T C^T\mu^{-1}\Gamma C\tilde{B}\Gamma^T \\
= & -A\beta - \beta A^T + A\Sigma^2 + \Sigma^2 A^T - \beta C^T\mu^{-1}C\beta.
\end{aligned}
$$

Hence $\beta(t) = B(t)$ for all $t \geq 0$. Let further $\tilde{Z}_0(t) \in \mathbb{R}^k$ satisfy

$$\mathrm{d}\tilde{Z}_0(t) = -\tilde{A}\tilde{Z}_0(t) + \tilde{B}(t)\tilde{C}^T\mu^{-1}(\,\mathrm{d}\tilde{Y}(t) - (\mu + \tilde{C}\tilde{Z}_0(t))\,\mathrm{d}t),$$

where the intensity of $\tilde{Y}(t)$ is assumed to be $\mu + \tilde{C}\tilde{Z}_0(t)$. Then $\Gamma\tilde{Z}_0(t)$ satisfies

$$
\begin{aligned}
\mathrm{d}\Gamma\tilde{Z}_0(t) = & -\Gamma\tilde{A}\tilde{Z}_0(t) + \Gamma\tilde{B}(t)\tilde{C}^T\mu^{-1}(\,\mathrm{d}\tilde{Y}(t) - (\mu + \tilde{C}\tilde{Z}_0(t))\,\mathrm{d}t) \\
= & -A\Gamma\tilde{Z}_0(t) + \Gamma\tilde{B}(t)\Gamma^T C^T\mu^{-1}(\,\mathrm{d}\tilde{Y}(t) - (\mu + C\Gamma\tilde{Z}_0(t))\,\mathrm{d}t) \\
= & -A\Gamma\tilde{Z}_0(t) + B(t)C^T\mu^{-1}(\,\mathrm{d}\tilde{Y}(t) - (\mu + C\Gamma\tilde{Z}_0(t))\,\mathrm{d}t).
\end{aligned}
$$

Hence for $\tilde{Z}_0(0)$ equal to the first $k-1$ entries of $Z_0(0)$, we have $\Gamma\tilde{Z}_0(t) = Z_0(t)$ for all $t \geq 0$ and in particular $\hat{Z}^{\mathrm{H}}(t) = \mu + \tilde{C}\tilde{Z}_0(t)$, hence $\tilde{Y}(t)$ is equal to $Y(t)$ in distribution.
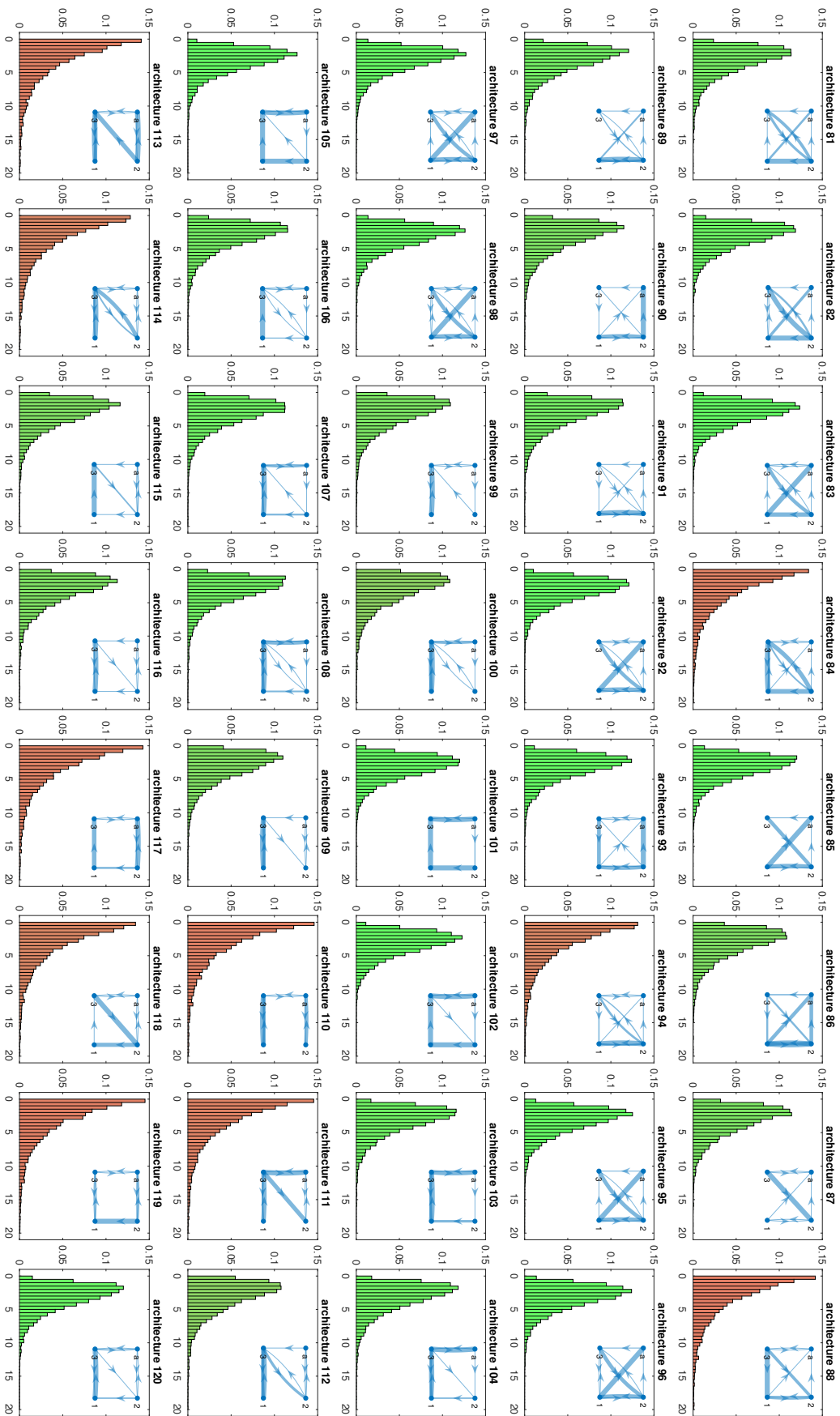
C   Promoter architectures

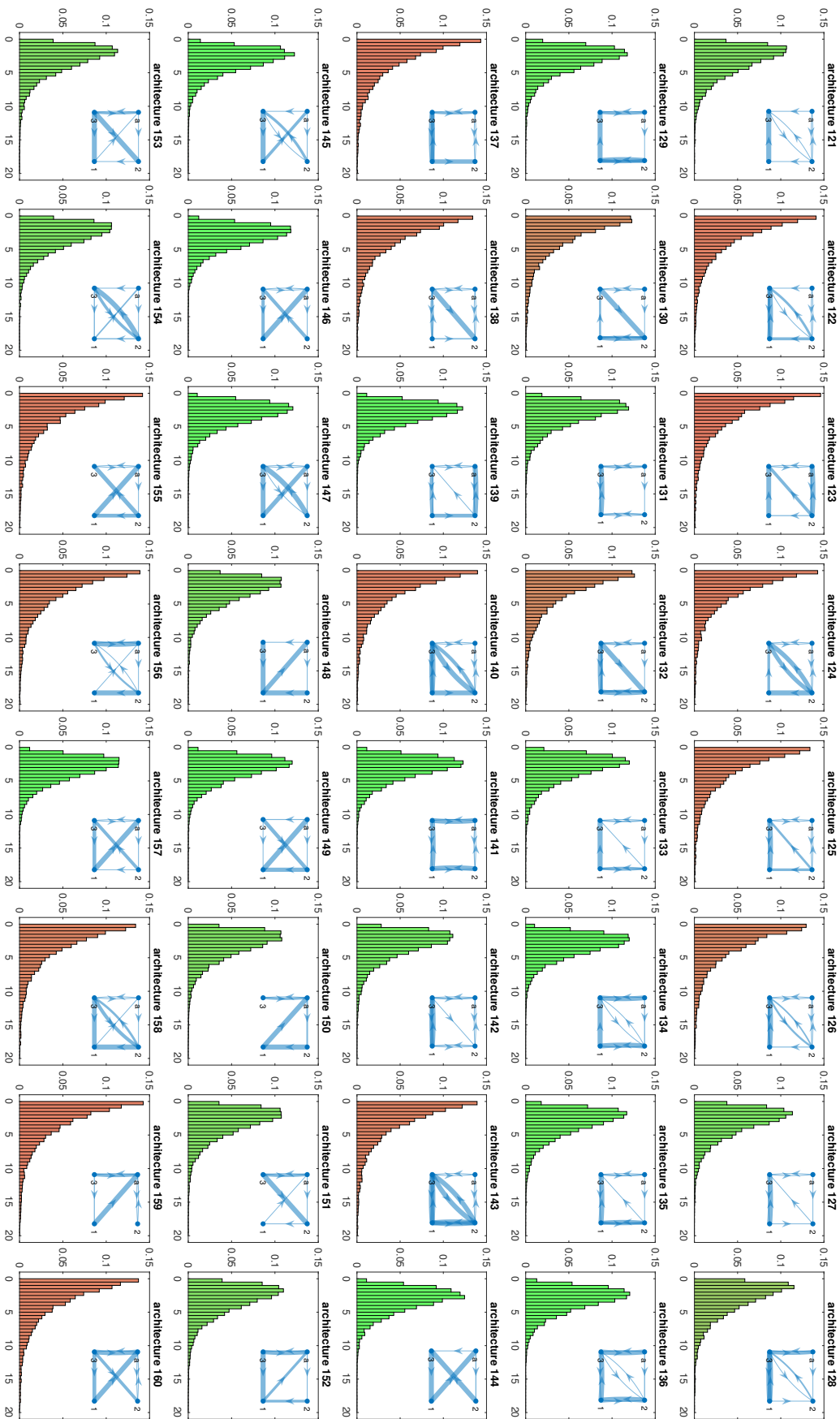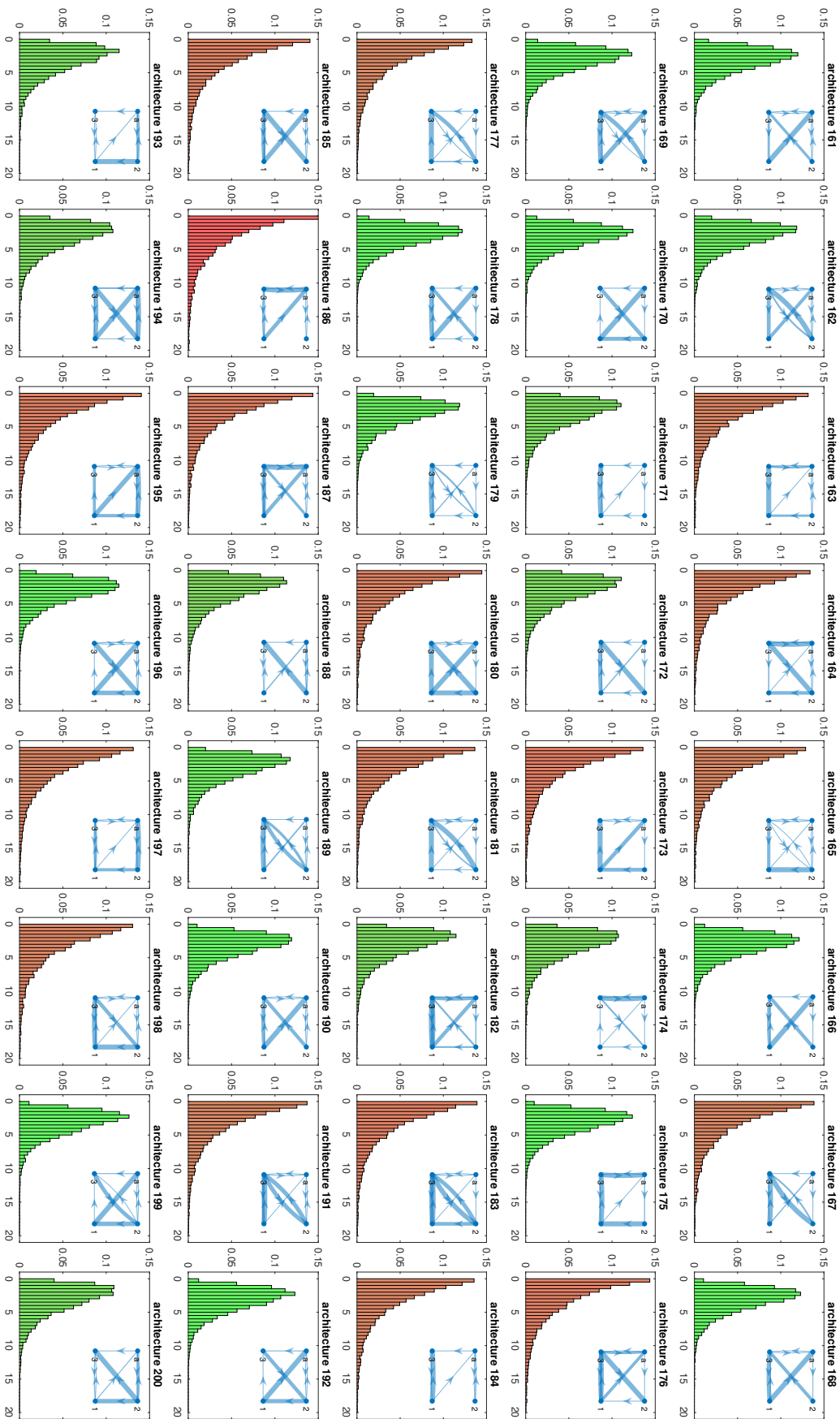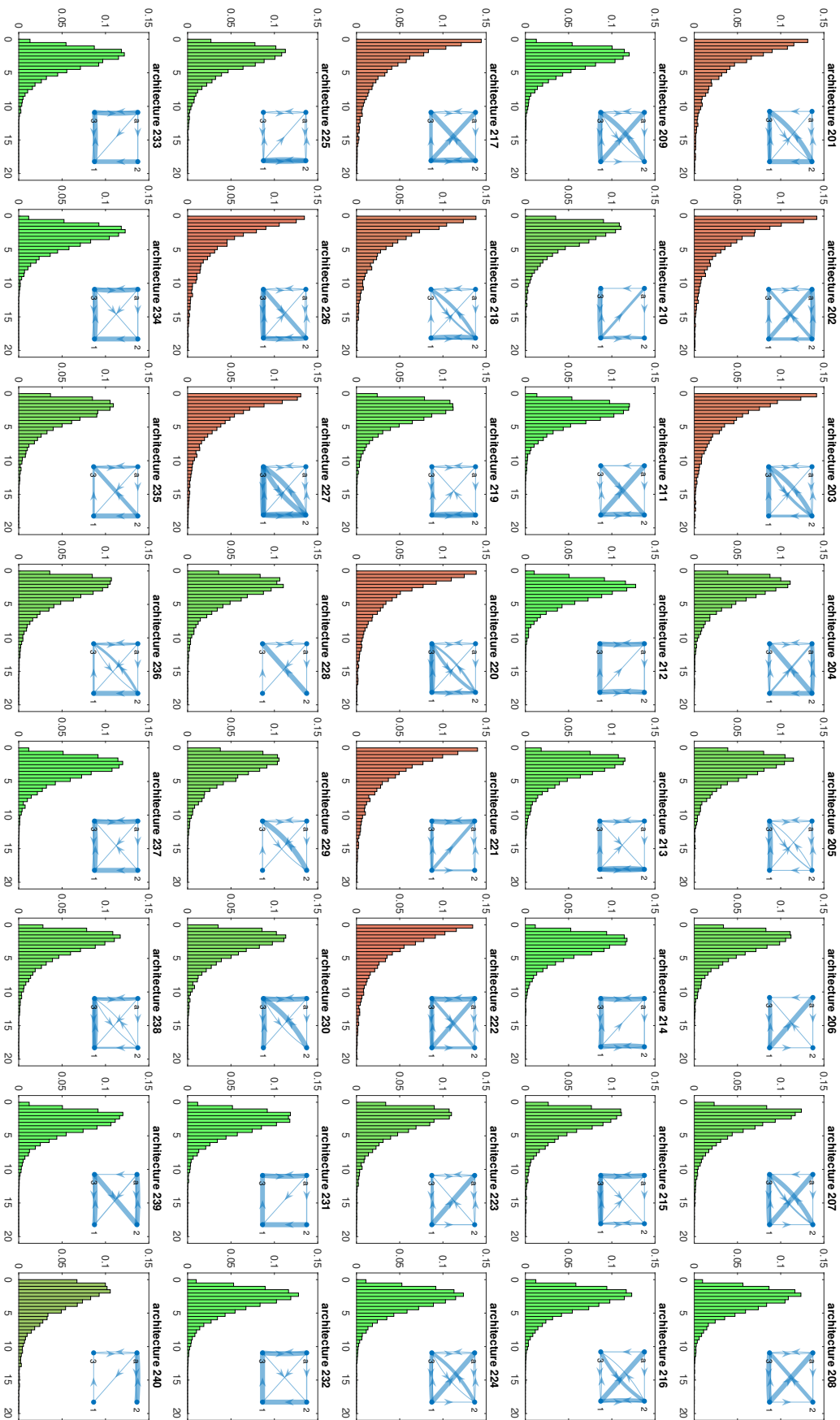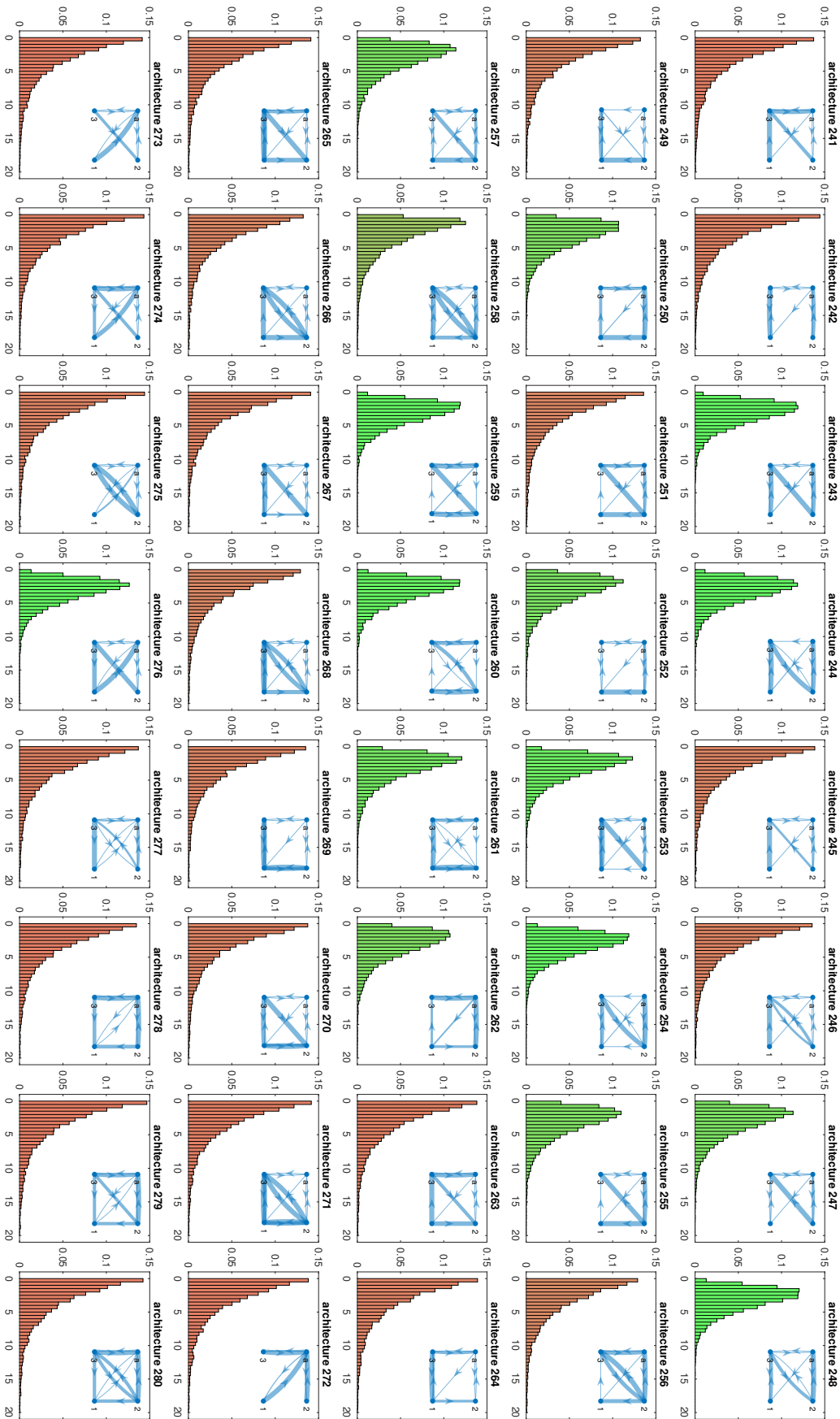Results from the experiment described in section 9.3.

OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5
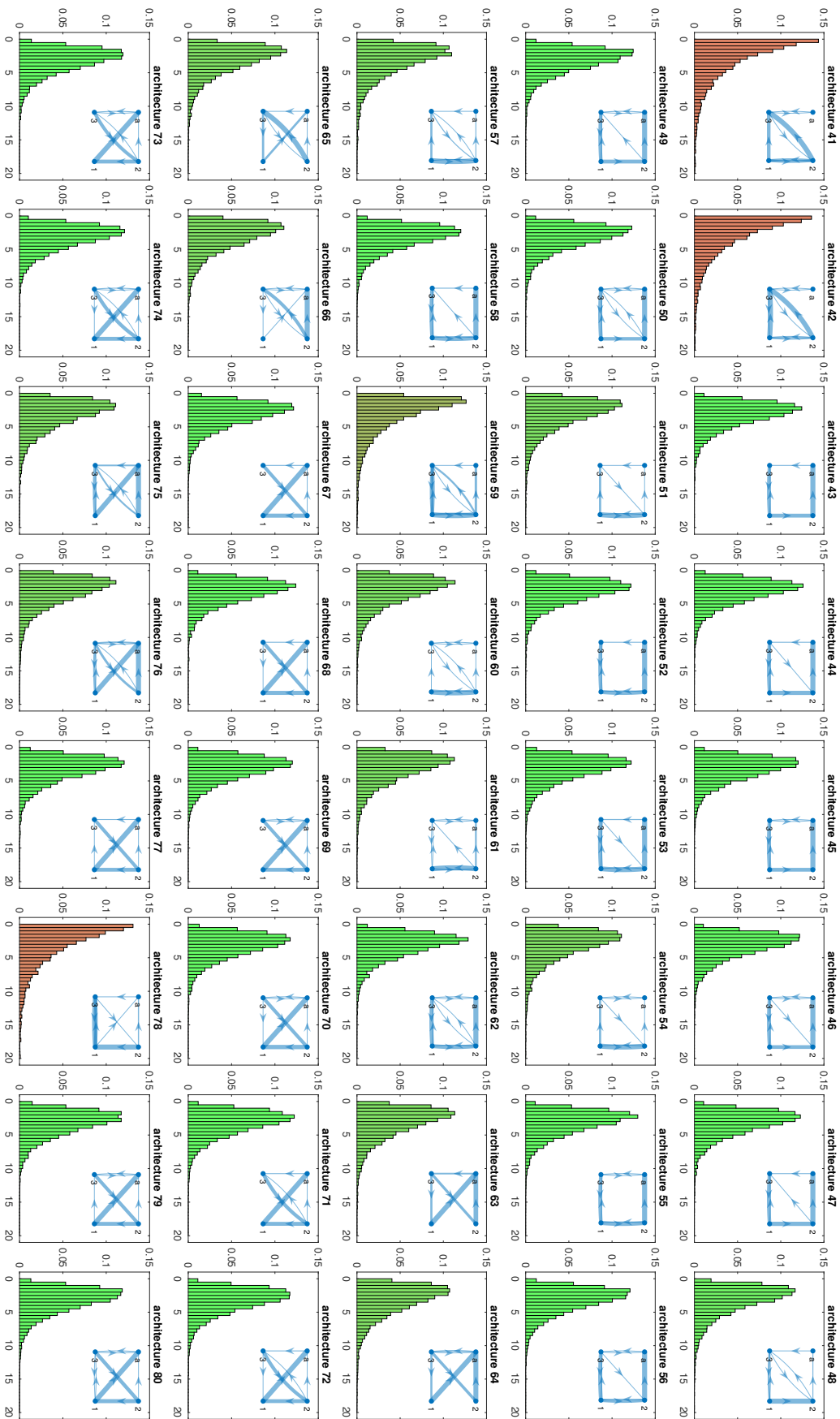
OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5

OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5

OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5

176

OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5

OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5

OFF-time distribution for 3 state architectures, OFF-mean = 3.3, ON-mean = 5
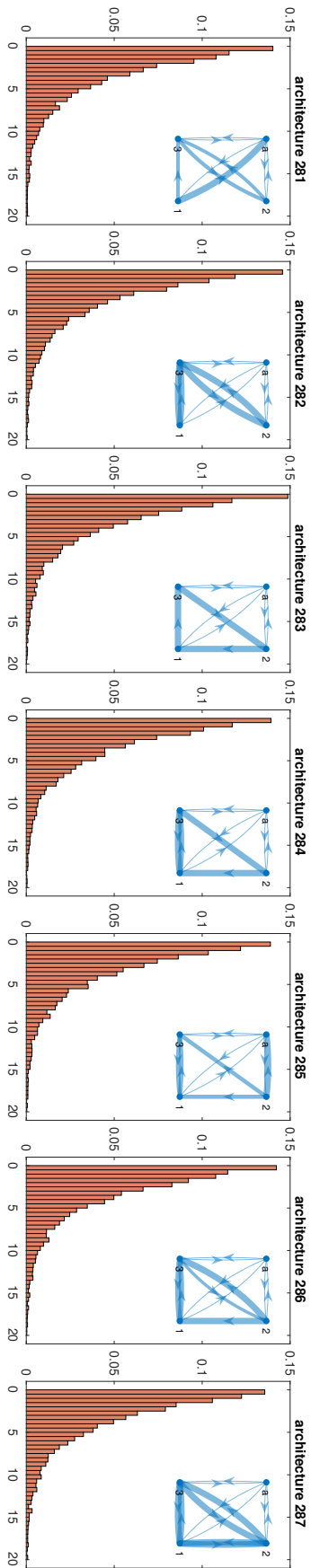
References

[1] Heinz-Peter Breuer and Francesco Petruccione. *The theory of open quantum systems*. Oxford University Press, 2002. ISBN: 9780198520634.

[2] Michael B Elowitz et al. "Stochastic gene expression in a single cell". In: *Science* 297.5584 (2002), pp. 1183–1186.

[3] Christoph Zechner and Heinz Koeppl. "Uncoupled analysis of stochastic reaction networks in fluctuating environments". In: *PLoS Computational Biology* 10.12 (Dec. 2014), e1003942. DOI: 10.1371/journal.pcbi.1003942.

[4] Stijn Sonneveld, Bram MP Verhagen, and Marvin E Tanenbaum. "Heterogeneity in mRNA translation". In: *Trends in Cell Biology* 30.8 (2020), pp. 606–618.

[5] Johan Paulsson. "Summing up the noise in gene networks". In: *Nature* 427.6973 (2004), pp. 415–418.

[6] Avigdor Eldar and Michael B Elowitz. "Functional roles for noise in genetic circuits". In: *Nature* 467.7312 (2010), pp. 167–173.

[7] Jina Ko et al. "Spatiotemporal multiplexed immunofluorescence imaging of living cells and tissues with bioorthogonal cycling of fluorescent probes". In: *Nature Biotechnology* 40.11 (2022), pp. 1654–1662.

[8] Nadia Sarfraz et al. "Visualizing orthogonal RNAs simultaneously in live mammalian cells by fluorescence lifetime imaging microscopy (FLIM)". In: *Nature Communications* 14.1 (2023), p. 867.

[9] Ru Zheng et al. "Multiplexed sequential imaging in living cells with orthogonal fluorogenic RNA aptamer/dye pairs". In: *bioRxiv* (2023).

[10] Christopher V Rao and Adam P Arkin. "Stochastic chemical kinetics and the quasi-steady-state assumption: Application to the Gillespie algorithm". In: *The Journal of Chemical Physics* 118.11 (2003), pp. 4999–5010.

[11] Yang Cao, Daniel T Gillespie, and Linda R Petzold. "The slow-scale stochastic simulation algorithm". In: *The Journal of Chemical Physics* 122.1 (2005), p. 014116.

[12] Kate E Dray et al. "GAMES: A dynamic model development workflow for rigorous characterization of synthetic genetic systems". In: *ACS Synthetic Biology* 11.2 (2022), pp. 1009–1029.

[13] Rok Grah, Benjamin Zoller, Gašper Tkačik, et al. "Nonequilibrium models of optimal enhancer function". In: *Proceedings of the National Academy of Sciences* 117.50 (2020), pp. 31614–31622.

[14] M. Santillán. "On the use of the Hill functions in mathematical models of gene regulatory networks". In: *Mathematical Modelling of Natural Phenomena* 3.2 (2008), pp. 85–97. DOI: 10.1051/mmnp:2008056.

[15] Yang Joon Kim et al. "Predictive modeling reveals that higher-order cooperativity drives transcriptional repression in a synthetic developmental enhancer". In: *Elife* 11 (2022), e73395.

[16] Roberto Chignola et al. "Thresholds, long delays and stability from generalized allosteric effect in protein networks". In: *Physica A: Statistical Mechanics and its Applications* 371.2 (2006), pp. 463–472.

[17] Javier Estrada et al. "Information integration and energy expenditure in gene regulation". In: *Cell* 166.1 (2016), pp. 234–244.

[18] Jean Peccoud and Bernard Ycart. "Markovian modeling of gene-product synthesis". In: *Theoretical population biology* 48.2 (1995), pp. 222–234.

[19] Edward Tunnacliffe and Jonathan R Chubb. "What is a transcriptional burst?" In: *Trends in Genetics* 36.4 (2020), pp. 288–297.

[20] Muir Morrison, Manuel Razo-Mejia, and Rob Phillips. "Reconciling kinetic and thermodynamic models of bacterial transcription". In: *PLoS Computational Biology* 17.1 (2021), e1008572.

[21] Alvaro Sanchez, Sandeep Choubey, and Jane Kondev. "Regulation of noise in gene expression". In: *Annual Review of Biophysics* 42 (2013), pp. 469–491.

[22] Benjamin Zoller et al. "Structure of silent transcription intervals and noise characteristics of mammalian genes". In: *Molecular Systems Biology* 11.7 (2015), p. 823.

[23] Virginia L Pimmett et al. "Quantitative imaging of transcription in living Drosophila embryos reveals the impact of core promoter motifs on promoter state dynamics". In: *Nature Communications* 12.1 (2021), p. 4504.

[24] Jiajun Zhang and Tianshou Zhou. "Promoter-mediated transcriptional dynamics". In: *Biophysical Journal* 106.2 (2014), pp. 479–488.

[25] Jarno Mäkelä, Vinodh Kandavalli, and Andre S Ribeiro. "Rate-limiting steps in transcription dictate sensitivity to variability in cellular components". In: *Scientific Reports* 7.1 (2017), p. 10588.

[26] Lisa Amrhein. "Stochastic Modeling of Heterogeneous Low-Input Gene Expression: Linking Single-Cell Probability Distributions to Transcription Mechanisms". PhD thesis. Technische Universität München, 2021.

[27] Meiling Chen et al. "Exact distributions for stochastic gene expression models with arbitrary promoter architecture and translational bursting". In: *Physical Review E* 105.1 (2022), p. 014405.

[28] Donald A McQuarrie. "Stochastic approach to chemical kinetics". In: *Journal of Applied Probability* 4.3 (1967), pp. 413–478.

[29] Anthony F Bartholomay. "A stochastic approach to statistical kinetics with application to enzyme kinetics". In: *Biochemistry* 1.2 (1962), pp. 223–230.

[30] Daniel T Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434.

[31] Ivan G Darvey, BW Ninham, and PJ Staff. "Stochastic models for second-order chemical reaction kinetics. The equilibrium state". In: *The Journal of Chemical Physics* 45.6 (1966), pp. 2145–2155.

[32] Tobias Jahnke. "On reduced models for the chemical master equation". In: *Multiscale Modeling & Simulation* 9.4 (2011), pp. 1646–1676.

[33] Stephan Menz et al. "Hybrid stochastic–deterministic solution of the chemical master equation". In: *Multiscale Modeling & Simulation* 10.4 (2012), pp. 1232–1262.

[34] Jan Hasenauer et al. "Method of conditional moments (MCM) for the chemical master equation". In: *Journal of Mathematical Biology* 69.3 (2014), pp. 687–735.

[35] Ramon Grima. "A study of the accuracy of moment-closure approximations for stochastic chemical kinetics". In: *The Journal of Chemical Physics* 136.15 (2012), p. 154105.

[36] David Schnoerr, Guido Sanguinetti, and Ramon Grima. "Comparison of different moment-closure approximations for stochastic chemical kinetics". In: *The Journal of Chemical Physics* 143.18 (2015), p. 185101.

[37] Eszter Lakatos et al. "Multivariate moment closure techniques for stochastic kinetic models". In: *The Journal of Chemical Physics* 143.9 (2015), p. 094107.

[38] Leo Bronstein and Heinz Koeppl. "A variational approach to moment-closure approximations for the kinetics of biomolecular reaction networks". In: *The Journal of Chemical Physics* 148.1 (2018), p. 014105.

[39] Abhyudai Singh and João P. Hespanha. "Approximate moment dynamics for chemically reacting systems". In: *IEEE Transactions on Automatic Control* 56.2 (2011), pp. 414–418.

[40] Christian Kuehn. "Moment closure—a brief review". In: *Control of Self-organizing Nonlinear Systems* (2016), pp. 253–271.

[41] Khem Raj Ghusinga et al. "Exact lower and upper bounds on stationary moments in stochastic biochemical systems". In: *Physical Biology* 14.4 (2017), 04LT01.

[42] Yuta Sakurai and Yutaka Hori. "A convex approach to steady state moment analysis for stochastic chemical reactions". In: *2017 IEEE 56th Annual Conference on Decision and Control (CDC)*. IEEE. 2017, pp. 1206–1211.

[43] Juan Kuntz et al. "Bounding the stationary distributions of the chemical master equation via mathematical programming". In: *The Journal of Chemical Physics* 151.3 (2019), p. 034109.

[44] Peter S Swain, Michael B Elowitz, and Eric D Siggia. "Intrinsic and extrinsic contributions to stochasticity in gene expression". In: *Proceedings of the National Academy of Sciences* 99.20 (2002), pp. 12795–12800.

[45] Andreas Hilfinger, Mark Chen, and Johan Paulsson. "Using temporal correlations and full distributions to separate intrinsic and extrinsic fluctuations in biological systems". In: *Physical Review Letters* 109.24 (Dec. 2012). DOI: `10.1103/PhysRevLett.109.248104`.

[46] Clive G. Bowsher and Peter S. Swain. "Identifying sources of variation and the flow of information in biochemical networks". In: *Proceedings of the National Academy of Sciences* 109.20 (2012). DOI: `10.1073/pnas.1119407109`.

[47] Filipe Tostevin, Wiet de Ronde, and Pieter Rein Ten Wolde. "Reliability of frequency and amplitude decoding in gene regulation". In: *Physical Review Letters* 108.10 (2012), p. 108104.

[48] Ankit Gupta and Mustafa Khammash. "Frequency spectra and the color of cellular noise". In: *Nature Communications* 13.1 (2022), p. 4305. DOI: `10.1038/s41467-022-31263-x`.

[49] James Holehouse, Abhishek Gupta, and Ramon Grima. "Steady-state fluctuations of a genetic feedback loop with fluctuating rate parameters using the unified colored noise approximation". In: *Journal of Physics A: Mathematical and Theoretical* 53.40 (2020), p. 405601.

[50] CA O'cinneide and P Purdue. "The M/M/infinity queue in a random environment". In: *Journal of Applied Probability* 23.1 (1986), pp. 175–184.

[51] Melike Baykal-Gursoy and Weihua Xiao. "Stochastic decomposition in M/M/infinity queues with Markov modulated service rates". In: *Queueing Systems* 48.1 (2004), pp. 75–88.

[52] Gennadi Falin. "The M/M/infinity queue in a random environment". In: *Queueing Systems* 58.1 (2008), pp. 65–76.

[53] David F Anderson and Thomas G Kurtz. *Stochastic analysis of biochemical systems*. Vol. 674. Springer, 2015.

[54] L Bronstein, Christoph Zechner, and Heinz Koeppl. "Bayesian inference of reaction kinetics from single-cell recordings across a heterogeneous cell population". In: *Methods* 85 (2015), pp. 22–35.

[55] Zhixing Cao and Ramon Grima. "Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells". In: *Proceedings of the National Academy of Sciences* 117.9 (2020), pp. 4682–4692. ISSN: 1091-6490. DOI: `10.1073/pnas.1910888117`.

[56] Tineke L Lenstra et al. "Transcription Dynamics in Living Cells". In: *Annual Review of Biophysics* (2016). DOI: `10.1146/annurev-biophys-062215-010838`.

[57] John C. Cox, Jonathan E. Ingersoll, and Stephen A. Ross. "A Theory of the Term Structure of Interest Rates". In: *Econometrica* 53.2 (1985), pp. 385–407. ISSN: 00129682, 14680262.

[58] Uri T Eden and Emery N Brown. "Continuous-time filters for state estimation from point process models of neural data". In: *Statistica Sinica* 18.4 (2008), p. 1293.

[59] Pierre Brémaud. *Point processes and queues: martingale dynamics*. Vol. 50. Springer, 1981.

[60] Daryl J. Daley and David Vere-Jones. *An introduction to the theory of point processes: volume I: Elementary theory and methods*. Springer, 2003.

[61] Yu M. Kabanov. "The capacity of a channel of the Poisson type". In: *Theory of Probability & Its Applications* 23.1 (1978), pp. 143–147.

[62] Per K Andersen et al. *Statistical models based on counting processes*. New York: Springer Verlag, 1993.

[63] Crispin Gardiner. *Stochastic methods: a handbook for the natural and social sciences, 4th ed.* Springer-Verlag, 2009.

[64] Donald L. Snyder and Michael I. Miller. *Random point processes in time and space*. 2nd edition. New York: Springer-Verlag, 1991.

[65] M.H.A. Davis. *Markov models and optimization*. Vol. 49. Monographs on Statistics and Applied Probability. Chapman & Hall/CRC, 1993.

[66] Pierre M. Bremaud. "A martingale approach to point-processes". PhD thesis. EECS Department, University of California, Berkeley, 1972. URL: http://www2.eecs.berkeley.edu/Pubs/TechRpts/1972/7335.html.

[67] Lorenzo Duso and Christoph Zechner. "Selected-node stochastic simulation algorithm". In: *The Journal of Chemical Physics* 148.16 (2018), p. 164108.

[68] C Radhakrishna Rao. "Minimum variance and the estimation of several parameters". In: *Mathematical Proceedings of the Cambridge Philosophical Society*. Vol. 43. 2. Cambridge University Press. 1947, pp. 280–283.

[69] David Blackwell. "Conditional expectation and unbiased sequential estimation". In: *The Annals of Mathematical Statistics* (1947), pp. 105–110.

[70] Jennifer AN Brophy and Christopher A Voigt. "Principles of genetic circuit design". In: *Nature Methods* 11.5 (2014), pp. 508–520.

[71] Ye Chen et al. "Genetic circuit design automation for yeast". In: *Nature Microbiology* 5.11 (2020), pp. 1349–1360.

[72] Carlos Sequeiros et al. "Automated Design of Synthetic Gene Circuits in the Presence of Molecular Noise". In: *ACS Synthetic Biology* (2023).

[73] Noah Davidsohn et al. "Accurate predictions of genetic circuit behavior from part characterization and modular composition". In: *ACS Synthetic Biology* 4.6 (2015), pp. 673–681.

[74] Alec AK Nielsen et al. "Genetic circuit design automation". In: *Science* 352.6281 (2016), aac7341.

[75] Tobias Schladt et al. "Automated Design of Robust Genetic Circuits: Structural Variants and Parameter Uncertainty". In: *ACS Synthetic Biology* 10.12 (Nov. 2021), pp. 3316–3329. DOI: 10.1021/acssynbio.1c00193.

[76] Timothy S Jones et al. "Genetic circuit design automation with Cello 2.0". In: *Nature Protocols* 17.4 (2022), pp. 1097–1113.

[77] Christoph Zechner, Soudeep Deb, and Heinz Koeppl. "Marginal dynamics of stochastic biochemical networks in random environments". In: *2013 European Control Conference (ECC)*. IEEE. 2013, pp. 4269–4274.

[78] Leo Bronstein and Heinz Koeppl. "Marginal process framework: A model reduction tool for Markov jump processes". In: *Physical Review E* 97.6 (June 2018), p. 062147.

[79] Hiroshi Kunita. "Asymptotic behavior of the nonlinear filtering errors of Markov processes". In: *Journal of Multivariate Analysis* 1.4 (1971), pp. 365–393.

[80] Kaan Öcal, Guido Sanguinetti, and Ramon Grima. "Model reduction for the chemical master equation: An information-theoretic approach". In: *The Journal of Chemical Physics* 158.11 (2023).

[81] Christoph Zechner et al. "Molecular circuits for dynamic noise filtering". In: *Proceedings of the National Academy of Sciences* 113.17 (2016), pp. 4729–4734.

[82] Shlomo Shamai and Amos Lapidoth. "Bounds on the capacity of a spectrally constrained Poisson channel". In: *IEEE Transactions on Information Theory* 39.1 (1993), pp. 19–29.

[83] Lorenzo Duso and Christoph Zechner. "Path mutual information for a class of biochemical reaction networks". In: *2019 IEEE 58th Conference on Decision and Control (CDC)*. IEEE. 2019, pp. 6610–6615.

[84] Anne-Lena Moor and Christoph Zechner. "Dynamic information transfer in stochastic biochemical networks". In: *Physical Review Research* 5.1 (2023), p. 013032.

[85] Rami Atar and Tsachy Weissman. "Mutual information, relative entropy, and estimation in the Poisson channel". In: *IEEE Transactions on Information Theory* 58.3 (2012), pp. 1302–1318.

[86] D Snyder. "Filtering and detection for doubly stochastic Poisson processes". In: *IEEE Transactions on Information Theory* 18.1 (1972), pp. 91–102.

[87] Muruhan Rathinam and Mingkai Yu. "State and parameter estimation from exact partial state observation in stochastic reaction networks". In: *The Journal of Chemical Physics* 154.3 (2021), p. 034103.

[88] Zhou Fang, Ankit Gupta, and Mustafa Khammash. "Stochastic filtering for multiscale stochastic reaction networks based on hybrid approximations". In: *Journal of Computational Physics* 467 (2022), p. 111441.

[89] Lirong Huang et al. "Reconstructing dynamic molecular states from single-cell time series". In: *Journal of The Royal Society Interface* 13.122 (2016), p. 20160533.

[90] Dongning Guo, Shlomo Shamai, and Sergio Verdú. "Mutual information and conditional mean estimation in Poisson channels". In: *IEEE Transactions on Information Theory* 54.5 (2008), pp. 1837–1849. ISSN: 1557-9654.

[91] Rudolf Emil Kalman et al. "Contributions to the theory of optimal control". In: *Bol. Soc. Mat. Mexicana* 5.2 (1960), pp. 102–119.

[92] Patrick B Warren, Sorin Tănase-Nicola, and Pieter Rein ten Wolde. "Exact results for noise power spectra in linear biochemical reaction networks". In: *The Journal of Chemical Physics* 125.14 (2006), p. 144904.

[93] Ankit Gupta, Corentin Briat, and Mustafa Khammash. "A scalable computational framework for establishing long-term behavior of stochastic reaction networks". In: *PLoS Computational Biology* 10.6 (2014), e1003669.

[94] Wolfram Liebermeister, Ulrike Baur, and Edda Klipp. "Biochemical network models simplified by balanced truncation". In: *The FEBS Journal* 272.16 (2005), pp. 4034–4043.

[95] Ethan A Mastny, Eric L Haseltine, and James B Rawlings. "Two classes of quasi-steady-state model reductions for stochastic kinetics". In: *The Journal of Chemical Physics* 127.9 (2007), p. 094106.

[96] J Blom et al. "Markov-modulated infinite-server queues with general service times". In: *Queueing Systems* 76.4 (2014), pp. 403–424.

[97] W Feller. *An introduction to probability theory and its applications, Vol. 2, J. Wiley*. Sons New York, 1971.

[98] David Roxbee Cox and Valerie Isham. *Point processes*. Vol. 12. CRC Press, 1980.

[99] Johannes Falk et al. "Context in synthetic biology: Memory effects of environments with monomolecular reactions". In: *The Journal of Chemical Physics* 150.2 (2019), p. 024106.

[100] Arjun Raj et al. "Stochastic mRNA synthesis in mammalian cells". In: *PLoS Biology* 4.10 (2006), e309.

[101] Lennart Ljung. *System identification: Theory for the user*. 2nd. Prentice Hall PTR, Dec. 1999.

[102] Peter S Maybeck. *Stochastic models, estimation, and control*. Academic Press, 1982.

[103] Julius S Bendat and Allan G Piersol. *Random data: analysis and measurement procedures*. 4th. Wiley series in probability and statistics. Hoboken, N.J. : Wiley, 2010. ISBN: 9781118032428.

[104] ALAN G. Hawkes. "Spectra of some self-exciting and mutually exciting point processes". In: *Biometrika* 58.1 (1971), pp. 83–90. ISSN: 0006-3444.

[105] Andrew Daw and Jamol Pender. "Queues driven by Hawkes processes". In: *Stochastic Systems* 8.3 (2018), pp. 192–229.

[106] Leo Bronstein. "Approximation and model reduction for the stochastic kinetics of reaction networks". PhD thesis. Technical University of Darmstadt, 2020.

[107] Odd Aalen. "Nonparametric inference for a family of counting processes". In: *The Annals of Statistics* (1978), pp. 701–726.

[108] Odd O Aalen. "Mixing distributions on a Markov chain". In: *Scandinavian Journal of Statistics* (1987), pp. 281–289.

[109] Pierre Brémaud and Laurent Massoulié. "Stability of nonlinear Hawkes processes". In: *The Annals of Probability* (1996), pp. 1563–1588.

[110] Yuval Harel, Ron Meir, and Manfred Opper. "A tractable approximation to optimal point process filtering: Application to neural encoding". In: *Advances in Neural Information Processing Systems* 28 (2015).

[111] Christian Wildner and Heinz Koeppl. "Moment-based variational inference for Markov jump processes". In: *International Conference on Machine Learning*. PMLR. 2019, pp. 6766–6775.

[112] Jan Grandell. "On stochastic processes generated by a stochastic intensity function". In: *Scandinavian Actuarial Journal* 3-4 (1971), pp. 204–240.

[113] H L. Van Trees. *Detection, estimation, and modulation theory, part I: Detection, estimation, and linear modulation theory*. Wiley, New York, 1968.

[114] Jonathan W Pillow, Yashar Ahmadian, and Liam Paninski. "Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains". In: *Neural Computation* 23.1 (2011), pp. 1–45.

[115] Daryl J Daley and David Vere-Jones. *An introduction to the theory of point processes: volume II: General theory and structure*. Springer, 2008.

[116] Maurice Stevenson Bartlett. "The spectral analysis of point processes". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 25.2 (1963), pp. 264–281.

[117] Kris V. Parag and Glenn Vinnicombe. "Point process analysis of noise in early invertebrate vision". In: *PLoS Computational Biology* 13.10 (2017), e1005687. DOI: `10.1371/journal.pcbi.1005687`.

[118] Kris V. Parag and Oliver G. Pybus. "Optimal point process filtering and estimation of the coalescent process". In: *Journal of Theoretical Biology* 421 (2017), pp. 153–167. ISSN: 0022-5193. DOI: `10.1016/j.jtbi.2017.04.001`.

[119] Hiroshi Kunita. "Ergodic properties of nonlinear filtering processes". In: *Spatial stochastic processes*. Springer, 1991, pp. 233–256.

[120] Amarjit Budhiraja. "Asymptotic stability, ergodicity and other asymptotic properties of the nonlinear filter". In: *Annales de l'IHP Probabilités et Statistiques*. Vol. 39. 6. 2003, pp. 919–941.

[121] D. R. Cox. *Renewal Theory*. London: Methuen and Co, New York: Wiley, 1962.

[122] I. Gohberg and Naum Krupnik. *One-dimensional linear singular integral equations: I. Introduction*. Springer Science & Business Media, 1992. ISBN: 9783764325848.

[123] OLV Costa. "Stationary distributions for piecewise-deterministic Markov processes". In: *Journal of Applied Probability* 27.1 (1990), pp. 60–73.

[124] Nikolaos Limnios and G. Oprisan. *Semi-Markov processes and reliability*. Springer Science & Business Media, 2001. ISBN: 9780817641962.

[125] Christian Maes, Karel Netočný, and Bram Wynants. "Dynamical fluctuations for semi-Markov processes". In: *Journal of Physics A: Mathematical and Theoretical* 42.36 (2009), p. 365002. ISSN: 1751-8113. DOI: `10.1088/1751-8113/42/36/365002`.

[126] Sean P Meyn and Richard L Tweedie. "Stability of Markovian processes II: Continuous-time processes and sampled chains". In: *Advances in Applied Probability* 25.3 (1993), pp. 487–517.

[127] Oswaldo LV Costa and François Dufour. "Stability and ergodicity of piecewise-deterministic Markov processes". In: *SIAM Journal on Control and Optimization* 47.2 (2008), pp. 1053–1077.

[128] Alain Durmus, Arnaud Guillin, and Pierre Monmarché. "Piecewise-deterministic Markov processes and their invariant measures". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*. Vol. 57. 3. Institut Henri Poincaré. 2021, pp. 1442–1475.

[129] Zbigniew Palmowski, Simon Pojer, and Stefan Thonhauser. "Exact asymptotics of ruin probabilities with linear Hawkes arrivals". In: *arXiv preprint arXiv:2304.03075* (2023).

[130] Richard L Tweedie. "Sufficient conditions for ergodicity and recurrence of Markov chains on a general state space". In: *Stochastic Processes and their Applications* 3.4 (1975), pp. 385–403.

[131] Søren Asmussen. *Applied probability and queues*. Vol. 2. Springer, 2003.

[132] Sean P Meyn and Richard L Tweedie. *Markov chains and stochastic stability*. Springer Science & Business Media, 2012.

[133] D. Vere-Jones and Y. Ogata. "On the moments of a self-correcting process". In: *Journal of Applied Probability* 21.2 (1984), pp. 335–342. ISSN: 0021-9002. DOI: `10.2307/3213644`.

[134] Stewart N Ethier and Thomas G Kurtz. *Markov processes: characterization and convergence*. John Wiley & Sons, 2009.

[135] Ioannis Lestas, Glenn Vinnicombe, and Johan Paulsson. "Fundamental limits on the suppression of molecular fluctuations". In: *Nature* 467.7312 (2010), pp. 174–178.

[136] Kris V. Parag. "On signalling and estimation limits for molecular birth-processes". In: *Journal of Theoretical Biology* 480 (2019), pp. 262–273. ISSN: 0022-5193.

[137] Gašper Tkačik, Curtis G. Callan, and William Bialek. "Information flow and optimization in transcriptional regulation". In: *Proceedings of the National Academy of Sciences* 105.34 (2008), pp. 12265–12270. ISSN: 0027-8424.

[138] Ryan Suderman et al. "Fundamental trade-offs between information flow in single cells and cellular populations". In: *Proceedings of the National Academy of Sciences* 114.22 (2017), pp. 5755–5760. ISSN: 0027-8424.

[139] Andrea Crisanti, Andrea de Martino, and Jonathan Fiorentino. "Statistics of optimal information flow in ensembles of regulatory motifs". In: *Physical Review. E* 97.2-1 (2018), p. 022407. DOI: `10.1103/physreve.97.022407`.

[140] Syed Ahmed Pasha and Victor Solo. "Computing the trajectory mutual information between a point process and an analog stochastic process". In: *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2012, pp. 4603–4606.

[141] Sarah Anhala Cepeda-Humerez, Jakob Ruess, and Gašper Tkačik. "Estimating information in time-varying signals". In: *PLoS Computational Biology* 15.9 (2019), e1007290.

[142] Manuel Reinhardt, Gašper Tkačik, and Pieter Rein Ten Wolde. "Path weight sampling: Exact Monte Carlo computation of the mutual information between stochastic trajectories". In: *Physical Review X* 13.4 (2023), p. 041017.

[143] Filipe Tostevin and Pieter Rein ten Wolde. "Mutual Information between Input and Output Trajectories of Biochemical Networks". In: *Phys. Rev. Lett.* 102.21 (2009), p. 218101. DOI: `10.1103/PhysRevLett.102.218101`.

[144] Tuhin Subhra Roy et al. "Information transmission in a two-step cascade: interplay of activation and repression". In: *Theory in Biosciences* 140.3 (2021), pp. 295–306. ISSN: 1611-7530. DOI: `10.1007/s12064-021-00357-3`.

[145] M. Davis. "Capacity and cutoff rate for Poisson-type channels". In: *IEEE Transactions on Information Theory* 26.6 (1980), pp. 710–715. ISSN: 1557-9654.

[146] D. Snyder and C. Georghiades. "Design of coding and modulation for power-efficient use of a band-limited optical channel". In: *IEEE Transactions on Communications* 31.4 (1983), pp. 560–565.

[147] Shlomo Shamai. "On the capacity of a direct-detection photon channel with intertransition-constrained binary input". In: *IEEE Transactions on Information Theory* 37.6 (1991), pp. 1540–1550. ISSN: 1557-9654.

[148] Robert S. Liptser and Albert N. Shiryaev. *Statistics of random processes: II. Applications*. Vol. 2. Springer-Verlag, New York, 1977-78.

[149] A Irle. "Asymptotic optimality of general sequential probability ratio tests". In: *Scandinavian Journal of Statistics* (1990), pp. 321–332.

[150] Andrew EB Lim and J George Shanthikumar. "Relative entropy, exponential utility, and robust dynamic pricing". In: *Operations Research* 55.2 (2007), pp. 198–214.

[151] Paola Tardelli. "Minimal martingale measure: Pricing and hedging in a pure jump model under restricted information". In: *Nonlinear Analysis: Theory, Methods & Applications* 71.12 (2009), e1771–e1787.

[152] Yannis Pantazis and Markos A Katsoulakis. "A relative entropy rate method for path space sensitivity analysis of stationary complex stochastic dynamics". In: *The Journal of Chemical Physics* 138.5 (2013), p. 054115.

[153] S. Shamai. "Capacity of a pulse amplitude modulated direct detection photon channel". In: *IEE Proceedings I - Communications, Speech and Vision* 137.6 (1990), pp. 424–430. ISSN: 0956-3776. DOI: `10.1049/ip-i-2.1990.0056`.

[154] Tobias Jahnke and Wilhelm Huisinga. "Solving the chemical master equation for monomolecular reaction systems analytically". In: *Journal of Mathematical Biology* 54 (2007), pp. 1–26.

[155] Eric Renshaw. "The high-order autocovariance structure of the telegraph wave". In: *Journal of Applied Probability* 25.4 (1988), pp. 744–751.

[156] Richard S Bucy. "Global theory of the Riccati equation". In: *Journal of Computer and System Sciences* 1.4 (1967), pp. 349–361.

[157] David F Anderson. "A modified next reaction method for simulating chemical systems with time dependent propensities and delays". In: *The Journal of Chemical Physics* 127.21 (2007), p. 214107.

[158] P. A. W Lewis and G. S. Shedler. "Simulation of nonhomogeneous Poisson processes by thinning". In: *Naval Research Logistics Quarterly* 26.3 (1979), pp. 403–413. ISSN: 1931-9193. DOI: `10.1002/nav.3800260304`.

[159] Margaritis Voliotis et al. "Stochastic simulation of biomolecular networks in dynamic environments". In: *PLoS Computational Biology* 12.6 (2016), e1004923.

[160] Angelika Andrzejewska, Małgorzata Zawadzka, and Katarzyna Pachulska-Wieczorek. "On the way to understanding the interplay between the RNA structure and functions in cells: a genome-wide perspective". In: *International Journal of Molecular Sciences* 21.18 (2020), p. 6770.

[161] Hervé Isambert. "The jerky and knotty dynamics of RNA". In: *Methods* 49.2 (2009), pp. 189–196.

[162] Anthony M Mustoe et al. "Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing". In: *Cell* 173.1 (2018), pp. 181–195.

[163] Marilyn Kozak. "Regulation of translation via mRNA structure in prokaryotes and eukaryotes". In: *Gene* 361 (2005), pp. 13–37.

[164] Agnieszka Chełkowska-Pauszek et al. "The role of RNA secondary structure in regulation of gene expression in bacteria". In: *International Journal of Molecular Sciences* 22.15 (2021), p. 7845.

[165] Alexander Serganov and Evgeny Nudler. "A decade of riboswitches". In: *Cell* 152.1 (2013), pp. 17–24.

[166] Hashim M Al-Hashimi and Nils G Walter. "RNA dynamics: it is about time". In: *Current Opinion in Structural Biology* 18.3 (2008), pp. 321–329.

[167] Laura R Ganser et al. "The roles of structural dynamics in the cellular functions of RNAs". In: *Nature Reviews Molecular Cell Biology* 20.8 (2019), pp. 474–489.

[168] Wilhelm Huisinga, Sean Meyn, and Christof Schütte. "Phase transitions and metastability in Markovian and molecular systems". In: *The Annals of Applied Probability* 14.1 (2004), pp. 419–458.

[169] Jérôme Hénin et al. "Enhanced sampling methods for molecular dynamics simulations". In: *arXiv preprint arXiv:2202.04164* (2022).

[170] David Z Bushhouse et al. "How does RNA fold dynamically?" In: *Journal of Molecular Biology* (2022), p. 167665.

[171] Elizabeth A Dethoff et al. "Functional complexity and regulation through RNA dynamics". In: *Nature* 482.7385 (2012), pp. 322–330.

[172] Aline Umuhire Juru, Neeraj N Patwardhan, and Amanda E Hargrove. "Understanding the contributions of conformational changes, thermodynamics, and kinetics of RNA–small molecule interactions". In: *ACS Chemical Biology* 14.5 (2019), pp. 824–838.

[173] Philipp Thomas, Ramon Grima, and Arthur V Straube. "Rigorous elimination of fast stochastic variables from the linear noise approximation using projection operators". In: *Physical Review E* 86.4 (2012), p. 041110.

[174] Katy J Rubin et al. "Memory effects in biochemical networks as the natural counterpart of extrinsic noise". In: *Journal of Theoretical Biology* 357 (2014), pp. 245–267.

[175] Barbara Bravi and Peter Sollich. "Statistical physics approaches to subnetwork dynamics in biochemical systems". In: *Physical Biology* 14.4 (2017), p. 045010.

[176] Barbara Bravi, Katy J Rubin, and Peter Sollich. "Systematic model reduction captures the dynamics of extrinsic noise in biochemical subnetworks". In: *The Journal of Chemical Physics* 153.2 (2020).

[177] Michael Strasser, Fabian J Theis, and Carsten Marr. "Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression". In: *Biophysical Journal* 102.1 (2012), pp. 19–29.

[178] David Oakes. "The Markovian self-exciting process". In: *Journal of Applied Probability* 12.1 (1975), pp. 69–77. ISSN: 0021-9002. DOI: 10.2307/3212408.

[179] Casper HL Beentjes, Ruben Perez-Carrasco, and Ramon Grima. "Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics". In: *Physical Review E* 101.3 (2020), p. 032403.

[180] Emma M Keizer et al. "Extending the linear-noise approximation to biochemical systems influenced by intrinsic noise and slow lognormally distributed extrinsic noise". In: *Physical Review E* 99.5 (2019), p. 052417.

[181] A Grassia. "On a family of distributions with argument between 0 and 1 obtained by transformation of the Gamma and derived compound distributions". In: *Australian Journal of Statistics* 19.2 (1977), pp. 108–114.

[182] Jeremy E. Purvis and Galit Lahav. "Encoding and decoding cellular information through signaling dynamics". In: *Cell* 152.5 (2013), pp. 945–956.

[183] Dhana Friedrich et al. "Stochastic transcription in the p53-mediated response to DNA damage is modulated by burst frequency". In: *Molecular Systems Biology* 15.12 (2019).

[184] Jangir Selimkhanov et al. "Accurate information transmission through dynamic biochemical signaling networks". In: *Science* 346.6215 (2014), pp. 1370–1373. ISSN: 0036-8075. DOI: 10.1126/science.1254933.

[185] Andrew Mugler, Aleksandra M. Walczak, and Chris H. Wiggins. "Spectral solutions to stochastic models of gene expression with bursts and regulation". In: *Phys. Rev. E* 80.4 (2009), p. 041921. DOI: 10.1103/PhysRevE.80.041921.

[186] Y. Nakahira et al. "Fundamental limits and achievable performance in biomolecular control". In: *2018 Annual American Control Conference (ACC)*. 2018, pp. 2707–2714. DOI: 10.23919/ACC.2018.8430933.

[187] Alok Maity and Roy Wollman. "Information transmission from NFkB signaling dynamics to gene expression". In: *PLoS Computational Biology* 16.8 (2020), e1008011. DOI: 10.1371/journal.pcbi.1008011.

[188] Tomasz Jetka et al. "Information-theoretic analysis of multivariate single-cell signaling responses". In: *PLoS Computational Biology* 15.7 (2019), e1007132. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1007132.

[189] Ryan Suderman and Eric J. Deeds. "Intrinsic limits of information transmission in biochemical signalling motifs". In: *Interface Focus* 8.6 (2018), p. 20180039. DOI: 10.1098/rsfs.2018.0039.

[190] Huahai Qiu, Bengong Zhang, and Tianshou Zhou. "Influence of Complex Promoter Structure on Gene Expression". In: *Journal of Systems Science and Complexity* 32.2 (2019), pp. 600–614. ISSN: 1559-7067. DOI: 10.1007/s11424-018-7224-7.

[191] O. Macchi and B. Picinbono. "Estimation and detection of weak optical signals". In: *IEEE Transactions on Information Theory* 18.5 (1972), pp. 562–573. ISSN: 1557-9654. DOI: `10.1109/TIT.1972.1054895`.

[192] Lawrence C. Evans. *Partial differential equations*. American Mathematical Society, 1998.

[193] Gašper Tkačik and Aleksandra M. Walczak. "Information transmission in genetic regulatory networks: a review". In: *Journal of Physics: Condensed Matter* 23.15 (2011), p. 153102.

[194] Alfonso Pérez-Escudero, Marta Rivera-Alba, and Gonzalo G. de Polavieja. "Structure of deviations from optimality in biological systems". In: *Proceedings of the National Academy of Sciences* 106.48 (2009), pp. 20544–20549. ISSN: 0027-8424.

[195] Patrick J Laub, Thomas Taimre, and Philip K Pollett. "Hawkes processes". In: *arXiv preprint arXiv:1507.02822* (2015).

[196] Manfred Opper. "An estimator for the relative entropy rate of path measures for stochastic differential equations". In: *Journal of Computational Physics* 330 (2017), pp. 127–133.

[197] Bogdan Ćmiel et al. "Generalised score distribution: underdispersed continuation of the beta-binomial distribution". In: *Statistical Papers* (2023), pp. 1–33.

[198] Anatol Kuczura. "The interrupted Poisson process as an overflow process". In: *Bell System Technical Journal* 52.3 (1973), pp. 437–448. DOI: `10.1002/j.1538-7305.1973.tb01971.x`.

[199] Rosa M Fernández-Alcalá et al. "Recursive linear estimation for doubly stochastic Poisson processes." In: *World Congress on Engineering*. Citeseer. 2007, pp. 894–897.

[200] Rosa M Fernández-Alcalá, Jesús Navarro-Moreno, and Juan Carlos Ruiz-Molina. "Statistical inference for doubly stochastic multichannel Poisson processes: A PCA approach". In: *Computational Statistics & Data Analysis* 53.12 (2009), pp. 4322–4331.

[201] Jörn Sass, Dorothee Westphal, and Ralf Wunderlich. "Diffusion approximations for randomly arriving expert opinions in a financial market with Gaussian drift". In: *Journal of Applied Probability* 58.1 (2021), pp. 197–216.

[202] Elena Sofia D'Ambrosio et al. "Filtered finite state projection method for the analysis and estimation of stochastic biochemical reaction networks". In: *bioRxiv* (2022), pp. 2022–10.

[203] Swati Dhobaley and Prashant Bhopale. "State estimation of yeast galactose pathway using extended Kalman filter". In: *2014 International Conference on Control, Instrumentation, Communication and Computational Technologies (ICCICCT)*. IEEE. 2014, pp. 1271–1274.

[204] Vilda Purutçuoğlu and Ernst Wit. "Exact and approximate stochastic simulations of the MAPK pathway and comparisons of simulations' results". In: *Journal of Integrative Bioinformatics* 3.2 (2006), pp. 199–211. DOI: `10.1515/jib-2006-38`.