IN SEARCH OF A COHERENT THEORETICAL FRAMEWORK

FOR STOCHASTIC GENE REGULATION

By

John J. Vastola

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Physics

May 14th, 2021

Nashville, Tennessee

Approved:

William R. Holmes, Ph.D.

Gregor Neuert, Ph.D.

Kalman Varga, Ph.D.

John P. Wikswo, Ph.D.

To the seekers and speakers of truth,
and all fires yet to be kindled

Acknowledgments

TABLE OF CONTENTS

6 Analytic solution of chemical master equations involving gene switching. I: Representation theory and diagrammatic approach to exact solution 140

List of Tables

List of Figures

Chapter 1

Invitation

Is there an equation that describes how tall an elephant grows? Sure; it might look something like

$$\dot{x}_i = \sum_{j,k} \frac{\prod_j c_{ijk} x_j^{n_{jk}}}{\sum_j d_{ijk} x_j^{n_{jk}}} - \gamma_i x_i \ . \tag{1.1}$$

Many features of biological systems—from obvious characteristics like height, to the complicated failure modes associated with disease—are ultimately due to interactions between genes. Different genes influence each other's expression (the number of a gene's corresponding RNA or proteins that exist in a single cell at some time), for example by interfering with or enhancing the transcription of the RNA associated with that gene. More colloquially, different genes turn each other 'on' or 'off'.

Equivalently, we can say that something like elephant height is ultimately determined by the dynamics of *gene networks*. We can consider two genes 'connected' if they influence each other's gene expression in some way, and write down mathematical models that instantiate this intuition in a quantitatively concrete form.

A classic example is the bistable switch, which describes two genes which inhibit each other (reduce gene expression) and activate themselves (increase gene expression). A plausible set of equations for this reads

$$\dot{x} = k_A \frac{x^n}{c^n + x^n} - k_I \frac{y^n}{c^n + y^n} - \gamma_x x$$
$$\dot{y} = k_A \frac{y^n}{c^n + y^n} - k_I \frac{x^n}{c^n + x^n} - \gamma_y y \ , \tag{1.2}$$

where x denotes the expression level (protein concentration, say) of gene X, y denotes the expression level of gene Y, k_A is the self-activation strength, k_I is the mutual inhibition strength, c is some characteristic concentration threshold that separates low concentrations from high concentrations, n quantifies how sharp the distinction between high and low concentrations is, and γ_x and γ_y are degradation rates (these terms must be present since, even without gene-gene interactions, proteins degrade after some time).

A model like this can be made more complicated or generic, perhaps by including more terms or by making weaker assumptions about parameters—after all, how much gene X inhibits gene Y need not be equal to how much gene Y inhibits gene X, or even take the same functional form. But the point is that we can write these models down as a first pass at understanding what's going on. We can analyze them, make predictions, compare them with experiment, refine them, and so on, in iterative fashion, until we have a satisfactory description of the dynamics of some gene network of interest.

The kind of models described by Eq. 1.1 and 1.2 are ordinary differential equation (ODE) models. They describe *deterministic* systems; two identically prepared initial states will always evolve into identical final states after some particular amount of time. But gene expression is known to be *stochastic*. In practice, two states that seem to be 'identically' prepared may evolve into very different states, because how the numbers of molecules inside cells change over time is partly random. A more technically useful way of stating this is: we cannot predict how many RNA or proteins a cell will have, but we *can* predict the *probability* it has this many or that many at some specific time in the future.

This randomness arises, in some sense, because of subtleties associated with defining the 'state' of a cell. We count cells as being in the same state if the numbers of certain reference molecules are the same (e.g. specific RNA or proteins of interest), but we care less about the motion and number of (for example) all water molecules inside a cell. Because the jostling about and interactions between every member of the complex soup of molecules inside the crowded interior of cells affects how the molecules we care about diffuse and interact, we get randomness mostly in the sense of statistical mechanics—due to incomplete information.

And this randomness can have real consequences. The ODE models above generically have different attractors / stable states. Once a cell reaches a state where the competition between forces that increase certain concentrations balance with all of those forces that decrease those concentrations, its concentrations stop changing entirely. If a perturbation pushes a cell slightly away from this state, it will return to it relatively quickly. But in 'noisy' models of gene regulation, just by chance a cell can be pushed into a completely different attractor entirely, and make what is called a noise-induced transition. This simple idea is thought to be used by eukaryotic developmental programs to correct cell fate errors [1], and by tumor cells to avoid drug treatments by randomly switching between sensitive and resistant states [2], among other things.

But there is also a completely different reason to care about noise. All things in the physical world are made of elementary particles (or more precisely, quantum fields), which can only interact in a small number of ways, and whose behavior over time is completely governed by the Schrödinger equation. In some sense, the project of understanding all physical phenomena can in principle be reduced to understanding the Schrödinger equation in various circumstances.

'In principle' may be the operative phrase, here. *In practice*, it is not necessary to solve the Schrödinger equation for a collection of 10^{23} interacting atoms in order to describe the trajectory of a ball in a weak gravitational field. Nevertheless, just knowing that such a reduction is in principle possible has lead to important physical insights about everything from the stability of matter [3, 4] to why systems tend toward thermal equilibrium [5]. More broadly, having a well-understood and well-defined theoretical framework sitting at the bottom of our hierarchy of models of physics allows us to reason quite clearly think about what is and is not within the realm of physical possibility.

Is there a grand unified theory of biology? Despite our best hopes, probably not. But there *is* something like the Schrödinger equation for gene regulation: an equation sitting at the bottom of our hierarchy of models, that includes every other model of gene expression

as a special case. I am talking about the so-called *chemical master equation*.

The chemical master equation (CME) describes how the probability of observing different numbers of molecules in some chemical system (e.g. a single cell) changes over time. It is *discrete*, because it treats molecule numbers as nonnegative integers rather than continuously varying concentrations, and *stochastic*.

A CME is determined by four things:

1. a list of chemical species

2. a list of chemical reactions

3. propensity functions / rate parameters for each chemical reaction

4. initial conditions.

If we think about CME dynamics like a play, to fully specify a model we need to specify its actors, the possible actions, how frequently each action tends to happen, and how the story begins.

It is worth making a point of this: *all* known models of gene regulation can be imagined to be approximate descriptions of some underlying CME model (see Chapter 2.4). Stochastic differential equation models can be derived via the chemical Langevin approximation or by taking the large volume limit (both of which are discussed in Ch. 5). Deterministic ordinary differential equation models can be derived from taking the low noise, local, or thermodynamic limits of *those* models. Boolean networks (deterministic or stochastic) can be derived by coarse-graining molecule number or concentration states. Markov models with finitely many states (e.g. each state is a different cell type) can also be derived by grouping different states together. All roads lead to the CME.

Just as our best physical theories identify which kinds of interactions are 'fundamental' (e.g. renormalizable interactions in quantum field theory, or conservative forces in classical physics) and which can only come about as an approximate description of more complicated dynamics, so too does the CME offer a strong sense of which kinds of interactions are 'fundamental' and which are not. Crudely, monomolecular (involving at most one input and one output molecule) and bimolecular (involving at most two input and two output molecules) reactions with elementary propensity functions (see Chapter 2.2) are fundamental, while other reactions must only be approximate descriptions of some more complex CME.

For an equation that offers a foundational theoretical framework for chemical kinetics and gene regulation, the CME—and particularly many of its mathematical properties—is surprisingly poorly studied. Most existing work on the CME focuses on approximating it so that it reduces to a more tractable model, or simulating it on a computer using variants of Gillespie's algorithm [6, 7, 8].

If we would like the CME to play the same role in gene regulation that the Schrödinger equation plays in physics, we need to understand it much better. This means that we should:

1. know how to **think** about it,

2. know how to **simplify** it, and

3. know how to **solve** it.

The aim of this thesis is to explore and offer partial progress toward the above three aims. Particular emphasis will be placed on exploring theoretical properties of the CME, and developing mathematical tools for working with it. It will turn out that the analogy between the Schrödinger equation and the CME is quite mathematically deep, and that many theoretical tools originally developed for studying the Schrödinger equation (e.g. path integral representations and ladder operators) are useful in this biological context too.

The structure of this thesis is as follows. In Chapter 2, we will introduce the CME as a theoretical framework uniquely appropriate for studying fundamental problems in gene regulation. This partly addresses how to *think* about the CME, and how to *simplify* it to obtain several common types of gene regulation models. In Chapter 3, we will introduce basic tools for working with the CME, including stochastic simulation methods and a number of analytic techniques. This covers the broad strokes of what was previously known about *solving* the CME.

In the remaining chapters, we introduce new mathematical techniques inspired by quantum mechanics for *solving* the CME, and obtain the most general solutions yet known. In Chapter 4, we use the Doi-Peliti path integral method to solve and study monomolecular reaction networks, a class of systems that are particularly suited to analytic approaches, and that we will encounter throughout this thesis. Using the same tools, we also formally treat systems whose reaction list consists of arbitrary combinations of zero and first order reactions, and obtain perhaps the most general known solution of the CME. In Chapter 5, we introduce a new state space path integral representation of the CME (distinct from but complementary to the Doi-Peliti one) and apply it to the problem of justifying continuous stochastic approximations of discrete stochastic dynamics. In Chapter 6, we show how a CME model of RNA transcription and splicing (an example of a monomolecular process) coupled to a switching gene can be solved using ladder operators and Feynman-like diagrams.

In Chapter 7, we show how the same CME models can be efficiently solved numerically using the path integral introduced in Chapter 5, and apply these efficient implementations to parameter inference on simulated data. The numerical methods we develop here far outperform competing methods like finite state projection, which has important implications for fitting stochastic models to single cell data. Namely, the large speedup in principle allows one to fit significantly more realistic models of transcription and splicing, and to interrogate at much higher resolution than before the associated kinetics. Finally, in Chapter 8, we conclude and reflect on various theoretical challenges related to the CME.

Bibliography

[1] William R. Holmes, Nabora Soledad Reyes de Mochel, Qixuan Wang, Huijing Du, Tao Peng, Michael Chiang, Olivier Cinquin, Ken Cho, and Qing Nie. Gene expression noise enhances robust organization of the early mammalian blastocyst. PLOS Computational Biology, 13(1):1–23, 01 2017.

[2] Mohit Kumar Jolly, Prakash Kulkarni, Keith Weninger, John Orban, and Herbert Levine. Phenotypic plasticity, bet-hedging, and androgen independence in prostate cancer: Role of non-genetic heterogeneity. Frontiers in Oncology, 8:50, 2018.

[3] Freeman J. Dyson and A. Lenard. Stability of matter. i. Journal of Mathematical Physics, 8(3):423–434, 1967.

[4] A. Lenard and Freeman J. Dyson. Stability of matter. ii. Journal of Mathematical Physics, 9(5):698–711, 1968.

[5] Joshua M Deutsch. Eigenstate thermalization hypothesis. Reports on Progress in Physics, 81(8):082001, jul 2018.

[6] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4):403 – 434, 1976.

[7] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81(25):2340–2361, 1977.

[8] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry, 58(1):35–55, 2007.

Chapter 2

Philosophy. Or: the chemical master equation in principle

2.1 On being the right scale

Our goal is to figure out some sort of foundation for gene networks. Ideally, we would like to use this foundation to understand—in a quantitative fashion—important consequences of gene regulation, like the ability for one kind of cell to change into another kind of cell, or the dynamics of transcription and splicing. How should we proceed?

All things are made of elementary particles, so one way to proceed would be to try to develop a theory of gene networks based on quantum field theory. While this would certainly be a noble goal, it is also a Sisyphean task par excellence; understanding a proton from first principles is hard enough. Thinking in terms of nonrelativistic quantum mechanics (electrons or atoms, perhaps) is not much better, given the well-known difficulties with understanding *ab initio* quantum chemistry. It does not take that many atoms to exceed the current capabilities of the world's best computers.

If not so fine a scale, at what scale should we work? And what are the most salient aspects of gene networks? What aspects are so crucially important, that it is absolutely imperative that we include them in our models? Another way to frame this question is: what is the state space of the models we are interested in? Do we care about the state of every electron inside a cell? Do we care about where molecules are? Must we keep track of every molecule, or only some of them? Do we care only how many molecules there are *on average*, or the exact number?

To some extent our decision is arbitrary, as one can always cook up examples where this or that feature of chemical or biological systems happens to be important. But we can at least base our decision on *something*. Consider the following four principles: *locality*, *experimental accessibility*, *experimental compatibility*, and *minimizing arbitrariness*.

Locality means that distant things, to good approximation, do not affect our system. For most purposes, it is sufficient to imagine only one cell and ignore the rest of the universe, or in practice to imagine just a few molecules in a single cell that interact with one another, and ignore the remainder. One can imagine a world in which every molecule inside a cell affected every other molecule, so it must be appreciated that this is a profound and important property of nature.

Experimental accessibility means that our framework should primarily concern itself with things that are measurable, or at least measurable in principle. We can count how many RNA are inside single cells; we cannot count the number of electrons associated with each RNA molecule. Even if we could, this information is almost certainly irrelevant for gene-network-scale dynamics. *Experimental compatibility* means that our framework includes

or can explain details of gene regulation known to be important. In particular, single cell experiments have established that discrete molecule number effects, along with the associated random fluctuations in molecule number, are important. For example, there are only about 10 tetrameric copies of *lac* repressor proteins per *E. coli* cell, so fluctuations in this number can have an outsized effect on lactose utilization [1, 2, 3]. A cell may make a different decision if there are 5 molecules instead of 10, which means accounting for this fact is *qualitatively* important.

Finally, one frustrating aspect of mathematical modeling is the proliferation of arbitrary choices. If we assumed *this* instead of *that*, how might our predictions change? In gene regulation, where microscopic models must usually be jettisoned in favor of phenomenological ones that there is some hope of connecting with data, this is particularly annoying; gene-gene interactions are often assumed to take on certain ad-hoc forms (especially Hill functions), and models are often made stochastic by tacking white noise terms onto ordinary differential equation models rather than following a more principled prescription. In quantum field theory, there is a sense in which the constraints of renormalizability and gauge symmetries fully constrain fundamental theories; we would like something analogous to be true for our gene network models.

Given these four principles, what is the right scale for a fundamental theory of gene regulation? First: by gene regulation, we really mean interactions between molecules that conspire to effect cell-level changes within a single cell. By the principle of locality, we can restrict ourselves to thinking about a single cell, or a collection of interacting cells. When important, environmental perturbations can be taken into account by their effect on parameters. The principle of experimental accessibility means that we can afford to ignore extremely fine differences between single cells, like the detailed arrangement of each atom or electron. For the most part, we can also afford to ignore spatial information (since even if it is modeled, it is hard to sample enough cells and locations within cells to obtain reliable spatial distributions), although compartmental information (e.g. whether a transcript is in the nucleus or the cytoplasm) may be important.

Because one can resolve individual molecules in single cell experiments (e.g. whether a cell has 10 or 11 RNA of some kind), and because even small differences in molecule number can matter, the principle of experimental compatibility forces our spatial and temporal scale to be fine enough to track the behavior of individual molecules. Most importantly, we should track the *number* of each molecule we are interested in, and how those numbers change over time. Usually, only some molecules are important for the logic of a gene regulatory circuit, and other molecules (or aggregates of molecules, like transcriptional machinery) can be treated indirectly via their effects on the number or parameters of the observable ones.

Thus far, we have decided to track the number of certain molecules (that are relevant to gene regulatory logic, like RNA and proteins) within single cells. Our framework should treat molecule numbers as discrete (nonnegative integers rather than continuous concentrations) and stochastic, with the randomness due to a mixture of quantum and thermal effects not explicitly modeled. The only remaining question is how to treat the molecule-molecule interactions that cause molecule numbers to change. By the principle of minimizing arbi-

trariness, we should avoid introducing interaction functions with many free parameters or an undetermined functional form.

Here, we can exploit a gift from Gillespie's seminal work on the microphysical basis of chemical kinetics [4, 5]. There are three relevant facts: first, real molecules change via monomolecular or bimolecular chemical reactions, with other kinds of reactions (like trimolecular) being only approximate descriptions; second, the interaction probabilities in each of these cases are completely determined by microphysical arguments; third, stochastic systems involving only these kinds of reactions are to good approximation Markovian/'memoryless'.

This allows us to establish a foundation for gene regulation based on the following model.

1. A system is assumed to have some number N of distinct chemical species (e.g. RNA, proteins). At any one time, the state of the system is completely specified by a tuple of nonnegative integers $\mathbf{x} = (x_1, ..., x_N) \in \mathbb{N}^N$.

2. The state of the system can change with time, and it is assumed that there are only some number M of ways that it can change. In particular, it is assumed that there are M possible chemical reactions that can occur, each of which are either monomolecular (involving at most one input and output) or bimolecular (involving at most two inputs and two outputs). In a small window of time, there is some probability that any one of these reactions occurs, and the functional form of these probabilities is completely determined by microphysical considerations.

3. The future state of the system depends only on its current state.

These restrictions precisely specify the theoretical framework associated with the *chemical master equation* (CME), our topic of study for the remainder of this thesis. This is our answer to the question: 'what is the right scale for the theoretical foundation of gene regulation?'

We should note that, in practice, these assumptions can be relaxed somewhat. The CME can deal with 'effective' interactions between molecules that do not take one of the fundamental forms identified by Gillespie, although it should be said that one can usually identify a 'fundamental' system that such effective descriptions approximate in some way. In similar spirit, the CME can conceive of systems that involve things like continuous species variables and non-Markovian dynamics—as approximations valid in some possibly singular limit. As we noted in the previous chapter, one attractive feature of the CME is that is subsumes all known gene regulation modeling frameworks as special cases, including ordinary differential equations, Boolean networks, Markov models, and so on. One must merely identify the appropriate limit for recovering a given approximate description.

2.2 Whence the chemical master equation?

Let us more precisely describe the chemical master equation framework we motivated in the previous section. In particular, we would like to 'derive' the equation itself given our assumptions.

Again, assume we have a system with N chemical species and M possible chemical reactions, whose state is specified by a tuple $\mathbf{x} = (x_1, ..., x_N) \in \mathbb{N}^N$. Let $\boldsymbol{\nu}_j$ denote the amount by which the jth reaction changes the state vector $\mathbf{x}$, i.e. reaction j firing causes the system to change according to $\mathbf{x} \to \mathbf{x} + \boldsymbol{\nu}_j$. Denote the probability per unit time of reaction j firing in the small time window $(t, t + \Delta t)$ by $a_j(\mathbf{x})$, where $\mathbf{x}$ is the state of the system at time t.

The central object of our theory is the probability distribution on state space, $P(\mathbf{x}, t)$, which indicates the probability that the system has state $\mathbf{x} \in \mathbb{N}^N$ at time $t \geq t_0$ (where t_0 is some initial time which can often arbitrarily be chosen to be zero). In order to have a well-defined theory, we must be able to quantitatively describe how $P(\mathbf{x}, t)$ changes in time. That is, given how the distribution of observing this or that many molecules looks *right now*, how will the distribution look *sometime in the future*?

Our evolution equation is essentially due to a combination of the conservation of probability and the Markovian property. For an infinitesimal amount of time Δt, we have that

$$P(\mathbf{x}, t + \Delta t) = P(\mathbf{x}, t) + \text{prob(arrived at } \mathbf{x} \text{ within } \Delta t) - \text{prob(left } \mathbf{x} \text{ within } \Delta t) \ . \quad (2.1)$$

Because the system is Markovian, the 'enter' and 'exit' probabilities on the right-hand side only depend on the state of the system at time t. If the state of the system was different from $\mathbf{x}$ at time t, the only way for it to 'enter' the state $\mathbf{x}$ within Δt is for one of the M chemical reactions to have fired (we are assuming Δt is sufficiently small that at *most* one reaction could have fired). Hence,

$$
\begin{aligned}
\text{prob(arrived at } \mathbf{x} \text{ within } \Delta t) &= \sum_{j=1}^{M} \text{prob(reaction } \mathbf{j} \text{ fired within } \Delta t) \\
&= \sum_{j=1}^{M} a_j(\mathbf{x} - \boldsymbol{\nu}_j)\Delta t \ P(\mathbf{x} - \boldsymbol{\nu}_j, t)
\end{aligned}
\quad (2.2)
$$

where we note that, *had* reaction j fired, the previous state of the system must have been exactly $\mathbf{x} - \boldsymbol{\nu}_j$. Meanwhile, the probability that the system *left* the state $\mathbf{x}$ within Δt means that it began in $\mathbf{x}$, and then one of the M reactions fired. This means that

$$\text{prob(left } \mathbf{x} \text{ within } \Delta t) = \sum_{j=1}^{M} a_j(\mathbf{x})\Delta t \ P(\mathbf{x}, t) \ . \quad (2.3)$$

Putting these together, we have

$$P(\mathbf{x}, t + \Delta t) = P(\mathbf{x}, t) + \Delta t \sum_{j=1}^{M} a_j(\mathbf{x} - \boldsymbol{\nu}_j)P(\mathbf{x} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{x})P(\mathbf{x}, t) \ . \quad (2.4)$$

9

Rearranging and taking the small time interval Δt to zero, we obtain the differential-difference equation

$$\frac{\partial P(\mathbf{x}, t)}{\partial t} = \sum_{j=1}^{M} a_j(\mathbf{x} - \boldsymbol{\nu}_j) P(\mathbf{x} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{x}) P(\mathbf{x}, t) \ . \tag{2.5}$$

This is the equation we mean when we talk about the chemical master equation. As the above derivation suggests, it says that the probability of being in state $\mathbf{x}$ changes both by the system *entering* $\mathbf{x}$ (because the system was in a different state, and one of the M chemical reactions occurred), and by the system *leaving* $\mathbf{x}$ (because the system was in state $\mathbf{x}$, and one of the M chemical reactions occurred).

As promised, microphysical arguments factor into determining the functional form of the functions $a_j(\mathbf{x})$, which we will henceforth call *propensity functions* [4, 5]. The following list exhausts every possible 'fundamental' type of reaction, and indicates the propensity function for each:

Reaction	Label	Propensity function
$\varnothing \xrightarrow{c} X$	zeroth order production	c
$X \xrightarrow{c} \varnothing$	first order degradation	$c\,x$
$X \xrightarrow{c} Y$	first order conversion	$c\,x$
$\varnothing \xrightarrow{c} X + X$	zeroth order identical pair production	c
$\varnothing \xrightarrow{c} X + Y$	zeroth order pair production	c
$X \xrightarrow{c} X + X$	first order auto-catalysis	$c\,x$
$X \xrightarrow{c} X + Y$	first order translation	$c\,x$
$X \xrightarrow{c} Y + Z$	first order splitting	$c\,x$
$X + X \xrightarrow{c} \varnothing$	second order self-annihilation	$c\binom{x}{2} = c\,\frac{x(x-1)}{2}$
$X + X \xrightarrow{c} X$	second order coagulation	$c\binom{x}{2}$
$X + X \xrightarrow{c} Y$	second order self-complexing	$c\binom{x}{2}$
$X + X \xrightarrow{c} X + Y$	second order self-catalyzed conversion	$c\binom{x}{2}$
$X + X \xrightarrow{c} Y + Z$	second order self-collision	$c\binom{x}{2}$
$X + Y \xrightarrow{c} \varnothing$	second order annihilation	$c\,xy$
$X + Y \xrightarrow{c} X$	second order catalyzed degradation	$c\,xy$
$X + Y \xrightarrow{c} Z$	second order complexing	$c\,xy$
$X + Y \xrightarrow{c} X + Z$	second order catalytic conversion	$c\,xy$
$X + Y \xrightarrow{c} X + X$	second order conformity	$c\,xy$
$X + Y \xrightarrow{c} Z + W$	second order collision	$c\,xy$

Table 2.1: A list of all 19 'fundamental' reaction types and their associated propensity functions. Each propensity function only depends on one free (real-valued) parameter $c > 0$, which quantifies how often that reaction happens. The first three are monomolecular, while the rest are bimolecular.

The important thing to notice is that the functional form of each propensity function only depends on the input molecule numbers, and that they are combinatorial rather than something more complicated. For example, the degradation reaction's propensity function is just counting the number of different molecules that could degrade (x, if that is the current number of X molecules). Similarly, the self-annihilation propensity function is counting the number of ways two X molecules could be picked to interact.

To summarize, a fundamental CME model is fully specified by:

- a list of $N \geq 1$ chemical species

- a list of $M \geq 1$ chemical reactions (drawn from Table 2.1), each of which is associated with a positive real rate parameter that quantifies how often that reaction happens

- an initial distribution $P_0(\mathbf{x})$ (i.e. the initial condition for Eq. 2.5).

2.3 Generic properties of the chemical master equation

What can we say in general about Eq. 2.5, divorced from the particulars of any specific system? First, it enforces probability conservation by construction, because

$$
\begin{aligned}
\frac{\partial}{\partial t} \sum_{\mathbf{x}} P(\mathbf{x}, t) &= \sum_{\mathbf{x}} \frac{\partial P(\mathbf{x}, t)}{\partial t} \\
&= \sum_{\mathbf{x}} \sum_{j=1}^{M} a_j(\mathbf{x} - \boldsymbol{\nu}_j) P(\mathbf{x} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{x}) P(\mathbf{x}, t) \\
&= \sum_{\mathbf{x}} \sum_{j=1}^{M} a_j(\mathbf{x}) P(\mathbf{x}, t) - a_j(\mathbf{x}) P(\mathbf{x}, t) \\
&= 0
\end{aligned}
\tag{2.6}
$$

where we have obtained the second to last line by reindexing the sum. Hence, if $P(\mathbf{x}, t)$ is normalized at one time, it is normalized for all time. Second, Eq. 2.5 is intimately related to the standard (deterministic, continuous) mass action prescription for chemical kinetics. While the CME captures the stochastic / single cell behavior of systems, the

11

average trajectory evolves according to

$$
\begin{aligned}
\frac{\partial \langle \mathbf{x} \rangle}{\partial t} &= \sum_{\mathbf{x}} \mathbf{x} \frac{\partial P(\mathbf{x}, t)}{\partial t} \\
&= \sum_{\mathbf{x}} \mathbf{x} \sum_{j=1}^{M} a_j(\mathbf{x} - \boldsymbol{\nu}_j) P(\mathbf{x} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{x}) P(\mathbf{x}, t) \\
&= \sum_{j=1}^{M} \sum_{\mathbf{x}} (\mathbf{x} + \boldsymbol{\nu}_j) a_j(\mathbf{x}) P(\mathbf{x}, t) - \mathbf{x}\, a_j(\mathbf{x}) P(\mathbf{x}, t) \\
&= \sum_{j=1}^{M} \boldsymbol{\nu}_j \langle a_j(\mathbf{x}) \rangle
\end{aligned}
\tag{2.7}
$$

which precisely constitute the reaction rate equations for this system. In this way, it is clear that mass action chemical kinetics is subsumed by the CME, and corresponds to either population-level averages or some kind of deterministic limit (e.g. the large volume limit Gillespie describes in [6]).

Third, essentially all CME models of biological interest have the property that

$$
P_{ss}(\mathbf{x}) := \lim_{t \to \infty} P(\mathbf{x}, t)
\tag{2.8}
$$

is well-defined (i.e. the limit exists and is a normalized probability distribution) and does not depend on the initial condition $P_0(\mathbf{x})$. This limiting distribution is called the *steady state distribution*, and in most cases this is the object amenable for comparison with experiment. This is for two reasons: experimental time scales are much longer than typical biological time scales (and many processes have been going on long before the experiment, and so have had ample time to equilibrate), so the infinite time limit is a good approximation; and a lack of sensitivity to initial conditions makes it a more robust model prediction. Mathematically, the existence of this limit has to do with whether there are reactions that counterbalance molecule number increases with molecule number decreases, and whether the system can be separated into two or more non-interacting subsystems. Also, it may not exist if there are persistent oscillations, e.g. in the case of models of the cell cycle (although in this case one could define an analogous object by taking the long time limit of the average over an oscillation period).

2.4 How the chemical master equation relates to other models

We have mentioned several times how the chemical master equation includes all other known models of gene regulation as special cases; it is worth specifying in a bit more detail exactly what we mean. In this section, we will closely examine a few representative relationships. But in the interest of brevity—the mathematics of these correspondences could easily fill an entire journal article—others we merely summarize.

It is most helpful to think about how the CME relates to ordinary differential equation (ODE) models, which are still the models most commonly used in practice. Schematically, these models often look something like:

$$\dot{x} = k \frac{x^n}{c^n + x^n} - \gamma x \ . \tag{2.9}$$

That is, the concentration $x(t)$ of some molecule changes deterministically in time due to a combination of gene-gene interactions (usually represented by one or more Hill functions) and a degradation term. Here, X 'activates' itself: k is the self-activation strength, c is the characteristic concentration threshold dividing 'high' levels of x from 'low' levels of x, and $n \in \{1, 2, 3, ...\}$ is the Hill coefficient. γ is the degradation rate.

We encounter two issues in trying to relate Eq. 2.9 to Eq. 2.5, the CME. First, while the degradation term represents something from our list of 'fundamental' reactions (Table 2.1), the Hill function term apparently does not. Second, there is no sign of probability in this model, and $x(t)$ varies continuously. Let us deal with these problems one at a time.

The Hill function is a phenomenological device usually intended to approximate a more detailed interaction between molecules. In this case, the meaning of 'X activates itself' might be something like: 'X proteins reversibly bind to the promoter of the X gene, increasing the transcription rate of X'. Perhaps this is a *cooperative* effect; up to n X proteins can bind to the X promoter, and each additional X protein modifies the transcription rate.

We can instantiate this idea by imagining that the X gene has $n + 1$ different states G_0, G_1, ..., G_n, each labeled according to how many X proteins are bound to the promoter. We have the list of chemical reactions

$$G_0 + X \underset{b_r^1}{\overset{b_f^1}{\rightleftharpoons}} G_1 \ , \ \ G_1 + X \underset{b_r^2}{\overset{b_f^2}{\rightleftharpoons}} G_2 \ , \ \ ... \ , \ \ G_{n-1} + X \underset{b_r^n}{\overset{b_f^n}{\rightleftharpoons}} G_n$$

$$G_i \xrightarrow{k_i} G_i + m$$

$$m \xrightarrow{k_p} m + X \tag{2.10}$$

$$m \xrightarrow{\gamma_m} \varnothing$$

$$X \xrightarrow{\gamma} \varnothing$$
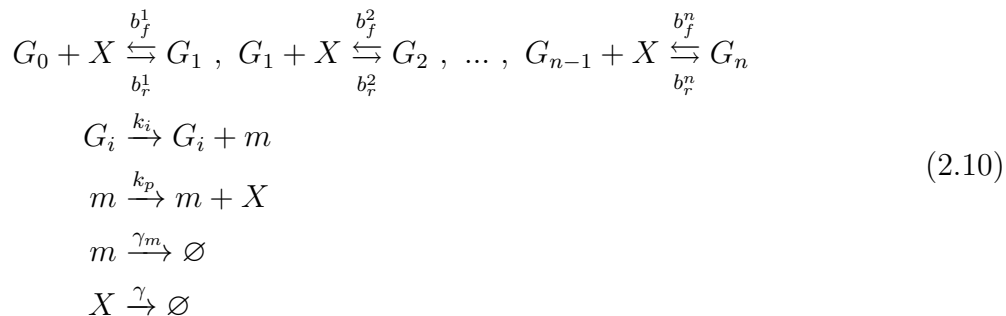
where m denotes the RNA produced by the X gene. In all, we have (i) reversible gene-protein binding reactions, (ii) transcription, (iii) translation, and (iv) degradation. Each of

these now represents a fundamental reaction type, and the corresponding set of mass action ODEs reads

$$
\begin{aligned}
\dot{g}_0 &= -b_f^1 g_0 x + b_r^1 g_1 \\
\dot{g}_i &= -b_f^{i+1} g_i x + b_r^{i+1} g_{i+1} - b_r^i g_i + b_f^i g_{i-1} x \\
\dot{g}_n &= -b_r^n g_n + b_f^n g_{n-1} x \\
\dot{m} &= \sum_{i=0}^{n} k_i g_i - \gamma_m m \\
\dot{x} &= k_p m - \gamma x + \sum_{i=1}^{n} -b_f^i g_{i-1} x + b_r^i g_i \; .
\end{aligned}
\tag{2.11}
$$

This certainly seems more complicated than what we started with—we went from a single ODE to $n + 3$! But it turns out that the system described by Eq. 2.11 can indeed be approximately described by Eq. 2.9, even if it is not obvious at first glance. To go from one to the other, we must exploit two things: the fact that we are only interested in the dynamics of x, and time scale separations.

Because we are only interested in the dynamics of x, we can eliminate the other variables however we like, as long as our elimination procedure can be reasonably well justified. One idea is to use the *quasi steady state approximation*. Generally, protein-promoter binding and transcription events happen somewhat more quickly than the time scales of protein translation and degradation. From the point of view of protein time scales, in fact, they can be approximated as equilibrating almost *instantaneously*. If we assume each of the protein-promoter binding reactions is approximately at equilibrium, we have

$$
b_f^{i+1} g_i x = b_r^{i+1} g_{i+1} \implies g_{i+1} = \left(\frac{b_f^{i+1}}{b_r^{i+1}} \right) g_i x
\tag{2.12}
$$

for all $i = 0, 1, ..., n-1$. If we define the coefficients c_i via

$$
\begin{aligned}
c_0 &:= 1 \\
c_i &:= \frac{b_f^1}{b_r^1} \cdots \frac{b_f^n}{b_r^n} \qquad (1 < i \le n)
\end{aligned}
\tag{2.13}
$$

then we can write each g_i in terms of x and g_0 as

$$
g_i = c_i x^i g_0 \; ,
\tag{2.14}
$$

meaning that we have successfully eliminated n variables. We can also eliminate g_0 by utilizing the constraint that $g_0 + g_1 + \cdots + g_n = G$ (i.e. the fraction of genes in each of the different gene states must sum to the total number of X genes), since

$$
g_0 + c_1 g_0 x + \cdots + c_n g_n x^n = g_0 \sum_{i=0}^{n} c_i x^i = G
$$

$$
\implies g_i = G \frac{c_i x^i}{\sum_{i=0}^{n} c_i x^i} \; .
\tag{2.15}
$$

14

Applying the quasi steady state approximation to the RNA dynamics is somewhat simpler, and yields

$$m = \frac{1}{\gamma_m} \sum_{i=0}^{n} k_i g_i = \frac{G}{\gamma_m} \frac{\sum_i k_i c_i x^i}{\sum_i c_i x^i} \ . \tag{2.16}$$

Having finally eliminated every variable but x, we find that x satisfies the ODE

$$\dot{x} = \frac{Gk_p}{\gamma_m} \frac{\sum_i k_i c_i x^i}{\sum_i c_i x^i} - \gamma x \ . \tag{2.17}$$

Under certain biologically reasonable assumptions, our complicated-looking rational function can be reasonably well approximated by a Hill function. In particular, we might suppose that: (i) transcription is 'all or nothing', so that each k_i aside from k_n is negligible; (ii) the fully bound gene state G_n is 'hard to reach' and 'hard to leave', so that each c_i with $i < n$ is very small but c_n is large. This leaves us with

$$\frac{Gk_p}{\gamma_m} \frac{\sum_i k_i c_i x^i}{\sum_i c_i x^i} \approx \frac{Gk_p}{\gamma_m} \frac{k_n c_n x^n}{1 + c_n x^n} = \frac{Gk_p k_n}{\gamma_m} \frac{x^n}{(1/c_n) + x^n} \tag{2.18}$$

which is the desired form modulo parameter redefinitions. But it should be kept in mind that such an approximation may not always be appropriate. In any case, what we have certainly learned is that quite a bit of biology can be hidden behind these Hill function terms.

This takes care of the non-fundamental propensity function problem; in essence, the solution is just to consider a different, more biologically grounded problem (usually with many more variables and parameters). What about the fact that this ODE model is deterministic and involves a continuous rather than discrete variable? This problem is, surprisingly, more easily dispatched.

As mentioned earlier, an ODE model either represents population-level averages or single cell dynamics in a certain limit. Both experimentally and biologically, the limiting behavior of single cell dynamics is both more relevant and more interesting (that is, in real organisms we have individual cells interacting, rather than hypothetical average cells interacting). What is the relevant limit?

It is just this issue that Gillespie treated in his analysis of the chemical Langevin equation [5], and of associated approximations like the large system volume limit [6]. We will discuss this in more detail in Chapter 5, but for now let us sketch the core idea.

When we are interested in relatively long time scales, so that many chemical reactions happen within a 'small' amount of time τ, to good approximation we can write

$$\dot{x} = f(x) + g(x)[\text{noise}] \tag{2.19}$$

where $f(x)$ and $g(x)$ depend on the CME being considered, and where the [noise] term usually represents fluctuations that are to good approximation a sum of Gaussian white noise terms. Especially as the number of x molecules gets larger, these fluctuations are less and less important compared to the average (deterministic) behavior, which means that an ODE model is a good approximation.

That covers the relationship between the CME and ODE models. What about other kinds of models? Although we emphasized earlier that it is not important for our purposes to track the spatial locations of individual molecules, the CME subsumes *reaction-diffusion models*, which do, as a special case. One can accomplish this by imagining a lattice on which chemical reactions can happen, and then taking the limit of infinitesimally small lattice spacing. For example, consider a system of molecules on a one-dimensional lattice that can randomly (i) diffuse, (ii) degrade, and (iii) be produced from a source at a specific location. This can be modeled using the CME

$$
\begin{aligned}
\frac{\partial P(\mathbf{X}, t)}{\partial t} = {} & \alpha \left[P(..., X_0 - 1, ..., t) - P(\mathbf{X}, t) \right] \\
& + \sum_i \frac{D}{(\Delta x)^2} \left[(X_i + 1) P(..., X_i + 1, X_{i+1} - 1, ..., t) - X_i P(\mathbf{X}, t) \right] \\
& + \frac{D}{(\Delta x)^2} \left[(X_i + 1) P(..., X_{i-1} - 1, X_i + 1, ..., t) - X_i P(\mathbf{X}, t) \right] \\
& + \gamma \left[(X_i + 1) P(..., X_i + 1, ..., t) - X_i P(\mathbf{X}, t) \right]
\end{aligned}
\tag{2.20}
$$

where $P(\mathbf{X}, t)$ is the probability that the system has state $\mathbf{X} = (..., X_{-2}, X_{-1}, X_0, X_1, X_2, ...)$ at time t, i.e. that lattice site i has $X_i \in \mathbb{N}$ molecules at time t. As it is now, this CME tracks the probability that the system has any possible number and configuration of molecules in space. If we think about the diffusion equation as describing distributions of spatial arrangements, this CME describes the *distribution of distributions of spatial arrangements*.

Usually, this is much more information than we want. If we are just interested in tracking how the most typical spatial arrangement of molecules evolves, we can use the mass action prescription and consider the ODEs

$$
\dot{X}_i = \frac{D}{(\Delta x)^2} \left[X_{i+1} - 2X_i + X_{i-1} \right] + \alpha \delta_{i,0} - \gamma X_i \ .
\tag{2.21}
$$

If we rewrite this in terms of the lattice spacing Δx, we must make the substitutions

$$
\begin{aligned}
i &\rightarrow i \Delta x \\
X_i &\rightarrow X(i \Delta x) \Delta x = X(x) \Delta x \ .
\end{aligned}
\tag{2.22}
$$

Then our ODEs read

$$
\frac{\partial X(x, t)}{\partial t} = \frac{D}{(\Delta x)^2} \left[X(x + \Delta x, t) - 2X(x, t) + X(x - \Delta x, t) \right] + \alpha \frac{\delta_{i\Delta x, 0}}{\Delta x} - \gamma X(x, t) \ .
\tag{2.23}
$$

Taking the $\Delta x \rightarrow 0$ limit yields the partial differential equation (PDE)

$$
\frac{\partial X(x, t)}{\partial t} = D \frac{\partial^2 X(x, t)}{\partial x^2} + \alpha \delta(x) - \gamma X(x, t)
\tag{2.24}
$$

which is an example of a reaction-diffusion equation. The nice thing about this approach to reaction-diffusion equations is that we understand *exactly* how the above equation relates

16

to an interpretable microscopic model, and can adjust our derivation to take more features into account if desired (for example, we could take fluctuations about this distribution into account as we did in Eq. 2.19).

In a different direction, we can consider the relationship between the CME and very coarse models, like Boolean networks and Markov models on finitely many states. The appropriateness of these kinds of approximations is usually determined by the *attractor structure* of our CME model, i.e. the regions of state space that will be occupied (in the long time limit) with high probability.

In the case of Markov models on finitely many states, our coarse graining procedure essentially lumps together all states sufficiently close to an attractor. For example, one can model stem cell development as a Markov chain by considering a model involving random transitions between cell states:

$$\text{Stem} \underset{k_r}{\overset{k_f}{\rightleftharpoons}} \text{Differentiated} . \tag{2.25}$$

An effective model like this may be appropriate if dynamics are dominated by two attractors (which biologically correspond to the different cell types), and if transitions between them happen on reasonable time scales. In principle, given a precise assignment of the microscopic states $\mathbf{x}$ to the macroscopic states 'Stem' and 'Differentiated', the CME model allows one to compute k_r and k_f from first principles. In practice, this is hopelessly difficult, and such parameters are usually treated as phenomenological numbers to be fit by experiment.

Boolean networks (either their deterministic or stochastic variants) work in much the same fashion. One lumps together states by treating molecule numbers as either 'high' or 'low', so that a state space $\mathbb{N}^N$ reduces[1] to a state space 2^N. Exactly where to make the cutoff depends on the details of the dynamics; in the case of a model like Eq. 2.9, the threshold concentration c would be the canonical choice since it corresponds to the half-max of the Hill function, and divides the 'high' attractor from the 'low' attractor. Then, instead of computing how an ODE changes in time, we create update rules (in either discrete or continuous time, and either synchronous or asynchronous) that approximately represent our ODEs or stochastic differential equations. For instance: *if* x is in the high state, *then* it will remain in the high state. *If* y is in the high state and x is in the high state, *then* x will next be in the low state. Such approximations, while crude, are probably necessary if one wants to examine large networks involving on the order of ten or a hundred genes. They have the advantage of being easier to compute than CME or ODE models, and much easier to parameterize.

A summary of some relationships between CME models and other modeling approaches is presented in Table 2.2. In the next chapter, we will move on to discussing how to work with the CME in practice—if we cannot actually solve it or use it to compute observables, all of this high-level discussion will have been in vain.

---

[1]Of course, one can make finer divisions of state space, for example by including a 'medium' state. This is just to illustrate the idea.

Modeling approach	Relationship to CME
Non-fundamental terms	Arise from approximating a more complicated CME with fundamental terms, and can often be justified by appeal to the quasi steady state approximation.
Stochastic differential equations	Time scale long enough for many chemical reactions to happen in a small amount of time τ; see discussion of chemical Langevin equation in Chapter 5. Can also work in terms of a system size expansion, but this is less reliable.
Ordinary differential equations	SDE approximation + large enough molecule numbers for fluctuations to be negligible.
Reaction-diffusion equations	Write down CME for diffusion and reactions on a lattice, consider the corresponding ODEs, and take the lattice spacing to zero.
Markov models on finitely many states	Coarse grain system according to attractors, e.g. if a state $\mathbf{x}$ is 'closest' to attractor A, identify $\mathbf{x}$ with A. Transition probabilities between coarse grained states in principle computable from CME.
Boolean networks	Coarse grain system according to characteristic concentration thresholds, e.g. if $x \geq c$, identify x with the 'high' state, or if $x < c$, identify x with the 'low' state. Approximate CME/SDE/ODE update rules by simpler Boolean variants. May also approximate time as discrete.

Table 2.2: A summary of the relationships between the CME and other modeling approaches sketched in this section.

Bibliography

[1] W. Gilbert and B. Müller-Hill. Isolation of the lac repressor. Proceedings of the National Academy of Sciences of the United States of America, 56(6):1891–1898, Dec 1966. 16591435[pmid].

[2] Benno Müller-Hill. The function of auxiliary operators. Molecular Microbiology, 29(1):13–18, 1998.

[3] Michail Stamatakis and Nikos V. Mantzaris. Comparison of deterministic and stochastic models of the lac operon genetic network. Biophysical Journal, 96(3):887–906, 2009.

[4] Daniel T. Gillespie. A rigorous derivation of the chemical master equation. Physica A: Statistical Mechanics and its Applications, 188(1):404 – 425, 1992.

[5] Daniel T. Gillespie. The chemical Langevin equation. <u>The Journal of Chemical Physics,</u> 113(1):297–306, 2000.

[6] Daniel T. Gillespie. Deterministic limit of stochastic chemical kinetics. <u>The Journal of Physical Chemistry B</u>, 113(6):1640–1644, 2009.

Chapter 3

Basic tools. Or: the chemical master equation in practice

3.1 First examples

In the previous chapter, we described our underlying philosophy, along with the ontology of the chemical master equation: it describes systems whose dynamics are *discrete* and *stochastic*, and cannot predict what *will* happen, but rather what is more or less *likely* to happen. It is especially suited for modeling how the numbers of RNA and proteins change within single cells.

As we described earlier, a CME model is defined by specifying four things:

1. a list of chemical species

2. a list of things that can happen to them

3. parameters/propensity functions that quantify how frequently they happen

4. initial conditions.

If one thinks about the CME dynamics like a play, the chemical species list specifies the 'actors', the chemical reaction list specifies the 'action', and the parameters specify stage directions (when and how often different things happen). A simple example of a CME model, involving two species and three reactions, is pictured in Fig. 3.1. The model describes RNA that is transcribed, spliced, and degraded.
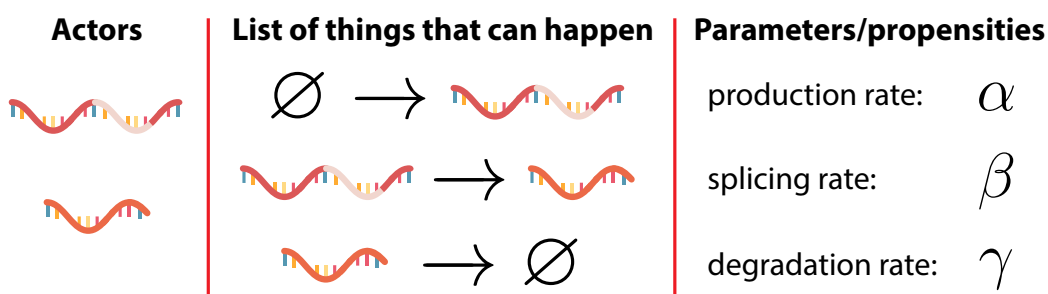
Figure 3.1: Requirements for specifying a CME model: a list of chemical species, and a list of parameterized reactions. Pictured is a model involving RNA transcription, splicing, and degradation.

This has all been very abstract so far; let us take a look at a few simple examples.

Example 1: Some simple CME models

The pure birth process:
$$R1 : \varnothing \xrightarrow{\alpha} X \tag{3.1}$$

- List of chemical species: $\{X\}$

- List of reactions: $\{R1\}$

- System state: $x \in \{0, 1, 2, ...\}$

- Propensity functions: $a_1(x) = \alpha$

- Stoichiometry vectors: $\nu_1 = +1$

The chemical birth-death process:

$$R1 : \varnothing \xrightarrow{\alpha} X$$
$$R2 : X \xrightarrow{\gamma} \varnothing \tag{3.2}$$

- List of chemical species: $\{X\}$

- List of reactions: $\{R1, R2\}$

- System state: $x \in \{0, 1, 2, ...\}$

- Propensity functions: $a_1(x) = \alpha$, $a_2(x) = \gamma\, x$

- Stoichiometry vectors: $\nu_1 = +1$, $\nu_2 = -1$

Transcription/splicing/degradation (pictured in Fig. 3.1):

$$R1 : \varnothing \xrightarrow{\alpha} X_1$$
$$R2 : X_1 \xrightarrow{\beta} X_2 \tag{3.3}$$
$$R3 : X_2 \xrightarrow{\gamma} \varnothing$$

- List of chemical species: $\{X_1, X_2\}$

- List of reactions: $\{R1, R2, R3\}$

- System state: $\mathbf{x} = (x_1, x_2) \in \mathbb{N}^2$

- Propensity functions: $a_1(\mathbf{x}) = \alpha$, $a_2(\mathbf{x}) = \beta\, x_1$, $a_3(\mathbf{x}) = \gamma\, x_2$

- Stoichiometry vectors: $\nu_1 = (+1, 0)$, $\nu_2 = (-1, +1)$, $\nu_3 = (0, -1)$

3.2 Simulating the CME

One way to solve a CME model (and, hence, determine its experimental predictions) is to simulate it many times, and use the simulation trajectories you found to compute the things you care about, like averages and probability distributions.

There are several methods for simulating a single stochastic trajectory. The simplest is the *rejection sampling method*, which involves taking small time steps of size Δt and checking whether or not a reaction happens at each time step (and if so, which one).

Using the fact that our propensity functions tell us the probability that any given reaction happens within a small amount of time Δt, we have the following algorithm.

Rejection sampling algorithm

i Generate a uniformly distributed random number r between 0 and 1.

ii If $r < a_1(\mathbf{x})\Delta t$, the first reaction happens. Update $\mathbf{x} \to \mathbf{x} + \boldsymbol{\nu}_1$, $t \to t + \Delta t$.
   If $\sum_{j=1}^{k-1} a_j(\mathbf{x})\Delta t \leq r < \sum_{j=1}^{k} a_j(\mathbf{x})\Delta t$, reaction k (for some $2 \leq k \leq M$) happens. Update $\mathbf{x} \to \mathbf{x} + \boldsymbol{\nu}_k$, $t \to t + \Delta t$.
   If $r \geq \sum_{j=1}^{M} a_j(\mathbf{x})\Delta t$, no reaction happens. Update $t \to t + \Delta t$.

iii Repeat.

As one might expect, this algorithm is incredibly wasteful, because for most time steps (assuming Δt is sufficiently small) nothing will happen. Is there a better way? Fortunately there is, thanks to the pioneering work of Gillespie [1, 2, 3]. The core idea of the *Gillespie algorithm* is to note that what we'd really like to know is when the *next* reaction happens, along with the probability that it is some *particular* reaction; advancing the system when it isn't changing is a waste of time and computational resources.

The first step towards discovering the Gillespie algorithm involves using what we know so far to write down the probability that no reaction happens in a small amount of time Δt. Since $a_j(\mathbf{x})\Delta t$ is the probability that reaction j happens, the probability that no reaction happens is

$$1 - \sum_{j=1}^{M} a_j(\mathbf{x})\Delta t \approx e^{-\sum_{j=1}^{M} a_j(\mathbf{x})\Delta t} \tag{3.4}$$

where we have used the fact that the left-hand side is equal to the first-order Taylor expansion of the right-hand side for sufficiently small Δt. Since no reaction happened in the small time Δt, $\mathbf{x}$ did not change; the probability that no reaction happens in the *next* small amount of time Δt *either* is

$$e^{-\sum_{j=1}^{M} a_j(\mathbf{x})\Delta t} e^{-\sum_{j=1}^{M} a_j(\mathbf{x})\Delta t} = e^{-2\sum_{j=1}^{M} a_j(\mathbf{x})\Delta t} \tag{3.5}$$

i.e. the expression above multiplied by itself. The probability that no reactions happen for an amount of time τ equal to taking n small time steps ($\tau = n\Delta t$) is

$$p(\text{wait } \tau) = e^{-\sum_{j=1}^{M} a_j(\mathbf{x})\Delta t} \cdots e^{-\sum_{j=1}^{M} a_j(\mathbf{x})\Delta t} = \exp\left\{-\sum_{j=1}^{M} a_j(\mathbf{x})\tau\right\} . \tag{3.6}$$

We can write this more succinctly by defining

$$a_0(\mathbf{x}) := \sum_{j=1}^{M} a_j(\mathbf{x}) . \tag{3.7}$$

Now that we know the probability that no reaction happens for some (not necessarily small) length of time τ, what's the probability that we have to wait a length of time τ *and* a reaction happens within a small amount of time Δt afterward? In other words, what's the chance that we have to wait for a length of time τ until the next reaction?

Since the jth propensity function tells us the probability (per unit time) that reaction j happens within a small amount of time Δt, the probability that we have a wait time τ *and* the next reaction is reaction j is given by

$$\begin{aligned} p(\text{wait } \tau \text{ and } j \text{ happens next}) &= a_j(\mathbf{x}) \, e^{-a_0(\mathbf{x})\tau} \\ &= \frac{a_j(\mathbf{x})}{a_0(\mathbf{x})} \, \cdot \, a_0(\mathbf{x}) e^{-a_0(\mathbf{x})\tau} \\ &= p(\text{choose reaction } j) \, \cdot \, p(\text{wait until } \tau) \end{aligned} \tag{3.8}$$

where the first factor can be interpreted as the probability we pick reaction j to happen next, and the second factor can be interpreted as the probability that our wait time is τ.

This yields the celebrated *Gillespie algorithm*: we can simulate our system by drawing our wait time from the exponential distribution above, and we can draw from a uniform distribution to determine which reaction fires next.

Although modern random sampling software packages have functions that allow one to draw from exponential distributions directly, we note that the inverse transform method for doing this generalizes to more complicated situations (and will be necessary when we consider time-dependent propensity functions). In particular, if the wait time $\tau \sim \text{Exp}(a_0)$, then we can accurately sample τ by drawing a uniformly distributed random variable $u \sim \text{Unif}[0,1]$ and computing

$$\tau = \frac{1}{a_0(\mathbf{x})} \log\left(\frac{1}{u}\right) . \tag{3.9}$$

The formula above comes from inverting the cumulative distribution function for an exponential distribution (i.e. $F(\tau) = 1 - e^{-a_0\tau}$). For future reference, we record the Gillespie algorithm in its entirety below.

Gillespie algorithm:

i Compute $a_0(\mathbf{x}) := \sum_{j=1}^{M} a_j(\mathbf{x})$.

ii Draw a random number $\tau \sim \mathrm{Exp}(a_0(\mathbf{x}))$; this is the length of time between the previous reaction and the next reaction.

iii Draw a random number $r \sim \mathrm{Unif}[0,1]$. If $\frac{1}{a_0} \sum_{j=1}^{k-1} a_j \leq r < \frac{1}{a_0} \sum_{j=1}^{k} a_j$, we say that reaction j fired. (Interpret $\sum_{j=1}^{0} a_j$ as zero.)

iv Update the state of the system: $(\mathbf{x}, t) \to (\mathbf{x} + \boldsymbol{\nu}_j, t + \tau)$.

v Repeat.

Example 2: Simulating the pure death process

Consider the system with just one reaction, a 'death' reaction, so that its reaction list reads

$$X \xrightarrow{\gamma} \varnothing . \tag{3.10}$$

This simple model is sometimes used to describe radioactive decay. Assuming γ is time-independent for simplicity, the propensity function for this system is

$$a(x) := \gamma\, x \tag{3.11}$$

where x is the current number of X molecules, and the stoichiometry vector is $\nu = -1$. Note the x-dependence; intuitively, the x-dependence comes from the fact that the death reaction can't happen if there are no X molecules left (i.e. $a(0) = 0$). Similarly, the probability that at least one molecule disappears within a small amount of time Δt increases when there are more molecules present. Let us simulate this system in two ways: first using the rejection sampling method, and then using the Gillespie algorithm.

Rejection sampling method

Assume that $\gamma = 1$, and that our initial number of molecules is $x_0 = 100$. First, we choose some small time step Δt; since we will be treating $\gamma x \Delta t$ as the probability that death happens in a given time step, we need $\gamma x_0 \Delta t$ to be less than 1. We will somewhat arbitrarily pick $\Delta t = 0.005$, so that $\gamma x_0 \Delta t = 0.5$.

At each time step, we will first compute the probabilities that (i) death happens, and that (ii) nothing happens. Then we will draw a random number $u \sim \mathrm{Unif}[0,1]$, and use that to decide which of the two possibilities happens for that time step. The table below summarizes the results of the first few steps of an example rejection sampling simulation.

We have used R(death) to denote the probability region that corresponds to picking that the death reaction happens, and R(nothing) to denote the probability region that corresponds to picking that nothing happens.

Step	Prev. x	R(death)	R(nothing)	u	Result	Time	x
0						0	100
1	100	0 - 0.5	0.5 - 1	0.948	nothing	0.005	100
2	100	0 - 0.5	0.5 - 1	0.227	death	0.01	99
3	99	0 - 0.495	0.495 - 1	0.594	nothing	0.015	99
4	99	0 - 0.495	0.495 - 1	0.428	death	0.02	98
5	98	0 - 0.49	0.49 - 1	0.764	nothing	0.025	98
6	98	0 - 0.49	0.49 - 1	0.003	death	0.03	97
7	97	0 - 0.485	0.485 - 1	0.357	death	0.035	96
8	96	0 - 0.48	0.48 - 1	0.91	nothing	0.04	96

From this table, we can immediately notice the problem with a rejection sampling approach: there are a lot of simulation steps where nothing happens! This problem would have been even worse had we chosen a smaller value for Δt.

Gillespie algorithm

Assume again that $\gamma = 1$ and $x_0 = 100$. The Gillespie algorithm does not require us to choose a step size Δt.

This time, each step of the simulation requires us to decide one thing: how long until the next reaction. Because our model has only one reaction, the next reaction is always guaranteed to be death. In order to decide when the next reaction happens, we draw a random number $u \sim \text{Unif}[0, 1]$, and compute

$$\tau = \frac{1}{\gamma\, x} \log\left(\frac{1}{u}\right) \ . \tag{3.12}$$

Notice that $\tau = \infty$ when $x = 0$; in other words, we must wait an infinitely long amount of time for the next reaction to happen when $x = 0$. This is as intended, because no more death reactions are possible when there are no more molecules left. The table below summarizes the results of the first few steps of an example Gillespie algorithm run.

25

Step	Prev. x	u	Wait time	Time	x
0				0	100
1	100	0.948	0.001	0.001	99
2	99	0.227	0.015	0.016	98
3	98	0.594	0.005	0.021	97
4	97	0.428	0.009	0.03	96
5	96	0.764	0.003	0.032	95
6	95	0.003	0.062	0.094	94
7	94	0.357	0.011	0.105	93
8	93	0.91	0.001	0.106	92

Because *something* is guaranteed to happen during each step of the Gillespie algorithm, this approach tends to run much faster in practice. Sample results for each approach are pictured below: as expected, they agree with each other.
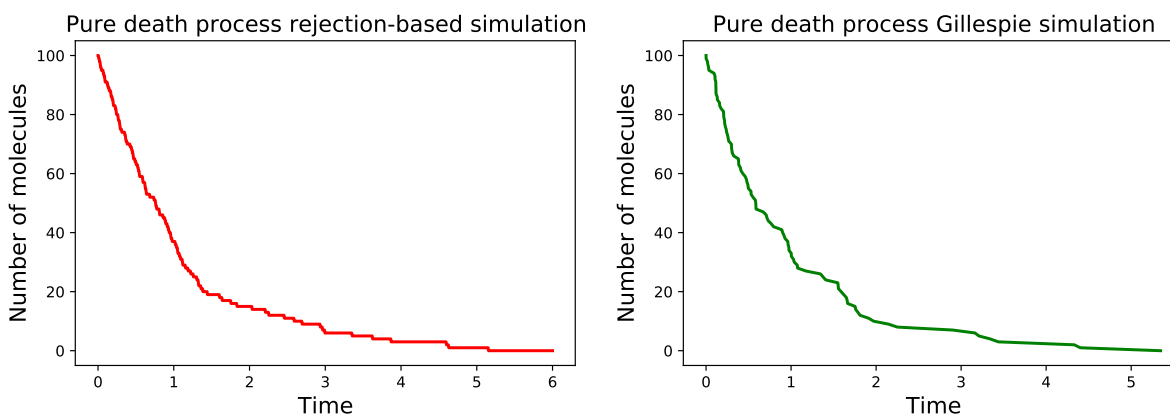
3.2.1 Simulating CMEs with time-dependent parameters

The Gillespie algorithm also works given reactions with time-dependent parameters, although working with them slightly more complicated. In that case, the waiting time distribution reads

$$p(\tau) = a_0(\mathbf{x}, t + \tau) \exp \left\{ - \int_t^{t+\tau} a_0(\mathbf{x}, t') \, dt' \right\} , \tag{3.13}$$

which we can learn by slightly generalizing the previous argument.

Since we are assuming the propensity functions $a_j(\mathbf{x}, t')$ are time-dependent, this waiting time distribution will typically not be a simple exponential distribution. In fact, the integral that appears in the argument of the exponential is generally not possible to evaluate in closed form.

But, in principle, we can use the inverse transform trick we described above to sample this distribution. This time, we must invert the cumulative distribution function for this more complicated distribution, which reads

$$F(\tau) = 1 - \exp\left\{ -\int_t^{t+\tau} a_0(\mathbf{x}, t') \, dt' \right\} . \tag{3.14}$$

Doing so, we can draw a $u \sim \text{Unif}[0, 1]$ and find the value of τ that satisfies

$$\int_t^{t+\tau} a_0(\mathbf{x}, t') \, dt' = \log\left(\frac{1}{u}\right) . \tag{3.15}$$

Numerically, this can be done using 'shooting'-type methods, where several values of τ are tried. As the name implies, if the value you tried is too high, adjust the next trial value to be smaller; if the value you tried it too small, adjust the next trial value to be larger. Eventually, you should find a value of τ that is correct up to a desired error tolerance. But there may be other methods that work faster when more is known about the time-dependence of the propensity functions.

For future reference, we record the Gillespie algorithm for time-dependent reaction rates below.

Gillespie algorithm (time-dependent reaction rates):

i Draw a random number $r_1 \sim \text{Unif}[0, 1]$. Find the value of τ that satisfies

$$\int_t^{t+\tau} a_0(\mathbf{x}, t') \, dt' = \log\left(\frac{1}{r_1}\right) . \tag{3.16}$$

This is the length of time between the previous reaction and the next reaction.

ii Draw a random number $r \sim \text{Unif}[0, 1]$.
If $\frac{1}{a_0(t+\tau)} \sum_{j=1}^{k-1} a_j(t+\tau) \leq r < \frac{1}{a_0(t+\tau)} \sum_{j=1}^{k} a_j(t+\tau)$, we say that reaction j fired.

iii Update the state of the system: $(\mathbf{x}, t) \rightarrow (\mathbf{x} + \boldsymbol{\nu}_j, t + \tau)$.

iv Repeat.

3.3 The CME as a matrix equation

Using variants of the Gillespie algorithm presented in the previous section, we can simulate any CME we like in order to determine how a system will behave. But this may be computationally expensive or slow, and it would be nice to have alternative methods that do not depend on simulations. The most naive approach to working with the CME is treating it as a matrix ODE. This works perfectly well when the system's state space is finite, as in the case of the gene that randomly switches between two states:

$$
\begin{aligned}
G_1 &\xrightarrow{k_{21}} G_2 \\
G_2 &\xrightarrow{k_{12}} G_1
\end{aligned}
\tag{3.17}
$$
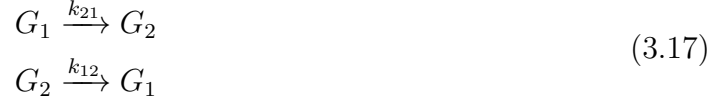
whose dynamics are described by the CME

$$
\begin{aligned}
\frac{\partial P(1,t)}{\partial t} &= -k_{21}P(1,t) + k_{12}P(2,t) \\
\frac{\partial P(2,t)}{\partial t} &= k_{21}P(1,t) - k_{12}P(2,t)
\end{aligned}
\tag{3.18}
$$

where $P(1,t)$ is the probability that the gene is in state 1 at time t, and $P(2,t)$ is the probability that the gene is in state 2 at time t. If we define the state vector

$$
\vec{P}(t) := \begin{pmatrix} P(1,t) \\ P(2,t) \end{pmatrix} ,
\tag{3.19}
$$

then we can write the CME in the form

$$
\dot{\vec{P}} = H\vec{P} ,
\tag{3.20}
$$

where H is the matrix

$$
H := \begin{pmatrix} -k_{21} & k_{12} \\ k_{21} & -k_{12} \end{pmatrix} .
\tag{3.21}
$$

By analogy with quantum mechanics, we will call this matrix the *Hamiltonian*, or the Hamiltonian operator. We will have much more to say about this system in Chapter 6, but for now we will just point out that it can be solved using a variety of conventional methods. For example, we can exponentiate H to find $\vec{P}(t)$, or we can diagonalize H and write down a solution in terms of its eigenvectors. If we are only interested in the steady state solution, as is often the case, we can use something like Gaussian elimination to compute the nullspace of H.

What if the state space of one's system is not finite (which is true most of the time)? For the sake of concreteness, consider the birth-death process ($\varnothing \xrightarrow{\alpha} X$, $X \xrightarrow{\gamma} \varnothing$), whose CME reads

$$
\frac{\partial P(x,t)}{\partial t} = \alpha \left[P(x-1,t) - P(x,t) \right] + \gamma \left[(x+1)P(x+1,t) - xP(x,t) \right] .
\tag{3.22}
$$

The infinite-dimensional matrix H in this case is

$$H = \begin{pmatrix} -\alpha & \gamma & & \\ \alpha & -\alpha - \gamma & 2\gamma & \\ & \alpha & -\alpha - 2\gamma & 3\gamma \\ & & & \ddots \end{pmatrix} . \tag{3.23}$$

One thing we could try would be to choose some cutoff state C, and *approximate* the solution $P(x, t)$ by treating H like an $(C + 1) \times (C + 1)$ matrix. For C sufficiently large (in particular, somewhat larger than where most of the support of the probability distribution is concentrated), we would expect this approximation to work fairly well. In the case of the birth-death process, and other processes whose molecule numbers do not tend to increase without bound, this approach does work well. In the case of something like the pure birth process, it will not, because there will always be some time after which most of the probability distribution is concentrated at states $y > C$.

This idea, plus a few technical tricks for carefully estimating the error introduced by the approximation, is the basis of the *finite state projection* approach [4, 5, 6] to solving the CME. Introduce a cutoff for state space, then solve using standard matrix ODE methods (especially sparse matrix exponentiation algorithms, which apply because H tends to consist mostly of zeros). The problem is that the method does not scale well as one considers larger state spaces. If there are N chemical species, one needs N cutoffs $C_1, ..., C_N$, and the size of the matrix will be $(C_1 + 1) \times \cdots \times (C_N + 1)$; if each C_i is on the order of ten or a hundred, this number gets large very quickly. While there are various ideas for optimizing finite state projection algorithms, the fundamental limitation for treating H like a matrix remains the sheer computational expense of dealing with very large matrices.

3.4 Conventional analytic approaches to solving the CME

We may as well face up to the fact that H is infinite-dimensional, and try to develop methods that respect this fundamental feature of the CME. Just as there are methods for solving PDEs other than discretization, we would like analogous theoretical methods for solving the CME. In this section, we will illustrate a few approaches in the context of the birth-death process, whose CME is Eq. 6.29.

3.4.1 Method of guessing

Much of what we know about how to solve differential equations comes from systematic guessing—and the method can also work here. It works best for finding the steady state probability distribution $P_{ss}(x)$, which is governed by the equation

$$0 = \alpha \left[P_{ss}(x - 1) - P_{ss}(x) \right] + \gamma \left[(x + 1) P_{ss}(x + 1) - x P_{ss}(x) \right] . \tag{3.24}$$

Rearranging, we have the simple recurrence relation

$$P_{ss}(x+1) = \frac{\left(x + \frac{\alpha}{\gamma}\right)P_{ss}(x) - \frac{\alpha}{\gamma}P_{ss}(x-1)}{(x+1)} \tag{3.25}$$

whose beginning ($x = 0$) looks like

$$P_{ss}(1) = \frac{\alpha}{\gamma}P(0) . \tag{3.26}$$

Continuing iteratively, we can find

$$P_{ss}(2) = \frac{(\alpha/\gamma)^2 P(0)}{2} , P_{ss}(3) = \frac{(\alpha/\gamma)^3 P(0)}{3 \cdot 2} , ... \tag{3.27}$$

This suggests the ansatz

$$P_{ss}(x) = (\text{const.})\frac{(\alpha/\gamma)^x}{x!} \tag{3.28}$$

which substitution into Eq. 3.25 shows is correct. Normalizing so that the total probability is 1, we obtain a Poisson distribution:

$$P_{ss}(x) = \frac{(\alpha/\gamma)^x e^{-\alpha/\gamma}}{x!} . \tag{3.29}$$

This method for computing P_{ss} works quite well when one is considering relatively simple one-dimensional (i.e. one species) systems. For systems involving more species, the recurrence relation will likely be hard to solve, so one must make a bold guess and hope for the best.

Knowing that the steady state distribution of the birth-death process is Poisson may lead us to boldly extrapolate, and try a function

$$P_{ss} \overset{?}{=} \frac{\mu(t)^x e^{-\mu(t)}}{x!} \tag{3.30}$$

for solving the full distribution. Substituting into Eq. 6.29 and simplifying yields the constraint

$$\dot{\mu} = \alpha - \gamma\mu \tag{3.31}$$

from which we obtain that

$$\mu(t) = \mu_0 e^{-\gamma t} + \frac{\alpha}{\gamma}\left(1 - e^{-\gamma t}\right) . \tag{3.32}$$

So we have solved Eq. 6.29 for at least *some* initial conditions (e.g. $P_0(x) = \delta_{x,0}$, which is a degenerate Poisson distribution and corresponds to $\mu_0 = 0$), but we have not solved it in general. The general form of the solution for arbitrary initial conditions, even for a problem this simple, is hard to guess.

3.4.2 Method of generating functions

Suppose we want a more principled approach. One idea is to convert the differential-difference equation into something more familiar, like a PDE. This can be done by considering the *generating function*

$$\psi(z,t) := \sum_{x=0}^{\infty} P(x,t)z^x \tag{3.33}$$

where $z \in \mathbb{C}$. If we can compute $\psi(z,t)$ for $z = e^{i\theta}$ (that is, for z on the complex unit circle), then we can compute $P(x,t)$ via an inverse Fourier transform, since

$$\int_{-\infty}^{\infty} \frac{d\theta}{2\pi} e^{-i\theta x} \psi(\theta,t) = \sum_y P(y,t) \int_{-\infty}^{\infty} \frac{d\theta}{2\pi} e^{-i\theta(x-y)} = \sum_y P(y,t)\delta(x-y) = P(x,t) \tag{3.34}$$

where we have used the usual integral representation of the Dirac delta function. Substituting the definition of the generating function into Eq. 6.29 and simplifying yields the PDE

$$\frac{\partial \psi(z,t)}{\partial t} = \alpha(z-1)\psi(z,t) - \gamma(z-1)\frac{\partial \psi(z,t)}{\partial z} . \tag{3.35}$$

This can be solved using techniques like the *method of characteristics*, which supposes that the solution lies along some parameterized curve, so that $\psi(z,t) = \psi(Z(s),T(s))$ for some value of s. Suppose this were true, and that

$$\frac{d\psi}{ds} = \frac{\partial \psi}{\partial T}\frac{\partial T}{\partial s} + \frac{\partial \psi}{\partial Z}\frac{\partial Z}{\partial s} = \frac{\partial \psi}{\partial T} + \frac{\partial \psi}{\partial Z}\gamma(z-1) . \tag{3.36}$$

Then our PDE becomes

$$\frac{d\psi}{ds} = \alpha(Z(s)-1)\psi , \tag{3.37}$$

which is relatively simple to solve. The constraints in Eq. 3.36 force

$$\begin{aligned} \frac{\partial T}{\partial s} &= 1 \\ \frac{\partial Z}{\partial s} &= \alpha(Z-1) . \end{aligned} \tag{3.38}$$

Solving for $T(s)$ and $Z(s)$, we find

$$\begin{aligned} T(s) &= s \\ Z(s) - 1 &= Ce^{\gamma s} \end{aligned} \tag{3.39}$$

where C is an arbitrary constant, and we have arbitrarily chosen $T(0) = 0$. Since $t = T(t)$, $T(0)$ corresponds to $t = 0$, so $s \in [0,t]$. We must also have $Z(s = t) = z$, so that $(z,t) =$

$(Z(s=t), T(s=t))$. Going back to Eq. 3.37, pushing forward yields

$$\frac{d\psi}{\psi} = \alpha C e^{\gamma s}$$

$$\implies \int_0^t \frac{d\psi}{\psi} ds = \alpha C \int_0^t e^{\gamma s} ds$$

$$\implies \log \frac{\psi(Z(t), T(t))}{\psi(Z(0), T(0))} = \frac{\alpha}{\gamma} C \left[e^{\gamma t} - 1 \right] = \frac{\alpha}{\gamma} (Z(t) - 1) e^{-\gamma t} \left[e^{\gamma t} - 1 \right]$$

$$\implies \psi(z,t) = \psi(Z(0), 0) \exp \left\{ \frac{\alpha}{\gamma} (z - 1) \left(1 - e^{-\gamma t} \right) \right\} . \tag{3.40}$$

The initial condition for the generating function depends on the initial condition for $P(x,t)$; because we can construct the solution for $P(x,t)$ provided that we know the transition probability $P(x,t; x_0, 0)$ for arbitrary $x_0 \in \mathbb{N}$, let us try to compute the transition probability by supposing $P_0(x) = \delta_{x,x_0}$, so that $\psi(z,0) = z^{x_0}$. Finally, we have

$$\psi(z,t) = Z(0)^{x_0} \exp \left\{ \frac{\alpha}{\gamma} (z - 1) \left(1 - e^{-\gamma t} \right) \right\}$$

$$= \left[1 + (z - 1) e^{-\gamma t} \right]^{x_0} \exp \left\{ \frac{\alpha}{\gamma} (z - 1) \left(1 - e^{-\gamma t} \right) \right\} . \tag{3.41}$$

In order to recover $P(x,t)$, we can either perform an inverse Fourier transform or match coefficients (i.e. $P(x,t)$ corresponds to the coefficient of z^x in Eq. 3.41). Because this problem is relatively simple, let us do the latter. We have

$$\psi(z,t) = \left[(1 - e^{-\gamma t}) + z e^{-\gamma t} \right]^{x_0} \exp \left\{ \frac{\alpha}{\gamma} (z - 1) \left(1 - e^{-\gamma t} \right) \right\}$$

$$= e^{-\frac{\alpha}{\gamma} \left(1 - e^{-\gamma t} \right)} \sum_{k=0}^{x_0} \binom{x_0}{k} \left[1 - e^{-\gamma t} \right]^{x_0 - k} \left[e^{-\gamma t} \right]^k \sum_{n=0}^{\infty} \frac{\left[\frac{\alpha}{\gamma} (1 - e^{-\gamma t}) \right]^n}{n!} z^{n+k} . \tag{3.42}$$

Cleverly reindexing and rearranging the sums, we obtain

$$\psi(z,t) = e^{-\frac{\alpha}{\gamma} \left(1 - e^{-\gamma t} \right)} \sum_{k=0}^{\min(x, x_0)} \binom{x_0}{k} \left[1 - e^{-\gamma t} \right]^{x_0 - k} \left[e^{-\gamma t} \right]^k \sum_{x=0}^{\infty} \frac{\left[\frac{\alpha}{\gamma} (1 - e^{-\gamma t}) \right]^{x-k}}{(x - k)!} z^x$$

$$= \sum_{x=0}^{\infty} z^x \, e^{-\frac{\alpha}{\gamma} \left(1 - e^{-\gamma t} \right)} \sum_{k=0}^{\min(x, x_0)} \binom{x_0}{k} \left[1 - e^{-\gamma t} \right]^{x_0 - k} \left[e^{-\gamma t} \right]^k \frac{\left[\frac{\alpha}{\gamma} (1 - e^{-\gamma t}) \right]^{x-k}}{(x - k)!} , \tag{3.43}$$

so that our final answer for the transition probability is

$$P(x,t; x_0, 0) = e^{-\frac{\alpha}{\gamma} \left(1 - e^{-\gamma t} \right)} \sum_{k=0}^{\min(x, x_0)} \binom{x_0}{k} \left[1 - e^{-\gamma t} \right]^{x_0 - k} \left[e^{-\gamma t} \right]^k \frac{\left[\frac{\alpha}{\gamma} (1 - e^{-\gamma t}) \right]^{x-k}}{(x - k)!} . \tag{3.44}$$

As a sanity check, we note that if $x_0 = 0$ (so that the initial condition is a degenerate Poisson distribution), we recover a Poisson distribution with a mean given by Eq. 3.32 with $\mu_0 = 0$.

At this point, we should make several observations. First, despite the birth-death process being ostensibly the *simplest* nontrivial and biologically relevant example of a CME model, finding its complete time-dependent solution was not simple. We shall have more to say about finding finding analytic solutions to the CME (e.g. in Chapter 4), but our difficulties here do not bode well for more realistic problems involving more than one species and more complicated interactions between species. Second, it would have been much easier to find the steady state solution; starting from Eq. 3.35, we could have set $\partial \psi / \partial t = 0$ and solved

$$\frac{\partial \psi_{ss}(z)}{\partial z} = \frac{\alpha}{\gamma} \psi_{ss}(z) \tag{3.45}$$

quickly to recover the Poisson distribution answer. That steady state solutions are significantly easier to find is also a generic fact about solving the CME.

Overall, this is the most widely used method, and it is generally effective (at least, for sufficiently simple systems that are amenable to an analytic approach in the first place). If one does not have a good guess prepared, trying to solve (or approximately solving) a generating function PDE like Eq. 3.35 is a good thing to try.

3.4.3 Method of Poisson representation

This method is closely related to the previous one, in the sense that it also involves converting the CME to a possibly more tractable PDE. The idea [7, 8, 9] is that we try to write $P(x, t)$ as a superposition of (typically infinitely many) Poisson distributions, i.e.

$$P(x, t) = \int_0^\infty dz \, \frac{z^x}{x!} e^{-z} q(z, t) \tag{3.46}$$

for some function $q(z, t)$. We convert Eq. 6.29 into an equation describing the time evolution of $q(z, t)$, solve for $q(z, t)$, and then Eq. 3.46 provides our solution for $P(x, t)$. Given a sufficiently well-behaved $q(x, t)$, we may also hope to manipulate Eq. 3.46 to get rid of the integral, so that our solution is even simpler-looking. Note that

$$
\begin{aligned}
P(x-1, t) &= \int_0^\infty \frac{z^{x-1}}{(x-1)!} e^{-z} q(z, t) \\
&= \int_0^\infty \frac{\partial}{\partial z} \left[\frac{z^x}{x!} \right] e^{-z} q(z, t) \\
&= \int_0^\infty \frac{z^x}{x!} \left[-\frac{\partial}{\partial z} \left(e^{-z} q(z, t) \right) \right] \\
&= \int_0^\infty \frac{z^x}{x!} e^{-z} \left[q(z, t) - \frac{\partial q(z, t)}{\partial z} \right] \quad .
\end{aligned}
\tag{3.47}
$$

In other words, in converting the CME into an equation for $q(z,t)$, we can substitute $P(x - 1, t) \to q - \frac{\partial q}{\partial z}$. By similar computations, we have:

$$P(x - 1, t) - P(x, t) \to -\frac{\partial q(z, t)}{\partial z}$$

$$(x + 1)P(x + 1, t) - xP(x, t) \to q(z, t) + z\frac{\partial q(z, t)}{\partial z} \ , \tag{3.48}$$

so $q(z, t)$ satisfies the PDE

$$\frac{\partial q(z, t)}{\partial t} = -\alpha\frac{\partial q(z, t)}{\partial z} + \gamma\left(q(z, t) + z\frac{\partial q(z, t)}{\partial z}\right) \ , \tag{3.49}$$

which we can rewrite as

$$\frac{1}{\gamma}\frac{\partial q(z, t)}{\partial t} - (z - \mu)\frac{\partial q(z, t)}{\partial z} = q(z, t) \ . \tag{3.50}$$

This is another PDE we can solve by the method of characteristics. Suppose $q(z, t) = q(Z(s), T(s))$, and obtain the constraints

$$\frac{\partial T(s)}{\partial s} = \frac{1}{\gamma}$$

$$\frac{\partial Z(s)}{\partial s} = \mu - Z(s) \tag{3.51}$$

which we can easily solve to find

$$T(s) = \frac{s}{\gamma}$$

$$Z(s) = Z_0 e^{-s} + \mu\left(1 - e^{-s}\right) \tag{3.52}$$

where we have arbitrarily chosen $T(0) = 0$, and Z_0 is an arbitrary constant. Since $T(s = \gamma t) = t$, we must have $Z(s = \gamma t) = z$. Also, if we the time corresponding to our initial condition is zero, we have $s \in [0, \gamma t]$. Continuing on,

$$\frac{dq}{ds} = \gamma q$$

$$\implies \int_0^{\gamma t}\frac{dq}{q} = \int_0^{\gamma t} ds$$

$$\implies \log\frac{q(z, t)}{q(Z(0), 0)} = \gamma t \tag{3.53}$$

$$\implies q(z, t) = q(Z(0), 0)e^{\gamma t} \ .$$

Noting that

$$z = Z(\gamma t) = Z_0 e^{-\gamma t} + \mu\left(1 - e^{-\gamma t}\right) \implies Z_0 = e^{\gamma t}\left[z - \mu\left(1 - e^{-\gamma t}\right)\right] \ , \tag{3.54}$$

all we have to do is determine $q(z,0)$ based on our initial condition. Assuming we are interested in the transition probability $P(x,t;x_0,0)$, we want $P_0(x) = \delta_{x,x_0}$. But it is not obvious how to satisfy

$$\delta_{x,x_0} = \int_0^\infty dz \frac{z^x}{x!} e^{-z} q(z,0) \ . \tag{3.55}$$

An inspired choice turns out to be

$$q(z,0) = \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-izp} \tag{3.56}$$

since

$$\delta_{x,x_0} = \int_0^\infty dz \int_{-\infty}^\infty \frac{dp}{2\pi} \frac{z^x}{x!} (1+ip)^{x_0} e^{-izp-z} \tag{3.57}$$

is an integral representation of the Kronecker delta function (incidentally, it is one we will use later in Chapter 4). Then our solution for $q(z,t)$ becomes

$$\begin{aligned} q(z,t) &= e^{\gamma t} \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-iZ(0)p} \\ &= e^{\gamma t} \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-ie^{\gamma t}\left[z-\mu\left(1-e^{-\gamma t}\right)\right]p} \ . \end{aligned} \tag{3.58}$$

Hence, using Eq. 3.46,

$$P(x,t;x_0,0) = e^{\gamma t} \int_0^\infty dz \frac{z^x}{x!} e^{-z} \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-ie^{\gamma t}\left[z-\mu\left(1-e^{-\gamma t}\right)\right]p} \ . \tag{3.59}$$

In this case, we can manipulate the integral to simplify it. Change variables to

$$w := e^{\gamma t}\left[z - \mu\left(1 - e^{-\gamma t}\right)\right] \tag{3.60}$$

so that $dw = e^{\gamma t}dz$, and our expression becomes

$$P(x,t;x_0,0) = e^{-\mu(1-e^{-\gamma t})} \int_{w_{min}}^\infty dw \frac{[e^{-\gamma t}w + \mu(1-e^{-\gamma t})]^x}{x!} e^{-e^{-\gamma t}w} \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-ipw} \tag{3.61}$$

where $w_{min} := -\mu\left(e^{\gamma t} - 1\right) \geq 0$. We can use a trick to write

$$\begin{aligned} \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-ipw} &= e^w \int_{-\infty}^\infty \frac{dp}{2\pi} \ (1+ip)^{x_0} e^{-(1+ip)w} \\ &= e^w \left(-\frac{\partial}{\partial w}\right)^{x_0} \int_{-\infty}^\infty \frac{dp}{2\pi} e^{-ipw} \\ &= e^w \left(-\frac{\partial}{\partial w}\right)^{x_0} e^{-w}\delta(w) \ . \end{aligned} \tag{3.62}$$

Integrating by parts, we have

$$e^{-\mu(1-e^{-\gamma t})} \int_{w_{min}}^{\infty} dw \frac{1}{x!} \left(\frac{\partial}{\partial w}\right)^{x_0} \left\{ \left[e^{-\gamma t} w + \mu(1-e^{-\gamma t})\right]^x e^{\left(1-e^{-\gamma t}\right)w} \right\} e^{-w} \delta(w)$$

$$= e^{-\mu(1-e^{-\gamma t})} \sum_{k=0}^{x} \binom{x}{k} \left(e^{-\gamma t}\right)^k \left[\mu(1-e^{-\gamma t})\right]^{x-k} \frac{1}{x!} \int_{w_{min}}^{\infty} dw \frac{1}{x!} \left(\frac{\partial}{\partial w}\right)^{x_0} \left\{ w^k e^{\left(1-e^{-\gamma t}\right)w} \right\} e^{-w} \delta(w) \ .$$

$$(3.63)$$

Since

$$\int_{w_{min}}^{\infty} dw \frac{1}{x!} \left(\frac{\partial}{\partial w}\right)^{x_0} \left\{ w^k e^{\left(1-e^{-\gamma t}\right)w} \right\} e^{-w} \delta(w) = \frac{x_0!}{(x_0-k)!} \left[\left(1-e^{-\gamma t}\right)\right]^{x_0-k} \qquad (3.64)$$

provided $k \leq x_0$, and zero otherwise, our final answer is

$$P(x,t;x_0,0) = e^{-\mu(1-e^{-\gamma t})} \sum_{k=0}^{\min(x,x_0)} \binom{x}{k} \left(e^{-\gamma t}\right)^k \left[\mu(1-e^{-\gamma t})\right]^{x-k} \frac{1}{x!} \frac{x_0!}{(x_0-k)!} \left[\left(1-e^{-\gamma t}\right)\right]^{x_0-k}$$

$$(3.65)$$

which is the same as Eq. 3.44. One thing we should note is that, while it was not that hard to derive Eq. 3.59, converting it into a more useful form was tedious. This seems to be a generic feature of various methods for solving the CME.

3.4.4   Method of eigenfunction expansion

This method should be familiar from the study of PDEs. The idea is that we find the eigenfunctions of H, expand the solution in terms of them (yielding a Fourier-like series), and exploiting some kind of orthogonality property of the eigenfunctions to compute the coefficients of the solution (*a la* exploiting the orthogonality of sines and cosines to compute Fourier coefficients).

Eigenfunction expansion can work, as we will see shortly, but oftentimes trying to find eigenfunctions is no less difficult than solving the full problem. Even if it is, one can run into issues in computing the Fourier-like coefficients of the solution if one has not established an orthogonality property satisfied by the eigenfunctions. And even if the method works, it may be because one happens to get lucky, and identify the eigenfunctions as a known class of special functions (as we are about to see). In case it is not clear, this method is often cumbersome and hard to generalize.

Consider again the birth-death process, and suppose that we already know that the steady state distribution is Poisson with mean $\mu := \alpha/\gamma$. The first step of this method is to write

$$P(x,t) = Q(x,t) P_{ss}(x) \qquad (3.66)$$

and substitute this ansatz into the CME, Eq. 6.29. Doing so, we find

$$\frac{\partial Q(x,t)}{\partial t} = \alpha \left[Q(x+1,t) - Q(x,t)\right] + \gamma x \left[Q(x-1,t) - Q(x,t)\right] \qquad (3.67)$$

36

which is a sort of mirror image of the CME (notice especially how the $x+1$ terms and $x-1$ terms swapped relative to the parameters they were next to in Eq. 6.29). Now, since we are looking for eigenfunctions, we write

$$Q(x,t) = Q_E(x)e^{-Et} \ , \tag{3.68}$$

and substitute to find

$$-EQ_E(x) = \alpha\left[Q_E(x+1) - Q_E(x)\right] + \gamma x\left[Q_E(x-1) - Q_E(x)\right] \ . \tag{3.69}$$

At this point, we have two tasks: find the allowed values of E, and determine each of the functions $Q_E(x)$. Let us deal with the allowed values of E first. Because the long time limit of the system's dynamics exists, we need $E > 0$ (otherwise, Eq. 3.68 would blow up as we take $t \to \infty$). Because, additionally, the long time limit of $P(x,t)$ is a *Poisson* distribution, we know that all moments exist. This should also be true of the eigenfunctions of H, the functions $Q_E(x)P_{ss}(x)$.

One way to study this constraint is to consider the function

$$\phi(z) := \sum_{x=0}^{\infty} \frac{(\mu z)^x}{x!} Q_E(x)e^{-\mu} \tag{3.70}$$

analogous to the generating function we defined earlier, but with extra factors to make the function on the inside match the Poisson distribution when $z = 1$. The purpose of introducing this function, as with introducing the generating function before, is to convert Eq. 3.69 to a more manageable PDE, and also to check to see if our finite moments constraint is satisfied. When we take z derivatives of ϕ and set $z = 1$, we recover factorial moments:

$$\langle x(x-1)\cdots(x-k+1)\rangle = \left.\frac{\partial^k \phi(z)}{\partial z^k}\right|_{z=1} = \sum_{x=0}^{\infty} [x(x-1)\cdots(x-k+1)]\frac{\mu^x}{x!} Q_E(x)e^{-\mu} \ . \tag{3.71}$$

After some algebra, we can find that $\phi(z)$ satisfies the PDE

$$\frac{E}{\gamma}\phi(z) = (z-1)\left(\frac{\partial \phi(z)}{\partial z} - \mu\phi(z)\right) \tag{3.72}$$

which we can solve to find

$$\phi(z) = e^{\mu(z-1)}(z-1)^{E/\gamma} \tag{3.73}$$

where the arbitrary constant has been fixed to make the $E = 0$ case correspond to the total probability of a Poisson distribution when $z = 1$ (i.e. $\phi(1) = 1$). Now we can notice something important: unless E/γ is a nonnegative integer, not all factorial moments will be finite, because taking sufficiently many derivatives and setting $z = 1$ yields

$$\left.\frac{\partial^k \phi(z)}{\partial z^k}\right|_{z=1} \sim \left.\frac{1}{(z-1)^{k-E/\gamma}}\right|_{z=1} = \infty \ . \tag{3.74}$$

Hence, it must be that E/γ is a nonnegative integer, i.e. that the allowed energies are

$$E_n = \gamma\, n \tag{3.75}$$

for $n \in \mathbb{N}$. Because we have a countably infinite number of allowed energies, we can relabel $Q_E(x) \to Q_n(x)$. Now we can proceed in two ways. Either we analyze our solution for $\phi(z)$ (Eq. 3.73) to extract the functional form of $Q_n(x)$, which we could do to find that

$$Q_n(x) = C_n(x, \mu) = \sum_{k=0}^{n} \binom{n}{k} (-1)^{n-k} \mu^{-k}\, x(x-1)\cdots(x-k+1) \tag{3.76}$$

where $C_n(x, \mu)$ is the nth Charlier polynomial. Alternatively, we could substitute $E_n = \gamma n$ into Eq. 3.69, and use prior special functions knowledge to notice that the recurrence relation we obtain is satisfied by the Charlier polynomials. Either way, we have eigenfunctions $P_n(x)$ given by

$$P_n(x) \propto Q_n(x) P_{ss}(x) = C_n(x, \mu) P_{ss}(x)\ . \tag{3.77}$$

The Charlier polynomials satisfy an orthogonality property

$$\sum_{x=0}^{\infty} C_n(x, \mu) C_m(x, \mu) \frac{\mu^x e^{-\mu}}{x!} = \delta_{n,m} \frac{n!}{\mu^n} \tag{3.78}$$

which we could either prove ourselves, or look up. Finally, we can write the general solution for $P(x, t)$ as an eigenfunction expansion

$$P(x, t) = \sum_{n=0}^{\infty} d_n C_n(x, \mu) P_{ss}(x) e^{-\gamma n t} \tag{3.79}$$

whose coefficients are determined by the initial condition $P(x, 0) = P_0(x)$. By the orthogonality property, we have

$$d_n = \frac{\mu^n}{n!} \sum_{x=0}^{\infty} C_n(x, \mu) P_0(x)\ . \tag{3.80}$$

If we are interested in the transition probability $P(x, t; x_0, 0)$ again, then we choose $P_0(x) = \delta_{x,x_0}$ and find that

$$P(x, t; x_0, 0) = \sum_{n=0}^{\infty} \frac{\mu^n}{n!} C_n(x_0, \mu) C_n(x, \mu) P_{ss}(x) e^{-\gamma n t}\ . \tag{3.81}$$

While this formula looks superficially different from our earlier solution (Eq. 3.44), one can use properties of Charlier polynomials to show that they are actually the same.

It seems like a fairly generic feature of the CME that the eigenfunctions of H will be discrete orthogonal polynomials times the steady state distribution. One also generally expects that the eigenvalues λ_n of H (or equivalently, the allowed energies $E_n = -\lambda_n$) are determined by the characteristic time scales of the problem—the degradation rate time scale

fixed by γ, in this case. But in general, the eigenvalues need not be equally spaced as they were here. In analyzing simple examples of bimolecular reaction networks, one quickly finds examples of systems with $E_n \sim n^2$.

Again, we should emphasize that we would have been severely handicapped had we not been able to identify the functions Q_n as the Charlier polynomials and been able to draw on the associated body of special functions knowledge. For a more complicated problem, whose answer does not yield a known family of orthogonal polynomials, one must prove all of their necessary properties from scratch in order to work with them. For example, it is a nontrivial task to derive the orthogonal polynomial properties necessary to show that our eigenfunction expansion agrees with Eq. 3.44.

3.5   Exotic analytic approaches to solving the CME

In the previous section, we described several typical analytic approaches to solving the CME. In this section—and the rest of this thesis—we will explore the use of a few 'exotic' ones, inspired by theoretical tools from quantum mechanics, that have various appealing features.

3.5.1   Path integrals

One of our primary tools will be the *path integral*, which expresses probability distributions $P(x,t)$ in terms of an infinite number of integrals. While this seems like it would not making solving the problem any easier, there are various tricks for evaluating these integrals, and there is one important advantage over a number of other methods: supposing one *can* evaluate these integrals, solving for $P(x,t)$ becomes relatively mechanical. Unlike in the case of eigenfunction expansion, for example, we do not need any a priori special functions knowledge to help us along. We only have to compute (usually relatively simple) integrals.

In this thesis, we will actually encounter several different kinds of path integrals: *state space* path integrals, and *coherent state* path integrals. State space path integrals express $P(x,t)$ in terms of all possible paths through state space, while coherent state path integrals express $P(x,t)$ as a superposition of distributions in some reference class. In the case of the CME, it is useful to try expressing $P(x,t)$ as a superposition of infinitely many Poisson distributions. This is the essence of the Doi-Peliti path integral approach, which is equivalent to the Poisson representation approach we described earlier. Of course, we need not restrict ourselves to Poisson distributions only; perhaps other distributions are better suited to certain problems.

Let us briefly review the idea of the (state space) path integral from quantum mechanics. The coherent state path integral from quantum mechanics, used to do things like write down a path integral description of spins, is similar in spirit if not in detail, so we will not discuss it here.

For simplicity, consider a single particle in one dimension subject to a (time and momentum-independent) potential V on the infinite line $(-\infty, \infty)$. The Schrödinger equation reads

$$i\hbar\frac{\partial\psi(x,t)}{\partial t} = \hat{H}\psi(x,t)$$

$$= \left[\frac{\hat{p}^2}{2m} + V(\hat{x})\right]\psi(x,t) \tag{3.82}$$

$$= -\frac{\hbar^2}{2m}\frac{\partial^2\psi(x,t)}{\partial x^2} + V(x)\psi(x,t) \ ,$$

where $\hat{H}$ is the Hamiltonian operator, $\hat{p}$ is the momentum operator, and $\hat{x}$ is the position operator. In Dirac's bra-ket notation, the Schrödinger equation reads

$$i\hbar\frac{\partial}{\partial t}\left|\psi\right\rangle = \hat{H}\left|\psi\right\rangle \ , \tag{3.83}$$

which can be formally solved via

$$\left|\psi(t_f)\right\rangle = e^{-\frac{i}{\hbar}\hat{H}(t_f-t_0)}\left|\psi(t_0)\right\rangle \tag{3.84}$$

where $\left|\psi(t_0)\right\rangle$ is the system's state at initial time t_0, and $\left|\psi(t_f)\right\rangle$ is the system's state at a later time t_f. To construct the path integral, two 'resolutions of the identity' are required:

$$1 = \int_{-\infty}^{\infty} dx \ \left|x\right\rangle\left\langle x\right| \tag{3.85}$$

$$1 = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dp \ \left|p\right\rangle\left\langle p\right| \ . \tag{3.86}$$

The formal solution can be rewritten suggestively by inserting many position eigenket resolutions of the identity (Eq. 3.85):

$$\left|\psi(t_f)\right\rangle = e^{-\frac{i}{\hbar}\hat{H}\Delta t}e^{-\frac{i}{\hbar}\hat{H}\Delta t}\cdots e^{-\frac{i}{\hbar}\hat{H}\Delta t}\left|\psi(t_0)\right\rangle$$

$$= \int dx_0 dx_1\cdots dx_N \left|x_N\right\rangle \tag{3.87}$$

$$\times \left\langle x_N\right|e^{-\frac{i}{\hbar}\hat{H}\Delta t}\left|x_{N-1}\right\rangle\left\langle x_{N-1}\right|e^{-\frac{i}{\hbar}\hat{H}\Delta t}\left|x_{N-2}\right\rangle$$

$$\times \cdots \left\langle x_1\right|e^{-\frac{i}{\hbar}\hat{H}\Delta t}\left|x_0\right\rangle\left\langle x_0|\psi(t_0)\right\rangle \ ,$$

where $\Delta t = (t_f - t_0)/N$. Then each small time propagator matrix element can be evaluated with the help of the momentum eigenket resolution of the identity (Eq. 3.86). The key part of the calculation is evaluating matrix elements like these:

$$\left\langle x_{j+1}|\hat{p}^2|x_j\right\rangle = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dp \ \left\langle x_{j+1}|\hat{p}^2|p\right\rangle\left\langle p|x_j\right\rangle$$

$$= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dp \ p^2 \left\langle x_{j+1}|p\right\rangle\left\langle p|x_j\right\rangle \tag{3.88}$$

$$= \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dp \ p^2 e^{\frac{i}{\hbar}p(x_{j+1}-x_j)} \ .$$

40

The rest of the calculation is fairly straightforward. If we take $|\psi(t_0)\rangle = |x_0\rangle$ for some $x_0 \in \mathbb{R}$, then we obtain a phase space path integral representation of $\psi(x, t_f; x_0, t_0) := \langle x|\psi(t_f)\rangle$ that reads

$$\psi = \lim_{N\to\infty} \int \frac{dx_1 dp_1}{2\pi\hbar} \cdots \frac{dx_{N-1} dp_{N-1}}{2\pi\hbar} \frac{dp_N}{2\pi\hbar} \exp\left\{ \frac{i}{\hbar} \sum_{j=1}^{N} p_j(x_j - x_{j-1}) - \left[\frac{p_j^2}{2m} + V(x_{j-1})\right] \Delta t \right\}$$
(3.89)

where $x_N := x$. This path integral is often written in the schematic form [10]

$$\psi(x, t_f; x_0, t_0) = \int \mathcal{D}[x(t)]\mathcal{D}[p(t)] \ \exp\left\{\frac{i}{\hbar} S[x, p]\right\}$$

$$S[x, p] = \int_{t_0}^{t_f} dt \ p(t)\dot{x}(t) - \frac{p(t)^2}{2m} - V(x(t)) \ .$$
(3.90)

One can integrate out the momentum variables in Eq. 3.89 to arrive at the usual configuration space path integral

$$\psi(x, t_f; x_0, t_0) = \lim_{N\to\infty} \int dx_1 \cdots dx_{N-1} \left(\frac{m}{2\pi i\hbar\Delta t}\right)^{N/2} \exp\left\{\frac{i}{\hbar}\Delta t \sum_{j=1}^{N} \frac{m}{2}\left(\frac{x_j - x_{j-1}}{\Delta t}\right)^2 - V(x_{j-1})\right\}$$
(3.91)

which has schematic form [11]

$$\psi(x, t_f; x_0, t_0) = \int \mathcal{D}[x(t)] \ \exp\left\{\frac{i}{\hbar} S[x(t)]\right\}$$

$$S[x] = \int_{t_0}^{t_f} dt \ \frac{1}{2}m\dot{x}(t)^2 - V(x(t)) \ .$$
(3.92)

The interpretation is that one is summing over all paths satisfying $x(t_0) = x_0$ and $x(t_f) = x$. For exact calculations, the discrete form of the phase space path integral (Eq. 3.89) is most useful. For example, let us try to solve for the propagator $\psi := \psi(x, t; x_0, t_0)$ of a free particle, which has $V = 0$. The path integral reads

$$\psi = \lim_{N\to\infty} \int \frac{dx_1 dp_1}{2\pi\hbar} \cdots \frac{dx_{N-1} dp_{N-1}}{2\pi\hbar} \frac{dp_N}{2\pi\hbar} \exp\left\{\frac{i}{\hbar} \sum_{j=1}^{N} p_j(x_j - x_{j-1}) - \frac{\Delta t}{2m} p_j^2\right\} \ .$$
(3.93)

We should do the integrals over the position variables x_j first; if we integrated out the momentum variables first, we would just recover the configuration space path integral. And as a general rule, configuration space path integral calculations are somewhat harder. Collecting terms involving x_j, we have integrals

$$\int_{-\infty}^{\infty} \frac{dx_j}{2\pi\hbar} e^{\frac{i}{\hbar}x_j(p_j - p_{j+1})} = \delta(p_j - p_{j+1})$$
(3.94)

41

for $j = 1, ..., N - 1$. Enforcing the delta function constraints, we have

$$\psi = \lim_{N \to \infty} \int_{-\infty}^{\infty} \frac{dp_N}{2\pi\hbar} \exp\left\{ \frac{i}{\hbar} p_N(x - x_0) - \frac{(t - t_0)}{2m} p_N^2 \right\} . \tag{3.95}$$

where we can drop the limit since N is now just a label. Doing this simple Gaussian integral, we obtain

$$\psi(x, t; x_0, t_0) = \sqrt{\frac{m}{2\pi\hbar i(t - t_0)}} \exp\left\{ \frac{im(x - x_0)^2}{2\hbar(t - t_0)} \right\} . \tag{3.96}$$

In general, path integrals that can be reduced to Gaussian integrals (like this one, and the harmonic oscillator) are easy to solve; problems involving more complicated functions may be solvable, but require various tricks.

Derivations of path integral representations of the CME analogous to the quantum mechanics derivation above are given in Chapter 4 and Chapter 5. See the appendix to Chapter 5 for elementary CME path integral calculations along the lines of the free particle problem we just solved. See Chapter 4 and Chapter 7 for detailed path integral derivations of the solutions of various CMEs involving monomolecular processes.

3.5.2 Ladder operators

Consider the Schrödinger equation of the one-dimensional harmonic oscillator, which in Hilbert space reads (in units where $\hbar = m = 1$)

$$i\frac{\partial}{\partial t} |\psi\rangle = \hat{H} |\psi\rangle = \left[\frac{\hat{p}^2}{2} + \frac{1}{2}\omega^2 \hat{x}^2 \right] |\psi\rangle . \tag{3.97}$$

The idea behind a ladder operator solution [12] to this problem is to try and 'factor' the Hamiltonian. This can be done by defining the operators

$$\hat{A} := \sqrt{\frac{\omega}{2}} \left(\hat{x} + \frac{i}{\omega}\hat{p} \right)$$
$$\hat{A}^\dagger := \sqrt{\frac{\omega}{2}} \left(\hat{x} - \frac{i}{\omega}\hat{p} \right) \tag{3.98}$$

which can be shown to satisfy the canonical commutation relation $[\hat{A}, \hat{A}^\dagger] = 1$. In terms of them, it turns out that we can write the Hamiltonian operator $\hat{H}$ as

$$\hat{H} = \omega \hat{A}^\dagger \hat{A} + \frac{\omega}{2} . \tag{3.99}$$

This decomposition has various extremely useful consequences—and in some sense, the oscillator is 'solved' once one identifies these operators, and how to write the Hamiltonian in terms of them. In particular, we immediately have that

$$[\hat{H}, \hat{A}] = -\omega\hat{A}$$
$$[\hat{H}, \hat{A}^\dagger] = \omega\hat{A}^\dagger \tag{3.100}$$

42

which can be used to determine the eigenvalues and eigenstates of $\hat{H}$. In particular, they can be used to show that if $|\psi_E\rangle$ is an eigenstate with energy E, then $\hat{A}^\dagger |\psi_E\rangle$ is an eigenstate with energy $E + \omega$, since

$$
\begin{aligned}
\hat{H}\hat{A}^\dagger |\psi_E\rangle &= \left\{ \hat{A}^\dagger \hat{H} + [\hat{H}, \hat{A}^\dagger] \right\} |\psi_E\rangle \\
&= (E + \omega)\hat{A}^\dagger |\psi_E\rangle .
\end{aligned}
\tag{3.101}
$$

Similarly, $\hat{A} |\psi_E\rangle$ is an eigenstate of $\hat{H}$ with energy $E - \omega$ when $|\psi_E\rangle$ is an eigenstate with energy E. This means that, given one eigenstate, we can use the ladder operators to construct many other eigenstates with different energies. Unless there is ' ground state' $|\psi_0\rangle$ for which

$$
\hat{A} |\psi_0\rangle = 0
\tag{3.102}
$$

we would be able to use the lowering operator $\hat{A}$ to create eigenstates of arbitrarily negative energy, which is not physical. This is also true if there are eigenstates that are 'off' the ladder, in the sense of not being of the form $(\hat{A}^\dagger)^n |\psi_0\rangle$ for some $n \in \mathbb{N}$ (since the ground state turns out to be unique). By a routine computation, we can find that the energy of the ground state is given by $E_0 = \frac{\omega}{2}$. Hence, the eigenvalues and eigenstates of $\hat{H}$ are given by

$$
\begin{aligned}
E_n &= \omega \left(n + \frac{1}{2} \right) \\
|\psi_n\rangle &= (\hat{A}^\dagger)^n |\psi_0\rangle .
\end{aligned}
\tag{3.103}
$$

From here, one can write the full solution for $|\psi(t)\rangle$ by writing it as an expansion in the eigenkets, computing coefficients by exploiting their orthonormality. To get the position space wave function, we could identify the position space ground state as a Gaussian, and thus find that the eigenfunctions are Hermite polynomials times a Gaussian. Even aside from just getting the solution, the ladder operators turn out to be useful for deriving various properties of it; for example, they allow the computation of moments, which normally can involve tedious integration, into simple algebra.

The availability of ladder operators as a solution technique depends highly on the structure of the problem. It is most suited when the eigenvalues of the Hamiltonian are linearly spaced. Although there are not many quantum mechanical problems that satisfy this property, it will turn out that there *are* many chemical kinetics problems that do. These include the monomolecular processes that we first study in Chapter 4, and many processes involving only first order reactions. See Chapter 6 for the application of ladder operators to solving CME models of RNA transcription and splicing. The extent to which a ladder operator approach may prove useful in exactly solving very complicated CMEs is not yet known.

Bibliography

[1] Daniel T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4):403 – 434, 1976.

[2] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81(25):2340–2361, 1977.

[3] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry, 58(1):35–55, 2007.

[4] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. The Journal of Chemical Physics, 124(4):044104, 2006.

[5] Slaven Peleš, Brian Munsky, and Mustafa Khammash. Reduction and solution of the chemical master equation using time scale separation and finite state projection. The Journal of Chemical Physics, 125(20):204104, 2006.

[6] Zachary Fox, Gregor Neuert, and Brian Munsky. Finite state projection based bounds to compare chemical master equation models using single-cell data. The Journal of Chemical Physics, 145(7):074101, 2016.

[7] C. W. Gardiner and S. Chaturvedi. The poisson representation. i. a new technique for chemical master equations. Journal of Statistical Physics, 17(6):429–468, Dec 1977.

[8] S. Chaturvedi and C. W. Gardiner. The poisson representation. ii two-time correlation functions. Journal of Statistical Physics, 18(5):501–522, May 1978.

[9] C. Gardiner. Stochastic Methods: A Handbook for the Natural and Social Sciences. Springer Series in Synergetics. Springer Berlin Heidelberg, 2009.

[10] R. Shankar. Principles of Quantum Mechanics. Springer US, 2nd edition, 2011.

[11] R.P. Feynman, A.R. Hibbs, and D.F. Styer. Quantum Mechanics and Path Integrals. Dover Books on Physics. Dover Publications, 2010.

[12] D.J. Griffiths and D.F. Schroeter. Introduction to Quantum Mechanics. Cambridge University Press, 2018.

Chapter 4

Path integral representations of the chemical master equation.
I: The Doi-Peliti coherent state path integral

How can we *solve* the CME? This chapter introduces the Doi-Peliti path integral technique as a fairly general solution method, and uses it to derive the most general known solution of the CME. This is the first of two chapters that introduce path integral methods; this method was previously known, while next chapter's method is new. Although this is not discussed in detail, it turns out that they are 'equally powerful' in some sense, and that the results in this chapter can also be derived using the other path integral method.

Abstract: The chemical master equation (CME) is a fundamental description of interacting molecules commonly used to model chemical kinetics and noisy gene regulatory networks. Exact time-dependent solutions of the CME—which typically consists of infinitely many coupled differential equations—are rare, and are valuable for numerical benchmarking and getting intuition for the behavior of more complicated systems. Jahnke and Huisinga's landmark calculation of the exact time-dependent solution of the CME for monomolecular reaction systems is one of the most general analytic results known; however, it is hard to generalize, because it relies crucially on properties of monomolecular reactions. In this paper, we rederive Jahnke and Huisinga's result on the time-dependent probability distribution and moments of monomolecular reaction systems using the Doi-Peliti path integral approach, which reduces solving the CME to evaluating many integrals. While the Doi-Peliti approach is less intuitive, it is also more mechanical, and hence easier to generalize. To illustrate how the Doi-Peliti approach can go beyond the method of Jahnke and Huisinga, we also find an explicit and exact time-dependent solution to a problem involving an autocatalytic reaction that Jahnke and Huisinga identified as not solvable using their method. We also find a formal exact time-dependent solution for any CME whose list of reactions involves only zero and first order reactions, which may be the most general result currently known.

4.1 Introduction

The chemical master equation (CME) provides a fundamental description of well-mixed molecules interacting with each other via a set of chemical reactions [1, 2, 3, 4, 5, 6, 7]. It models dynamics that are discrete (the state of the system is a set of nonnegative integers) and stochastic (chemical reactions occur with some probability). The CME has recently enjoyed tremendous success as a framework for understanding noisy single cell data [8, 9, 10, 11, 12, 13, 14], particularly in simple model organisms like yeast where techniques like single-molecule Fluorescence *in situ* Hybridization (smFISH) allow RNA molecule numbers to be counted accurately [15, 16, 17]. Outside of cell and molecular biology, master equations have been successfully used to model population dynamics [18, 19, 20], traffic [21, 22, 23], and gas phase chemical kinetics [24, 25, 26], among other things.

Although it is very useful for defining discrete stochastic models, the CME generally cannot be solved directly. One typically resorts to an approximate approach, like using Gillespie's algorithm [27, 28] to extract information from many brute force simulations, or using finite state projection [29, 30, 10], or partitioning the system (e.g. low versus high copy number, slow versus fast time scale) [31, 32, 33, 34, 35, 36], or solving a continuous approximation to the CME like the chemical Langevin equation [3, 37, 38, 39].

Unsurprisingly, exact time-dependent solutions (as opposed to steady state solutions) of the CME are particularly rare, and have only been computed for specific cases. McQuarrie [1] describes some of the early attempts: in 1940, Max Delbrück evaluated the CME for the autocatalytic reaction $S \to S+S$ [40]; in 1954, Renyi solved the binding reaction $A+B \to C$ [41]; in 1960, Ishida solved the death reaction $S \to \varnothing$ and presented the first CME solution with time-dependent rates [42]; in 1963 and 1964, McQuarrie et al. solved many simple systems (including $A + A \to B$ and $A + B \to C$) using the method of generating functions [43, 44].

The situation did not change appreciably until Jahnke and Huisinga's landmark paper [45], more than forty years later. Their 2007 paper constituted a major advance in our collective understanding of the CME; they were able to solve the CME for a system with an *arbitrary* number of species experiencing an *arbitrary* number of reactions whose rates have *arbitrary* time-dependence, *provided that* the reactions consisted of some combination of birth ($\varnothing \to S_k$), death ($S_j \to \varnothing$), and conversion ($S_j \to S_k$)[1]. The shocking generality of their result, as well as the explicitness of the solution they wrote down (in Theorem 1 of that paper), was powerful.

Since 2007, there have been few new results of the same generality. Reis et al. [46] extend Jahnke and Huisinga's result by considering hierarchical first-order reaction networks (which allow a certain subset of first-order reactions that is strictly larger than the set of monomolecular reactions). However, there has not been (for example) any result on the solution to general first-order reactions, or general bimolecular reactions. At present, even finding the exact solutions of simple systems that involve bimolecular reactions is nontrivial:

---

[1]These systems are called "monomolecular" because all allowed reactions have at most one molecule as input, and at most one molecule as output.

the work of Laurenzi ($A + B \leftrightarrow C$) [47], as well as Arslan and Laurenzi ($A + B \leftrightarrow A + A$) [48] are two examples.

One drawback of Jahnke and Huisinga's paper is that it essentially relied on guessing the solution. It was well-known that Poisson and multinomial distributions solved the CME in special cases, and that these distributions had certain desirable properties (e.g. a Poisson distribution stays a Poisson distribution, and a multinomial distribution stays a multinomial distribution; see Sec. 3 of their paper). To derive their Theorem 1, these properties were exploited, along with the fact that only monomolecular reactions were considered. Of course, their method completely breaks down for a system that is only slightly more complicated; as they point out in Sec. 6, adding an autocatalytic reaction $S \rightarrow S + S$ to a system they can easily solve manages to make it beyond the scope of their results.

Hence, it would be nice if there was a method to obtain their classic result that did not rely on systematic guessing. In this paper, we offer the Doi-Peliti path integral approach to solving the CME as one such method. The Doi-Peliti approach allows one to 'turn the crank', so to speak, and generate a time-dependent solution of the CME through a straightforward but difficult calculation. Importantly, it is system-agnostic: one does not need to know properties like 'Poisson distributions stay Poisson', or assume the solution takes a certain form.

Doi-Peliti field theory—which emerged from the pioneering papers of researchers like Doi [49, 50], Peliti [51, 52, 53], and Grassberger [54, 55, 56, 57]—reframes solving the CME as a field theory problem. This enables the use of powerful approximation schemes, like the renormalization group and diagrammatic perturbation theory [58, 59, 60, 61, 62, 63, 64, 65]. While Doi-Peliti field theory is still somewhat obscure in mathematical biology, it has seen the occasional application: e.g. to understand population dynamics given colored noise [66], age dependent branching processes [67, 68], and large deviations in gene regulatory networks [69, 20]. Although not Doi-Peliti, a qualitatively similar path integral has been used to solve the CME for a multistep transcription and translation process [70].

We will use Doi-Peliti field theory to rederive Jahnke and Huisinga's Theorem 1. Moreover, in order to show that the Doi-Peliti path integral approach is strictly *more* powerful than the one used by Jahnke and Huisinga, we use it to exactly solve a problem they said their method could not, as well as a far more general problem (the CME whose list of reactions consists of any combination of zero and first order reactions). We solve these additional problems in complete generality, and obtain exact time-dependent solutions assuming rates with arbitrary time-dependence.

The paper is organized as follows. In Sec. 4.2, we state the problems we will solve, as well as our main results on their solutions. This includes reviewing the monomolecular CME and Jahnke and Huisinga's solution of it. In Sec. 4.3, we review the generating function formulation of the CME, and develop the basic machinery of the Doi-Peliti approach to solving the differential equation satisfied by the generating function. This includes introducing several important concepts (including coherent states and two inner products) and constructing the Doi-Peliti path integral. We also describe how information like transition probabilities, moments, and (the usual) generating functions can be extracted from the (bra-

ket) generating function. Sec. 4.4 through 4.6 contain the Doi-Peliti calculations that back up our main results, with each section focusing on one of the problems introduced in Sec. 4.2: (i) monomolecular systems, (ii) one species birth-death-autocatalysis, and (iii) arbitrary combinations of zero and first order reactions. Sec. 4.7 describes an alternative view of the propagator, the central object to be calculated in the Doi-Peliti approach; in particular, we show that our results can also be derived by solving a certain partial differential equation (PDE) using the method of characteristics. Finally, in Sec. 7.6, we discuss the merits and drawbacks of the Doi-Peliti approach to solving the CME, and speculate on how it could be further utilized.

4.2   Problem statements and main results

In this section, we introduce in detail the specific problems we will solve, and we present our main results regarding their solutions. If the reader is *only* interested in the answers, and not how they were obtained, then in some sense the rest of the paper is irrelevant; the validity of any given solution can usually be checked by tedious but straightforward substitution into the CME.

We present results for several systems, in order of increasing complexity: the chemical birth-death process, monomolecular reactions, single species birth-death-autocatalysis, and arbitrary combinations of zero and first order reactions with arbitrarily many species. While the chemical birth-death process is a kind of monomolecular reaction system, we include it here to give mathematicians new to the CME a toy example (that can be stated with minimal notational baggage) of the kind of results we are seeking.

4.2.1   The chemical birth-death process as a prototype

The chemical birth-death process is simple enough to be biologically relevant (it can be used to model how the number of a single type of mRNA or protein in a single cell changes stochastically with time in the absence of significant regulation [6, 7, 71]), but complicated enough to have nontrivial dynamics (the number of molecules does not increase without bound or shrink to zero in the long time limit, allowing there to exist a Poisson-like steady state probability distribution). It is linear (in several distinct but related senses of the word), allowing its associated CME to be exactly solved via a variety of methods (e.g. separation of variables and ladder operators [72], and via a path integral different from the one we will discuss here [39]).

It is characterized by the chemical reactions

$$\varnothing \xrightarrow{k} S$$
$$S \xrightarrow{\gamma} \varnothing$$

(4.1)

where $k(t)$ and $\gamma(t)$ parameterize the small time birth and death rates, respectively[2]. The time-dependence of these parameters is allowed to be arbitrary as long as they both remain nonnegative for all times. The corresponding CME reads

$$\frac{\partial P(x,t)}{\partial t} = k(t)\left[P(x-1,t) - P(x,t)\right]$$
$$+ \gamma(t)\left[(x+1)P(x+1,t) - xP(x,t)\right] \tag{4.2}$$

where $P(x,t)$ is the probability that the state of the system is $x \in \mathbb{N} := \{0,1,2,...\}$ at time $t \geq t_0$. We are particularly interested in the transition probability $P(x,t;\xi,t_0)$, i.e. the solution of Eq. 4.2 whose probability distribution is initially certain ($P(x,t_0) = \delta(x - \xi)$ for some $\xi \in \mathbb{N}$); if the transition probability is known, the solution to Eq. 4.2 for an arbitrary initial distribution $P_0(x)$ can be written

$$P(x,t) = \sum_{y=0}^{\infty} P(x,t;y,t_0)P_0(y) \ . \tag{4.3}$$

In practice, we are also interested in several other properties of the solution. In particular, we are interested in the long time behavior described by the steady state probability distribution

$$P_{ss}(x) := \lim_{t\to\infty} P(x,t;\xi,t_0) \ , \tag{4.4}$$

moments like

$$\langle x(t) \rangle := \sum_{x=0}^{\infty} xP(x,t)$$
$$\langle x(t)[x(t) - 1] \rangle := \sum_{x=0}^{\infty} x(x-1)P(x,t) \ , \tag{4.5}$$

and the real or complex-valued probability generating function

$$\psi(g,t) := \sum_{x=0}^{\infty} P(x,t) \ g^x \ , \tag{4.6}$$

whose derivatives correspond to various moments of interest. In fact, knowing the generating function is equivalent to knowing $P(x,t)$, and its time evolution is described by a PDE analogous to the CME:

$$\frac{\partial \psi(g,t)}{\partial t} = k(t)[g-1]\psi(g,t) - \gamma(t)[g-1]\frac{\partial \psi(g,t)}{\partial g} \ . \tag{4.7}$$

---

[2]In particular, if the system has x molecules of species S at time t, the probability that the birth reaction happens in a window of time $[t, t+\Delta t)$ is approximately $k(t)\Delta t$, and the probability that the death reaction happens is approximately $\gamma(t)x\Delta t$. See Gillespie [3] for more details on the interpretation of and formalism underlying the CME.

The initial condition corresponding to $P(x, t_0) = \delta(x - \xi)$ is $\psi(g, t_0) = g^\xi$, as can be verified using the definition of ψ. We can also take its long time limit, which we will denote by $\psi_{ss}(g)$.

The main result for the chemical birth-death process is the following.

Theorem 1 (Chemical birth-death process). *Let $\lambda(t)$ and $w(t)$ be the solutions of*

$$\dot{\lambda} = k - \gamma\lambda \ , \ \ \lambda(t_0) = 0$$
$$\dot{w} = -\gamma w \ , \ \ w(t_0) = 1 \ . \tag{4.8}$$

Then if $P(x, t_0) = \delta(x - \xi)$ for some $\xi \in \mathbb{N}$, we have:

i.

$$P(x, t; \xi, t_0) = \sum_{k=0}^{\min(x,\xi)} \frac{\lambda(t)^{x-k} e^{-\lambda(t)}}{(x - k)!} \binom{\xi}{k} w(t)^k [1 - w(t)]^{\xi-k} \tag{4.9}$$

ii.

$$\langle x(t) \rangle = \xi w(t) + \lambda(t) \tag{4.10}$$

iii.

$$\langle x(t)[x(t) - 1] \rangle = w(t)^2 \xi(\xi - 1) + 2\lambda(t)w(t)\xi + \lambda(t)^2 \tag{4.11}$$

iv.

$$\psi(g, t) = [1 + (g - 1)w(t)]^\xi \, e^{(g-1)\lambda(t)} \tag{4.12}$$

Proof. All results can be obtained independently of one another using the Doi-Peliti approach described in the following sections. Alternatively, one can verify directly that $P(x, t; \xi, t_0)$ satisfies Eq. 4.2 or that $\psi(g, t)$ satisfies Eq. 4.7, and then obtain the rest of the results by brute force calculation. □

These results simplify tremendously in the long time limit if k and γ are time-independent, essentially because $P(x, t)$ reduces to a Poisson distribution regardless of one's choice of initial distribution $P(x, t_0)$.

Corollary 1 (Long time behavior of chemical birth-death process). *Let k and γ be time-independent, and define $\mu := k/\gamma$. In the long time limit, we have:*

i.

$$P_{ss}(x) = \frac{\mu^x e^{-\mu}}{x!} \tag{4.13}$$

ii.

$$\langle x \rangle = \mu \tag{4.14}$$

iii.

$$\langle x[x - 1] \rangle = \mu^2 \tag{4.15}$$

iv.

$$\psi_{ss}(g) = e^{(g-1)\mu} \tag{4.16}$$

Proof. Take the $t \to \infty$ limit of the previous results, noting that $w(t) \to 0$ and $\lambda(t) \to \mu$. $\quad\square$

Results on the chemical birth-death process are far from new. Still, we ask the reader to keep these results in the back of their mind as we go on to discuss the analogous results for more complicated systems. Because the chemical birth-death process is so fundamental, more complicated results often reduce to these in the appropriate limit.

4.2.2 Monomolecular results

Heuristically, monomolecular reaction systems are the minimal multi-species generalization of the chemical birth-death process. Like the chemical birth-death process, they exhibit a certain kind of linearity; we will see that their solutions are completely determined by a system of linear ordinary differential equations (ODEs).

Let us define them. Consider a system with n chemical species $S_1, ..., S_n$, whose reaction list reads

$$
\begin{aligned}
S_j &\xrightarrow{c_{jk}} S_k & j &\neq k \\
\varnothing &\xrightarrow{c_{0k}} S_k & k &= 1, ..., n \\
S_j &\xrightarrow{c_{j0}} \varnothing & j &= 1, ..., n
\end{aligned}
\tag{4.17}
$$
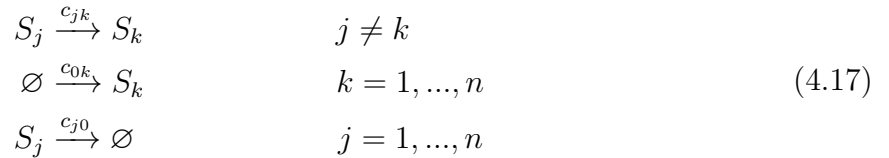
i.e. all possible monomolecular reactions (birth, death, and conversion) are allowed[3]. Note that the rates are allowed to have arbitrary time-dependence as long as $c_{jk}(t) \geq 0$ for all j, k and all times $t \geq t_0$. The corresponding CME reads

$$
\begin{aligned}
\frac{\partial P(\mathbf{x}, t)}{\partial t} =& \sum_{k=1}^{n} c_{0k}(t) \left[P(\mathbf{x} - \boldsymbol{\epsilon}_k, t) - P(\mathbf{x}, t) \right] \\
&+ \sum_{k=1}^{n} c_{k0}(t) \left[(x_k + 1) P(\mathbf{x} + \boldsymbol{\epsilon}_k, t) - x_k P(\mathbf{x}, t) \right] \\
&+ \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[(x_j + 1) P(\mathbf{x} + \boldsymbol{\epsilon}_j - \boldsymbol{\epsilon}_k, t) - x_j P(\mathbf{x}, t) \right]
\end{aligned}
\tag{4.18}
$$

where $P(\mathbf{x}, t)$ is the probability that the state of the system is $\mathbf{x} := (x_1, ..., x_n)^T \in \mathbb{N}^n$ at time $t \geq t_0$, and where $\boldsymbol{\epsilon}_k$ is the n-dimensional vector with a 1 in the kth place and zeros everywhere else.

The exact solution to Eq. 4.18, given the initial condition $P(\mathbf{x}, t_0) = \delta(\mathbf{x} - \boldsymbol{\xi})$ for some vector $\boldsymbol{\xi} := (\xi_1, ..., \xi_n)^T \in \mathbb{N}^n$, is reported in Theorem 1 of Jahnke and Huisinga [45]. In order to state their solution, we will need some notation.

---

[3]The one exception is the trivial conversion reaction $S_k \to S_k$, which is disallowed because including it would be pointless. To ease notation (i.e. to avoid writing $j \neq k$ many times), we follow Jahnke and Huisinga and define $c_{kk} := 0$ for all $k = 1, ..., n$.

Define the matrix $\mathbf{A}(t)$ and vector $\mathbf{b}(t)$ by

$$A_{jk}(t) := c_{kj}(t) \quad \text{for } j \neq k \geq 1$$

$$A_{kk}(t) := -\sum_{j=0}^{n} c_{kj}(t) \quad \text{for } 1 \leq k \leq n \tag{4.19}$$

$$\mathbf{b}(t) := \begin{pmatrix} c_{01}(t) & c_{02}(t) & \cdots & c_{0n}(t) \end{pmatrix}^{T} .$$

The deterministic reaction rate equations corresponding to our reaction list can be written in terms of $\mathbf{A}(t)$ and $\mathbf{b}(t)$ as

$$\dot{\mathbf{x}} = \mathbf{A}(t)\mathbf{x} + \mathbf{b}(t) . \tag{4.20}$$

Because Eq. 4.20 is linear, the solution with initial condition $\boldsymbol{\xi} = (\xi_1, ..., \xi_n)$ can be written as

$$\mathbf{x}(t) = \sum_{k=1}^{n} \xi_k \mathbf{w}^{(k)}(t) + \boldsymbol{\lambda}(t) \tag{4.21}$$

where the vectors $\mathbf{w}^{(1)}(t), ..., \mathbf{w}^{(n)}(t)$ and $\boldsymbol{\lambda}(t)$ are defined as

$$\dot{\mathbf{w}}^{(k)} = \mathbf{A}(t)\mathbf{w}^{(k)} , \quad \mathbf{w}^{(k)}(t_0) = \boldsymbol{\epsilon}_k$$

$$\dot{\boldsymbol{\lambda}} = \mathbf{A}(t)\boldsymbol{\lambda} + \mathbf{b}(t) , \quad \boldsymbol{\lambda}(t_0) = \mathbf{0} . \tag{4.22}$$

As we will shortly observe, the solution to the deterministic reaction rate equations is intimately related to the solution of the CME (at least for monomolecular reactions).

Now define the 1-norm of a vector $\mathbf{x}$ as

$$|\mathbf{x}| := \sum_{k=1}^{n} |x_k| . \tag{4.23}$$

Define, because they will appear throughout this paper, multi-dimensional generalizations of powers, factorials, sums, integrals, and derivatives:

$$\mathbf{v}^{\mathbf{x}} := v_1^{x_1} \cdots v_n^{x_n}$$

$$\mathbf{x}! := x_1! \cdots x_n!$$

$$\sum_{\mathbf{x}} := \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \tag{4.24}$$

$$\int d\mathbf{x} := \int dx_1 \cdots \int dx_n$$

$$\left(\frac{d}{d\mathbf{z}}\right)^{\mathbf{x}} := \left(\frac{d}{dz_1}\right)^{x_1} \cdots \left(\frac{d}{dz_n}\right)^{x_n} .$$

Using the above shorthand, we can define the product Poisson distribution as

$$\mathcal{P}(\mathbf{x}, \boldsymbol{\lambda}) := \frac{\lambda_1^{x_1}}{x_1!} \cdots \frac{\lambda_n^{x_n}}{x_n!} e^{-|\boldsymbol{\lambda}|} = \frac{\boldsymbol{\lambda}^{\mathbf{x}}}{\mathbf{x}!} e^{-|\boldsymbol{\lambda}|} , \tag{4.25}$$

the multinomial distribution as

$$
\begin{aligned}
\mathcal{M}(\mathbf{x}, N, \mathbf{w}) &:= \frac{N! \; [1 - |\mathbf{w}|]^{N - |\mathbf{x}|}}{(N - |\mathbf{x}|)!} \frac{w_1^{x_1}}{x_1!} \cdots \frac{w_n^{x_n}}{x_n!} \quad \text{if } |\mathbf{x}| \le N \;\; \text{and} \;\; x \in \mathbb{N}^N \\
&= \frac{N! \; [1 - |\mathbf{w}|]^{N - |\mathbf{x}|}}{(N - |\mathbf{x}|)!} \frac{\mathbf{w}^{\mathbf{x}}}{\mathbf{x}!} \quad \text{if } |\mathbf{x}| \le N \;\; \text{and} \;\; x \in \mathbb{N}^N \;,
\end{aligned}
\tag{4.26}
$$

and the convolution of two probability distributions as

$$
P_1(\mathbf{x}) \star P_2(\mathbf{x}) := \sum_{\mathbf{z}} P_1(\mathbf{z}) P_2(\mathbf{x} - \mathbf{z}) = \sum_{\mathbf{z}} P_1(\mathbf{x} - \mathbf{z}) P_2(\mathbf{z})
\tag{4.27}
$$

where the sum is over all $\mathbf{z} \in \mathbb{N}^n$ such that $\mathbf{x} - \mathbf{z} \in \mathbb{N}^n$. As in the one-dimensional case, we can also define the probability generating function

$$
\psi(\mathbf{g}, t) := \sum_{\mathbf{x}} P(\mathbf{x}, t) \, \mathbf{g}^{\mathbf{x}}
\tag{4.28}
$$

which satisfies the PDE

$$
\begin{aligned}
\frac{\partial \psi(\mathbf{g}, t)}{\partial t} = &\sum_{k=1}^{n} c_{0k}(t) \, [g_k - 1] \, \psi(\mathbf{g}, t) \\
&- \sum_{k=1}^{n} c_{k0}(t) \, [g_k - 1] \, \frac{\partial \psi(\mathbf{g}, t)}{\partial g_k} \\
&+ \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \, [g_k - g_j] \, \frac{\partial \psi(\mathbf{g}, t)}{\partial g_j} \;.
\end{aligned}
\tag{4.29}
$$

With all of that notation defined, we are ready to state the main result for monomolecular reaction systems, which was originally proved by Jahnke and Huisinga[4].

Theorem 2 (Jahnke-Huisinga monomolecular). *Let $\boldsymbol{\lambda}(t)$ and $\mathbf{w}(t)$ be the solutions of*

$$
\begin{aligned}
\dot{\boldsymbol{\lambda}} &= \mathbf{A}(t)\boldsymbol{\lambda} + \mathbf{b}(t) \;, \quad \boldsymbol{\lambda}(t_0) = \mathbf{0} \\
\dot{\mathbf{w}}^{(k)} &= \mathbf{A}(t)\mathbf{w}^{(k)} \;, \quad \mathbf{w}^{(k)}(t_0) = \boldsymbol{\epsilon}_k \;.
\end{aligned}
\tag{4.30}
$$

Then if $P(\mathbf{x}, t_0) = \delta(\mathbf{x} - \boldsymbol{\xi})$ for some $\boldsymbol{\xi} \in \mathbb{N}^n$, we have:

i.

$$
P(\mathbf{x}, t; \boldsymbol{\xi}, t_0) = \mathcal{P}(\mathbf{x}, \boldsymbol{\lambda}(t)) \star \mathcal{M}(\mathbf{x}, \xi_1, \mathbf{w}^{(1)}(t)) \star \cdots \star \mathcal{M}(\mathbf{x}, \xi_n, \mathbf{w}^{(n)}(t))
\tag{4.31}
$$

ii.

$$
\langle x_j(t) \rangle = \sum_{k=1}^{n} \xi_k w_j^{(k)}(t) + \lambda_j(t)
\tag{4.32}
$$

---

[4] The generating function was not directly computed by them, but is essentially trivial to compute given their results.

iii.

$$\text{Cov}(x_j, x_\ell) = \begin{cases} \sum_{k=1}^{n} \xi_k w_j^{(k)} \left[1 - w_j^{(k)} \right] + \lambda_j & j = \ell \\ -\sum_{k=1}^{n} \xi_k w_j^{(k)} w_\ell^{(k)} & j \neq \ell \end{cases}. \tag{4.33}$$

iv.

$$\psi(\mathbf{g}, t) = \prod_{k=1}^{n} \left[1 + (\mathbf{g} - \mathbf{1}) \cdot \mathbf{w}^{(k)}(t) \right]^{\xi_k} e^{(\mathbf{g}-\mathbf{1}) \cdot \boldsymbol{\lambda}(t)} \tag{4.34}$$

Proof. All results can be obtained independently of one another using the Doi-Peliti approach described in the following sections. Alternatively, one can verify directly that $P(\mathbf{x}, t; \boldsymbol{\xi}, t_0)$ satisfies Eq. 4.18 or that $\psi(\mathbf{g}, t)$ satisfies Eq. 4.29 with the correct initial condition, and then obtain the rest of the results by brute force calculation. □

4.2.3 Birth-death-autocatalysis results

In section 6 of their classic paper [45], Jahnke and Huisinga solve the CME corresponding to the autocatalytic reaction $S \to S + S$ exactly; however, they note that adding birth and death reactions yields a system not amenable to their approach. In this section, we present the exact time-dependent solution to this problem, whose reactions read

$$\varnothing \xrightarrow{k} S$$
$$S \xrightarrow{\gamma} \varnothing \tag{4.35}$$
$$S \xrightarrow{c} S + S$$

where the parameters controlling the rates of birth, death, and autocatalysis are all allowed to have arbitrary time-dependence as long as they are nonnegative for all times. The CME reads

$$\begin{aligned} \frac{\partial P(x, t)}{\partial t} &= k(t) \left[P(x-1, t) - P(x, t) \right] \\ &+ \gamma(t) \left[(x+1) P(x+1, t) - x P(x, t) \right] \\ &+ c(t) \left[(x-1) P(x-1, t) - x P(x, t) \right] \end{aligned} \tag{4.36}$$

where $P(x, t)$ is the probability that the state of the system is $x \in \mathbb{N}$ at time $t \geq t_0$. Meanwhile, the PDE satisfied by the probability generating function ψ reads

$$\begin{aligned} \frac{\partial \psi(g, t)}{\partial t} &= k(t)[g-1]\psi(g, t) - \gamma(t)[g-1]\frac{\partial \psi(g, t)}{\partial g} \\ &+ c(t)[g-1]g\frac{\partial \psi(g, t)}{\partial g} \end{aligned} \tag{4.37}$$

Our main result on the solution of this system is the following.

Theorem 3 (Birth-death-autocatalysis). *Let $q(s)$ and $w(s)$ be the solutions of*

$$
\begin{aligned}
\dot{q} &= \left[c(t - s + t_0) - \gamma(t - s + t_0) \right] q(s) + ic(t - s + t_0)\, q(s)^2 \ , \quad q(t_0) = p_f \\
\dot{w} &= \left[c(t - s + t_0) - \gamma(t - s + t_0) \right] w(s) \ , \quad w(t_0) = 1
\end{aligned}
\tag{4.38}
$$

for arbitrary $p_f \in \mathbb{R}$, which can be explicitly written as

$$
q(s) = \frac{w(s)}{\frac{1}{p_f} - i \int_{t_0}^{s} c(t - t' + t_0) w(t')\, dt'}
\tag{4.39}
$$

$$
w(s) = e^{\int_{t_0}^{s} c(t - t' + t_0) - \gamma(t - t' + t_0)\, dt'} \ .
$$

Then if $P(x, t_0) = \delta(x - \xi)$ for some $\xi \in \mathbb{N}$, we have:

$$
P(x, t; \xi, t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dp_f \ \frac{\left[1 + iq(t) \right]^{\xi} e^{i \int_{t_0}^{t} k(t - s + t_0) q(s) ds}}{(1 + ip_f)^{x+1}}
$$

$$
\psi(g, t) = \left[1 + \frac{w(t)}{\frac{1}{g-1} - \int_{t_0}^{t} c(t - t' + t_0) w(t')\, dt'} \right]^{\xi} \times
\tag{4.40}
$$

$$
\times \exp\left\{ \int_{t_0}^{t} \frac{k(t - s + t_0) w(s)}{\frac{1}{g-1} - \int_{t_0}^{s} c(t - t' + t_0) w(t')\, dt'} \, ds \right\} \ .
$$

Moreover, if the parameters k, γ, and c are all time-independent and non-zero, the function $w(s)$ is explicitly

$$
w(s) = e^{(c - \gamma)(s - t_0)}
\tag{4.41}
$$

and the transition probability can be rewritten as

$$
P = \left(\frac{\frac{\gamma}{c} - 1}{\frac{\gamma}{c} - w} \right)^{k/c} \frac{(1 - w)^{x - \xi}}{\left(\frac{\gamma}{c} - w \right)^{x}} \times
$$

$$
\times \sum_{j=0}^{\xi} \binom{\xi}{j} \frac{(j + k/c)_x}{x!} \left[1 - \frac{\gamma}{c} w \right]^{\xi - j} \left[\frac{w \left(\frac{\gamma}{c} - 1 \right)^2}{\frac{\gamma}{c} - w} \right]^{j}
\tag{4.42}
$$

where $(y)_x := (y)(y + 1) \cdots (y + x - 1)$ is the Pochhammer symbol/rising factorial. The generating function in this case reduces to

$$
\psi(g) = \left[\frac{1 + (g - 1)\frac{c - \gamma w}{c - \gamma}}{1 - (g - 1)\frac{c(w - 1)}{c - \gamma}} \right]^{\xi} \frac{1}{\left[1 - (g - 1)\frac{c(w - 1)}{c - \gamma} \right]^{k/c}} \ .
\tag{4.43}
$$

Proof. The transition probability and probability generating function can be obtained independently of one another using the Doi-Peliti approach described in the following sections. Alternatively, one can verify directly that the probability generating function solves Eq. 4.37, and use the definition of the generating function to find the transition probability. Because the expression for the transition probability is somewhat complicated, verifying it directly is not recommended. $\square$

It is expected that this solution reduces to familiar distributions in certain limits; in particular, as Jahnke and Huisinga originally point out, it should interpolate between a binomial distribution, a Poisson distribution, and a negative binomial distribution. Indeed it does, with these special cases corresponding to the $k = c = 0$ (pure death), $\gamma = c = 0$ (pure birth), and $k = \gamma = 0$ (pure autocatalysis) limits, respectively. We formalize this in the following corollary.

Corollary 2 (Limiting behavior of the birth-death-autocatalysis transition probability). *The transition probability $P(x, t; x_0, t_0)$ becomes (i) binomial in the limit that $k = c = 0$, (ii) Poisson in the limit that $\gamma = c = 0$, and (iii) negative binomial in the limit that $k = \gamma = 0$. That is,*

$$\lim_{k,c \to 0} P(x, t; x_0, t_0) = \binom{\xi}{x} [w(t)]^x [1 - w(t)]^{\xi - x} \tag{4.44}$$

for $x \leq \xi$ and 0 otherwise, i.e. a binomial distribution;

$$\lim_{\gamma,c \to 0} P(x, t; x_0, t_0) = \frac{\lambda(t)^{x-\xi} e^{-\lambda(t)}}{(x - \xi)!} \tag{4.45}$$

for $x \geq \xi$ and 0 otherwise, i.e. a (shifted) Poisson distribution; and

$$\lim_{k,\gamma \to 0} P(x, t; x_0, t_0) = \binom{x-1}{\xi-1} [w(t)]^\xi [1 - w(t)]^{x-\xi} \tag{4.46}$$

which is nonzero only for $x \geq \xi$. As Jahnke and Huisinga note in their Sec. 6, this is a shifted negative binomial distribution.

Proof. These limits can be taken directly, and are relatively straightforward; see Sec. 4.5 for the calculations. $\qquad\square$

It is also true that the solution of this problem reduces to a birth-death process in the $c \to 0$ limit (i.e. the limit in which autocatalysis no longer happens). This limit is harder to take properly (it is not clear how to take it using $P(x, t; x_0, t_0)$, so the generating function must be used), and the resulting distribution is less special than the Poisson, binomial, or negative binomial distributions, but it is still interesting; for these reasons, we formalize it as its own corollary below.

Corollary 3 (No autocatalysis limit reduces to birth-death). *In the $c \to 0$ limit, the transition probability $P(x, t; x_0, t_0)$ and generating function $\psi(g, t)$ reduce to the ones for the chemical birth-death process described in Theorem 1.*

Proof. The easiest way to do this is to show that the generating functions correspond in the $c \to 0$ limit. Because this is straightforward, we will not show the calculation explicitly. $\qquad\square$

If autocatalysis happens tends to happen more frequently than degradation (i.e. if $c > \gamma$), the number of molecules almost surely blows up to infinity in the long time limit. However, if degradation tends to overtake autocatalysis (i.e. if $\gamma > c$), then the steady state distribution exists and is nontrivial.

Corollary 4 (Long time behavior of birth-death-autocatalysis). *Let k, γ, and c be time-independent, and suppose that $\gamma > c$. In the long time limit, we have:*

$$P_{ss}(x) = \left(\frac{\gamma - c}{\gamma}\right)^{k/c} \frac{\left(\frac{c}{\gamma}\right)^x}{x!} \left(\frac{k}{c}\right)_x$$

$$\psi_{ss}(g) = \left(\frac{\gamma - c}{\gamma}\right)^{k/c} \frac{1}{\left[1 - \frac{cg}{\gamma}\right]^{k/c}} \,.$$

(4.47)

Moreover, these reduce to the P_{ss} and ψ_{ss} for the chemical birth-death process in the $c \to 0$ limit, i.e.

$$\left(\frac{\gamma - c}{\gamma}\right)^{k/c} \frac{\left(\frac{c}{\gamma}\right)^x}{x!} \left(\frac{k}{c}\right)_x \to \frac{\mu^x}{x!} e^{-\mu}$$

$$\left(\frac{\gamma - c}{\gamma}\right)^{k/c} \frac{1}{\left[1 - \frac{cg}{\gamma}\right]^{k/c}} \to e^{(g-1)\mu} \,.$$

(4.48)

Proof. The simplest way to find P_{ss} is to solve the steady state CME directly (i.e. set $\partial P/\partial t = 0$ and solve the resulting recurrence relation). Alternatively, noting that $w \to 0$, one can straightforwardly take the $t \to \infty$ limit of our result from Theorem 3. The $c \to 0$ is also easy to take. $\qquad\square$

4.2.4 Zero and first order reactions

While obtaining the solution to the birth-death-autocatalysis system is certainly interesting, the problem itself is quite special: it is one-dimensional, and involves only three chemical reactions. What else can the Doi-Peliti approach be used to solve? What kind of sets of chemical reactions are tractable?

The full potential of the Doi-Peliti approach is not clear. As a partial answer to this question, however, we offer a result of somewhat shocking generality: a formal solution to the CME of any system whose reaction list only contains zero and first order reactions. By zero order reactions, we mean reactions like

$$\varnothing \to S_k$$
$$\varnothing \to S_k + S_\ell$$
$$\varnothing \to S_k + S_\ell + S_r$$
$$\varnothing \to S_k + S_\ell + S_r + \cdots$$

(4.49)

and so on, i.e. reactions requiring no molecules as input. By first order reactions, we mean

reactions like

$$
\begin{aligned}
S_j &\to \varnothing \\
S_j &\to S_k \\
S_j &\to S_k + S_\ell \\
S_j &\to S_k + S_\ell + S_r \\
S_j &\to S_k + S_\ell + S_r + \cdots
\end{aligned}
\tag{4.50}
$$
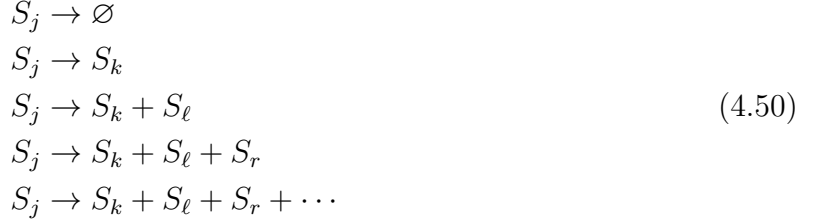
and so on, i.e. reactions requiring exactly one molecule as input. The birth reactions described in the previous sections are examples of zero order reactions, while the death and conversion reactions are examples of first order reactions. Other biologically relevant examples of first order reactions include catalytic production ($S_j \to S_j + S_k$, $j \neq k$) and splitting ($S_j \to S_k + S_\ell$, $j \neq k$, $j \neq \ell$).

The list of all possible zero and first order reactions also includes many reactions that are almost certainly *not* biologically relevant—for example, the reaction where one molecule splits into *exactly* one hundred molecules with no intermediate splitting.

The CME of this system is somewhat tedious to write down, so we will instead note that the PDE satisfied by the generating function can be written in the form

$$
\begin{aligned}
\frac{\partial \psi(\mathbf{g}, t)}{\partial t} &= \sum_{\nu_1, \ldots, \nu_n} \alpha_{\nu_1, \ldots, \nu_n}(t) \, (g_1 - 1)^{\nu_1} \cdots (g_n - 1)^{\nu_n} \, \psi(\mathbf{g}, t) \\
&\quad + \sum_{k} \sum_{\nu_1, \ldots, \nu_n} \beta^k_{\nu_1, \ldots, \nu_n}(t) \, (g_1 - 1)^{\nu_1} \cdots (g_n - 1)^{\nu_n} \, \frac{\partial \psi(\mathbf{g}, t)}{\partial g_k}
\end{aligned}
\tag{4.51}
$$

where the precise form of the coefficients $\alpha_{\nu_1, \ldots, \nu_n}(t)$ and $\beta^j_{\nu_1, \ldots, \nu_n}(t)$ are determined by the details of one's list of reactions. Our main result for this class of systems is the following.

Theorem 4 (Arbitrary combinations of zero and first order reactions)**.** *Let* $\mathbf{q}(s)$ *be the solution of*

$$
\dot{q}_j(s) = -i \sum_{\nu_1, \ldots, \nu_n} \beta^j_{\nu_1, \ldots, \nu_n}(t - s + t_0) \, [i q_1(s)]^{\nu_1} \cdots [i q_n(s)]^{\nu_n} \quad, \quad q_j(t_0) = p^f_j
\tag{4.52}
$$

for some $\mathbf{p}^f \in \mathbb{R}^n$*, with* $s \in [t_0, t]$*. Then if* $P(\mathbf{x}, t_0) = \delta(\mathbf{x} - \boldsymbol{\xi})$ *for some* $\boldsymbol{\xi} \in \mathbb{N}^n$*, we have*

$$
P = \int_{\mathbb{R}^n} \frac{d\mathbf{p}^f}{(2\pi)^n} \frac{[\mathbf{1} + i\mathbf{q}(t)]^{\boldsymbol{\xi}} \, e^{\int_{t_0}^{t} \sum \alpha_{\nu_1, \ldots, \nu_n}(t - s + t_0)[i q_1(s)]^{\nu_1} \cdots [i q_n(s)]^{\nu_n} \, ds}}{(\mathbf{1} + i\mathbf{p}^f)^{\mathbf{x} + \mathbf{1}}}
\tag{4.53}
$$

for the transition probability, and

$$
\psi(\mathbf{g}, t) = [\mathbf{1} + i\mathbf{q}(t)]^{\boldsymbol{\xi}} \times \\
\times \, e^{\int_{t_0}^{t} \sum \alpha_{\nu_1, \ldots, \nu_n}(t - s + t_0)[i q_1(s)]^{\nu_1} \cdots [i q_n(s)]^{\nu_n} \, ds} \Big|_{\mathbf{p}^f = -i(\mathbf{g} - \mathbf{1})}
\tag{4.54}
$$

for the probability generating function.

Proof. In principle, the transition probability or generating function could be verified by substitution directly; however, this seems very difficult. A better way is discussed in Sec. 4.7.

$\square$

In some sense all other results in this paper are corollaries of this result; still, it is helpful to study the simpler cases in their own right, both to double-check the correctness of this more general result, and to develop a sense for how to derive these solutions.

While this result is perhaps shockingly general, it is also incredibly formal. For most systems of interest, it is likely that reducing the problem of solving the CME (an infinite number of coupled linear ODEs) to the problem of solving n coupled nonlinear ODEs is not much of an improvement. However, in some cases legitimate simplification seems possible: monomolecular systems and the birth-death-autocatalysis system addressed in the previous two theorems are clearly examples. It is not clear whether there are large classes of more complicated systems for which the ODEs given by Eq. 4.52 are solvable, but searching for them seems like a promising avenue for future research on solving the CME.

While we have not pursued this question, it is also possible that solving these ODEs numerically could yield new insights for efficiently solving the CME on computers.

4.3   Reframing the problem and basic Doi-Peliti formalism

4.3.1   A brief digression on notation

Although it is not normally used in the study of stochastic processes, it is the author's strong belief that the bra-ket notation originally developed for quantum mechanics is most appropriate here. Because this notation makes it harder for most mathematicians to read this paper, here we will briefly argue why this is necessary. The reader who wishes to review the basics of bra-ket notation, and see how it compares to more standard mathematical notation for things like vector spaces and inner products, should refer to Appendix 4.A.

Given that the problems we are attempting to solve are quite complicated, carefully choosing notation is important; a bad choice of notation would clutter our already complicated arguments, making them nearly impossible to understand. We would like notation that (i) *generalizes* cleanly to complicated systems in arbitrarily many dimensions; (ii) *simplifies* the construction of the Doi-Peliti path integral; and (iii) is *suggestive* of the operations we want to take, and not suggestive of operations that are not valid.

Let us say more about each heuristic requirement:

1. **Generalizes**: We will usually work with systems for which there are n distinct chemical species, where $n \geq 1$ is some positive integer. In the next section, we will see that this forces us to work in a Hilbert space where each basis vector can be identified with an

element of $\mathbb{N}^n$. We will eventually need notation for each basis vector, as well as for sums over $\mathbb{N}^n$, integrals over $\mathbb{R}^n$, and eigenvectors with eigenvalues $\mathbf{z} \in \mathbb{C}^n$. Denoting basis vectors, eigenvectors, and things like sums and integrals cleanly in arbitrarily many dimensions is easy using bra-ket notation.

2. **Simplifies**: Constructing the Doi-Peliti path integral involves using many identity operators/resolutions of the identity (see Sec. 4.3.5). This is cleanest with bra-ket notation, and using alternative notation obfuscates these steps.

3. **Suggestive/not confusing**: We will have to compute many inner products, as well as different kinds of inner products. Bra-ket notation allows them to be denoted simply, e.g. the inner product of $|x\rangle$ and $|y\rangle$ is $\langle x|y \rangle$. If we used generating function notation, where we have g^x instead of $|x\rangle$ and g^y instead of $|y\rangle$, we would have to define strange operations like $g^x \cdot g^y = x! \delta(x - y)$. Moreover, this notation suggests operations like $g^x g^y = g^{x+y}$ are valid, although they are not. Vector notation (using e.g. e_x and e_y to denote basis vectors) would be somewhat confusing, because we are already considering vectors like $\mathbf{x} \in \mathbb{N}^n$ to denote particular states of our system.

Aside from issues of notation, there is a 'deeper' reason this path integral requires special notation, whereas for others (see e.g. [69]) standard notation and a Chapman-Kolmogorov-based argument suffices. Most of the time, when path integrals are applied to stochastic processes or mathematical biology, what one is *really* doing is applying the Chapman-Kolmogorov equation many times. This has the interpretation that one is imagining all possible paths from one state to another state and appropriately discretizing them.

This kind of path integral is qualitatively different. It involves expanding an abstract object (rather than the transition probability itself) in terms of coherent states (which we will define later), which are themselves kind of abstract Poisson-like distributions. There does not seem to be the same obvious interpretation linking this path integral to the Chapman-Kolmogorov equation, or to all possible paths through state space (the space of all possible configurations of the system, i.e. $\mathbb{N}^n$ for a system with n distinct chemical species).

4.3.2   Hilbert space, the generating function, and basic operators

In order to apply the Doi-Peliti technique, we first need to rewrite the CME in terms of states and operators in a certain Hilbert space. Consider an infinite-dimensional Hilbert space spanned by the $|\mathbf{x}\rangle$ states/basis vectors (where $\mathbf{x} = (x_1, ..., x_n)^T \in \mathbb{N}^n$), in which an arbitrary state $|\phi\rangle$ is written

$$|\phi\rangle = \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} c(\mathbf{x}) \, |\mathbf{x}\rangle \tag{4.55}$$

for some generally complex-valued coefficients $c(\mathbf{x})$. These states can be added and multiplied by (complex) scalars in the usual way. There is one basis vector for every possible state

of the system (recall that the state space of the system is $\mathbb{N}^n$), e.g. $|00\cdots0\rangle$, $|01\cdots0\rangle$, $|20,45,1,\cdots10\rangle$, and so on; one interpretation of these objects is that they encode a certain generalization of probability distributions on the state space, given that they assign every $\mathbf{x} \in \mathbb{N}^n$ a complex number.

It should be noted that the basis vectors cannot be combined since they represent distinct directions in the Hilbert space, i.e. $|\mathbf{x}\rangle + |\mathbf{y}\rangle \neq |\mathbf{x} + \mathbf{y}\rangle$. We will denote the zero vector by 0, which we emphasize for clarity's sake is distinct from the basis vector $|\mathbf{0}\rangle$ (e.g. $|\mathbf{0}\rangle + 0 = |\mathbf{0}\rangle$). The relevant inner products (without which this would just be a vector space) will be described in a few sections.

To ease notation, we remind the reader that we will write

$$\sum_{\mathbf{x}} := \sum_{x_1=0}^{\infty} \cdots \sum_{x_n=0}^{\infty} \ . \tag{4.56}$$

The state we are principally interested in is the generating function, which is essentially the function $\psi(\mathbf{g}, t)$ described earlier, but using different notation. For more information on their correspondence, see Appendix 4.A.

Definition 1. *The generating function is defined to be the state*

$$|\psi(t)\rangle := \sum_{\mathbf{x}} P(\mathbf{x}, t) \, |\mathbf{x}\rangle \tag{4.57}$$

where $P(\mathbf{x}, t)$ is some solution to the CME (i.e. its precise form depends on the chosen initial condition $P(x, t_0)$).

For the rest of this paper, we will only be concerned with the case where $P(\mathbf{x}, t_0) = \delta(\mathbf{x} - \boldsymbol{\xi})$ for some $\boldsymbol{\xi} \in \mathbb{N}^n$, so we will always assume that $|\psi(t_0)\rangle = |\boldsymbol{\xi}\rangle$.

Because the generating function $|\psi(t)\rangle$ depends on $P(\mathbf{x}, t)$, whose dynamics are controlled by the CME, $|\psi(t)\rangle$ also has dynamics; we can write the equation controlling them (its 'equation of motion') in the form

$$\frac{\partial |\psi\rangle}{\partial t} = \hat{H} |\psi\rangle \tag{4.58}$$

where the Hamiltonian operator $\hat{H}$ is a linear operator whose precise form depends on the CME. For the reader familiar with quantum mechanics, this is analogous to the equation of motion for a quantum mechanical state. In any case, it is this equation that we will solve instead of the CME.

It may or may not be helpful for the reader to think of $\hat{H}$ as a (possibly infinite-dimensional) matrix. Although it is infinite-dimensional in essentially every case we care about in this paper, it would literally be a matrix if we were solving a CME with a finite state space. One example of a problem with a finite state space is the pure conversion process ($A \leftrightarrow B$), which involves A molecules and B molecules randomly converting between each other; it has a finite state space because the total number of molecules remains constant.

Pressing the analogy between $\hat{H}$ and matrices, we have the usual formal solution for the generating function $|\psi(t)\rangle$ in terms of the (time-ordered) exponential of $\hat{H}$.

Proposition 1 (Formal solution for the generating function)**.** *The equation of motion for the generating function $|\psi(t)\rangle$ (Eq. 4.58) has the formal solution*

$$
\begin{aligned}
|\psi(t)\rangle &= \hat{T}e^{\int_{t_0}^t \hat{H}(t')dt'}|\psi(t_0)\rangle \\
&= \sum_{j=0}^\infty \frac{1}{j!}\int_{t_0}^t \cdots \int_{t_0}^t \hat{T}\left[\hat{H}(t_1)\cdots\hat{H}(t_n)\right]\,dt_1\cdots dt_n \\
&= 1 + \int_{t_0}^t \hat{H}(t_1)dt_1 + \frac{1}{2}\int_{t_0}^t\int_{t_0}^t \hat{T}\left[\hat{H}(t_1)\hat{H}(t_2)\right]dt_1 dt_2 + \cdots
\end{aligned}
\tag{4.59}
$$

where $\hat{T}$ is the time-ordering symbol, whose action on a product of operators is defined to be

$$
\hat{T}\left[\hat{\mathcal{A}}_1(t)\hat{\mathcal{A}}_2(t')\right] := \left\{
\begin{array}{ll}
\hat{\mathcal{A}}_1(t)\hat{\mathcal{A}}_2(t') & t \geq t' \\
\hat{\mathcal{A}}_2(t')\hat{\mathcal{A}}_1(t) & t < t'
\end{array}
\right. .
\tag{4.60}
$$

Proof. Substitute this expression for the generating function directly into Eq. 4.58. While the presence of the time-ordering symbol makes this more subtle than it would be in the case of a time-independent Hamiltonian (i.e. in the case where the reaction parameters were all time-independent), this exercise is standard, so we will not go through this in detail. $\square$

Corollary 5. *For time-independent $\hat{H}$, the above formal solution reduces to*

$$
|\psi(t)\rangle = e^{\hat{H}(t-t_0)}|\psi(t_0)\rangle .
\tag{4.61}
$$

Proof. One can either show that this solves the equation of motion directly, or simplify the result above. $\square$

Fortunately, we will never have to work with a time-ordered exponential of operators directly. The first salient consequence of the formal solution for us is that

$$
|\psi(t+\Delta t)\rangle \approx \left[1 + \hat{H}(t)\Delta t\right]|\psi(t)\rangle
\tag{4.62}
$$

for sufficiently small Δt, with the approximation becoming exact in the $\Delta t \to 0$ limit. Notice that this also matches what we would find by naively approximating the time derivative with a finite difference in Eq. 4.58.

This formal solution motivates defining the time evolution operator, which carries the solution at time t_1 (the state $|\psi(t_1)\rangle$) to the solution at time t_2 (the state $|\psi(t_2)\rangle$).

Definition 2. *The time evolution operator $\hat{U}(t_2, t_1)$ is defined as*

$$
\begin{aligned}
\hat{U}(t_2, t_1) &:= \hat{T}e^{\int_{t_1}^{t_2}\hat{H}(t')dt'} \\
&= 1 + \int_{t_1}^{t_2}\hat{H}(s_1)ds_1 + \frac{1}{2}\int_{t_1}^{t_2}\int_{t_1}^{t_2}\hat{T}\left[\hat{H}(s_1)\hat{H}(s_2)\right]ds_1 ds_2 + \cdots
\end{aligned}
\tag{4.63}
$$

for any two times $t_1 \leq t_2$. In terms of the time evolution operator, the formal solution for $|\psi(t)\rangle$ can be written

$$
|\psi(t)\rangle = \hat{U}(t, t_0)|\psi(t_0)\rangle .
\tag{4.64}
$$

The second salient consequnce of Proposition 1 is that this operator has an important composition property.

Proposition 2 (Composition property of the time evolution operator)**.** *The time evolution operator $\hat{U}$ has the property that*

$$\hat{U}(t_2, t_1) = \hat{U}(t_2, t')\hat{U}(t', t_1) \tag{4.65}$$

for any time t' with $t_1 \leq t' \leq t_2$.

Proof. This is most easily seen using the infinite series form of the time evolution operator $\hat{U}$, by expanding both sides and showing that they match at each order. $\square$

Because we are interested in the dynamics of the generating function $|\psi(t)\rangle$, we need to introduce operators to act on it. In particular, we introduce the creation and annihilation operators, which we will later use to write the Hamiltonian operator.

Definition 3. *Define the annihilation and creation operators $\hat{a}_j$ and $\hat{\pi}_j$ for all $j = 1, ..., n$ as the operators whose action on a basis vector $|\mathbf{x}\rangle$ is*

$$\begin{aligned} \hat{a}_j |\mathbf{x}\rangle &= x_j |\mathbf{x} - \boldsymbol{\epsilon}_j\rangle \\ \hat{\pi}_j |\mathbf{x}\rangle &= |\mathbf{x} + \boldsymbol{\epsilon}_j\rangle \end{aligned} \tag{4.66}$$

where we remind the reader that $\boldsymbol{\epsilon}_j$ is the n-dimensional vector with a 1 in the jth place and zeros everywhere else.

It is easy to show that these operators satisfy the commutation relations analogous to those seen in quantum mechanics (for example, in the ladder operator treatment of the harmonic oscillator [73], or in the canonical quantization approach to quantum field theory [74]). These properties will be used in calculations a few times throughout this paper.

Proposition 3. *Recall that, for two operators $\hat{\mathcal{A}}_1$ and $\hat{\mathcal{A}}_2$, their commutator is defined to be $[\hat{\mathcal{A}}_1, \hat{\mathcal{A}}_2] := \hat{\mathcal{A}}_1\hat{\mathcal{A}}_2 - \hat{\mathcal{A}}_2\hat{\mathcal{A}}_1$. The creation and annihilation operators satisfy the commutation relations*

$$[\hat{a}_j, \hat{\pi}_k] = \delta(j - k) \ , \ \ [\hat{a}_j, \hat{a}_k] = [\hat{\pi}_j, \hat{\pi}_k] = 0 \ . \tag{4.67}$$

Proof. Use their definitions to straightforwardly show this. $\square$

In essence, the Doi-Peliti approach to solving Eq. 4.58 involves using many coherent state 'resolutions of the identity' (a phrase we will define shortly) to rewrite Eq. 4.59 as a coherent state path integral. Once that path integral is evaluated, quantities like moments and $P(\mathbf{x}, t)$ can be recovered by manipulating the path integral solution in specific ways. In order to follow this prescription, we will need to define coherent states, define inner products, and construct associated resolutions of the identity; that is our next task.

4.3.3 Coherent states

Because we will be expressing the Hamiltonian operator in terms of creation and annihilation operators, it is convenient to work in terms of states that behave simply when acted upon by these operators. These are the so-called coherent states, which are often used to study the semiclassical limit of quantum mechanics. Here, we will only care about them for their algebraic properties; while their biological meaning is not completely obscure (they are essentially states that correspond to Poisson distributions), thinking about it is not necessary in what follows.

Definition 4. *Let* $\mathbf{z} = (z_1, ..., z_n)^T \in \mathbb{C}^n$. *A coherent state is a state*

$$|\mathbf{z}\rangle := \sum_{\mathbf{y}} c(\mathbf{y}) \ |\mathbf{y}\rangle \tag{4.68}$$

satisfying

$$
\begin{aligned}
\hat{a}_j |\mathbf{z}\rangle &= z_j |\mathbf{z}\rangle && \text{for all } j = 1, ..., n \\
\sum_{\mathbf{y}} c(\mathbf{y}) &= 1
\end{aligned}
\tag{4.69}
$$

i.e. it is an eigenstate of all annihilation operators $\hat{a}_j$, *and it has a specific normalization.*

By imposing the eigenstate constraint on an arbitrary state, it is straightforward to determine the coefficients $c(\mathbf{y})$ explicitly. Coherent states can also be written in terms of a specific combination of creation operators acting on the 'vacuum' state $|\mathbf{0}\rangle$. We make these statements more precise in the following proposition.

Proposition 4. *The coherent state* $|\mathbf{z}\rangle$ *can explicitly be written in the following two equivalent forms:*

i.

$$|\mathbf{z}\rangle = \sum_{\mathbf{y}} \frac{z_1^{y_1} \cdots z_n^{y_n}}{y_1! \cdots y_n!} e^{-(z_1 + \cdots + z_n)} \ |\mathbf{y}\rangle = \sum_{\mathbf{y}} \frac{\mathbf{z}^{\mathbf{y}}}{\mathbf{y}!} e^{-\mathbf{z} \cdot \mathbf{1}} \ |\mathbf{y}\rangle \tag{4.70}$$

ii.

$$|\mathbf{z}\rangle = \sum_{\mathbf{y}} \frac{[z_1(\hat{\pi}_1 - 1)]^{y_1} \cdots [z_n(\hat{\pi}_n - 1)]^{y_n}}{y_1! \cdots y_n!} |\mathbf{0}\rangle = e^{\mathbf{z} \cdot (\hat{\boldsymbol{\pi}} - \mathbf{1})} |\mathbf{0}\rangle \tag{4.71}$$

where $\hat{\boldsymbol{\pi}} := (\hat{\pi}_1, ..., \hat{\pi}_n)^T$.

Proof. Showing (i) is straightforward. To show (ii), first note that $[\hat{a}_j, (\hat{\pi}_j - 1)^y] = y(\hat{\pi}_j - 1)^{y-1}$ for all $y \in \mathbb{N}$, a useful commutator result that can be proved by induction. Using this, along

with the facts that $\hat{a}_j$ commutes with $\hat{\pi}_k$ for $k \neq j$ and the $\hat{\pi}_k$ all commute with each other, we have

$$
\begin{aligned}
\hat{a}_j \left| z \right\rangle &= e^{\mathbf{z} \cdot (\hat{\boldsymbol{\pi}} - \mathbf{1}) - z_j(\hat{\pi}_j - 1)} \sum_{y_j=0}^{\infty} \frac{z_j^{y_j}}{y_j!} \hat{a}_j (\hat{\pi}_j - 1)^{y_j} \left| \mathbf{0} \right\rangle \\
&= e^{\mathbf{z} \cdot (\hat{\boldsymbol{\pi}} - \mathbf{1}) - z_j(\hat{\pi}_j - 1)} \sum_{y_j=0}^{\infty} \frac{z_j^{y_j}}{y_j!} \left\{ (\hat{\pi}_j - 1)^{y_j} \hat{a}_j + y_j(\hat{\pi}_j - 1)^{y_j - 1} \right\} \left| \mathbf{0} \right\rangle \\
&= z_j \; e^{\mathbf{z} \cdot (\hat{\boldsymbol{\pi}} - \mathbf{1}) - z_j(\hat{\pi}_j - 1)} \sum_{y_j=1}^{\infty} \frac{z_j^{y_j - 1} (\hat{\pi}_j - 1)^{y_j - 1}}{(y_j - 1)!} \left| \mathbf{0} \right\rangle \\
&= z_j \left| \mathbf{z} \right\rangle \; .
\end{aligned}
\tag{4.72}
$$

Hence, this expression satisfies the eigenstate constraint. Noting that eigenstates are unique up to a proportionality constant, to show that it satisfies the normalization constraint (and hence is the same as the expression given by (i)), observe that

$$
\begin{aligned}
e^{\mathbf{z} \cdot (\hat{\boldsymbol{\pi}} - \mathbf{1})} \left| \mathbf{0} \right\rangle &= \sum_{\mathbf{y}} \frac{z_1^{y_1} \cdots z_n^{y_n} \left[(-1)^{y_1 + \cdots + y_n} + \cdots \right]}{y_1! \cdots y_n!} \left| \mathbf{0} \right\rangle \\
&= e^{-\mathbf{z} \cdot \mathbf{1}} \left| \mathbf{0} \right\rangle + \cdots
\end{aligned}
\tag{4.73}
$$

i.e. the coefficient of $\left| \mathbf{0} \right\rangle$ is $e^{-\mathbf{z} \cdot \mathbf{1}}$, because every other term in the above expansion contains a creation operator. Because this is the same as the coefficient of $\left| \mathbf{0} \right\rangle$ in (i), we have equivalence.
□

In what follows, we will generally reserve the letters $\mathbf{z}$ and $\mathbf{p}$ for coherent states.

4.3.4 Inner products

Now we will define two inner products on our Hilbert space: the exclusive product, and the Grassberger-Scheunert product. Both were introduced by Grassberger and Scheunert in a 1980 paper that clearly describes their motivation and properties [54]; we are calling their "inclusive" inner product the Grassberger-Scheunert product to recognize their contribution.

Briefly, the exclusive product is useful for computing $P(\mathbf{x}, t)$, while the Grassberger-Scheunert product is useful for simplifying path integral calculations (specifically, we avoid having to perform a "Doi shift" [75, 76]; see Eq. 3.4 of Peliti [51] for an example of the Doi shift) and computing moments. We will use both inner products in solving the CME.

In this section and the following sections, the reader should keep in mind that $\langle \mathbf{x} | \hat{\mathcal{A}}(t) | \mathbf{y} \rangle$, where $\hat{\mathcal{A}}(t)$ is some possibly time-dependent operator, means the same as $\langle \mathbf{e_x}, \hat{\mathcal{A}}(t) \mathbf{e_y} \rangle$ in more standard notation. See Appendix 4.A for more details.

Definition 5. *Let* $\left| \mathbf{x} \right\rangle$ *and* $\left| \mathbf{y} \right\rangle$ *be basis vectors. Their exclusive product is defined to be*

$$
\langle \mathbf{x} | \mathbf{y} \rangle_{ex} := \mathbf{x}! \; \delta(\mathbf{x} - \mathbf{y}) \; .
\tag{4.74}
$$

Extending this by linearity, define the exclusive product of two arbitrary states $|\phi_1\rangle$ and $|\phi_2\rangle$ as (c.f. Eq. 4.55)

$$\langle\phi_2|\phi_1\rangle_{ex} = \sum_{\mathbf{x}} \mathbf{x}! \; c_2^*(\mathbf{x})c_1(\mathbf{x}) \; . \tag{4.75}$$

Definition 6. *Let $|\mathbf{x}\rangle$ and $|\mathbf{y}\rangle$ be basis vectors, and define $\hat{\mathbf{a}} := (\hat{a}_1,...,\hat{a}_n)^T$. Their Grassberger-Scheunert product is defined to be*

$$\langle\mathbf{x}|\mathbf{y}\rangle := \langle\mathbf{x}|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{y}\rangle_{ex} = \sum_{\mathbf{k}} \frac{\mathbf{x}! \; \mathbf{y}!}{(\mathbf{x}-\mathbf{k})! \; (\mathbf{y}-\mathbf{k})! \; \mathbf{k}!} \tag{4.76}$$

where the sum on the right is over all values of $\mathbf{k} \in \mathbb{N}^n$ with $k_j \leq \min(x_j, y_j)$ for all $j = 1,...,n$. Extending this by linearity, define the Grassberger-Scheunert product of two arbitrary states $|\phi_1\rangle$ and $|\phi_2\rangle$ as

$$\langle\phi_2|\phi_1\rangle = \langle\phi_2|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\phi_1\rangle_{ex} \; . \tag{4.77}$$

While it is not obvious just from looking at them, it is straightforward to show that the operator-based and sum-based definitions are equivalent (see Grassberger and Scheunert [54] and the appendix to Peliti [51]).

The primary reason these inner products are useful to define is that the creation and annihilation operators behave well under Hermitian conjugation with respect to them.

Proposition 5 (Hermitian conjugates of creation and annihilation operators). *Let $|\mathbf{x}\rangle$ and $|\mathbf{y}\rangle$ be basis vectors. With respect to the exclusive product, $\hat{a}_j$ and $\hat{\pi}_j$ are Hermitian conjugates of each other for all $j = 1,...,n$, i.e.*

$$(\hat{a}_j)^\dagger = \hat{\pi}_j$$
$$\langle\mathbf{x}|\hat{a}_j|\mathbf{y}\rangle_{ex} = \langle\mathbf{y}|\hat{\pi}_j|\mathbf{x}\rangle_{ex} \; . \tag{4.78}$$

With respect to the Grassberger-Scheunert product, the Hermitian conjugate of $\hat{a}_j$ is

$$(\hat{a}_j)^\dagger = \hat{\pi}_j - 1 \tag{4.79}$$

for all $j = 1,...,n$, i.e.

$$\langle\mathbf{x}|\hat{a}_j|\mathbf{y}\rangle = \langle\mathbf{y}|\hat{\pi}_j - 1|\mathbf{x}\rangle \; . \tag{4.80}$$

Proof. Showing that $\hat{\pi}_j$ and $\hat{a}_j$ are Hermitian conjugates with respect to the exclusive product is straightforward given their definitions, so we will show that $(\hat{a}_j)^\dagger = \hat{\pi}_j - 1$ with respect to the Grassberger-Scheunert product.

Recall the result mentioned in the proof of Proposition 4 that $[\hat{a}_j, (\hat{\pi}_j - 1)^y] = y(\hat{\pi}_j - 1)^{y-1}$ for all $y \in \mathbb{N}$. Using just the same argument, one can show $[\hat{a}_j, \hat{\pi}_j^y] = y(\hat{\pi}_j)^{y-1}$ for all $y \in \mathbb{N}$. This, in turn, can be used to prove that

$$e^{\hat{\pi}_j}\hat{a}_j = (\hat{a}_j - 1)e^{\hat{\pi}_j} \; . \tag{4.81}$$

Let $|\mathbf{x}\rangle$ and $|\mathbf{y}\rangle$ be arbitrary basis vectors. Now we can say that

$$
\begin{aligned}
\langle\mathbf{x}|\hat{a}_j|\mathbf{y}\rangle &= \langle\mathbf{x}|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}\hat{a}_j|\mathbf{y}\rangle_{ex} \\
&= \langle\mathbf{x}|e^{\hat{\pi}_j}\hat{a}_j\ e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}-\hat{\pi}_j}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{y}\rangle_{ex} \\
&= \langle\mathbf{x}|(\hat{a}_j-1)e^{\hat{\pi}_j}\ e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}-\hat{\pi}_j}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{y}\rangle_{ex} \\
&= \langle\mathbf{x}+\boldsymbol{\epsilon}_j|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{y}\rangle_{ex} - \langle\mathbf{x}|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{y}\rangle_{ex} \\
&= \langle\mathbf{x}+\boldsymbol{\epsilon}_j|\mathbf{y}\rangle - \langle\mathbf{x}|\mathbf{y}\rangle
\end{aligned}
\tag{4.82}
$$

where we have used the fact that $\hat{\pi}_j$ and $\hat{a}_j$ are Hermitian conjugates with respect to the exclusive product in the next to last step. But this is the same as

$$
\langle\mathbf{y}|\hat{\pi}_j-1|\mathbf{x}\rangle = \langle\mathbf{y}|\mathbf{x}+\boldsymbol{\epsilon}_j\rangle - \langle\mathbf{y}|\mathbf{x}\rangle
\tag{4.83}
$$

because the Grassberger-Scheunert product of two basis vectors is symmetric. Hence, $\hat{a}_j$ and $\hat{\pi}_j-1$ are Hermitian conjugates with respect to the Grassberger-Scheunert product. $\qquad\square$

Now let us compute some inner products that we will use later.

Proposition 6 (Useful inner products)**.** *Let $|\mathbf{x}\rangle$ be a basis vector, and let $|\mathbf{z}\rangle$ and $|\mathbf{p}\rangle$ be coherent states. Then*

i.

$$
\langle\mathbf{x}|\mathbf{z}\rangle_{ex} = \mathbf{z}^{\mathbf{x}}e^{-\mathbf{z}\cdot\mathbf{1}}
\tag{4.84}
$$

ii.

$$
\langle\mathbf{x}|\mathbf{z}\rangle = (\mathbf{1}+\mathbf{z})^{\mathbf{x}}
\tag{4.85}
$$

iii.

$$
\langle\mathbf{p}|\mathbf{z}\rangle_{ex} = e^{\mathbf{p}^*\cdot\mathbf{z}-(\mathbf{p}^*+\mathbf{z})\cdot\mathbf{1}}\ .
\tag{4.86}
$$

iv.

$$
\langle\mathbf{p}|\mathbf{z}\rangle = e^{\mathbf{p}^*\cdot\mathbf{z}}
\tag{4.87}
$$

Moreover, we remind the reader that other results (e.g. $\langle\mathbf{z}|\mathbf{x}\rangle = (\mathbf{1}+\mathbf{z}^*)^{\mathbf{x}}$) can be obtained from the above ones by taking a complex conjugate.

Proof. First, the exclusive product of a basis state $|\mathbf{x}\rangle$ with a coherent state $|\mathbf{z}\rangle$ is

$$
\langle\mathbf{x}|\mathbf{z}\rangle_{ex} = \sum_{\mathbf{y}}\frac{\mathbf{z}^{\mathbf{y}}}{\mathbf{y}!}e^{-\mathbf{z}\cdot\mathbf{1}}\ \langle\mathbf{x}|\mathbf{y}\rangle_{ex} = \sum_{\mathbf{y}}\frac{\mathbf{z}^{\mathbf{y}}}{\mathbf{y}!}e^{-\mathbf{z}\cdot\mathbf{1}}\ \mathbf{x}!\ \delta_{\mathbf{xy}} = \mathbf{z}^{\mathbf{x}}e^{-\mathbf{z}\cdot\mathbf{1}}\ .
\tag{4.88}
$$

Next, the Grassberger-Scheunert product of a basis state $|\mathbf{x}\rangle$ with a coherent state $|\mathbf{z}\rangle$ is

$$
\begin{aligned}
\langle\mathbf{x}|\mathbf{z}\rangle &= \langle\mathbf{x}|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{z}\rangle_{ex} \\
&= e^{\mathbf{z}\cdot\mathbf{1}}\langle\mathbf{x}|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}}|\mathbf{z}\rangle_{ex} \\
&= e^{(\mathbf{z}+\mathbf{1})\cdot\mathbf{1}}\langle\mathbf{x}|e^{(\mathbf{z}+\mathbf{1})\cdot(\hat{\boldsymbol{\pi}}-\mathbf{1})}|\mathbf{0}\rangle_{ex} \\
&= e^{(\mathbf{z}+\mathbf{1})\cdot\mathbf{1}}\langle\mathbf{x}|\mathbf{z}+\mathbf{1}\rangle_{ex} \\
&= (\mathbf{1}+\mathbf{z})^{\mathbf{x}}
\end{aligned}
\tag{4.89}
$$

where we have used that $|\mathbf{z}\rangle$ is an eigenstate of the annihilation operators $\hat{a}_j$, the operator representation of $|\mathbf{z}\rangle$ from Proposition 4, and Eq. 4.88. The exclusive product of two coherent states is

$$\langle \mathbf{p}|\mathbf{z}\rangle_{ex} = \sum_{\mathbf{y}} \frac{(\mathbf{p}^*)^{\mathbf{y}}}{\mathbf{y}!} e^{-\mathbf{p}^* \cdot \mathbf{1}} \quad \langle \mathbf{y}|\mathbf{z}\rangle_{ex} = \sum_{\mathbf{y}} \frac{(\mathbf{p}^*)^{\mathbf{y}} \, \mathbf{z}^{\mathbf{y}}}{\mathbf{y}!} e^{-(\mathbf{p}^*+\mathbf{z})\cdot\mathbf{1}} = e^{\mathbf{p}^*\cdot\mathbf{z} - (\mathbf{p}^*+\mathbf{z})\cdot\mathbf{1}} \; . \tag{4.90}$$

Finally, the Grassberger-Scheunert product of two coherent states is

$$\langle \mathbf{p}|\mathbf{z}\rangle = \langle \mathbf{p}|e^{\hat{\boldsymbol{\pi}}\cdot\mathbf{1}} e^{\hat{\mathbf{a}}\cdot\mathbf{1}}|\mathbf{z}\rangle_{ex} = e^{(\mathbf{p}^*+\mathbf{z})\cdot\mathbf{1}} \langle \mathbf{p}|\mathbf{z}\rangle_{ex} = e^{\mathbf{p}^*\cdot\mathbf{z}} \tag{4.91}$$

where we have used Eq. 4.90. $\qquad\square$

4.3.5 Resolution of the identity

The phrase 'resolution of the identity' refers to a useful way to write the identity operator. In our case, we would like to write the identity operator in terms of coherent states, which will allow us to construct the Doi-Peliti path integral. The relevant proposition, using coherent states and the Grassberger-Scheunert product, is the following.

Proposition 7 (Identity operator in terms of coherent states). *Let $|\mathbf{x}\rangle$ be a basis vector, and $|\mathbf{z}\rangle$ and $|-i\mathbf{p}\rangle$ be coherent states. Then*

$$|\mathbf{x}\rangle = \int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n} \; |\mathbf{z}\rangle \langle -i\mathbf{p}|\mathbf{x}\rangle \, e^{-i\mathbf{z}\cdot\mathbf{p}} \tag{4.92}$$

i.e.

$$1 = \int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n} \; |\mathbf{z}\rangle \langle -i\mathbf{p}| \, e^{-i\mathbf{z}\cdot\mathbf{p}} \tag{4.93}$$

is the identity operator (because the relationship holds for basis vectors, it holds for all states by linearity).

Proof. To establish Eq. 4.93, first observe that

$$
\begin{aligned}
&\int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n} \; |\mathbf{z}\rangle \langle -i\mathbf{p}|\mathbf{x}\rangle \, e^{-i\mathbf{z}\cdot\mathbf{p}} \\
&= \int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n} \; |\mathbf{z}\rangle \, (\mathbf{1}+i\mathbf{p})^{\mathbf{x}} \, e^{-i\mathbf{z}\cdot\mathbf{p}} \\
&= \sum_{\mathbf{y}} \frac{1}{\mathbf{y}!} \, |\mathbf{y}\rangle \int_{[0,\infty)^n} d\mathbf{z} \, \mathbf{z}^{\mathbf{y}} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n} \; (\mathbf{1}+i\mathbf{p})^{\mathbf{x}} \, e^{-\mathbf{z}\cdot(\mathbf{1}+i\mathbf{p})} \\
&= \sum_{\mathbf{y}} \frac{1}{\mathbf{y}!} \, |\mathbf{y}\rangle \int_{[0,\infty)^n} d\mathbf{z} \, \mathbf{z}^{\mathbf{y}} \left(-\frac{d}{d\mathbf{z}}\right)^{\mathbf{x}} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n} \; e^{-\mathbf{z}\cdot(\mathbf{1}+i\mathbf{p})}
\end{aligned}
\tag{4.94}
$$

for all basis kets $|\mathbf{x}\rangle$, where we remind the reader of the shorthand

$$\left(\frac{d}{d\mathbf{z}}\right)^{\mathbf{x}} := \left(\frac{d}{dz_1}\right)^{x_1} \cdots \left(\frac{d}{dz_n}\right)^{x_n} \tag{4.95}$$

used to ease notation. Integrate the last line of Eq. 4.94 by parts to obtain

$$\sum_{\mathbf{y}} \frac{1}{\mathbf{y}!} \ |\mathbf{y}\rangle \int_{[0,\infty)^n} d\mathbf{z} \ \left(\frac{d}{d\mathbf{z}}\right)^{\mathbf{x}} [\mathbf{z}^{\mathbf{y}}] \ e^{-\mathbf{z}} \delta(\mathbf{z})$$

$$= \sum_{\mathbf{y}} \frac{1}{\mathbf{y}!} \ |\mathbf{y}\rangle \ \mathbf{x}! \ \delta(\mathbf{y} - \mathbf{x}) \tag{4.96}$$

$$= |\mathbf{x}\rangle$$

which confirms that Eq. 4.93 is a resolution of the identity. $\qquad\square$

We can use the coherent state resolution of the identity (Eq. 4.93) we just constructed to rewrite our formal solution for $|\psi(t)\rangle$ (Eq. 4.59). Applying it twice, we have the following result.

Corollary 6 (Generating function in terms of coherent states). *The generating function can be written in the form*

$$|\psi(t)\rangle = \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \ |\mathbf{z}^f\rangle \ \langle -i\mathbf{p}^f |\hat{U}(t,t_0)|\mathbf{z}^0\rangle \ \langle -i\mathbf{p}^0 |\psi(t_0)\rangle \ e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} \ . \tag{4.97}$$

Proof. Apply Proposition 7 to the formal solution for $|\psi(t)\rangle$ (c.f. Eq. 4.64) twice. $\qquad\square$

The object that appears in the middle of this expression is sufficiently important that it deserves its own name.

Definition 7. *The propagator is defined as the matrix element*

$$U(i\mathbf{p}^f, t; \mathbf{z}^0, t_0) := \left\langle -i\mathbf{p}^f |\hat{U}(t,t_0)|\mathbf{z}^0\right\rangle \ . \tag{4.98}$$

where $\left|-i\mathbf{p}^f\right\rangle$ and $|\mathbf{z}^0\rangle$ are coherent states, and $\hat{U}(t,t_0)$ is the time evolution operator. Usually, we will refer to it using the abbreviated notation $U(i\mathbf{p}^f, \mathbf{z}^0)$.

Now we will construct a coherent state path integral expression for the propagator—this is one of the most important equations in this paper, as it forms the basis of Doi-Peliti path integral calculations.

Proposition 8 (Path integral expression for the propagator). *The propagator $U(i\mathbf{p}^f, t; \mathbf{z}^0, t_0)$ is equal to the path integral*

$$U = \lim_{N\to\infty} \int \prod_{\ell=1}^{N-1} \frac{d\mathbf{z}^\ell d\mathbf{p}^\ell}{(2\pi)^n} \exp\left\{ \sum_{\ell=1}^{N-1} -i\mathbf{p}^\ell \cdot (\mathbf{z}^\ell - \mathbf{z}^{\ell-1}) + \Delta t \ \mathcal{H}(i\mathbf{p}^\ell, \mathbf{z}^{\ell-1}, t_{\ell-1}) \right.$$
$$\left. + \Delta t \ \mathcal{H}(i\mathbf{p}^f, \mathbf{z}^{N-1}, t_{\ell-1}) + i\mathbf{p}^f \cdot \mathbf{z}^{N-1} \right\} \tag{4.99}$$

where $\Delta t := (t - t_0)/N$ and the Hamiltonian kernel $\mathcal{H}$ is defined as

$$\mathcal{H}(i\mathbf{p}, \mathbf{z}, t) := \langle -i\mathbf{p} | \hat{H}(t) | \mathbf{z} \rangle \ e^{-i\mathbf{p} \cdot \mathbf{z}} \tag{4.100}$$

where $|-i\mathbf{p}\rangle$ and $|\mathbf{z}\rangle$ are coherent states with $\mathbf{p}, \mathbf{z} \in \mathbb{R}^n$.

Proof. First write the time evolution operator $U(t, t_0)$ as a product of many time evolution operators using the composition property (Eq. 4.65):

$$\hat{U}(t, t_0) = \hat{U}(t, t_{N-1}) \hat{U}(t_{N-1}, t_{N-2}) \cdots \hat{U}(t_1, t_0) \tag{4.101}$$

where $t_\ell := t_0 + \ell \Delta t$ for $\ell = 0, ..., N$, and $\Delta t := (t - t_0)/N$. Now insert $(N - 1)$ resolutions of the identity to write

$$U = \int \prod_{\ell=1}^{N-1} \frac{d\mathbf{z}^\ell d\mathbf{p}^\ell}{(2\pi)^n} \langle -i\mathbf{p}^f | \hat{U}(t, t_{N-1}) | \mathbf{z}^{\ell-1} \rangle \cdots \langle -i\mathbf{p}^1 | \hat{U}(t_1, t_0) | \mathbf{z}^0 \rangle \ e^{-i \sum_{\ell=1}^{N-1} \mathbf{p}^\ell \cdot \mathbf{z}^\ell} . \tag{4.102}$$

To arrive at our desired path integral, all we must do is compute the matrix elements in the above equation. Assuming that N is large enough that Δt is very small, we have that

$$\hat{U}(t_\ell, t_{\ell-1}) \approx 1 + \hat{H}(t_{\ell-1}) \Delta t \tag{4.103}$$

i.e. $\hat{U}$ is equal to its first order Taylor expansion. Moreover, this inequality becomes exact in the $N \to \infty$ limit. Using this,

$$\langle -i\mathbf{p}^\ell | \hat{U}(t_\ell, t_{\ell-1}) | \mathbf{z}^{\ell-1} \rangle \approx e^{i\mathbf{p}^\ell \cdot \mathbf{z}^{\ell-1}} + \Delta t \langle -i\mathbf{p}^\ell | \hat{H}(t_{\ell-1}) | \mathbf{z}^{\ell-1} \rangle . \tag{4.104}$$

By the definition of the Hamiltonian kernel,

$$\begin{aligned}
\langle -i\mathbf{p}^\ell | \hat{U}(t_\ell, t_{\ell-1}) | \mathbf{z}^{\ell-1} \rangle &\approx e^{i\mathbf{p}^\ell \cdot \mathbf{z}^{\ell-1}} \left[1 + \mathcal{H}(i\mathbf{p}^\ell, \mathbf{z}^{\ell-1}, t_{\ell-1}) \Delta t \right] \\
&\approx e^{i\mathbf{p}^\ell \cdot \mathbf{z}^{\ell-1} + \Delta t \mathcal{H}(i\mathbf{p}^\ell, \mathbf{z}^{\ell-1}, t_{\ell-1})}
\end{aligned} \tag{4.105}$$

where we have again used the fact that Δt is small. Putting all of these matrix elements together, our final coherent state path integral expression for $U(i\mathbf{p}^f, \mathbf{z}^0)$ reads

$$\begin{aligned}
U = \lim_{N \to \infty} \int \prod_{\ell=1}^{N-1} \frac{d\mathbf{z}^\ell d\mathbf{p}^\ell}{(2\pi)^n} \exp \Bigg\{ &\sum_{\ell=1}^{N-1} -i\mathbf{p}^\ell \cdot (\mathbf{z}^\ell - \mathbf{z}^{\ell-1}) + \Delta t \ \mathcal{H}(i\mathbf{p}^\ell, \mathbf{z}^{\ell-1}, t_{\ell-1}) \\
&+ \Delta t \ \mathcal{H}(i\mathbf{p}^f, \mathbf{z}^{N-1}, t_{\ell-1}) + i\mathbf{p}^f \cdot \mathbf{z}^{N-1} \Bigg\}
\end{aligned} \tag{4.106}$$

where the $N \to \infty$ limit must be taken so that the approximation we made in Eq. 4.103 becomes exact. $\qquad \square$

4.3.6   Grassberger-Scheunert creation operators

As we noted in Sec. 4.3.4, the Hermitian conjugate of the annihilation operator $\hat{a}_j$ with respect to the Grassberger-Scheunert product is $\hat{\pi}_j - 1$ for all $j = 1, ..., n$. Motivated by this, we define the Grassberger-Scheunert creation operators.

Definition 8. *The Grassberger-Scheunert creation operators are defined to be*

$$\hat{a}_j^+ := \hat{\pi}_j - 1 \tag{4.107}$$

for all $j = 1, ..., n$.

In the rest of the article, we will take 'creation operator' without qualification to mean one of these operators.

All Hamiltonians we consider may be expressed in terms of creation operators and annihilation operators. For example, the Hamiltonian operator corresponding to the chemical birth-death process (c.f. Sec. 4.2.1) can be shown to read

$$\hat{H} = k(t)\hat{a}^+ - \gamma(t)\hat{a}^+\hat{a} \tag{4.108}$$

which is a specific case of a result we derive later (see Sec. 4.4). Note that this expression is 'normal ordered'—all creation operators are to the left of all annihilation operators. For all (possibly time-dependent) operators $\hat{\mathcal{A}}(t)$ in this form, i.e.

$$\hat{\mathcal{A}}(t) := \sum_{\nu_1,...,\nu_n,\rho_1,...,\rho_n} d_{\rho_1,...,\rho_n}^{\nu_1,...,\nu_n}(t) \, (\hat{a}_1^+)^{\nu_1} \cdots (\hat{a}_n^+)^{\nu_n} (\hat{a}_1)^{\rho_1} \cdots (\hat{a}_n)^{\rho_n} \, , \tag{4.109}$$

coherent state matrix elements are easily evaluated by exploiting that $(\hat{a}_j)^\dagger = \hat{a}_j^+$ and that the coherent states are eigenstates of the annihilation operators. The particular result is the following.

Proposition 9 (Coherent state matrix elements of normal ordered operators)**.** *Let $|\mathbf{z}\rangle$ and $|\mathbf{p}\rangle$ be coherent states, and $\hat{\mathcal{A}}(t)$ be an arbitrary (possibly time-dependent) operator that is normal ordered (i.e. all creation operators are to the left of all annihilation operators). The coherent state matrix element $\langle\mathbf{p}|\hat{\mathcal{A}}(t)|\mathbf{z}\rangle$ can be evaluated to be*

$$\langle\mathbf{p}|\hat{\mathcal{A}}(t)|\mathbf{z}\rangle = e^{\mathbf{p}^*\cdot\mathbf{z}} \sum_{\nu_1,...,\nu_n,\rho_1,...,\rho_n} d_{\rho_1,...,\rho_n}^{\nu_1,...,\nu_n}(t) \, (p_1^*)^{\nu_1} \cdots (p_n^*)^{\nu_n} (z_1)^{\rho_1} \cdots (z_n)^{\rho_n} \, . \tag{4.110}$$

Proof. Use the linearity of the inner product to take the sum out, then use the facts that $(\hat{a}_j)^\dagger = \hat{a}_j^+$, $\hat{a}_k |\mathbf{z}\rangle = z_k |\mathbf{z}\rangle$, and $\langle\mathbf{p}| \hat{a}_k^+ = p_k^* \langle\mathbf{p}|$. Finally, note that $\langle\mathbf{p}|\mathbf{z}\rangle = e^{\mathbf{p}^*\cdot\mathbf{z}}$. $\square$

We will use this result in the calculation sections to derive many Hamiltonian kernels.

4.3.7 Probability distribution and moments

We need some way to extract information (like the transition probability $P(\mathbf{x}, t)$ or factorial moments) from the generating function $|\psi(t)\rangle$. It turns out that we can achieve this using the exclusive product [51] and Grassberger-Scheunert product [54].

Proposition 10 (Extracting transition probabilities and moments from the generating function). *Transition probabilities can be obtained from the generating function using the exclusive product, and factorial moments can be obtained from the generating function using the Grassberger-Scheunert product and the annihilation operators. In particular,*

$$P(\mathbf{x}, t) = \frac{\langle \mathbf{x} | \psi(t) \rangle_{ex}}{\mathbf{x}!} \tag{4.111}$$

and

$$
\begin{aligned}
\langle x_j(t) \rangle &= \langle \mathbf{0} | \hat{a}_j | \psi(t) \rangle \\
\langle x_j(t) x_k(t) \rangle &= \langle \mathbf{0} | \hat{a}_j \hat{a}_k | \psi(t) \rangle \\
\langle x_j(t)[x_j(t) - 1] \rangle &= \langle \mathbf{0} | \hat{a}_j^2 | \psi(t) \rangle \\
\langle x_j(t)[x_j(t) - 1][x_j(t) - 2] \rangle &= \langle \mathbf{0} | \hat{a}_j^3 | \psi(t) \rangle \ .
\end{aligned}
\tag{4.112}
$$

Proof. By the definition of the exclusive product,

$$\frac{\langle \mathbf{x} | \psi(t) \rangle_{ex}}{\mathbf{x}!} = \sum_{\mathbf{y}} P(\mathbf{y}, t) \frac{\langle \mathbf{x} | \mathbf{y} \rangle_{ex}}{\mathbf{x}!} = \sum_{\mathbf{y}} P(\mathbf{y}, t) \delta(\mathbf{x} - \mathbf{y}) = P(\mathbf{x}, t) \ . \tag{4.113}$$

By the explicit definition of the Grassberger-Scheunert product of two basis vectors (c.f. Eq. 4.76), note that $\langle \mathbf{0} | \mathbf{x} \rangle$ for all $\mathbf{x} \in \mathbb{N}^n$. Then

$$
\begin{aligned}
\langle \mathbf{0} | \hat{a}_j | \psi(t) \rangle &= \langle \mathbf{0} | \sum_{\mathbf{x}} P(\mathbf{x}, t) \hat{a}_j | \mathbf{x} \rangle \\
&= \langle \mathbf{0} | \sum_{\mathbf{x}} P(\mathbf{x}, t) x_j | \mathbf{x} - \boldsymbol{\epsilon}_j \rangle \\
&= \sum_{\mathbf{x}} P(\mathbf{x}, t) x_j \langle \mathbf{0} | \mathbf{x} - \boldsymbol{\epsilon}_j \rangle \\
&= \sum_{\mathbf{x}} x_j P(\mathbf{x}, t) \\
&= \langle x_j(t) \rangle \ .
\end{aligned}
\tag{4.114}
$$

The other expectation value formulas can be demonstrated in a similar fashion. $\qquad \square$

With this done, we have all of the machinery necessary to solve the problems identified in Sec. 4.2.

4.4 Monomolecular calculations

In this section, we present the calculations relevant to proving the formulas from Theorems 1 and 2 using the Doi-Peliti approach. First, we derive the Hamiltonian operator and use it to compute the Hamiltonian kernel. Then we evaluate the path integral expression for the propagator $U(i\mathbf{p}^f, \mathbf{z}^0)$. Finally, we use the explicit form of the propagator to derive the transition probability and several moments. We do not explicitly show how to compute the generating function directly from the propagator, because it is very similar to the other calculations.

4.4.1 The Hamiltonian operator and Hamiltonian kernel

We would like an equation equivalent to the CME (Eq. 4.18) that is satisfied by the generating function $|\psi(t)\rangle$ (Eq. 6.32). Doing so involves a straightforward calculation, which we spell out here for the sake of illustration.

Lemma 1. *The Hamiltonian operator corresponding to the monomolecular CME (Eq. 4.18) is*

$$
\begin{aligned}
\hat{H} = {} & \sum_{k=1}^{n} c_{0k}(t) \left[\hat{\pi}_k - 1 \right] - \sum_{k=1}^{n} c_{k0}(t) \left[\hat{\pi}_k - 1 \right] \hat{a}_k \\
& + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[\hat{\pi}_k - \hat{\pi}_j \right] \hat{a}_j \ .
\end{aligned}
\tag{4.115}
$$

Proof. First, take the time derivative of $|\psi(t)\rangle$:

$$
\begin{aligned}
\frac{\partial |\psi\rangle}{\partial t} = {} & \sum_{\mathbf{x}} \frac{\partial P(\mathbf{x}, t)}{\partial t} |\mathbf{x}\rangle \\
= {} & \sum_{\mathbf{x}} \Bigg\{ \sum_{k=1}^{n} c_{0k}(t) \left[P(\mathbf{x} - \boldsymbol{\epsilon}_k, t) - P(\mathbf{x}, t) \right] \\
& + \sum_{k=1}^{n} c_{k0}(t) \left[(x_k + 1) P(\mathbf{x} + \boldsymbol{\epsilon}_k, t) - x_k P(\mathbf{x}, t) \right] \\
& + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[(x_j + 1) P(\mathbf{x} + \boldsymbol{\epsilon}_j - \boldsymbol{\epsilon}_k, t) - x_j P(\mathbf{x}, t) \right] \Bigg\} |\mathbf{x}\rangle
\end{aligned}
\tag{4.116}
$$

where we have used Eq. 4.18. Reindex the sums over $\mathbf{x}$ so that this expression reads

$$\frac{\partial \left| \psi \right\rangle}{\partial t} = \sum_{\mathbf{x}} \left\{ \sum_{k=1}^{n} c_{0k}(t) \left[\left| \mathbf{x} + \boldsymbol{\epsilon}_k \right\rangle - \left| \mathbf{x} \right\rangle \right] \right.$$

$$+ \sum_{k=1}^{n} c_{k0}(t) \left[x_k \left| \mathbf{x} - \boldsymbol{\epsilon}_k \right\rangle - x_k \left| \mathbf{x} \right\rangle \right] \qquad (4.117)$$

$$\left. + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[x_j \left| \mathbf{x} - \boldsymbol{\epsilon}_j + \boldsymbol{\epsilon}_k \right\rangle - x_j \left| \mathbf{x} \right\rangle \right] \right\} P(\mathbf{x}, t) .$$

Using the creation and annihilation operators we defined earlier, the right-hand side can be written as

$$\sum_{\mathbf{x}} \left\{ \sum_{k=1}^{n} c_{0k}(t) \left[\hat{\pi}_k - 1 \right] + \sum_{k=1}^{n} c_{k0}(t) \left[\hat{a}_k - \hat{\pi}_k \hat{a}_k \right] \right.$$

$$\left. + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[\hat{a}_j \hat{\pi}_k - \hat{\pi}_j \hat{a}_j \right] \right\} P(\mathbf{x}, t) \left| \mathbf{x} \right\rangle$$

$$= \left\{ \sum_{k=1}^{n} c_{0k}(t) \left[\hat{\pi}_k - 1 \right] + \sum_{k=1}^{n} c_{k0}(t) \left[\hat{a}_k - \hat{\pi}_k \hat{a}_k \right] \right. \qquad (4.118)$$

$$\left. + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[\hat{a}_j \hat{\pi}_k - \hat{\pi}_j \hat{a}_j \right] \right\} \left| \psi(t) \right\rangle .$$

Comparing this with the definition of the Hamiltonian operator (c.f. Eq. 4.58), we have our result. $\qquad \square$

The Hamiltonian can be written more compactly in terms of the Grassberger-Scheunert creation operators:

Corollary 7. *In terms of the Grassberger-Scheunert creation operator, the Hamiltonian is*

$$\hat{H} = \sum_{k=1}^{n} c_{0k}(t) \hat{a}_k^+ - \sum_{k=1}^{n} c_{k0}(t) \hat{a}_k^+ \hat{a}_k + \sum_{j=1}^{n} \sum_{k=1}^{n} c_{jk}(t) \left[\hat{a}_k^+ - \hat{a}_j^+ \right] \hat{a}_j . \qquad (4.119)$$

Proof. Start with the result above and make the identification $\hat{a}_j^+ = \hat{\pi}_j - 1$. $\qquad \square$

Note that this expression is 'normal ordered'—all creation operators are to the left of all annihilation operators. This allows us to use Proposition 9 to compute the Hamiltonian kernel.

Corollary 8. *The Hamiltonian kernel for the monomolecular CME is*

$$-i\mathcal{H}(i\mathbf{p}^\ell, \mathbf{z}^{\ell-1}, t_{\ell-1}) = \sum_{k=1}^{n} c_{0k}(t_{\ell-1}) p_k^\ell - \sum_{k=1}^{n} c_{k0}(t_{\ell-1}) p_k^\ell z_k^{\ell-1}$$

$$+ \sum_{j=1}^{n}\sum_{k=1}^{n} c_{jk}(t_{\ell-1}) \left[p_k^\ell - p_j^\ell \right] z_j^{\ell-1} \ . \tag{4.120}$$

Proof. Make the identifications $\hat{a}_j^+ \to i p_j^\ell$ and $\hat{a}_j \to z_j^{\ell-1}$ in the Hamiltonian above. $\qquad\square$

4.4.2 Evaluating the propagator path integral

In this section, we will evaluate the path integral expression for the propagator $U(i\mathbf{p}^f, \mathbf{z}^0)$ (Eq. 4.99) given our specific dynamics, which are captured by the Hamiltonian kernel $\mathcal{H}$ (Eq. 4.120).

Lemma 2 (Monomolecular propagator)**.** *The propagator for the monomolecular system is*

$$U(i\mathbf{p}^f, \mathbf{z}^0) = e^{i\mathbf{p}^f \cdot \mathbf{z}(t)} \tag{4.121}$$

where

$$\mathbf{z}(t) := \sum_{k=1}^{n} z_k^0 \mathbf{w}^{(k)}(t) + \boldsymbol{\lambda}(t) \tag{4.122}$$

with $\mathbf{w}^{(k)}(t)$ and $\boldsymbol{\lambda}(t)$ as defined in Theorem 2.

Proof. Begin with the path integral expression for U (Eq. 4.99). Let us first integrate over the p_k^ℓ (where $\ell \in \{1, ..., N-1\}$ and $k \in \{1, ..., n\}$). For fixed ℓ and k, these integrals look like

$$\int_{-\infty}^{\infty} \frac{dp_k^\ell}{2\pi} \exp\left\{ -ip_k^\ell \left[(z_k^\ell - z_k^{\ell-1}) - \Delta t \left(c_{0k}^{\ell-1} - c_{k0}^{\ell-1} z_k^{\ell-1} + \sum_{j=1}^{n} c_{jk}^{\ell-1} z_j^{\ell-1} - c_{kj}^{\ell-1} z_k^{\ell-1} \right) \right] \right\} \tag{4.123}$$

where $c_{jk}^{\ell-1}$ is shorthand for $c_{jk}(t_{\ell-1})$. Using the usual integral representation of the Dirac delta function, these integrals are easily done to obtain $n \cdot (N-1)$ delta function constraints:

$$\delta\left[(z_k^\ell - z_k^{\ell-1}) - \Delta t \left(c_{0k}^{\ell-1} - c_{k0}^{\ell-1} z_k^{\ell-1} + \sum_{j=1}^{n} c_{jk}^{\ell-1} z_j^{\ell-1} - c_{kj}^{\ell-1} z_k^{\ell-1} \right) \right] \ . \tag{4.124}$$

Fortunately, that is *exactly* how many integrals we have left to do. Notice that the constraints force

$$z_k^\ell = z_k^{\ell-1} + \Delta t \left(c_{0k}^{\ell-1} - c_{k0}^{\ell-1} z_k^{\ell-1} + \sum_{j=1}^{n} c_{jk}^{\ell-1} z_j^{\ell-1} - c_{kj}^{\ell-1} z_k^{\ell-1} \right) \tag{4.125}$$

which exactly corresponds to taking an Euler time step given the deterministic dynamics described by the reaction rate equations, Eq. 4.20. What remains of our calculation is to evaluate

$$U = \lim_{N \to \infty} \exp \left\{ \Delta t \, \mathcal{H}(i\mathbf{p}^f, \mathbf{z}^{N-1}, t_{\ell-1}) + i\mathbf{p}^f \cdot \mathbf{z}^{N-1} \right\} \tag{4.126}$$

given Eq. 4.125, the constraint on $\mathbf{z}^{N-1}$ relating it (via $(N-1)$ Euler time steps) to $\mathbf{z}^0$. We have

$$
\begin{aligned}
& i\mathbf{p}^f \cdot \mathbf{z}^{N-1} + \Delta t \, \mathcal{H}(i\mathbf{p}^f, \mathbf{z}^{N-1}, t_{\ell-1}) \\
=& i \sum_{k=1}^{n} p_k^f \left\{ z_k^{N-1} + \Delta t \left[c_{0k}^{N-1} - c_{k0}^{N-1} z_k^{N-1} + \sum_{j=1}^{n} c_{jk}^{N-1} z_j^{N-1} - c_{kj}^{N-1} z_k^{N-1} \right] \right\} \\
=& i \sum_{k=1}^{n} p_k^f z_k^N
\end{aligned}
\tag{4.127}
$$

where we define z_k^N as the result of taking N time steps of length Δt according to Eq. 4.125 given the initial condition z_k^0. In the $N \to \infty$ limit, $z_k^N \to z_k(t)$, where $z_k(t)$ is defined as the kth component of the solution to Eq. 4.20. As described in Sec. 4.2, $\mathbf{z}(t)$ can be decomposed in terms of $\boldsymbol{\lambda}(t)$ and the $\mathbf{w}^{(k)}(t)$.

□

While it may seem that this path integral calculation was completely trivial, that is mostly because we put in the legwork to define and characterize the Grassberger-Scheunert product beforehand. Had we used the exclusive product to construct our path integral, we would either have to perform a hard to justify Doi shift, or deal with extra terms after enforcing the delta function constraints.

Now that we have computed the propagator $U(i\mathbf{p}^f, \mathbf{z}^0)$, we can relate it to the generating function $|\psi(t)\rangle$ using Corollary 6. Then, using Proposition 10, the generating function can be used to compute $P(\mathbf{x}, t)$ and various moments. Because the transition probability and moment calculations are somewhat involved, we first present them for the one species system (i.e. the chemical birth-death process).

4.4.3 One species transition probability derivation

Lemma 3 (One species monomolecular transition probability). *For the single species monomolecular system (i.e. the chemical birth-death process), the transition probability $P(x, t; \xi, t_0)$ is*

$$
\begin{aligned}
P &= \sum_{k=0}^{\min(x,\xi)} \left[\frac{\lambda(t)^{x-k} e^{-\lambda(t)}}{(x-k)!} \right] \left[\binom{\xi}{k} w(t)^k [1 - w(t)]^{\xi-k} \right] \\
&= \mathcal{P}(x, \lambda(t)) \star \mathcal{M}(x, \xi, w(t))
\end{aligned}
\tag{4.128}
$$

where $w(t)$ and $\lambda(t)$ are as defined in Theorem 1.

76

Proof. Recall that, since $P(x, t_0) = \delta(x - \xi)$ for some $\xi \geq 0$,

$$|\psi(t_0)\rangle = |\xi\rangle \ . \tag{4.129}$$

Using Eq. 4.111 and Eq. 4.97, we have

$$
\begin{aligned}
P(x, t; \xi, t_0) &= \frac{1}{x!} \int \frac{dz^f dp^f}{2\pi} \frac{dz^0 dp^0}{2\pi} \langle x | z^f \rangle_{ex} \ U(ip^f, z^0) \langle -ip^0 | \psi(t_0) \rangle \ e^{-ip^0 z^0 - ip^f z^f} \\
&= \frac{1}{x!} \int \frac{dz^f dp^f}{2\pi} \frac{dz^0 dp^0}{2\pi} \left(z^f\right)^x e^{-z^f} e^{ip^f z(t)} (1 + ip^0)^\xi \ e^{-ip^0 z^0 - ip^f z^f} \ .
\end{aligned}
\tag{4.130}
$$

The integral over p^f is easily done:

$$\int_{-\infty}^{\infty} \frac{dp^f}{2\pi} e^{ip^f[z(t) - z^f]} = \delta(z(t) - z^f) \ . \tag{4.131}$$

Enforcing the delta function constraint removes the integral over z^f. Since $z(t) = z^0 w(t) + \lambda(t)$,

$$
\begin{aligned}
P &= \frac{1}{x!} \int \frac{dz^0 dp^0}{2\pi} \left[z^0 w(t) + \lambda(t)\right]^x e^{-[z^0 w(t) + \lambda(t)]} (1 + ip^0)^\xi \ e^{-ip^0 z^0} \\
&= \frac{e^{-\lambda(t)}}{x!} \int \frac{dz^0 dp^0}{2\pi} \left[z^0 w(t) + \lambda(t)\right]^x e^{z^0 [1 - w(t)]} (1 + ip^0)^\xi \ e^{-z^0 [1 + ip^0]} \ .
\end{aligned}
\tag{4.132}
$$

This can be rewritten as

$$
\begin{aligned}
P &= \frac{e^{-\lambda(t)}}{x!} \int \frac{dz^0 dp^0}{2\pi} \left[z^0 w(t) + \lambda(t)\right]^x e^{z^0 [1 - w(t)]} \left(-\frac{d}{dz^0}\right)^\xi e^{-z^0 [1 + ip^0]} \\
&= \frac{e^{-\lambda(t)}}{x!} \int \frac{dz^0 dp^0}{2\pi} \left(\frac{d}{dz^0}\right)^\xi \left\{\left[z^0 w(t) + \lambda(t)\right]^x e^{z^0 [1 - w(t)]}\right\} e^{-z^0 [1 + ip^0]}
\end{aligned}
\tag{4.133}
$$

where we integrated by parts in the second step. The p^0 integral can now be done:

$$\int_{-\infty}^{\infty} \frac{dp^0}{2\pi} e^{-ip^0 z^0} = \delta(z^0) \ . \tag{4.134}$$

We now have

$$P = \frac{e^{-\lambda(t)}}{x!} \int_0^\infty dz^0 \left(\frac{d}{dz^0}\right)^\xi \left\{\left[z^0 w(t) + \lambda(t)\right]^x e^{z^0 [1 - w(t)]}\right\} e^{-z^0} \delta(z^0) \ . \tag{4.135}$$

If we can evaluate the derivative, then we can easily evaluate the integral using the delta function. Using the binomial theorem,

$$\left[z^0 w(t) + \lambda(t)\right]^x = \sum_{k=0}^x \binom{x}{k} w(t)^k \lambda(t)^{x-k} \left(z^0\right)^k \ . \tag{4.136}$$

Since

$$\left(z^0\right)^k e^{z^0[1-w(t)]} = \sum_{j=0}^{\infty} \frac{(z^0)^{j+k} \left[1 - w(t)\right]^j}{j!} \ , \tag{4.137}$$

the derivative of a specific term is

$$\left(\frac{d}{dz^0}\right)^\xi \left\{\left(z^0\right)^k e^{z^0[1-w(t)]}\right\} = \sum_{j=0}^{\infty} \frac{(j+k)(j+k-1)\cdots(j+k-\xi+1)\,(z^0)^{j+k-\xi}\left[1 - w(t)\right]^j}{j!} \ . \tag{4.138}$$

When enforcing the delta function constraint that $z^0 = 0$, all terms will disappear from this series except for the constant term. The constant term is the term with $j + k = \xi$, which reads

$$\frac{\xi!}{(\xi - k)!}[1 - w(t)]^{\xi-k}\theta(\xi - k) \tag{4.139}$$

where the step function θ, defined as

$$\theta(\xi - k) := \begin{cases} 1 & k \leq \xi \\ 0 & k > \xi \end{cases} \tag{4.140}$$

must be there since the result will be zero if $k > \xi$. Hence,

$$\begin{aligned}
P &= \frac{e^{-\lambda(t)}}{x!} \sum_{k=0}^{x} \binom{x}{k} w(t)^k \lambda(t)^{x-k} \frac{\xi!}{(\xi - k)!}[1 - w(t)]^{\xi-k}\theta(\xi - k) \\
&= e^{-\lambda(t)} \sum_{k=0}^{\min(x,\xi)} \binom{\xi}{k} w(t)^k \lambda(t)^{x-k} \frac{1}{(x - k)!}[1 - w(t)]^{\xi-k} \\
&= \sum_{k=0}^{\min(x,\xi)} \left[\frac{\lambda(t)^{x-k}e^{-\lambda(t)}}{(x - k)!}\right] \left[\binom{\xi}{k} w(t)^k[1 - w(t)]^{\xi-k}\right] \\
&= \mathcal{P}(x, \lambda(t)) \star \mathcal{M}(x, \xi, w(t))
\end{aligned} \tag{4.141}$$

as desired. $\qquad\qquad\square$

4.4.4 General transition probability derivation

Lemma 4 (Monomolecular transition probability). *For the general monomolecular system, the transition probability $P(\mathbf{x}, t; \boldsymbol{\xi}, t_0)$ is*

$$P = \mathcal{P}(\mathbf{x}, \boldsymbol{\lambda}(t)) \star \mathcal{M}(\mathbf{x}, \xi_1, \mathbf{w}^{(1)}(t)) \star \cdots \star \mathcal{M}(\mathbf{x}, \xi_n, \mathbf{w}^{(n)}(t)) \tag{4.142}$$

where $\boldsymbol{\lambda}(t)$ and the $\mathbf{w}^{(j)}(t)$ are as defined in Theorem 2.

Proof. The general case proceeds analogously to the one species case. The main difference is that we must do the appropriate multivariable generalization of each of the steps in the

previous subsection (e.g. use the multinomial theorem instead of the binomial theorem). Since $P(\mathbf{x}, t_0) = \delta(\mathbf{x} - \boldsymbol{\xi})$,

$$|\psi(t_0)\rangle = |\boldsymbol{\xi}\rangle \ . \tag{4.143}$$

Using Eq. 4.111 and Eq. 4.97,

$$P(\mathbf{x}, t; \boldsymbol{\xi}, t_0) = \frac{1}{\mathbf{x}!} \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \langle \mathbf{x}|\mathbf{z}^f \rangle_{ex} \ U(i\mathbf{p}^f, \mathbf{z}^0) \langle -i\mathbf{p}^0|\psi(t_0)\rangle \ e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f}$$

$$= \frac{1}{\mathbf{x}!} \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left(\mathbf{z}^f\right)^{\mathbf{x}} e^{-\mathbf{z}^f \cdot \mathbf{1}} \ e^{i\mathbf{p}^f \cdot \mathbf{z}(t)} (\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}} \ e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} \ . \tag{4.144}$$

The integrals over $p_1^f, ..., p_n^f$ yield delta functions:

$$\int \frac{d\mathbf{p}^f}{(2\pi)^n} e^{i\mathbf{p}^f \cdot [\mathbf{z}(t) - \mathbf{z}^f]} = \delta(z_1(t) - z_1^f) \cdots \delta(z_n(t) - z_n^f) = \delta(\mathbf{z}(t) - \mathbf{z}^f) \ . \tag{4.145}$$

Enforcing the delta function constraints removes the integrals over $z_1^f, ..., z_n^f$. Using Eq. 4.122,

$$P = \frac{1}{\mathbf{x}!} \int \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left[\sum_{k=1}^n z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda}\right]^{\mathbf{x}} e^{-[\sum_{k=1}^n z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda}] \cdot \mathbf{1}} \ (\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}} \ e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0}$$

$$= \frac{e^{-|\boldsymbol{\lambda}(t)|}}{\mathbf{x}!} \int \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left[\sum_{k=1}^n z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda}\right]^{\mathbf{x}} e^{\sum_{k=1}^n z_k^0 (1 - |\mathbf{w}^{(k)}|)} \ (\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}} \ e^{-\mathbf{z}^0 \cdot [\mathbf{1} + i\mathbf{p}^0]} \ . \tag{4.146}$$

Reusing the notation we used earlier to denote many derivatives with respect to each variable (Eq. 4.95), we can rewrite this result as

$$P = \frac{e^{-|\boldsymbol{\lambda}(t)|}}{\mathbf{x}!} \int \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left[\sum_{k=1}^n z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda}\right]^{\mathbf{x}} e^{\sum_{k=1}^n z_k^0 (1 - |\mathbf{w}^{(k)}|)} \ \left(-\frac{d}{d\mathbf{z}^0}\right)^{\boldsymbol{\xi}} e^{-\mathbf{z}^0 \cdot [\mathbf{1} + i\mathbf{p}^0]}$$

$$= \frac{e^{-|\boldsymbol{\lambda}(t)|}}{\mathbf{x}!} \int \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left(\frac{d}{d\mathbf{z}^0}\right)^{\boldsymbol{\xi}} \left\{\left[\sum_{k=1}^n z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda}\right]^{\mathbf{x}} e^{\sum_{k=1}^n z_k^0 (1 - |\mathbf{w}^{(k)}|)}\right\} e^{-\mathbf{z}^0 \cdot [\mathbf{1} + i\mathbf{p}^0]} \tag{4.147}$$

where we integrated by parts many times in the second step. The $p_1^0, ..., p_n^0$ integrals can now be done:

$$\int \frac{d\mathbf{p}^0}{(2\pi)^n} e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0} = \delta(z_1^0) \cdots \delta(z_n^0) = \delta(\mathbf{z}^0) \ . \tag{4.148}$$

We now have

$$P = \frac{e^{-|\boldsymbol{\lambda}(t)|}}{\mathbf{x}!} \int d\mathbf{z}^0 \left(\frac{d}{d\mathbf{z}^0}\right)^{\boldsymbol{\xi}} \left\{\left[\sum_{k=1}^n z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda}\right]^{\mathbf{x}} e^{\sum_{k=1}^n z_k^0 (1 - |\mathbf{w}^{(k)}|)}\right\} e^{-\mathbf{z}^0 \cdot \mathbf{1}} \delta(\mathbf{z}^0) \ . \tag{4.149}$$

If we can evaluate the derivative, then we can easily evaluate the integral using the delta function. Recall that

$$\left[\sum_{k=1}^{n} z_k^0 \mathbf{w}^{(k)} + \boldsymbol{\lambda} \right]^{\mathbf{x}} = \left[\sum_{k=1}^{n} z_k^0 w_1^{(k)} + \lambda_1 \right]^{x_1} \cdots \left[\sum_{k=1}^{n} z_k^0 w_n^{(k)} + \lambda_n \right]^{x_n} . \tag{4.150}$$

Using the multinomial theorem,

$$\left[\sum_{k=1}^{n} z_k^0 w_j^{(k)} + \lambda_j \right]^{x_j} = \sum_{v_1^j + \cdots v_{n+1}^j = x_j} \binom{x_j}{v_1^j \cdots v_{n+1}^j} \left[z_1^0 w_j^{(1)} \right]^{v_1^j} \cdots \left[z_n^0 w_j^{(n)} \right]^{v_n^j} [\lambda_j]^{v_{n+1}^j} \tag{4.151}$$

for each $j = 1, ..., n$. Write $|\mathbf{v}_\ell| := v_\ell^1 + \cdots + v_\ell^n$. Putting these multinomial expansions together, our integral now involves computing n expressions of the form

$$\left(\frac{d}{dz_\ell^0} \right)^{\xi_\ell} \left\{ \left[z_\ell^0 \right]^{|\mathbf{v}_\ell|} e^{z_\ell^0 \left(1 - |\mathbf{w}^{(\ell)}| \right)} \right\} \Bigg|_{z_\ell^0 = 0} = \frac{\xi_\ell!}{(\xi_\ell - |\mathbf{v}_\ell|)!} \left(1 - |\mathbf{w}^{(\ell)}| \right)^{\xi_\ell - |\mathbf{v}_\ell|} \theta(\xi_\ell - |\mathbf{v}_\ell|) \tag{4.152}$$

where we have used the result from earlier (Eq. 4.139) to evaluate it. When enforcing the delta function constraint that $z_\ell^0 = 0$ for all $\ell = 1, ..., n$, we get

$$\frac{e^{-|\boldsymbol{\lambda}(t)|}}{\mathbf{x}!} \sum_{v_k^j} \left\{ \prod_{j=1}^{n} \binom{x_j}{v_1^j \cdots v_{n+1}^j} \left[w_j^{(1)} \right]^{v_1^j} \cdots \left[w_j^{(n)} \right]^{v_n^j} [\lambda_j]^{v_{n+1}^j} \frac{\xi_j!}{(\xi_j - |\mathbf{v}_j|)!} \left(1 - |\mathbf{w}^{(j)}| \right)^{\xi_j - |\mathbf{v}_j|} \theta(\xi_j - |\mathbf{v}_j|) \right\} \tag{4.153}$$

for P. This is the final result, but let us rewrite it so that we recover the result from Theorem 1 (Eq. 4.31) of Jahnke and Huisinga's paper. Note that

$$e^{-|\boldsymbol{\lambda}(t)|} \prod_{j=1}^{n} \frac{[\lambda_j]^{v_{n+1}^j}}{v_{n+1}^j!} = \frac{\boldsymbol{\lambda}(t)^{\mathbf{v}_{n+1}}}{\mathbf{v}_{n+1}!} e^{-|\boldsymbol{\lambda}(t)|} = \mathcal{P}(\mathbf{v}_{n+1}, \boldsymbol{\lambda}(t)) . \tag{4.154}$$

Also,

$$\frac{\xi_k! \left(1 - |\mathbf{w}^{(k)}| \right)^{\xi_k - |\mathbf{v}_k|}}{(\xi_k - |\mathbf{v}_k|)!} \theta(\xi_k - |\mathbf{v}_k|) \prod_{j=1}^{n} \frac{\left[w_j^{(k)} \right]^{v_k^j}}{v_k^j!}$$

$$= \frac{\xi_k! \left(1 - |\mathbf{w}^{(k)}| \right)^{\xi_k - |\mathbf{v}_k|}}{(\xi_k - |\mathbf{v}_k|)!} \theta(\xi_k - |\mathbf{v}_k|) \frac{\left[\mathbf{w}^{(k)} \right]^{\mathbf{v}_k}}{\mathbf{v}_k!} \tag{4.155}$$

$$= \mathcal{M}(\mathbf{v}_k, \xi_k, \mathbf{w}^{(k)}) .$$

We are left with

$$P = \sum_{v_k^j} \mathcal{P}(\mathbf{v}_{n+1}, \boldsymbol{\lambda}(t)) \, \mathcal{M}(\mathbf{v}_1, \xi_1, \mathbf{w}^{(1)}) \cdots \mathcal{M}(\mathbf{v}_n, \xi_n, \mathbf{w}^{(n)})$$

$$= \sum_{v_k^j} \mathcal{P}(\mathbf{x} - \mathbf{v}_1 - \cdots - \mathbf{v}_n, \boldsymbol{\lambda}(t)) \, \mathcal{M}(\mathbf{v}_1, \xi_1, \mathbf{w}^{(1)}) \cdots \mathcal{M}(\mathbf{v}_n, \xi_n, \mathbf{w}^{(n)}) \tag{4.156}$$

$$= \mathcal{P}(\mathbf{x}, \boldsymbol{\lambda}(t)) \star \mathcal{M}(\mathbf{x}, \xi_1, \mathbf{w}^{(1)}(t)) \star \cdots \star \mathcal{M}(\mathbf{x}, \xi_n, \mathbf{w}^{(n)}(t))$$

80

which matches Eq. 4.31. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\;$ $\square$

If we wanted to compute the moments of $P(\mathbf{x}, t)$, we could just use Eq. 4.31 and carry out the calculation directly; however, the Doi-Peliti approach offers a way to compute moments which bypasses $P(\mathbf{x}, t)$ completely. In other words, if we are *only* interested in moments, the work from the previous section is unnecessary. Instead, we can use Proposition 10. As with the previous calculation, we will warm up with the one species case before treating the multi-species case.

4.4.5   One species moments derivation

Lemma 5 (One species monomolecular moments). *For the single species monomolecular system (i.e. the chemical birth-death process), the first and second factorial moments are*

$$
\begin{aligned}
\langle x(t) \rangle &= \xi w(t) + \lambda(t) \\
\langle x(t)[x(t) - 1] \rangle &= w(t)^2 \xi(\xi - 1) + 2\lambda(t)w(t)\xi + \lambda(t)^2
\end{aligned}
\tag{4.157}
$$

where $w(t)$ and $\lambda(t)$ are as defined in Theorem 1.

Proof. Using Eq. 4.112,

$$
\begin{aligned}
\langle x(t) \rangle &= \langle 0 | \hat{a} | \psi(t) \rangle \\
&= \int \frac{dz^f dp^f}{2\pi} \frac{dz^0 dp^0}{2\pi} \langle 0 | \hat{a} | z^f \rangle \; U(ip^f, z^0) \langle -ip^0 | \psi(t_0) \rangle \; e^{-ip^0 z^0 - ip^f z^f} \\
&= \int \frac{dz^f dp^f}{2\pi} \frac{dz^0 dp^0}{2\pi} z^f \; e^{ip^f z(t)} (1 + ip^0)^\xi \; e^{-ip^0 z^0 - ip^f z^f} \; .
\end{aligned}
\tag{4.158}
$$

The p^f, z^f, and p^0 integrals can be done as in Sec. 4.4.3, leaving

$$
\langle x(t) \rangle = \int_0^\infty dz^0 \; \left(\frac{d}{dz^0} \right)^\xi \left\{ \left[z^0 w(t) + \lambda(t) \right] e^{z^0} \right\} \; e^{-z^0} \delta(z^0) \; .
\tag{4.159}
$$

The derivative is easily evaluated, and we obtain

$$
\langle x(t) \rangle = \int_0^\infty dz^0 \; \left[\xi w(t) e^{z^0} + z(t) e^{z^0} \right] \; e^{-z^0} \delta(z^0) = \xi w(t) + \lambda(t)
\tag{4.160}
$$

which is just the solution to the one species reaction rate equation with $x(t_0) = \xi$, just as expected. The second factorial moment can be computed in similar fashion:

$$
\begin{aligned}
\langle x(t)[x(t) - 1] \rangle &= \langle 0 | \hat{a}^2 | \psi(t) \rangle \\
&= \int \frac{dz^f dp^f}{2\pi} \frac{dz^0 dp^0}{2\pi} \left(z^f \right)^2 \; e^{ip^f z(t)} (1 + ip^0)^\xi \; e^{-ip^0 z^0 - ip^f z^f} \\
&= \int_0^\infty dz^0 \; \left(\frac{d}{dz^0} \right)^\xi \left\{ \left[z^0 w(t) + \lambda(t) \right]^2 e^{z^0} \right\} \; e^{-z^0} \delta(z^0) \\
&= w(t)^2 \xi(\xi - 1) + 2\lambda(t)w(t)\xi + \lambda(t)^2 \; .
\end{aligned}
\tag{4.161}
$$

$\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad$ $\square$

Higher factorial moments can be computed in exactly the same way.

4.4.6   General moments derivation

Unlike in the one species case, there are many first moments: $\langle x_1(t)\rangle, ..., \langle x_n(t)\rangle$. There are also many second moments. To summarize them usefully, we compute the covariance matrix elements (i.e. $\mathrm{Cov}(x_j, x_\ell) := \langle x_j(t)x_\ell(t)\rangle - \langle x_j(t)\rangle\langle x_\ell(t)\rangle$ for all pairs of j and ℓ).

Lemma 6 (Monomolecular moments). *For the general monomolecular system, the first moments, second factorial moments, and covariance matrix elements are given by*

$$\langle x_j(t)\rangle = \sum_{k=1}^{n} \xi_k w_j^{(k)}(t) + \lambda_j(t) \qquad j = 1, ..., n$$

$$\langle x_j x_\ell\rangle = \sum_{k=1}^{n}\sum_{k'=1}^{n} \xi_k \xi_{k'} w_j^{(k)} w_\ell^{(k')}$$
$$+ \sum_{k=1}^{n} \xi_k \left[w_j^{(k)}\lambda_\ell + w_\ell^{(k)}\lambda_j - w_j^{(k)}w_\ell^{(k)}\right] + \lambda_j\lambda_\ell \qquad j \neq \ell$$

$$\langle x_j(t)[x_j(t) - 1]\rangle = \sum_{k=1}^{n}\sum_{k'=1}^{n} \xi_k \xi_{k'} w_j^{(k)} w_j^{(k')}$$
$$+ \sum_{k=1}^{n} \xi_k \left[2w_j^{(k)}\lambda_j - \left(w_j^{(k)}\right)^2\right] + \lambda_j^2 \qquad j = 1, ..., n$$

$$\mathrm{Cov}(x_j, x_\ell) = \begin{cases} \sum_{k=1}^{n} \xi_k w_j^{(k)}\left[1 - w_j^{(k)}\right] + \lambda_j & j = \ell \\ -\sum_{k=1}^{n} \xi_k w_j^{(k)} w_\ell^{(k)} & j \neq \ell \end{cases}$$

(4.162)

where $\boldsymbol{\lambda}(t)$ and the $\mathbf{w}^{(j)}(t)$ are as defined in Theorem 2.

Proof. Picking a specific x_j and using Eq. 4.112, we have

$$\langle x_j(t)\rangle = \langle\mathbf{0}|\hat{a}_j|\psi(t)\rangle$$
$$= \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \langle\mathbf{0}|\hat{a}_j|\mathbf{z}^f\rangle \; U(i\mathbf{p}^f, \mathbf{z}^0)\langle-i\mathbf{p}^0|\psi(t_0)\rangle \; e^{-i\mathbf{p}^0\cdot\mathbf{z}^0 - i\mathbf{p}^f\cdot\mathbf{z}^f}$$
$$= \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} z_j^f \; e^{i\mathbf{p}^f\cdot\mathbf{z}(t)}(\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}} \; e^{-i\mathbf{p}^0\cdot\mathbf{z}^0 - i\mathbf{p}^f\cdot\mathbf{z}^f} \; .$$

(4.163)

The $\mathbf{p}^f$, $\mathbf{z}^f$, and $\mathbf{p}^0$ integrals can be done as in Sec. 4.4.4, yielding

$$\langle x_j(t)\rangle = \int d\mathbf{z}^0 \; \left(\frac{d}{d\mathbf{z}^0}\right)^{\boldsymbol{\xi}} \left\{ \left[\sum_{k=1}^{n} z_k^0 w_j^{(k)} + \lambda_j\right] e^{\mathbf{z}^0\cdot\mathbf{1}}\right\} \; e^{-\mathbf{z}^0\cdot\mathbf{1}}\delta(\mathbf{z}^0)$$
$$= \sum_{k=1}^{n} \xi_k w_j^{(k)}(t) + \lambda_j(t)$$

(4.164)

which is the *j*th component of the solution to Eq. 4.20 with $\mathbf{x}(t_0) = \boldsymbol{\xi}$.

Let us compute $\langle x_j(t)x_\ell(t)\rangle$ for $j \neq \ell$. To start off,

$$
\begin{aligned}
\langle x_j(t)x_\ell(t)\rangle &= \langle \mathbf{0}|\hat{a}_j\hat{a}_\ell|\psi(t)\rangle \\
&= \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \langle \mathbf{0}|\hat{a}_j\hat{a}_\ell|\mathbf{z}^f\rangle \; U(i\mathbf{p}^f,\mathbf{z}^0)\langle -i\mathbf{p}^0|\psi(t_0)\rangle \; e^{-i\mathbf{p}^0\cdot\mathbf{z}^0 - i\mathbf{p}^f\cdot\mathbf{z}^f} \\
&= \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} z_j^f z_\ell^f \; e^{i\mathbf{p}^f\cdot\mathbf{z}(t)}(\mathbf{1}+i\mathbf{p}^0)^{\boldsymbol{\xi}} \; e^{-i\mathbf{p}^0\cdot\mathbf{z}^0 - i\mathbf{p}^f\cdot\mathbf{z}^f} \; .
\end{aligned}
$$
(4.165)

Proceeding as we just did, we obtain

$$
\begin{aligned}
\langle x_j x_\ell\rangle &= \int d\mathbf{z}^0 \left(\frac{d}{d\mathbf{z}^0}\right)^{\boldsymbol{\xi}} \left\{ \left[\sum_{k=1}^n z_k^0 w_j^{(k)} + \lambda_j\right]\left[\sum_{k'=1}^n z_{k'}^0 w_\ell^{(k')} + \lambda_\ell\right] e^{\mathbf{z}^0\cdot\mathbf{1}}\right\} e^{-\mathbf{z}^0\cdot\mathbf{1}}\delta(\mathbf{z}^0) \\
&= \sum_{k=1}^n \sum_{k'=1}^n \xi_k \xi_{k'} w_j^{(k)} w_\ell^{(k')} + \sum_{k=1}^n \xi_k \left[w_j^{(k)}\lambda_\ell + w_\ell^{(k)}\lambda_j - w_j^{(k)} w_\ell^{(k)}\right] + \lambda_j \lambda_\ell \; .
\end{aligned}
$$
(4.166)

For the similar case $j = \ell$, we obtain

$$
\begin{aligned}
\langle x_j(t)[x_j(t)-1]\rangle &= \int d\mathbf{z}^0 \left(\frac{d}{d\mathbf{z}^0}\right)^{\boldsymbol{\xi}} \left\{ \left[\sum_{k=1}^n z_k^0 w_j^{(k)} + \lambda_j\right]^2 e^{\mathbf{z}^0\cdot\mathbf{1}}\right\} e^{-\mathbf{z}^0\cdot\mathbf{1}}\delta(\mathbf{z}^0) \\
&= \sum_{k=1}^n \sum_{k'=1}^n \xi_k \xi_{k'} w_j^{(k)} w_j^{(k')} + \sum_{k=1}^n \xi_k \left[2w_j^{(k)}\lambda_j - \left(w_j^{(k)}\right)^2\right] + \lambda_j^2 \; .
\end{aligned}
$$
(4.167)

Putting these results together, we find that the covariance of x_j and x_ℓ is

$$
\mathrm{Cov}(x_j, x_\ell) = \begin{cases} \sum_{k=1}^n \xi_k w_j^{(k)}\left[1 - w_j^{(k)}\right] + \lambda_j & j = \ell \\ -\sum_{k=1}^n \xi_k w_j^{(k)} w_\ell^{(k)} & j \neq \ell \end{cases} \; .
$$
(4.168)

Hence, we have recovered the moment results from Sec. 4.2 of Jahnke and Huisinga. $\qquad\square$

4.5 Birth-death-autocatalysis calculations

In this section, we present the calculations relevant to proving the formulas from Theorem 3 on the birth-death-autocatalysis system. First, we present the Hamiltonian operator and kernel. Then we evaluate the path integral expression for the propagator $U(i\mathbf{p}^f, \mathbf{z}^0)$. Finally, we use the explicit form of the propagator to derive the transition probability, and several interesting limiting forms of it. We do not explicitly show how to compute the generating function directly from the propagator, because it is very similar to the other calculations.

4.5.1 Evaluating the propagator

We can straightforwardly go from the CME (Eq. 4.36) to the Hamiltonian operator and Hamiltonian kernel.

Lemma 7. *The Hamiltonian operator corresponding to the birth-death-autocatalysis CME (Eq. 4.36) is*

$$\hat{H} = \hat{a}^+ \left[k + (c - \gamma)\hat{a} + c\ \hat{a}^+\hat{a} \right] \ . \tag{4.169}$$

Proof. Starting with Eq. 4.36, follow the argument from Lemma 1, and then substitute in the Grassberger-Scheunert creation operator. □

Corollary 9. *The Hamiltonian kernel for the birth-death-autocatalysis CME is*

$$\mathcal{H}(ip, z, t) = ip\left[k + (c - \gamma)z\right] - c\ p^2 z \ . \tag{4.170}$$

Proof. Make the identifications $\hat{a}^+ \to ip$ and $\hat{a} \to z$ in the Hamiltonian above. □

Now we must compute the propagator, a calculation which turns out to be somewhat involved.

Lemma 8 (Birth-death-autocatalysis propagator)**.** *The propagator for the birth-death-autocatalysis system is*

$$U(ip_f, z_0) = \exp\left\{ iz_0 q(t) + i \int_{t_0}^{t} k(s)q(t - s + t_0)\ ds \right\} \tag{4.171}$$

where $q(s)$ is as in Theorem 3.

Proof. The path integral expression for the propagator $U(ip_f, z_0)$ is

$$U = \lim_{N \to \infty} \int \prod_{\ell=1}^{N-1} \frac{dz_\ell dp_\ell}{2\pi} \exp\left\{ \sum_{\ell=1}^{N-1} -ip_\ell(z_\ell - z_{\ell-1}) + \Delta t\mathcal{H}(ip_\ell, z_{\ell-1}, t_{\ell-1}) \right.$$
$$\left. + \Delta t\mathcal{H}(ip_f, z_{N-1}, t_{\ell-1}) + ip_f z_{N-1} \right\} \tag{4.172}$$

where we have used slightly different notation than before since there is only one chemical species. In order to evaluate this path integral, first integrate over each z_ℓ, and then integrate over each p_ℓ. Collecting terms containing z_ℓ, the integral over each z_ℓ looks like

$$\int_0^{\infty} \frac{dz_\ell}{2\pi}\ \exp\left\{ z_\ell \left[-c_\ell \Delta t\ p_{\ell+1}^2 + i(c_\ell - \gamma_\ell)\Delta t\ p_{\ell+1} - i(p_\ell - p_{\ell+1}) \right] \right\}$$
$$= \frac{1}{2\pi i} \frac{1}{(p_\ell - p_{\ell+1}) - \Delta t \left[(c_\ell - \gamma_\ell)\ p_{\ell+1} + ic_\ell\ p_{\ell+1}^2 \right]}\ . \tag{4.173}$$

The integrals over p_ℓ can now be done—but they must be done in a specific order. Do the integral over p_{N-1}, then p_{N-2}, and so on, until the integral over p_1 has been done. Each of these integrals is schematically

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} dp_\ell\ \frac{f(p_\ell)}{(p_\ell - p_{\ell+1}) - \Delta t \left[(c_\ell - \gamma_\ell)\ p_{\ell+1} + ic_\ell\ p_{\ell+1}^2 \right]} \tag{4.174}$$

where the function $f(p_\ell)$ has no poles. This means that each integral can be evaluated using Cauchy's integral formula, so that the net effect of doing them is to enforce the $(N-1)$ constraints

$$p_\ell = p_{\ell+1} + \Delta t \left[(c_\ell - \gamma_\ell) \, p_{\ell+1} + ic_\ell \, p_{\ell+1}^2 \right] \tag{4.175}$$

on the p_ℓ for $\ell = 1, ..., N-1$. There are no more integrals to do, so all that remains is to evaluate what's left of the propagator using these constraints. Eq. 4.175 looks like an Euler time step, although it is 'backwards'—we go from $p_{\ell+1}$ to p_ℓ instead of the other way around. Define $q_{N-\ell} := p_\ell$ so that it reads

$$q_{N-\ell} = q_{N-\ell-1} + \Delta t \left[(c_\ell - \gamma_\ell) \, q_{N-\ell-1} + ic_\ell \, q_{N-\ell-1}^2 \right] \ . \tag{4.176}$$

Choosing $\ell = N - n$, we find

$$q_n = q_{n-1} + \Delta t \left[(c_{N-n} - \gamma_{N-n}) \, q_{n-1} + ic_{N-n} \, q_{n-1}^2 \right] \ . \tag{4.177}$$

This corresponds to dynamics

$$\dot{q}(s) = \left[c(t - s + t_0) - \gamma(t - s + t_0) \right] q(s) + ic(t - s + t_0) \, q(s)^2 \tag{4.178}$$

where $s \in [t_0, t]$ and $q(t_0) = p_f$. As can be verified by substitution, Eq. 4.178 is solved by

$$q(s) = \frac{w(s)}{\frac{1}{p_f} - i \int_{t_0}^{s} c(t - t' + t_0) w(t') \ dt'} \tag{4.179}$$

where $w(t)$ is the solution to

$$\dot{w}(s) = \left[c(t - s + t_0) - \gamma(t - s + t_0) \right] w(s) \tag{4.180}$$

with $w(t_0) = 1$ (c.f. Eq. 4.30), i.e.

$$w(s) = e^{\int_{t_0}^{s} c(t - t' + t_0) - \gamma(t - t' + t_0) \ dt'} \ . \tag{4.181}$$

The continuous limit of Eq. 4.175 is then $p(s) := q(t - s + t_0)$. With that done, the propagator with most terms integrated out reads

$$U(ip_f, z_0) = \lim_{N \to \infty} \exp \left\{ i \sum_{\ell=1}^{N-1} k_{\ell-1} p_\ell \, \Delta t + ip_1 z_0 + \Delta t \left[ip_1 (c_0 - \gamma_0) z_0 - c_0 p_1^2 z_0 \right] \right\} \ . \tag{4.182}$$

The term on the right is just another Euler time step, so we can write it as

$$iz_0 \left\{ p_1 + \Delta t \left[p_1 (c_0 - \gamma_0) - c_0 p_1^2 \right] \right\} = iz_0 p_0 \tag{4.183}$$

where we define

$$p_0 := p_1 + \Delta t \left[p_1 (c_0 - \gamma_0) - c_0 p_1^2 \right] \ . \tag{4.184}$$

In the limit as $N \to \infty$, $p_0 \to p(t_0) = q(t)$. The term on the left is just a Riemann sum:

$$\sum_{\ell=1}^{N-1} k_{\ell-1} p_\ell \, \Delta t \approx \int_{t_0}^{t} k(s) p(s) \, ds = \int_{t_0}^{t} k(s) q(t - s + t_0) \, ds \ . \tag{4.185}$$

Hence, our final answer for the propagator U is

$$U(ip_f, z_0) = \exp\left\{ iz_0 q(t) + i \int_{t_0}^{t} k(t - s + t_0) q(s) \, ds \right\} \tag{4.186}$$

where we have reparameterized the integral on the right to swap s and $(t - s + t_0)$. $\qquad\square$

As an aside, we note that this calculation closely resembles the Martin-Siggia-Rose-Janssen-De Dominicis path integral computation from our earlier paper [77]: in particular, many applications of Cauchy's integral formula and another 'backwards' Euler time step constraint are both involved.

4.5.2 Deriving the transition probability

As in Sec. 4.4.3 and 4.4.4, we will use the propagator derived in the previous section to derive an expression for the transition probability $P(x, t; \xi, t_0)$.

Lemma 9 (Birth-death-autocatalysis transition probability). *For the birth-death-autocatalysis system, the transition probability $P(x, t; \xi, t_0)$ is*

$$P(x, t; \xi, t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dp_f \, \frac{[1 + iq(t)]^{\xi} \, e^{i \int_{t_0}^{t} k(t - s + t_0) q(s) ds}}{(1 + ip_f)^{x+1}} \tag{4.187}$$

where $q(s)$ is as in Theorem 3.

Proof. Since $P(x, t_0) = \delta(x - \xi)$, we have $|\psi_0\rangle = |\xi\rangle$. Using Eq. 4.111 and Eq. 4.97,

$$
\begin{aligned}
P(x, t; \xi, t_0) &= \frac{1}{x!} \int \frac{dz_f dp_f}{2\pi} \frac{dz_0 dp_0}{2\pi} \, \langle x | z_f \rangle_{ex} \, U(ip_f, z_0) \, \langle -ip_0 | \psi(t_0) \rangle \, e^{-ip_0 z_0 - ip_f z_f} \\
&= \frac{1}{x!} \int \frac{dz_f dp_f}{2\pi} \frac{dz_0 dp_0}{2\pi} \, (z_f)^x \, e^{-z_f} \, e^{iz_0 q(t) + i \int_{t_0}^{t} k(t - s + t_0) q(s) \, ds} (1 + ip_0)^{\xi} \, e^{-ip_0 z_0 - ip_f z_f} \ .
\end{aligned}
\tag{4.188}
$$

The integral over z_0 is

$$\int_0^{\infty} \frac{dz_0}{2\pi} \, e^{-iz_0 [p_0 - q(t)]} = \frac{1}{2\pi i} \frac{1}{p_0 - q(t)} \ . \tag{4.189}$$

The integral over p_0 can be performed using Cauchy's integral formula:

$$\frac{1}{2\pi i} \int_{-\infty}^{\infty} dp_0 \, \frac{(1 + ip_0)^{\xi}}{p_0 - q(t)} = [1 + iq(t)]^{\xi} \ . \tag{4.190}$$

86

The integral over z_f can be recognized as a Laplace transform:

$$\int_0^\infty dz_f \ (z_f)^x \ e^{-z_f[1+ip_f]} = \frac{x!}{(1+ip_f)^{x+1}} \ . \tag{4.191}$$

Putting these together, we obtain the desired result. □

We will leave our solution in this form, since it is difficult to evaluate the contour integral without knowing the explicit time-dependence of the rates. In the next few sections, we will examine a few special cases.

4.5.3 Time-independent rates

Lemma 10 (Birth-death-autocatalysis transition probability for time-independent rates). *Suppose the parameters k, γ, and c are all time-independent and non-zero. Then the transition probability can be rewritten as*

$$P = \left(\frac{\frac{\gamma}{c}-1}{\frac{\gamma}{c}-w}\right)^{k/c} \frac{(1-w)^{x-\xi}}{\left(\frac{\gamma}{c}-w\right)^x} \times$$

$$\times \sum_{j=0}^{\xi} \binom{\xi}{j} \frac{(j+k/c)_x}{x!} \left[1-\frac{\gamma}{c}w\right]^{\xi-j} \left[\frac{w\left(\frac{\gamma}{c}-1\right)^2}{\frac{\gamma}{c}-w}\right]^j \tag{4.192}$$

where $(y)_x := (y)(y+1)\cdots(y+x-1)$ is the Pochhammer symbol/rising factorial, and where $w(t) = e^{-(\gamma-c)(t-t_0)}$.

Proof. In this case, $q(t)$ reads

$$\dot{q} = [c-\gamma]\, q + ic\, q^2$$

$$q(t) = \frac{e^{(c-\gamma)T}}{\frac{1}{p_f} - i\frac{c}{c-\gamma}\left[e^{(c-\gamma)T}-1\right]} = \frac{w(t)}{\frac{1}{p_f} - i\frac{c}{c-\gamma}\left[w(t)-1\right]} \tag{4.193}$$

where $T := t-t_0$. We have

$$\int_{t_0}^t q(s)\ ds = \frac{i}{c}\log\left\{1 - \frac{ic}{c-\gamma}\left[e^{(c-\gamma)T}-1\right]p_f\right\} \tag{4.194}$$

so that the convolution term from the propagator reads

$$e^{ik\int_{t_0}^t q(s)\ ds} = \frac{1}{\left[1 - \frac{ic}{c-\gamma}\left[e^{(c-\gamma)T}-1\right]p_f\right]^{k/c}} = \frac{1}{[1-iB(t)p_f]^{k/c}} \tag{4.195}$$

where we define

$$B(t) := \frac{c}{c-\gamma}\left[w(t)-1\right]\ . \tag{4.196}$$

87

It is important to note that Eq. 4.195 has no poles in the upper half-plane (the region around which we are integrating), regardless of whether $c - \gamma > 0$, $c - \gamma < 0$, or $c = \gamma$. Next,

$$1 + iq(t) = 1 + \frac{iw(t)p_f}{1 - iB(t)p_f} = \left[1 - \frac{w(t)}{B(t)}\right] + \frac{w(t)}{B(t)}\frac{1}{[1 - iB(t)p_f]} \qquad (4.197)$$

so that

$$[1 + iq(t)]^\xi = \sum_{j=0}^{\xi} \binom{\xi}{j} \left[1 - \frac{w(t)}{B(t)}\right]^{\xi-j} \left(\frac{w(t)}{B(t)}\right)^j \frac{1}{[1 - iB(t)p_f]^j}. \qquad (4.198)$$

Putting all these results together, our expression for the transition probability is

$$P = \sum_{j=0}^{\xi} \binom{\xi}{j} \left[1 - \frac{w(t)}{B(t)}\right]^{\xi-j} \left(\frac{w(t)}{B(t)}\right)^j \frac{1}{x! \, i^x}\left\{\frac{x!}{2\pi i}\int_{-\infty}^{\infty} dp_f \frac{1}{[1 - iB(t)p_f]^{j+k/c}}\frac{1}{(p_f - i)^{x+1}}\right\}. \qquad (4.199)$$

Since

$$\frac{d^x}{dp_f^x}\left[\frac{1}{[1 - iB(t)p_f]^{j+k/c}}\right]_{p=i} = \frac{i^x B(t)^x}{[1 + B(t)]^{j+k/c+x}}\left(j + \frac{k}{c}\right)\left(j + \frac{k}{c} + 1\right)\cdots\left(j + \frac{k}{c} + x - 1\right) \qquad (4.200)$$

we have

$$P = \sum_{j=0}^{\xi} \binom{\xi}{j}\left[1 - \frac{w(t)}{B(t)}\right]^{\xi-j}\left(\frac{w(t)}{B(t)}\right)^j \frac{(j + k/c)_x}{x!}\frac{B(t)^x}{[1 + B(t)]^{j+k/c+x}}$$

$$= \left(\frac{1 - \frac{\gamma}{c}}{w - \frac{\gamma}{c}}\right)^{k/c}\left(\frac{w - 1}{w - \frac{\gamma}{c}}\right)^x \sum_{j=0}^{\xi}\binom{\xi}{j}\frac{(j + k/c)_x}{x!}\left[1 - \left(1 - \frac{\gamma}{c}\right)\frac{w}{w - 1}\right]^{\xi-j}\left[\frac{w\left(1 - \frac{\gamma}{c}\right)^2}{(w-1)(w - \frac{\gamma}{c})}\right]^j$$

$$= \left(\frac{\frac{\gamma}{c} - 1}{\frac{\gamma}{c} - w}\right)^{k/c}\frac{(1 - w)^{x-\xi}}{\left(\frac{\gamma}{c} - w\right)^x} \times$$

$$\times \sum_{j=0}^{\xi}\binom{\xi}{j}\frac{(j + k/c)_x}{x!}\left[1 - \frac{\gamma}{c}w\right]^{\xi-j}\left[\frac{w\left(\frac{\gamma}{c} - 1\right)^2}{\frac{\gamma}{c} - w}\right]^j \qquad (4.201)$$

where $(y)_x := (y)(y+1)\cdots(y+x-1)$ is the Pochhammer symbol/rising factorial. This can also be written in terms of the hypergeometric function $_2F_1(a, b; c; x)$. $\qquad \square$

4.5.4 Binomial, Poisson, and negative binomial special cases

Corollary 2. Return to the original contour integral for time-dependent rates (Eq. 4.40), and set $k = c = 0$, but leave the time-dependence of $\gamma(t)$ arbitrary. We have

$$w(t) := \exp\left[-\int_{t_0}^{t}\gamma(t')dt'\right] \qquad (4.202)$$

$$q(t) = w(t)p_f \tag{4.203}$$

$$P(x,t;\xi,t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dp_f \; \frac{[1 + iw(t)p_f]^\xi}{(1 + ip_f)^{x+1}} \; . \tag{4.204}$$

The function in the numerator has no poles, so the contour integral can easily be evaluated using Cauchy's integral formula. The result is

$$P(x,t;\xi,t_0) = \binom{\xi}{x} [w(t)]^x [1 - w(t)]^{\xi-x} \tag{4.205}$$

for $x \leq \xi$ and 0 otherwise, i.e. a binomial distribution.

Return to the original contour integral for time-dependent rates (Eq. 4.40), and set $\gamma = c = 0$, but leave the time-dependence of $k(t)$ arbitrary. We have

$$\lambda(t) := \int_{t_0}^{t} k(t')dt' \tag{4.206}$$

$$q(t) = p_f \tag{4.207}$$

$$P(x,t;\xi,t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dp_f \; \frac{e^{i\lambda(t)p_f}}{(1 + ip_f)^{x+1-\xi}} \; . \tag{4.208}$$

This contour integral can be evaluated using either Cauchy's integral formula or a table of integrals (c.f. Gradshteyn and Ryzhik [78] ET I 118(3), in section 3.382, on pg. 365). The result is

$$P(x,t;\xi,t_0) = \frac{\lambda(t)^{x-\xi}e^{-\lambda(t)}}{(x - \xi)!} \tag{4.209}$$

for $x \geq \xi$ and 0 otherwise, i.e. a (shifted) Poisson distribution.

Return to the original contour integral for time-dependent rates (Eq. 4.40), and set $k = \gamma = 0$, but leave the time-dependence of $c(t)$ arbitrary. In this case, we will define $w(t)$ differently from before as

$$w(t) := \exp\left[-\int_{t_0}^{t} c(t')dt' \right] \tag{4.210}$$

i.e. as the reciprocal of what we previously called $w(t)$. This is to match the result from Jahnke and Huisinga. Now we have

$$q(t) = \frac{w(t)^{-1}}{\frac{1}{p_f} - i\left[w(t)^{-1} - 1\right]} \tag{4.211}$$

$$P(x,t;\xi,t_0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} dp_f \; \frac{1}{(1 + ip_f)^{x-\xi+1}} \frac{1}{[1 - i(w(t)^{-1} - 1)p_f]^\xi} \; . \tag{4.212}$$

The term on the right has no poles in the upper half-plane, so we can evaluate it using Cauchy's formula to find

$$P(x,t;\xi,t_0) = \binom{x-1}{\xi-1} [w(t)]^\xi [1 - w(t)]^{x-\xi} \tag{4.213}$$

which is nonzero only for $x \geq \xi$, i.e. we have a shifted negative binomial distribution. □

4.6 Zero and first order calculations

In this section, we sketch the calculations necessary for proving the formulas from Theorem 4 on a system with arbitrary combinations of zero and first order reactions. First, we present the Hamiltonian operator and kernel. Then we sketch how the path integral expression for the propagator $U(i\mathbf{p}^f, \mathbf{z}^0)$ may be evaluated. Finally, we use the explicit form of the propagator to derive the transition probability and generating function.

Writing down the CME directly is difficult; however, we *do* know that a CME involving only zero or first order reactions only as terms proportional to Grassberger-Scheunert creation operators, and has terms with at most one annihilation operator. We can assume it takes the form (c.f. Eq. 4.51)

$$
\hat{H}(t) = \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t) \left(\hat{a}_1^+\right)^{\nu_1} \cdots \left(\hat{a}_n^+\right)^{\nu_n}
$$
$$
+ \sum_k \sum_{\nu_1,\ldots,\nu_n} \beta_{\nu_1,\ldots,\nu_n}^k(t) \left(\hat{a}_1^+\right)^{\nu_1} \cdots \left(\hat{a}_n^+\right)^{\nu_n} \hat{a}_k
$$
(4.214)

for some coefficients $\alpha_{\nu_1,\ldots,\nu_n}(t)$ and $\beta_{\nu_1,\ldots,\nu_n}^j(t)$ that are determined by the details of one's list of reactions. The corresponding Hamiltonian kernel is

$$
\mathcal{H}(i\mathbf{p}, \mathbf{z}, t) = \langle -i\mathbf{p}|\hat{H}(t)|\mathbf{z}\rangle
$$
$$
= \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t) (ip_1)^{\nu_1} \cdots (ip_n)^{\nu_n}
$$
$$
+ \sum_k \sum_{\nu_1,\ldots,\nu_n} \beta_{\nu_1,\ldots,\nu_n}^k(t) (ip_1)^{\nu_1} \cdots (ip_n)^{\nu_n} z_k \ .
$$
(4.215)

Now we can compute the propagator corresponding to this kernel.

Lemma 11 (Zero and first order reactions propagator)**.** *The propagator for the system with arbitrary combinations of zero and first order reactions is*

$$
U = \exp\left\{ i\, \mathbf{z}^0 \cdot \mathbf{q}(t) + \int_{t_0}^t \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t-s+t_0) \left[iq_1(s)\right]^{\nu_1} \cdots \left[iq_n(s)\right]^{\nu_n} \, ds \right\}
$$
(4.216)

where $q(s)$ is as defined in Theorem 4.

Proof. We will only sketch this proof, because the argument is exactly the same as the one presented in Lemma 8—the notation is just more cluttered, because we are now dealing with a mult-species system and an arbitrarily large list of reactions. One may notice, from a careful look at that prior argument, that its success did not depend on the detailed features of the birth-death-autocatalysis system at all; it only depended on the Hamiltonian containing terms at most first order in annihilation operators (i.e. no terms like $\hat{a}_j \hat{a}_k$ or $(\hat{a}_j)^6$ appear).

Since this is also true in the current case, we can rerun that argument to find that the propagator can be written in terms of the solution $\mathbf{q}(t)$ to

$$\dot{q}_j(s) = -i \sum_{\nu_1,\ldots,\nu_n} \beta^j_{\nu_1,\ldots,\nu_n}(t-s+t_0) \, [iq_1(s)]^{\nu_1} \cdots [iq_n(s)]^{\nu_n} \tag{4.217}$$

satisfying the initial condition $q_j(t_0) = p^f_j$. As before, the final propagator has two terms. There is one term that comes from $\mathbf{p}^0 = \mathbf{q}(t)$ coupling to $\mathbf{z}^0$, and another term (due to the terms in the Hamiltonian involving no annihilation operators) that becomes a convolution integral.

$\square$

Next, we will derive the transition probability.

Lemma 12 (Zero and first order reactions transition probability). *The transition probability for the system with arbitrary combinations of zero and first order reactions is*

$$P = \int_{\mathbb{R}^n} \frac{d\mathbf{p}^f}{(2\pi)^n} \, \frac{[\mathbf{1} + i\mathbf{q}(t)]^{\boldsymbol{\xi}} \, e^{\int_{t_0}^t \sum \alpha_{\nu_1,\ldots,\nu_n}(t-s+t_0)[iq_1(s)]^{\nu_1}\cdots[iq_n(s)]^{\nu_n} \, ds}}{(\mathbf{1} + i\mathbf{p}^f)^{\mathbf{x}+\mathbf{1}}} \tag{4.218}$$

where $q(s)$ is as in Theorem 4.

Proof. Begin with the expression for the generating function in terms of the propagator U (c.f. Corollary 6). As in Sec. 4.4.4, we have

$$\begin{aligned}
P =& \frac{1}{\mathbf{x}!} \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left\langle \mathbf{x} \middle| \mathbf{z}^f \right\rangle_{ex} U(i\mathbf{p}^f, \mathbf{z}^0) \left\langle -i\mathbf{p}^0 \middle| \psi(t_0) \right\rangle \, e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} \\
=& \frac{1}{\mathbf{x}!} \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \left(\mathbf{z}^f \right)^{\mathbf{x}} e^{-\mathbf{z}^f \cdot \mathbf{1}} \, e^{i\mathbf{z}^0 \cdot \mathbf{q}(t)} (\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}} \\
& \times \; e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} e^{\int_{t_0}^t \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t-s+t_0)[iq_1(s)]^{\nu_1}\cdots[iq_n(s)]^{\nu_n} \, ds} .
\end{aligned} \tag{4.219}$$

The integral over $\mathbf{z}^0$ is

$$\int d\mathbf{z}^0 \, e^{-i\mathbf{z}^0 \cdot (\mathbf{p}^0 - \mathbf{q}(t))} = \frac{1}{i^n (\mathbf{p}^0 - \mathbf{q}(t))^{\mathbf{1}}} , \tag{4.220}$$

the integral over $\mathbf{p}^0$ is

$$\int \frac{d\mathbf{p}^0}{(2\pi i)^n} \frac{(\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}}}{(\mathbf{p}^0 - \mathbf{q}(t))^{\mathbf{1}}} = (\mathbf{1} + i\mathbf{q}(t))^{\boldsymbol{\xi}} , \tag{4.221}$$

and the integral over $\mathbf{z}^f$ is

$$\int d\mathbf{z}^f \, \frac{\left(\mathbf{z}^f \right)^{\mathbf{x}}}{\mathbf{x}!} e^{-\mathbf{z}^f \cdot (\mathbf{1} + i\mathbf{p}^f)} = \frac{1}{(\mathbf{1} + i\mathbf{p}^f)^{\mathbf{x}+\mathbf{1}}} , \tag{4.222}$$

leaving only the desired integral. $\square$

Finally, we will derive the generating function.

Lemma 13 (Zero and first order reactions generating function)**.** *The generating function for the system with arbitrary combinations of zero and first order reactions is*

$$\psi(\mathbf{g}, t) = [\mathbf{1} + i\mathbf{q}(t)]^{\boldsymbol{\xi}} \times$$
$$\times \left. e^{\int_{t_0}^{t} \sum \alpha_{\nu_1, \dots, \nu_n}(t-s+t_0)[iq_1(s)]^{\nu_1} \cdots [iq_n(s)]^{\nu_n} \, ds} \right|_{\mathbf{p}^f = -i(\mathbf{g}-\mathbf{1})} \tag{4.223}$$

where $q(s)$ is as in Theorem 4.

Proof. Begin again with the expression for the generating function in terms of the propagator U (c.f. Corollary 6). For proving this result, it is convenient to switch over to analytic notation, in which

$$\begin{aligned}
|\psi(t)\rangle &\to \psi(\mathbf{g}, t) \\
|\mathbf{x}\rangle &\to \mathbf{g}^{\mathbf{x}} \\
|\mathbf{z}\rangle &\to e^{\mathbf{z} \cdot (\mathbf{g}-\mathbf{1})} \ .
\end{aligned} \tag{4.224}$$

In particular, we have

$$\psi = \int \frac{d\mathbf{z}^f \, d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \ e^{\mathbf{z}^f \cdot (\mathbf{g}-\mathbf{1})} e^{i\mathbf{z}^0 \cdot \mathbf{q}(t)} (\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}} \tag{4.225}$$
$$\times \ e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} e^{\int_{t_0}^{t} \sum_{\nu_1, \dots, \nu_n} \alpha_{\nu_1, \dots, \nu_n}(t-s+t_0)[iq_1(s)]^{\nu_1} \cdots [iq_n(s)]^{\nu_n} \, ds} \ .$$

The integral over $\mathbf{z}^0$ is

$$\int d\mathbf{z}^0 \ e^{-i\mathbf{z}^0 \cdot (\mathbf{p}^0 - \mathbf{q}(t))} = \frac{1}{i^n (\mathbf{p}^0 - \mathbf{q}(t))^{\mathbf{1}}} \ , \tag{4.226}$$

the integral over $\mathbf{p}^0$ is

$$\int \frac{d\mathbf{p}^0}{(2\pi i)^n} \frac{(\mathbf{1} + i\mathbf{p}^0)^{\boldsymbol{\xi}}}{(\mathbf{p}^0 - \mathbf{q}(t))^{\mathbf{1}}} = (\mathbf{1} + i\mathbf{q}(t))^{\boldsymbol{\xi}} \ , \tag{4.227}$$

and the integral over $\mathbf{z}^f$ is

$$\int d\mathbf{z}^f \ e^{-\mathbf{z}^f \cdot \left[-(\mathbf{g}-\mathbf{1}) + i\mathbf{p}^f \right]} = \frac{1}{\left[-(\mathbf{g}-\mathbf{1}) + i\mathbf{p}^f \right]^{\mathbf{1}}} \ , \tag{4.228}$$

leaving only the integral

$$\int \frac{d\mathbf{p}^f}{(2\pi i)^n} \frac{[\mathbf{1} + i\mathbf{q}(t)]^{\boldsymbol{\xi}} e^{\int_{t_0}^{t} \sum_{\nu_1, \dots, \nu_n} \alpha_{\nu_1, \dots, \nu_n}(t-s+t_0)[iq_1(s)]^{\nu_1} \cdots [iq_n(s)]^{\nu_n} \, ds}}{[\mathbf{p}^f + i(\mathbf{g}-\mathbf{1})]^{\mathbf{1}}} \ . \tag{4.229}$$

This integral is a simple contour integral (the numerator is the exponential of a function analytic in $\mathbf{p}^f$), whose evaluation via Cauchy's integral formula corresponds to the desired result. □

4.7 Another view of the propagator

The mess of formalism aside, a coarse view of what we have been doing is that we have been calculating the propagator U, which we remind the reader is defined via

$$U(i\mathbf{p}^f, \mathbf{z}^0) := \langle -i\mathbf{p}^f | \hat{U}(t, t_0) | \mathbf{z}^0 \rangle \tag{4.230}$$

where $|\mathbf{z}^0\rangle$ and $\left| -i\mathbf{p}^f \right\rangle$ are coherent states. We computed U by evaluating many integrals, and then used the formula (c.f. Corollary 6)

$$|\psi(t)\rangle = \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \; |\mathbf{z}^f\rangle \, U(i\mathbf{p}^f, \mathbf{z}^0) \, \langle -i\mathbf{p}^0 | \psi(t_0) \rangle \, e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} \tag{4.231}$$

to recover the generating function $|\psi(t)\rangle$. This expression for $|\psi(t)\rangle$ is then suitably manipulated to directly recover other objects of interest, like moments or transition probabilities.

Given the relatively simple-looking results we have derived for U (c.f. Lemmas 2, 8, and 11), one may wonder whether there is another way to derive it—in particular, does U satisfy some PDE?

In the following, it will be more convenient to switch to a more standard notation for the probability generating function:

$$|\psi(t)\rangle = \sum_{\mathbf{x}} P(\mathbf{x}, t) \, |\mathbf{x}\rangle \rightarrow \psi(\mathbf{g}, t) := \sum_{\mathbf{x}} P(\mathbf{x}, t) \, \mathbf{g}^{\mathbf{x}} \tag{4.232}$$

which really just amounts to the replacement $|\mathbf{x}\rangle \rightarrow \mathbf{g}^{\mathbf{x}}$. This is related to our notation by taking the Grassberger-Scheunert product of the generating function with a coherent state $|\mathbf{g} - \mathbf{1}\rangle$ for some $\mathbf{g} \in \mathbb{R}^n$:

$$\langle \mathbf{g} - \mathbf{1} | \psi(t) \rangle = \sum_{\mathbf{x}} P(\mathbf{x}, t) \, \langle \mathbf{g} - \mathbf{1} | \mathbf{x} \rangle = \sum_{\mathbf{x}} P(\mathbf{x}, t)(\mathbf{1} + (\mathbf{g} - \mathbf{1}))^{\mathbf{x}} = \psi(\mathbf{g}, t) \; . \tag{4.233}$$

In this notation, the relationship between the generating function and the propagator reads

$$\psi(\mathbf{g}, t) = \int \frac{d\mathbf{z}^f d\mathbf{p}^f}{(2\pi)^n} \frac{d\mathbf{z}^0 d\mathbf{p}^0}{(2\pi)^n} \; e^{\mathbf{z}^f \cdot (\mathbf{g} - \mathbf{1})} U(i\mathbf{p}^f, \mathbf{z}^0) \, \langle -i\mathbf{p}^0 | \psi(t_0) \rangle \, e^{-i\mathbf{p}^0 \cdot \mathbf{z}^0 - i\mathbf{p}^f \cdot \mathbf{z}^f} \; . \tag{4.234}$$

Recall that the generating function $\psi(\mathbf{g}, t)$ satisfies a partial differential equation. For simplicity, suppose we are dealing with the chemical birth-death process, for which the relevant PDE reads

$$\frac{\partial \psi}{\partial t} = k(t)[g - 1]\psi - \gamma(t)[g - 1]\frac{\partial \psi}{\partial g} \; . \tag{4.235}$$

Substituting this into (the one-dimensional version of) Eq. 4.234, the right-hand side reads

$$\int \frac{dz_f dp_f}{2\pi} \frac{dz_0 dp_0}{2\pi} U \left\{ k[g - 1] - \gamma[g - 1]z_f \right\} \; e^{z_f(g-1)} e^{-i p_f z_f} \, \langle -i p_0 | \psi(t_0) \rangle \, e^{-i p_0 z_0} \; . \tag{4.236}$$

93

But note that

$$
(g-1)e^{z_f(g-1)} = \frac{\partial}{\partial z_f}e^{z_f(g-1)}
$$

$$
z_f e^{-ip_f z_f} = i\frac{\partial}{\partial p_f}e^{-ip_f z_f} \ .
$$

(4.237)

Using these identities, integrating by parts, and freely removing boundary terms, the right-hand side now reads

$$
\int \frac{dz_f dp_f}{2\pi}\frac{dz_0 dp_0}{2\pi}\left\{ip_f\left[kU + i\gamma\frac{\partial U}{\partial p_f}\right]\right\} \ e^{z_f(g-1)}e^{-ip_f z_f}\left\langle -ip_0|\psi(t_0)\right\rangle e^{-ip_0 z_0} \ .
$$

(4.238)

This suggests that the expression given by Eq. 4.234 will solve the equation of motion for $|\psi(t)\rangle$ (Eq. 4.58) if

$$
\frac{\partial U(ip_f, z_0)}{\partial t} = ip_f\left[kU(ip_f, z_0) + i\gamma\frac{\partial U(ip_f, z_0)}{\partial p_f}\right] \ .
$$

(4.239)

It is easy to verify that our expression for the propagator of the chemical birth-death process (c.f. Lemma 2) does solve this PDE.

We can generalize this enough for our purposes, although it should be clear that this correspondence holds for any CME (and not just ones involving only zero and first order reactions).

Proposition 11 (Propagator PDE)**.** *If the generating function $\psi(\mathbf{g}, t)$ satisfies the PDE given by Eq. 4.51, then the propagator $U(i\mathbf{p}^f, t; \mathbf{z}^0, t_0)$ satisfies a PDE*

$$
\begin{aligned}
\frac{\partial U}{\partial t} &= \sum_{\nu_1,\dots,\nu_n} \alpha_{\nu_1,\dots,\nu_n}(t)\left[ip_1^f\right]^{\nu_1}\cdots\left[ip_n^f\right]^{\nu_n} U \\
&\quad - i\sum_k \sum_{\nu_1,\dots,\nu_n} \beta_{\nu_1,\dots,\nu_n}^k(t)\left[ip_1^f\right]^{\nu_1}\cdots\left[ip_n^f\right]^{\nu_n}\frac{\partial U}{\partial p_k^f}
\end{aligned}
$$

(4.240)

with initial condition $U(i\mathbf{p}^f, t_0; \mathbf{z}^0, t_0) = \exp\!\left(i\mathbf{z}^0 \cdot \mathbf{p}^f\right)$ for arbitrary $\mathbf{p}^f, \mathbf{z}^0 \in \mathbb{R}^n$.

Proof. Integrate by parts as in the one-dimensional example. The initial condition comes from the definition of U:

$$
U(i\mathbf{p}^f, t_0; \mathbf{z}^0, t_0) = \left\langle -i\mathbf{p}^f\big|\hat{U}(t_0, t_0)\big|\mathbf{z}^0\right\rangle = \left\langle -i\mathbf{p}^f\big|\mathbf{z}^0\right\rangle = e^{i\mathbf{z}^0 \cdot \mathbf{p}^f} \ .
$$

(4.241)

□

At this point, we should note that the PDE satisfied by the propagator (Eq. 4.240) and the PDE satisfied by the generating function (Eq. 4.51) are equivalent up to a change of variables (i.e. $\mathbf{g} - \mathbf{1} \to i\mathbf{p}^f$). Does this mean that the propagator, along with the entire Doi-Peliti artifice we have constructed, is extraneous?

While this is a reasonable question to ask, the answer is probably no. It is easy to see that our expressions for the propagator and our expressions for the generating function have tended to look somewhat different, with the latter almost always being more complicated. The main reason for this difference seems to be that the propagator's initial condition is much simpler than the initial condition for the generating function PDE, which usually permits finding explicit solutions of the propagator PDE.

Now that we have this result, how can we connect it with the propagator solution we found in Lemma 11 (for arbitrary combinations of zero and first order reactions, which includes all other propagators considered in this paper as special cases)? It turns out that there is a straightforward way to do this using the method of characteristics, a standard approach for solving first order PDEs like the one above.

The method involves supposing that the relevant independent variables (in this case, $\mathbf{p}^f$ and t) lie along some parameterized curve. For a pedagogical example applying this method to solve a toy problem in chemical kinetics (the chemical birth-death process with additive noise), see [72].

Lemma 14 (Method of characteristics solution)**.** *The propagator for the system with arbitrary combinations of zero and first order reactions matches the one given by Lemma 11.*

Proof. Suppose (where we use slightly different notation here, because only the initial condition of the PDE depends on $\mathbf{z}^0$ and t_0) that $\mathbf{p}^f$ and t lie along curves parameterized by some parameter s, so that

$$
\begin{aligned}
\frac{\partial}{\partial s}\left[U(\mathbf{p}^f(s), t(s)\right] =& \frac{\partial U}{\partial t}\frac{\partial t}{\partial s} + \sum_{k=1}^{n}\frac{\partial U}{\partial p_k^f}\frac{\partial p_k^f}{\partial s} \\
=& -\frac{\partial U}{\partial t} - i\sum_k \sum_{\nu_1,\dots,\nu_n} \beta_{\nu_1,\dots,\nu_n}^k(t)\left[ip_1^f\right]^{\nu_1}\cdots\left[ip_n^f\right]^{\nu_n}\frac{\partial U}{\partial p_k^f} .
\end{aligned}
\tag{4.242}
$$

Choose the curve so that

$$
\begin{aligned}
\frac{\partial t}{\partial s} &= -1 \\
\frac{\partial p_k^f}{\partial s} &= -i\sum_{\nu_1,\dots,\nu_n} \beta_{\nu_1,\dots,\nu_n}^k(t(s))\left[ip_1^f(s)\right]^{\nu_1}\cdots\left[ip_n^f(s)\right]^{\nu_n} .
\end{aligned}
\tag{4.243}
$$

Suppose that we are interested in $U(i\mathbf{p}^f, t_f; \mathbf{z}^0, t_0)$ for some particular final time t_f. Solving the equation for $t(s)$, we have

$$
t(s) = t_f - s + t_0
\tag{4.244}
$$

where the arbitrary constant was chosen so that $s \in [t_0, t_f]$ with $t(t_0) = t_f$ and $t(t_f) = t_0$. Then the equation determining $\mathbf{p}^f(s)$ reads

$$
\frac{\partial p_k^f}{\partial s} = -i\sum_{\nu_1,\dots,\nu_n} \beta_{\nu_1,\dots,\nu_n}^k(t_f - s + t_0)\left[ip_1^f(s)\right]^{\nu_1}\cdots\left[ip_n^f(s)\right]^{\nu_n} .
\tag{4.245}
$$

Notice that this is exactly the same as the equation satisfied by $\mathbf{q}(s)$ (see Theorem 4). Moreover, our $\mathbf{p}^f(s)$ and $\mathbf{q}(s)$ satisfy the same initial condition: $\mathbf{p}^f(s = t_0) = \mathbf{p}^f(t(s) = t_f) = \mathbf{p}^f$, since the symbol $\mathbf{p}^f$ means the value corresponding to the evaluation of $U(i\mathbf{p}^f, t_f; \mathbf{z}^0, t_0)$ at the final time t_f. This point is somewhat subtle, so convince yourself of it before going forward.

Hence, we can make the identification $\mathbf{p}^f(s) \to \mathbf{q}(s)$. This means our PDE for U now reads

$$\frac{\partial U}{\partial s} = \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t) \, [iq_1]^{\nu_1} \cdots [iq_n]^{\nu_n} \, U \ . \tag{4.246}$$

Solving this as usual, we have the solution

$$U(s) = C \exp\left\{ \int_{t_0}^s \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t - s + t_0) \, [iq_1(s)]^{\nu_1} \cdots [iq_n(s)]^{\nu_n} \ ds \right\} \tag{4.247}$$

for some constant C that depends on the initial condition. Implement the initial condition for $s = t_0$ (i.e. $t(s) = t_f$), noting that $\mathbf{p}^f(s = t_0) = \mathbf{q}(s = t_f)$. Finally, evaluate $U(s)$ at $s = t_f$ to obtain

$$U(t_f) = \exp\left\{ i\mathbf{z}^0 \cdot \mathbf{q}(t_f) + \int_{t_0}^{t_f} \sum_{\nu_1,\ldots,\nu_n} \alpha_{\nu_1,\ldots,\nu_n}(t - s + t_0) \, [iq_1(s)]^{\nu_1} \cdots [iq_n(s)]^{\nu_n} \ ds \right\}$$
$$\tag{4.248}$$

which is the desired answer. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad\Box$

After all this, it is natural to ask whether the path integral calculations were necessary if the answer for the propagator can be determined by solving a relatively simple PDE. The author can only note that he was able to come up with this alternative approach only after carefully studying the path integral answer. It is likely that there are other cases where one can 'turn the crank' to determine the path integral answer, and then justify that answer using some more conventional method after one realizes why it takes its precise form.

4.8 Discussion

The strength of the Doi-Peliti approach—that calculations require nothing more clever than evaluating many integrals—is probably also its primary weakness. In Jahnke and Huisinga's original paper, they began with proofs of partial results that offered intuition for why their main result is true: in short, Poisson remains Poisson, and multinomial remains multinomial. In contrast, our calculation does not seem to offer such insight en route to the full solution. This may make it easier to generalize to other kinds of systems (as we did in Sec. 4.5 and Sec. 4.6), but it is a little unsatisfying.

Still, the Doi-Peliti approach *was* able to generate a solution in a nontrivial case where Jahnke and Huisinga's approach broke down, and we showed that it can offer solutions in far more general and nontrivial cases in Sec. 4.6. While the calculation is likely to be tedious, it seems possible that the Doi-Peliti approach could also be used to find explicit generating functions and transition probabilities (i.e. involving the explicit solution for $\mathbf{q}(t)$) for suitable generalizations of (for example) the birth-death-autocatalysis system, like one that involves many birth reactions, death reactions, and reactions of the form $S_j \to S_k + S_\ell$. It is not clear what new insights are necessary to solve explicitly for $\mathbf{q}(t)$ in cases like this.

Another obvious objection to the Doi-Peliti approach is that it is not entirely mathematically rigorous: in rederiving Jahnke and Huisinga's result, we freely swapped many improper integrals, frequently utilized the integral representation of the Dirac delta function, and so on. But we did get answers, and the method is likely to yield answers for problems that other methods cannot currently solve. If nothing else, the Doi-Peliti approach can be used as a tool to generate answers, which can be justified as rigorously correct using some other method (e.g. by showing that they solve the CME directly).

While we did not resort to approximations in this paper, it is worth noting that utilizing Doi-Peliti path integrals enables the use of powerful perturbative and asymptotic expansions. For most systems of interest in mathematical biology (e.g. gene networks with many species and interactions), this is the way in which the Doi-Peliti approach can be practically applied. See Weber and Frey [76], and Assaf and Meerson [20], for recent reviews discussing approximation techniques related to path integral descriptions of the CME.

The Doi-Peliti path integral is just one example of a stochastic path integral [76, 79]. The Onsager-Machlup [80, 81, 82, 83] and Martin-Siggia-Rose-Janssen-De Dominicis [84, 85, 86, 87, 83] path integrals are two other examples, which offer an alternative to the Fokker-Planck equation in the same way the Doi-Peliti path integral is an alternative to the CME. While exact computations of these path integrals are also tedious, they are just as mechanical— one can 'turn the crank' and generate answer, without relying on (for example) *a priori* knowledge of special functions to solve differential equations [72, 77].

4.9   Conclusion

We rederived Jahnke and Huisinga's classic result on monomolecular reaction systems using the Doi-Peliti coherent state path integral approach, which reduces solving the CME to the computation of many integrals. In addition, we also derived an explicit exact time-dependent solution to a problem involving an autocatalytic reaction that was beyond the scope of Jahnke and Huisinga's method, and a formal exact solution for systems involving arbitrary combinations of zero and first order reactions. We hope that our calculations, as well as our detailed description of the Doi-Peliti formalism, help make the Doi-Peliti method more accessible to mathematical biologists studying the CME.

4.A   Quantum vs standard notation

In this paper, we make abundant use of Dirac's bra-ket notation for vectors and inner products. While this notation is standard in quantum mechanics, it is less often used in areas with a more strictly mathematical bent. At the beginning of Sec. 4.3, we listed a few important reasons for this choice; to reiterate, it eases notation, makes it easy to repeatedly apply the identity operator (c.f. the derivation of the path integral expression for the propagator U), and is suggestive for the inner products we are using.

In this appendix, we will briefly review it and compare it with notation more commonly used in linear algebra and stochastic processes, so that this paper can be easily read by mathematicians unfamiliar with quantum mechanical notation.

For now, we will work in one dimension for simplicity. Consider a complex vector space V with a countable basis e_0, e_1, e_2, ..., so that an arbitrary state in this space reads

$$\phi = \sum_{x=0}^{\infty} c(x) e_x \tag{4.249}$$

for some complex coefficients $c(x)$. In terms of bra-ket notation, we would denote the basis vectors (also called 'kets' or 'states') by $|0\rangle, |1\rangle, |2\rangle$, ... and an arbitrary state by

$$|\phi\rangle = \sum_{x=0}^{\infty} c(x) |x\rangle \ , \tag{4.250}$$

which essentially amounts to the identifications $e_x \to |x\rangle$ and $\phi \to |\phi\rangle$.

Define the inner product $\langle e_x, e_y \rangle := \delta(x - y)$ for all $x, y \in \mathbb{N}$, and extend it to arbitrary states by linearity. Using bra-ket notation, we would write

$$\langle x|y\rangle = \delta(x - y) \ . \tag{4.251}$$

At this point, there are not yet any significant differences between the two choices of notation. The significant differences begin when we consider linear functionals like the functional $L_y :$ $V \to \mathbb{C}$ defined by its action on a basis vector e_x:

$$L_y(e_x) := \langle e_y, e_x \rangle \ . \tag{4.252}$$

Using bra-ket notation, we would denote L_y by $\langle y|$ (this is called a 'bra'), and L_y acting on e_x by $\langle y|x\rangle$ (the inner product is sometimes called a 'bra-ket'). This allows us to represent Fourier-like identities like

$$\phi = \sum_{y=0}^{\infty} \langle e_y, \phi \rangle e_y \tag{4.253}$$

via

$$|\phi\rangle = \sum_{y=0}^{\infty} \langle y|\phi\rangle |y\rangle \ , \tag{4.254}$$

or more succinctly by defining the operator

$$1 = \sum_{y=0}^{\infty} |y\rangle \langle y| \tag{4.255}$$

which by definition is equal to the identity operator. Equations like these are often called 'resolutions of the identity', because they recast the identity operator in some convenient form. The notation above is meant to be highly suggestive; one can imagine it 'bumping into' a vector/state $|\phi\rangle$ from the left to recover Eq. 4.254.

This notation also makes it easy to repeatedly apply resolutions of the identity, and to see what the result will be. Compare

$$\phi = \sum_{y_1, y_2, y_3} \langle e_{y_3}, e_{y_2} \rangle \langle e_{y_2}, e_{y_1} \rangle \langle e_{y_1}, \phi \rangle e_{y_3} \tag{4.256}$$

to

$$|\phi\rangle = \sum_{y_1, y_2, y_3} |y_3\rangle \langle y_3|y_2\rangle \langle y_2|y_1\rangle \langle y_1|\phi\rangle \ . \tag{4.257}$$

The above can be obtained simply by inserting Eq. 4.255 next to $|\phi\rangle$ many times.

One helpful feature of bra-ket notation is that eigenvectors are traditionally labeled by their eigenvalues. For example, if $\hat{\mathcal{A}}\phi = \lambda\phi$, it is traditional to write ϕ as

$$\phi \rightarrow |\lambda\rangle \ , \tag{4.258}$$

so that $\hat{\mathcal{A}}|\lambda\rangle = \lambda|\lambda\rangle$. We used this throughout the paper to denote coherent states, which we defined to be eigenstates of the annihilation operators.

Matrix elements—expressions of the form $\langle \phi_2, \hat{\mathcal{A}}\phi_1 \rangle$ for two vectors ϕ_1 and ϕ_2 and some operator $\hat{\mathcal{A}}$—are denoted by

$$\langle \phi_2 | \hat{\mathcal{A}} | \phi_1 \rangle \ . \tag{4.259}$$

This notation is convenient when we are computing matrix elements involving operators and their eigenstates. For example, let $\hat{a}$ be an operator, $\hat{a}^{\dagger}$ be its Hermitian conjugate, and $\phi_1 \rightarrow |\lambda_1\rangle$ and $\phi_2 \rightarrow |\lambda_2\rangle$ be eigenstates with eigenvalues λ_1 and λ_2, respectively. Then on the one hand, we have

$$\langle \phi_2, \hat{a}^{\dagger}\hat{a} \ \phi_1 \rangle = \langle \hat{a}\phi_2, \hat{a}\phi_1 \rangle = \lambda_2^* \lambda_1 \langle \phi_2, \phi_1 \rangle \tag{4.260}$$

in standard notation. On the other hand, we have

$$\langle \lambda_2 | \hat{a}^{\dagger}\hat{a} | \lambda_1 \rangle = \lambda_2^* \lambda_1 \langle \lambda_2 | \lambda_1 \rangle \tag{4.261}$$

using bra-ket notation, where we imagine $\hat{a}^{\dagger}$ 'acting to the left' and $\hat{a}$ 'acting to the right'.

That is about all there is to say about the correspondence between bra-ket notation and typical vector space notation. One should keep in mind that the strength of bra-ket notation is in repeatedly applying the identity operator/resolutions of the identity, which is required

to construct the Doi-Peliti path integral. The correspondence is summarized (for arbitrary dimensions, using the notation introduced in Sec. 4.2) in Table 4.1.

A few words should also be said about the relationship between our generating function and its usual analytic function form. We remind the reader that both are defined (in one dimension again, for simplicity) via

$$\psi(g,t) = \sum_{x=0}^{\infty} P(x,t)\ g^x \qquad |\psi\rangle = \sum_{x=0}^{\infty} P(x,t)\ |x\rangle \tag{4.262}$$

where $P(x,t)$ is a solution to the CME. These expressions are completely equivalent, up to the identification $g^x \to |x\rangle$. In fact, the equations of motion they satisfy exactly correspond. For example, in the case of the chemical birth-death process, we remind the reader that $\psi(g,t)$ satisfies the PDE

$$\frac{\partial \psi(g,t)}{\partial t} = k(t)[g-1]\psi(g,t) - \gamma(t)[g-1]\frac{\partial \psi(g,t)}{\partial g} \tag{4.263}$$

whereas $|\psi\rangle$ satisfies the equation

$$\frac{\partial |\psi\rangle}{\partial t} = \hat{H} |\psi\rangle \tag{4.264}$$

where in this case the Hamiltonian operator $\hat{H}$ is given (in terms of our original creation and annihilation operators) by

$$\hat{H} = k(\hat{\pi} - 1) - \gamma(\hat{\pi} - 1)\hat{a}\ . \tag{4.265}$$

This is the same as the above PDE, provided one makes the identifications

$$\hat{\pi} \to g$$
$$\hat{a} \to \frac{\partial}{\partial g}\ . \tag{4.266}$$

It turns out that these identifications work more generally (for arbitrary numbers of dimensions, and an arbitrary list of reactions). Although they are equivalent, one form of the generating function is often more convenient to use than the other. In our case, we use the Hilbert space form for almost the entirety of this paper, because it allows us to exploit bra-ket notation to denote applying many resolutions of the identity (c.f. Sec. 4.3.5), and to work straightforwardly in terms of matrix elements of the Hamiltonian.

Finally, we should say that the coherent state resolution of the identity we used many times to construct the Doi-Peliti path integral (c.f. Sec. 4.3.5) can be written in terms of ordinary functions as

$$g^x = \int_0^{\infty} dz \int_{-\infty}^{\infty} \frac{dp}{2\pi} e^{z(g-1)}(1+ip)^x e^{-izp} \tag{4.267}$$

in one dimension, and

$$\mathbf{g^x} = \int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \frac{d\mathbf{p}}{(2\pi)^n}\ e^{\mathbf{z}\cdot(\mathbf{g}-\mathbf{1})}(1+i\mathbf{p})^{\mathbf{x}} e^{-i\mathbf{z}\cdot\mathbf{p}} \tag{4.268}$$

in arbitrarily many dimensions. However, attempting to construct the path integral using this notation instead of bra-ket notation is significantly messier, so we have avoided it.

Object	Bra-ket notation	Standard notation
basis vector/ket	$|\mathbf{x}\rangle$	$\mathbf{e_x}$
linear functional/bra	$\langle\mathbf{x}|$	$L_\mathbf{x} : \mathbf{e_y} \mapsto \langle \mathbf{e_x}, \mathbf{e_y}\rangle$
zero vector	0	0
arbitrary state	$|\phi\rangle = \sum_\mathbf{x} c(\mathbf{x})\,|\mathbf{x}\rangle$	$\phi = \sum_\mathbf{x} c(\mathbf{x})\,\mathbf{e_x}$
inner product	$\langle\mathbf{x}|\mathbf{y}\rangle$	$\langle \mathbf{e_x}, \mathbf{e_y}\rangle$
operator matrix element	$\langle\mathbf{x}|\mathcal{A}|\mathbf{y}\rangle$	$\langle \mathbf{e_x}, \mathcal{A}\,\mathbf{e_y}\rangle = \langle \mathcal{A}^\dagger\,\mathbf{e_x}, \mathbf{e_y}\rangle$
generating function	$|\psi(t)\rangle = \sum_\mathbf{x} P(\mathbf{x},t)\,|\mathbf{x}\rangle$	$\psi(t) = \sum_\mathbf{x} P(\mathbf{x},t)\,\mathbf{e_x}$
coherent state (c.s.)	$|\mathbf{z}\rangle = \sum_\mathbf{y} \dfrac{\mathbf{z^y}}{\mathbf{y}!}e^{-\mathbf{z}\cdot\mathbf{1}}\,|\mathbf{y}\rangle$	$\mathrm{cs}(\mathbf{z}) = \sum_\mathbf{y} \dfrac{\mathbf{z^y}}{\mathbf{y}!}e^{-\mathbf{z}\cdot\mathbf{1}}\,\mathbf{e_y}$
c.s. identity operator	$|\mathbf{x}\rangle = \int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \dfrac{d\mathbf{p}}{(2\pi)^n}\,|\mathbf{z}\rangle\,\langle -i\mathbf{p}|\mathbf{x}\rangle\,e^{-i\mathbf{z}\cdot\mathbf{p}}$	$\mathbf{e_x} = \int_{[0,\infty)^n} d\mathbf{z} \int_{\mathbb{R}^n} \dfrac{d\mathbf{p}}{(2\pi)^n}\,\mathrm{cs}(\mathbf{z})\langle \mathrm{cs}(-i\mathbf{p}), \mathbf{e_x}\rangle e^{-i\mathbf{z}\cdot\mathbf{p}}$

Table 4.1: Let $\mathbf{x} \in \mathbb{R}^n$, and let the notation be as in Sec. 4.2.2 (e.g. $\mathbf{x}! := x_1! \cdots x_n!$). This table summarizes the correspondence between quantum and standard notation for several objects discussed in this appendix, as well as objects discussed elsewhere in this paper (e.g. coherent states).

Bibliography

[1] D.A. McQuarrie, Journal of Applied Probability **4**(3), 413–478 (1967). DOI 10.2307/3212214

[2] D.T. Gillespie, Physica A: Statistical Mechanics and its Applications **188**(1), 404 (1992). DOI https://doi.org/10.1016/0378-4371(92)90283-V. URL http://www.sciencedirect.com/science/article/pii/037843719290283V

[3] D.T. Gillespie, The Journal of Chemical Physics **113**(1), 297 (2000)

[4] D.T. Gillespie, Annual Review of Physical Chemistry **58**(1), 35 (2007)

[5] D.T. Gillespie, A. Hellander, L.R. Petzold, The Journal of Chemical Physics **138**(17), 170901 (2013)

[6] Z. Fox, B. Munsky, arXiv e-prints arXiv:1708.09264 (2017)

[7] B. Munsky, W.S. Hlavacek, L.S. Tsimring (eds.), Quantitative Biology: Theory, Computational Methods, and Models (The MIT Press, 2018)

[8] G. Neuert, B. Munsky, R.Z. Tan, L. Teytelman, M. Khammash, A. van Oudenaarden, Science **339**(6119), 584 (2013). DOI 10.1126/science.1231456

[9] B. Munsky, Z. Fox, G. Neuert, Methods **85**, 12 (2015). Inferring Gene Regulatory Interactions from Quantitative High-Throughput Measurements

[10] Z. Fox, G. Neuert, B. Munsky, The Journal of Chemical Physics **145**(7), 074101 (2016)

[11] B. Munsky, G. Li, Z.R. Fox, D.P. Shepherd, G. Neuert, Proceedings of the National Academy of Sciences **115**(29), 7533 (2018). DOI 10.1073/pnas.1804060115

[12] L. Weber, W. Raymond, B. Munsky, Physical Biology **15**(5), 055001 (2018)

[13] Z.R. Fox, B. Munsky, PLOS Computational Biology **15**(1), 1 (2019)

[14] Z.R. Fox, G. Neuert, B. Munsky, bioRxiv (2019). DOI 10.1101/812479. URL `https://www.biorxiv.org/content/early/2019/10/21/812479`

[15] A. Raj, P. van den Bogaard, S.A. Rifkin, A. van Oudenaarden, S. Tyagi, Nature Methods **5**(10), 877 (2008)

[16] A.M. Femino, F.S. Fay, K. Fogarty, R.H. Singer, Science **280**(5363), 585 (1998)

[17] S. Rahman, D. Zenklusen, in Imaging Gene Expression: Methods and Protocols, ed. by Y. Shav-Tal (Humana Press, Totowa, NJ, 2013), pp. 33–46

[18] O. Ovaskainen, B. Meerson, Trends in Ecology & Evolution **25**(11), 643 (2010)

[19] A. Melbinger, J. Cremer, E. Frey, Phys. Rev. Lett. **105**, 178101 (2010)

[20] M. Assaf, B. Meerson, Journal of Physics A: Mathematical and Theoretical **50**(26), 263001 (2017)

[21] Kai Nagel, Michael Schreckenberg, J. Phys. I France **2**(12), 2221 (1992)

[22] R. Mahnke, N. Pieret, Phys. Rev. E **56**, 2666 (1997)

[23] R. Mahnke, J. Kaupužs, I. Lubashevsky, Physics Reports **408**(1), 1 (2005)

[24] J.A. Miller, S.J. Klippenstein, The Journal of Physical Chemistry A **110**(36), 10528 (2006)

[25] D.R. Glowacki, C.H. Liang, C. Morley, M.J. Pilling, S.H. Robertson, The Journal of Physical Chemistry A **116**(38), 9545 (2012)

[26] A.W. Jasper, K.M. Pelzer, J.A. Miller, E. Kamarchik, L.B. Harding, S.J. Klippenstein, Science **346**(6214), 1212 (2014)

[27] D.T. Gillespie, Journal of Computational Physics **22**(4), 403 (1976). DOI https://doi.org/10.1016/0021-9991(76)90041-3. URL `http://www.sciencedirect.com/science/article/pii/0021999176900413`

[28] D.T. Gillespie, The Journal of Physical Chemistry **81**(25), 2340 (1977)

[29] B. Munsky, M. Khammash, The Journal of Chemical Physics **124**(4), 044104 (2006)

[30] S. Peleš, B. Munsky, M. Khammash, The Journal of Chemical Physics **125**(20), 204104 (2006)

[31] L.A. Harris, P. Clancy, The Journal of Chemical Physics **125**(14), 144107 (2006)

[32] L.A. Harris, A.M. Piccirilli, E.R. Majusiak, P. Clancy, Phys. Rev. E **79**, 051906 (2009). DOI 10.1103/PhysRevE.79.051906. URL https://link.aps.org/doi/10.1103/PhysRevE.79.051906

[33] K.A. Iyengar, L.A. Harris, P. Clancy, The Journal of Chemical Physics **132**(9), 094101 (2010)

[34] P. Bokes, J.R. King, A.T.A. Wood, M. Loose, Journal of Mathematical Biology **65**(3), 493 (2012)

[35] J. Hasenauer, V. Wolf, A. Kazeroonian, F.J. Theis, Journal of Mathematical Biology **69**(3), 687 (2014)

[36] X. Kan, C.H. Lee, H.G. Othmer, Journal of Mathematical Biology **73**(5), 1081 (2016)

[37] D.T. Gillespie, The Journal of Physical Chemistry A **106**(20), 5063 (2002)

[38] R. Grima, P. Thomas, A.V. Straube, The Journal of Chemical Physics **135**(8), 084103 (2011)

[39] J.J. Vastola, W.R. Holmes, Phys. Rev. E **101**, 032417 (2020)

[40] M. Delbrück, The Journal of Chemical Physics **8**(1), 120 (1940)

[41] A. Rényi, Magyar Tud. Akad. Alkalm. Mat. Int. Közl **2**, 93 (1954)

[42] K. Ishida, Bulletin of the Chemical Society of Japan **33**(8), 1030 (1960)

[43] D.A. McQuarrie, The Journal of Chemical Physics **38**(2), 433 (1963)

[44] D.A. McQuarrie, C.J. Jachimowski, M.E. Russell, The Journal of Chemical Physics **40**(10), 2914 (1964)

[45] T. Jahnke, W. Huisinga, Journal of Mathematical Biology **54**(1), 1 (2007)

[46] M. Reis, J.A. Kromer, E. Klipp, Journal of Mathematical Biology **77**(2), 377 (2018)

[47] I.J. Laurenzi, The Journal of Chemical Physics **113**(8), 3315 (2000)

[48] E. Arslan, I.J. Laurenzi, The Journal of Chemical Physics **128**(1), 015101 (2008)

[49] M. Doi, J. Phys. A **9**(9), 1479 (1976). DOI 10.1088/0305-4470/9/9/009. URL https://doi.org/10.1088%2F0305-4470%2F9%2F9%2F009

[50] M. Doi, J. Phys. A **9**, 1465 (1976). DOI 10.1088/0305-4470/9/9/008

[51] L. Peliti, Journal de Physique **46**(9), 1469 (1985). DOI 10.1051/jphys: 019850046090146900. URL `http://www.edpsciences.org/10.1051/jphys: 019850046090146900`

[52] L. Peliti, Y.C. Zhang, Journal de Physique Lettres **46**(24), 1151 (1985)

[53] L. Peliti, Journal of Physics A: Mathematical and General **19**(6), L365 (1986)

[54] P. Grassberger, M. Scheunert, Fortschritte der Physik **28**(10), 547 (1980). DOI 10.1002/ prop.19800281004

[55] P. Grassberger, Zeitschrift für Physik B Condensed Matter **47**(4), 365 (1982)

[56] J.L. Cardy, P. Grassberger, Journal of Physics A: Mathematical and General **18**(6), L267 (1985)

[57] P. Grassberger, Journal of Physics A: Mathematical and General **22**(23), L1103 (1989)

[58] D.C. Mattis, M.L. Glasser, Rev. Mod. Phys. **70**, 979 (1998). DOI 10.1103/RevModPhys. 70.979. URL `https://link.aps.org/doi/10.1103/RevModPhys.70.979`

[59] B.P. Lee, Journal of Physics A: Mathematical and General **27**(8), 2633 (1994)

[60] B.P. Lee, J. Cardy, Phys. Rev. E **50**, R3287 (1994)

[61] B.P. Lee, J. Cardy, Journal of Statistical Physics **80**(5), 971 (1995)

[62] F. van Wijland, K. Oerding, H. Hilhorst, Physica A: Statistical Mechanics and its Applications **251**(1), 179 (1998)

[63] L. Canet, B. Delamotte, O. Deloubrière, N. Wschebor, Phys. Rev. Lett. **92**, 195703 (2004)

[64] L. Canet, Journal of Physics A: Mathematical and General **39**(25), 7901 (2006)

[65] U.C. Täuber, M. Howard, B.P. Vollmayr-Lee, Journal of Physics A: Mathematical and General **38**(17), R79 (2005)

[66] T. Fung, J.P. O'Dwyer, R.A. Chisholm, Journal of Mathematical Biology **74**(1), 289 (2017)

[67] C.D. Greenman, T. Chou, Phys. Rev. E **93**, 012112 (2016)

[68] C.D. Greenman, Journal of Statistical Mechanics: Theory and Experiment **2017**(3), 033101 (2017)

[69] P.C. Bressloff, <u>Stochastic Processes in Cell Biology</u> (Springer, 2014). DOI 10.1007/978-3-319-08488-6

[70] J. Albert, Journal of Mathematical Biology (2019)

[71] P.C. Bressloff, Journal of Physics A: Mathematical and Theoretical **50**(13), 133001 (2017)

[72] J.J. Vastola, arXiv e-prints arXiv:1910.09117 (2019)

[73] D. Griffiths, D. Schroeter, <u>Introduction to Quantum Mechanics</u> (Cambridge University Press, 2018)

[74] M.D. Schwartz, <u>Quantum Field Theory and the Standard Model</u> (Cambridge University Press, 2014). URL `http://www.cambridge.org/us/academic/subjects/physics/theoretical-physics-and-mathematical-physics/quantum-field-theory-and-standard-model`

[75] J. Cardy, G. Falkovich, K. Gawedzki, <u>Reaction-diffusion processes</u> (Cambridge University Press, 2008), p. 108–161. London Mathematical Society Lecture Note Series. DOI 10.1017/CBO9780511812149.004

[76] M.F. Weber, E. Frey, Reports on Progress in Physics **80**(4), 046601 (2017). DOI 10.1088/1361-6633/aa5ae2. URL `http://stacks.iop.org/0034-4885/80/i=4/a=046601?key=crossref.02abaf744081951aaaca3bafec0e1284`

[77] J.J. Vastola, arXiv e-prints arXiv:1910.10807 (2019)

[78] I.S. Gradshteyn, I.M. Ryzhik, <u>Table of integrals, series, and products</u> (Academic press, 2014)

[79] J.J. Vastola, W.R. Holmes, arXiv e-prints arXiv:1909.12990 (2019)

[80] L. Onsager, S. Machlup, Phys. Rev. **91**, 1505 (1953). DOI 10.1103/PhysRev.91.1505. URL `https://link.aps.org/doi/10.1103/PhysRev.91.1505`

[81] S. Machlup, L. Onsager, Phys. Rev. **91**, 1512 (1953). DOI 10.1103/PhysRev.91.1512. URL `https://link.aps.org/doi/10.1103/PhysRev.91.1512`

[82] R. Graham, Zeitschrift für Physik B Condensed Matter **26**(3), 281 (1977). DOI 10.1007/BF01312935. URL `https://doi.org/10.1007/BF01312935`

[83] J.A. Hertz, Y. Roudi, P. Sollich, Journal of Physics A: Mathematical and Theoretical **50**(3), 033001 (2016). DOI 10.1088/1751-8121/50/3/033001. URL `https://doi.org/10.1088%2F1751-8121%2F50%2F3%2F033001`

[84] P.C. Martin, E.D. Siggia, H.A. Rose, Phys. Rev. A **8**, 423 (1973). DOI 10.1103/PhysRevA.8.423. URL `https://link.aps.org/doi/10.1103/PhysRevA.8.423`

[85] H.K. Janssen, Zeitschrift für Physik B Condensed Matter **23**(4), 377 (1976). DOI 10.1007/BF01316547. URL https://doi.org/10.1007/BF01316547

[86] DE DOMINICIS, C., J. Phys. Colloques **37**(C1), C1 (1976). DOI 10.1051/jphyscol: 1976138. URL https://doi.org/10.1051/jphyscol:1976138

[87] C. De Dominicis, L. Peliti, Phys. Rev. B **18**, 353 (1978). DOI 10.1103/PhysRevB.18.353. URL https://link.aps.org/doi/10.1103/PhysRevB.18.353

Chapter 5

Path integral representations of the chemical master equation.
II: A new state space path integral and application to coarse-graining

How can we *solve* the CME, and how can we *simplify* it? This is the second of two chapters that introduce path integral methods; this method is new, and is directly analogous to state space path integrals (i.e. the usual ones) from quantum mechanics. This path integral is both useful for solving the CME (see Appendix A of this chapter) and for simplifying it. The main thrust here is to justify the most common approximation in chemical kinetics/gene regulation: approximating continuous dynamics as discrete. As noted at the beginning of the previous chapter, this path integral is just as 'powerful' as the Doi-Peliti one, although it has the advantage of requiring less formalism.

Abstract: In 2000, Gillespie rehabilitated the chemical Langevin equation (CLE) by describing two conditions that must be satisfied for it yield a valid approximation of the chemical master equation (CME). In this work, we construct an original path integral description of the CME, and show how applying Gillespie's two conditions to it directly leads to a path integral equivalent to the CLE. We compare this approach to the path integral equivalent of a large system size derivation, and show that they are qualitatively different. In particular, both approaches involve converting many sums into many integrals, and the difference between the two methods is essentially the difference between using the Euler-Maclaurin formula and using Riemann sums. Our results shed light on how path integrals can be used to conceptualize coarse-graining biochemical systems, and are readily generalizable.

5.1 Introduction

Gillespie's classic paper [1] on how to derive the chemical Langevin equation (CLE) from the chemical master equation (CME) proceeds differently than by naively truncating the Kramers-Moyal expansion of the CME [2, 3, 4] or by invoking the largeness of the system volume Ω *a la* van Kampen [5, 6]; instead, he argues based on the existence of a *time scale* with certain properties. In particular, his derivation completes avoids rewriting discrete number variables n as concentration variables $x := n/\Omega$.

By writing down two precise conditions that control the validity of the CLE (to be reviewed in Sec. 5.2), he rehabilitated it as a well-founded approach to approximating stochastic dynamics described by the CME (in the face of ostensible no-go results like the Pawula theorem [7]), and directly inspired the tau-leaping algorithm [8] and its many modifications [9, 10, 11, 12, 13, 14] for speeding up numerical simulations of biochemical reactions.

Path integrals offer a way to think about stochastic processes that is somewhat independent from the usual differential equations perspective [15]. This means that—at least in principle—there should be a way to translate Gillespie's derivation into path integral language. Because path integrals (along with associated technology like the renormalization group [16, 17, 18, 19]) are known to be useful for understanding coarse-grained descriptions of systems (e.g. effective field theories [20, 19]), such a translation should contribute meaningfully to our understanding of how to intelligently coarse-grain biochemical systems.

In this paper, we show how Gillespie's two conditions translate to a path integral-based derivation of the chemical Langevin equation. Our approach here builds upon the path integral descriptions of Langevin/Fokker-Planck equations described in [15]. We will proceed with little mathematical rigor (as is typical in physics), but with enough clarity that our arguments could in principle be made mathematically precise.

The paper is organized as follows. In Sec. 5.2, we review Gillespie's derivation of the CLE. In Sec. 7.3, we construct a path integral description of CME dynamics. In Sec. 5.4, we apply Gillespie's conditions to our path integral formulation to obtain the CLE, and also discuss an alternative method based on a large system volume argument. Finally, we discuss consequences of our work for understanding coarse-grained biochemical systems in Sec. 7.6.

5.2 Review of Gillespie's chemical Langevin equation derivation

In this section, we review Gillespie's derivation [1] of the chemical Langevin equation from the chemical master equation. We use the same notation Gillespie used in his paper, although we will not require that the same physical assumptions (i.e. well-stirred, dilute chemicals in a fixed volume and at constant temperature) hold, because the derivation does not depend on them.

Consider a system with N species and M reactions. Denote the propensity function of

the jth reaction by a_j, and the corresponding stoichiometry vector by $\boldsymbol{\nu}_j$. The chemical master equation reads

$$\frac{\partial P(\mathbf{n}, t)}{\partial t} = \sum_{j=1}^{M} a_j(\mathbf{n} - \boldsymbol{\nu}_j)P(\mathbf{n} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{n})P(\mathbf{n}, t) \tag{5.1}$$

where $P(\mathbf{n}, t)$ is the probability that the state of the system is $\mathbf{n} = (n_1, ..., n_N) \in \mathbb{N}^N$ at time t.

Gillespie's derivation requires the existence of a time scale τ for which the following two conditions hold:

 i The propensity functions do not change their values appreciably,
   i.e. $a_j(\mathbf{n}(t)) \approx a_j(\mathbf{n}(t'))$ for all j and all $t' \in [t, t + \tau]$.

 ii The average number of firings of each reaction over a time τ is much larger than 1.

Due to their connection with the tau-leaping algorithm [8, 9, 10, 11, 12, 13, 14] for approximately simulating CME dynamics, Gillespie later called these the first leap condition and the second leap condition [21]. They are in practice easily satisfied in the case of large molecule numbers, and they are exactly satisfied in the thermodynamic limit [22], where the system volume Ω is taken to infinity while keeping all concentrations fixed.

Consider $n_i(t)$, the number of molecules corresponding to species i at time t. It changes in a small time Δt according to

$$n_i(t + \Delta t) = n_i(t) + \sum_{j=1}^{M} \nu_{ji} K_j(a_j, \Delta t) \ , \tag{5.2}$$

where ν_{ji} is the ith component of the stoichiometry vector $\boldsymbol{\nu}_j$ (i.e. the change in number of species i due to reaction j firing once), and $K_j(a_j, \Delta t)$ is a random variable that describes the number of times reaction j fires in Δt.

For an arbitrary CME and arbitrary length of time Δt, K_j might be taken from a complicated distribution. But if condition (i) holds in a length of time τ, each reaction fires independently of each other reaction, because no reactions significantly change any propensity functions. Because (by definition) the probability of reaction j firing in an infinitesimal time dt is $a_j(\mathbf{n}(t))dt$, and because that probability will not significantly change during the time length τ, the number of times reaction j fires in τ is well-approximated as a Poisson random variable with mean $a_j(\mathbf{n}(t))\tau$, which we will denote by $\mathcal{P}_j(a_j(\mathbf{n}(t))\tau)$.

This means that when condition (i) holds we can write the time evolution of $n_i(t)$ over a length of time τ as

$$n_i(t + \tau) = n_i(t) + \sum_{j=1}^{M} \nu_{ji} \mathcal{P}_j(a_j(\mathbf{n}(t))\tau) \ . \tag{5.3}$$

This equation is the basis for the tau-leaping approach first described by Gillespie in 2001 [8], and later modified and extended by himself and others [9, 10, 11, 12, 13, 14].

If condition (ii) holds, the average number of times reaction j fires in τ (i.e. $a_j(\mathbf{n}(t))\tau$) is much larger than 1, so the Poisson random variables are well-approximated by normal random variables:

$$\mathcal{P}_j(a_j(\mathbf{n}(t))\tau) \approx \mathcal{N}_j(a_j(\mathbf{n}(t))\tau, a_j(\mathbf{n}(t))\tau) , \qquad (5.4)$$

where $\mathcal{N}_j(a_j(\mathbf{n}(t))\tau, a_j(\mathbf{n}(t))\tau)$ is a normal random variable with mean and variance both equal to $a_j(\mathbf{n}(t))\tau$. If we also note that each normal random variable can be decomposed as

$$\mathcal{N}_j(a_j(\mathbf{n}(t))\tau, a_j(\mathbf{n}(t))\tau) = a_j(\mathbf{n}(t))\tau + \sqrt{a_j(\mathbf{n}(t))\tau}\, \mathcal{N}_j(0,1) , \qquad (5.5)$$

we can write the time evolution of $n_i(t)$ in a time τ as

$$n_i(t+\tau) = n_i(t) + \sum_{j=1}^{M} \nu_{ji} a_j(\mathbf{n}(t))\tau + \sum_{j=1}^{M} \nu_{ji} \sqrt{a_j(\mathbf{n}(t))\tau}\, \mathcal{N}_j(0,1) . \qquad (5.6)$$

Because this equation has the form of an Euler-Maruyama time step, we can identify the dynamics of the system on the time scale τ with the set of N stochastic differential equations (SDEs)

$$\dot{x}_i = \sum_{j=1}^{M} \nu_{ji} a_j(\mathbf{x}) + \sum_{j=1}^{M} \nu_{ji} \sqrt{a_j(\mathbf{x})}\, \Gamma_j , \qquad (5.7)$$

where the Γ_j are M independent Gaussian white noise terms, and where we have relabeled each n_i as x_i to emphasize that we are now working with continuous variables.

Our chemical Langevin equation corresponds to a chemical Fokker-Planck equation

$$\frac{\partial P(\mathbf{x},t)}{\partial t} = \sum_{i=1}^{N} -\frac{\partial}{\partial x_i}\left[\left(\sum_{j=1}^{M}\nu_{ji}a_j(\mathbf{x})\right)P(\mathbf{x},t)\right] + \frac{1}{2}\sum_{i=1}^{N}\sum_{i'=1}^{N}\frac{\partial^2}{\partial x_i \partial x_{i'}}\left[\left(\sum_{j=1}^{M}\nu_{ji}\nu_{ji'}a_j(\mathbf{x})\right)P(\mathbf{x},t)\right]$$
$$(5.8)$$

which serves an approximation to the CME (Eq. 5.1). As Gillespie notes, this is exactly what one would get from truncating the Kramers-Moyal expansion of the CME at second order, so his derivation in some sense justifies the naive one.

The CLE (Eq. 5.7), and the associated chemical Fokker-Planck equation (Eq. 5.8) describing how the system's probability density will evolve in time, are not without problems. They generically predict negative concentrations [23] (although the hope is that the system has a negligibly small probability of occupying these states, and this is often borne out in practice), can be inaccurate for systems far from equilibrium [24], may not always exhibit multistability when the CME is multistable [25], and can give rise to nonphysical probability currents at equilibrium [26].

Despite these shortcomings, utilizing the CLE can help speed up simulations of CME dynamics when some species have large molecule numbers [27, 28, 29, 30] or when there is a clear separation of time scales [31, 32, 33]. Moreover, alternative schemes like the deterministic reaction rate equations and the linear noise approximation [6] can profitably

be viewed as approximations to the CLE [34], and moment-closure approximations have comparable accuracy [35].

The CLE, and Langevin equations more generally, have become standard approaches to modeling noisy gene regulation [36, 37, 38, 39, 40]. They have also been used to analyze noise-driven oscillations [41], model intracellular calcium dynamics [42, 43, 44], study ion-channel gating [45], and understand spiking neurons [46]. While only approximate, the CLE is unquestionably useful.

5.3 Path integral formulation of CME dynamics

Although path integrals [47] are most well-known in the context of quantum mechanics and quantum field theory [48, 49, 50, 19, 51, 52], they have also proven useful for understanding classical stochastic phenomena like Brownian motion [53, 54, 55], conformational transitions [56, 57, 58], quantitative finance [59, 60, 50], population dynamics [61, 62, 63, 64, 65], neuron firing [66, 67, 68, 69, 70, 71, 72], gene regulation [36, 73, 74, 75, 76, 77], and chemical kinetics [78, 79, 80, 81, 82].

In this section, we will develop a straightforward path integral formulation of chemical master equation dynamics. Our path integral is constructed to closely resemble the formalism we used to describe SDE/Fokker-Planck dynamics in [15]. To our knowledge, it is original, although certain aspects also resemble the approach used by Lazarescu et al. [83]. The approach presented in this section is somewhat distinct from the often used Doi-Peliti approach [84, 85, 86, 87], which involves integrating over so-called coherent states and yields integrals instead of sums.

5.3.1 States and operators

Our main objective is to solve the CME, Eq. 5.1. Instead of solving it directly, we will solve a related problem phrased in terms of states and operators in a certain Hilbert space; this allows us to construct a path integral just as one does in quantum mechanics.

Consider an infinite-dimensional Hilbert space spanned by the $|\mathbf{n}\rangle$ vectors (where $\mathbf{n} = (n_1, ..., n_N) \in \mathbb{N}^N$), in which an arbitrary state $|\phi\rangle$ is written

$$|\phi\rangle = \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} c(\mathbf{n}) |\mathbf{n}\rangle \tag{5.9}$$

for some generally complex-valued coefficients $c(\mathbf{n})$. To ease notation, we will write

$$\sum_{\mathbf{n}} := \sum_{n_1=0}^{\infty} \cdots \sum_{n_N=0}^{\infty} \tag{5.10}$$

so that an arbitrary state reads

$$|\phi\rangle = \sum_{\mathbf{n}} c(\mathbf{n}) |\mathbf{n}\rangle \ . \tag{5.11}$$

Define an inner product in this space by

$$\langle \mathbf{m}|\mathbf{n}\rangle = \delta_{\mathbf{mn}} \tag{5.12}$$

for all basis vectors $|\mathbf{m}\rangle$ and $|\mathbf{n}\rangle$, so that the inner product of two arbitrary states reads

$$\langle \phi_2|\phi_1\rangle = \sum_{\mathbf{n}} c_2^*(\mathbf{n})c_1(\mathbf{n}) \ . \tag{5.13}$$

Using the inner product defined by Eq. 5.12 and Eq. 5.13, we can show that there is a resolution of the identity

$$1 = \sum_{\mathbf{n}} |\mathbf{n}\rangle \langle \mathbf{n}| \tag{5.14}$$

since $\langle \mathbf{n}|\phi\rangle = c(\mathbf{n})$. Define the state operators $\hat{n}_i$ by

$$\hat{n}_i |\mathbf{n}\rangle := n_i |\mathbf{n}\rangle \tag{5.15}$$

for all $i = 1, ..., N$. We will associate any function $f(\mathbf{n}) = f(n_1, ..., n_N)$ with the operator $f(\hat{\mathbf{n}})$, whose action on a basis vector $|\mathbf{n}\rangle$ is

$$f(\hat{\mathbf{n}}) |\mathbf{n}\rangle := f(\hat{n}_1, ..., \hat{n}_N) |\mathbf{n}\rangle \tag{5.16}$$

where there is no operator ordering ambiguity because the $\hat{n}_i$ all commute with one another. Also define the propensity function operators $\hat{a}_j$ via

$$\hat{a}_j |\mathbf{n}\rangle := a_j(\boldsymbol{n}) |\mathbf{n} + \boldsymbol{\nu}_j\rangle \tag{5.17}$$

for all $j = 1, ..., M$, where $\boldsymbol{\nu}_j$ denotes the stoichiometry vector of the jth reaction.

5.3.2 Generating function and equation of motion

In the spirit of Peliti[86], define the generating function

$$|\psi(t)\rangle := \sum_{\mathbf{n}} P(\mathbf{n}, t) |\mathbf{n}\rangle \tag{5.18}$$

where, as in the previous section, $P(\mathbf{n}, t)$ is the probability that the state of the system is $\mathbf{n} = (n_1, ..., n_N)$ at time t. Note that

$$\begin{aligned}
\frac{\partial |\psi\rangle}{\partial t} &= \sum_{\mathbf{n}} \frac{\partial P(\mathbf{n}, t)}{\partial t} |\mathbf{n}\rangle \\
&= \sum_{\mathbf{n}} \left[\sum_{j=1}^{M} a_j(\mathbf{n} - \boldsymbol{\nu}_j)P(\mathbf{n} - \boldsymbol{\nu}_j, t) - a_j(\mathbf{n})P(\mathbf{n}, t) \right] |\mathbf{n}\rangle \\
&= \sum_{\mathbf{n}} \sum_{j=1}^{M} a_j(\mathbf{n})P(\mathbf{n}, t) |\mathbf{n} + \boldsymbol{\nu}_j\rangle - \sum_{\mathbf{n}} \sum_{j=1}^{M} a_j(\mathbf{n})P(\mathbf{n}, t) |\mathbf{n}\rangle
\end{aligned} \tag{5.19}$$

112

where we have reindexed the left sum in the last step. Now we have

$$\frac{\partial |\psi\rangle}{\partial t} = \sum_{\mathbf{n}} \left[\sum_{j=1}^{M} a_j(\mathbf{n}) |\mathbf{n} + \boldsymbol{\nu}_j\rangle - a_j(\mathbf{n}) |\mathbf{n}\rangle \right] P(\mathbf{n}, t)$$

$$= \sum_{\mathbf{n}} \left[\sum_{j=1}^{M} \hat{a}_j |\mathbf{n}\rangle - a_j(\mathbf{n}) \right] P(\mathbf{n}, t) |\mathbf{n}\rangle \qquad (5.20)$$

$$= \left[\sum_{j=1}^{M} \hat{a}_j - a_j(\hat{\mathbf{n}}) \right] |\psi\rangle \ .$$

If we define the operator

$$\hat{H} := \sum_{j=1}^{M} \hat{a}_j - a_j(\hat{\mathbf{n}}) \ , \qquad (5.21)$$

which we will call (in analogy with quantum mechanics) the Hamiltonian, then we can write the equation describing the time evolution of the generating function as

$$\frac{\partial |\psi\rangle}{\partial t} = \hat{H} |\psi\rangle \ . \qquad (5.22)$$

It is this equation that we will solve instead of the CME; since $\langle \mathbf{n} | \psi(t) \rangle = P(\mathbf{n}, t)$, a solution to the CME can be extracted out of a solution to this equation.

5.3.3 Deriving the CME path integral

The formal solution to Eq. 5.22 is

$$|\psi(t_f)\rangle = e^{\hat{H}(t_f - t_0)} |\psi(t_0)\rangle \ . \qquad (5.23)$$

At this point (following the usual procedure for deriving path integrals [15]), we write the length of time $(t_f - t_0)$ as $T\Delta t$ for some large number of time steps T, split the propagator into many pieces, and insert many resolutions of the identity:

$$|\psi(t_f)\rangle = e^{\hat{H}\Delta t} \cdots e^{\hat{H}\Delta t} |\psi(t_0)\rangle$$

$$= \sum_{\mathbf{n}_0} \cdots \sum_{\mathbf{n}_T} |\mathbf{n}_T\rangle \langle \mathbf{n}_T | e^{\hat{H}\Delta t} |\mathbf{n}_{T-1}\rangle \cdots \langle \mathbf{n}_1 | e^{\hat{H}\Delta t} |\mathbf{n}_0\rangle \langle \mathbf{n}_0 | \psi(t_0)\rangle \ . \qquad (5.24)$$

We are specifically interested in the transition probability $P(\mathbf{n}_f, t_f; \mathbf{n}_0, t_0)$. To obtain an expression for it, note that if $|\psi(t_0)\rangle = |\mathbf{n}_0\rangle$, then $P(\mathbf{n}_f, t_f; \mathbf{n}_0, t_0) = \langle \mathbf{n}_f | \psi(t_f)\rangle$. Hence, we have

$$P(\mathbf{n}_f, t_f; \mathbf{n}_0, t_0) = \sum_{\mathbf{n}_1} \cdots \sum_{\mathbf{n}_{T-1}} \langle \mathbf{n}_T | e^{\hat{H}\Delta t} |\mathbf{n}_{T-1}\rangle \cdots \langle \mathbf{n}_1 | e^{\hat{H}\Delta t} |\mathbf{n}_0\rangle \ . \qquad (5.25)$$

where $\mathbf{n}_T = \mathbf{n}_f$. Now we just need to evaluate these matrix elements and put them together. Choose Δt sufficiently small so that

$$\langle \mathbf{n}_k | e^{\hat{H}\Delta t} | \mathbf{n}_{k-1} \rangle \approx \langle \mathbf{n}_k | 1 + \hat{H}\Delta t | \mathbf{n}_{k-1} \rangle$$
$$= \delta_{\mathbf{n}_k, \mathbf{n}_{k-1}} + \langle \mathbf{n}_k | \hat{H} | \mathbf{n}_{k-1} \rangle \, \Delta t \ . \tag{5.26}$$

We will take $\Delta t \to 0$ at the end of the calculation, so this equality will hold exactly. Using the specific form of $\hat{H}$ (Eq. 5.21), we have

$$\langle \mathbf{n}_k | \hat{H} | \mathbf{n}_{k-1} \rangle = \langle \mathbf{n}_k | \sum_{j=1}^{M} \hat{a}_j - a_j(\hat{\mathbf{n}}) | \mathbf{n}_{k-1} \rangle$$
$$= \sum_{j=1}^{M} \langle \mathbf{n}_k | \hat{a}_j - a_j(\hat{\mathbf{n}}) | \mathbf{n}_{k-1} \rangle$$
$$= \sum_{j=1}^{M} a_j(\mathbf{n}_{k-1}) \left[\langle \mathbf{n}_k | \mathbf{n}_{k-1} + \boldsymbol{\nu}_j \rangle - \langle \mathbf{n}_k | \mathbf{n}_{k-1} \rangle \right]$$
$$= \sum_{j=1}^{M} a_j(\mathbf{n}_{k-1}) \left[\delta_{\mathbf{n}_k, \mathbf{n}_{k-1}+\boldsymbol{\nu}_j} - \delta_{\mathbf{n}_k, \mathbf{n}_{k-1}} \right] \ . \tag{5.27}$$

Recall that the usual integral representation of the Dirac delta function reads

$$\delta_{\mathbf{m},\mathbf{n}} = \int \frac{d\mathbf{p}}{(2\pi)^N} \, e^{-i\mathbf{p}\cdot(\mathbf{m}-\mathbf{n})} \ , \tag{5.28}$$

where $d\mathbf{p} = dp_1 \cdots dp_N$ and each p_i is integrated over the whole real line. Using this representation, $\langle \mathbf{n}_k | \hat{H} | \mathbf{n}_{k-1} \rangle$ becomes

$$\int \frac{d\mathbf{p}_k}{(2\pi)^N} \, e^{-i\mathbf{p}_k\cdot(\mathbf{n}_k-\mathbf{n}_{k-1})} \left\{ \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k\cdot\boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \right\} \tag{5.29}$$

where we have labeled the integration variable $\mathbf{p}_k$ to anticipate there being one integral for each matrix element in the final answer. Using Eq. 5.26, $\langle \mathbf{n}_k | e^{\hat{H}\Delta t} | \mathbf{n}_{k-1} \rangle$ is approximately equal to

$$\int \frac{d\mathbf{p}_k}{(2\pi)^N} \, e^{-i\mathbf{p}_k\cdot(\mathbf{n}_k-\mathbf{n}_{k-1})} \left\{ 1 + \Delta t \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k\cdot\boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \right\} \ . \tag{5.30}$$

Noting that Δt is small enough for the bracketed expression to be approximately equal to the corresponding exponential, our final expression for $\langle \mathbf{n}_k | e^{\hat{H}\Delta t} | \mathbf{n}_{k-1} \rangle$ becomes

$$\int \frac{d\mathbf{p}_k}{(2\pi)^N} \, e^{-i\mathbf{p}_k\cdot(\mathbf{n}_k-\mathbf{n}_{k-1})+\Delta t \sum_{j=1}^{M}[\exp(i\mathbf{p}_k\cdot\boldsymbol{\nu}_j)-1]a_j(\mathbf{n}_{k-1})} \ . \tag{5.31}$$

Using Eq. 5.25 and Eq. 5.31, we find that $P(\mathbf{n}_f, t_f; \mathbf{n}_0, t_0)$ can be written as the path integral

$$
P = \lim_{T \to \infty} \sum_{\mathbf{n}_1} \cdots \sum_{\mathbf{n}_{T-1}} \int \frac{d\mathbf{p}_1}{(2\pi)^N} \cdots \int \frac{d\mathbf{p}_T}{(2\pi)^N} \exp\left\{ \sum_{k=1}^{T} -i\mathbf{p}_k \cdot (\mathbf{n}_k - \mathbf{n}_{k-1}) + \Delta t \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \right\}
$$

$$(5.32)$$

which resembles the MSRJD (Martin-Siggia-Rose-De Dominicis) path integral description [88, 89, 90, 91, 92] of the Fokker-Planck equation. Again, while the Doi-Peliti path integral involves integrating over coherent states, this path integral involves integrating over every possible discrete path through $\mathbb{N}^N$ that goes from $\mathbf{n}_0$ to $\mathbf{n}_f$.

Although our primary interest in this paper is to use Eq. 5.32 to derive the CLE, this path integral representation of the CME has utility in its own right. See Appendix 5.A for how it can be used to exactly solve for time-dependent transition probabilities associated with simple chemical reaction systems.

5.4 Path integral derivation of the chemical Langevin equation

In this section, we reinterpret Gillespie's derivation of the CLE in the context of stochastic path integrals, and show explicitly how his two conditions translate in the path integral context. Our central tool will be the Euler-Maclaurin formula [93, 94], which allows one to approximate sums as integrals (plus correction terms). It says that

$$
\sum_{n=a}^{b} f(n) \sim \int_a^b f(x)\, dx + \frac{f(b) + f(a)}{2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} \left[f^{(2k-1)}(b) - f^{(2k-1)}(a) \right]
$$

$$(5.33)$$

where B_{2k} is the $(2k)$th Bernoulli number, and the "$\sim$" symbol is to indicate that we are to interpret the right-hand side as an asymptotic expansion (generically, the infinite sum may not be convergent, but retaining a finite number of terms still usually provides a good approximation to the left-hand side).

The Euler-Maclaurin formula is not an unfamiliar tool in chemical physics, given that it is often used to approximate partition functions [95, 96, 97] to good accuracy in certain regimes (e.g. the high temperature limit). It has also been used for other interesting purposes, like computing Fermi-Dirac integrals [98], and proving the asymptotic equivalence of two descriptions of Coulombic systems in certain potentials [99].

Roughly speaking, we will proceed as follows. Condition (i) will allow us to approximate each sum in Eq. 5.32 as an integral, and to argue that the correction terms are small; meanwhile, condition (ii) will allow us to Taylor expand the $\exp(i\mathbf{p}_k \cdot \boldsymbol{\nu}_j)$ terms in Eq. 5.32 to second order in the momenta $\mathbf{p}_k$. The result of these two approximations will be a MSRJD path integral, which we know from studies of stochastic path integrals [15] to be equivalent to a system of Langevin equations. In particular, it will be equivalent to Eq. 5.7, the CLE.

5.4.1 Only some paths satisfy Gillespie's conditions

Gillespie's first condition (see Sec. 5.2) says that, in a period of time τ, the propensity functions do not change appreciably. Upon some reflection, we realize that this cannot be true for *all* possible trajectories the system might have, assuming the propensity functions have some state-dependence (which, in general, they do). In principle, it is possible that the number of molecules of some species jumps between 1 and 10^{100}, wildly and irregularly, so that there does not exist any time scale on which the propensity functions do not change appreciably. Indeed, all sorts of crazy trajectories are possible *in principle*—but they are overwhelmingly unlikely in practice.

While there certainly exist crazy and pathological paths for which it is hard or impossible to find a time scale τ that satisfies Gillespie's first condition, the requirement is not so stringent for most of the trajectories the system might take. In other words, we will suppose that the first condition is satisfied for the *dominant paths* rather than for all paths.

A similar argument applies to the second condition. This means that, in applying our two conditions, we will no longer be summing over all possible paths (c.f. Eq. 5.32). Instead, we will be summing over all possible paths that satisfy the two conditions, a collection which we will assume includes the dominant or most likely paths.

If we are not summing over *all* possible paths, what does our region of integration look like? To understand this, it is helpful to consider the simple case of a CME with one species and one reaction. Label the number of that species by n, and the propensity function of the single reaction by a.

Imagine starting the system in the state with n_0 molecules and thinking about where it will go (i.e. all possible states n_1) in the next time length τ. For the dominant paths, we assume that the difference $|a(n_1) - a(n_0)|$ is small, so that the propensity function did not change appreciably. But what do we mean by "appreciably"?

In a paper showing his two conditions hold in the thermodynamic limit [22], Gillespie assumed that his first condition meant

$$\frac{|a(n_1) - a(n_0)|}{a(n_0)} \ll 1 \tag{5.34}$$

i.e. that the change in the propensity function on the time scale τ is negligible compared to the size of its original value. This matches the intuition we have about what constitutes a negligible change in population size: for example, if the population size changed by 100 molecules, but the total number of molecules is on the order of 10^5, we imagine that change not to be noticeable.

Here, we can be a little bit more precise than Eq. 5.34. We generally assume that our propensity functions are nicely behaved—in particular, that they are continuous, that they are infinitely differentiable, and that we may freely Taylor expand them. That is, we assume the a_j are analytic functions throughout our domain. Because most propensity functions of interest are polynomials (or at worst, rational functions like Hill functions), and because expressions like the Kramers-Moyal expansion already assume the a_j are smooth, these assumptions do not turn out to be particularly strong.

116

Suppose $a(n_0) > 0$, which is always true in the regime we care about, since we will usually need n sufficiently large; generic monomolecular and bimolecular propensity functions have zeros at $n = 0$ and $n = 1$. Because a is continuous, for any $\epsilon > 0$ we can find a $\delta > 0$ such that

$$|a(n_1) - a(n_0)| < \epsilon a(n_0) \tag{5.35}$$

provided $|n_1 - n_0| < \delta$. In order for the correction terms that arise from applying the Euler-Maclaurin formula (Eq. 5.33) to be negligible, we also want to bound the derivatives of a in a similar fashion.

Analogous conditions apply in the general case, where the a_j may be functions of more than one variable. The moral of the story is that, because of the assumed smooth behavior of the propensity functions, we can find a region where they (and their derivatives) do not vary appreciably. In the simple one-dimensional case, this is an 'interval' $[n_0 - \delta_0^-, n_0 + \delta_0^+] \subseteq \mathbb{N}$ (where we let $\delta_0^- \neq \delta_0^+$ in general since we need $n_0 - \delta_0^-$ and $n_0 + \delta_0^+$ to both be natural numbers); in general, this is the intersection of an open set with a lattice: $U_0(\delta) \cap \mathbb{N}^N \subseteq \mathbb{N}^N$. For convenience, we will use U_0 to denote both the open set and its lattice intersection.

Hence, for a one-dimensional system, we restrict ourselves to paths

$$\sum_{n_1=0}^{\infty} \cdots \sum_{n_{N-1}=0}^{\infty} \to \sum_{n_1=n_0-\delta_0^-}^{n_0+\delta_0^+} \cdots \sum_{n_{N-1}=n_{N-2}-\delta_{N-2}^-}^{n_{N-2}+\delta_{N-2}^+} \tag{5.36}$$

where the δ_i^+ and δ_i^-, as in the discussion above, are chosen so that the propensity functions and their derivatives vary within acceptable bounds. For an arbitrary CME, we restrict ourselves to paths

$$\sum_{\mathbf{n}_1} \cdots \sum_{\mathbf{n}_{T-1}} \to \sum_{\mathbf{n}_1 \in U_0} \cdots \sum_{\mathbf{n}_{T-1} \in U_{T-2}} \tag{5.37}$$

where the sets $U_i \subseteq \mathbb{N}^N$ are chosen similarly.

5.4.2 Coarse-graining time

There is another 'philosophical' point we need to address. Earlier, we imagined breaking up the propagator into T time steps of length Δt, and choosing T to be large enough (or equivalently, Δt to be small enough) that each piece of the propagator was well-approximated by its first-order Taylor expansion (c.f. Eq. 5.26). However, Gillespie's two conditions only apply on the 'coarser' time scale τ. How do we go from time steps of size Δt to time steps of size τ in Eq. 5.32?

There are two straightforward ways we can imagine. The simpler way is to say that, since we are in the business of making approximations *anyway*, we may as well make the approximation that Eq. 5.32 is valid on the time scale τ, and that the terms we neglected when Taylor expanding the propagator do not matter much in the regime where Gillespie's conditions apply.

But there is a more intellectually honest way to proceed. Suppose we originally broke the propagator into $S \cdot T$ time steps, for some natural number S large enough for our derivation

to go through without issue. This means that the time step in our path integral has size $\Delta t := t/(S \cdot T)$. We want to rewrite our path integral in terms of a 'macroscopic' time scale $\tau := t/T$, which corresponds to breaking up the overall time t into T time steps of length τ. Schematically, this means we want to make the following identifications:

$$\mathbf{n}_0 \xrightarrow{\Delta t} \mathbf{n}_1 \xrightarrow{\Delta t} \cdots \xrightarrow{\Delta t} \mathbf{n}_S \ : \ \mathbf{n}_0 \xrightarrow{\tau} \mathbf{n}_1$$

$$\mathbf{n}_S \xrightarrow{\Delta t} \mathbf{n}_{S+1} \xrightarrow{\Delta t} \cdots \xrightarrow{\Delta t} \mathbf{n}_{2S} \ : \ \mathbf{n}_1 \xrightarrow{\tau} \mathbf{n}_2$$

$$\vdots$$

$$\mathbf{n}_{S \cdot (T-1)} \xrightarrow{\Delta t} \mathbf{n}_{S \cdot (T-1)+1} \xrightarrow{\Delta t} \cdots \xrightarrow{\Delta t} \mathbf{n}_{S \cdot T} \ : \ \mathbf{n}_{T-1} \xrightarrow{\tau} \mathbf{n}_T$$

$$(5.38)$$

The argument of the exponential in Eq. 5.32 reads

$$\sum_{k=1}^{S \cdot T} -i\mathbf{p}_k \cdot (\mathbf{n}_k - \mathbf{n}_{k-1}) + \Delta t \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \ . \tag{5.39}$$

Consider the following small piece of this expression:

$$\sum_{k=1}^{S \cdot T} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \ . \tag{5.40}$$

Assuming (on the dominant paths) that the propensity function a_j only changes appreciably on the time scale $\tau = S\Delta t$, we can make the approximation that

$$a_j(\mathbf{n}_0) \approx a_j(\mathbf{n}_1) \approx \cdots \approx a_j(\mathbf{n}_{S-1})$$

$$a_j(\mathbf{n}_S) \approx a_j(\mathbf{n}_{S+1}) \approx \cdots \approx a_j(\mathbf{n}_{2S-1})$$

$$\vdots$$

$$a_j(\mathbf{n}_{S \cdot (T-1)}) \approx a_j(\mathbf{n}_{S \cdot (T-1)+1}) \approx \cdots \approx a_j(\mathbf{n}_{S \cdot T-1})$$

$$(5.41)$$

and rewrite Eq. 5.40 in terms of $a_j(\mathbf{n}_0), a_j(\mathbf{n}_S), a_j(\mathbf{n}_{2S}), ..., a_j(\mathbf{n}_{S \cdot T})$ only. This means that the only places the 'intermediate' time steps (e.g. $\mathbf{n}_1, ..., \mathbf{n}_{S-1}$, or $\mathbf{n}_{S+1}, ..., \mathbf{n}_{2S-1}$) will appear are in the piece that reads

$$\sum_{k=1}^{S \cdot T} -i\mathbf{p}_k \cdot (\mathbf{n}_k - \mathbf{n}_{k-1}) \ . \tag{5.42}$$

Happily, this means that all of the intermediate time steps can be summed over. For example,

$$\sum_{\mathbf{n}_1} \cdots \sum_{\mathbf{n}_{S-1}} \exp\left\{ \sum_{k=1}^{S} -i\mathbf{p}_k \cdot (\mathbf{n}_k - \mathbf{n}_{k-1}) \right\} \approx \delta(\mathbf{p}_1 - \mathbf{p}_2)\delta(\mathbf{p}_2 - \mathbf{p}_3) \cdots \delta(\mathbf{p}_{S-1} - \mathbf{p}_S) \tag{5.43}$$

where the right-hand side is approximate because, due to our restriction of the sum domain in the previous section, the sum representation of the Dirac delta function

$$\frac{1}{(2\pi)^N} \sum_{\mathbf{n}} \exp\left\{ -i\mathbf{n} \cdot (\mathbf{p} - \mathbf{p}') \right\} = \delta(\mathbf{p} - \mathbf{p}') \tag{5.44}$$

118

only approximately applies. After summing over all intermediate time steps and integrating out extraneous $\mathbf{p}_k$ using the delta functions that appear, Eq. 5.40 reads

$$
\begin{aligned}
&\sum_{k=1}^{T} -i\mathbf{p}_k \cdot (\mathbf{n}_k - \mathbf{n}_{k-1}) + S\Delta t \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \\
&= \sum_{k=1}^{T} -i\mathbf{p}_k \cdot (\mathbf{n}_k - \mathbf{n}_{k-1}) + \tau \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1}) \ .
\end{aligned}
\tag{5.45}
$$

Hence, using Gillespie's first condition, we have successfully gone from a path integral with time scale Δt to a path integral with a 'coarser' time scale τ.

5.4.3 Applying condition 1

In this section, we will apply condition (i) in order to convert the sums in Eq. 5.32 to integrals. After restricting our domain to the dominant paths (see Sec. 5.4.1) and coarse-graining time (see Sec. 5.4.2), our current path integral description of the CME reads

$$
P \approx \sum_{\mathbf{n}_1 \in U_0} \cdots \sum_{\mathbf{n}_{T-1} \in U_{T-2}} \int \frac{d\mathbf{p}_1}{(2\pi)^N} \cdots \int \frac{d\mathbf{p}_T}{(2\pi)^N} \ \exp\{-S\tau\}
\tag{5.46}
$$

where we recall that the sets $U_0, ..., U_{T-2}$ cover all trajectories on which Gillespie's two conditions apply, and where we have defined the function (which we can call the "action", in analogy with quantum mechanics)

$$
S := \sum_{k=1}^{T} i\mathbf{p}_k \cdot \left(\frac{\mathbf{n}_k - \mathbf{n}_{k-1}}{\tau} \right) - \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1})
\tag{5.47}
$$

to ease notation. We will proceed using the Euler-Maclaurin formula (Eq. 5.33). As a starting point, consider Eq. 5.46 in one-dimension:

$$
P \approx \sum_{n_1 = n_0 - \delta_0^-}^{n_0 + \delta_0^+} \cdots \sum_{n_{N-1} = n_{N-2} - \delta_{N-2}^-}^{n_{N-2} + \delta_{N-2}^+} \int \frac{dp_1}{2\pi} \cdots \int \frac{dp_T}{2\pi} \ \exp\{-S\tau\}
\tag{5.48}
$$

where the δ_i^- and δ_i^+ are as described in Sec. 5.4.1. Using the Euler-Maclaurin formula, we have

$$
\begin{aligned}
&\sum_{n_1 = n_0 - \delta_0^-}^{n_0 + \delta_0^+} \exp\{-S\tau\} \\
&\approx \int_{n_0 - \delta_0^-}^{n_0 + \delta_0^+} \exp\{-S\tau\} \ dn_1 + \frac{e^{-S(n_0 + \delta_0^+)\tau} + e^{-S(n_0 - \delta_0^-)\tau}}{2} + \sum_{k=1}^{\infty} \frac{B_{2k}}{(2k)!} \frac{d^{2k-1}}{dn_1^{2k-1}} \left[e^{-S\tau} \right]_{n_0 - \delta_0^-}^{n_0 + \delta_0^+} \ .
\end{aligned}
\tag{5.49}
$$

119

Now we need to argue that the correction terms can safely be neglected. Define $\delta := \delta_0^+ + \delta_0^-$. Because the propensity functions don't change vary much in the interval $[n_0 - \delta_0^-, n_0 + \delta_0^+]$ (by Gillespie's first condition), the integral term is roughly

$$\exp\left\{-S(n_0)\tau\right\}\delta \ . \tag{5.50}$$

Meanwhile, the next term is roughly

$$\exp\left\{-S(n_0)\tau\right\} \tag{5.51}$$

which should be negligible compared to the first as long as $\delta \gg 1$. This should certainly be true; if $\delta \sim 1$, our conditions are either too strict, or we are in a regime with too small molecule numbers.

Because the propensity functions a_j do not change much (and because the a_j are nicely behaved, usually monotonic functions in the regime we care about), they are approximately 'flat'. This means that their derivatives $a_j^{(2k-1)}$ are small. For example,

$$\frac{d}{dn_1}\left[e^{-S\tau}\right] = \left[i(p_2 - p_1) + \tau \sum_{j=1}^{M}\left[e^{ip_k\nu_j} - 1\right]a_j'\right]e^{-S\tau} \tag{5.52}$$

so

$$
\begin{aligned}
&\frac{d}{dn_1}\left[e^{-S\tau}\right]_{n_0-\delta_0^-}^{n_0+\delta_0^+} \\
&= \tau \sum_{j=1}^{M}\left[e^{ip_k\nu_j} - 1\right]\left[a_j'(n_0 + \delta_0^+)e^{-S(n_0+\delta_0^+)\tau} - a_j'(n_0 - \delta_0^-)e^{-S(n_0-\delta_0^-)\tau}\right] \\
&\approx \tau e^{-S(n_0)\tau} \sum_{j=1}^{M}\left[e^{ip_k\nu_j} - 1\right]\left[a_j'(n_0 + \delta_0^+) - a_j'(n_0 - \delta_0^-)\right] \\
&\approx 0 \ .
\end{aligned}
\tag{5.53}
$$

In summary, we have

$$\sum_{n_1=n_0-\delta_0^-}^{n_0+\delta_0^+}\exp\{-S\tau\} \approx \int_{n_0-\delta_0^-}^{n_0+\delta_0^+}\exp\{-S\tau\}\ dn_1 \tag{5.54}$$

which means we've successfully converted a sum into an integral. Apply this argument many more times to obtain

$$
\begin{aligned}
&\sum_{n_1=n_0-\delta_0^-}^{n_0+\delta_0^+} \cdots \sum_{n_{T-1}=n_{T-2}-\delta_{T-2}^-}^{n_{T-2}+\delta_{T-2}^+}\exp\{-S\tau\} \\
&\approx \int_{n_0-\delta_0^-}^{n_0+\delta_0^+} dn_1 \cdots \int_{n_{T-2}-\delta_{T-2}^-}^{n_{T-2}+\delta_{T-2}^+} dn_{T-1}\ \exp\{-S\tau\} \ .
\end{aligned}
\tag{5.55}
$$

A similar argument applies to the N species path integral (Eq. 5.46); the only difference is that the Euler-Maclaurin formula must be applied N times for each time step, because we would like to convert N sums to an N-variable integral.

Alternatively, one can argue using the appropriate many sum generalization of the Euler-Maclaurin formula (Eq. 5.33). There is some literature on generalizations of it to sums over polytopes [100, 101, 102] (schematically, shapes in N-dimensional space whose vertices we can imagine as living in $\mathbb{Z}^N$). The main challenge for this approach would be to show that satisfying Gillespie's first condition corresponds to satisfying the requirements associated with the approximation being accurate (which are somewhat more technical than those for the single sum Euler-Maclaurin formula).

The end result of all this is

$$P \approx \int_{U_0} d\mathbf{x}_1 \cdots \int_{U_{T-2}} d\mathbf{x}_{T-1} \int \frac{d\mathbf{p}_1}{(2\pi)^N} \cdots \int \frac{d\mathbf{p}_T}{(2\pi)^N} \ \exp\{-S\tau\} \ . \tag{5.56}$$

where we have relabeled each $\mathbf{n}_k$ as $\mathbf{x}_k$ to (as in Sec. 5.2) emphasize that we are now working with continuous variables. We remark that, if not for the bounds, we would have a Kramers-Moyal path integral (see Sec. V of our earlier paper [15]).

5.4.4 Applying condition 2

Consider the terms in the action S (Eq. 5.47) that look like

$$\left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1\right] a_j(\mathbf{x}_{k-1})\tau \ . \tag{5.57}$$

Condition (ii) tells us that, for the dominant paths, $a_j(\mathbf{x}_{k-1})\tau \gg 1$. In particular, we will assume that it is *so* large that Taylor expanding the term it is multiplied by will have a negligible effect on the overall value[1], i.e.

$$\left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1\right] a_j(\mathbf{x}_{k-1})\tau$$
$$\approx \left[i\mathbf{p}_k \cdot \boldsymbol{\nu}_j - \frac{1}{2}\sum_{\ell=1}^{N}\sum_{\ell'=1}^{N} p_k^\ell p_k^{\ell'} \nu_{j\ell}\nu_{j\ell'}\right] a_j(\mathbf{x}_{k-1})\tau \tag{5.58}$$

where p_k^ℓ is the ℓ-th component of the vector $\mathbf{p}_k$. Thus, we finally obtain

$$S \approx \sum_{k=1}^{T} i\mathbf{p}_k \cdot \left[\frac{\mathbf{x}_k - \mathbf{x}_{k-1}}{\tau} - \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j(\mathbf{x}_{k-1})\right] + \frac{1}{2}\sum_{\ell=1}^{N}\sum_{\ell'=1}^{N}\sum_{j=1}^{M} p_k^\ell p_k^{\ell'} \nu_{j\ell}\nu_{j\ell'} a_j(\mathbf{x}_{k-1}) \tag{5.59}$$

which looks just like the action for the MSRJD path integral (see Sec. V of [15]) corresponding to the chemical Fokker-Planck equation (Eq. 5.8). Our final result for the whole path

---

[1]See Appendix 5.B for a somewhat more rigorous argument justifying this approximation.

integral reads

$$P \approx \int_{U_0} d\mathbf{x}_1 \cdots \int_{U_{T-2}} d\mathbf{x}_{T-1} \int \frac{d\mathbf{p}_1}{(2\pi)^N} \cdots \int \frac{d\mathbf{p}_T}{(2\pi)^N}$$
$$\exp\left\{ -\sum_{k=1}^{T} \left[i\mathbf{p}_k \cdot \left(\frac{\mathbf{x}_k - \mathbf{x}_{k-1}}{\tau} - \sum_{j=1}^{M} \boldsymbol{\nu}_j a_j(\mathbf{x}_{k-1}) \right) + \frac{1}{2} \sum_{\ell=1}^{N} \sum_{\ell'=1}^{N} \sum_{j=1}^{M} p_k^{\ell} p_k^{\ell'} \nu_{j\ell} \nu_{j\ell'} a_j(\mathbf{x}_{k-1}) \right] \tau \right\} \tag{5.60}$$

which looks like the usual Fokker-Planck path integral but with restricted integration bounds.

The appearance of Eq. 5.60 can be compacted somewhat if we define the diffusion tensor $D_{\ell\ell'}$:

$$D_{\ell\ell'}(\mathbf{x}) := \frac{1}{2} \sum_{j=1}^{M} \nu_{j\ell} \nu_{j\ell'} a_j(\mathbf{x}) \ . \tag{5.61}$$

At the CLE/Fokker-Planck level, the diffusion tensor captures all information about a system's noise. It must be positive semidefinite for the Fokker-Planck equation and its corresponding path integral to make sense [103, 104].

Finish the derivation by enlarging our integration domain as much as possible (while keeping the diffusion tensor positive semidefinite), assuming that permitting these additional paths does not substantially contribute to transition probabilities, since they were small enough to neglect in the first place. In general, we do not expect that the appropriate domain for our new continuous variables will be $[0, \infty)^N$, despite the fact that our original domain was $\mathbb{N}^N$. For example, the chemical Langevin equation [1] for the birth-death process (with birth rate k, death rate γ, and steady state mean $\mu := k/\gamma$) reads

$$\dot{x} = k - \gamma x + \sqrt{k + \gamma x}\ \eta(t) \tag{5.62}$$

and is naturally defined on $[-\mu, \infty)$, because there is always some nonzero probability that the noise term will push the system into negative concentrations while its magnitude is greater than or equal to zero, i.e. when $k + \gamma x = \gamma(\mu + x) \geq 0$.

5.4.5   Comparison with the system volume approach

We have shown in the previous few sections how Gillespie's derivation works in a path integral context. Because Gillespie himself [1] compared his approach to ones which rely upon the largeness of the system volume Ω, it is interesting to do that here also. Let us translate the typical system volume approach into path integral language, and see how it compares with the approach we described earlier.

Consider again a CME with N species and M reactions (Eq. 5.1), but this time with the additional physical context that the chemicals interact inside a very large volume Ω. Suppose we rewrite the CME in terms of concentration variables $x_i := n_i/\Omega$ for all $i = 1, ..., N$. The change in variables will lead to the probability density function $P(\mathbf{n}, t)$ increasing by a factor

of Ω^N:

$$P(\mathbf{n}, t)d\mathbf{n} = \Omega^N P(\mathbf{n}, t)d\mathbf{x} = P(\mathbf{x}, t)d\mathbf{x}$$
$$\implies P(\mathbf{x}, t) = \Omega^N P(\mathbf{n}, t) \,. \tag{5.63}$$

Gillespie used rigorous microphysical arguments [105, 106, 107, 1] to show that the volume-dependence of the propensity functions for monomolecular, bimolecular, and trimolecular reactions goes like

$$a_j(\mathbf{n}) = \Omega \, \tilde{a}_j(\mathbf{x}) \tag{5.64}$$

where the adjusted propensity functions $\tilde{a}_j$ are volume-independent. Using Eq. 5.63 and Eq. 5.64, our original CME path integral (Eq. 5.32) can be rewritten as

$$P = \lim_{T \to \infty} \Omega^N \sum_{\mathbf{n}_1} \cdots \sum_{\mathbf{n}_{T-1}} \int \frac{d\mathbf{p}_1}{(2\pi)^N} \cdots \int \frac{d\mathbf{p}_T}{(2\pi)^N}$$
$$\exp\left\{ \sum_{k=1}^{T} -i\Omega \, \mathbf{p}_k \cdot (\mathbf{x}_k - \mathbf{x}_{k-1}) + \Omega\Delta t \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] \tilde{a}_j(\mathbf{x}_{k-1}) \right\} \,. \tag{5.65}$$

Now add in $T - 1$ factors of Ω^N / Ω^N:

$$P = \lim_{T \to \infty} \left[\frac{1}{\Omega^N} \sum_{\mathbf{n}_1} \right] \cdots \left[\frac{1}{\Omega^N} \sum_{\mathbf{n}_{T-1}} \right] \int \left(\frac{\Omega}{2\pi} \right)^N d\mathbf{p}_1 \cdots \int \left(\frac{\Omega}{2\pi} \right)^N d\mathbf{p}_T$$
$$\exp\left\{ \sum_{k=1}^{T} -i\Omega \, \mathbf{p}_k \cdot (\mathbf{x}_k - \mathbf{x}_{k-1}) + \Omega\Delta t \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] \tilde{a}_j(\mathbf{x}_{k-1}) \right\} \,. \tag{5.66}$$

Riemann sums will play the role that the Euler-Maclaurin formula did (i.e. converting sums to integrals) in our earlier derivation. Recall that the (right endpoint) Riemann sum for a function f on $[0, b]$ reads [108]

$$\int_0^b f(x)dx \approx \sum_{i=0}^{N} f(i\Delta x)\Delta x \tag{5.67}$$

where $\Delta x = b/N$. If we take $b \to \infty$ and $N \to \infty$ in such a way that Δx remains constant, we can write

$$\int_0^\infty f(x)dx \approx \sum_{i=0}^{\infty} f(i\Delta x)\Delta x \,. \tag{5.68}$$

The corresponding N-dimensional result is

$$\int_0^\infty dx_1 \cdots \int_0^\infty dx_N \, f(x)$$
$$\approx \sum_{i_1=0}^{\infty} \cdots \sum_{i_N=0}^{\infty} f(i_1\Delta x, \cdots, i_N\Delta x)(\Delta x)^N \,. \tag{5.69}$$

Since the inverse system volume $1/\Omega$ seems to play the role of Δx in Eq. 5.66, we can use this Riemann sum result to approximate each sum as

$$\frac{1}{\Omega^N} \sum_{\mathbf{n}} \approx \int_0^\infty dx_1 \cdots \int_0^\infty dx_N \tag{5.70}$$

so that our path integral is now

$$
P = \lim_{T \to \infty} \int d\mathbf{x}_1 \cdots \int d\mathbf{x}_{T-1} \int \left(\frac{\Omega}{2\pi}\right)^N d\mathbf{p}_1 \cdots \int \left(\frac{\Omega}{2\pi}\right)^N d\mathbf{p}_T
$$
$$
\exp\left\{ \sum_{k=1}^T -i\Omega \, \mathbf{p}_k \cdot (\mathbf{x}_k - \mathbf{x}_{k-1}) + \Omega \Delta t \sum_{j=1}^M \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] \tilde{a}_j(\mathbf{x}_{k-1}) \right\}. \tag{5.71}
$$

Now we can argue just as we did in Sec. 5.4.4. Because we are taking Ω to be extraordinarily large in the thermodynamic limit,

$$
\Omega \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] \tilde{a}_j(\mathbf{x}_{k-1}) \tau
$$
$$
\approx \Omega \left[i\mathbf{p}_k \cdot \boldsymbol{\nu}_j - \frac{1}{2} \sum_{\ell=1}^N \sum_{\ell'=1}^N p_k^\ell p_k^{\ell'} \nu_{j\ell} \nu_{j\ell'} \right] \tilde{a}_j(\mathbf{x}_{k-1}) \tau \tag{5.72}
$$

i.e. Ω is so large that above term does not change much in value when Taylor expanded to second order in $\mathbf{p}_k$. Finally, we have

$$
P = \lim_{T \to \infty} \int d\mathbf{x}_1 \cdots \int d\mathbf{x}_{T-1} \int \left(\frac{\Omega}{2\pi}\right)^N d\mathbf{p}_1 \cdots \int \left(\frac{\Omega}{2\pi}\right)^N d\mathbf{p}_T
$$
$$
\exp\left\{ -\Omega \sum_{k=1}^T \left[i\mathbf{p}_k \cdot \left(\frac{\mathbf{x}_k - \mathbf{x}_{k-1}}{\Delta t} - \sum_{j=1}^M \boldsymbol{\nu}_j \tilde{a}_j(\mathbf{x}_{k-1}) \right) + \frac{1}{2} \sum_{\ell=1}^N \sum_{\ell'=1}^N \sum_{j=1}^M p_k^\ell p_k^{\ell'} \nu_{j\ell} \nu_{j\ell'} \tilde{a}_j(\mathbf{x}_{k-1}) \right] \Delta t \right\} \tag{5.73}
$$

which is the same as the result from Sec. 5.4.4 (c.f. Eq. 5.59) but with additional factors of Ω. It also exactly matches the system volume MSRJD path integral for the Fokker-Planck equation (c.f. Eq. 94 in Sec. V of [15]). In other words, we have indeed derived a path integral equivalent to a set of Langevin equations/a Fokker-Planck equation. Moreover, it is equivalent to the *same* set of Langevin equations that Eq. 5.59 is (as is easily seen after changing back from concentration variables to the original number variables)—although the integration bounds on the path integral are different here.

Given that this approach was significantly simpler (in both a technical and conceptual sense), why bother with Gillespie's derivation? There are a few good reasons.

- The approximations provided by Eq. 5.68 and Eq. 5.69 are more mathematically dubious than the Euler-Maclaurin formula (Eq. 5.33), which is well-studied and has precisely expressed error bounds.

- The thermodynamic limit may not apply to most biochemical systems of interest, given that molecule numbers are often large but not overwhelmingly so, and that the system volume (for example, of a cell) is not large enough to prevent crowding [109, 110, 111, 112] and boundary effects [113, 114, 115] from being important.

- The system volume approach only *applies* when our CME describes a well-stirred, dilute mix of chemicals held at fixed temperature in a very large box—but the CLE is known to be a useful approximate description of all sorts of other stochastic systems (e.g. spiking neurons, fluctuating population dynamics models, stock options). In these other situations, there is no clear notion of a control parameter analogous to Ω.

- The system volume approach misses the subtlety of the integration bounds associated with the chemical Langevin/chemical Fokker-Planck equations; as we pointed out at the end of the previous section, it is a nontrivial issue that the domain of the approximating CLE will generally not be $[0, \infty)^N$.

5.5 Discussion

We constructed an original path integral description of the CME, and applied Gillespie's conditions (suitably interpreted) to it in order to derive a path integral known to be equivalent to the CLE. In some sense, the difference between the system size approach and Gillespie's approach to deriving the CLE is the difference between approximating sums as integrals via Riemann sums, and via the Euler-Maclaurin formula. As discussed at the end of the previous section, while both approximation techniques can be valid in the appropriate circumstances, the Euler-Maclaurin formula is more generally applicable and has better characterized correction terms.

It is interesting to note that, although we began with an exact path integral that involved taking the limit $\Delta t \to 0$ (Eq. 5.32), we coarse-grained time to end up with a path integral with fixed time step τ that does not get taken to zero (Eq. 5.60). This leads to another sense in which the CLE is only an approximate description, since a true CLE/Fokker-Planck path integral (see [15]) also involves taking the limit $\Delta t \to 0$. However, the idea of a 'macroscopic' timescale was addressed by Gillespie himself in his original paper [1]. There, he offered an analogy to current in an electric circuit: we can freely write and manipulate the derivative $I := dq/dt$, and think about the limit $dt \to 0$, *provided* we understand that we are not taking it to be *so* small that shot noise effects start to matter.

Because our argument applied to each reaction/propensity function separately, it can in principle be used construct path integrals for hybrid systems. In other words, just as Harris et al. [28, 29, 30] do, we can suppose that Gillespie's two conditions apply only to a subset of all reactions or species, and construct a path integral in which some species/reactions are treated CLE-style, while others are treated CME-style. Indeed, there should be a path integral way

to view all of the hybrid constructions—based on molecule numbers or separations of time scales—referenced in Sec. 5.2. These path integrals could then be used to extract large deviation results.

It is unclear if Gillespie's conditions could be applied to the Doi-Peliti path integral [84, 85, 86, 87] in order to recover a CLE-equivalent path integral. Part of the difficulty is that the Doi-Peliti construction involves integrating over coherent states, which contribute integrals over the whole real line in the expression for the propagator [86]; it is not necessarily straightforward to associate these with sums or integrals over state space.

Although we only used the path integral representation of the CME constructed in Sec. 7.3 to derive the CLE, it can also be used as a tool in its own right for the same purposes other path integral representations are often used for: namely, finding exact solutions [116, 117, 118], constructing asymptotic or perturbative approximations to transition probabilities and moments [119], computing least action paths associated with particular state transitions, and enabling a variational method for numerically computing transition probabilities and least action paths [70]. In cases where Gillespie's first condition applies, and the domain of the path integral can be restricted, the numerical evaluation of Eq. 5.32 becomes even easier. Though our argument does not offer a constructive prescription for the restricted integration domains $U_0, ..., U_{T-2}$, one can in principle bootstrap the path integral by running Gillespie algorithm simulations beforehand to estimate reasonable domains.

5.6   Conclusion

The chemical Langevin equation is usually derived using Gillespie's two conditions, or large system volume arguments; as we described, both methods have clear path integral analogues. Our results suggest that path integrals offer a useful and mathematically precise way of thinking about the relationship between different levels of approximation (e.g. CME and CLE), and about coarse-graining biochemical models more generally.

5.A   Sample path integral calculations

One can get a feel for a given path integral by using it to exactly calculate transition probabilities for simple systems. In this appendix, we use the path integral described by Eq. 5.32 to compute exact time-dependent transition probabilities for a (i) pure birth process, a (ii) pure death process, and a (iii) chemical birth-death process. For more information on the chemical birth-death process, and for analogous path integral calculations that are valid in

the limit where molecule numbers can be treated as continuous, see [116] and [117].

These systems can also be solved using the Doi-Peliti path integral, the most commonly used path integral description of the CME—see Vastola [118] for a guide. It is interesting to note that the transition probability derivations presented in this appendix are simpler in some ways; for example, one does not need to introduce coherent states or special scalar products, and transition probabilities can be computed without computing the generating function first.

The following path integral calculations involve the evaluation of many contour integrals. The resolution of identity introduced in Eq. 5.28, which leads to the 'momentum' integrals going from negative infinity to positive infinity, yields contour integrals with an infinite number of poles. In order to make these calculations somewhat simpler, we instead use the equally valid resolution of the identity given by

$$\delta(n_1 - n_2) = \int_{-\pi}^{\pi} \frac{dp}{2\pi} \, e^{-ip(n_1 - n_2)} \tag{5.74}$$

in one dimension, and

$$\delta(\mathbf{n}_1 - \mathbf{n}_2) = \int_{-\pi}^{\pi} \cdots \int_{-\pi}^{\pi} \frac{dp_1 \cdots dp_N}{(2\pi)^N} \, e^{-i\mathbf{p}\cdot(\mathbf{n}_1 - \mathbf{n}_2)} \tag{5.75}$$

in more than one dimension. This means that our CME path integral is still given by Eq. 5.32, but with the $\mathbf{p}_k$ integrals all going from $-\pi$ to π.

5.A.1 The pure birth process

The pure birth process models a species that is randomly created at some rate. It is characterized by the chemical reaction

$$\varnothing \xrightarrow{k} X \tag{5.76}$$

where k is the birth rate. This reaction corresponds to the CME

$$\frac{\partial P(n,t)}{\partial t} = k \left[P(n-1,t) - P(n,t) \right] \tag{5.77}$$

where $P(n,t)$ is the probability that the system has n X molecules at time t (with $n \in \{0,1,2,...\}$). For this system, our path integral for the transition probability $P(n_f,t;n_0,0)$ (c.f. Eq. 5.32) reads

$$P = \lim_{T \to \infty} \sum_{n_1} \cdots \sum_{n_{T-1}} \int \frac{dp_1}{2\pi} \cdots \int \frac{dp_T}{2\pi} \, \exp\left\{ \sum_{\ell=1}^{T} -ip_\ell(n_\ell - n_{\ell-1}) + k\Delta t \left(e^{ip_\ell} - 1 \right) \right\} \tag{5.78}$$

where $\Delta t := t/T$. First, since

$$- k\Delta t \sum_{\ell=1}^{T} = -kt \ , \tag{5.79}$$

127

the overall answer has a prefactor $\exp(-kt)$. Next, organizing terms by the n_ℓ, we have

$$ip_1n_0 - ip_Tn_f + in_1(p_2 - p_1) + in_2(p_3 - p_2) + \cdots + in_{T-1}(p_T - p_{T-1}) \ . \tag{5.80}$$

Each sum over n_ℓ (for $\ell = 1, ..., T-1$) just corresponds to summing a geometric series:

$$\sum_{n_\ell=0}^{\infty} \left[e^{i(p_{\ell+1}-p_\ell)} \right]^{n_\ell} = \frac{1}{1 - e^{i(p_{\ell+1}-p_\ell)}} \ . \tag{5.81}$$

Now consider the integral over p_1, which reads

$$\int_{-\pi}^{\pi} \frac{dp_1}{2\pi} \frac{e^{ip_1n_0 + k\Delta t e^{ip_1}}}{1 - e^{i(p_2-p_1)}} \ . \tag{5.82}$$

Switching variables to $z = e^{ip_1}$, we have the contour integral

$$\oint \frac{dz}{2\pi i} \frac{z^{n_0}e^{k\Delta t z}}{z - e^{ip_2}} \ , \tag{5.83}$$

whose integrand has a simple pole at e^{ip_2}. Although Cauchy's theorem does not technically apply in this case, since the pole lies on the circular contour, we can imagine deforming the contour or integrand slightly (e.g. by taking $e^{ip_2} \to (1 - \epsilon)e^{ip_2}$) so that Cauchy's theorem does apply, allowing us to evaluate the integral in the usual way[2]. Doing so, we obtain

$$e^{ip_2n_0 + k\Delta t e^{ip_2}} \ , \tag{5.84}$$

which indicates that the overall effect of doing the contour integral was to implement the constraint that $p_1 = p_2$. Similarly, the overall effect of doing the integral over p_ℓ (for $\ell = 2, ..., T-1$), which schematically reads

$$\int_{-\pi}^{\pi} \frac{dp_\ell}{2\pi} \frac{f(p_\ell)}{1 - e^{i(p_{\ell+1}-p_\ell)}} \tag{5.85}$$

for some function $f(p_\ell)$, is to implement the constraint that $p_\ell = p_{\ell+1}$. Hence, after doing the integrals over $p_1, ..., p_{T-1}$, we have

$$\int_{-\pi}^{\pi} \frac{dp_T}{2\pi} e^{-ip_T(n_f - n_0) + kte^{ip_T}} \ . \tag{5.86}$$

Define $\Delta n := n_f - n_0$ and change variables to $z = e^{ip_T}$. Then we have

$$\oint \frac{dz}{2\pi i} \frac{e^{ktz}}{z^{\Delta n+1}} \ . \tag{5.87}$$

---

[2]These arguments can be made more rigorous if one wishes. Because path integrals are somewhat mathematically dubious in the first place, regularization techniques like these are required for the integrals involved to be well-defined. Ultimately, the proof is 'in the pudding': if we get the right answer, which can easily be verified by substituting it directly into Eq. 5.77, these abuses can be excused.

For $\Delta n < 0$, this integral has no poles, so $P(n_f, t; n_0, 0) = 0$ in that case. For $\Delta n \geq 0$, a standard application of Cauchy's integral formula yields that the result is

$$\frac{1}{\Delta n!} \frac{d^{\Delta n}}{dz^{\Delta n}} \left[e^{ktz} \right]_{z=0} = \frac{(kt)^{\Delta n}}{\Delta n!} \ . \tag{5.88}$$

Hence, including the prefactor e^{-kt}, our result for the transition probability is

$$P(n_f, t; n_0, 0) = \frac{(kt)^{\Delta n} e^{-kt}}{\Delta n!} \tag{5.89}$$

i.e. a Poisson distribution.

5.A.2 The pure death process

The pure death process models a species that randomly degrades at some rate. It is characterized by the chemical reaction

$$X \xrightarrow{\gamma} \varnothing \tag{5.90}$$

where γ is the death rate. This reaction corresponds to the CME

$$\frac{\partial P(n, t)}{\partial t} = \gamma \left[(n+1)P(n+1, t) - nP(n, t) \right] \tag{5.91}$$

where $P(n, t)$ is the probability that the system has n X molecules at time t (with $n \in \{0, 1, 2, ...\}$). This time, our path integral for the transition probability $P(n_f, t; n_0, 0)$ reads

$$P = \lim_{T \to \infty} \sum_{n_1} \cdots \sum_{n_{T-1}} \int \frac{dp_1}{2\pi} \cdots \int \frac{dp_T}{2\pi} \ \exp\left\{ \sum_{\ell=1}^{T} -ip_\ell(n_\ell - n_{\ell-1}) + \gamma \Delta t \ n_{\ell-1} \left(e^{-ip_\ell} - 1 \right) \right\} \ . \tag{5.92}$$

The terms that involve n_ℓ (for some $\ell = 1, ..., T-1$) look like

$$i n_\ell \left[p_{\ell+1} - p_\ell - i\gamma \Delta t \left(e^{-ip_{\ell+1}} - 1 \right) \right] \ , \tag{5.93}$$

so the sum over n_ℓ yields

$$\sum_{n_\ell=0}^{\infty} \left[e^{i \left[p_{\ell+1} - p_\ell - i\gamma \Delta t \left(e^{-ip_{\ell+1}} - 1 \right) \right]} \right]^{n_\ell} = \frac{1}{1 - e^{i \left[p_{\ell+1} - p_\ell - i\gamma \Delta t \left(e^{-ip_{\ell+1}} - 1 \right) \right]}} \ . \tag{5.94}$$

The integral over p_1 reads

$$\int_{-\pi}^{\pi} \frac{dp_1}{2\pi} \frac{e^{ip_1 n_0 + \gamma \Delta t n_0 \left(e^{-ip_1} - 1 \right)}}{1 - e^{i \left[p_2 - p_1 - i\gamma \Delta t \left(e^{-ip_2} - 1 \right) \right]}} \ . \tag{5.95}$$

We can argue just as in the previous calculation to show that the net effect of doing this integral is to implement the constraint

$$p_1 = p_2 - i\gamma \Delta t \left(e^{-ip_2} - 1 \right) \ . \tag{5.96}$$

129

Similarly, the effect of doing the integrals over $p_2, ..., p_{T-1}$ is to implement the constraints

$$p_\ell = p_{\ell+1} - i\gamma\Delta t \left(e^{-ip_{\ell+1}} - 1\right) \tag{5.97}$$

for $\ell = 2, ..., T-1$. While this recurrence relation is probably not solvable in closed form, we can do a trick to evaluate p_ℓ (for $\ell = 1, ..., T-1$) in terms of p_T (the only momentum variable not yet integrated over). Notice that Eq. 5.97 looks like a 'backwards' Euler time step corresponding to the ordinary differential equation (ODE)

$$\dot{p} = -i\gamma \left(e^{-ip} - 1\right) . \tag{5.98}$$

This approximation becomes exact in the $\Delta t \to 0$ limit, which is the limit we are interested in. Eq. 5.98 has the solution

$$e^{ip(t)} = 1 - Ce^{-\gamma t} , \tag{5.99}$$

where C must be determined from the initial condition $p(0) = p_T$ (since the ODE runs 'backwards'). Doing so, we have

$$e^{ip_0} = 1 - (1 - e^{ip_T})e^{-\gamma t} \tag{5.100}$$

in the small Δt limit, where p_0 is defined via

$$e^{ip_0} := e^{ip_1 + \gamma\Delta t\left(e^{-ip_1} - 1\right)} . \tag{5.101}$$

The intuition is that p_0 is what we reach after starting at p_T and taking T time steps of size Δt. Our last integral reads

$$\int_{-\pi}^{\pi} \frac{dp_T}{2\pi} e^{ip_0 n_0} e^{-ip_T n_T} = \int_{-\pi}^{\pi} \frac{dp_T}{2\pi} \left[(1 - e^{-\gamma t}) + e^{-\gamma t}e^{ip_T}\right]^{n_0} e^{-ip_T n_T} . \tag{5.102}$$

Define $q := 1 - e^{-\gamma t}$ and change variables to $z = e^{ip_T}$ so that we have

$$\oint \frac{dz}{2\pi i} \frac{[q + (1-q)z]^{n_0}}{z^{n_T+1}} = \sum_{j=0}^{n_0} \binom{n_0}{j} (1-q)^j q^{n_0-j} \oint \frac{dz}{2\pi i} \frac{1}{z^{n_T-j+1}} . \tag{5.103}$$

The contour integral is nonzero only for the term with $j = n_T$. Then $P(n_f, t; n_0, 0) = 0$ for $n_T > n_0$, and

$$P(n_f, t; n_0, 0) = \binom{n_0}{n_f} \left[e^{-\gamma t}\right]^{n_f} \left[1 - e^{-\gamma t}\right]^{n_0 - n_f} \tag{5.104}$$

otherwise, i.e. we have a binomial distribution.

130

5.A.3 The chemical birth-death process

The chemical birth-death process models a species that is randomly created and randomly degrades. Its list of chemical reactions is

$$\varnothing \xrightarrow{k} X$$
$$X \xrightarrow{\gamma} \varnothing$$

(5.105)

where k is the birth rate and γ is the death rate. This reaction corresponds to the CME

$$\frac{\partial P(n,t)}{\partial t} = k\left[P(n-1,t) - P(n,t)\right] + \gamma\left[(n+1)P(n+1,t) - nP(n,t)\right]$$

(5.106)

where $P(n,t)$ is the probability that the system has n X molecules at time t (with $n \in \{0,1,2,...\}$). This path integral reads

$$\lim_{T\to\infty} \sum_{n_1}\cdots\sum_{n_{T-1}} \int \frac{dp_1}{2\pi} \cdots \int \frac{dp_T}{2\pi} \; \exp\left\{\sum_{\ell=1}^{T} -ip_\ell(n_\ell - n_{\ell-1}) + k\Delta t\left(e^{ip_\ell} - 1\right) + \gamma\Delta t \; n_{\ell-1}\left(e^{-ip_\ell} - 1\right)\right\}.$$

(5.107)

The terms that involve n_ℓ (for some $\ell = 1, ..., T-1$) look just the same as in the previous subsection, so the sums over the n_ℓ evaluate to the same answer. The same ODE constraint is also enforced by the integrals over the p_ℓ, leaving the only difference between this problem and the previous one in the evaluation of the final integral. We have

$$\int_{-\pi}^{\pi} \frac{dp_T}{2\pi} \; e^{ip_0 n_0} e^{-ip_T n_T} e^{k\sum_{\ell=1}^{T}\left(e^{ip_\ell}-1\right)\Delta t}$$

$$= \int_{-\pi}^{\pi} \frac{dp_T}{2\pi} \; \left[q + (1-q)e^{ip_T}\right]^{n_0} e^{-ip_T n_T} e^{k\sum_{\ell=1}^{T}\left(e^{ip_\ell}-1\right)\Delta t}$$

(5.108)

where p_0 is defined as in Eq. 5.101, and $q := 1 - e^{-\gamma t}$ as before. Observe that there is a term in the integrand, due to the birth reaction, that was not present in Eq. 5.102. In order to proceed, we must evaluate that term in the small Δt limit. Note,

$$k\sum_{\ell=1}^{T}\left(e^{ip_\ell} - 1\right)\Delta t \approx k\int_{0}^{t} dt\;\left(e^{ip(t)} - 1\right) = -\left[1 - e^{ip_T}\right]\mu q$$

(5.109)

where we have exploited the link between Riemann sums and integrals, used the expression for $p(t)$ from Eq. 5.99, and defined $\mu := k/\gamma$. Now our integral reads

$$e^{-\mu q} \int_{-\pi}^{\pi} \frac{dp_T}{2\pi} \; \left[q + (1-q)e^{ip_T}\right]^{n_0} e^{-ip_T n_T} e^{\mu q e^{ip_T}} .$$

(5.110)

Changing variables to $z = e^{ip_T}$ yields

$$e^{-\mu q} \oint \frac{dz}{2\pi i} \frac{[q + (1-q)z]^{n_0} e^{\mu q z}}{z^{n_T+1}} = e^{-\mu q} \sum_{j=0}^{n_0} \binom{n_0}{j} q^{n_0-j}(1-q)^j \oint \frac{dz}{2\pi i} \frac{e^{\mu q z}}{z^{n_T-j+1}}$$

(5.111)

131

which can be evaluated via Cauchy's integral formula to obtain the result

$$P(n_f, t; n_0, 0) = \sum_{j=0}^{\min(n_0, n_f)} \binom{n_0}{j} q^{n_0 - j}(1 - q)^j \cdot \frac{(\mu q)^{n_f - j} e^{-\mu q}}{(n_f - j)!} \ . \tag{5.112}$$

This agrees with the result obtained using the Doi-Peliti path integral (c.f. Eq. 77 of [118]), although the calculation was arguably less complicated here.

5.B   An alternative argument for Taylor expanding the action

In section 5.4.4, we argued that Gillespie's second condition allows us to say that

$$\left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{x}_{k-1})\tau$$
$$\approx \left[i\mathbf{p}_k \cdot \boldsymbol{\nu}_j - \frac{1}{2} \sum_{\ell=1}^{N} \sum_{\ell'=1}^{N} p_k^\ell p_k^{\ell'} \nu_{j\ell} \nu_{j\ell'} \right] a_j(\mathbf{x}_{k-1})\tau \ . \tag{5.113}$$

In this appendix, we present a slightly more careful alternative argument for the validity of this important approximation. Because the term we are approximating appears in the argument of the path integral's exponential (c.f. Eq. 5.32), we are *really* trying to say that

$$\exp\left\{ \sum_{j=1}^{M} \left[e^{i\mathbf{p}_k \cdot \boldsymbol{\nu}_j} - 1 \right] a_j(\mathbf{n}_{k-1})\tau \right\}$$
$$\approx \exp\left\{ i\mathbf{p}_k \cdot \left(\sum_{j=1}^{M} \boldsymbol{\nu}_j a_j(\mathbf{n}_{k-1})\tau \right) - \frac{1}{2} \sum_{j=1}^{M} (\mathbf{p}_k \cdot \boldsymbol{\nu}_j)^2 a_j(\mathbf{n}_{k-1})\tau \right\} \ . \tag{5.114}$$

Let us now show that this approximation holds. To ease notation, rewrite this as

$$\exp\left\{ \sum_{j=1}^{M} \left[e^{i\mathbf{p} \cdot \boldsymbol{\nu}_j} - 1 \right] a_j \right\}$$
$$\approx \exp\left\{ i\mathbf{p} \cdot \left(\sum_{j=1}^{M} \boldsymbol{\nu}_j a_j \right) - \frac{1}{2} \sum_{j=1}^{M} (\mathbf{p} \cdot \boldsymbol{\nu}_j)^2 a_j \right\} \ , \tag{5.115}$$

where we have removed the superfluous index on $\mathbf{p}$, and used the shorthand $a_j := a_j(\mathbf{n}_{k-1})\tau$. Start by Taylor expanding the exponentials:

$$\exp\left\{\sum_{j=1}^{M}\left[e^{i\mathbf{p}\cdot\boldsymbol{\nu}_j} - 1\right]a_j\right\}$$

$$=e^{-(a_1+\cdots a_M)}e^{e^{i\mathbf{p}\cdot\boldsymbol{\nu}_1}a_1}\cdots e^{e^{i\mathbf{p}\cdot\boldsymbol{\nu}_M}a_M} \tag{5.116}$$

$$= \sum_{x_1,\ldots,x_M} \frac{(a_1)^{x_1}}{x_1!}e^{-a_1}\cdots\frac{(a_M)^{x_M}}{x_M!}e^{-a_M}e^{i\mathbf{p}\cdot(\boldsymbol{\nu}_1 x_1+\cdots+\boldsymbol{\nu}_M x_M)}$$

where the sum over each x_j runs from zero to infinity. Notice that the factor corresponding to the jth reaction looks like a Poisson distribution; since $a_j \gg 1$ (condition (ii)), it can be approximated as Gaussian in the usual way:

$$\frac{(a_j)^{x_j}}{x_j!}e^{-a_j} \approx \frac{1}{\sqrt{2\pi a_j}}e^{-\frac{(x_j-a_j)^2}{2a_j}}\ . \tag{5.117}$$

Now the sum over each x_j can be approximated as an integral using the Euler-Maclaurin formula, and we can expand the domain of integration without significantly changing the result since the Gaussian function is sharply peaked:

$$\sum_{x_j=0}^{\infty} \frac{1}{\sqrt{2\pi a_j}}e^{-\frac{(x_j-a_j)^2}{2a_j}+i\mathbf{p}\cdot\boldsymbol{\nu}_j x_j}$$

$$\approx \int_0^{\infty} dx_j\ \frac{1}{\sqrt{2\pi a_j}}e^{-\frac{(x_j-a_j)^2}{2a_j}+i\mathbf{p}\cdot\boldsymbol{\nu}_j x_j}$$

$$\approx \int_{-\infty}^{\infty} dx_j\ \frac{1}{\sqrt{2\pi a_j}}e^{-\frac{(x_j-a_j)^2}{2a_j}+i\mathbf{p}\cdot\boldsymbol{\nu}_j x_j} \tag{5.118}$$

$$=e^{i\mathbf{p}\cdot\boldsymbol{\nu}_j a_j-\frac{(\mathbf{p}\cdot\boldsymbol{\nu}_j)^2}{2}a_j}\ .$$

Putting these factors together for each j yields Eq. 5.115, the desired approximation. While the argument we just presented is somewhat more technical than the one in the main text, it also mirrors Gillespie's original derivation of the CLE more closely: we essentially approximated many Poisson distributions as normal distributions, just as Gillespie did (c.f. Eq. 5.4).

133

Bibliography

[1] D. T. Gillespie, The Journal of Chemical Physics **113**, 297 (2000).

[2] H. Risken and H. D. Vollmer, Zeitschrift für Physik B Condensed Matter **66**, 257 (1987).

[3] J. Łuczka, P. Hänggi, and A. Gadomski, Phys. Rev. E **51**, 2933 (1995).

[4] H. Risken, "Fokker-planck equation," in The Fokker-Planck Equation (Springer Berlin Heidelberg, Berlin, Heidelberg, 1996) pp. 63–95.

[5] N. G. v. Kampen, Canadian Journal of Physics **39**, 551 (1961).

[6] N. G. V. Kampen, Stochastic Processes in Physics and Chemistry (Third Edition), third edition ed., North-Holland Personal Library (Elsevier, Amsterdam, 2007).

[7] R. F. Pawula, Phys. Rev. **162**, 186 (1967).

[8] D. T. Gillespie, The Journal of Chemical Physics **115**, 1716 (2001).

[9] D. T. Gillespie and L. R. Petzold, The Journal of Chemical Physics **119**, 8229 (2003).

[10] M. Rathinam, L. R. Petzold, Y. Cao, and D. T. Gillespie, The Journal of Chemical Physics **119**, 12784 (2003).

[11] Y. Cao, L. R. Petzold, M. Rathinam, and D. T. Gillespie, The Journal of Chemical Physics **121**, 12169 (2004).

[12] Y. Cao, D. T. Gillespie, and L. R. Petzold, The Journal of Chemical Physics **123**, 054104 (2005).

[13] Y. Cao, D. T. Gillespie, and L. R. Petzold, The Journal of Chemical Physics **124**, 044109 (2006).

[14] Y. Cao, D. T. Gillespie, and L. R. Petzold, The Journal of Chemical Physics **126**, 224101 (2007).

[15] J. J. Vastola and W. R. Holmes, arXiv e-prints , arXiv:1909.12990 (2019), arXiv:1909.12990 [cond-mat.stat-mech] .

[16] K. G. Wilson, Phys. Rev. B **4**, 3174 (1971).

[17] K. G. Wilson, Phys. Rev. B **4**, 3184 (1971).

[18] M. E. Peskin and D. V. Schroeder, An Introduction to quantum field theory (Addison-Wesley, Reading, USA, 1995).

[19] M. D. Schwartz, Quantum Field Theory and the Standard Model (Cambridge University Press, 2014).

[20] H. Georgi, Ann. Rev. Nucl. Part. Sci. **43**, 209 (1993).

[21] D. T. Gillespie, A. Hellander, and L. R. Petzold, The Journal of Chemical Physics **138**, 170901 (2013).

[22] D. T. Gillespie, The Journal of Physical Chemistry B **113**, 1640 (2009).

[23] D. Schnoerr, G. Sanguinetti, and R. Grima, The Journal of Chemical Physics **141**, 024103 (2014).

[24] R. Grima, P. Thomas, and A. V. Straube, The Journal of Chemical Physics **135**, 084103 (2011).

[25] A. Duncan, S. Liao, T. Vejchodský, R. Erban, and R. Grima, Phys. Rev. E **91**, 042111 (2015).

[26] A. Ceccato and D. Frezzato, The Journal of Chemical Physics **148**, 064114 (2018).

[27] S. Winkelmann and C. Schutte, The Journal of Chemical Physics **147**, 114115 (2017).

[28] L. A. Harris and P. Clancy, The Journal of Chemical Physics **125**, 144107 (2006).

[29] L. A. Harris, A. M. Piccirilli, E. R. Majusiak, and P. Clancy, Phys. Rev. E **79**, 051906 (2009).

[30] K. A. Iyengar, L. A. Harris, and P. Clancy, The Journal of Chemical Physics **132**, 094101 (2010).

[31] E. L. Haseltine and J. B. Rawlings, The Journal of Chemical Physics **117**, 6959 (2002).

[32] H. Salis and Y. Kaznessis, The Journal of Chemical Physics **122**, 054103 (2005).

[33] F. Wu, T. Tian, J. B. Rawlings, and G. Yin, The Journal of Chemical Physics **144**, 174112 (2016).

[34] E. W. Wallace, D. T. Gillespie, K. R. Sanft, and L. R. Petzold, IET Syst Biol **6**, 102 (2012).

[35] R. Grima, The Journal of Chemical Physics **136**, 154105 (2012).

[36] J. Wang, K. Zhang, L. Xu, and E. Wang, Proceedings of the National Academy of Sciences **108**, 8257 (2011).

[37] J. X. Zhou, M. D. S. Aliyu, E. Aurell, and S. Huang, Journal of the Royal Society, Interface **9**, 3539 (2012).

[38] P. Zhou and T. Li, The Journal of Chemical Physics **144**, 094109 (2016).

[39] M. L. Simpson, C. D. Cox, and G. S. Sayler, Journal of Theoretical Biology **229**, 383 (2004).

[40] M. Hinczewski and D. Thirumalai, The Journal of Physical Chemistry B **120**, 6166 (2016).

[41] T. Brett and T. Galla, The Journal of Chemical Physics **140**, 124112 (2014).

[42] T. Choi, M. R. Maurya, D. M. Tartakovsky, and S. Subramaniam, The Journal of Chemical Physics **133**, 165101 (2010).

[43] S. Rudiger, Physics Reports **534**, 39 (2014), stochastic models of intracellular calcium signals.

[44] H. Li, Z. Hou, and H. Xin, Phys. Rev. E **71**, 061916 (2005).

[45] Y. Huang, S. Rudiger, and J. Shuai, Physical Biology **12**, 061001 (2015).

[46] E. Wallace, M. Benayoun, W. van Drongelen, and J. D. Cowan, PLOS ONE **6**, 1 (2011).

[47] R. P. Feynman, Reviews of Modern Physics **20**, 367 (1948).

[48] R. Feynman, A. Hibbs, and D. Styer, Quantum Mechanics and Path Integrals, Dover Books on Physics (Dover Publications, 2010).

[49] L. S. Schulman, Techniques and Applications of Path Integration (Dover Publications, 1981).

[50] H. Kleinert, Path Integrals in..., 5th ed. (World Scientific, 2009).

[51] R. Shankar, Principles of Quantum Mechanics, 2nd ed. (Springer US, 2011).

[52] R. Shankar, Quantum Field Theory and Condensed Matter: An Introduction (Cambridge University Press, 2017).

[53] A. Caldeira and A. Leggett, Physica A: Statistical Mechanics and its Applications **121**, 587 (1983).

[54] A. W. C. Lau and T. C. Lubensky, Phys. Rev. E **76**, 011123 (2007).

[55] K. Asheichyk, A. P. Solon, C. M. Rohwer, and M. Krüger, The Journal of Chemical Physics **150**, 144111 (2019).

[56] J. Wang, K. Zhang, H. Lu, and E. Wang, Biophysical Journal **89**, 1612 (2005).

[57] J. Wang, K. Zhang, H. Lu, and E. Wang, Phys. Rev. Lett. **96**, 168101 (2006).

[58] S. Orioli, S. a Beccara, and P. Faccioli, The Journal of Chemical Physics **147**, 064108 (2017).

[59] V. Linetsky, Computational Economics **11**, 129 (1997).

[60] L. Ingber, Physica A: Statistical Mechanics and its Applications **283**, 529 (2000).

[61] T. Spanio, J. Hidalgo, and M. A. Muñoz, Phys. Rev. E **96**, 042301 (2017).

[62] A. Kamenev, B. Meerson, and B. Shklovskii, Phys. Rev. Lett. **101**, 268103 (2008).

[63] U. C. Täuber, Journal of Physics: Conference Series **319**, 012019 (2011).

[64] U. C. Täuber, Journal of Physics A: Mathematical and Theoretical **45**, 405002 (2012).

[65] S. R. Serrao and U. C. Täuber, Journal of Physics A: Mathematical and Theoretical **50**, 404005 (2017).

[66] G. K. Ocker, K. Josić, E. Shea-Brown, and M. A. Buice, PLOS Computational Biology **13**, 1 (2017).

[67] A. Iomin, Phys. Rev. E **100**, 010104(R) (2019).

[68] R. Baravalle, O. A. Rosso, and F. Montani, Physica A: Statistical Mechanics and its Applications **486**, 986 (2017).

[69] P. C. Bressloff, The Journal of Mathematical Neuroscience (JMN) **5**, 4 (2015).

[70] P. C. Bressloff and J. M. Newby, Phys. Rev. E **89**, 042701 (2014).

[71] S.-W. Qiu and C. C. Chow, Phys. Rev. E **98**, 062414 (2018).

[72] M. Buice and C. Chow, Frontiers in Computational Neuroscience **7**, 162 (2013).

[73] C. Li and J. Wang, Journal of The Royal Society Interface **11**, 20140774 (2014).

[74] C. Li, T. Hong, and Q. Nie, Phys. Chem. Chem. Phys. **18**, 17949 (2016).

[75] C. Li and J. Wang, PLOS Computational Biology **9**, 1 (2013).

[76] J. Wang, K. Zhang, and E. Wang, The Journal of Chemical Physics **133**, 125103 (2010).

[77] J. Newby, arXiv e-prints , arXiv:1412.8446 (2014), arXiv:1412.8446 [q-bio.MN] .

[78] V. Elgart and A. Kamenev, Phys. Rev. E **70**, 041106 (2004).

[79] V. Elgart and A. Kamenev, Phys. Rev. E **74**, 041101 (2006).

[80] T. Li and F. Lin, Journal of Physics A: Mathematical and Theoretical **49**, 135204 (2016).

[81] F. Benitez, C. Duclut, H. Chaté, B. Delamotte, I. Dornic, and M. A. Muñoz, Phys. Rev. Lett. **117**, 100601 (2016).

[82] P. Thomas, C. Fleck, R. Grima, and N. Popović, Journal of Physics A: Mathematical and Theoretical **47**, 455007 (2014).

[83] A. Lazarescu, T. Cossetto, G. Falasco, and M. Esposito, The Journal of Chemical Physics **151**, 064117 (2019).

[84] M. Doi, J. Phys. A **9**, 1479 (1976).

[85] M. Doi, J. Phys. A **9**, 1465 (1976).

[86] L. Peliti, Journal de Physique **46**, 1469 (1985).

[87] D. C. Mattis and M. L. Glasser, Rev. Mod. Phys. **70**, 979 (1998).

[88] P. C. Martin, E. D. Siggia, and H. A. Rose, Phys. Rev. A **8**, 423 (1973).

[89] H.-K. Janssen, Zeitschrift für Physik B Condensed Matter **23**, 377 (1976).

[90] DE DOMINICIS, C., J. Phys. Colloques **37**, C1 (1976).

[91] C. De Dominicis and L. Peliti, Phys. Rev. B **18**, 353 (1978).

[92] J. A. Hertz, Y. Roudi, and P. Sollich, Journal of Physics A: Mathematical and Theoretical **50**, 033001 (2016).

[93] M. Abramowitz, Handbook of Mathematical Functions (Dover Publications, Inc., New York, NY, USA, 1974).

[94] V. Lampret, Mathematics Magazine **74**, 109 (2001).

[95] R. Pathria and P. D. Beale, Statistical Mechanics (Third Edition), third edition ed. (Academic Press, Boston, 2011).

[96] R. Kayser and J. E. Kilpatrick, The Journal of Chemical Physics **68**, 1511 (1978).

[97] Y. Liao and B. Miao, The Journal of Chemical Physics **142**, 164903 (2015).

[98] G. Rzadkowski and S. Lepkowski, Journal of Scientific Computing **35**, 63 (2008).

[99] J. Cioslowski and J. Albin, The Journal of Chemical Physics **139**, 114109 (2013).

[100] Y. Karshon, S. Sternberg, and J. Weitsman, Proceedings of the National Academy of Sciences **100**, 426 (2003).

[101] V. Guillemin and S. Sternberg, Annales de l'Institut Fourier **57**, 2183 (2007).

[102] T. Tate, Journal of Functional Analysis **260**, 501 (2011).

[103] C. Gardiner, Stochastic Methods: A Handbook for the Natural and Social Sciences, Springer Series in Synergetics (Springer Berlin Heidelberg, 2009).

[104] H. Risken, The Fokker-Planck equation: methods of solution and applications, 2nd ed., Springer series in synergetics No. v. 18 (Springer-Verlag, New York, 1996).

[105] D. T. Gillespie, Journal of Computational Physics **22**, 403 (1976).

[106] D. T. Gillespie, Physica A: Statistical Mechanics and its Applications **188**, 404 (1992).

[107] D. T. Gillespie, Markov Processes: An Introduction for Physical Scientists (Elsevier Science, 1992).

[108] W. Rudin, Principles of mathematical analysis, 3rd ed. (McGraw-Hill Book Co., New York-Auckland-Düsseldorf, 1976) pp. x+342, international Series in Pure and Applied Mathematics.

[109] R. Ellis, Trends in Biochemical Sciences **26**, 597 (2001).

[110] A. P. Minton, Journal of Biological Chemistry **276**, 10577 (2001).

[111] F. Höfling and T. Franosch, Reports on Progress in Physics **76**, 046602 (2013).

[112] G. Rivas and A. P. Minton, Trends in Biochemical Sciences **41**, 970 (2016).

[113] Y. A. Komarova, I. A. Vorobjev, and G. G. Borisy, Journal of Cell Science **115**, 3527 (2002).

[114] J. Buceta, Journal of The Royal Society Interface **14**, 20170316 (2017).

[115] W. R. Holmes, Biophysical Journal **116**, 1538 (2019).

[116] J. J. Vastola, arXiv e-prints , arXiv:1910.09117 (2019), arXiv:1910.09117 [q-bio.QM] .

[117] J. J. Vastola, arXiv e-prints , arXiv:1910.10807 (2019), arXiv:1910.10807 [cond-mat.stat-mech] .

[118] J. J. Vastola, arXiv e-prints , arXiv:1911.00978 (2019), arXiv:1911.00978 [q-bio.QM] .

[119] C. C. Chow and M. A. Buice, The Journal of Mathematical Neuroscience **5**, 8 (2015).

Chapter 6

Analytic solution of chemical master equations involving gene switching.
I: Representation theory and diagrammatic approach to exact solution

How can we *solve* the CME? This chapter introduces a ladder operator technique, and finds that the dynamics of RNA production and degradation looks 'bosonic', and that the dynamics of a gene switching between two states of transcriptional activity looks 'fermionic'. Treating systems involving both monomolecular reactions (e.g. RNA splicing and degradation) and gene switching is important for explaining single cell RNA counts data, so theoretical advances in understanding these kinds of problems may be particularly valuable.

Abstract: The chemical master equation (CME), which describes the discrete and stochastic molecule number dynamics associated with biological processes like transcription, is difficult to solve analytically. It is particularly hard to solve for models involving bursting/gene switching, a biological feature that tends to produce heavy-tailed single cell RNA counts distributions. In this paper, we present a novel method for computing exact and analytic solutions to the CME in such cases, and use these results to explore approximate solutions valid in different parameter regimes, and to compute observables of interest. Our method leverages tools inspired by quantum mechanics, including ladder operators and Feynman-like diagrams, and establishes close formal parallels between the dynamics of bursty transcription, and the dynamics of bosons interacting with a single fermion. We focus on two problems: (i) the chemical birth-death process coupled to a switching gene/the telegraph model, and (ii) a model of transcription and multistep splicing involving a switching gene and an arbitrary number of downstream splicing steps. We work out many special cases, and exhaustively explore the special functionology associated with these problems. This is Part I in a two-part series of papers; in Part II, we explore an alternative solution approach that is more useful for numerically solving these problems, and apply it to parameter inference on simulated RNA counts data.

6.1 Introduction

This is the first in a series of two papers on solving chemical master equations involving a switching gene, particularly in the context of bursty models of transcription and splicing. This paper describes a novel theoretical approach for solving these problems analytically, and explores the mathematical properties of the solutions we obtain in detail; the second paper [1] describes an alternative approach that is more appropriate for efficiently solving these problems on a computer. In other words, this paper tells us how to think about the solutions to these models, while the next paper offers a fast and stable way to work with them in practice—especially in the context of parameter inference.

The chemical master equation (CME) describes how the number of molecules in a system (e.g. RNA in a single cell) changes over time [2, 3, 4, 5, 6, 7, 8, 9, 10]. It treats molecule numbers as discrete rather than continuous, and dynamics as stochastic rather than deterministic; especially when there are small numbers of molecules, fluctuations due to the randomness associated with diffusion and molecular interactions become important.

The equation itself—which can either be described as a differential-difference equation, or as a countably infinite number of coupled ordinary differential equations—is typically challenging to solve both analytically and numerically. Monomolecular reaction networks, treated by Jahnke and Huisinga in a landmark 2007 paper [11, 12], comprise the largest class of systems for which analytic solutions to the CME are known. But this class does not include systems with many biologically common types of chemical reactions, like protein synthesis and binding, whose presence in a system's reaction list yields CMEs that can at present be solved only in very special cases [13] (see e.g. [2, 14, 15, 16, 17, 12, 18, 19, 20]).

One biological feature that yields CMEs that are hard to solve analytically is *bursting* [21, 22, 23, 24, 25, 26, 27, 28, 29]. Bursting refers to the fact that RNA transcription does not seem to happen at a continuous rate in many contexts, but instead in discrete bursts. This is usually modeled by assuming that genes have at least two states that they randomly switch between, and transcribe at different rates when in different states.

Recent improvements in fluorescence and sequencing-based experimental methods have enabled the identification of sub-molecular features of individual RNA [30, 31] in single cells. Specifically, one can detect individual introns, which provides the ability to interrogate the dynamics of RNA splicing—long implicated in the physiological control of gene expression [32, 33, 34, 35, 36]. If the kinetics of these systems are to be understood in terms of a mechanistic model, one must be able to solve CMEs involving both splicing and gene switching, since bursting is known to be one of the salient features.

In this paper, we present a new and fairly general approach to solving CMEs involving gene switching, and first illustrate our approach by exactly solving a telegraph model-like problem that describes mRNA production and degradation given a gene that randomly switches between two states. This is a particularly useful benchmark, as the problem has been solved before [37, 38, 39, 40, 41]. We then use our method to tackle a much more complicated problem: deriving the full steady state distribution of a model of bursty transcription and splicing, assuming a linear path graph and arbitrarily many splicing steps. This path graph

model is relatively simple, but plausible and directly applicable to modeling genes that have few introns and a single significant isoform. We validate our work using standard numerical approaches (exact stochastic simulations [42, 43, 5] and finite state projection [44, 45]).

The paper is organized as follows. In Sec. 6.2, we present our main results and outline the theoretical approach described in later sections. This section is intended to allow readers to appreciate the utility of this new method, and to use the results we have derived with it, without having to enmesh themselves in its technical details. In Sec. 6.3, we describe a ladder operator-based approach to solving the chemical birth-death process and pure gene switching, and derive the ladder operators that allow us to effectively reframe the coupled problem. In Sec. 6.4, we construct a diagrammatic series solution to the coupled problem, and explain how to draw the relevant Feynman-like diagrams. In Sec. 6.5, we examine special cases of our series solution, and show how to extract them from the more general result. In Sec. 6.6, we study the high copy number/continuous concentration limit of our solution, and show that it corresponds to solving a certain Fokker-Planck equation coupled to a switching gene. In Sec. 6.7, we generalize our theoretical approach to tackle the multistep splicing problem. In Sec. 7.2, we consider the problem of implementing our solutions on a computer, and validate them against numerical approaches. Finally, we discuss how our approach could be fruitfully extended and applied to study more general splicing dynamics in Sec. 7.6.

6.2   Main results and outline of approach

In this section, we summarize our main analytic results. Readers only interested in our results, and not in methodological details, may find this section the most useful.

We have analytically solved two problems: (i) a telegraph model-like problem involving one RNA species and a switching gene, and (ii) a generalization of it intended to model the dynamics of bursty transcription with multiple downstream splicing steps (Fig 6.1). The problem statements and main results for each are presented here in their own subsections (with additional, more specialized results in Sec. 6.4 - 6.7). The third subsection provides a brief guide to the three kinds of qualitative solution behavior one can expect. The last subsection provides a roadmap for the new theoretical methodology explored in the rest of the paper.
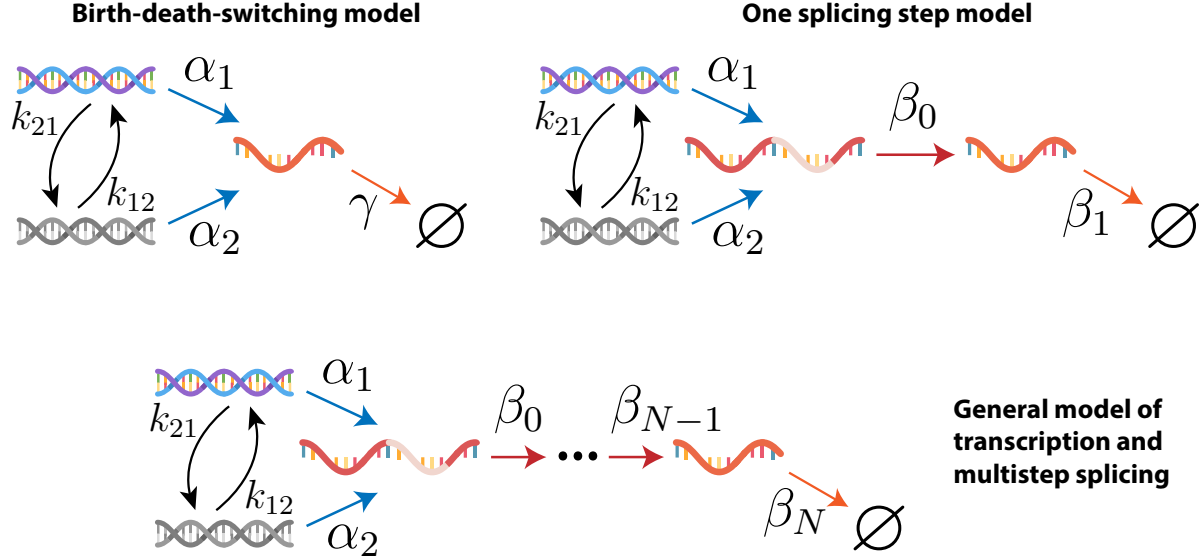
Figure 6.1: The models we will examine in this paper. The 1 species version is the birth-death-switching model (usually called the telegraph model elsewhere). We generalize it to include some arbitrary number N of downstream splicing steps, so that the most general model has $N + 1$ distinct species of RNA overall.

6.2.1 Chemical birth-death process coupled to a switching gene

Problem statement

Let X denote the RNA species, and G_1 and G_2 the two possible gene states. Using the formalism of the chemical master equation (CME), RNA production and degradation coupled to a switching gene can be modeled using the list of chemical reactions

$$
\begin{aligned}
G_1 &\xrightarrow{k_{21}} G_2 & G_2 &\xrightarrow{k_{12}} G_1 \\
G_1 &\xrightarrow{\alpha_1} G_1 + X & G_2 &\xrightarrow{\alpha_2} G_2 + X \\
X &\xrightarrow{\gamma} \varnothing
\end{aligned}
\tag{6.1}
$$
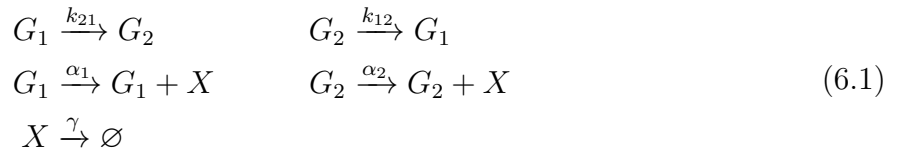
where k_{21} and k_{12} parameterize the rates of gene switching, α_1 and α_2 parameterize the transcription rate in each gene state, and γ parameterizes the RNA's degradation rate. This model is often called the 'telegraph model', especially when one of the production rates is zero. For reasons that will become clear in the following sections, we prefer to think of it as two distinct problems coupled together: the chemical birth-death process (describing RNA production and degradation given a constitutively active gene) coupled to a switching gene.

The stochastic dynamics of this system are completely characterized by the probability distribution $P(x, S, t)$, which indicates the probability that the system has $x \in \mathbb{N} := \{0, 1, 2, ...\}$ RNA molecules and is in gene state $S \in \{1, 2\}$ at some time t (given some specified initial condition $P(x, S, 0)$ which has no effect on steady state information). The

probability distributions $P(x, 1, t)$ and $P(x, 2, t)$ are determined by the CME

$$
\begin{aligned}
\frac{\partial P(x, 1, t)}{\partial t} = & - k_{21} P(x, 1, t) + k_{12} P(x, 2, t) \\
& + \alpha_1 \left[P(x-1, 1, t) - P(x, 1, t) \right] + \gamma \left[(x+1) P(x+1, 1, t) - x P(x, 1, t) \right] \\
\frac{\partial P(x, 2, t)}{\partial t} = & \; k_{21} P(x, 1, t) - k_{12} P(x, 2, t) \\
& + \alpha_2 \left[P(x-1, 2, t) - P(x, 2, t) \right] + \gamma \left[(x+1) P(x+1, 2, t) - x P(x, 2, t) \right]
\end{aligned}
\tag{6.2}
$$

which is known to be difficult to solve analytically. Heuristically, the difficulty comes from coupling two qualitatively very different kinds of stochastic processes: a discrete switching process on two states, and a biased random walk on a countably infinite molecular number space.

We will not compute $P(x, 1, t)$ or $P(x, 2, t)$ in full generality, but rather the functions

$$
\begin{aligned}
P_{ss}(x) & := \lim_{t \to \infty} P(x, 1, t) + P(x, 2, t) \\
\psi_{ss}(g) & := \sum_{x=0}^{\infty} P_{ss}(x) g^x
\end{aligned}
\tag{6.3}
$$

i.e. the steady state probability distribution and steady state (analytic) generating function for a complex variable g, both marginalized over gene state. Several distinct objects labeled 'generating functions' will appear throughout this paper, so we distinguish this one with the adjective 'analytic' to avoid confusion.

Steady state results

Define the derived quantities

$$
\begin{aligned}
\alpha_{eff} & := \alpha_1 \left(\frac{k_{12}}{s} \right) + \alpha_2 \left(\frac{k_{21}}{s} \right) \\
\mu & := \frac{\alpha_1 \left(\frac{k_{12}}{s} \right) + \alpha_2 \left(\frac{k_{21}}{s} \right)}{\gamma} = \frac{\alpha_{eff}}{\gamma} \\
\Delta\alpha & := \alpha_1 - \alpha_2 \\
c_3 & := \frac{k_{12} k_{21}}{s^2} \\
c_4 & := \frac{k_{21} - k_{12}}{s} \; ,
\end{aligned}
\tag{6.4}
$$

where α_{eff} is the effective transcription rate (which takes into account the average amount of time spent in each gene state in the long time limit), μ is the effective mean molecule number, $\Delta\alpha$ is the production rate difference, and c_3 and c_4 are non-dimensional coefficients that appear in our solution. Also define the 2×2 transfer matrix T (c.f. Sec. 6.4.2) as

$$
T = \begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ c_3 & c_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{k_{12} k_{21}}{s^2} & \frac{k_{21} - k_{12}}{s} \end{pmatrix} .
\tag{6.5}
$$

144

Using $C_k(x, \mu)$ to denote the kth Charlier polynomial (see Appendix 6.A for the definition and properties of these objects), the exact analytic results for $P_{ss}(x)$ and $\psi_{ss}(g)$ can be written

$$\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} = 1 + \sum_{k=2}^{\infty} (\Delta\alpha)^k \, C_k(x,\mu) \sum_{i_1,\dots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma)} \frac{T_{i_{k-1}i_{k-2}}}{[(k-1)\gamma + i_{k-1}s]} \cdots \frac{T_{i_10}}{(\gamma + i_1 s)}$$

$$= 1 + (\Delta\alpha)^2 C_2(x,\mu) \cdot \frac{c_3}{(s+\gamma)2\gamma} + (\Delta\alpha)^3 C_3(x,\mu) \cdot \frac{c_3 c_4}{(s+\gamma)(s+2\gamma)3\gamma} + \cdots \tag{6.6}$$

$$\frac{\psi_{ss}(g)}{e^{\mu(g-1)}} = 1 + \sum_{k=2}^{\infty} [(\Delta\alpha)(g-1)]^k \sum_{i_1,\dots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_10}}{(\gamma + i_1 s)}$$

$$= 1 + \frac{[(\Delta\alpha)(g-1)]^2 c_3}{(s+\gamma)2\gamma} + \frac{[(\Delta\alpha)(g-1)]^3 c_3 c_4}{(s+\gamma)(s+2\gamma)3\gamma} + \cdots \tag{6.7}$$

where $\text{Poiss}(x,\mu)$ denotes the Poisson distribution

$$\text{Poiss}(x,\mu) := \frac{\mu^x e^{-\mu}}{x!} \ . \tag{6.8}$$

See Fig. 6.2 for plots in three representative cases, and a comparison against numerical results obtained via finite state projection [44, 45]. Although it is not obvious, our results agree with previously published results on this problem, e.g. results due to Iyer-Biswas et al. [39], Huang et al. [40], and Cao and Grima [41]; see Appendix 6.E for a mathematical proof of this. For special cases and limits of these formulas, see Sec. 6.5.

The mean and variance of this distribution are given by

$$\langle x \rangle = \mu$$

$$\sigma_x^2 = \langle x^2 \rangle - \langle x \rangle^2 = \mu + \frac{(\Delta\alpha)^2 c_3}{(s+\gamma)\gamma} \ . \tag{6.9}$$

6.2.2 Multistep splicing coupled to a switching gene

Problem statement

Let X_0 denote unspliced/nascent RNA, X_1 denote RNA that has experienced one splicing step (i.e. one intron has been removed), X_2 denote RNA that has experienced two splicing steps, and so on. Assuming there are N splicing steps, let X_N denote the fully processed/mature RNA. As before, let G_1 and G_2 denote the two possible gene states. Transcription coupled to a switching gene, with multiple downstream splicing steps, can be modeled using the list of chemical reactions

$$\begin{aligned} G_1 &\xrightarrow{k_{21}} G_2 & G_2 &\xrightarrow{k_{12}} G_1 \\ G_1 &\xrightarrow{\alpha_1} G_1 + X_0 & G_2 &\xrightarrow{\alpha_2} G_2 + X_0 \\ X_0 &\xrightarrow{\beta_0} X_1 \xrightarrow{\beta_1} X_2 \xrightarrow{\beta_2} \cdots \xrightarrow{\beta_{N-1}} X_N \xrightarrow{\beta_N} \varnothing \end{aligned} \tag{6.10}$$
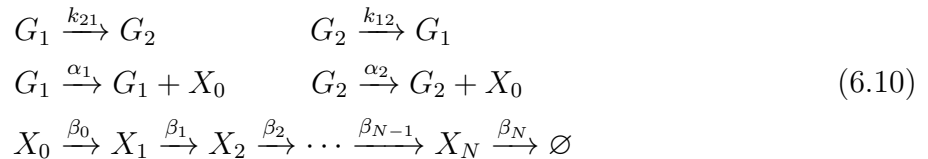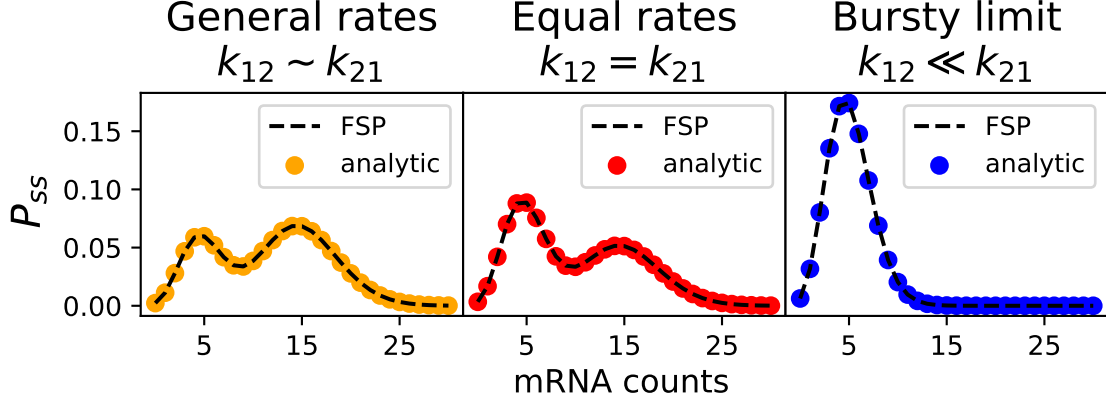
Figure 6.2: The birth-death-switching steady state distribution $P_{ss}(x)$ in three representative cases: general rates ($k_{12} \sim k_{21}$), equal rates ($k_{12} = k_{21}$), and very unequal rates ($k_{12} \ll k_{21}$). The black dotted lines correspond to the finite state projection (FSP) result, while the colored dots correspond to the result from our solution approach.

where k_{21} and k_{12} parameterize the rates of gene switching, α_1 and α_2 parameterize the transcription rate in each gene state, β_i parameterizes the rate of the ith splicing step (for $0 \le i < N$), and β_N parameterizes the mature RNA's degradation rate.

Denote the state of the system using $\mathbf{x} := (x_0, x_1, ..., x_N) \in \mathbb{N}^{N+1}$ and $S \in \{1, 2\}$. Use $\boldsymbol{\epsilon}_j$ to denote the vector with a 1 in the jth place and zeros elsewhere. The CME corresponding to this model reads

$$
\begin{aligned}
\frac{\partial P(\mathbf{x}, 1, t)}{\partial t} =& -k_{21}P(\mathbf{x}, 1, t) + k_{12}P(\mathbf{x}, 2, t) \\
& + \alpha_1 \left[P(\mathbf{x} - \boldsymbol{\epsilon}_0, 1, t) - P(\mathbf{x}, 1, t) \right] \\
& + \sum_{j=0}^{N-1} \beta_j \left[(x_j + 1)P(\mathbf{x} + \boldsymbol{\epsilon}_j - \boldsymbol{\epsilon}_{j+1}, 1, t) - x_j P(\mathbf{x}, 1, t) \right] \\
& + \beta_N \left[(x_N + 1)P(\mathbf{x} + \boldsymbol{\epsilon}_N, 1, t) - x_N P(\mathbf{x}, 1, t) \right] \\
\frac{\partial P(\mathbf{x}, 2, t)}{\partial t} =& \, k_{21}P(\mathbf{x}, 1, t) - k_{12}P(\mathbf{x}, 2, t) \\
& + \alpha_2 \left[P(\mathbf{x} - \boldsymbol{\epsilon}_0, 2, t) - P(\mathbf{x}, 2, t) \right] \\
& + \sum_{j=0}^{N-1} \beta_j \left[(x_j + 1)P(\mathbf{x} + \boldsymbol{\epsilon}_j - \boldsymbol{\epsilon}_{j+1}, 2, t) - x_j P(\mathbf{x}, 2, t) \right] \\
& + \beta_N \left[(x_N + 1)P(\mathbf{x} + \boldsymbol{\epsilon}_N, 2, t) - x_N P(\mathbf{x}, 2, t) \right] \ .
\end{aligned}
\tag{6.11}
$$

For this more complicated many-variable problem, it is helpful to introduce the notation

$$\mathbf{v^x} := v_0^{x_0} \cdots v_N^{x_N}$$

$$\mathbf{x}! := x_0! \cdots x_N!$$

$$\sum_{\mathbf{x}} := \sum_{x_0=0}^{\infty} \cdots \sum_{x_N=0}^{\infty} \tag{6.12}$$

$$\int d\mathbf{x} := \int dx_0 \cdots \int dx_N .$$

We aim to compute the steady state probability distribution and (analytic) generating function both marginalized over gene state, i.e.

$$P_{ss}(\mathbf{x}) := \lim_{t \to \infty} P(\mathbf{x}, 1, t) + P(\mathbf{x}, 2, t)$$

$$\psi_{ss}(\mathbf{g}) := \sum_{\mathbf{x}} P_{ss}(\mathbf{x}) \, \mathbf{g^x} \tag{6.13}$$

where $\mathbf{g} \in \mathbb{C}^{N+1}$.

Steady state results

We will reuse several derived quantities from Sec. 6.2.1. One that must be modified is the effective mean, which in this case is a vector $\boldsymbol{\mu} := (\mu_0, \mu_1, ..., \mu_N)$ whose components are defined as

$$\mu_i = \frac{\alpha_{eff}}{\beta_i} . \tag{6.14}$$

We can use it to define the multivariate Poisson distribution

$$\text{Poiss}(\mathbf{x}, \boldsymbol{\mu}) := \frac{\boldsymbol{\mu^x} e^{-\boldsymbol{\mu} \cdot \mathbf{1}}}{\mathbf{x}!} = \frac{\mu_0^{x_0} e^{-\mu_0}}{x_0!} \cdots \frac{\mu_N^{x_N} e^{-\mu_N}}{x_N!} . \tag{6.15}$$

We must also define the vector $\mathbf{q} = (q_0, q_1, ..., q_N)$, and the family of vectors $\mathbf{v}^{(k)}$ (for $0 \leq k \leq N$). The entries of the former are defined via

$$q_0 = 1$$

$$q_1 = -\frac{\beta_0}{\beta_1 - \beta_0}$$

$$q_2 = \frac{\beta_0 \beta_1}{(\beta_2 - \beta_0)(\beta_2 - \beta_1)} \tag{6.16}$$

$$q_j = (-1)^j \frac{\beta_0 \cdots \beta_{j-1}}{(\beta_j - \beta_0) \cdots (\beta_j - \beta_{j-1})} \quad (\, 1 \leq j \leq N \,)$$

while the entries of the latter are defined via

$$v_j^{(k)} = \begin{cases} 0 & j < k \\ 1 & j = k \\ \frac{\beta_k \cdots \beta_{j-1}}{(\beta_{k+1} - \beta_k) \cdots (\beta_j - \beta_k)} & j > k . \end{cases} \tag{6.17}$$

Use $V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu})$ to denote the multistep orthogonal polynomial associated with the nonnegative integer vector $\mathbf{n} \in \mathbb{N}^{N+1}$; these polynomials are new, and are defined and discussed in Appendix 6.D. Our exact analytic result for $P_{ss}(\mathbf{x})$ is

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} = 1 + \sum_{k=2}^{\infty} \sum_{\text{paths } \mathbf{j}} (\Delta\alpha)^k \, \mathbf{q^n} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \sum_{i_1,\dots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{[\beta_{j_1} + \dots + \beta_{j_k}]} \cdots \frac{T_{i_10}}{[\beta_{j_1} + i_1 s]} \,,$$

$$(6.18)$$

where the sum requires some explanation. For each $k \geq 2$, we sum over all paths $\mathbf{j}$ of length k with elements in $\{0, 1, ..., N\}$, i.e. $\mathbf{j} := (j_1, ..., j_k) \in \{0, 1, ..., N\}^k$. The vector $\mathbf{n} := (n_0, n_1, ..., n_N)$ counts the number of times that each integer i appears in a path, i.e.

$$n_r = \sum_{i \text{ s.t. } j_i = r} 1 \qquad (6.19)$$

for all $0 \leq r \leq N$. For example, in the case of two splicing steps ($N = 1$), for $k = 1$ there are just two paths: $\mathbf{j}^{(1)} = (0)$ and $\mathbf{j}^{(2)} = (1)$. For $k = 2$, there are four:

$$\begin{aligned}
\mathbf{j}^{(1)} &= (0, 0) \\
\mathbf{j}^{(2)} &= (0, 1) \\
\mathbf{j}^{(3)} &= (1, 0) \\
\mathbf{j}^{(4)} &= (1, 1) \ .
\end{aligned} \qquad (6.20)$$

In general, for N splicing steps and a set length $k \geq 1$, there will be $(N+1)^k$ distinct paths that must be summed over. The (analytic) generating function is

$$\frac{\psi_{ss}(\mathbf{g})}{e^{\boldsymbol{\mu} \cdot (\mathbf{g}-1)}} = 1 + \sum_{k=2}^{\infty} (\Delta\alpha)^k \sum_{\text{paths } \mathbf{j}^{(k)}} \prod_{m=0}^{N} \left[q_m (\mathbf{g}-1) \cdot \mathbf{v}^{(m)} \right]^{n_m} \sum_{i_1,\dots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{[\beta_{j_1} + \dots + \beta_{j_k}]} \cdots \frac{T_{i_10}}{[\beta_{j_1} + i_1 s]}$$

$$= 1 + (\Delta\alpha)^2 \sum_{a=0}^{N} \sum_{b=0}^{N} \left[q_a (\mathbf{g}-1) \cdot \mathbf{v}^{(a)} \right] \left[q_b (\mathbf{g}-1) \cdot \mathbf{v}^{(b)} \right] \frac{c_3}{(\beta_a + \beta_b)(\beta_a + s)} + \cdots .$$

$$(6.21)$$

For special cases and limits, see Sec. 6.7.4. The means, variances, and covariances of this distribution are

$$\langle x_i \rangle = \mu_i$$

$$\sigma_i^2 = \langle x_i^2 \rangle - \langle x_i \rangle^2 = \mu_i + (\Delta\alpha)^2 \sum_{a=0}^{N} \sum_{b=0}^{N} \frac{2 q_a q_b v_i^{(a)} v_i^{(b)} c_3}{(\beta_a + \beta_b)(\beta_a + s)} \qquad (6.22)$$

$$\text{Cov}(x_i, x_j) = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle = (\Delta\alpha)^2 \sum_{a=0}^{N} \sum_{b=0}^{N} \frac{q_a q_b v_i^{(a)} v_j^{(b)} c_3}{(\beta_a + \beta_b)(\beta_a + s)} \ .$$

148

6.2.3 Brief guide to qualitative solution behaviors

For developing intuition about the behavior of these distributions, it is helpful to keep three edge cases in mind: (i) the Poisson limit, (ii) the Poisson mixture limit, and (iii) the negative binomial limit. Our exact solutions provably reduce to each of these three special cases in various biologically plausible limits (see Sec. 6.5 and 6.7.4), and can be viewed as interpolating between them in intermediate parameter regimes.

When the sum of the gene switching rates, $s = k_{12} + k_{21}$, is much larger than the degradation and splicing rates, switching happens so quickly that the RNA only 'feels' the effective transcription rate α_{eff}. In this parameter regime, the steady state solutions reduce to Poisson distributions—exactly the behavior we would expect if there were no gene switching.

When s is much smaller than the degradation and splicing rates, switching happens so rarely that the system behaves like a superposition of two birth-death/monomolecular processes: one with transcription rate α_1, and the other with transcription rate α_2. In this parameter regime, the steady state solutions reduce to the Poisson mixture

$$P_{mix}(\mathbf{x}) = \frac{k_{12}}{s}\text{Poiss}(\mathbf{x}, \boldsymbol{\mu}_1) + \frac{k_{21}}{s}\text{Poiss}(\mathbf{x}, \boldsymbol{\mu}_2) \tag{6.23}$$

where the k_{12}/s and k_{21}/s factors represent the typical fractions of time the system spends in each gene state.

Finally, the bursty limit is defined by the conditions:

$$\begin{aligned} \alpha_1 &\gg 1 \\ \alpha_2 &= 0 \\ k_{21} &\gg 1 \quad k_{21} \gg k_{12} \end{aligned} \tag{6.24}$$

where taking α_1 and k_{21} large is done while keeping α_1/k_{21} held fixed. This yields a model with burst size $b := \alpha_1/k_{21}$ and burst frequency k_{12}, which has a heavy-tailed steady state distribution that is approximately negative binomial [28, 46]. For example, the birth-death-switching distribution approximately becomes

$$P_{nb}(x) = \binom{x + r - 1}{x}(1 - p)^r p^x \tag{6.25}$$

where

$$\begin{aligned} r &:= \frac{k_{12}}{\gamma} \\ p &:= \frac{b}{1 + b} \ . \end{aligned} \tag{6.26}$$

We emphasize again that it will be helpful throughout the paper to think of the somewhat complicated general analytic solutions as interpolating between these three extremes. Indeed,

as Fig. 6.3 suggests, this can be viewed somewhat literally. If we define the ad-hoc measures

$$'\text{burstiness}' := \frac{\Delta \alpha}{\alpha_1 + \alpha_2} \cdot \frac{k_{12} k_{21}}{(k_{12} + k_{21})^2}$$
$$'\text{switching vs degradation}' := \frac{s}{s + \gamma} \tag{6.27}$$

and compare the exact birth-death-switching solutions for various randomly sampled parameter sets to each of these three limit distributions, we can see that the parameter space can be relatively cleanly divided up into three regions.
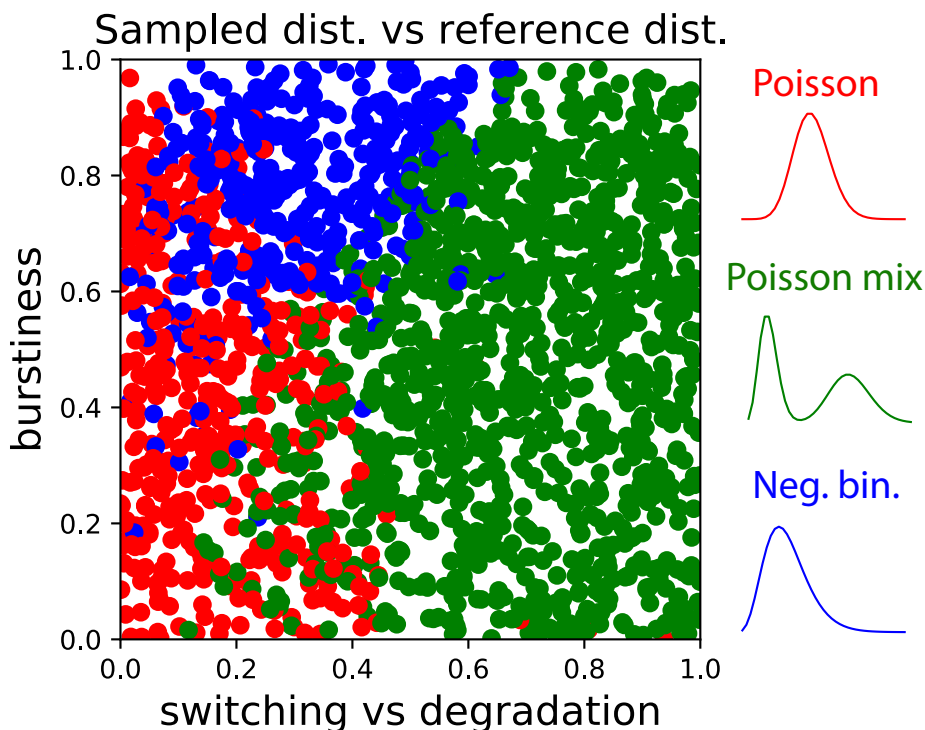
Figure 6.3: The birth-death-switching steady state distribution for many randomly sampled ($N = 2000$) parameter sets compared to three reference distributions (Poisson, Poisson mixture, negative binomial). After sampling a parameter set and computing the true $P_{ss}(x)$, the L2 distance between it and each reference distribution was computed. Each parameter set was plotted in 'switching vs degradation'-'burstiness' space and colored according to the closest reference distribution. Red: Poisson, Green: Poisson mixture, Blue: negative binomial.

6.2.4 Outline of approach to analytic solution

The theoretical methodology used to obtain these results is new, and uses ideas inspired by quantum mechanics. Instead of directly trying to solve each CME (Eq. 6.2 and Eq. 6.11),

1. Find ladder operators (1) birth-death $\quad \hat{A}\ ,\ \hat{A}^+$

(2) switching gene $\quad \hat{B}\ ,\ \hat{B}^+$

$$\hat{H} = \hat{H}_s + \hat{H}_{bd} + \hat{H}_{int}$$

2. Represent combined problem in terms of them

3. Solve via diagrams

$$P_{ss} = \mathrm{Poisson} \times (1 + \wedge + \diagup\diagdown + \cdots)$$

Figure 6.4: The three main steps of our theoretical approach. First, we identify ladder operators for the RNA-only problem and the gene-switching-only problem. Then we use those ladder operators to formulate the CME of the full problem as a specific coupling of the individual problems. Finally, we exploit properties of the ladder operators to obtain a series solution, which can be viewed as a sum of Feynman diagrams.

we reframe the problem in terms of states and operators in a certain Hilbert space. In this more abstract space, we must compute the generating function (instead of the probability distribution), whose dynamics are determined by an operator that 'looks' like a Hamiltonian. In this way, solving the CME is reduced to solving an abstract Schrödinger-like equation.

But this Schrödinger-like equation is, naively, just as difficult to solve as the original CME. The key insight is to split the problem into two sepa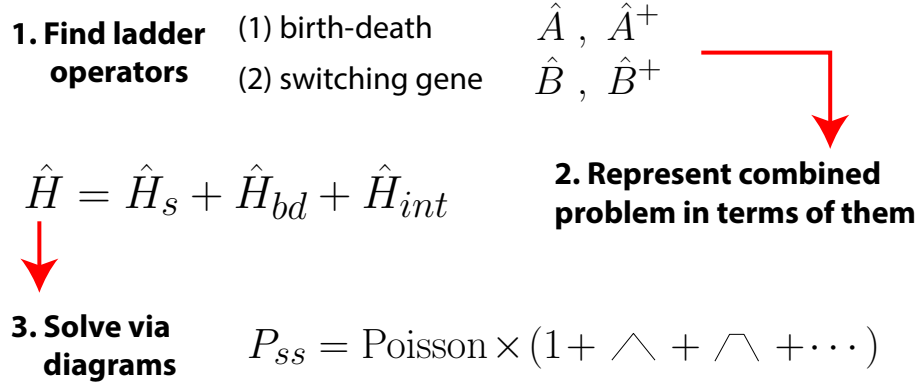rate pieces (the dynamics of RNA by itself, and the dynamics of a switching gene by itself), solve each individual piece in a very particular way, and then pose the original problem as a certain coupling of the individual problems.

The aforementioned 'particular way' is to solve the individual problems in terms of ladder operators, which are mathematically analogous to those used in the algebraic treatment of the quantum harmonic oscillator [47] and in quantum field theory [48]. The dynamics of RNA by itself can be solved in terms of 'bosonic' ladder operators, while the dynamics of a switching gene by itself can be solved in terms of 'fermionic' ladder operators. Posing the full coupled problem in terms of these operators makes it amenable to the usual methods of quantum field theory, and enables a diagrammatic approach to computing quantities of interest. It also essentially reduces many computations (e.g. of moments) to straightforward algebra.

To summarize, our general strategy involves three steps (Fig. 6.4):

1. Find ladder operators for RNA dynamics (assuming a constitutively active gene). Find ladder operators for pure gene switching.

2. Express the Hamiltonian of the coupled problem in terms of these ladder operators.

3. Solve the coupled problem perturbatively/diagrammatically to all orders.

151

In the following sections, we carry out this plan. We first focus on the birth-death-switching problem because it is less encumbered by notational baggage, and so better illustrates our method. After establishing the method by solving this problem, we go on to solve the multistep splicing problem to demonstrate the power of this new approach.

6.3   Preliminary representation theory

In this section, we study the representation theory of RNA production and degradation by itself, and gene switching by itself. We will find that the former dynamics can be described by bosonic ladder operators, while the latter dynamics can be described by fermionic ladder operators. In the last subsection, we show how to use these ladder operators to couple the two individual problems and obtain the Hamiltonian operator corresponding to Eq. 6.2.

6.3.1   Representation theory of chemical birth-death process

In this section, we study the chemical birth-death process—a simple stochastic model of RNA being produced and degraded, without any additional features like gene switching or regulation—and in particular how to work with it in terms of ladder operators that make various theoretical features (e.g. the spectrum, the relationship between eigenstates, and computing expectation values) transparent. These ladder operators and the morass of formalism surrounding them will then be used to couple the birth-death process to a switching gene in Sec. 6.4.

Let X denote the RNA species as before. The chemical birth-death process is defined by the list of reactions

$$\varnothing \xrightarrow{\alpha} X$$
$$X \xrightarrow{\gamma} \varnothing$$

(6.28)

where α parameterizes the rate of RNA production and γ parameterizes the rate of degradation. For simplicity, we assume both α and γ are time-independent. The corresponding CME is

$$\frac{\partial P(x,t)}{\partial t} = \alpha \left[P(x-1,t) - P(x,t) \right] + \gamma \left[(x+1)P(x+1,t) - xP(x,t) \right]$$

(6.29)

where $P(x,t)$ is the probability that the system has $x \in \{0, 1, 2, ...\}$ RNA molecules at time t.

In order to think about this problem in terms of ladder operators, we must first reframe it in terms of states within a certain Hilbert space, and operators that act on those states. Consider a Hilbert space spanned by the basis kets $|0\rangle, |1\rangle, ...$, where an arbitrary state $|\phi\rangle$

can be written

$$|\phi\rangle := \sum_{x=0}^{\infty} c(x) \ |x\rangle \tag{6.30}$$

for some generally complex coefficients $c(x)$. The operators that act on these states will generally be written in terms of the creation and annihilation operators

$$\begin{aligned}
\hat{\pi} \,|x\rangle &:= |x + 1\rangle \\
\hat{a} \,|x\rangle &:= x \ |x - 1\rangle
\end{aligned} \tag{6.31}$$

which satisfy the familiar commutation relation $[\hat{a}, \hat{\pi}] = 1$. In the reframed problem, we work with the generating function

$$|\psi\rangle := \sum_{x=0}^{\infty} P(x, t) \ |x\rangle \tag{6.32}$$

instead of the probability density $P(x, t)$; by construction, it contains equivalent information, so that knowing one is equivalent to knowing the other. Just as $P(x, t)$ satisfies the CME given by Eq. 6.29, it is easy to show that the generating function $|\psi\rangle$ satisfies the equation of motion

$$\frac{\partial \, |\psi\rangle}{\partial t} = \hat{H}_{bd} \, |\psi\rangle \tag{6.33}$$

where the Hamiltonian operator $\hat{H}_{bd}$ is defined as

$$\hat{H}_{bd} := \alpha \left(\hat{\pi} - 1 \right) + \gamma \left(\hat{a} - \hat{\pi}\hat{a} \right) \ . \tag{6.34}$$

For technical reasons, it will be slightly more convenient to work in terms of the Grassberger-Scheunert creation operator $\hat{a}^+ := \hat{\pi} - 1$ [49, 12], in terms of which the Hamiltonian operator reads

$$\hat{H}_{bd} := \alpha \ \hat{a}^+ - \gamma \ \hat{a}^+\hat{a} \ . \tag{6.35}$$

Now we will proceed with the ladder operator solution of the problem. We are seeking two operators $\hat{A}$ and $\hat{A}^+$ for which the following hold:

1. **Hamiltonian decomposition:** $\hat{H}_{bd}$ is proportional to $\hat{A}^+\hat{A}$.

2. **Commutation relation:** The pair satisfies $[\hat{A}, \hat{A}^+] = 1$.

3. **Hermitian conjugates:** There is an inner product with respect to which $\hat{A}$ and $\hat{A}^+$ are Hermitian conjugates.[1]

---

[1]This last requirement is optional, and is mainly necessary if one wishes to use ladder operators to derive transient (as opposed to steady state) solutions.

Although such a pair of operators may not exist for a generic choice of $\hat{H}$, their existence is easy to establish for this problem. Define $\mu := \alpha/\gamma$ (which turns out to be the mean molecule number at steady state), and consider the operators

$$\hat{A}^+ := \sqrt{\mu}\,\hat{a}^+$$
$$\hat{A} := \frac{1}{\sqrt{\mu}}\,(\hat{a} - \mu)\ . \tag{6.36}$$

Their fundamental commutation relation $[\hat{A}, \hat{A}^+] = 1$ is easy to verify. In terms of these two operators, the Hamiltonian can be written

$$\hat{H}_{bd} = -\gamma\ \hat{A}^+\hat{A}\ . \tag{6.37}$$

This decomposition makes it easy to show that the commutation relations

$$[\hat{A}, \hat{H}_{bd}] = -\gamma\hat{A}$$
$$[\hat{A}^+, \hat{H}_{bd}] = \gamma\hat{A}^+ \tag{6.38}$$

hold. They can be used to argue [47, 50] that there is a ground state $|0\rangle$ satisfying

$$\hat{A}\,|0\rangle = 0$$
$$|0\rangle = \sum_{x=0}^{\infty} P_{ss}(x)\,|x\rangle = \sum_{x=0}^{\infty} \frac{\mu^x}{x!}e^{-\mu}\,|x\rangle\ , \tag{6.39}$$

where the normalization of $|0\rangle$ was chosen so that its coefficients sum to 1 (since they correspond to the values of the steady state probability distribution $P_{ss}(x)$), and that the eigenvalues of the Hamiltonian operator are $-E_n$, where

$$E_n = \gamma\ n \tag{6.40}$$

for $n = 0, 1, 2, ...$ Hence, we can write the eigenstates of $\hat{H}_{bd}$ as $|n\rangle$ for $n \in \mathbb{N}$ (e.g. $|0\rangle$ is the ground state, and $|1\rangle$ if the first excited state), so that

$$\hat{H}_{bd}\,|n\rangle = -\gamma\ n\,|n\rangle\ . \tag{6.41}$$

Moreover, using the facts that[2] $[\hat{A}, (\hat{A}^+)^n] = n(\hat{A}^+)^{n-1}$ and $\hat{A}\,|0\rangle = 0$, we have

$$\begin{aligned}
\hat{H}_{bd}(\hat{A}^+)^n\,|0\rangle &= -\gamma\hat{A}^+\hat{A}(\hat{A}^+)^n\,|0\rangle \\
&= -\gamma\hat{A}^+\left[(\hat{A}^+)^n\hat{A} + n(\hat{A}^+)^{n-1}\right]|0\rangle \\
&= -\gamma\ n(\hat{A}^+)^n\,|0\rangle
\end{aligned} \tag{6.42}$$

i.e. that

$$|n\rangle \propto (\hat{A}^+)^n\,|0\rangle\ . \tag{6.43}$$

---

[2]This is easy to show by induction, using the fundamental commutation relation between $\hat{A}$ and $\hat{A}^+$.

In order to fix the normalization of the eigenstates $|n\rangle$, we must decide on an inner product. For this problem, a natural choice turns out to be

$$\langle x|y\rangle := \frac{\delta_{xy}}{P_{ss}(x)} = \frac{\delta_{xy} \; x! \; e^{\mu}}{\mu^x} \; . \tag{6.44}$$

It is straightforward to show that $\hat{A}$ and $\hat{A}^+$ are Hermitian conjugates with respect to this inner product. On the one hand,

$$\begin{aligned}
\langle x|\hat{A}^+|y\rangle &= \sqrt{\mu} \; \langle x|\hat{a}^+ - 1|y\rangle \\
&= \sqrt{\mu} \; [\; \langle x|y+1\rangle - \langle x|y\rangle \;] \\
&= \sqrt{\mu} \; \left[\frac{\delta_{x,y+1} \; x! \; e^{\mu}}{\mu^x} - \frac{\delta_{xy} \; x! \; e^{\mu}}{\mu^x} \right] \; .
\end{aligned} \tag{6.45}$$

On the other hand,

$$\begin{aligned}
\langle y|\hat{A}|x\rangle &= \frac{1}{\sqrt{\mu}} \; \langle y|\hat{a} - \mu|x\rangle \\
&= \frac{1}{\sqrt{\mu}} \; [\; x \langle y|x-1\rangle - \mu \langle y|x\rangle \;] \\
&= \frac{1}{\sqrt{\mu}} \; \left[x \frac{\delta_{x-1,y} \; (x-1)! \; e^{\mu}}{\mu^{x-1}} - \mu \frac{\delta_{xy} \; x! \; e^{\mu}}{\mu^x} \right] \\
&= \sqrt{\mu} \; \left[\frac{\delta_{x,y+1} \; x! \; e^{\mu}}{\mu^x} - \frac{\delta_{xy} \; x! \; e^{\mu}}{\mu^x} \right] \\
&= \langle x|\hat{A}^+|y\rangle \; .
\end{aligned} \tag{6.46}$$

Because $\hat{A}$ and $\hat{A}^+$ are Hermitian conjugates, $\hat{H}_{bd}$ is Hermitian, since

$$\left(\hat{H}_{bd}\right)^{\dagger} = -\gamma \left(\hat{A}^+\hat{A}\right)^{\dagger} = -\gamma \left(\hat{A}\right)^{\dagger} \left(\hat{A}^+\right)^{\dagger} = -\gamma \hat{A}^+\hat{A} = \hat{H}_{bd} \; . \tag{6.47}$$

This allows us to use a standard argument to show that different eigenstates are orthogonal:

$$\langle m|n\rangle = -\frac{1}{E_n} \langle m|\hat{H}_{bd}|n\rangle = -\frac{1}{E_n} \langle n|\hat{H}_{bd}|m\rangle^* = \frac{E_m}{E_n} \langle n|m\rangle^* = \frac{E_m}{E_n} \langle m|n\rangle \tag{6.48}$$

which forces $\langle m|n\rangle = 0$ since $E_m \neq E_n$. In order to normalize these eigenstates, so that $\langle n|n\rangle = 1$ for all $n \in \mathbb{N}$, we note that

$$\langle n|n\rangle = \langle 0|\hat{A}^n(\hat{A}^+)^n|0\rangle = n! \langle 0|0\rangle = n! \; . \tag{6.49}$$

This implies the correct normalization for the eigenkets is

$$|n\rangle := \frac{(\hat{A}^+)^n}{\sqrt{n!}} \; |0\rangle \; . \tag{6.50}$$

155

Using this definition, we can show that

$$\hat{A}^+ \ket{n} = \sqrt{n+1} \ket{n+1}$$
$$\hat{A} \ket{n} = \sqrt{n} \ket{n-1} \ ,$$

(6.51)

properties that will be useful later.

All of this legwork we put into understanding the eigenstates of $\hat{H}_{bd}$ makes finally solving the equation of motion, Eq. 6.33, relatively simple. Because the eigenstates as we have defined them are orthonormal, we have the resolution of the identity

$$1 = \sum_{n=0}^{\infty} \ket{n} \bra{n} \ .$$

(6.52)

The formal solution of Eq. 6.33 is

$$\ket{\psi(t)} = e^{\hat{H}_{bd}(t-t_0)} \ket{\psi_0}$$

(6.53)

where $\ket{\psi_0} = \ket{\psi(t_0)}$ is the initial generating function. Applying Eq. 6.52 to our formal solution, we have

$$\begin{aligned}
\ket{\psi(t)} &= \sum_{n=0}^{\infty} e^{\hat{H}_{bd}(t-t_0)} \ket{n} \braket{n|\psi_0} \\
&= \sum_{n=0}^{\infty} e^{-E_n(t-t_0)} \ket{n} \braket{n|\psi_0} \\
&= \sum_{n=0}^{\infty} \braket{n|\psi_0} \ket{n} \ e^{-\gamma n(t-t_0)} \ .
\end{aligned}$$

(6.54)

We are particularly interested in computing the transition probability $P(x,t;x_0,t_0)$ for arbitrary $x_0 \in \mathbb{N}$, in which case $\ket{\psi_0} = \ket{x_0}$. Denoting the coefficients of $\ket{n}$ by $P_n(x)$, note that

$$\braket{n|x_0} = \frac{P_n(x_0)}{P_{ss}(x_0)}$$

(6.55)

by definition. It can be shown, using the definition of $\ket{n}$ from Eq. 6.50 and the fact that the coefficients of $\ket{0}$ are known (c.f. Eq. 6.39), that

$$P_n(x) = \sqrt{\frac{\mu^n}{n!}} \ C_n(x,\mu) P_{ss}(x)$$

(6.56)

where $C_n(x,\mu)$ are the Charlier polynomials, a family of discrete orthogonal polynomials whose properties are discussed at length in Appendix 6.A. They play a role analogous to the one the Hermite polynomials play in the solution of the quantum harmonic oscillator. Hence, the solution to Eq. 6.33 with $\ket{\psi_0} = \ket{x_0}$ is

$$\ket{\psi(t)} = \sum_{n=0}^{\infty} \sqrt{\frac{\mu^n}{n!}} \ C_n(x_0,\mu) \ket{n} \ e^{-\gamma n(t-t_0)} \ .$$

(6.57)

The corresponding probability distribution $P(x, t; x_0, t_0)$ can be obtained in two ways: (i) by appealing to the definition of $|\psi(t)\rangle$, or (ii) by invoking the Euclidean product. The Euclidean product is another useful inner product, and defined on two basis kets as

$$\langle x|y\rangle_{Eu} = \delta_{x,y} \tag{6.58}$$

and extended to more general kets by linearity. The expression $\langle x|\psi(t)\rangle_{Eu}$ essentially 'picks out' the coefficient of $|\psi(t)\rangle$ corresponding to $P(x, t)$; although this is identical to just picking out the x coefficient of the generating function, using this notation makes the mathematical analogy between position-space wave functions and probability distributions more transparent. Finally,

$$
\begin{aligned}
P(x, t; x_0, t_0) &= \langle x|\psi(t)\rangle_{Eu} \\
&= \sum_{n=0}^{\infty} \sqrt{\frac{\mu^n}{n!}} \; C_n(x_0, \mu) \; \langle x|n\rangle_{Eu} \; e^{-\gamma n(t-t_0)} \\
&= P_{ss}(x) \sum_{n=0}^{\infty} \frac{\mu^n}{n!} \; C_n(x_0, \mu) C_n(x, \mu) \; e^{-\gamma n(t-t_0)} \; .
\end{aligned}
\tag{6.59}
$$

The more familiar expression for $P(x, t; x_0, t_0)$, given by [11, 12]

$$P(x, t; x_0, t_0) = \sum_{j=0}^{\min(x_0, x)} \binom{x_0}{j} q^{x_0-j}(1-q)^j \; \cdot \; \frac{(\mu q)^{x-j} e^{-\mu q}}{(x-j)!} \tag{6.60}$$

where $q(t) := 1 - e^{-\gamma(t-t_0)}$, can be obtained from Eq. 6.59 using Charlier polynomial properties discussed in Appendix 6.A; for example, one can either use the integral representation given by Eq. 6.211, or the Charlier polynomial generating function given by Eq. 6.209.

Particular features of the solution, like moments, can be obtained by expressing certain operators in terms of ladder operators and then using their algebraic properties to compute the result. For example, suppose we want to compute the mean $\langle x \rangle$ as a function of time. First, note that

$$\hat{a} |\psi(t)\rangle = \sum_{x=0}^{\infty} P(x, t) \; x \; |x - 1\rangle \; , \tag{6.61}$$

so

$$\langle 0|\hat{a}|\psi(t)\rangle = \sum_{x=0}^{\infty} P(x, t) x = \langle x \rangle \tag{6.62}$$

where we have used the fact that the inner product of the vacuum state with any basis ket

$|x\rangle$ is $\langle 0|x\rangle = 1$. Now, since $\hat{a} = \sqrt{\mu}\,\hat{A} + \mu$, we can compute

$$
\begin{aligned}
\langle x \rangle &= \langle 0|(\sqrt{\mu}\,\hat{A} + \mu)|\psi(t)\rangle \\
&= \sum_{n=0}^{\infty} \sqrt{\frac{\mu^n}{n!}}\, C_n(x_0, \mu)\ \langle 0|(\sqrt{\mu}\,\hat{A} + \mu)|n\rangle\ e^{-\gamma n(t-t_0)} \\
&= \sum_{n=0}^{\infty} \sqrt{\frac{\mu^n}{n!}}\, C_n(x_0, \mu)\ \left[\sqrt{\mu}\,\langle 0|\hat{A}|n\rangle + \mu\,\langle 0|n\rangle\right]\ e^{-\gamma n(t-t_0)} \\
&= \sum_{n=0}^{\infty} \sqrt{\frac{\mu^n}{n!}}\, C_n(x_0, \mu)\ \left[\sqrt{\mu n}\,\langle 0|n-1\rangle + \mu\,\langle 0|n\rangle\right]\ e^{-\gamma n(t-t_0)} \\
&= \sum_{n=0}^{\infty} \sqrt{\frac{\mu^n}{n!}}\, C_n(x_0, \mu)\ \left[\sqrt{\mu n}\,\delta_{n,1} + \mu\,\delta_{n,0}\right]\ e^{-\gamma n(t-t_0)} \\
&= (x_0 - \mu)\,e^{-\gamma(t-t_0)} + \mu
\end{aligned}
\tag{6.63}
$$

where we have used the fact that $C_1(x_0, \mu) = (x_0/\mu) - 1$.

Parenthetically, we may note that the use of these ladder operators to understand the birth-death process evokes striking parallels with the quantum harmonic oscillator; the analogy is even more striking in the continuous limit, where Hermite polynomials replace the Charlier polynomials (see Sec. 6.6 and [50]). In some sense, a single chemical species being randomly produced and degraded behaves like a free boson. More generally, we will see in Sec. 6.7 that N chemical species being produced and degraded behave like N free bosons.

6.3.2 Representation theory of pure gene switching

In this section, we will consider the stochastic dynamics of a switching gene by itself. A careful study of the switching gene uncovers ladder operators that will be useful when coupling switching dynamics to the birth-death process in the next section.

Consider a gene with two states, which we will label as G_1 and G_2. Our list of chemical reactions is

$$
\begin{aligned}
G_1 &\xrightarrow{k_{21}} G_2 \\
G_2 &\xrightarrow{k_{12}} G_1
\end{aligned}
\tag{6.64}
$$

and the corresponding CME reads

$$
\begin{aligned}
\frac{\partial P(1,t)}{\partial t} &= -k_{21}P(1,t) + k_{12}P(2,t) \\
\frac{\partial P(2,t)}{\partial t} &= k_{21}P(1,t) - k_{12}P(2,t)
\end{aligned}
\tag{6.65}
$$

where $P(1,t)$ is the probability that the gene is in state 1 at time t, and $P(2,t)$ is the probability that the gene is in state 2 at time t. If we define the state vector

$$
\vec{P}(t) := \begin{pmatrix} P(1,t) \\ P(2,t) \end{pmatrix},
\tag{6.66}
$$

then we can write the CME in the form

$$\dot{\vec{P}} = \hat{H}_s \vec{P} \ , \tag{6.67}$$

where the switching Hamiltonian $\hat{H}_s$ is defined to be

$$\hat{H}_s := \begin{pmatrix} -k_{21} & k_{12} \\ k_{21} & -k_{12} \end{pmatrix} \ . \tag{6.68}$$

One important property of the matrix $\hat{H}_s$ is that it is *infinitesimal stochastic* (see [51] for more discussion of this property), which means that $1^T \hat{H}_s = 0^T$, or equivalently that the columns of $\hat{H}_s$ sum to zero. This property guarantees probability conservation:

$$\frac{d}{dt}\left(1^T \vec{P}\right) = 1^T \frac{d\vec{P}}{dt} = 1^T \hat{H}_s \vec{P} = 0^T \vec{P} = 0 \ . \tag{6.69}$$

It is easy enough to find the exact time-dependent solution to this system by diagonalizing $\hat{H}_s$. Its eigenvalues are $\lambda_0 = 0$ and $\lambda_1 = -s$, where $s := k_{12} + k_{21}$ is the gene switching rate sum. The corresponding eigenvectors are

$$\vec{v}_0 = \begin{pmatrix} \frac{k_{12}}{k_{12}+k_{21}} \\ \frac{k_{21}}{k_{12}+k_{21}} \end{pmatrix} \quad \vec{v}_1 = \begin{pmatrix} 1 \\ -1 \end{pmatrix} \ . \tag{6.70}$$

Hence, the exact solution is

$$\vec{P}(t) = \vec{v}_0 + c \ \vec{v}_1 \ e^{-st} \tag{6.71}$$

where the constant c depends upon one's initial condition.

A more interesting task is to try and express $\hat{H}_s$ in terms of ladder operators that will make its solution structure more explicit. Recall that the anticommutator of two matrices M_1 and M_2 is defined as $\{M_1, M_2\} := M_1 M_2 + M_2 M_1$. We are seeking two operators $\hat{B}$ and $\hat{B}^+$ for which the following hold:

1. **Hamiltonian decomposition:** $\hat{H}_s$ is proportional to $\hat{B}^+ \hat{B}$.

2. **Commutation relations:** The pair satisfies $\{\hat{B}, \hat{B}\} = \{\hat{B}^+, \hat{B}^+\} = 0$ and $\{\hat{B}, \hat{B}^+\} = 1$. Here, 1 is shorthand for the 2×2 identity matrix.

3. **Hermitian conjugates:** There is an inner product with respect to which $\hat{B}$ and $\hat{B}^+$ are Hermitian conjugates.

By trial and error[3], we can find that the operators

$$\hat{B} := \begin{pmatrix} \frac{k_{12}k_{21}}{s^2} & -\frac{k_{12}^2}{s^2} \\ \frac{k_{21}^2}{s^2} & -\frac{k_{12}k_{21}}{s^2} \end{pmatrix}$$

$$\hat{B}^+ := \begin{pmatrix} 1 & 1 \\ -1 & -1 \end{pmatrix} \tag{6.72}$$

---

[3]For a more principled route to determining them, see Appendix 6.B.

suffice. In terms of them, the switching Hamiltonian $\hat{H}_s$ can be written

$$\hat{H}_s = -s \ \hat{B}^+\hat{B} \ . \tag{6.73}$$

One nice feature of $\hat{B}$ and $\hat{B}^+$ is that their action on the eigenvectors of $\hat{H}_s$ is particularly simple:

$$B\vec{v}_0 = 0 \qquad\qquad B\vec{v}_1 = \vec{v}_0 \tag{6.74}$$

$$B^+\vec{v}_0 = \vec{v}_1 \qquad\qquad B^+\vec{v}_1 = 0 \ . \tag{6.75}$$

These are properties that we will repeatedly exploit later. One can define a natural inner product on $\mathbb{R}^2$ using the symmetric and positive-definite weight matrix

$$\hat{W} = \begin{pmatrix} 1 + \frac{k_{21}^2}{s^2} & 1 - \frac{k_{12}k_{21}}{s^2} \\ 1 - \frac{k_{12}k_{21}}{s^2} & 1 + \frac{k_{12}^2}{s^2} \end{pmatrix} \ . \tag{6.76}$$

Note that, since k_{12} and k_{21} are nonnegative, $0 \leq (k_{12}k_{21})/s^2 \leq 1/4$, which makes the entries of $\hat{W}$ nonnegative. For more discussion on the appropriateness of this weight matrix, along with some mathematical motivation for considering these kinds of ladder operators, see Appendix 6.B. For now, assuming our choice of $\hat{W}$ is well-motivated, we define the inner product of two vectors $\vec{z}_1, \vec{z}_2 \in \mathbb{R}^2$ via

$$\langle \vec{z}_1, \vec{z}_2 \rangle := \vec{z}_1^T W \vec{z}_2 \ . \tag{6.77}$$

With respect to this inner product, $\hat{B}$ and $\hat{B}^+$ are Hermitian conjugates (see Appendix 6.B for the mathematical details), i.e. we have $\hat{W}\hat{B} = (\hat{B}^+)^T\hat{W}$ and $\hat{W}\hat{B}^+ = (\hat{B})^T\hat{W}$.

Recalling that the eigenvectors of $\hat{H}_s$ and $\hat{H}_s^T$ are orthogonal with respect to the usual dot product offers another way to view this inner product. That is, given the eigenvectors $\vec{w}_0$ and $\vec{w}_1$ (which correspond to the eigenvalues $\lambda_0 = 0$ and $\lambda_1 = -s$ since $\hat{H}_s$ and $\hat{H}_s^T$ have the same eigenvalues), we have $\vec{v}_i^T W = \vec{w}_i^T$. For completeness' sake, we note that

$$\vec{w}_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \vec{w}_1 = \begin{pmatrix} \frac{k_{21}}{k_{12}+k_{21}} \\ -\frac{k_{12}}{k_{12}+k_{21}} \end{pmatrix} \ . \tag{6.78}$$

Because applying $\vec{w}_0^T$ on the left corresponds to summing over each gene state, we will end up using it to marginalize over gene state in Sec. 6.5.

It will be helpful to recast this inner product in terms of bra-ket notation; to that end, we identify the eigenvectors $\vec{v}_0$ and $\vec{v}_1$ with $|0\rangle$ and $|1\rangle$, $\vec{w}_0$ and $\vec{w}_1$ with $\langle 0|$ and $\langle 1|$, and note that

$$\begin{aligned} \langle 0|1\rangle &= 0 \\ \langle 0|0\rangle &= \langle 1|1\rangle = 1 \end{aligned} \tag{6.79}$$

i.e. they constitute an orthonormal basis.

It is interesting to note that the switching gene behaves much like a single spin-1/2 fermion. Given that many RNA species being produced and degraded behave like many free bosons (see Sec. 6.7), it is tempting to speculate that the dynamics of switching between many gene states corresponds to the dynamics of many free fermions. As we discuss at the end of Appendix 6.B, this generically does not seem to be true, at least in a straightforward generalization of our approach.

6.3.3 Coupling a switching gene to the birth-death process

We would like to couple a switching gene to the chemical birth-death process; in particular, we would like to make the production rate dependent on the current gene state. This means that, instead of the production rate being a constant parameter α, we can view it as an operator

$$\hat{\alpha} := \begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} \tag{6.80}$$

where α_1 and α_2 are the distinct production rates we first introduced in Eq. 7.1. We stress that we *do not* assume anything about α_1 or α_2 (e.g. that they are distinct, or that one is significantly larger than the other), even though these assumptions apply in the biologically interesting case of a bursty gene.

This section's central trick is to note that the matrix $\hat{\alpha}$ can be written in terms of ladder operators, since the set $\{1, \hat{B}, \hat{B}^+, \hat{B}^+\hat{B}\}$ forms a basis for the space of all real 2×2 matrices. We will use this fact to write

$$\begin{pmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{pmatrix} = r_1 + r_2\hat{B} + r_3\hat{B}^+ + r_4\hat{B}^+\hat{B} \tag{6.81}$$

for some coefficients r_1, r_2, r_3, and r_4. To determine these coefficients, we recall from the previous subsection the facts that (i) the actions of $\hat{B}$ and $\hat{B}^+$ on the eigenvectors of $\hat{H}_s$ are easy to compute (c.f. Eq. 6.74 and Eq. 6.75), and (ii) the eigenvectors of $\hat{H}_s$ are orthonormal with respect to the inner product we defined (c.f. Eq. 6.79). For example,

$$\begin{aligned} \hat{\alpha}\,|0\rangle &= r_1\,|0\rangle + r_2\hat{B}\,|0\rangle + r_3\hat{B}^+\,|0\rangle + r_4\hat{B}^+\hat{B}\,|0\rangle \\ &= r_1\,|0\rangle + r_3\,|1\rangle \\ \implies r_1 &= \langle 0|\hat{\alpha}|0\rangle \ . \end{aligned} \tag{6.82}$$

Similarly,

$$\begin{aligned} r_2 &= \langle 0|\hat{\alpha}|1\rangle \\ r_3 &= \langle 1|\hat{\alpha}|0\rangle \\ r_4 &= \langle 1|\hat{\alpha}|1\rangle - r_1 \ . \end{aligned} \tag{6.83}$$

It should be noted that, although we are interested here in the case where the production rate depends on the current gene state in the sense specified above, this method can in principle be used to tackle many possible generalizations of this problem.

Our particular coefficients read

$$\begin{aligned} r_1 &= \frac{\alpha_1 k_{12} + \alpha_2 k_{21}}{s} \\ r_2 &= \alpha_1 - \alpha_2 \\ r_3 &= (\alpha_1 - \alpha_2)\frac{k_{12}k_{21}}{s^2} \\ r_4 &= (\alpha_1 - \alpha_2)\frac{k_{21} - k_{12}}{s} \ . \end{aligned} \tag{6.84}$$

If we define the effective production rate

$$\alpha_{eff} := \frac{\alpha_1 k_{12} + \alpha_2 k_{21}}{s} \ , \tag{6.85}$$

then we can write the matrix $\hat{\alpha}$ in terms of ladder operators as

$$\hat{\alpha} = \alpha_{eff} + (\alpha_1 - \alpha_2) \left[B + \frac{k_{12}k_{21}}{s^2} B^+ + \frac{k_{21} - k_{12}}{s} B^+ B \right] \ . \tag{6.86}$$

With this done, we can write out the Hamiltonian operator for our full, coupled problem[4]. It reads

$$
\begin{aligned}
\hat{H} &= -s\hat{B}^+\hat{B} + \hat{\alpha}\ \hat{a}^+ - \gamma\ \hat{a}^+\hat{a} \\
&= -s\hat{B}^+\hat{B} - \gamma\hat{A}^+\hat{A} + (\alpha_1 - \alpha_2)\ \hat{a}^+ \left[\hat{B} + \frac{k_{12}k_{21}}{s^2}\hat{B}^+ + \frac{k_{21} - k_{12}}{s}\hat{B}^+\hat{B} \right] \\
&= -s\hat{B}^+\hat{B} - \gamma\hat{A}^+\hat{A} + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\hat{A}^+ \left[\hat{B} + \frac{k_{12}k_{21}}{s^2}\hat{B}^+ + \frac{k_{21} - k_{12}}{s}\hat{B}^+\hat{B} \right] \\
&= \hat{H}_s + \hat{H}_{bd} + \hat{H}_{int}
\end{aligned}
\tag{6.87}
$$

where we define the interaction Hamiltonian $\hat{H}_{int}$ as

$$\hat{H}_{int} := \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\hat{A}^+ \left[\hat{B} + \frac{k_{12}k_{21}}{s^2}\hat{B}^+ + \frac{k_{21} - k_{12}}{s}\hat{B}^+\hat{B} \right] \tag{6.88}$$

along with the shorthand notations

$$
\begin{aligned}
\mu &:= \frac{\alpha_{eff}}{\gamma} \\
\hat{A}^+ &= \sqrt{\mu}\left(a^+ - 1\right) \\
\hat{A} &= \frac{1}{\sqrt{\mu}}\left(a - \mu\right) \ .
\end{aligned}
\tag{6.89}
$$

The parameter μ that appears in this Hamiltonian is actually the effective mean—a sort of average of what the mean would be given the typical amount of time the system spends in each gene state. It should not be conflated with α_1/γ or α_2/γ in what follows.

Interestingly, $\hat{H}_{int}$ looks like an interaction term with interaction strength proportional to $(\alpha_1 - \alpha_2)$, the difference between the two production rates. In particular, since the $\hat{A}$ operators are 'bosonic', and the $\hat{B}$ operators are 'fermionic', our full Hamiltonian looks like it describes one boson interacting with one fermion. When the production rates are equal (i.e. when $\alpha_1 = \alpha_2$), the interaction term disappears, and the birth-death process and the gene switching dynamics become uncoupled. This makes sense—from the point of view of the birth-death process, it doesn't matter what the current gene state is as long as it doesn't affect the production rate.

---

[4]This can be derived from the full CME, Eq. 6.2.

It is natural to speculate that the Hamiltonian describing a birth-death process coupled to a gene with N gene states would correspond to one boson interacting with $N-1$ fermions (which also possibly interact with each other). Coupling additional birth-death processes (for example, representing gene-gene interactions or downstream products of the original RNA) may look like having more bosons. It is not completely clear what the representation theory of $\hat{H}_s$ would look like in these more general cases, or even if a ladder operator view would be appropriate. Regrettably, the representation theory of infinitesimal stochastic matrices seems not particularly well-studied; for example, it is not clear to the authors that infinitesimal stochastic matrices (or even restricted classes of them, like irreducible infinitesimal stochastic matrices) have a spectral theorem that would permit eigenvector decompositions similar to the one we have used here.

6.4 Diagrammatic approach to exact solution

In this section, we use the ladder operators identified in the previous section to develop a diagrammatic approach to obtaining the exact steady state solution of CMEs involving switching (and in particular, of Eq. 6.2).

6.4.1 Constructing eigenstates of full Hamiltonian

Consider a state formed by naively combining the eigenstates of the two original problems[5], which we will denote by $|n; g\rangle$, and which satisfies

$$
\begin{aligned}
\hat{H}_s \,|n; 0\rangle &= 0 \\
\hat{H}_s \,|n; 1\rangle &= -s \,|n; 1\rangle \\
\hat{H}_{bd} \,|n; g\rangle &= -\gamma n \,|n; g\rangle \ .
\end{aligned}
\tag{6.90}
$$

The action of our ladder operators on this state is

$$
\begin{aligned}
\hat{B} \,|n; 0\rangle &= \hat{B}^+ \,|n; 1\rangle = 0 \\
\hat{B}^+ \,|n; 0\rangle &= |n; 1\rangle \\
\hat{B} \,|n; 1\rangle &= |n; 0\rangle \\
\hat{A} \,|n; g\rangle &= \sqrt{n} \,|n-1; g\rangle \\
\hat{A}^+ \,|n; g\rangle &= \sqrt{n+1} \,|n+1; g\rangle \ .
\end{aligned}
\tag{6.91}
$$

---

[5]The birth-death problem has a countably infinite number of eigenstates, which can be indexed by a natural number n, and the switching gene has two, which we label using 0 and 1. Note that these two eigenstates are distinct from the two gene states, which are labeled using 1 and 2.

All of this means that we can compute the action of $\hat{H}$ on these states. It is

$$\hat{H}\left|n;0\right\rangle = -\gamma n\left|n;0\right\rangle + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\frac{k_{12}k_{21}}{s^2}\sqrt{n+1}\left|n+1;1\right\rangle$$

$$\hat{H}\left|n;1\right\rangle = -(s+\gamma n)\left|n;1\right\rangle + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\sqrt{n+1}\left[\left|n+1;0\right\rangle + \frac{k_{21}-k_{12}}{s}\left|n+1;1\right\rangle\right]\ . \tag{6.92}$$

From the above, it should be clear that our states $\left|n;g\right\rangle$ are 'almost' eigenstates of $\hat{H}$, if it were not for the term proportional to $(\alpha_1 - \alpha_2)$ due to the interaction $\hat{H}_{int}$ between the bosonic and fermionic ladder operators. If α_1 is very close to α_2, this term would be small, and it may be reasonable to approximately consider the states $\left|n;g\right\rangle$ as eigenstates.

But this suggests something interesting. Consider the action of $\hat{H}$ on the 'next' states, indexed by $n+1$:

$$\hat{H}\left|n+1;0\right\rangle = -\gamma(n+1)\left|n+1;0\right\rangle + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\frac{k_{12}k_{21}}{s^2}\sqrt{n+2}\left|n+2;1\right\rangle$$

$$\hat{H}\left|n+1;1\right\rangle = -[s+\gamma(n+1)]\left|n+1;1\right\rangle + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\sqrt{n+2}\left[\left|n+2;0\right\rangle + \frac{k_{21}-k_{12}}{s}\left|n+2;1\right\rangle\right]\ . \tag{6.93}$$

Note that

$$\hat{H}\left\{\left|n;0\right\rangle + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\frac{k_{12}k_{21}}{s^2(s+\gamma)}\sqrt{n+1}\left|n+1;1\right\rangle\right\}$$

$$= -\gamma n\left\{\left|n;0\right\rangle + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\frac{k_{12}k_{21}}{s^2(s+\gamma)}\sqrt{n+1}\left|n+1;1\right\rangle\right\} \tag{6.94}$$

$$+ \left[\frac{(\alpha_1 - \alpha_2)}{\sqrt{\mu}}\right]^2\frac{k_{12}k_{21}}{s^2(s+\gamma)}\sqrt{(n+1)(n+2)}\left[\left|n+2;0\right\rangle + \frac{k_{21}-k_{12}}{s}\left|n+2;1\right\rangle\right]$$

which is to even *better* approximation an eigenstate of $\hat{H}$ when α_1 is close to α_2. It seems like this process may be continued to obtain an infinite series in powers of $(\alpha_1 - \alpha_2)$—and indeed it can. To show how it can, first let us ease notation by writing

$$d := \frac{\alpha_1 - \alpha_2}{\sqrt{\mu}}$$

$$c_3 := \frac{k_{12}k_{21}}{s^2} \tag{6.95}$$

$$c_4 := \frac{k_{21}-k_{12}}{s}$$

so that the interaction Hamiltonian $\hat{H}_{int}$ can be written as

$$\hat{H}_{int} = d\ \hat{A}^+\left[\hat{B} + c_3\hat{B}^+ + c_4\hat{B}^+\hat{B}\right]\ . \tag{6.96}$$

164

If we write our result for the eigenstate $|n\rangle$ of the full coupled problem (which we label as such since it has eigenvalue $-\gamma n$) as

$$|n\rangle := |n;0\rangle + d\sqrt{n+1}\left\{\; q^0_{1,0}\,|n+1;0\rangle + q^0_{1,1}\,|n+1;1\rangle \;\right\} + \cdots$$

$$= |n;0\rangle + \sum_{k=1}^{\infty} d^k \sqrt{(n+1)_k} \left\{\; q^0_{k,0}\,|n+k;0\rangle + q^0_{k,1}\,|n+k;1\rangle \;\right\} \tag{6.97}$$

for some coefficients $q^0_{k,g}$ determined by the procedure above, with $(n+1)_k := (n+1)\cdots(n+k)$, we find that the coefficients satisfy the recurrence relations

$$k\gamma\; q^0_{k,0} = q^0_{k-1,1}$$
$$(k\gamma + s)\; q^0_{k,1} = c_3\; q^0_{k-1,0} + c_4\; q^0_{k-1,1} \tag{6.98}$$

for all $k \geq 1$, with the initial conditions $q^0_{0,0} = 1$ and $q^0_{0,1} = 0$.

Similarly, we can construct eigenstates $|n + s/\gamma\rangle$ (which we label as such since they have eigenvalues $-\gamma\,(n + s/\gamma)\,$) as an infinite series

$$|n + s/\gamma\rangle := |n;1\rangle + d\sqrt{n+1}\left\{\; q^1_{1,0}\,|n+1;0\rangle + q^1_{1,1}\,|n+1;1\rangle \;\right\} + \cdots$$

$$= |n;1\rangle + \sum_{k=1}^{\infty} d^k \sqrt{(n+1)_k} \left\{\; q^1_{k,0}\,|n+k,0\rangle + q^1_{k,1}\,|n+k;1\rangle \;\right\} \tag{6.99}$$

whose coefficients are determined by the recurrence relations

$$(k\gamma - s)\; q^1_{k,0} = q^1_{k-1,1}$$
$$k\gamma\; q^1_{k,1} = c_3\; q^1_{k-1,0} + c_4\; q^1_{k-1,1} \tag{6.100}$$

for all $k \geq 1$, with the initial conditions $q^1_{0,0} = 0$ and $q^1_{0,1} = 1$.

The above recurrences have closed form solutions, but their expressions are somewhat cumbersome; fortunately, there is an interesting diagrammatic interpretation of the coefficients $q^0_{k,g}$ and $q^1_{k,g}$. We will explore these points in the next two subsections.

For now, let us stop and note that we have established that there are eigenstates of $\hat{H}$ with eigenvalues $-\gamma\, n - s\, g$ for all $n \in \mathbb{N}$ and $g \in \{0,1\}$. Although it is not clear how to prove this mathematically, it is likely that this is the *complete* collection of eigenstates, so that the full spectrum of $\hat{H}$ is given by

$$E_{n,g} = \gamma\, n + s\, g \tag{6.101}$$

for $n \in \mathbb{N}$ and $g \in \{0,1\}$. One nice feature of this result is that it offers a natural way to understand previous observations regarding the relative importance of different time scales in this problem. For example, Iyer-Biswas et al. [39] point out (in our notation) that $\min(\gamma, s)$ determines the time scale of relaxation to steady state. This makes sense, because $-\min(\gamma, s)$ is precisely the smallest nonzero eigenvalue.

6.4.2 The transfer matrix

Closed form solutions to the recurrence relations given by Eq. 6.98 and 6.100 can be written in terms of a certain 2×2 matrix, which we will call the transfer matrix T. First, consider that Eq. 6.98 can be rewritten in the form

$$\begin{pmatrix} q_{k,0}^0 \\ q_{k,1}^0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{k\gamma} \\ \frac{c_3}{k\gamma+s} & \frac{c_4}{k\gamma+s} \end{pmatrix} \begin{pmatrix} q_{k-1,0}^0 \\ q_{k-1,1}^0 \end{pmatrix} = \begin{pmatrix} \frac{1}{k\gamma} & 0 \\ 0 & \frac{1}{k\gamma+s} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ c_3 & c_4 \end{pmatrix} \begin{pmatrix} q_{k-1,0}^0 \\ q_{k-1,1}^0 \end{pmatrix} . \tag{6.102}$$

Similarly, Eq. 6.100 can be rewritten in the form

$$\begin{pmatrix} q_{k,0}^1 \\ q_{k,1}^1 \end{pmatrix} = \begin{pmatrix} \frac{1}{k\gamma-s} & 0 \\ 0 & \frac{1}{k\gamma} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ c_3 & c_4 \end{pmatrix} \begin{pmatrix} q_{k-1,0}^1 \\ q_{k-1,1}^1 \end{pmatrix} . \tag{6.103}$$

This motivates defining the matrix T whose entries are

$$T_{00} = 0$$
$$T_{01} = 1$$
$$T_{10} = c_3 = \frac{k_{12}k_{21}}{s^2} \tag{6.104}$$
$$T_{11} = c_4 = \frac{k_{21} - k_{12}}{s} ,$$

i.e.

$$T = \begin{pmatrix} T_{00} & T_{01} \\ T_{10} & T_{11} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ c_3 & c_4 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{k_{12}k_{21}}{s^2} & \frac{k_{21}-k_{12}}{s} \end{pmatrix} . \tag{6.105}$$

Keeping in mind the initial condition, we can now straightforwardly compute the coefficients $q_{k,0}^0$ via

$$\begin{pmatrix} q_{k,0}^0 \\ q_{k,1}^0 \end{pmatrix} = \begin{pmatrix} 0 & \frac{1}{k\gamma} \\ \frac{c_3}{k\gamma+s} & \frac{c_4}{k\gamma+s} \end{pmatrix} \cdots \begin{pmatrix} 0 & \frac{1}{\gamma} \\ \frac{c_3}{\gamma+s} & \frac{c_4}{\gamma+s} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} . \tag{6.106}$$

In other words, $q_{k,0}^0$ is the 00 (upper left) entry of a k-fold product of 2×2 matrices, and $q_{k,1}^0$ is the 10 (bottom left) entry. By the same argument, $q_{k,0}^1$ and $q_{k,1}^1$ are the 01 and 11 entries of a k-fold product. Explicitly,

$$q_{k,0}^0 (0 \to 0 \text{ in } k \text{ steps}) = \sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_m i_{m-1}}}{(m\gamma + i_m s)} \cdots \frac{T_{i_2 i_1}}{(2\gamma + i_2 s)} \frac{T_{i_1 0}}{(\gamma + i_1 s)}$$

$$q_{k,1}^0 (0 \to 1 \text{ in } k \text{ steps}) = \sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{1i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_m i_{m-1}}}{(m\gamma + i_m s)} \cdots \frac{T_{i_2 i_1}}{(2\gamma + i_2 s)} \frac{T_{i_1 0}}{(\gamma + i_1 s)}$$

$$q_{k,0}^1 (1 \to 0 \text{ in } k \text{ steps}) = \sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma - s)} \cdots \frac{T_{i_m i_{m-1}}}{(m\gamma - s + i_m s)} \cdots \frac{T_{i_1 1}}{(\gamma - s + i_1 s)} \tag{6.107}$$

$$q_{k,1}^1 (1 \to 1 \text{ in } k \text{ steps}) = \sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{1i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_m i_{m-1}}}{(m\gamma - s + i_m s)} \cdots \frac{T_{i_1 1}}{(\gamma - s + i_1 s)} .$$

These are the cumbersome expressions referred to in the previous subsection. In the next subsection, we offer a diagrammatic interpretation of them.

6.4.3 Feynman rules

It turns out that we can associate with each coefficient $q_{k,b}^a$ a diagram with k lines and $k+1$ vertices. We obtain the following Feynman rules for computing the Hilbert space/bra-ket representation of the eigenstates.

Hilbert space Feynman rules

In order to compute the coefficient of $|n+k;g\rangle$ in the infinite series expansion of $|n\rangle$ (c.f. Eq. 6.97), draw all valid diagrams going from 0 to $g \in \{0,1\}$ in k steps according to the following rules, and add the numbers corresponding to each diagram.

1. **Set up grid**: Write out positions 0 through k from left to right. Draw two parallel horizontal lines above these labels to denote the 0 and 1 gene eigenstates. The diagram will consist of $k+1$ vertices, each located at a horizontal $\{0, ..., k\}$ position and vertical gene eigenstate (lower or upper) position, and lines connecting those vertices.

2. **Draw lines**: Place the first vertex at horizontal position 0 and on the bottom row. Fill in the following positions from left to right. There are three possible moves: (i) if at 0, you must next go to 1; (ii) if at 1, you can go to 0 next; (iii) if at 1, you can stay at 1. If $g = 0$, the last vertex must be on the bottom row. If $g = 1$, the last vertex must be on the top row.

3. **Numerical factors**: Associate each move/line with a numerical factor. In particular, associate the move from position $m-1$ to position m with the factor:

   - $0 \to 1$ *flip*: $\dfrac{c_3}{s + m\gamma}$
   - $1 \to 1$ *stay*: $\dfrac{c_4}{s + m\gamma}$
   - $1 \to 0$ *flip*: $\dfrac{1}{m\gamma}$

4. **Tack on generic factors**: Multiply the numbers associated with each line together, along with the generic factors $d^k \sqrt{(n+1)\cdots(n+k)}$, to get the number corresponding to the diagram you drew.

The steps are displayed in the context of a specific example in Fig. 6.5. See Tables 6.2 and 6.3 in Appendix 6.C for the values of many low-order Feynman diagrams.

Analogous rules can be written for computing $q_{k,g}^1$: one must start on the upper row instead of the lower row, and the denominators associated with step 3 are slightly different, since we would have to make the substitutions $s + m\gamma \to m\gamma$ and $m\gamma \to m\gamma - s$.

1.

0   1   2   3   0   1   2   3

2.   **3.** $\dfrac{c_3}{s+\gamma}$ $\dfrac{c_4}{s+2\gamma}$ $\dfrac{1}{3\gamma}$

0   1   2   3   0   1   2   3

4.
$$= \frac{c_3 c_4}{(s+\gamma)(s+2\gamma)3\gamma} \times \begin{pmatrix} \text{generic} \\ \text{factors} \end{pmatrix}$$
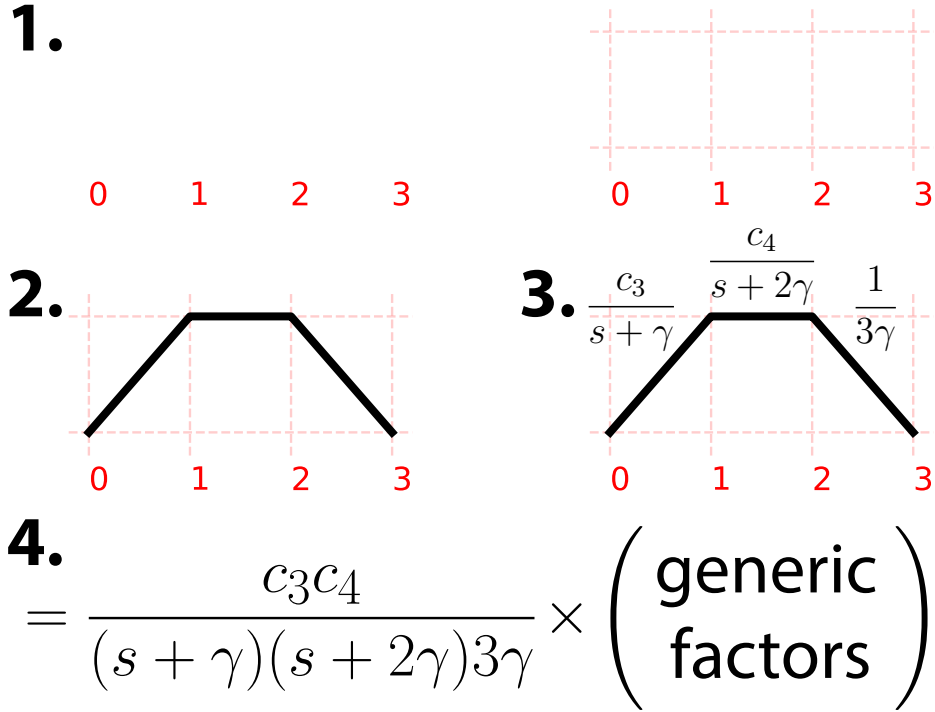
Figure 6.5: How to draw a Feynman diagram. Here, we illustrate the process using the unique diagram that goes from 0 to 0 in three steps. Step 1: Draw the grid. Step 2: Draw the lines for your diagram. Step 3: Associate each line with numerical factors. Step 4: Multiply the numbers for each line together, and tack on generic factors.

6.5   Special cases and limits

In this section, we examine various special cases of the result we found in the previous section. Of particular biological relevance are the steady state probability distribution and various special cases of it, including the limiting distributions described earlier in Sec. 6.2.3.

To summarize our work from the previous section, we found that $\hat{H}$ has eigenstates $|n\rangle$ (where $n \in \mathbb{N}$) with energies $E_n = \gamma n$ which can be written in terms of the naive eigenstates $|n; g\rangle$ via

$$
\begin{aligned}
|n\rangle &= |n; 0\rangle + \sum_{k=1}^{\infty} \sum_{i_k=0,1} d^k \sqrt{(n+1)_k}\; q^0_{k,i_k} |n+k; i_k\rangle \\
&= |n; 0\rangle + \sum_{k=1}^{\infty} d^k \sqrt{(n+1)_k} \sum_{i_1,\dots,i_k=0,1} \frac{T_{i_k i_{k-1}}}{(k\gamma + i_k s)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} |n+k; i_k\rangle \;,
\end{aligned}
\tag{6.108}
$$

and eigenstates $|n + s/\gamma\rangle$ with energies $E_n = \gamma n + s$ that can be written in terms of naive

168

eigenstates as

$$|n + s/\gamma\rangle = |n; 1\rangle + \sum_{k=1}^{\infty} \sum_{i_k=0,1} d^k \sqrt{(n+1)_k}\, q^1_{k,i_k} |n + k; i_k\rangle$$

$$= |n; 1\rangle + \sum_{k=1}^{\infty} d^k \sqrt{(n+1)_k} \sum_{i_1,\ldots,i_k=0,1} \frac{T_{i_k i_{k-1}}}{(k\gamma - s + i_k s)} \cdots \frac{T_{i_1 1}}{(\gamma - s + i_1 s)} |n + k; i_k\rangle \ . \tag{6.109}$$

For comparison with experiment, we are most interested in a special eigenstate: the eigenstate with eigenvalue 0, which corresponds to the steady state probability distribution. Specializing the above, we have the 'vacuum state'

$$|0\rangle = |0; 0\rangle + \sum_{k=1}^{\infty} d^k \sqrt{k!} \sum_{i_1,\ldots,i_k=0,1} \frac{T_{i_k i_{k-1}}}{(k\gamma + i_k s)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} |k; i_k\rangle \ . \tag{6.110}$$

6.5.1 Steady state probability distribution

We can extract what we *really* care about—the probability distribution corresponding to Eq. 6.110—by either picking out coefficients, or invoking the Euclidean product as in Sec. 6.3.1. For the states $|x; S\rangle$ (where x represents molecule number and $S \in \{1, 2\}$ represents gene state, rather than gene eigenstate), from which the naive eigenkets $|n; g\rangle$ can be constructed, the Euclidean product can be defined via

$$\langle x_1; S_1 | x_2; S_2 \rangle_{Eu} := \delta_{x_1,x_2} \delta_{S_1,S_2} \tag{6.111}$$

so that

$$\langle x; S | n; g \rangle_{Eu} = \sqrt{\frac{\mu^n}{n!}}\, C_n(x, \mu) \text{Poiss}(x, \mu)(\vec{v}_g)_S \ . \tag{6.112}$$

Either way, we find

$$\frac{P_{ss}(x, \vec{S})}{\text{Poiss}(x, \mu)} = \vec{v}_0 + \sum_{k=1}^{\infty} (\Delta\alpha)^k\, C_k(x, \mu) \sum_{i_1,\ldots,i_k=0,1} \frac{T_{i_k i_{k-1}}}{(k\gamma + i_k s)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} \vec{v}_{i_k} \tag{6.113}$$

where $\Delta\alpha := \alpha_1 - \alpha_2$ is the production rate difference. We can marginalize this result over gene state by applying $\vec{w}_0^T = (1, 1)^T$ on the left; doing so, we obtain

$$\frac{P_{ss}(x)}{\text{Poiss}(x, \mu)} = 1 + \sum_{k=2}^{\infty} (\Delta\alpha)^k\, C_k(x, \mu) \sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{0 i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} \ . \tag{6.114}$$

The corresponding result for the (analytic) generating function, which is straightforward to compute from this formula, is presented in Sec. 6.2.1. If we are only interested in the steady state distribution marginalized over gene state, we can modify the Feynman rules we presented earlier to compute it instead of the Hilbert space coefficients. These can be viewed as the Feynman rules for 'molecular number space' rather than Hilbert space.

169

Molecule number space Feynman rules

In order to compute the terms of order $(\Delta\alpha)^k$ in the infinite series expansion of $P_{ss}(x)/\text{Poiss}(x,\mu)$ (c.f. Eq. 6.114), draw all valid diagrams going from 0 to 0 in k steps according to the following rules, and add the numbers corresponding to each diagram.

1. **Set up grid**: Write out positions 0 through k from left to right. Draw two parallel horizontal lines above these labels to denote the 0 and 1 gene eigenstates. The diagram will consist of $k+1$ vertices, each located at a horizontal $\{0, ..., k\}$ position and vertical gene eigenstate (lower or upper) position, and lines connecting those vertices.

2. **Draw lines**: Start on the bottom row (0). There are three possible moves: (i) if at 0, you must next go to 1; (ii) if at 1, you can go to 0 next; (iii) if at 1, you can stay at 1. The last vertex must be on the bottom row.

3. **Numerical factors**: Associate each move/line with a numerical factor. In particular, associate the move from position $m-1$ to position m with the factor:

   - $0 \to 1$ *flip*: $\dfrac{c_3}{s + m\gamma}$

   - $1 \to 1$ *stay*: $\dfrac{c_4}{s + m\gamma}$

   - $1 \to 0$ *flip*: $\dfrac{1}{m\gamma}$

4. **Tack on generic factors**: Multiply the numbers associated with each line together, along with the generic factors $(\Delta\alpha)^k C_k(x,\mu)$, to get the number corresponding to the diagram you drew.

It appears that the appropriate correlation function for this problem is given by the function

$$\langle 0|\hat{A}^2|0\rangle = \frac{\sigma_x^2 - \mu}{\mu} = \frac{(\Delta\alpha)^2 c_3}{(s+\gamma)\gamma} \leq \frac{(\Delta\alpha)^2}{4\gamma^2} \tag{6.115}$$

where the inner product is defined analogously to before (c.f. Eq. 6.44) on molecule number basis kets via

$$\langle x|y\rangle := \frac{\delta_{xy}}{P_{ss}(x)} \ . \tag{6.116}$$

This function measures how much the steady state distribution deviates from a Poisson distribution; it corresponds to the first nontrivial $0 \to 0$ diagram, which can be seen in the upper left corner of Fig. 6.6. It may be possible to view this as the propagator for some sort of quasiparticle (a transcripton?), but the advantages of such a view are not completely clear. In the next few sections, we examine biologically relevant special cases and limits of Eq. 6.114.

First few diagrams for computing the steady state probability
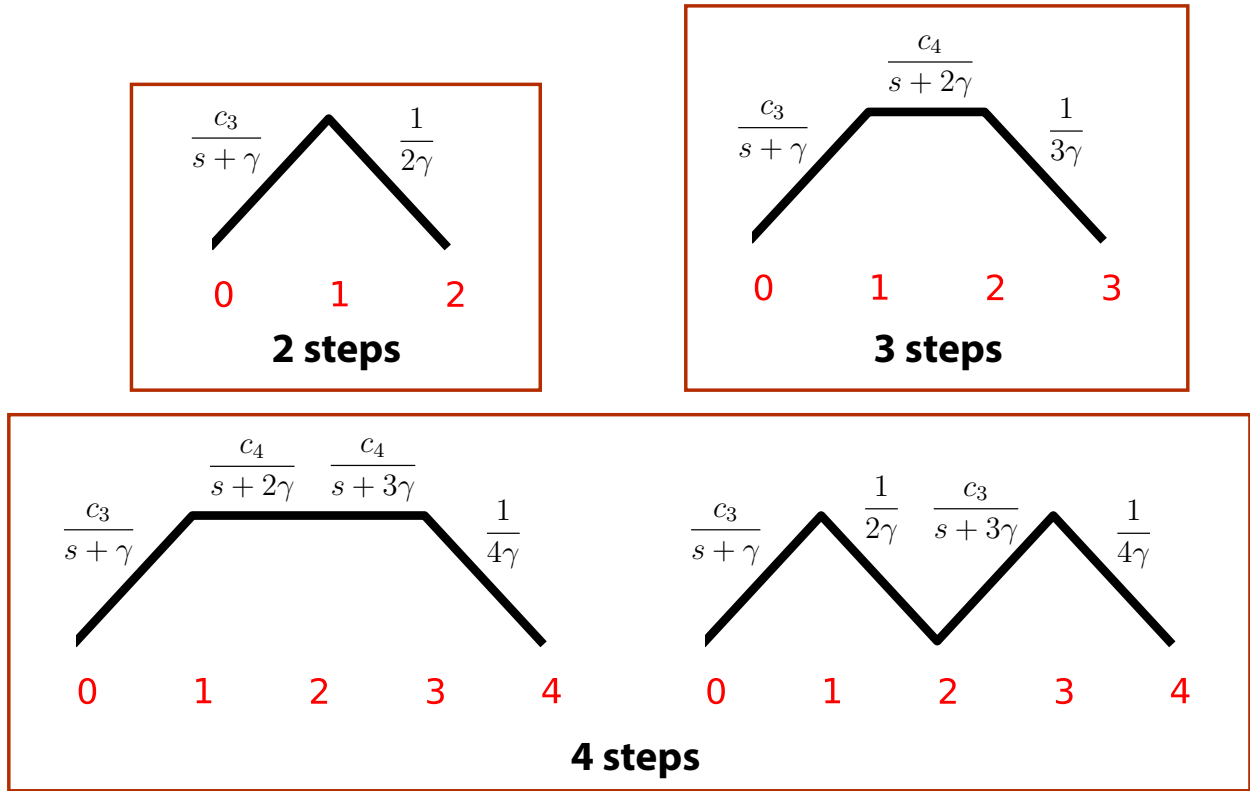
$$P_{ss} = \text{Poisson} \times \left(1 + \wedge + \diagup\wedge + \cdots\right)$$

Figure 6.6: First few Feynman diagrams contributing to $P_{ss}(x)$, with the numerical factors corresponding to each line shown explicitly. The overall result for $P_{ss}(x)$ is obtained by multiplying the sum of all diagrams by a Poisson distribution.

6.5.2 Equal switching rates

In the special case of equal switching rates, where $k_{12} = k_{21}$, the exact solution simplifies considerably. In this regime, we have

$$c_3 = \frac{k_{12}k_{21}}{s^2} = \frac{1}{4}$$
$$c_4 = \frac{k_{21} - k_{12}}{s} = 0$$

(6.117)

which means that all Feynman diagrams involving a $1 \to 1$ 'stay' line vanish. This leaves only diagrams of the form $0 \to 1$, $0 \to 1 \to 0$, $0 \to 1 \to 0 \to 1$, and so on, i.e. the 'zigzag'

diagrams (see Fig. 6.7). In particular, for k even we have

$$q_{k,0}^0 = \frac{c_3^{k/2}}{(s+\gamma)2\gamma(s+3\gamma)\cdots[s+(k-1)\gamma]k\gamma}$$

$$q_{k,1}^0 = 0$$

$$q_{k,0}^1 = 0 \tag{6.118}$$

$$q_{k,1}^1 = \frac{c_3^{k/2}}{(\gamma-s)2\gamma(3\gamma-s)\cdots[(k-1)\gamma-s]k\gamma}$$

and for k odd we have

$$q_{k,0}^0 = 0$$

$$q_{k,1}^0 = \frac{c_3^{(k+1)/2}}{(s+\gamma)2\gamma(s+3\gamma)\cdots[(k-1)\gamma](s+k\gamma)}$$

$$q_{k,0}^1 = \frac{c_3^{(k-1)/2}}{(\gamma-s)2\gamma(3\gamma-s)\cdots(k-1)\gamma(k\gamma-s)} \tag{6.119}$$

$$q_{k,1}^1 = 0 \ .$$

This allows us to write explicit formulas like

$$|n\rangle = |n,0\rangle + \sum_{k \text{ even}>0}^{\infty} \frac{d^k\sqrt{(n+1)_k}\,c_3^{k/2}}{(s+\gamma)2\gamma(s+3\gamma)\cdots[s+(k-1)\gamma]k\gamma}\,|n+k;0\rangle$$

$$+ \sum_{k \text{ odd}}^{\infty} \frac{d^k\sqrt{(n+1)_k}\,c_3^{(k+1)/2}}{(s+\gamma)2\gamma(s+3\gamma)\cdots[(k-1)\gamma](s+k\gamma)}\,|n+k;1\rangle \ . \tag{6.120}$$

The corresponding formula for the steady state probability distribution marginalized over gene state is

$$\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} = 1 + \sum_{k=2,4,\ldots}^{\infty}(\Delta\alpha)^k \frac{c_3^{k/2}}{(s+\gamma)2\gamma(s+3\gamma)\cdots[s+(k-1)\gamma]k\gamma}C_k(x,\mu) \ . \tag{6.121}$$

Rewriting it in terms of Gamma functions, we have the formula

$$\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} = \Gamma\left(\frac{s}{2\gamma}+\frac{1}{2}\right)\sum_{m=0}^{\infty}\left[\frac{\Delta\alpha}{4\gamma}\right]^{2m}\frac{C_{2m}(x,\mu)}{m!\,\Gamma\left(\frac{s}{2\gamma}+\frac{1}{2}+m\right)} \ , \tag{6.122}$$

which makes it easy to compute that the steady state (analytic) generating function is

$$\psi_{ss}(g) = e^{\mu(g-1)}\Gamma\left(\frac{s}{2\gamma}+\frac{1}{2}\right)\left[\frac{4\gamma}{\Delta\alpha(g-1)}\right]^{\frac{s}{2\gamma}-\frac{1}{2}}I_{\frac{s}{2\gamma}-\frac{1}{2}}\left(\frac{\Delta\alpha}{2\gamma}(g-1)\right) \tag{6.123}$$

where $I_\nu(z)$ is the modified Bessel function of the first kind.

In the case of equal switching rates, for sufficiently different α_1 and α_2 the distribution is generically bimodal with unequally sized peaks. In the limit of very large α_1 and α_2 (i.e. the continuous concentration limit discussed in Sec. 6.6), this asymmetry disappears, and P_{ss} becomes completely symmetric about the effective mean μ (see Fig. 6.8).

	Special case	Relevant diagrams	Example

zigzag diagrams

$$k_{12} = k_{21} \qquad \wedge + \wedge\wedge + \wedge\wedge\wedge + \cdots$$

balance beam diagrams

$$k_{12} \ll k_{21} \qquad \wedge + \frown\wedge + \frown\wedge + \cdots$$
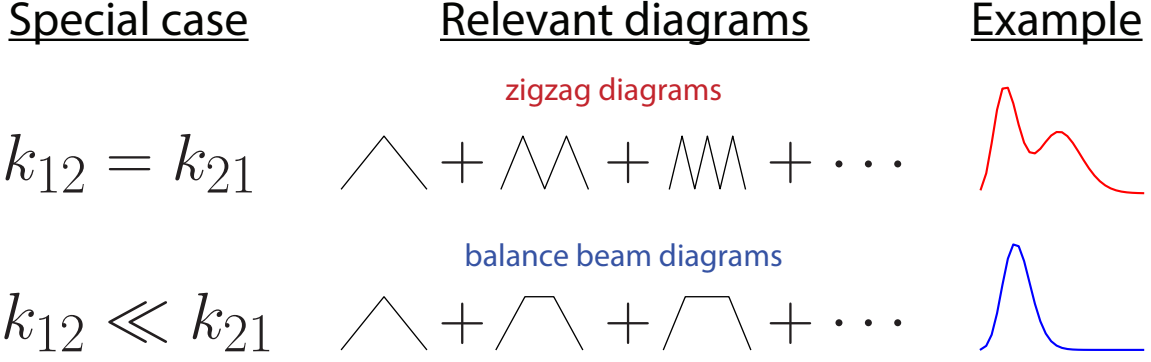
Figure 6.7: Dominant Feynman diagrams in two special cases. When the gene switching rates are equal ($k_{12} = k_{21}$), only the 'zigzag' diagrams are nonzero. When the gene switching rates are very unequal (e.g. $k_{12} \ll k_{21}$), the 'balance beam' diagrams contribute the leading terms.

6.5.3 Very unequal switching rates

Suppose that k_{12} and k_{21} are very different—as is often the case in practice with bursty transcription, where one gene state is considered 'active' and the other 'inactive'. Without loss of generality, let $k_{21} \gg k_{12}$, so that the $1 \to 2$ transition happens much more readily than the $2 \to 1$ transition. Define $r := k_{12}/k_{21}$ and note $r \ll 1$. In this regime, we are interested in finding the steady state probability distribution to first order in r. Note that

$$
\begin{aligned}
c_3 &= \frac{k_{12}k_{21}}{(k_{12} + k_{21})^2} = \frac{r}{1 + 2r + r^2} \approx r \\
c_4 &= \frac{k_{21} - k_{12}}{k_{21} + k_{12}} = \frac{1 - r}{1 + r} \approx 1 - 2r
\end{aligned}
\tag{6.124}
$$

to first order approximation in r. Hence, $c_3 \sim r$ and $c_4 \sim 1$ at leading order. Then the dominant diagrams contributing to the steady state probability sum have first step $0 \to 1$, all intermediate steps $1 \to 1$, and last step $1 \to 0$.

There exists such a diagram for all $k \geq 2$, and its value is always (to first order in r) proportional to r. We will call these 'balance beam' diagrams (see Fig. 6.7 for a visual). Then we have

$$
\begin{aligned}
\sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} &\approx \frac{T_{01}}{(k\gamma)} \cdots \frac{T_{11}}{(2\gamma + s)} \frac{T_{10}}{(\gamma + s)} \\
&\approx \frac{r}{\gamma^k} \frac{1}{[(s/\gamma) + 1][(s/\gamma) + 2] \cdots [(s/\gamma) + (k-1)]k} \\
&= \frac{r}{\gamma^k} \frac{(s/\gamma)}{\left(\frac{s}{\gamma}\right)_k},
\end{aligned}
\tag{6.125}
$$

so that our approximate formula for $P_{ss}(x)$ in the very unequal switching rates regime reads

$$\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} \approx 1 + \frac{sr}{\gamma} \sum_{k=2}^{\infty} \left(\frac{\Delta\alpha}{\gamma}\right)^k C_k(x,\mu) \; \frac{1}{\left(\frac{s}{\gamma}\right)_k} \; . \tag{6.126}$$

Like the negative binomial distribution, this distribution is somewhat heavy-tailed. This extra variance is thought to be one of the more dramatic consequences of gene switching for empirical single cell RNA counts distributions.

It should also be noted that this 'dominant' diagram kind of analysis recalls what one often does in field theory, e.g. in studies of condensed matter systems [52]. This is one of the advantages of this Feynman diagram way of thinking: various approximations can be justified by observing in some regime that certain diagrams are dominant, and others can be ignored.

The very unequal switching rates limit is closely related to the *bursty* / negative binomial limit we mentioned in Sec. 6.2.3. Recall that the relevant parameter conditions are

$$\alpha_2 = 0$$
$$\alpha_1 \to \infty, \; k_{21} \to \infty$$
$$b := \frac{\alpha_1}{k_{21}} \text{ held fixed} \tag{6.127}$$
$$\omega := \frac{k_{12}}{\gamma} \text{ held fixed}$$

where b is called the burst size and ω is called the burst frequency. Unfortunately, because this limit requires keeping α_1/k_{21} held fixed, we cannot simply take the relevant limit of Eq. 6.126. Instead, we would have to start again with Eq. 6.114.

It turns out that a somewhat slicker way to recover the negative binomial limit is to use the hypergeometric form of the (analytic) generating function, which we derive from Eq. 6.114 in Appendix 6.E. For $\alpha_2 = 0$, it reads

$$\psi_{ss}(g) = \sum_{n=0}^{\infty} \frac{\left[\frac{\alpha_1}{\gamma}(g-1)\right]^n}{n!} \frac{\left(\frac{k_{12}}{\gamma}\right)_n}{\left(\frac{s}{\gamma}\right)_n} \; . \tag{6.128}$$

First, note that

$$\begin{aligned}
\left(\frac{s}{\gamma}\right)_n &= \frac{(k_{12}+k_{21})}{\gamma}\left(\frac{(k_{12}+k_{21})}{\gamma}+1\right)\cdots\left(\frac{(k_{12}+k_{21})}{\gamma}+n-1\right) \\
&\approx \frac{k_{21}}{\gamma}\left(\frac{k_{21}}{\gamma}+1\right)\cdots\left(\frac{k_{21}}{\gamma}+n-1\right) \\
&\approx \left(\frac{k_{21}}{\gamma}\right)^n
\end{aligned} \tag{6.129}$$

174

in the large k_{21} limit. Taking α_1 and k_{21} large while keeping b and ω held fixed,

$$
\begin{aligned}
\psi_{ss}(g) &\approx \sum_{n=0}^{\infty} \frac{\left[\frac{\alpha_1}{\gamma}(g-1)\right]^n}{n!} \frac{\left(\frac{k_{12}}{\gamma}\right)_n}{\left(\frac{k_{21}}{\gamma}\right)^n} \\
&= \sum_{n=0}^{\infty} \frac{[b(g-1)]^n}{n!} (\omega)_n \\
&= \left[\frac{1}{1-b(g-1)}\right]^{\omega}
\end{aligned}
\tag{6.130}
$$

which is precisely the generating function of a negative binomial distribution with $p = b/(1+b)$ and $r = \omega$.

6.5.4 Switching much faster than degradation

If gene switching tends to happen much more frequently than degradation, so that $s \gg \gamma$, we can approximate the Feynman denominators $(m\gamma + i_m s) \approx i_m s$ where $i_m = 1$, yielding an asymptotic series in powers of $1/s$:

$$
\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} \approx 1 + \frac{1}{s}\left[(\Delta\alpha)^2 C_2 \frac{c_3}{2\gamma}\right] + \frac{1}{s^2}\left[(\Delta\alpha)^3 C_3 \frac{c_3 c_4}{3\gamma} + (\Delta\alpha)^4 C_4 \frac{c_3^2}{8\gamma^2}\right] + \cdots .
\tag{6.131}
$$

In the $s/\gamma \to \infty$ limit, the distribution becomes exactly Poisson with mean μ. Intuitively, gene switching happens so fast in this regime that the gene product only 'feels' the effective transcription rate given by α_{eff}, and the problem reduces to the familiar chemical birth-death process.

6.5.5 Switching much slower than degradation

If degradation tends to happen much more frequently than gene switching, so that $\gamma \gg s$, we can approximate the Feynman denominators $(m\gamma + i_m s) \approx m\gamma$, yielding

$$
\begin{aligned}
\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} &\approx 1 + \sum_{k=1}^{\infty} \left[\frac{\Delta\alpha}{\gamma}\right]^k \frac{1}{k!} C_k(x,\mu) \sum_{i_1,\dots,i_{k-1}=0,1} T_{0 i_{k-1}} \cdots T_{i_1 0} \\
&\overset{00}{=} \sum_{k=0}^{\infty} \left[\frac{\Delta\alpha}{\gamma}T\right]^k \frac{1}{k!} C_k(x,\mu)
\end{aligned}
\tag{6.132}
$$

where the notation $\overset{00}{=}$ means that the LHS is the 00 (upper left) entry of the 2×2 matrix described by the RHS. We can sum this in closed form using the known formula for the generating function of the Charlier polynomials (c.f. Eq. 6.209 in Appendix 6.A) to obtain the beautiful approximate formula

$$
\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} \overset{00}{\approx} \left[1 + \frac{\Delta\alpha}{\gamma\mu}T\right]^x e^{-(\Delta\alpha)T/\gamma} .
\tag{6.133}
$$

Recall from Sec. 6.2.3 that, when $\gamma \gg s$, we expect the steady state probability distribution to look like a mixture of two Poisson distributions (with means α_1/γ and α_2/γ). Because switching is slow compared to the dynamics of production and degradation, on short time scales the system looks like a birth-death process; the effect of switching is to move from one Poisson distribution to the other.

Although this formula appears superficially different, it turns out to be *exactly the same* as a Poisson mixture. To see why, note that we can write

$$T = \begin{pmatrix} 0 & 1 \\ f(1-f) & 1-2f \end{pmatrix} \tag{6.134}$$

where $f := k_{12}/s$. Diagonalizing T, we find that

$$T = \begin{pmatrix} 1 & 1 \\ -f & 1-f \end{pmatrix} \begin{pmatrix} -f & 0 \\ 0 & 1-f \end{pmatrix} \begin{pmatrix} 1-f & -1 \\ f & 1 \end{pmatrix} . \tag{6.135}$$

This representation makes it easy to take powers of T, which we can use to evaluate Eq. 6.132 in a different way. We have

$$\left(T^k\right)_{00} = f(1-f)^k + (1-f)(-f)^k \tag{6.136}$$

which means

$$\begin{aligned}
\frac{P_{ss}(x)}{\text{Poiss}(x,\mu)} &\approx \sum_{k=0}^{\infty} \left[\frac{\Delta\alpha}{\gamma}\right]^k \left\{ f(1-f)^k + (1-f)(-f)^k \right\} \frac{1}{k!} C_k(x,\mu) \\
&= f \sum_{k=0}^{\infty} \left[\frac{\Delta\alpha}{\gamma}(1-f)\right]^k \frac{1}{k!} C_k(x,\mu) + (1-f) \sum_{k=0}^{\infty} \left[\frac{\Delta\alpha}{\gamma}(-f)\right]^k \frac{1}{k!} C_k(x,\mu) \\
&= f \left(1 + \frac{(1-f)\frac{\Delta\alpha}{\gamma}}{\mu}\right)^x e^{-(1-f)\frac{\Delta\alpha}{\gamma}} + (1-f)\left(1 - \frac{f\frac{\Delta\alpha}{\gamma}}{\mu}\right)^x e^{f\frac{\Delta\alpha}{\gamma}}
\end{aligned} \tag{6.137}$$

where we used the generating function of the Charlier polynomials (Eq. 6.209) in the last step. Simplifying, we obtain

$$\begin{aligned}
P_{ss}(x) &\approx \frac{\mu^x}{x!} e^{-\mu} \left[\frac{k_{12}}{s}\left(1 + \frac{k_{21}}{s}\frac{\Delta\alpha}{\gamma}\right)^x e^{-\frac{k_{21}}{s}\frac{\Delta\alpha}{\gamma}} + \frac{k_{21}}{s}\left(1 - \frac{k_{12}}{s}\frac{\Delta\alpha}{\gamma}\right)^x e^{\frac{k_{12}}{s}\frac{\Delta\alpha}{\gamma}} \right] \\
&= \frac{k_{12}}{s}\frac{(\alpha_1/\gamma)^x}{x!} e^{-\alpha_1/\gamma} + \frac{k_{21}}{s}\frac{(\alpha_2/\gamma)^x}{x!} e^{-\alpha_2/\gamma}
\end{aligned} \tag{6.138}$$

i.e. a Poisson mixture.

6.6   The continuous concentration limit

Remarkably, by solving the birth-death-switching CME we have also solved a qualitatively very different problem almost for free: the problem of coupling an Ornstein-Uhlenbeck-like process to a switching gene, which corresponds to the continuous concentration limit of the discrete birth-death process. The same ladder operator/diagrammatic approach we have just followed works, with only minor changes.

The chemical birth-death process with additive noise [50] is governed by the Fokker-Planck equation

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}\left[(\alpha - \gamma x)P(x,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 P(x,t)}{\partial x^2} \tag{6.139}$$

where $P(x,t)$ is the probability that the concentration of X molecules in the system is $x \in \mathbb{R}$ at time t, and where $\sigma > 0$ is called the additive noise coefficient. It is equivalent to the Ornstein-Uhlenbeck process up to a simple change of variables, which is most transparently seen by noting that its stochastic trajectories follow the stochastic differential equation (SDE)

$$\dot{x} = \alpha - \gamma x + \sigma\eta(t) \tag{6.140}$$

with $\eta(t)$ a Gaussian white noise term. It is biologically interesting because, in the regime where typical molecule numbers are large, it is reasonable to approximate the discrete molecule number dynamics of the chemical birth-death process as the continuous concentration dynamics described by this Fokker-Planck equation[6].

If we coupled this process to a switching gene, we would obtain the Fokker-Planck equations

$$\frac{\partial P(x,1,t)}{\partial t} = -k_{21}P(x,1,t) + k_{12}P(x,2,t) - \frac{\partial}{\partial x}\left[(\alpha_1 - \gamma x)P(x,1,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 P(x,1,t)}{\partial x^2}$$
$$\frac{\partial P(x,2,t)}{\partial t} = k_{21}P(x,1,t) - k_{12}P(x,2,t) - \frac{\partial}{\partial x}\left[(\alpha_2 - \gamma x)P(x,2,t)\right] + \frac{\sigma^2}{2}\frac{\partial^2 P(x,2,t)}{\partial x^2}$$
$$\tag{6.141}$$

with $P(x,i,t)$ denoting the probability that the system has concentration $x \in \mathbb{R}$ and gene state $i \in \{1,2\}$ at time t.

Analogously to before, we can first consider the problem without a switching gene (Eq. 6.139), and define the generating function via

$$|\phi\rangle = \int_{-\infty}^{\infty} dx\ P(x,t)\ |x\rangle\ . \tag{6.142}$$

If we introduce operators $\hat{x}, \hat{p}$ that act as

$$\hat{x}\,|\phi\rangle := \int_{-\infty}^{\infty} dx\ x\ c(x)\,|x\rangle$$
$$\hat{p}\,|\phi\rangle := \int_{-\infty}^{\infty} dx\ -\frac{\partial c(x)}{\partial x}\,|x\rangle \tag{6.143}$$

---

[6]Negative concentrations are allowed but are overwhelmingly improbable in the regime where this continuous model is used, i.e. assuming $\alpha/\gamma \gg 1$ so that typical molecule numbers are large.

on a general state $|\phi\rangle$, we can write the equation of motion satisfied by the generating function in the form

$$\frac{\partial |\psi\rangle}{\partial t} = \hat{H}_{bda} |\psi\rangle \tag{6.144}$$

where the Hamiltonian operator is defined as

$$\hat{H}_{bda} = \hat{p}\left(\alpha - \gamma\hat{x}\right) + \frac{\sigma^2}{2}\,\hat{p}^2 \ . \tag{6.145}$$

We can go on to derive ladder operators

$$\hat{A} := \frac{1}{\sqrt{\sigma^2/2\gamma}}\left[\hat{x} - \mu - \frac{\sigma^2}{2\gamma}\hat{p}\right]$$

$$\hat{A}^+ := \sqrt{\frac{\sigma^2}{2\gamma}}\,\hat{p} \tag{6.146}$$

whose properties match those discussed in Sec. 6.3.1, which includes that we can write the Hamiltonian in terms of them as

$$\hat{H} = -\gamma\,\hat{A}^+\hat{A} \ . \tag{6.147}$$

With that done, we have ladder operators for the continuous birth-death problem, as well as the ones derived in Sec. 6.3.2 for the switching gene problem. We can pose the coupled problem in terms of them, as in Sec. 6.3.3, via

$$
\begin{aligned}
\hat{H} &= -s\hat{B}^+\hat{B} + \hat{p}\left[\hat{\alpha} - \gamma\hat{x}\right] + \frac{\sigma^2}{2}\hat{p}^2 \\
&= -s\hat{B}^+\hat{B} - \gamma\hat{A}^+\hat{A} + (\alpha_1 - \alpha_2)\,\hat{p}\,\left[\hat{B} + \frac{k_{12}k_{21}}{s^2}\hat{B}^+ + \frac{k_{21} - k_{12}}{s}\hat{B}^+\hat{B}\right] \\
&= -s\hat{B}^+\hat{B} - \gamma\hat{A}^+\hat{A} + \frac{(\alpha_1 - \alpha_2)}{\sqrt{\sigma^2/2\gamma}}\,\hat{A}^+\,\left[\hat{B} + \frac{k_{12}k_{21}}{s^2}\hat{B}^+ + \frac{k_{21} - k_{12}}{s}\hat{B}^+\hat{B}\right] \\
&= \hat{H}_s + \hat{H}_{bda} + \hat{H}_{int} \ .
\end{aligned}
\tag{6.148}
$$

As before, we can simplify notation by defining the quantities

$$
\begin{aligned}
d &:= \frac{\alpha_1 - \alpha_2}{\sqrt{\sigma^2/2\gamma}} \\
c_3 &:= \frac{k_{12}k_{21}}{s^2} \\
c_4 &:= \frac{k_{21} - k_{12}}{s}
\end{aligned}
\tag{6.149}
$$

so that the interaction Hamiltonian $\hat{H}_{int}$ can be written as

$$\hat{H}_{int} = d\,\hat{A}^+\left[\hat{B} + c_3\hat{B}^+ + c_4\hat{B}^+\hat{B}\right] \ . \tag{6.150}$$

178

But this is precisely the same interaction Hamiltonian we solved earlier (c.f. Eq. 6.96)! Because the method from here on out works in Hilbert space, rather than molecule number space or concentration space, we can go on derive Feynman rules of exactly the same form.

We must do something slightly different only when we go from the generating function to the probability distribution. The analogue to the Euclidean product (c.f. Eq. 6.58) is

$$\langle \phi_1 | \phi_2 \rangle_{Eu} := \int dx \, c_1^*(x) c_2(x) \tag{6.151}$$

and the relevant results (given no gene switching) can be summarized as

$$\langle x | \psi \rangle_{Eu} = P(x, t)$$
$$\langle x | n \rangle_{Eu} = \sqrt{\frac{1}{2^n n!}} H_n \left(\sqrt{\frac{\gamma}{\sigma^2}} (x - \mu) \right) \cdot \text{Gauss} \left(x, \mu, \frac{\sigma^2}{2\gamma} \right) \tag{6.152}$$
$$\text{Gauss}(x, m, s^2) := \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}}$$

where H_n denotes the nth Hermite polynomial. Continuing as before, we find[7]

$$\frac{P_{ss}(x, \vec{g})}{\text{Gauss}\left(x, \mu, \frac{\sigma^2}{2\gamma}\right)} = \vec{v}_0 + \sum_{k=1}^{\infty} \left[\frac{\Delta\alpha}{\sqrt{\sigma^2/\gamma}} \right]^k H_k \left(\sqrt{\frac{\gamma}{\sigma^2}} (x - \mu) \right) \sum_{i_1,\dots,i_k=0,1} \frac{T_{i_k i_{k-1}}}{(k\gamma + i_k s)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} \vec{v}_{i_k} . \tag{6.153}$$

If we marginalize over gene state, we obtain

$$\frac{P_{ss}(x)}{\text{Gauss}\left(x, \mu, \frac{\sigma^2}{2\gamma}\right)} = 1 + \sum_{k=2}^{\infty} \left[\frac{\Delta\alpha}{\sqrt{\sigma^2/\gamma}} \right]^k H_k \left(\sqrt{\frac{\gamma}{\sigma^2}} (x - \mu) \right) \sum_{i_1,\dots,i_{k-1}=0,1} \frac{T_{0 i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} . \tag{6.154}$$

One easily notes that the only difference between these results and the ones for the discrete birth-death process are the substitutions

$$\text{Poiss}(x, \mu) \to \text{Gauss} \left(x, \mu, \frac{\sigma^2}{2\gamma} \right)$$
$$C_k(x, \mu) \to \left[\frac{1}{\sqrt{\sigma^2/\gamma}} \right]^k H_k \left(\sqrt{\frac{\gamma}{\sigma^2}} (x - \mu) \right) . \tag{6.155}$$

---

[7]See [53] for another interesting series expansion involving the Hermite polynomials. Thomas and Grima study higher-order terms in the system size expansion of the CME; the Hermite polynomials seem to appear essentially due to fact that (most) arbitrary CMEs can be locally approximated as a multivariate birth-death process with additive noise.

Because the birth-death process with additive noise (with the specific additive noise constant $\sigma = \sqrt{2\alpha}$) is the large μ limit of the discrete birth-death process (see Fig. 6.8), there is an alternative way to get this result. One merely needs to know that

$$C_k(x, \mu) \xrightarrow{\mu \gg 1} \left[\frac{1}{\sqrt{2\mu}}\right]^k H_k\left(\frac{x - \mu}{\sqrt{2\mu}}\right) \tag{6.156}$$

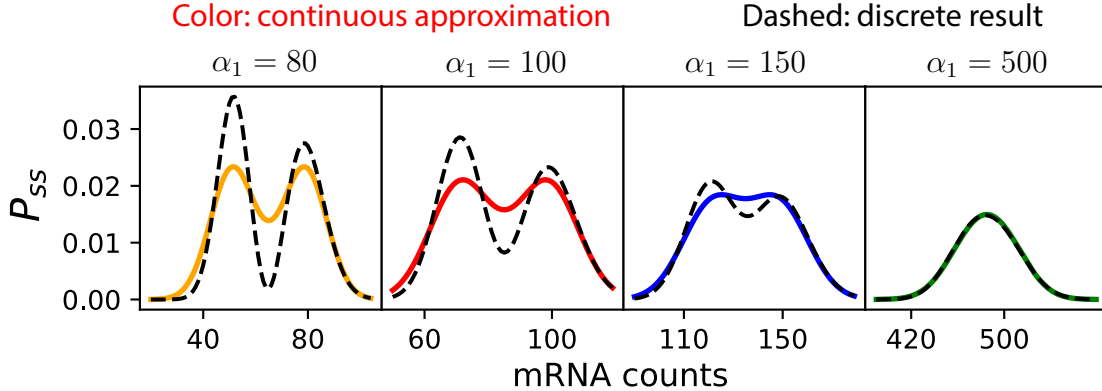so that Eq. 6.114 becomes Eq. 6.154 (with $\sigma = \sqrt{2\alpha}$).

Figure 6.8: Comparison of the continuous approximation to the discrete birth-death-switching result for equal switching rates as α_1 is varied. $\alpha_2 = \alpha_1 - 30$, $\gamma = 1$, $k_{12} = k_{21} = 0.1$. The continuous result is symmetric about the effective mean μ in all cases; the discrete result approaches the continuous result as α_1 becomes large.

It is interesting to note that problems of similar form appear elsewhere. For example, by specializing the parameter values of this problem, one finds that Eq. 6.153 and Eq. 6.154 can exactly describe the solution of an active matter run-and-tumble motion problem [54]. The precise correspondence is described in Table 6.1 below.

Gene regulation problem	Run-and-tumble problem
Concentration x	Position x
Gene state $i \in \{1, 2\}$	Right-moving or left-moving
$P(x, 1, t)$ and $P(x, 2, t)$	$P_\phi(x, t)$ and $P_\psi(x, t)$
Transcription rates α_1, α_2	Self-propulsion speeds $\frac{w}{2}, -\frac{w}{2}$
Degradation rate γ	Spring constant k
Gene switching rates k_{12}, k_{21}	Direction switching rates $\frac{\alpha}{2}, \frac{\alpha}{2}$
Effective mean concentration $\mu = \frac{\alpha_1 k_{12} + \alpha_2 k_{21}}{\gamma(k_{12} + k_{21})}$	Effective mean position $\mu = 0$

Table 6.1: The parameter correspondence between the continuous birth-death-switching problem considered here, and the run-and-tumble problem considered by Garcia-Millan and Pruessner [54].

6.7 Generalization to multistep splicing

Fortuitously, our approach to solving the birth-death-switching problem—a problem that has been solved before [39, 40, 41], albeit using very different methods—straightforwardly generalizes, so that we can solve a significantly more complicated problem that has not been solved before. In this section, we consider more realistic transcription dynamics, involving an arbitrary number of splicing steps that occur in a fixed order, coupled to a switching gene.

Our plan of attack is the same as the one detailed in Sec. 6.2.4. The ladder operators for pure gene switching are the same as before, so we will first find the ladder operators for the dynamics of transcription and multistep splicing given a constitutively active gene. The list of reactions and CME for the coupled problem are as stated in Sec. 6.2.2 (c.f. Eq. 6.11).

6.7.1 Representation theory of multistep splicing

Assuming a constitutively active gene, the reaction list for transcription and multistep splicing (with N splicing steps that occur in a fixed order) reads

$$
\begin{aligned}
\varnothing &\xrightarrow{\alpha} X_0 \\
X_0 &\xrightarrow{\beta_0} X_1 \\
X_1 &\xrightarrow{\beta_1} X_2 \\
&\vdots \\
X_{N-1} &\xrightarrow{\beta_{N-1}} X_N \\
X_N &\xrightarrow{\beta_N} \varnothing
\end{aligned}
\tag{6.157}
$$

where α parameterizes the transcription rate, β_i parameterizes the rate of the ith splicing step (for $0 \leq i < N$), and β_N parameterizes the mature RNA's degradation rate. For simplicity, we assume that the β_i are all distinct. Denote the state of this system using the vector $\mathbf{x} := (x_0, x_1, ..., x_N) \in \mathbb{N}^N$. If we use $\boldsymbol{\epsilon}_j$ to denote the vector with a 1 in the jth place and zeros elsewhere, we can write the CME for this system as

$$
\begin{aligned}
\frac{\partial P(\mathbf{x}, t)}{\partial t} =\ & \alpha \left[P(\mathbf{x} - \boldsymbol{\epsilon}_0, t) - P(\mathbf{x}, t) \right] \\
& + \sum_{j=0}^{N-1} \beta_j \left[(x_j + 1) P(\mathbf{x} + \boldsymbol{\epsilon}_j - \boldsymbol{\epsilon}_{j+1}, t) - x_j P(\mathbf{x}, t) \right] \\
& + \beta_N \left[(x_N + 1) P(\mathbf{x} + \boldsymbol{\epsilon}_N, t) - x_N P(\mathbf{x}, t) \right] \ .
\end{aligned}
\tag{6.158}
$$

As before, we will reframe the problem in terms of a generating function in a certain Hilbert space, where the dynamics are completely determined by a Hamiltonian operator. In terms

of the Grassberger-Scheunert [49, 12] creation operators ($\hat{a}_i^+ := \hat{\pi}_i - 1$), the Hamiltonian operator $\hat{H}_{ms}$ of the multistep splicing problem reads

$$\hat{H}_{ms} := \alpha\,\hat{a}_0^+ + \beta_0\left(\hat{a}_1^+ - \hat{a}_0^+\right)\hat{a}_0 + \beta_1\left(\hat{a}_2^+ - \hat{a}_1^+\right)\hat{a}_1 + \cdots + \beta_{N-1}\left(\hat{a}_N^+ - \hat{a}_{N-1}^+\right)\hat{a}_{N-1} - \beta_N\hat{a}_N^+\hat{a}_N$$
$$= \alpha\,\hat{a}_0^+ + \sum_{j=0}^{N-1}\beta_j\left(\hat{a}_{j+1}^+ - \hat{a}_j^+\right)\hat{a}_j - \beta_N\hat{a}_N^+\hat{a}_N\ .$$

$$(6.159)$$

We are seeking operators satisfying the commutation relations (c.f. Eq. 6.38)

$$[\hat{A}, \hat{H}_{ms}] = -\lambda\hat{A}$$
$$[\hat{A}^+, \hat{H}_{ms}] = \lambda\hat{A}^+$$

$$(6.160)$$

where $\lambda > 0$ is some constant with units of inverse time. A reasonable strategy is to compute many commutators by hand, and combine them by trial and error in order to construct the desired operators. Some commutator results that are useful for this purpose are:

$$\begin{aligned}
[\hat{a}_j^+, \hat{H}] &= \beta_j\,\hat{a}_j^+ - \beta_j\,\hat{a}_{j+1}^+ & j = 0, ..., N-1 \\
[\hat{a}_N^+, \hat{H}] &= \beta_N\,\hat{a}_N^+ \\
[\hat{a}_j, \hat{H}] &= -\beta_j\,\hat{a}_j + \beta_{j-1}\,\hat{a}_{j-1} & j = 1, ..., N \\
[\hat{a}_0, \hat{H}] &= \alpha - \beta_0\hat{a}_0 = -\beta_0\left(\hat{a}_0 - \frac{\alpha}{\beta_0}\right)\ .
\end{aligned}$$

$$(6.161)$$

From the above, it is immediately clear that $\hat{a}_N^+$ is an up ladder operator (with constant β_N), and $\hat{a}_0 - (\alpha/\beta_0)$ is a down ladder operator (with constant β_0).

To find the rest, we can note that $\hat{a}_{N-1}^+$ is an up ladder operator up to a correction involving $\hat{a}_N^+$; making that correction allows us to find an up ladder operator (with constant β_{N-1}) that is a linear combination of $\hat{a}_{N-1}^+$ and $\hat{a}_N^+$. Similarly, we can find an up ladder operator (with constant β_{N-2}) that is a linear combination of $\hat{a}_{N-2}^+$, $\hat{a}_{N-1}^+$, and $\hat{a}_N^+$. In general, we have up ladder operators $\hat{A}_i^+$ (for $i = 0, ..., N$), where

$$\hat{A}_N^+ := \hat{a}_N^+$$
$$\hat{A}_{N-1}^+ := \hat{a}_{N-1}^+ + \frac{\beta_{N-1}}{\beta_N - \beta_{N-1}}\,\hat{a}_N^+$$
$$\hat{A}_{N-2}^+ := \hat{a}_{N-2}^+ + \frac{\beta_{N-2}}{\beta_{N-1} - \beta_{N-2}}\,\hat{a}_{N-1}^+ + \frac{\beta_{N-1}\,\beta_{N-2}}{(\beta_{N-1} - \beta_{N-2})(\beta_N - \beta_{N-2})}\,\hat{a}_N^+$$
$$\hat{A}_i^+ := \hat{a}_i^+ + \sum_{j=i+1}^{N}\frac{\beta_i\cdots\beta_{j-1}}{(\beta_j - \beta_i)\cdots(\beta_{i+1} - \beta_i)}\,\hat{a}_j^+ \qquad i = 0, 1, ..., N.$$

$$(6.162)$$

One can apply the same argument to derive the down ladder operators, starting from $\hat{A}_0$

and constructing $\hat{A}_N$ last. We obtain

$$\hat{A}_0 := \hat{a}_0 - \frac{\alpha}{\beta_0}$$

$$\hat{A}_1 := \hat{a}_1 - \frac{\beta_0}{\beta_1 - \beta_0}\ \hat{a}_0 + \frac{\alpha\ \beta_0}{\beta_1(\beta_1 - \beta_0)}$$

$$\hat{A}_2 := \hat{a}_2 - \frac{\beta_1}{\beta_2 - \beta_1}\ \hat{a}_1 + \frac{\beta_0\ \beta_1}{(\beta_2 - \beta_1)(\beta_2 - \beta_0)}\ \hat{a}_0 - \frac{\alpha\ \beta_0\ \beta_1}{\beta_2(\beta_2 - \beta_0)(\beta_2 - \beta_1)} \tag{6.163}$$

$$\hat{A}_i := \hat{a}_i + \frac{(-1)^{i+1}\alpha\ \beta_0 \cdots \beta_{i-1}}{\beta_i(\beta_i - \beta_{i-1}) \cdots (\beta_i - \beta_0)} + \sum_{j=0}^{i-1}(-1)^{i-j}\frac{\beta_j \cdots \beta_{i-1}}{(\beta_i - \beta_{i-1}) \cdots (\beta_i - \beta_j)}\ \hat{a}_j\ .$$

These ladder operators satisfy the important commutation properties:

$$[\hat{A}_j, \hat{H}_{ms}] = -\beta_j\ \hat{A}_j$$
$$[\hat{A}_j^+, \hat{H}_{ms}] = \beta_j\ \hat{A}_j^+ \tag{6.164}$$
$$[\hat{A}_i, \hat{A}_j^+] = \delta_{i,j}$$

for all $i, j = 0, 1, ..., N$. We can write the Hamiltonian in terms of them as

$$\hat{H}_{ms} = -\beta_0\ \hat{A}_0^+\hat{A}_0 - \beta_1\ \hat{A}_1^+\hat{A}_1 - \cdots - \beta_N\ \hat{A}_N^+\hat{A}_N\ , \tag{6.165}$$

which looks like the Hamiltonian of $(N+1)$ uncoupled harmonic oscillators or non-interacting bosons. The energies of this Hamiltonian are

$$E_{n_0,n_1,...,n_N} = \beta_0\ n_0 + \beta_1\ n_1 + \cdots + \beta_N\ n_N \tag{6.166}$$

with each $n_j \in \mathbb{N}$. Using ladder operators or other methods, it is clear that the ground state $|0\rangle$ (which has energy $E_{0,...,0} = 0$) corresponds to the steady state probability distribution

$$\text{Poiss}(\mathbf{x}, \boldsymbol{\mu}) := \frac{\boldsymbol{\mu}^{\mathbf{x}}e^{-\boldsymbol{\mu}\cdot\mathbf{1}}}{\mathbf{x}!} = \frac{\mu_0^{x_0}e^{-\mu_0}}{x_0!} \cdots \frac{\mu_N^{x_N}e^{-\mu_N}}{x_N!} \tag{6.167}$$

where the components of the vector $\boldsymbol{\mu}$ are $\mu_i = \alpha/\beta_i$ for all $0 \le i \le N$.

In what follows, it will be helpful to consider simple cases in order to get intuition for the more general case. The 'toy' case of one splicing step ($N = 1$, two distinct species) has ladder operators

$$\hat{A}_0 := \hat{a}_0 - \frac{\alpha}{\beta_0}$$

$$\hat{A}_1 := \hat{a}_1 - \frac{\beta_0}{\beta_1 - \beta_0}\ \hat{a}_0 + \frac{\alpha\ \beta_0}{\beta_1(\beta_1 - \beta_0)}$$

$$\hat{A}_0^+ := \hat{a}_0^+ + \frac{\beta_0}{\beta_1 - \beta_0}\ \hat{a}_1^+ \tag{6.168}$$

$$\hat{A}_1^+ := \hat{a}_1^+\ .$$

Properties like the canonical commutation relations and ladder operator decomposition of the Hamiltonian are straightforward to verify in this case. The case of two splicing steps ($N = 2$, three distinct species) has ladder operators

$$\hat{A}_0 := \hat{a}_0 - \frac{\alpha}{\beta_0}$$

$$\hat{A}_1 := \hat{a}_1 - \frac{\beta_0}{\beta_1 - \beta_0} \, \hat{a}_0 + \frac{\alpha \, \beta_0}{\beta_1(\beta_1 - \beta_0)}$$

$$\hat{A}_2 := \hat{a}_2 - \frac{\beta_1}{\beta_2 - \beta_1} \, \hat{a}_1 + \frac{\beta_0 \, \beta_1}{(\beta_2 - \beta_1)(\beta_2 - \beta_0)} \, \hat{a}_0 - \frac{\alpha \, \beta_0 \, \beta_1}{\beta_2(\beta_2 - \beta_0)(\beta_2 - \beta_1)} \tag{6.169}$$

$$\hat{A}_0^+ := \hat{a}_0^+ + \frac{\beta_0}{\beta_1 - \beta_0} \, \hat{a}_1^+ + \frac{\beta_0 \, \beta_1}{(\beta_1 - \beta_0)(\beta_2 - \beta_0)} \, \hat{a}_2^+$$

$$\hat{A}_1^+ := \hat{a}_1^+ + \frac{\beta_1}{\beta_2 - \beta_1} \, \hat{a}_2^+$$

$$\hat{A}_2^+ := \hat{a}_2^+ \ .$$

The reader may notice that, in these ladder operators, there are no analogues of the $\sqrt{\mu}$ factors that appear in Eq. 6.36. Those factors appeared earlier in order to make $\hat{A}$ and $\hat{A}^+$ Hermitian conjugates with respect to a certain inner product. Because we will only solve for the steady state probability distribution, such an inner product, along with guaranteeing the up and down ladder operators are Hermitian conjugates of one another, is not necessary. Furthermore, it is not clear what the 'correct' choice of inner product is in this case, because choosing one directly analogous to Eq. 6.44 (but using Eq. 6.167 instead of a one-dimensional Poisson distribution) does not seem to work.

6.7.2 Diagrammatic approach to exact solution

We will now use our ladder operators to reframe the coupled problem as in Sec. 6.3.3, and proceed with a diagrammatic approach to the exact solution as in 6.4. The Hamiltonian of the coupled problems reads

$$\hat{H} = -s\hat{B}^+\hat{B} + \hat{\alpha} \, \hat{a}_0^+ + \sum_{j=0}^{N-1} \beta_j \left(\hat{a}_{j+1}^+ - \hat{a}_j^+ \right) \hat{a}_j - \beta_N \hat{a}_N^+ \hat{a}_N$$

$$= -s\hat{B}^+\hat{B} - \sum_{j=0}^{N} \beta_j \hat{A}_j^+ \hat{A}_j + (\Delta\alpha) \, \hat{a}_0^+ \left[B + c_3 B^+ + c_4 B^+ B \right] \ . \tag{6.170}$$

All we need to do in order to proceed is to express $\hat{a}_0^+$ in terms of our ladder operators. In the case of one splicing step, this is easy:

$$\hat{a}_0^+ = \hat{A}_0^+ - \frac{\beta_0}{\beta_1 - \beta_0} \hat{A}_1^+ \tag{6.171}$$

In case of two splicing steps:

$$\hat{a}_0^+ = \hat{A}_0^+ - \frac{\beta_0}{\beta_1 - \beta_0}\hat{A}_1^+ + \frac{\beta_0 \beta_1}{(\beta_2 - \beta_0)(\beta_2 - \beta_1)}\hat{A}_2^+ \ . \tag{6.172}$$

In general, for N splicing steps, we have

$$\hat{a}_0^+ = \hat{A}_0^+ + \sum_{j=1}^{N}(-1)^j \frac{\beta_0 \cdots \beta_{j-1}}{(\beta_j - \beta_0)\cdots(\beta_j - \beta_{j-1})}\ \hat{A}_j^+ \ . \tag{6.173}$$

To ease notation, we can write

$$\hat{a}_0^+ = \sum_{j=0}^{N} q_j \hat{A}_j^+ \tag{6.174}$$

and define the vector $\mathbf{q} = (q_0, ..., q_N)$ via

$$
\begin{aligned}
q_0 &= 1 \\
q_1 &= -\frac{\beta_0}{\beta_1 - \beta_0} \\
q_2 &= \frac{\beta_0 \beta_1}{(\beta_2 - \beta_0)(\beta_2 - \beta_1)} \\
q_j &= (-1)^j \frac{\beta_0 \cdots \beta_{j-1}}{(\beta_j - \beta_0)\cdots(\beta_j - \beta_{j-1})}\ (\ 1 \le j \le N\ )\ .
\end{aligned}
\tag{6.175}
$$

Hence, we have

$$
\begin{aligned}
\hat{H} &= -s\hat{B}^+\hat{B} - \sum_{j=0}^{N}\beta_j \hat{A}_j^+\hat{A}_j + (\Delta\alpha)\sum_{j=0}^{N} q_j \hat{A}_j^+ \left[B + c_3 B^+ + c_4 B^+ B\right] \\
&= \hat{H}_s + \hat{H}_{ms} + \hat{H}_{int}
\end{aligned}
\tag{6.176}
$$

where the interaction Hamiltonian $\hat{H}_{int}$ is defined to be

$$\hat{H}_{int} = (\Delta\alpha)\sum_{j=0}^{N} q_j \hat{A}_j^+ \left[B + c_3 B^+ + c_4 B^+ B\right]\ . \tag{6.177}$$

As before, we will work in terms of the naive eigenstates $|\mathbf{n}; g\rangle = |n_0, n_1, ..., n_N; g\rangle$. The action of our ladder operators on these states is easily computed. We seek to calculate the 'ground state', which we will write as

$$|\bar{0}\rangle := |0 + \cdots + 0\rangle = |\mathbf{0}; 0\rangle + \sum_{\mathbf{n}\neq\mathbf{0}}\sum_{g=0,1} \mathbf{q}^{\mathbf{n}}\sqrt{\mathbf{n}!}\ Q(\mathbf{n}, g)\ |\mathbf{n}; g\rangle \tag{6.178}$$

for some coefficients $Q(\mathbf{n}, g)$. Start by observing that

$$\hat{H}\,|\mathbf{0}; 0\rangle = (\Delta\alpha)c_3 \sum_{j=0}^{N} q_j\,|\boldsymbol{\epsilon}_j; 1\rangle\ . \tag{6.179}$$

Next,

$$
\begin{aligned}
\hat{H} \ket{\boldsymbol{\epsilon}_i; 1} = &- (s + \beta_i) \ket{\boldsymbol{\epsilon}_i; 1} + (\Delta\alpha) q_i \sqrt{2} \left[\ket{2\boldsymbol{\epsilon}_i; 0} + c_4 \ket{2\boldsymbol{\epsilon}_i; 1} \right] \\
&+ (\Delta\alpha) \sum_{j \neq i} q_j \left[\ket{\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_j; 0} + c_4 \ket{\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_j; 1} \right] .
\end{aligned}
\tag{6.180}
$$

This means that, to second order in $(\Delta\alpha)$, the vacuum state is

$$
\begin{aligned}
\ket{\bar{0}} \approx \ &\ket{\mathbf{0}; 0} + (\Delta\alpha) \sum_{i=0}^{N} q_i \frac{T_{10}}{s + \beta_i} \ket{\boldsymbol{\epsilon}_i; 1} \\
&+ (\Delta\alpha)^2 \sqrt{2} \sum_{i=0}^{N} (q_i)^2 \left[\frac{T_{01} T_{10}}{2\beta_i (s + \beta_i)} \ket{2\boldsymbol{\epsilon}_i; 0} + \frac{T_{11} T_{10}}{(s + 2\beta_i)(s + \beta_i)} \ket{2\boldsymbol{\epsilon}_i; 1} \right] \\
&+ (\Delta\alpha)^2 \sum_{i \neq j} q_i q_j \left[\frac{T_{01} T_{10} \ket{\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_j; 0}}{(\beta_j + \beta_i)(s + \beta_i)} + \frac{T_{11} T_{10} \ket{\boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_j; 1}}{(s + \beta_j + \beta_i)(s + \beta_i)} \right] .
\end{aligned}
\tag{6.181}
$$

The qualitative difference between this problem and the one-dimensional problem we solved earlier, at least as far as this diagrammatic approach goes, is quickly becoming clear. Earlier, we had to think about paths through gene eigenstate space (e.g. one diagram might correspond to $0 \to 1 \to 0$) when writing down our solution, but the 'path' through RNA eigenstate space was simple. In the case of the vacuum state, it went $0 \to 1 \to 2 \to \cdots$, with each increase corresponding to another application of the up ladder operator $\hat{A}^+$.

But in this case, each time we increment the RNA eigenstate number vector, we have $(N+1)$ choices of up ladder operator. For example, among the second order terms there is one that corresponds to first applying $\hat{A}_0^+$, and then applying $\hat{A}_1^+$. The order of traversal matters, because this is distinct from the term whose path involves first applying $\hat{A}_1^+$, and then $\hat{A}_0^+$—as one can see from the Feynman diagram denominators.

Continuing this procedure, we find that

$$
\ket{\bar{0}} = \ket{\mathbf{0}; 0} + \sum_{k=1}^{\infty} \sum_{\text{paths } \mathbf{j}} (\Delta\alpha)^k \mathbf{q^n} \sqrt{\mathbf{n}!} \sum_{i_1, \dots, i_k} \frac{T_{i_k i_{k-1}}}{[\beta_{j_1} + \cdots + \beta_{j_k} + i_k s]} \cdots \frac{T_{i_1 0} \ket{\mathbf{n}; i_k}}{[\beta_{j_1} + i_1 s]}
\tag{6.182}
$$

where the sum over paths should be understood as follows. For each $k \geq 1$, we sum over all paths $\mathbf{j}$ of length k with elements in $\{0, 1, ..., N\}$, i.e. $\mathbf{j} := (j_1, ..., j_k) \in \{0, 1, ..., N\}^k$. The vector $\mathbf{n} := (n_0, n_1, ..., n_N)$ counts the number of times that each integer i appears in a path, i.e.

$$
n_r = \sum_{i \text{ s.t. } j_i = r} 1
\tag{6.183}
$$

for all $0 \leq r \leq N$. The Hilbert space Feynman rules we obtain are almost the same as before, up to this change involving considering paths through RNA eigenstate space.

Hilbert space Feynman rules (multistep version)

In order to compute the coefficient of $|\mathbf{n}; g\rangle$ in the infinite series expansion of $|\bar{0}\rangle$ (c.f. Eq. 6.182), draw all valid diagrams going from 0 to $g \in \{0, 1\}$ in $k = n_0 + \cdots n_N$ steps according to the following rules, and add the numbers corresponding to each diagram.

1. **Set up grid**: Write out positions 0 through k from left to right. Draw two parallel horizontal lines above these labels to denote the 0 and 1 gene eigenstates. The diagram will consist of $k+1$ vertices, each located at a horizontal $\{0, ..., k\}$ position and vertical gene eigenstate (lower or upper) position, and lines connecting those vertices.

2. **Draw lines**: Place the first vertex at horizontal position 0 and on the bottom row. Fill in the following positions from left to right. There are three possible moves: (i) if at 0, you must next go to 1; (ii) if at 1, you can go to 0 next; (iii) if at 1, you can stay at 1. If $g = 0$, the last vertex must be on the bottom row. If $g = 1$, the last vertex must be on the top row.

3. **Numerical factors**: Associate each move/line with a numerical factor. In particular, associate the move from position $m - 1$ to position m with the factor:

   - $0 \to 1$ *flip*: $\dfrac{c_3}{s + \beta_{j_1} + \cdots + \beta_{j_m}}$
   - $1 \to 1$ *stay*: $\dfrac{c_4}{s + \beta_{j_1} + \cdots + \beta_{j_m}}$
   - $1 \to 0$ *flip*: $\dfrac{1}{\beta_{j_1} + \cdots + \beta_{j_m}}$

4. **Sum over RNA paths**: Write down all possible paths on $\{0, 1, ..., N\}^k$ of length k in which each integer i appears exactly n_i times. Evaluate your diagram for all of these paths (by multiplying the numbers associated with each line together) and sum the contributions due to each path.

5. **Tack on generic factors**: Multiply in the generic factors $(\alpha)^k \mathbf{q^n} \sqrt{\mathbf{n!}}$ to get the number corresponding to the diagram you drew.

6.7.3   Steady state probability distribution

Let us determine the steady state probability distribution corresponding to Eq. 6.182. As before, we can either invoke the Euclidean product or manually pick out coefficients in molecule number space. For the states $|\mathbf{x}; S\rangle$ (where $\mathbf{x}$ represents molecule number and $S \in \{1, 2\}$ represents gene state, rather than gene eigenstate), from which the naive eigenkets $|\mathbf{n}; g\rangle$ can be constructed, the Euclidean product can be defined via

$$\langle \mathbf{x}_1; S_1 | \mathbf{x}_2; S_2 \rangle_{Eu} := \delta_{\mathbf{x}_1, \mathbf{x}_2} \delta_{S_1, S_2} \tag{6.184}$$

so that

$$\langle \mathbf{x}; S | \mathbf{n}; g \rangle_{Eu} = \frac{1}{\sqrt{\mathbf{n}!}} \, V_\mathbf{n}(\mathbf{x}, \boldsymbol{\mu}) \text{Poiss}(\mathbf{x}, \boldsymbol{\mu}) \, (\vec{v}_g)_S \qquad (6.185)$$

where $V_\mathbf{n}(\mathbf{x}, \boldsymbol{\mu})$ denotes a novel generalization of the Charlier polynomials that we define and discuss in Appendix 6.D. Using this, we have

$$\frac{P_{ss}(\mathbf{x}, \vec{S})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} = \vec{v}_0 + \sum_{k=1}^{\infty} \sum_{\text{paths } \mathbf{j}} (\Delta\alpha)^k \, \mathbf{q}^\mathbf{n} V_\mathbf{n}(\mathbf{x}, \boldsymbol{\mu}) \sum_{i_1, \dots, i_k} \frac{T_{i_k i_{k-1}}}{[\beta_{j_1} + \cdots + \beta_{j_k} + i_k s]} \cdots \frac{T_{i_1 0} \, \vec{v}_{i_k}}{[\beta_{j_1} + i_1 s]} \, .$$
$$(6.186)$$

Marginalizing over gene state,

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} = 1 + \sum_{k=2}^{\infty} \sum_{\text{paths } \mathbf{j}} (\Delta\alpha)^k \, \mathbf{q}^\mathbf{n} V_\mathbf{n}(\mathbf{x}, \boldsymbol{\mu}) \sum_{i_1, \dots, i_{k-1}} \frac{T_{0 i_{k-1}}}{[\beta_{j_1} + \cdots + \beta_{j_k}]} \cdots \frac{T_{i_1 0}}{[\beta_{j_1} + i_1 s]} \, .$$
$$(6.187)$$

See Fig. 6.9 for some representative distributions in the $N = 1$ case, and how they compare to numerical results obtained by finite state projection [44, 45]. It is interesting to note that this result is the same as our one species solution (c.f. Eq. 6.114) up to the following replacements:

$$\sum_{k=2}^{\infty} \to \sum_{k=2}^{\infty} \sum_{\text{paths } \mathbf{j}}$$
$$d^k \to (\Delta\alpha)^k \, (q_0)^{n_0} \cdots (q_N)^{n_N}$$
$$k\gamma \to \beta_{j_1} + \cdots + \beta_{j_k} \qquad (6.188)$$
$$\text{Poiss}(x, \mu) \to \text{Poiss}(\mathbf{x}, \boldsymbol{\mu})$$
$$C_k(x, \mu) \to V_{n_0, \dots, n_N}(\mathbf{x}, \boldsymbol{\mu}) \, .$$

We can write down our last set of Feynman rules for computing this steady state probability distribution directly.

Molecule number space Feynman rules (multistep version)

In order to compute the terms of order $(\Delta\alpha)^k$ in the infinite series expansion of $P_{ss}(\mathbf{x})/\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})$ (c.f. Eq. 6.187), draw all valid diagrams going from 0 to 0 in k steps according to the following rules, and add the numbers corresponding to each diagram.

1. **Set up grid**: Write out positions 0 through k from left to right. Draw two parallel horizontal lines above these labels to denote the 0 and 1 gene eigenstates. The diagram will consist of $k+1$ vertices, each located at a horizontal $\{0, ..., k\}$ position and vertical gene eigenstate (lower or upper) position, and lines connecting those vertices.

2. **Draw lines**: Place the first vertex at horizontal position 0 and on the bottom row. Fill in the following positions from left to right. There are three possible moves: (i) if at 0, you must next go to 1; (ii) if at 1, you can go to 0 next; (iii) if at 1, you can stay at 1. If $g = 0$, the last vertex must be on the bottom row. If $g = 1$, the last vertex must be on the top row.

3. **Numerical factors**: Associate each move/line with a numerical factor. In particular, associate the move from position $m - 1$ to position m with the factor:

   - $0 \to 1$ *flip*: $\dfrac{c_3}{s + \beta_{j_1} + \cdots + \beta_{j_m}}$
   - $1 \to 1$ *stay*: $\dfrac{c_4}{s + \beta_{j_1} + \cdots + \beta_{j_m}}$
   - $1 \to 0$ *flip*: $\dfrac{1}{\beta_{j_1} + \cdots + \beta_{j_m}}$

4. **Sum over RNA paths + generic factors**: Write down all possible paths on $\{0, 1, ..., N\}^k$ of length k. Evaluate your diagram for all of these paths (by multiplying the numbers associated with each line together) and sum the contributions due to each path with the weights $(\Delta\alpha)^k \mathbf{q}^{\mathbf{n}} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu})$ where $\mathbf{n} = (n_0, ..., n_N)$ counts the number of times each integer i appears in a given path.

6.7.4 Special cases

In this section, we will specialize our result for the steady state probability distribution in various ways, just as we did earlier for the one species problem in Sec. 6.5.

If we want to marginalize over $x_1, ..., x_N$, leaving only the distribution for the number x_0 of nascent RNA, we need the formula

$$\sum_{x_1,...,x_N} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \, \mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) = (\delta_{n_1,0} \cdots \delta_{n_N,0}) \, C_{n_0}(x_0, \mu_0) \, \mathrm{Poiss}(x_0, \mu_0) \,. \tag{6.189}$$

Only the terms with $n_1 = \cdots = n_N = 0$ survive, converting the remaining polynomials to Charlier polynomials and eliminating any paths other than the ones with all zeros. We end up with

$$\frac{P_{ss}(x_0)}{\mathrm{Poiss}(x_0, \mu_0)} = 1 + \sum_{k=2}^{\infty} (\Delta\alpha)^k \, C_{n_0}(x_0, \mu_0) \sum_{i_1,...,i_{k-1}} \frac{T_{0i_{k-1}}}{(k\beta_0)} \cdots \frac{T_{i_10}}{(\beta_0 + i_1 s)} \tag{6.190}$$

i.e. the result is the same as what we found in the one species problem (c.f. Eq. 6.114). This makes sense, because the dynamics downstream of the nascent RNA species X_0 should have no impact on its distribution. More generally, if we were to marginalize over $x_{m+1}, ..., x_N$, leaving only $x_0, ..., x_m$, the distribution would match the distribution for the multistep problem with m splicing steps.
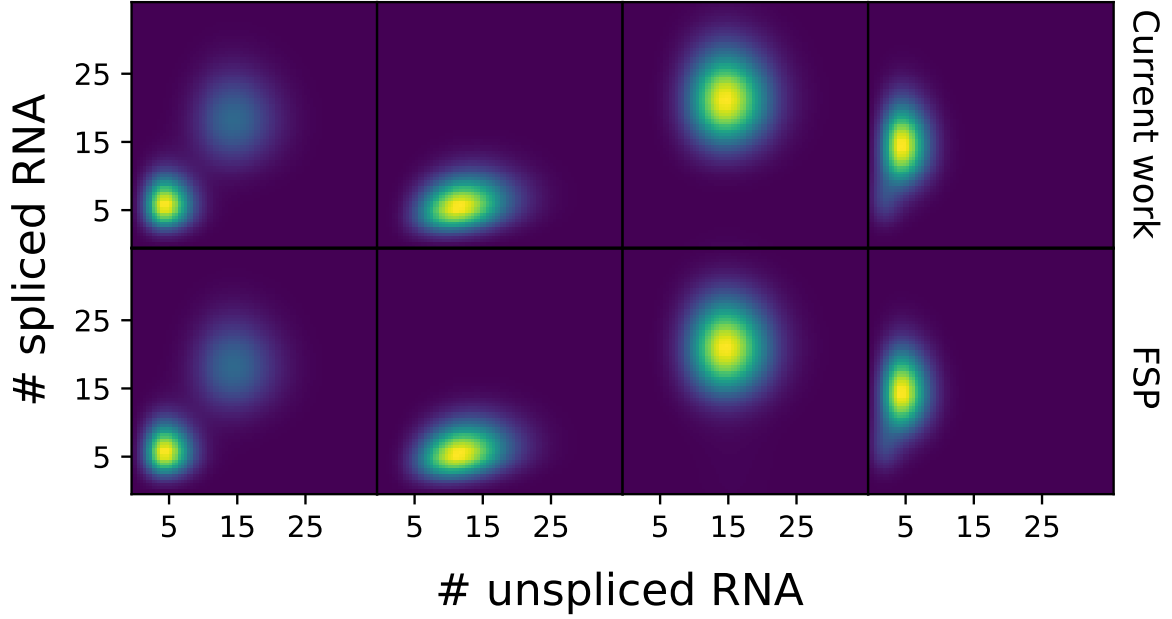
189

Figure 6.9: Comparison of the 2 species/1 splicing step (i.e. $N = 1$) result described by Eq. 6.187 against finite state projection (FSP) in a variety of parameter conditions. Unspliced RNA corresponds to x_0, while spliced RNA corresponds to x_1. In these cases, our formula matches orthogonal numerical results quite well.

Meanwhile, if we want to marginalize over $x_0, ..., x_{N-1}$, leaving only the distribution for the number x_N of mature/fully processed RNA, we need the formula

$$\sum_{x_0,...,x_{N-1}} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \, \text{Poiss}(\mathbf{x}, \boldsymbol{\mu}) = C_{n_N}(x_N, \mu_N) \, \text{Poiss}(x_N, \mu_N) \prod_{k=0}^{N} \left[v_N^{(k)} \right]^{n_k} . \tag{6.191}$$

Since all of the $v_N^{(k)}$ are nonzero, no paths vanish this time. We have

$$\frac{P_{ss}(x_N)}{\text{Poiss}(x_N, \mu_N)} = 1 + \sum_{k=2}^{\infty} \sum_{\text{paths } \mathbf{j}} (\Delta \alpha)^k \, \mathbf{q}^{\mathbf{n}} C_{n_N}(x_N, \mu_N) \times$$
$$\times \prod_{k=0}^{N} \left[v_N^{(k)} \right]^{n_k} \sum_{i_1,...,i_{k-1}} \frac{T_{0 i_{k-1}}}{[\beta_{j_1} + \cdots + \beta_{j_k}]} \cdots \frac{T_{i_1 0}}{[\beta_{j_1} + i_1 s]} . \tag{6.192}$$

Qualitatively, this formula indicates that the distribution of the mature RNA species X_N *is* affected by all of the preceding splicing steps—including how many steps there were, and the associated rates. This observation is less trivial than it at first seems, since the marginal steady state distribution of X_N for a constitutively active gene (which is a Poisson distribution) is completely independent of upstream splicing dynamics.

Let us move on to special parameter regimes of the full joint distribution. In the case of very unequal switching rates ($k_{21} \gg k_{12}$), we can argue as in Sec. 6.5.3 that the 'balance beam' Feynman diagrams contribute the most to first order in the switching rate ratio $r := k_{12}/k_{21}$. The end result is analogous:

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} = 1 + \sum_{k=2}^{\infty} \sum_{\text{paths } \mathbf{j}} (\Delta\alpha)^k \, \mathbf{q^n} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \times$$
$$\times \frac{r}{\left(\beta_{j_1} + \cdots + \beta_{j_k}\right)\left(\beta_{j_1} + \cdots + \beta_{j_{k-1}} + s\right)\cdots\left(\beta_{j_1} + s\right)} \; . \tag{6.193}$$

If switching is much faster than all of the β_i (for $0 \le i \le N$), we can again write an asymptotic series in $1/s$:

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} \approx 1 + \frac{1}{s}(\Delta\alpha)^2 c_3 \sum_{i,j} \frac{\mathbf{q^n} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu})}{\beta_i + \beta_j} + \mathcal{O}\left(\frac{1}{s^2}\right) \tag{6.194}$$

where $\mathbf{n} := \boldsymbol{\epsilon}_i + \boldsymbol{\epsilon}_j$. Finally, if switching is much slower than all of the β_i, we can approximate the Feynman denominators as $(\beta_{j_1} + \cdots \beta_{j_m} + i_m s) \approx (\beta_{j_1} + \cdots \beta_{j_m})$ and use the combinatorial result[8] that

$$\sum_{\text{paths } \mathbf{j} \, , \, \mathbf{n} \text{ fixed}} \frac{1}{\left(\beta_{j_1} + \cdots + \beta_{j_k}\right)\cdots\left(\beta_{j_1} + \beta_{j_2}\right)\left(\beta_{j_1}\right)} = \frac{1}{\mathbf{n}!} \frac{1}{\beta_0^{n_0} \cdots \beta_N^{n_N}} \tag{6.195}$$

where we consider only paths of length k with the same sum vector $\mathbf{n}$. Then our steady state probability becomes

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} = 1 + \sum_{\mathbf{n}} \frac{1}{\mathbf{n}!} \prod_{j=0}^{N} \left[\frac{(\Delta\alpha)q_j}{\beta_j}T\right]^{n_j} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \; . \tag{6.196}$$

Using the formula for the generating function of these orthogonal polynomials (c.f. Eq. 6.246 in Appendix 6.D), we can sum this in closed form to obtain

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} \overset{00}{\approx} \prod_{j=0}^{N} \left[1 + \frac{(\Delta\alpha)}{\mu_j}T \sum_k v_j^{(k)} \frac{q_k}{\beta_k}\right]^{x_j} e^{-(\Delta\alpha)T \sum_k v_j^{(k)} \frac{q_k}{\beta_k}} \; . \tag{6.197}$$

Using a straightforward generalization of the argument from Sec. 6.5.5, we can show that this interesting-looking formula is exactly equal to a Poisson mixture, i.e. that

$$P_{ss}(\mathbf{x}) \approx \frac{k_{12}}{s}\text{Poiss}(\mathbf{x}, \boldsymbol{\mu}_1) + \frac{k_{21}}{s}\text{Poiss}(\mathbf{x}, \boldsymbol{\mu}_2) \tag{6.198}$$

in this limit.

---

[8]This can be proved by induction.

6.8 Numerical implementation and validation

In this section, we explain how to implement the formulas presented in the previous sections efficiently, and discuss typical numerical behavior.

The most important thing to note is that, while our results are formally correct (see e.g. Appendix 6.E for a mathematical proof of the agreement between our birth-death-switching result and the previously known result), they are likely not very useful for numerically implementing the solution to these problems in large swaths of parameter space. In particular, they are frequently numerically unstable, with the more general formulas presented in the previous section somewhat more unstable than the birth-death-switching formulas.

This is essentially due to two reasons. The first reason is that our solutions were constructed as perturbative solutions in powers of $\Delta\alpha$; when this is not small—as in the case in most real situations—our implementations behave particularly poorly (Fig. 6.10a) because one must compute a very large number of series terms in order to closely approximate the correct answer. One is also likely to encounter overflow errors in taking large powers of $\Delta\alpha$.

One might wonder whether our formulas are asymptotic expansions of the true result, which would mean they formally diverge. This is not so. For example, in the birth-death-switching case we have that $C_n \sim (x/\mu)^n$, and that the Feynman diagram sums go like $1/(\gamma^n n!)$. Very roughly, we have

$$\frac{P_{ss}(x)}{\text{Poiss}(x, \mu)} \sim \sum_{n=0}^{\infty} \left[\frac{\Delta\alpha \cdot x}{\gamma \cdot \mu}\right]^n \frac{1}{n!} = \exp\left[\frac{\Delta\alpha \cdot x}{\alpha_{eff}}\right] . \tag{6.199}$$

The analysis is similar in the multistep case, where we instead obtain

$$\frac{P_{ss}(\mathbf{x})}{\text{Poiss}(\mathbf{x}, \boldsymbol{\mu})} \sim \exp\left[\sum_{j=0}^{N} \frac{\Delta\alpha \cdot q_j x_j}{\alpha_{eff}}\right] . \tag{6.200}$$

The second reason is that additional instability arises due to the proliferation of time scales (i.e. the β_i) in the multistep problem. When the β_i are all comparable, all of the Feynman diagrams contribute, leading to a large number of terms being required to converge to the correct answer.

Despite these instabilities, the formulas are still usable if their regime of validity is kept in mind. We have found that computing the analytic generating function (Eq. 6.7 and Eq. 6.21), and then inverse fast Fourier transforming, tends to produce more stable results than using the state space formulas directly. This has the added advantage that one can bypass numerically computing the orthogonal polynomials.

We explore runtime performance and stability in quantitative fashion for the birth-death-switching solution in Fig. 6.10. Because the birth-death-switching problem has been solved before (in particular in [40], by Huang et al.), we can assess the accuracy of our solution formula (Eq. 6.6) by comparing it to their hypergeometric solution formula. By sampling physiologically plausible parameters and modulating the approximation order, we can compute the precision and computational complexity of the solution; the details of the sampling procedure are described in Appendix 6.F.

The diagrammatic solution exhibits substantially weaker state space size dependence than the hypergeometric solution (Fig. 6.10a), with nearly constant time complexity compared to the $\mathcal{O}(N^{1.00})$ complexity of the Huang solution in the relevant domain. The runtime dependence on order of approximation is approximately linear (Fig. 6.10b). In the low-$\Delta\alpha$ region, the diagrammatic solution recapitulates the analytical solution quite well, as quantified using the Kullback-Leibler and Kolmogorov-Smirnov divergence measures (Fig. 6.10c-d). Therefore, the algorithm performance suggests natural applications to the design of adaptive solvers, with the diagrammatic solution used as much as possible in the low-$\Delta\alpha$ regime. Finally, we note that the slow-switching Poisson mixture approximation given in Eq. 6.198 generally underperforms the diagrammatic solution.

But, as we might have anticipated, the discrepancy between the 'ground truth' Huang result and our solution formula increases sharply as $\Delta\alpha$ gets large. This is clear from the large Kullback-Leibler divergence (Fig. 6.10c) and Kolmogorov-Smirnov distance (Fig. 6.10d) values in this regime.

In conclusion, the solution formulas have proven useful tools for examining the qualitative behavior of the models in various regimes (e.g. as we did in Sec. 6.5 and Sec. 6.7.4), and can be used to obtain reliable theoretical results for observables like moments. But the series themselves are numerically poorly conditioned, and can be somewhat slow in the multistep case (where one must sum over many 'paths'), and so are not appropriate for tasks like parameter inference on RNA counts data. In the second part of this two part article [1], we present an alternative approach that has better numerical properties.
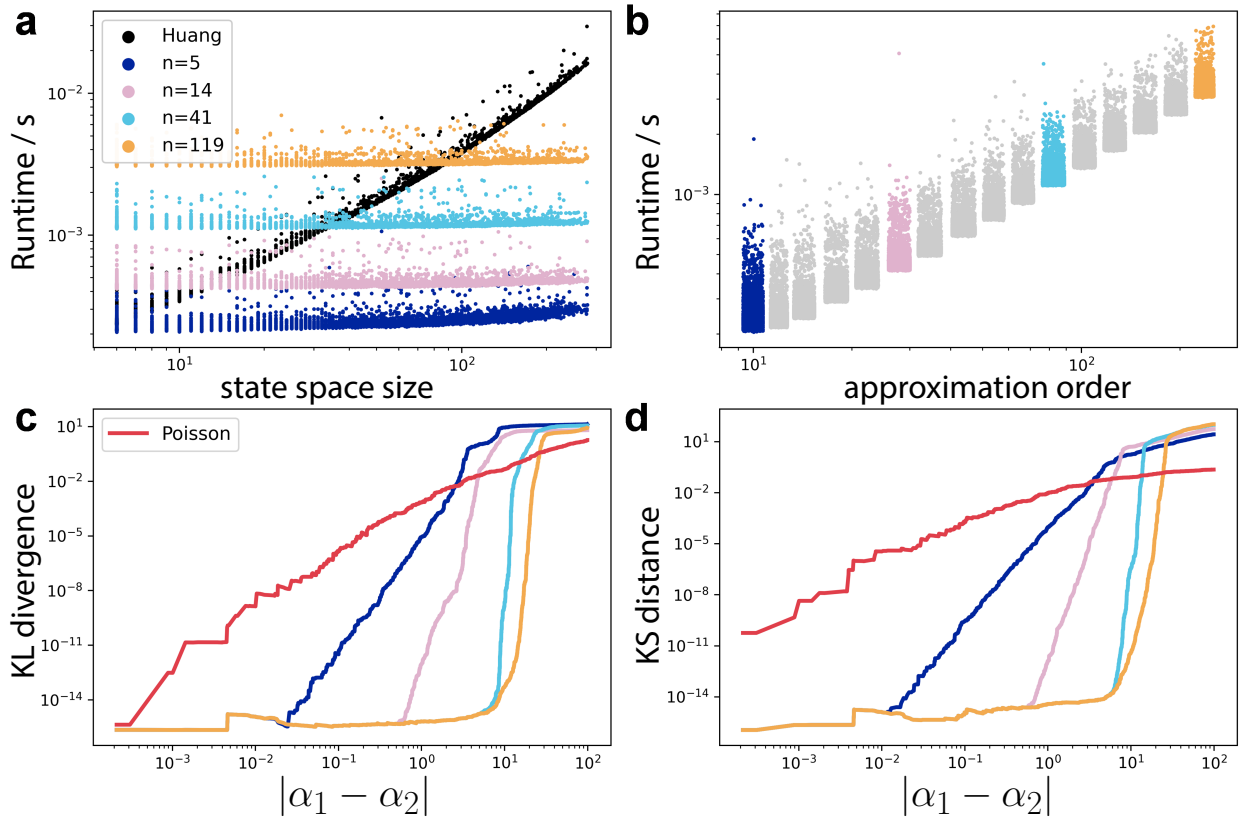
Figure 6.10: Exploration of birth-death-switching solution performance and stability. For many randomly sampled parameter sets, we evaluated the effective state space size, runtime, and accuracy. Each dot corresponds to a different parameter set. See Appendix 6.F for details on the parameter sampling procedure and definition of state space size. (a) Runtime comparison between the ground truth hypergeometric solution from Huang et al. [40] and the diagrammatic estimates as a function of state space size, computing only the first n terms of Eq. 6.6. (b) Runtime comparison for the diagrammatic estimates as a function of approximation order. (c) Cumulative 98th percentile of the Kullback-Leibler divergence between estimates and ground truth as a function of $|\Delta\alpha|$ (colors correspond to colors in (b); red represents trivial Poisson solution). (d) Cumulative 98th percentile of the Kolmogorov-Smirnov divergence between estimates and ground truth as a function of $|\Delta\alpha|$ (colors correspond to colors in (b)). The last two panels confirm that increasing $\Delta\alpha$ makes convergence much slower, which in practice could mean inaccurate numerical results when computing Eq. 6.6 to fixed order.

6.9 Discussion

Using a novel theoretical methodology inspired by quantum mechanics, we have solved several complicated problems in stochastic chemical kinetics exactly: the chemical birth-death process coupled to a switching gene, the chemical birth-death process with additive noise coupled to a switching gene (the continuous limit of the former model), and a more realistic model of transcription involving a switching gene and an arbitrary number of downstream splicing steps. We also uncovered tantalizing formal connections between chemical kinetics and quantum physics: the dynamics of a system involving RNA production and degradation coupled to a switching gene looks much like the dynamics of a system involving non-interacting bosons coupled to a fermion, for example. Unfortunately, the instability of these solution formulas renders them inadequate for practical purposes like parameter inference. We will address this deficiency further in part II [1].

There are various potential directions for generalizing this approach, some of which we have pointed out earlier in the paper. Most obviously, the idea of expressing a system's Hamiltonian in terms of 'natural' ladder operators, and rewriting it in terms of 'free' Hamiltonians and an interaction term, is broadly applicable—whether or not a switching gene is involved. In the context of stochastic chemical kinetics and gene regulation, birth-death-like processes like the ones we have studied here may be an appropriate choice of free Hamiltonian, with the interaction terms corresponding to dynamics we currently do not know how to solve exactly. Notably, in most cases we cannot exactly solve dynamics involving molecular binding/unbinding [13]. One expects nonlinear operator products like $\hat{A}_1 \hat{A}_2 \hat{A}_3^+$ to appear in the interaction term in such cases.

Closer in spirit to our current results, one expects the ability to generalize to (i) more than two gene states, and (ii) more realistic transcription and splicing dynamics. Multistate models have been solved in at least one special case [55], but it is expected that such a generalization would be highly nontrivial. It is not immediately obvious how to generalize to more than two gene states while maintaining nice anti-commutation properties for the switching-associated ladder operators; we offer some reflection on this problem towards the end of Appendix 6.B. On the other hand, generalizing to more complicated models of transcription and splicing, involving branching splicing topologies and even alternative splicing, seems more straightforward. The main difficulty would be understanding the representation theory of this more general problem, which in some sense means reckoning with the representation theory of so-called monomolecular reaction networks [11, 12] (the natural multi-species generalization of the birth-death process).

One unexpected strength of our method is that the problem it treats is fairly abstract. Although we took our main CME (Eq. 6.11) to represent RNA dynamics coupled to a switching gene, almost the exact same equation can be used to describe a more granular model of transcription involving RNA polymerase molecules that hop sequentially along a gene [56]. Also, as pointed out at the end of Sec. 6.6, the continuous limit of our model corresponds exactly to a model of active matter run-and-tumble motion [54]. This suggests our techniques may be useful for solving a broad range of stochastic physical and biological

problems.

Lastly, we must emphasize the surprising applicability of quantum mechanics-like operator-based approaches, and the natural appearance of objects that resemble Feynman diagrams. There are probably many more results (in both stochastic gene regulation and studies of active matter dynamics) that can be derived by exploiting useful parallels between quantum mechanics and stochastic chemical reaction dynamics, as several pieces of past work have suggested [51, 57, 50, 58, 59].

Acknowledgments

Authors' contributions

J.J.V. and G.G. conceived of the work, studied the numerical properties of the analytic solutions, and wrote the manuscript. J.J.V. came up with the solution method and worked out the associated mathematics. L.P. and W.R.H. reviewed and edited the manuscript.

6.A Charlier polynomials

In this appendix, we will define the Charlier polynomials and describe their properties. It is useful to view them as a discrete analogue to the more familiar Hermite polynomials; in particular, the Charlier polynomials relate to the Poisson distribution the same way that Hermite polynomials relate to the normal distribution. Moreover, they reduce to the Hermite polynomials in the same limit that the Poisson distribution reduces to a normal distribution.

Given a parameter $\mu \geq 0$ and a nonnegative integer n, define the nth Charlier polynomial $C_n(x, \mu)$ for all $x \geq 0$ via the Rodrigues formula

$$C_n(x, \mu) := \frac{1}{\mathrm{Poiss}(x, \mu)} \left[-\nabla \right]^n \mathrm{Poiss}(x, \mu) \tag{6.201}$$

where we recall that

$$\mathrm{Poiss}(x, \mu) := \frac{\mu^x e^{-\mu}}{x!} \ , \tag{6.202}$$

and where ∇ is the backward difference operator that acts on a discrete-valued function f according to

$$\nabla f(x) := f(x) - f(x-1) \ . \tag{6.203}$$

For example, to obtain $C_1(x, \mu)$, we compute

$$C_1(x, \mu) = \frac{\mathrm{Poiss}(x-1, \mu) - \mathrm{Poiss}(x, \mu)}{\mathrm{Poiss}(x, \mu)} = \frac{\frac{x}{\mu}\mathrm{Poiss}(x, \mu) - \mathrm{Poiss}(x, \mu)}{\mathrm{Poiss}(x, \mu)} = \frac{x}{\mu} - 1 \ . \tag{6.204}$$

From the point of view of our ladder operator solution of the chemical birth-death process, this definition of $C_n(x, \mu)$ arises from taking

$$\langle x | n \rangle = \langle x | \frac{(\hat{A}^+)^n}{\sqrt{n!}} \ | 0 \rangle = \sqrt{\frac{\mu^n}{n!}} \langle x | (\hat{\pi} - 1)^n \ | 0 \rangle \tag{6.205}$$

and defining

$$C_n(x, \mu) := \langle x | (\hat{\pi} - 1)^n \ | 0 \rangle \ . \tag{6.206}$$

In our convention, the first few Charlier polynomials are

$$\begin{aligned}
C_0(x, \mu) &= 1 \\
C_1(x, \mu) &= \frac{x}{\mu} - 1 \\
C_2(x, \mu) &= \frac{x^2}{\mu^2} - \left[\frac{1}{\mu^2} + \frac{2}{\mu} \right] x + 1 \\
C_3(x, \mu) &= \frac{x^3}{\mu^3} - \frac{3}{\mu^2} \left[1 + \frac{1}{\mu} \right] x^2 + \frac{1}{\mu} \left[3 + \frac{3}{\mu} + \frac{2}{\mu^2} \right] x - 1 \ .
\end{aligned} \tag{6.207}$$

In general, the nth Charlier polynomial will be an nth order polynomial in x which asymptotically goes like $(x/\mu)^n$.
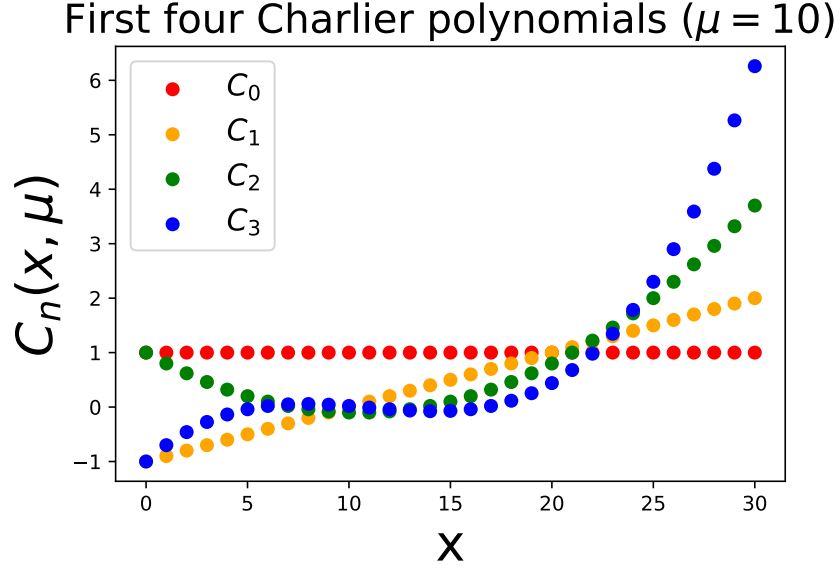
First four Charlier polynomials ($\mu = 10$)

Figure 6.11: The first few Charlier polynomials (C_0, C_1, C_2, C_3) for $\mu = 10$. For x somewhat larger than μ, we can see that $C_n \sim (x/\mu)^n$.

Two results are particularly useful for deriving various properties of the Charlier polynomials. One is its generating function $G_C(t, x)$ (not to be confused with probability generating function seen elsewhere in this paper), defined via

$$G_C(t, x) := \sum_{n=0}^{\infty} \frac{C_n(x, \mu)}{n!} \, t^n \; . \tag{6.208}$$

Using the Rodrigues formula (Eq. 6.201), we can show that

$$
\begin{aligned}
G_C(t, x) &= \frac{1}{\text{Poiss}(x, \mu)} \sum_{n=0}^{\infty} \frac{[-\nabla \, t]^n}{n!} \, \text{Poiss}(x, \mu) \\
&= \frac{1}{\text{Poiss}(x, \mu)} e^{-\nabla \, t} \, \text{Poiss}(x, \mu) \\
&= \frac{1}{\text{Poiss}(x, \mu)} e^{-t} \sum_{n=0}^{x} \frac{\text{Poiss}(x - n, \mu) \, t^n}{n!} \\
&= \frac{1}{\text{Poiss}(x, \mu)} e^{-t} \left(1 + \frac{t}{\mu}\right)^x \text{Poiss}(x, \mu) \\
&= \left(1 + \frac{t}{\mu}\right)^x e^{-t} \; .
\end{aligned}
\tag{6.209}
$$

Another useful result comes from a novel integral representation of the Poisson distribution[9]

$$\text{Poiss}(x, \mu) = \int_{-\infty}^{\infty} \frac{dz}{2\pi} \frac{e^{-i\mu z}}{(1 - iz)^{x+1}} \ . \tag{6.210}$$

Applying the Rodrigues formula to it, we can derive the integral representation

$$C_n(x, \mu) = \frac{1}{\text{Poiss}(x, \mu)} \int_{-\infty}^{\infty} \frac{dz}{2\pi} \frac{(-iz)^n e^{-i\mu z}}{(1 - iz)^{x+1}} \tag{6.211}$$

which, incidentally, offers another method for explicitly computing $G_C(t, x)$.

One property of the Charlier polynomials that we will use is that

$$\sum_{x=0}^{\infty} C_n(x, \mu) g^x \frac{\mu^x e^{-\mu}}{x!} = (g - 1)^n e^{\mu(g-1)} \ , \tag{6.212}$$

a result that can be derived using our expression for $G_C(t, x)$. To do so, note that

$$\begin{aligned}
\sum_{n=0}^{\infty} \sum_{x=0}^{\infty} C_n(x, \mu) g^x \frac{\mu^x e^{-\mu}}{x!} \frac{t^n}{n!} &= \sum_{x=0}^{\infty} g^x \frac{(\mu + t)^x}{x!} e^{-t-\mu} \\
&= \exp\left\{ g\left(\mu + t\right) - t - \mu \right\} \\
&= e^{\mu(g-1)} \sum_{n=0}^{\infty} \frac{(g-1)^n}{n!} t^n
\end{aligned} \tag{6.213}$$

and match the coefficients of t^n. This property, along with our formula for $G_C(t, x)$, can be used to derive what is probably the most important property of the Charlier polynomials: that they are orthogonal with respect to a Poisson weight function, i.e.

$$\sum_{x=0}^{\infty} C_n(x, \mu) C_m(x, \mu) \text{Poiss}(x, \mu) = \frac{n!}{\mu^n} \delta_{n,m} \ . \tag{6.214}$$

This orthogonality relation offers an alternative way to view the resolution of the identity described by Eq. 6.52.

Other properties that are useful to establish include that the Charlier polynomials obey a recurrence relation in x, and that they reduce to Hermite polynomials in the large μ limit. To establish the recurrence relation, it is easiest to solve the birth-death CME (Eq. 6.29) via separation of variables. Given that we already know the allowed energies E_n (c.f. Eq. 6.40) and steady state distribution (c.f. Eq. 6.39), we can substitute the ansatz

$$P(x, t) = C_n(x, \mu) \text{Poiss}(x, \mu) e^{-\gamma n t} \tag{6.215}$$

into the time-dependent CME to derive (after simplifying)

$$\begin{aligned}
-n C_n(x, \mu) &= \mu \left[C_n(x + 1, \mu) - C_n(x, \mu) \right] \\
&\quad + x \left[C_n(x - 1, \mu) - C_n(x, \mu) \right] \ .
\end{aligned} \tag{6.216}$$

---

[9]See Gradshteyn and Ryzhik [60] (ET I 118(3), in section 3.382, on pg. 365).

In addition to offering another way to define the Charlier polynomials, this recurrence relation also offers one numerically efficient way to determine their values. Meanwhile, the large μ behavior of the Charlier polynomials is described by

$$
\begin{aligned}
(\sqrt{\mu})^n \, C_n(x, \mu) &\xrightarrow{\mu \gg 1} \left(\frac{1}{\sqrt{2}} \right)^n H_n \left(\frac{x - \mu}{\sqrt{2\mu}} \right) \\
G_C(\sqrt{\mu} \, t, x) &\xrightarrow{\mu \gg 1} G_H \left(\frac{t}{\sqrt{2}}, \frac{x - \mu}{\sqrt{2\mu}} \right) \ ,
\end{aligned}
\tag{6.217}
$$

where $G_H(t, y)$ denotes the generating function of the (physicists') Hermite polynomials $H_n(x)$, which reads

$$
G_H(t, y) := \sum_{n=0}^{\infty} \frac{H_n(y)}{n!} \, t^n = e^{2yt - t^2} \ .
\tag{6.218}
$$

This can be established by straightforwardly approximating $G_C(\sqrt{\mu} \, t, x)$. For large μ, we have that

$$
\begin{aligned}
G_C(\sqrt{\mu} \, t, x) &= e^{-\sqrt{\mu} t} \left(1 + \frac{t}{\sqrt{\mu}} \right)^x \\
&= e^{-\sqrt{\mu} t} e^{x \log \left[1 + \frac{t}{\sqrt{\mu}} \right]} \\
&\approx e^{-\sqrt{\mu} t} e^{x \left[\frac{t}{\sqrt{\mu}} - \frac{t^2}{2\mu} \right]}
\end{aligned}
\tag{6.219}
$$

where we have expanded the logarithm to second order. Approximating $x \approx \mu$ in the t^2 term, we have

$$
\begin{aligned}
e^{-\sqrt{\mu} t} e^{x \left[\frac{t}{\sqrt{\mu}} - \frac{t^2}{2\mu} \right]} &\approx e^{-\sqrt{\mu} t} e^{\frac{xt}{\sqrt{\mu}} - \frac{t^2}{2}} \\
&= \exp \left\{ \frac{x - \mu}{\sqrt{\mu}} t - \frac{t^2}{2} \right\} \\
&= \exp \left\{ 2 \frac{x - \mu}{\sqrt{2\mu}} \left(\frac{t}{\sqrt{2}} \right) - \left(\frac{t}{\sqrt{2}} \right)^2 \right\} \\
&= G_H \left(\frac{t}{\sqrt{2}}, \frac{x - \mu}{\sqrt{2\mu}} \right)
\end{aligned}
\tag{6.220}
$$

as desired. The reader may notice that the approximations that were necessary here are the same as the ones required to show that a Poisson distribution reduces to a normal distribution in the large μ limit, i.e. it is necessary in both cases to truncate the Taylor series of a logarithm and assume that $x \approx \mu$ in part of the formula.

6.B   More on switching gene representation theory

Although the appearance of fermionic ladder operators in the solution of pure gene switching dynamics may at first seem mysterious, there is a systematic approach to finding these operators and deriving their properties. In this appendix, we discuss this approach, and point towards how it might be generalized for problems involving more than two gene states.

The key insight is that the matrix[10] H_s is infinitesimal stochastic, which means its columns sum to zero. This forces one of the eigenvalues of H_s to be zero, and the other (assuming $H_s \neq 0$) to be strictly negative. Because its two eigenvalues are distinct, it is diagonalizable, and can be written in the form

$$H_s = -s \, QDQ^{-1} \tag{6.221}$$

where we are using $-s$ to denote the negative eigenvalue of H_s, and where D is the diagonal matrix

$$D := \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} . \tag{6.222}$$

Trivially, we can observe that

$$D = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \tilde{B}^+ \tilde{B} \tag{6.223}$$

where we have defined the matrices $\tilde{B}$ and $\tilde{B}^+$. Various properties of $\tilde{B}$ and $\tilde{B}^+$, including their commutation relations, are obvious:

$$
\begin{aligned}
\tilde{B} e_0 &= 0 & \tilde{B} e_1 &= e_0 \\
\tilde{B}^+ e_0 &= e_1 & \tilde{B}^+ e_1 &= 0 \\
\{\tilde{B}, \tilde{B}\} &= \tilde{B}^2 + \tilde{B}^2 = 0 \\
\{\tilde{B}^+, \tilde{B}^+\} &= (\tilde{B}^+)^2 + (\tilde{B}^+)^2 = 0 \\
\{\tilde{B}, \tilde{B}^+\} &= \tilde{B}\tilde{B}^+ + \tilde{B}^+ \tilde{B} = I
\end{aligned}
\tag{6.224}
$$

where $e_0 := (1,0)^T$ and $e_1 := (0,1)^T$. These anticommutation properties are identical to those satisfied by B and B^+, and the action of $\tilde{B}$ and $\tilde{B}^+$ on e_0 and e_1 looks like the action of B and B^+ on the eigenvectors of H_s (c.f. Eq. 6.74 and Eq. 6.75). This is no accident; noting that

$$H_s = -s \, Q\tilde{B}^+ \tilde{B} Q^{-1} = -s \, (Q\tilde{B}^+Q^{-1})(Q\tilde{B}Q^{-1}) , \tag{6.225}$$

we can define $B := Q\tilde{B}Q^{-1}$ and $B^+ := Q\tilde{B}^+Q^{-1}$, i.e. B and B^+ are just a similarity transformation away from $\tilde{B}$ and $\tilde{B}^+$. Various properties of B and B^+ follow trivially from the properties of $\tilde{B}$ and $\tilde{B}^+$ listed in Eq. 6.224. For example,

$$Q\tilde{B}Q^{-1}Qe_1 = Qe_0 \implies Bv_1 = v_0 , \tag{6.226}$$

---

[10]In this appendix, we will relax our convention of putting hats on matrices and arrows on vectors.

since Q constitutes a change of basis, e.g. $Qe_0 = v_0$.

A 'natural' weight matrix W can be motivated by this idea that our system is simple up to a change of basis. To force the eigenvectors v_i of H_s to be orthogonal, we can note that the standard basis vectors e_i are orthogonal, and that Q changes from one basis set to the other. In particular,

$$\delta_{i,j} = e_i^T e_j = (Q^{-1}v_i)^T(Q^{-1}v_j) = v_i^T(Q^{-T}Q^{-1})v_j \ , \tag{6.227}$$

so we should define W to be the symmetric matrix

$$W := Q^{-T}Q^{-1} \ . \tag{6.228}$$

It is easy to see that B and B^+ are Hermitian conjugates with respect to the inner product induced by this weight matrix, e.g.

$$WB = Q^{-T}Q^{-1}Q\tilde{B}Q^{-1} = Q^{-T}\tilde{B}Q^TQ^{-T}Q^{-1} = Q^{-T}(\tilde{B}^+)^TQ^TQ^{-T}Q^{-1} = (B^+)^TW \ . \tag{6.229}$$

In principle, it is straightforward to extend this idea to represent gene switching dynamics involving more than two gene states. For example, for a three state gene switching problem we might generically expect H_s to have three distinct eigenvalues: zero, and two distinct strictly negative eigenvalues (because, as before, H_s is infinitesimal stochastic). In that case, H_s can be diagonalized so that

$$
\begin{aligned}
H_s &= Q \begin{pmatrix} 0 & 0 & 0 \\ 0 & -\lambda_1 & 0 \\ 0 & 0 & -\lambda_2 \end{pmatrix} Q^{-1} \\
&= -\lambda_1 \, Q \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^{-1} - \lambda_2 \, Q \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} Q^{-1} \\
&= -\lambda_1 \, Q \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^{-1} - \lambda_2 \, Q \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} Q^{-1} \\
&= -\lambda_1 \, Q\tilde{B}_1^+\tilde{B}_1Q^{-1} - \lambda_2 \, Q\tilde{B}_2^+\tilde{B}_2Q^{-1}
\end{aligned}
\tag{6.230}
$$

and we obtain matrices $\tilde{B}_i$ and $\tilde{B}_i^+$ (for $i = 1, 2$) with similar properties as before. If we use δ_{ij} to denote the matrix with a 1 in the (i, j)th place and zeros elsewhere, we can write

$$
\begin{aligned}
\tilde{B}_1 &:= \delta_{01} & \tilde{B}_1^+ &:= \delta_{10} \\
\tilde{B}_2 &:= \delta_{02} & \tilde{B}_2^+ &:= \delta_{20} \ ,
\end{aligned}
\tag{6.231}
$$

and note that the generalization to $(N + 1)$ gene states is clear.

If constructed in this way, one should be careful to note that the two sets anticommute with themselves, but not necessarily with *each other*. For example, while $\tilde{B}_1\tilde{B}_1^+ + \tilde{B}_1^+\tilde{B}_1 = I$, we have

$$\tilde{B}_1\tilde{B}_2^+ + \tilde{B}_2^+\tilde{B}_1 = \delta_{01}\delta_{20} + \delta_{20}\delta_{01} = \delta_{21} \neq 0 \tag{6.232}$$

which is naively not what we would expect from 'independent' fermionic operators. Using the same similarity transformation idea as before, we can define matrices

$$B_i := Q\tilde{B}_i Q^{-1}$$
$$B_i^+ := Q\tilde{B}_i^+ Q^{-1}$$

(6.233)

that inherit the properties of the $\tilde{B}_i$ and $\tilde{B}_i^+$, so that

$$H_s = -\lambda_1 \, B_1^+ B_1 - \lambda_2 \, B_2^+ B_2 \, ,$$

(6.234)

which looks like a decomposition of H_s in terms of two 'independent' fermions. Just like the set $\{I, B, B^+, B^+B\}$ formed a basis for the set of all 2×2 matrices, the set

$$\{I, B_1, B_2, B_1^+, B_2^+, B_1^+ B_1, B_2^+ B_2, B_2^+ B_1, B_1^+ B_2\}$$

(6.235)

forms a basis for the set of all 3×3 matrices. Supposing there are $(N + 1)$ gene states and that H_s has distinct eigenvalues, we could write

$$H_s = -\lambda_1 \, B_1^+ B_1 - \cdots - \lambda_N \, B_N^+ B_N$$

(6.236)

and note that a basis for the space of all $(N + 1) \times (N + 1)$ matrices can be constructed using B_i, B_i^+, and $B_i^+ B_j$ (for all i, j), for a total of $N^2 + 2N + 1 = (N + 1)^2$ basis elements.

6.C Tables of low order Feynman diagrams
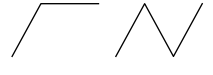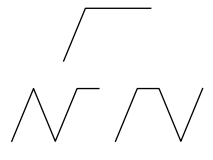
Coeff.	Values	Relevant diagrams
$q_{1,0}^0$	0	no diagrams
$q_{1,1}^0$	$\dfrac{c_3}{s+\gamma}$	
$q_{2,0}^0$	$\dfrac{c_3}{(s+\gamma)2\gamma}$	
$q_{2,1}^0$	$\dfrac{c_3 c_4}{(s+\gamma)(s+2\gamma)}$	
$q_{3,0}^0$	$\dfrac{c_3 c_4}{(s+\gamma)(s+2\gamma)3\gamma}$	
$q_{3,1}^0$	$\dfrac{c_3 c_4^2}{(s+\gamma)(s+2\gamma)(s+3\gamma)} + \dfrac{c_3^2}{(s+\gamma)2\gamma(s+3\gamma)}$	
$q_{4,0}^0$	$\dfrac{c_3 c_4^2}{(s+\gamma)(s+2\gamma)(s+3\gamma)4\gamma} + \dfrac{c_3^2}{(s+\gamma)2\gamma(s+3\gamma)4\gamma}$	
$q_{4,1}^0$	$\dfrac{c_3 c_4^3}{(s+\gamma)(s+2\gamma)(s+3\gamma)(s+4\gamma)}$ $+\dfrac{c_3^2 c_4}{(s+\gamma)2\gamma(s+3\gamma)(s+4\gamma)} + \dfrac{c_3^2 c_4}{(s+\gamma)(s+2\gamma)3\gamma(s+4\gamma)}$	

Table 6.2: The first few coefficients $q_{k,g}^0$ and the corresponding diagrams.
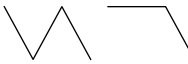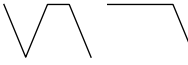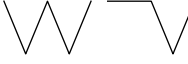
Coeff.	Values	Relevant diagrams
$q^1_{1,0}$	$\dfrac{1}{\gamma - s}$	
$q^1_{1,1}$	$\dfrac{c_4}{\gamma}$	
$q^1_{2,0}$	$\dfrac{c_4}{\gamma(2\gamma - s)}$	
$q^1_{2,1}$	$\dfrac{c_3}{(\gamma - s)2\gamma} + \dfrac{c_4^2}{\gamma \cdot 2\gamma}$	
$q^1_{3,0}$	$\dfrac{c_3}{(\gamma - s)2\gamma(3\gamma - s)} + \dfrac{c_4^2}{\gamma \cdot 2\gamma(3\gamma - s)}$	
$q^1_{3,1}$	$\dfrac{c_3 c_4}{(\gamma - s)2\gamma \cdot 3\gamma} + \dfrac{c_3 c_4}{\gamma(2\gamma - s)3\gamma} + \dfrac{c_4^3}{\gamma \cdot 2\gamma \cdot 3\gamma}$	
$q^1_{4,0}$	$\dfrac{c_3 c_4}{\gamma(2\gamma - s)3\gamma(4\gamma - s)}$ $+\dfrac{c_3 c_4}{(\gamma - s)2\gamma \cdot 3\gamma(4\gamma - s)} + \dfrac{c_4^3}{\gamma \cdot 2\gamma \cdot 3\gamma(4\gamma - s)}$	
$q^1_{4,1}$	$\dfrac{c_4^4}{\gamma \cdot 2\gamma \cdot 3\gamma \cdot 4\gamma} + \dfrac{c_3 c_4^2}{(\gamma - s)2\gamma \cdot 3\gamma \cdot 4\gamma}$ $+\dfrac{c_3 c_4^2}{\gamma(2\gamma - s)3\gamma \cdot 4\gamma}$ $+\dfrac{c_3^2}{(\gamma - s)2\gamma(3\gamma - s)4\gamma} + \dfrac{c_3 c_4^2}{\gamma \cdot 2\gamma(3\gamma - s)4\gamma}$	

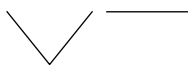Table 6.3: The first few coefficients $q^1_{k,g}$ and the corresponding diagrams.

6.D Multistep orthogonal polynomials

In this appendix, we define a new family of orthogonal polynomials that appear in the solution of the multistep splicing dynamics problem, and discuss their properties. They generalize the Charlier polynomials in a natural way, and many Charlier polynomial properties have analogues in this more general context.

Fix a natural number $N \geq 0$ and mean parameter $\boldsymbol{\mu} := (\mu_0, ..., \mu_N) \in \mathbb{R}^{N+1}$. These polynomials will be functions of $\mathbf{x} := (x_0, ..., x_N) \in \mathbb{N}^{N+1}$. Denote the polynomial associated with the integer vector $\mathbf{n} := (n_0, n_1, ..., n_N) \in \mathbb{N}^{N+1}$ by $V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu})$. We assume we have 'ladder operators' made up of linear combinations of finite difference operators

$$\hat{L}_i^+ := \sum_{j=0}^{N} v_j^{(i)} \left(-\nabla_j \right) \qquad\qquad i = 0, 1, ..., N \qquad (6.237)$$

where $\mathbf{v}^{(i)}$ contains the coefficients of the ith ladder operator, e.g.

$$\hat{A}_0^+ = \hat{a}_0^+ + \frac{\beta_0}{\beta_1 - \beta_0} \, \hat{a}_1^+ + \frac{\beta_0 \, \beta_1}{(\beta_1 - \beta_0)(\beta_2 - \beta_0)} \, \hat{a}_2^+$$

$$\implies \mathbf{v}^{(0)} = \left(1 \quad \frac{\beta_0}{\beta_1 - \beta_0} \quad \frac{\beta_0 \, \beta_1}{(\beta_1 - \beta_0)(\beta_2 - \beta_0)} \right)^T , \qquad (6.238)$$

and where the backward difference operator ∇_i acts on a function $f(\mathbf{x})$ according to

$$\nabla_i f(\mathbf{x}) := f(\mathbf{x}) - f(x_0, ..., x_i - 1, ..., x_N) . \qquad (6.239)$$

The simplest way to define these polynomials is via the Rodrigues formula

$$V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) := \frac{1}{\mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu})} \left[\prod_{k=0}^{N} \left(\hat{L}_k^+ \right)^{n_k} \right] \mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) . \qquad (6.240)$$

This definition exactly corresponds to the fact that the eigenstates of the multistep splicing problem can be constructed by repeatedly applying the up ladder operators:

$$|\mathbf{n}\rangle = \frac{(\hat{A}_0^+)^{n_0}}{\sqrt{n_0!}} \cdots \frac{(\hat{A}_N^+)^{n_N}}{\sqrt{n_N!}} |\mathbf{0}\rangle \qquad (6.241)$$

where we have used a mostly arbitrary normalization convention. In the $N = 0$ case, these polynomials are just the Charlier polynomials. The first few polynomials in the $N = 1$ case

are given by

$$V_{00} = 1$$

$$V_{0,1} = \frac{x_1}{\mu_1} - 1$$

$$V_{1,0} = \frac{x_0}{\mu_0} + \frac{\beta_0}{\beta_1 - \beta_0} \frac{x_1}{\mu_1} - \frac{\beta_1}{\beta_1 - \beta_0}$$

$$V_{0,2} = \frac{x_1^2}{\mu_1^2} - \left[\frac{1}{\mu_1^2} + \frac{2}{\mu_1}\right] x_1 + 1$$

$$V_{1,1} = \frac{x_0 x_1}{\mu_0 \mu_1} + \frac{\beta_0}{\beta_1 - \beta_0} \frac{x_1(x_1 - 1)}{\mu_1^2} - \frac{x_0}{\mu_0} - \frac{(\beta_1 + \beta_0)}{\beta_1 - \beta_0} \frac{x_1}{\mu_1} + \frac{\beta_1}{\beta_1 - \beta_0}$$

$$V_{2,0} = \frac{x_0(x_0 - 1)}{\mu_0^2} + \left(\frac{\beta_0}{\beta_1 - \beta_0}\right)^2 \frac{x_1(x_1 - 1)}{\mu_1^2} + \frac{2\beta_0}{\beta_1 - \beta_0} \frac{x_0}{\mu_0} \frac{x_1}{\mu_1}$$

$$- \frac{2\beta_1}{\beta_1 - \beta_0} \frac{x_0}{\mu_0} - \frac{2\beta_0\beta_1}{(\beta_1 - \beta_0)^2} \frac{x_1}{\mu_1} + \left(\frac{\beta_1}{\beta_1 - \beta_0}\right)^2 .$$

$$(6.242)$$

In general, $V_{\mathbf{n}}$ will be a polynomial of order n_i with respect to the variable x_i, and of order $|\mathbf{n}| = n_0 + \cdots + n_N$ overall. The leading term asymptotically goes like

$$\left(\frac{x_0}{\mu_0}\right)^{n_0} \cdots \left(\frac{x_N}{\mu_N}\right)^{n_N} \tag{6.243}$$

but other terms of the same order can appear (e.g. a term that goes like x_1^2 appears in V_{11}). As with the Charlier polynomials, two helpful results are the generating function and a particular integral representation. Noting the representation of a multivariate Poisson distribution via

$$\begin{aligned}
\mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) &= \int \frac{d\mathbf{z}}{(2\pi)^{N+1}} \frac{e^{-i\boldsymbol{\mu}\cdot\mathbf{z}}}{(1 - i\mathbf{z})^{\mathbf{x}+\mathbf{1}}} \\
&= \int \frac{dz_0 \cdots dz_N}{(2\pi)^{N+1}} \frac{e^{-i\mu_0 z_0 - \cdots - i\mu_N z_N}}{(1 - iz_0)^{x_0+1} \cdots (1 - iz_N)^{x_N+1}} ,
\end{aligned} \tag{6.244}$$

we can apply the Rodrigues formula to find the integral representation

$$\begin{aligned}
V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) &= \frac{1}{\mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu})} \int \frac{d\mathbf{z}}{(2\pi)^{N+1}} \frac{\prod_{k=0}^{N}\left[-i\sum_{j=0}^{N} v_j^{(k)} z_j\right]^{n_k} e^{-i\boldsymbol{\mu}\cdot\mathbf{z}}}{(1 - i\mathbf{z})^{\mathbf{x}+\mathbf{1}}} \\
&= \frac{1}{\mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu})} \int \frac{dz_0 \cdots dz_N}{(2\pi)^{N+1}} \frac{\prod_{k=0}^{N}\left[-i\sum_{j=0}^{N} v_j^{(k)} z_j\right]^{n_k} e^{-i\mu_0 z_0 - \cdots - i\mu_N z_N}}{(1 - iz_0)^{x_0+1} \cdots (1 - iz_N)^{x_N+1}} .
\end{aligned} \tag{6.245}$$

This can be used to derive the generating function

$$G(\mathbf{z}) = \sum_{\mathbf{n}} \frac{V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu})}{\mathbf{n}!} \mathbf{z}^{\mathbf{n}}$$

$$= \prod_{j=0}^{N} \left[1 + \frac{1}{\mu_j} \sum_k v_j^{(k)} z_k \right]^{x_j} e^{-\sum_k v_j^{(k)} z_k} \tag{6.246}$$

$$= \prod_{j=0}^{N} \left[1 + \frac{1}{\mu_j} \sum_k v_j^{(k)} z_k \right]^{x_j} e^{-\sum_k v_j^{(k)} z_k} \ .$$

This, in turn, can be used to derive the helpful identity

$$\sum_{\mathbf{x}} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \, \mathbf{g}^{\mathbf{x}} \, \mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) = \prod_{k=0}^{N} \left[\sum_{j=0}^{N} (g_j - 1) v_j^{(k)} \right]^{n_k} e^{\mu_k(g_k - 1)}$$

$$= e^{\boldsymbol{\mu} \cdot (\mathbf{g} - \mathbf{1})} \prod_{k=0}^{N} \left[(\mathbf{g} - \mathbf{1}) \cdot \mathbf{v}^{(k)} \right]^{n_k} \tag{6.247}$$

which can be used both to find the (analytic) generating function in the multistep problem (c.f. Eq. 6.21) and to marginalize over variables by setting different $g_i = 1$. To derive marginal distributions in Sec. 6.7.4, we use this fact to obtain the formulas

$$\sum_{x_1, \ldots, x_N} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \, \mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) = (\delta_{n_1, 0} \cdots \delta_{n_N, 0}) C_{n_0}(x_0, \mu_0) \, \mathrm{Poiss}(x_0, \mu_0) \tag{6.248}$$

and

$$\sum_{x_0, \ldots, x_{N-1}} V_{\mathbf{n}}(\mathbf{x}, \boldsymbol{\mu}) \, \mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) = \prod_{k=0}^{N} \left[v_N^{(k)} \right]^{n_k} C_{n_N}(x_N, \mu_N) \, \mathrm{Poiss}(x_N, \mu_N) \ . \tag{6.249}$$

The formulas we have derived so far can be used to determine various special values of these polynomials. For example, using the generating function (Eq. 6.246) we can find that

$$V_{\mathbf{n}}(\mathbf{0}, \boldsymbol{\mu}) = \left(-\sum_{j=0}^{N} v_j^{(0)} \right)^{n_0} \cdots \left(-\sum_{j=0}^{N} v_j^{(N)} \right)^{n_N} \ . \tag{6.250}$$

Using Eq. 6.247, we can find that

$$V_{0,0,\ldots,0,n_N}(\mathbf{x}, \boldsymbol{\mu}) = C_{n_N}(x_N, \mu_N) \ . \tag{6.251}$$

What about orthogonality? Are $V_{\mathbf{n}}$ and $V_{\mathbf{n}'}$ orthogonal for $\mathbf{n} \neq \mathbf{n}'$? Perhaps surprisingly the answer is no in general; for example, in the $N = 1$ case, V_{10} and V_{01} are not orthogonal with respect to a Poisson weight function:

$$\sum_{x_0, x_1} V_{10}(\mathbf{x}, \boldsymbol{\mu}) V_{01}(\mathbf{x}, \boldsymbol{\mu}) \mathrm{Poiss}(\mathbf{x}, \boldsymbol{\mu}) = \frac{\beta_0}{\beta_1 - \beta_0} \frac{1}{\mu_1} \neq 0 \ . \tag{6.252}$$

It *is* true, however, that $V_{\mathbf{n}}$ and $V_{\mathbf{n}'}$ are orthogonal when $|\mathbf{n}| \neq |\mathbf{n}'|$ (which is not true in the above case, where we have $0 + 1 = 1 + 0$). This can be argued using the generating function (Eq. 6.246).

6.E   Consistency of current results with previous results

Previous work on the birth-death-switching problem (Iyer-Biswas et al. [39], Huang et al. [40], and Cao and Grima [41]) found an exact analytic solution for the steady state generating function $\psi_{ss}(g)$ given by

$$\psi_{ss}(g) = e^{\frac{\alpha_2}{\gamma}(g-1)} \, {}_1F_1\left(\frac{k_{12}}{\gamma}, \frac{s}{\gamma}; \frac{\Delta\alpha}{\gamma}(g-1)\right) = e^{\frac{\alpha_2}{\gamma}(g-1)} \sum_{n=0}^{\infty} \frac{\left[\frac{\Delta\alpha}{\gamma}(g-1)\right]^n}{n!} \frac{\left(\frac{k_{12}}{\gamma}\right)_n}{\left(\frac{s}{\gamma}\right)_n} \quad (6.253)$$

where ${}_1F_1$ denotes the confluent hypergeometric function of the first kind, and $(a)_n := a(a + 1)\cdots(a+n-1)$ denotes the Pochhammer symbol/rising factorial. Naively, this result looks quite different from ours, and does not involve Charlier polynomials or Feynman diagrams at all. We have checked the numerical consistency of our work with previous work elsewhere in this paper; in this appendix, we show that our result is mathematically identical to this one.

First, we recall that our result for the probability generating function is

$$\psi_{ss}(g) = e^{\mu(g-1)} \left\{ 1 + \sum_{k=1}^{\infty} [\Delta\alpha(g-1)]^k \sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} \right\} . \quad (6.254)$$

The Feynman diagram part of this formula can be rewritten in terms of the transfer matrix T and another matrix, which we arbitrarily denote using R:

$$R := \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} . \quad (6.255)$$

Specifically, it is the 00 (upper left) entry of a certain product of 2×2 matrices, i.e.

$$\sum_{i_1,\ldots,i_{k-1}=0,1} \frac{T_{0i_{k-1}}}{(k\gamma)} \cdots \frac{T_{i_1 0}}{(\gamma + i_1 s)} \overset{00}{=} \frac{[kI + rR]\, T \cdots [I + rR]\, T}{\gamma^k \, k! \, (1+r)_k} \quad (6.256)$$

where $r := s/\gamma$ and I is the 2×2 identity matrix. Also note that

$$\mu = \frac{\alpha_1 k_{12} + \alpha_2 k_{21}}{\gamma s} = \frac{\alpha_1 k_{12} - \alpha_2 k_{12} + \alpha_2 k_{12} + \alpha_2 k_{21}}{\gamma s} = \frac{\Delta\alpha}{\gamma} x + \frac{\alpha_2}{\gamma} \quad (6.257)$$

where in this context we define $x := k_{12}/s$. Now we can write

$$\frac{\psi_{ss}(g)}{e^{\frac{\alpha_2}{\gamma}(g-1)}} \overset{00}{=} e^{\frac{\Delta\alpha}{\gamma}x(g-1)} \sum_{k=0}^{\infty} \frac{\left[\frac{\Delta\alpha(g-1)}{\gamma}\right]^k}{k!} \frac{[kI + rR]\,T \cdots [I + rR]\,T}{(1+r)_k} . \tag{6.258}$$

Let us rewrite the right-hand side by multiplying out the two infinite series involved term by term (in powers of $(g-1)$). The kth term d_k of the (Cauchy) product reads

$$
\begin{aligned}
d_k &= \sum_{\ell=0}^{k} \frac{\left[\frac{\Delta\alpha}{\gamma}x(g-1)\right]^{k-\ell}}{(k-\ell)!} \frac{\left[\frac{\Delta\alpha(g-1)}{\gamma}\right]^{\ell}}{\ell!} \frac{[\ell + rR]\,T \cdots [1 + rR]\,T}{(\ell+r)\cdots(1+r)} \\
&= \left[\frac{\Delta\alpha}{\gamma}(g-1)\right]^k \sum_{\ell=0}^{k} \frac{x^{k-\ell}}{(k-\ell)!}\frac{1}{\ell!} \frac{[\ell + rR]\,T \cdots [1 + rR]\,T}{(\ell+r)\cdots(1+r)} \\
&= \frac{\left[\frac{\Delta\alpha}{\gamma}(g-1)\right]^k}{k!} \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \frac{[\ell + rR]\,T \cdots [1 + rR]\,T}{(\ell+r)\cdots(1+r)} .
\end{aligned}
\tag{6.259}
$$

We can simplify this using the surprising identity

$$\frac{rx(rx+1)\cdots(rx+k-1)}{r(r+1)\cdots(r+k-1)} \overset{00}{=} \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \frac{[\ell + rR]\,T \cdots [1 + rR]\,T}{(\ell+r)\cdots(1+r)} . \tag{6.260}$$

Doing so, we have

$$
\begin{aligned}
\psi_{ss}(g) &= e^{\frac{\alpha_2}{\gamma}(g-1)} \sum_{k=0}^{\infty} \frac{\left[\frac{\Delta\alpha}{\gamma}(g-1)\right]^k}{k!} \frac{rx(rx+1)\cdots(rx+k-1)}{r(r+1)\cdots(r+k-1)} \\
&= e^{\frac{\alpha_2}{\gamma}(g-1)}\, {}_1F_1\left(rx, r; \frac{\Delta\alpha}{\gamma}(g-1)\right)
\end{aligned}
\tag{6.261}
$$

which matches the previously published result. The remainder of this appendix is dedicated to proving the unusual identity that allows us to make this simplification.

Lemma. Let $f(\ell)$ and $g(\ell)$ denote the 00 and 10 entries of the ℓ-fold matrix product

$$\frac{[\ell + rR]\,T \cdots [1 + rR]\,T}{(\ell+r)\cdots(1+r)} , \tag{6.262}$$

where

$$
\begin{aligned}
T &:= \begin{pmatrix} 0 & 1 \\ x(1-x) & 1-2x \end{pmatrix} \\
R &:= \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} .
\end{aligned}
\tag{6.263}
$$

Define

$$S_{00}(k) := \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} f(\ell)$$

$$S_{10}(k) := \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} g(\ell) \ .$$

(6.264)

We have that

$$S_{00}(k) := \frac{rx(rx+1)\cdots(rx+k-1)}{r(r+1)\cdots(r+k-1)}$$

$$S_{10}(k) := \frac{k(1-x)}{r+k} S_{00}(k)$$

(6.265)

for all integers $k \geq 0$ (using the convention that the empty product is 1, so that $S_{00}(0) = 1$).

Proof. First, by definition, we note that $f(\ell)$ and $g(\ell)$ satisfy the recurrence relations

$$f(\ell) = g(\ell - 1)$$

$$g(\ell) = \frac{\ell}{\ell + r} [x(1-x)f(\ell-1) + (1-2x)g(\ell-1)]$$

(6.266)

for all $\ell \geq 1$. Next, note that

$$f(0) = 1$$

$$f(1) = 0$$

$$f(2) = \frac{x(1-x)}{r+1}$$

$$g(0) = 0$$

$$g(1) = \frac{x(1-x)}{r+1}$$

$$g(2) = \frac{2x(1-x)(1-2x)}{(r+1)(r+2)} \ .$$

(6.267)

We will proceed by induction. The $k = 0$ case is trivially true. The $k = 1$ case is easy to show:

$$S_{00}(1) = xf(0) + f(1) = x$$

$$S_{10}(1) = xg(0) + g(1) = \frac{x(1-x)}{r+1} \ .$$

(6.268)

Assume our formulas for S_{00} and S_{10} hold up to some k. Now we have

$$S_{00}(k+1) = \sum_{\ell=0}^{k+1} \binom{k+1}{\ell} x^{k+1-\ell} f(\ell)$$

$$= x^{k+1} + f(k+1) + \sum_{\ell=1}^{k} \left[\binom{k}{\ell} + \binom{k}{\ell-1} \right] x^{k+1-\ell} f(\ell)$$

(6.269)

211

where we have used Pascal's identity. Rearranging, this becomes

$$
\begin{aligned}
S_{00}(k+1) &= x \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} f(\ell) + \sum_{\ell=1}^{k+1} \binom{k}{\ell-1} x^{k+1-\ell} f(\ell) \\
&= x S_{00}(k) + \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} f(\ell+1) \\
&= x S_{00}(k) + \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} g(\ell) \\
&= x S_{00}(k) + S_{10}(k) \\
&= \left[x + \frac{k(1-x)}{r+k} \right] S_{00}(k) \\
&= \left[\frac{rx+k}{r+k} \right] S_{00}(k)
\end{aligned}
\tag{6.270}
$$

which is exactly what we wanted. Now we will perform the induction step for S_{10}; it is somewhat more involved. It begins as we argued for S_{00}:

$$
\begin{aligned}
S_{10}(k+1) &= \sum_{\ell=0}^{k+1} \binom{k+1}{\ell} x^{k+1-\ell} g(\ell) \\
&= x^{k+1} + g(k+1) + \sum_{\ell=1}^{k} \left[\binom{k}{\ell} + \binom{k}{\ell-1} \right] x^{k+1-\ell} g(\ell) \\
&= x \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} g(\ell) + \sum_{\ell=1}^{k+1} \binom{k}{\ell-1} x^{k+1-\ell} g(\ell) \\
&= x S_{10}(k) + \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} g(\ell+1)
\end{aligned}
\tag{6.271}
$$

where we have again used Pascal's identity. The sum on the right is somewhat tricky to

evaluate; denote it by S'. Using the recurrence relation relating $g(\ell+1)$ to $f(\ell)$ and $g(\ell)$,

$$S' = \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \frac{\ell+1}{\ell+1+r} [x(1-x)f(\ell) + (1-2x)g(\ell)]$$

$$= \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \left[1 - \frac{r}{\ell+1+r}\right] [x(1-x)f(\ell) + (1-2x)g(\ell)]$$

$$= x(1-x)S_{00}(k) + (1-2x)S_{10}(k)$$

$$- \frac{r}{k+1+r} \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \frac{k+1+r}{\ell+1+r} [x(1-x)f(\ell) + (1-2x)g(\ell)]$$

$$= x(1-x)S_{00}(k) + (1-2x)S_{10}(k)$$

$$- \frac{r}{k+1+r} \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \left[\frac{k+1+r}{\ell+1+r} - 1 + 1\right] [x(1-x)f(\ell) + (1-2x)g(\ell)]$$

$$= x(1-x)S_{00}(k) + (1-2x)S_{10}(k)$$

$$- \frac{r}{k+1+r} \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \left[\frac{k-\ell}{\ell+1}\frac{\ell+1}{\ell+1+r} + 1\right] [x(1-x)f(\ell) + (1-2x)g(\ell)]$$

$$= [x(1-x)S_{00}(k) + (1-2x)S_{10}(k)]\left[1 - \frac{r}{k+1+r}\right]$$

$$- \frac{r}{k+1+r} \sum_{\ell=0}^{k} \binom{k}{\ell} x^{k-\ell} \frac{k-\ell}{\ell+1}\frac{\ell+1}{\ell+1+r} [x(1-x)f(\ell) + (1-2x)g(\ell)]$$

$$= [x(1-x)S_{00}(k) + (1-2x)S_{10}(k)]\left[1 - \frac{r}{k+1+r}\right]$$

$$- \frac{r}{k+1+r} \sum_{\ell=-1}^{k-1} \frac{k!}{(k-\ell)!\,\ell!} x^{k-\ell} \frac{k-\ell}{\ell+1} g(\ell+1)$$

$$= [x(1-x)S_{00}(k) + (1-2x)S_{10}(k)]\left[1 - \frac{r}{k+1+r}\right]$$

$$- \frac{r}{k+1+r} \sum_{\ell=-1}^{k-1} \frac{k!}{(k-\ell-1)!\,(\ell+1)!} x^{k-\ell} \, g(\ell+1)$$

$$= [x(1-x)S_{00}(k) + (1-2x)S_{10}(k)]\left[1 - \frac{r}{k+1+r}\right] - \frac{rx}{k+1+r}S_{10}(k)$$

$$= x(1-x)S_{00}(k) + (1-2x)S_{10}(k) - \frac{r}{k+1+r}(1-x)[xS_{00}(k) + S_{10}(k)] \ .$$

$$\text{(6.272)}$$

Adding all the terms together yields that

$$S_{10}(k+1) = xS_{10}(k) + x(1-x)S_{00}(k) + (1-2x)S_{10}(k) - \frac{r}{r+k+1}(1-x)\left[xS_{00}(k) + S_{10}(k)\right]$$

$$= (1-x)\left[xS_{00}(k) + S_{10}(k)\right] - \frac{r}{r+k+1}(1-x)\left[xS_{00}(k) + S_{10}(k)\right]$$

$$= \frac{k+1}{r+k+1}(1-x)\left[xS_{00}(k) + S_{10}(k)\right]$$

$$= \frac{k+1}{r+k+1}(1-x)\left[x + \frac{k(1-x)}{r+k}\right]S_{00}(k)$$

$$= \frac{k+1}{r+k+1}(1-x)\frac{rx+k}{r+k}S_{00}(k)$$

$$(6.273)$$

which suffices to establish our result. $\square$

6.F   Numerical validation details

The validation of a method for the solution of a given system requires its characterization under three criteria. Firstly, the runtime must be benchmarked against a known solution; we implement the solution due to Huang et al. [40] (specifically, the hypergeometric generating function solution given by Eq. 6.253). Secondly, the bounds on the error between the estimate and the ground truth distribution must be estimated, particularly as a function of the parameter domain. Finally, the switching gene problem necessitates the criterion of nontriviality. Specifically, in parameter regimes with very slow gene state transition rates, the solution is given by a mixture of two Poisson distributions, given by Eq. 6.198. Therefore, it is necessary to confirm that the method outperforms this trivial and computationally facile solution.

We explore the stationary solution of the chemical master equation. Since the time variable is immaterial, the five-parameter system reduces to a four-parameter one (i.e. we have $\alpha_1, \alpha_2, \gamma, k_{12}, k_{21}$). We set the degradation rate γ to 1 with no loss of generality. Physically, the production timescale is assumed to be rather shorter than the degradation timescale, whereas the transition timescale is assumed to be longer. Therefore, we draw the α parameters from a log-uniform distribution on $[10^{-1}, 10^2]$ and the k parameters from a log-uniform distribution on $[10^{-3}, 10^0]$.

Aside from the parameter values, the evaluation of the distributions requires a domain, i.e., a distribution support $\{0, 1, 2, ..., m\}$. One approach simulates the system for N cells, considers the observations $Y_1, ..., Y_N$ collected after equilibration, and uses $m \leftarrow \max_j Y_j$. This method is rather natural, but problematic in practice, because simulation is substantially more computationally expensive than the computation of distributions by any method.

Instead, we adopt a heuristic method, inspired by previous literature in the field [61]. Specifically, given a distribution with mean μ and standard deviation σ, we implement $m \leftarrow \max(5, \mu + 4\sigma)$. Since μ and σ have simple analytical expressions for the two-state switch (see Sec. 6.2.1), this procedure is substantially less computationally intensive. To benchmark this choice of meta-parameters, we selected 1000 parameters and simulated $N = 100$ cells for each. Only 1.7% of m produced by the moment-based estimate were lower than those produced by the simulation-based estimate, mostly restricted to low m (Fig. 6.12). Therefore, we adopt $\max(5, \mu + 4\sigma)$ as a conservative estimate for the state space upper bound.
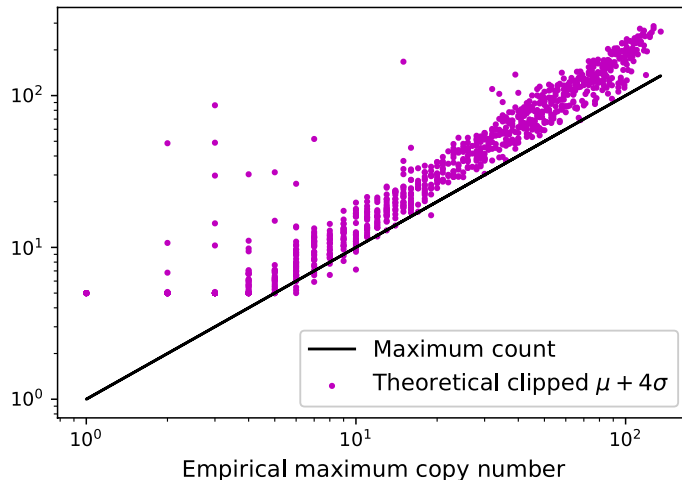

Figure 6.12: Demonstration of the performance of the moment-based threshold against the simulation maximum-based threshold for state space definition.

Thus, given a parameter vector and a state space $\{0, ..., m\}$, we compute the solution due to Huang et al. as the non-trivial ground truth, as well as the Poisson solution as the trivial estimate, and compare them to the diagrammatic solution for various orders of approximation. The ground truth solution requires $m + 1$ evaluations of the hypergeometric function; therefore, we expect the runtime to be a strong function of the state space size. Furthermore, the approximation runtime is a function of the approximation order. We compute and report both dependence trends.

The divergence from the ground truth affords a choice of possible measures. We use the Kullback-Leibler divergence and the Kolmogorov-Smirnov distance, with probabilities fixed to a minimum value of 10^{-16}. However, raw divergences are somewhat uninformative for the purpose of determining robust domains. Given the trend of roughly monotonically increasing divergence as a function of $|\alpha_1 - \alpha_2|$ (as seen in Fig. 6.10), a natural choice reports the maximum of all divergences up to a particular value of $|\alpha_1 - \alpha_2|$. Intuitively, such a function of $|\Delta\alpha|$ reports an estimate for an upper bound on error. For a given approximation order and choice of acceptable error bound, the function immediately provides the upper bound of the $|\Delta\alpha|$ domain that meets this error threshold.

The maximum is not a particularly robust statistic, and is unstable to outliers. Instead, we report the 98th percentile of all divergences up to a particular value of $|\Delta\alpha|$. Therefore, such a trace is an estimate of the 98% confidence interval for the maximum error in the domain. As the sampling density increases, this interval converges to the true one with $O(M^{-1/2})$ in the number of parameters sampled. We use 5000 parameter samples to estimate the intervals.

This procedure lends itself especially easily to quantitative comparison with the trivial Poisson solution. Specifically, if the divergence between ground truth and the diagrammatic estimate is lower than the divergence between ground truth and the Poisson mixture estimate, the diagrammatic estimate quantitatively outperforms the accuracy of the Poisson solution.

Bibliography

[1] John J. Vastola, Gennady Gorin, Lior Pachter, and William R. Holmes. Analytic solution of chemical master equations involving gene switching. II: Path integral approach to exact solution and applications to parameter inference. in preparation, 2021.

[2] Donald A. McQuarrie. Stochastic approach to chemical kinetics. Journal of Applied Probability, 4(3):413–478, 1967.

[3] Daniel T. Gillespie. A rigorous derivation of the chemical master equation. Physica A: Statistical Mechanics and its Applications, 188(1):404 – 425, 1992.

[4] Daniel T. Gillespie. The chemical Langevin equation. The Journal of Chemical Physics, 113(1):297–306, 2000.

[5] Daniel T. Gillespie. Stochastic simulation of chemical kinetics. Annual Review of Physical Chemistry, 58(1):35–55, 2007. PMID: 17037977.

[6] Daniel T. Gillespie, Andreas Hellander, and Linda R. Petzold. Perspective: Stochastic algorithms for chemical kinetics. The Journal of Chemical Physics, 138(17):170901, 2013.

[7] Zachary Fox and Brian Munsky. Stochasticity or Noise in Biochemical Reactions. arXiv e-prints, page arXiv:1708.09264, Aug 2017.

[8] Brian Munsky, William S. Hlavacek, and Lev S. Tsimring, editors. Quantitative Biology: Theory, Computational Methods, and Models. The MIT Press, 2018.

[9] Radek Erban and S. Jonathan Chapman. Stochastic Modelling of Reaction–Diffusion Processes. Cambridge Texts in Applied Mathematics. Cambridge University Press, 2020.

[10] John J. Vastola and William R. Holmes. Chemical Langevin equation: A path-integral view of Gillespie's derivation. Phys. Rev. E, 101:032417, Mar 2020.

[11] Tobias Jahnke and Wilhelm Huisinga. Solving the chemical master equation for monomolecular reaction systems analytically. Journal of Mathematical Biology, 54(1):1–26, Jan 2007.

[12] John J. Vastola. Solving the chemical master equation for monomolecular reaction systems analytically: a Doi-Peliti path integral view. arXiv e-prints, page arXiv:1911.00978, November 2019.

[13] David Schnoerr, Guido Sanguinetti, and Ramon Grima. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. Journal of Physics A: Mathematical and Theoretical, 50(9):093001, jan 2017.

[14] Ian J. Laurenzi. An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics. The Journal of Chemical Physics, 113(8):3315–3322, 2000.

[15] Erdem Arslan and Ian J. Laurenzi. Kinetics of autocatalysis in small systems. The Journal of Chemical Physics, 128(1):015101, 2008.

[16] Pavol Bokes, John R. King, Andrew T. A. Wood, and Matthew Loose. Exact and approximate distributions of protein and mrna levels in the low-copy regime of gene expression. Journal of Mathematical Biology, 64(5):829–854, Apr 2012.

[17] Hodjat Pendar, Thierry Platini, and Rahul V. Kulkarni. Exact protein distributions for stochastic models of gene expression using partitioning of poisson processes. Phys. Rev. E, 87:042720, Apr 2013.

[18] Casper H. L. Beentjes, Ruben Perez-Carrasco, and Ramon Grima. Exact solution of stochastic gene expression models with bursting, cell cycle and replication dynamics. Phys. Rev. E, 101:032403, Mar 2020.

[19] Zhixing Cao, Tatiana Filatova, Diego A. Oyarzún, and Ramon Grima. A stochastic model of gene expression with polymerase recruitment and pause release. Biophysical Journal, 119(5):1002 – 1014, 2020.

[20] Zhixing Cao and Ramon Grima. Analytical distributions for detailed models of stochastic gene expression in eukaryotic cells. Proceedings of the National Academy of Sciences, 117(9):4682–4692, 2020.

[21] Mark D. Robinson and Gordon K. Smyth. Moderated statistical tests for assessing differences in tag abundance. Bioinformatics, 23(21):2881–2887, 09 2007.

[22] Mark D. Robinson and Gordon K. Smyth. Small-sample estimation of negative binomial dispersion, with applications to SAGE data. Biostatistics, 9(2):321–332, 08 2007.

[23] Yanming Di, Daniel W Schafer, Jason S Cumbie, and Jeff H Chang. The nbp negative binomial model for assessing differential gene expression from rna-seq. Statistical Applications in Genetics and Molecular Biology, 10(1), 12 May. 2011.

[24] Tao Jia and Rahul V. Kulkarni. Intrinsic noise in stochastic models of gene expression with molecular memory and bursting. Phys. Rev. Lett., 106:058102, Feb 2011.

[25] Ann L. Oberg, Brian M. Bot, Diane E. Grill, Gregory A. Poland, and Terry M. Therneau. Technical and biological variance structure in mrna-seq data: life in the real world. BMC Genomics, 13(1):304, Jul 2012.

[26] Niraj Kumar, Thierry Platini, and Rahul V. Kulkarni. Exact distributions for stochastic gene expression models with bursting and feedback. Phys. Rev. Lett., 113:268105, Dec 2014.

[27] Niraj Kumar, Abhyudai Singh, and Rahul V. Kulkarni. Transcriptional bursting in gene expression: Analytical results for general stochastic models. PLOS Computational Biology, 11(10):1–22, 10 2015.

[28] Lisa Amrhein, Kumar Harsha, and Christiane Fuchs. A mechanistic model for the negative binomial distribution of single-cell mrna counts. bioRxiv, 2019.

[29] Gennady Gorin and Lior Pachter. Special function methods for bursty models of transcription. Phys. Rev. E, 102:022409, Aug 2020.

[30] Sheel Shah, Yodai Takei, Wen Zhou, Eric Lubeck, Jina Yun, Chee-Huat Linus Eng, Noushin Koulena, Christopher Cronin, Christoph Karp, Eric J. Liaw, Mina Amin, and Long Cai. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell, 174(2):363–376.e16, July 2018.

[31] Heather L. Drexler, Karine Choquet, and L. Stirling Churchman. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Molecular Cell, 77(5):985–998.e8, March 2020.

[32] Hervé Le Hir, Ajit Nott, and Melissa J. Moore. How introns influence and enhance eukaryotic gene expression. Trends in Biochemical Sciences, 28(4):215 – 220, 2003.

[33] Stefan Stamm, Shani Ben-Ari, Ilona Rafalska, Yesheng Tang, Zhaiyi Zhang, Debra Toiber, T.A. Thanaraj, and Hermona Soreq. Function of alternative splicing. Gene, 344:1 – 20, 2005.

[34] Jennifer C. Long and Javier F. Caceres. The SR protein family of splicing factors: master regulators of gene expression. Biochemical Journal, 417(1):15–27, 12 2008.

[35] Qianliang Wang and Tianshou Zhou. Alternative-splicing-mediated gene expression. Phys. Rev. E, 89:012713, Jan 2014.

[36] Kian Huat Lim, Zhou Han, Hyun Yong Jeon, Jacob Kach, Enxuan Jing, Sebastien Weyn-Vanhentenryck, Mikaela Downs, Anna Corrionero, Raymond Oh, Juergen Scharner, Aditya Venkatesh, Sophina Ji, Gene Liau, Barry Ticho, Huw Nash, and Isabel Aznarez. Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. Nature Communications, 11(1):3501, Jul 2020.

[37] J. Peccoud and B. Ycart. Markovian modeling of gene-product synthesis. Theoretical Population Biology, 48(2):222 – 234, 1995.

[38] Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. Stochastic mrna synthesis in mammalian cells. PLOS Biology, 4(10):1–13, 09 2006.

[39] Srividya Iyer-Biswas, F. Hayot, and C. Jayaprakash. Stochasticity of gene products from transcriptional pulsing. Phys. Rev. E, 79:031911, Mar 2009.

[40] Lifang Huang, Zhanjiang Yuan, Peijiang Liu, and Tianshou Zhou. Effects of promoter leakage on dynamics of gene expression. BMC Systems Biology, 9(1):16, Mar 2015.

[41] Zhixing Cao and Ramon Grima. Linear mapping approximation of gene regulatory networks with stochastic dynamics. Nature Communications, 9(1):3305, Aug 2018.

[42] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4):403 – 434, 1976.

[43] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81(25):2340–2361, 1977.

[44] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. The Journal of Chemical Physics, 124(4):044104, 2006.

[45] Zachary R Fox and Brian Munsky. The finite state projection based fisher information matrix approach to estimate information and optimize single-cell experiments. PLOS Computational Biology, 15(1):1–23, 01 2019.

[46] Arjun Raj, Charles S Peskin, Daniel Tranchina, Diana Y Vargas, and Sanjay Tyagi. Stochastic mRNA Synthesis in Mammalian Cells. PLoS Biology, 4(10):e309, September 2006.

[47] D.J. Griffiths and D.F. Schroeter. Introduction to Quantum Mechanics. Cambridge University Press, 2018.

[48] Matthew D. Schwartz. Quantum Field Theory and the Standard Model. Cambridge University Press, 2014.

[49] P. Grassberger and M. Scheunert. Fock-space methods for identical classical objects. Fortschritte der Physik, 28(10):547–578, 1980.

[50] John J. Vastola. The chemical birth-death process with additive noise. arXiv e-prints, page arXiv:1910.09117, October 2019.

[51] John Baez and Jacob D Biamonte. Quantum Techniques in Stochastic Mechanics. World Scientific, 2018.

[52] Richard D Mattuck. A guide to Feynman diagrams in the many-body problem. Dover Publications, 2 edition, 1992.

[53] Philipp Thomas and Ramon Grima. Approximate probability distributions of the master equation. Phys. Rev. E, 92:012120, Jul 2015.

[54] Rosalba Garcia-Millan and Gunnar Pruessner. Run-and-tumble motion: field theory and entropy production. arXiv e-prints, page arXiv:2012.02900, December 2020.

[55] Tianshou Zhou and Jiajun Zhang. Analytical Results for a Multistate Gene Model. SIAM Journal on Applied Mathematics, 72(3):789–818, January 2012.

[56] Tatiana Filatova, Nikola Popovic, and Ramon Grima. Statistics of nascent and mature rna fluctuations in a stochastic model of transcriptional initiation, elongation, pausing, and termination. Bulletin of Mathematical Biology, 83(1):3, Dec 2020.

[57] John J. Vastola and William R. Holmes. Stochastic path integrals can be derived like quantum mechanical path integrals. arXiv e-prints, page arXiv:1909.12990, September 2019.

[58] Paul C. Bressloff. Construction of stochastic hybrid path integrals using "quantum-mechanical" operators. arXiv e-prints, page arXiv:2012.07770, December 2020.

[59] Paul C. Bressloff. Spin coherent states and stochastic hybrid path integrals. arXiv e-prints, page arXiv:2102.03878, February 2021.

[60] Izrail Solomonovich Gradshteyn and Iosif Moiseevich Ryzhik. Table of integrals, series, and products. Academic press, 2014.

[61] Ankit Gupta, Jan Mikelson, and Mustafa Khammash. A finite state projection algorithm for the stationary solution of the chemical master equation. The Journal of Chemical Physics, 147(15):154101, 2017.

Chapter 7

Analytic solution of chemical master equations involving gene switching.
II: Path integral approach to exact solution and applications to parameter inference

How can we *solve* the CME? Unlike in the previous chapters, here we are concerned with developing an efficient *numerical* solution technique. We leverage the novel path integral technique introduced in Chapter 5 to reduce the computation of steady state probability distributions to (i) solving ODEs, and then (ii) doing an inverse Fourier transform. For this specific class of problems (monomolecular reactions + gene switching), we greatly outperform competing approaches like finite state projection and stochastic simulations. Because this approach is significantly faster than previously known approaches, it in principle allows one to fit more complicated/realistic models to single cell RNA counts data.

Abstract: This is the second in a series of two papers on solving chemical master equations involving gene switching. In the previous paper, we used a ladder operator approach inspired by quantum mechanics to obtain exact analytic solutions to two biologically interesting problems. However, the solutions we obtained were in the form of slowly converging infinite series, which are not practical for purposes like parameter inference. In this paper, we used a completely orthogonal theoretical approach—a state space path integral representation of the chemical master equation—in order to derive different analytic solutions more appropriate for numerical implementations. We show that this approach yields fast and stable implementations of the steady state solutions of these problems, which outperform alternative approaches like finite state projection, and which can in principle be used to interrogate transcription and splicing dynamics given single cell RNA counts data. We illustrate this using simulated data from models of bursty transcription and splicing involving one, two, and three species.

7.1 Introduction

This is the second in a series of two papers on solving chemical master equations involving a switching gene, particularly in the context of bursty models of transcription and splicing. In the previous paper, we used a ladder operator method inspired by quantum mechanics to derive exact perturbative solutions to a number of problems. We found solutions that were informative, but not numerically stable. In this paper, we use a completely orthogonal path integral method to solve the same problems, and develop numerically fast and efficient implementations of these solutions. In other words, the previous paper tells us how to think about the solutions to these models; this paper will offer a fast way to work with them in practice, especially for purposes like parameter inference.

Parameter inference is currently a salient topic in biophysics and single cell biology, especially given recent advances in obtaining single cell RNA counts data using fluorescence or sequencing-based approaches. RNA counts can be recorded accurately enough to fit probability distributions, which can be connected to mechanistic chemical master equation (CME) models to learn about the kinetics and structure of the underlying biological processes [1]. Given new work that has recorded not just how many RNA are in single cells, but how many RNA with different numbers of introns removed are present [2, 3], it is in principle possible to apply mechanistic models to counts data to learn about the dynamics of splicing—and in turn, how splicing controls gene expression [4, 5, 6, 7, 8]. How many splicing steps are typical? How quickly are introns removed, and how much does this differ across introns and genes? Are introns usually removed in a fixed order, or a random order?

The ability to identify splicing topologies and parameters has far-reaching implications. The estimation of splicing rates using a discrete and stochastic CME model produces results orthogonal to conventional methods for probing intron removal kinetics [9], with broader applicability to transcriptome-wide, label-free single cell RNA-seq data. The explicit description of intermediate species permits finer identification of gene locus kinetics [10]. Finally, the ability to simulate a system forward in time offers opportunities for mapping cellular development and decisions [11].
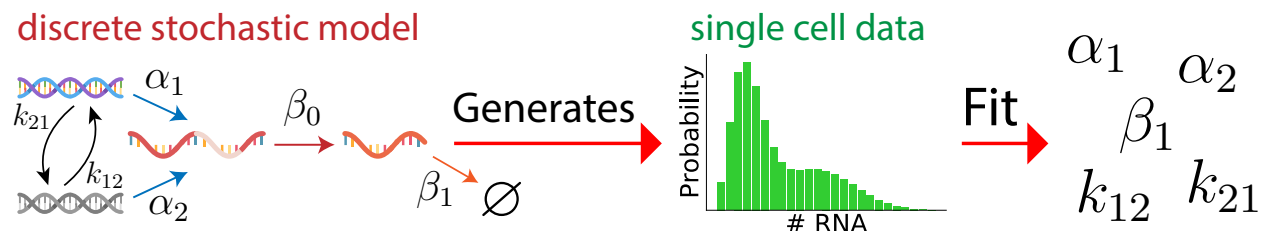

Figure 7.1: The general problem considered in this paper. Given (steady state) simulated RNA counts data generated by a CME model of bursty transcription and splicing, we must correctly infer (non-dimensionalized) model parameters. The two species model is pictured on the left, but our theoretical approach works for any number of splicing steps.

In this work, we use our new path integral-based technique to explore the parameter inference problem in the context of simple stochastic models of bursty transcription and splicing (Fig. 7.1), and impose a number of restrictions to make it tractable. First, we use 'clean' simulated data, and assume the correct model is known *a priori*; this means that our task is only to correctly infer parameters, and to do so in a way that runs quickly, is numerically stable, and scales relatively well as we consider increasingly complex models. Second, we assume that introns are spliced out in a fixed order (rather than in a random or semi-random order), i.e. that we have a 'linear' path graph. Finally, we assume one downstream RNA product, i.e. that there is no alternative splicing.

This paper is organized as follows. In Sec. 7.2, we briefly discuss utility of different kinds of solutions to the CME for purposes like parameter inference, and the importance of analytic approaches. In Sec. 7.3, we derive a set of ordinary differential equations (ODEs) behind our new efficient numerical solution of the CME for this class of systems. In Sec. 7.4, we validate our results and benchmark performance against other methods (like stochastic simulations and finite state projection). In Sec. 7.5, we apply our new method to do fast parameter inference on simulated data for the birth-death switching and multistep splicing systems. In Sec. 7.6, we discuss implications and possible generalizations.

7.2 Numerical solutions of the CME

CME models are discrete and stochastic, and can only predict the *probability* of observing different numbers of RNA inside a single cell at some future time, rather than the specific number. In order to use these models to learn about real single cell data—especially single cell RNA counts data—we need efficient ways to generate model predictions. In particular, we need to generate the steady state probability distributions predicted by a CME model for many different parameter sets, and compare them with experimental counts distributions. The parameter set for which the CME distribution most closely matches the empirical distribution is considered the solution of the inference problem[1].

Model predictions typically needs to be generated for some large number of parameter sets (thousands, tens of thousands, or more), so inefficient features of implementations accumulate and make the fitting of models with many parameters in reasonable amounts of time unrealistic. Without efficient implementations, only the simplest models can be compared against data, limiting how much we can learn about the underlying structure and kinetics.

It is helpful to think of approaches to solving CME models as falling into three heuristic classes: (i) simulation-based, (ii) solver-based, and (iii) analytic. Simulation-based methods employ stochastic simulations (usually variants of the Gillespie algorithm/stochastic simu-

---

[1]Of course, the solution to an inference problem may not be unique, especially when the model in question is 'sloppy' [12, 13, 14]. Although this issue must be dealt with carefully in general, it does not seem to be significant here.

lation algorithm [15, 16]) to sample the steady state probability distribution. Solver-based methods, like finite state projection (FSP) [17, 18, 19], do not involve simulations but instead attempt to solve the CME itself (or something equivalent to it) directly through a numerical procedure. Analytic methods involve using various mathematical techniques (as we did in the previous part), frequently method of characteristics solutions to a partial differential equation describing the time evolution of the generating function, to write down a closed-form solution to the problem.
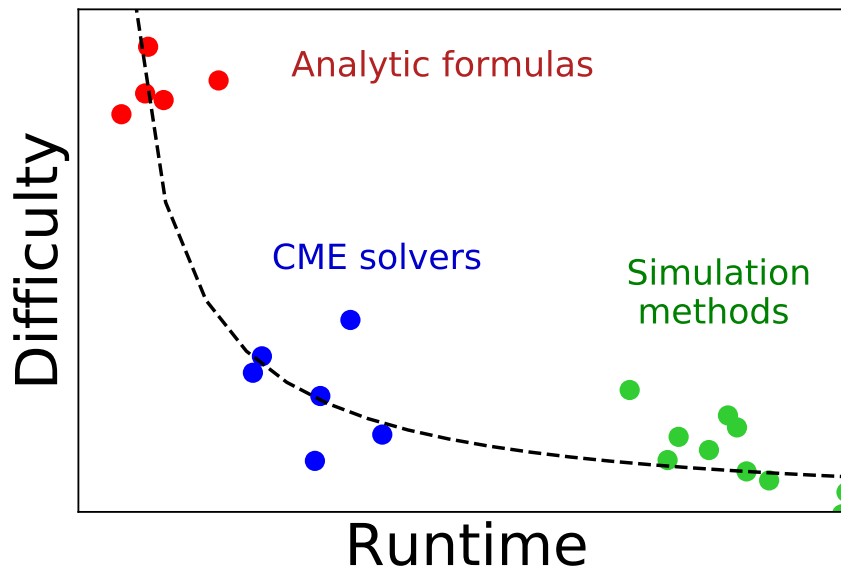

Figure 7.2: Implementation difficulty versus runtime. Simulation methods (green dots) are easy to implement, but tend to be slow at generating CME solutions. CME solver methods (blue dots) are harder to implement, and more problem-specific, but run faster. Analytic methods (red dots) are the fastest possible approach, but exact solution formulas tend to be extremely challenging to derive.

In choosing between these methods, one encounters two important trade-offs. The first is between implementation difficulty and runtime. While stochastic simulation methods are very easy to code, and while code for simulating one model can easily be adapted to simulate another, they tend to be a very slow way to generate steady state probability distributions. FSP, the paradigmatic numerical solver method, is both faster than stochastic simulation approaches and somewhat harder to implement. Its code is also more problem-specific. Because it involves exponentiating (or finding the null space of) a generically large matrix, it scales somewhat poorly as state space size increases. Analytic approaches vary, but their runtime is often quite fast: their runtime can scale linearly with state space size, or even be approximately constant.

The other important trade-off is stability: although the analytic infinite series solution

for the birth-death-switching problem presented Part I can in principle be computed very quickly, it is not numerically stable, making it useless for parameter inference. It is harder to make precise assertions about the stability of various solution approaches, but stochastic simulations are generally quite stable (provided one runs enough simulations), while analytic solutions can be more variable in their stability.

<div align="center">

Solve ODEs to get generating function

$$\dot{Y}_1 = -k_{21}Y_1 + k_{21}Y_2 + \alpha_1 \left[\sum_i c_i e^{-\beta_i s} \right] Y_1$$

$$\dot{Y}_2 = k_{12}Y_1 - k_{12}Y_2 + \alpha_2 \left[\sum_i c_i e^{-\beta_i s} \right] Y_2$$

iFFT $\longrightarrow$

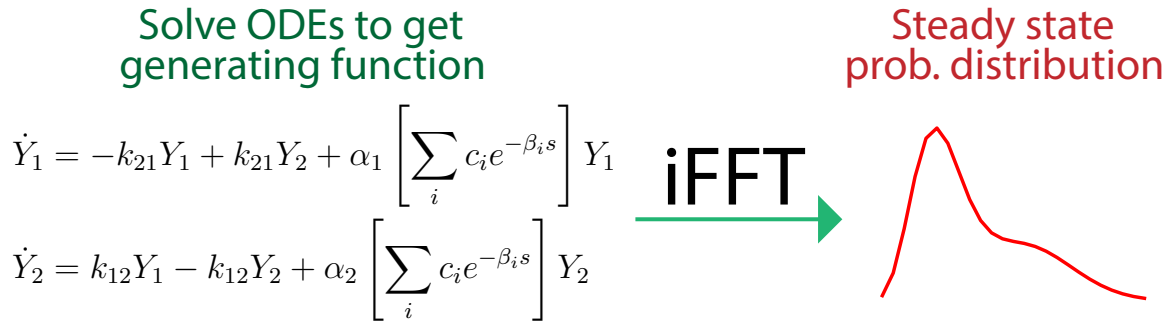**Steady state prob. distribution**

</div>

Figure 7.3: Our novel approach to numerically solving the CME for models of transcription and splicing. We use a state space path integral representation of the CME to derive ODEs satisfied by the generating function, which we solve numerically. Then we take an inverse fast Fourier transform to obtain the steady state probability distribution.
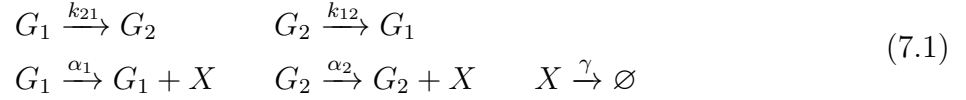
In this paper, we take a hybrid approach in order to balance speed with stability. Using a state space path integral representation of the CME [20], we find that the generating function of the probability distribution (essentially, its discrete Fourier transform) satisfies a certain set of ODEs. Instead of solving these ODEs analytically (which is likely to yield an unstable series solution), we solve them numerically. Then we take the inverse fast Fourier transform to obtain our result (Fig. 7.3). Because one can integrate ODEs and take an inverse Fourier transform relatively quickly, the implementation is still very fast.

7.3   Path integral approach to analytic solutions

In this section, we present the path integral computations that lead to our numerical solver algorithm. First, in Sec. 7.3.1, we define and present the calculations for solving the one species birth-death-switching problem. While they are a special case of the more general problem, seeing the derivation in a simpler context is helpful for developing a sense of how the calculations work. Then, in Sec. 7.3.2, we define and present the calculations for solving the general multi-species problem, involving bursty transcription and arbitrarily many splicing steps. In Sec. 7.3.3, we simplify the multistep solution in the cases of one splicing step and two splicing steps.

7.3.1 Solution to birth-death-switching via path integral method

Let X denote the RNA species, and G_1 and G_2 the two possible gene states. RNA production and degradation coupled to a switching gene can be modeled using the list of chemical reactions

$$
\begin{aligned}
G_1 &\xrightarrow{k_{21}} G_2 & G_2 &\xrightarrow{k_{12}} G_1 \\
G_1 &\xrightarrow{\alpha_1} G_1 + X & G_2 &\xrightarrow{\alpha_2} G_2 + X & X &\xrightarrow{\gamma} \varnothing
\end{aligned}
\tag{7.1}
$$

where k_{21} and k_{12} parameterize the rates of gene switching, α_1 and α_2 parameterize the transcription rate in each gene state, and γ parameterizes the RNA's degradation rate.

In this treatment of the birth-death-switching problem, it is helpful to imagine it as involving three distinct species: RNA ($x \in \mathbb{N}$), the number of genes in the G_1 state ($a \in \mathbb{N}$), and the number of genes in the G_2 state ($b \in \mathbb{N}$). Even though we are only interested in the case with $a + b = 1$ (since the gene is either in one state or the other), assuming $a, b \in \mathbb{N}$ is helpful for certain technical reasons.

Let $P := P(x, a, b, t; x_0, a_0, b_0, t_0)$ denote the probability of transitioning from a state with (x_0, a_0, b_0) at time t_0 to a state with (x, a, b) at time t. Without loss of generality, we will assume $t_0 = 0$. Using the state space path integral developed in [20], we can write P as the expression

$$
\begin{aligned}
P = \sum_{x_1,\ldots,x_{T-1}} \sum_{a_1,\ldots,a_{T-1}} \sum_{b_1,\ldots,b_{T-1}} \int \frac{dp_1 dA_1 dB_1 \cdots dp_T dA_T dB_T}{(2\pi)^T (2\pi)^T (2\pi)^T} \\
\exp\Bigg\{ \sum_{\ell=1}^{T} -ip_\ell(x_\ell - x_{\ell-1}) - iA_\ell(a_\ell - a_{\ell-1}) - iB_\ell(b_\ell - b_{\ell-1}) \\
+ k_{21} a_{\ell-1} \Delta t \left(e^{i(B_\ell - A_\ell)} - 1 \right) + k_{12} b_{\ell-1} \Delta t \left(e^{i(A_\ell - B_\ell)} - 1 \right) \\
+ \alpha_1 a_{\ell-1} \Delta t \left(e^{ip_\ell} - 1 \right) + \alpha_2 b_{\ell-1} \Delta t \left(e^{ip_\ell} - 1 \right) + \gamma\, x_{\ell-1} \Delta t \left(e^{-ip_\ell} - 1 \right) \Bigg\}
\end{aligned}
\tag{7.2}
$$

where we have gone from $t_0 = 0$ to t in T steps, with $\Delta t := (t - t_0)/T$.

The early parts of this calculation closely mirror the birth-death calculations presented in Appendix A of [20]. First, collect terms proportional to x_ℓ (for $\ell = 1, \ldots, T-1$) and sum over x_ℓ:

$$
\begin{aligned}
&\sum_{x_\ell=0}^{\infty} \exp\left\{ -ix_\ell \left[p_\ell - p_{\ell+1} + i\gamma\Delta t \left(e^{-ip_{\ell+1}} - 1 \right) \right] \right\} \\
&= \frac{1}{1 - e^{-ix_\ell \left[p_\ell - p_{\ell+1} + i\gamma\Delta t \left(e^{-ip_{\ell+1}} - 1 \right) \right]}}\ .
\end{aligned}
\tag{7.3}
$$

The integrals over p_ℓ (for $\ell = 1, \ldots, T-1$), as in Appendix A of [20], schematically take the form

$$
\int_{-\pi}^{\pi} \frac{dp_\ell}{2\pi} \frac{f(p_\ell)}{1 - e^{-ix_\ell \left[p_\ell - p_{\ell+1} + i\gamma\Delta t \left(e^{-ip_{\ell+1}} - 1 \right) \right]}}\ .
\tag{7.4}
$$

If we treat this as a contour integral[2] and use Cauchy's integral formula to evaluate it, we end up enforcing the constraint

$$p_\ell = p_{\ell+1} - \Delta t i \gamma \left(e^{-ip_{\ell+1}} - 1 \right) \tag{7.5}$$

for $\ell = 1, ..., T - 1$, which makes each of $p_1, ..., p_{T-1}$ functions of p_T. In the small Δt limit we are interested in, this constraint becomes the (backwards) ODE

$$\dot{p} = -i\gamma \left(e^{-ip} - 1 \right) \tag{7.6}$$

with initial condition $p(0) = p_T$. If we define $q := e^{ip} - 1$, this simplifies to

$$\dot{q} = i e^{ip} \dot{p} = \gamma \left(1 - e^{ip} \right) = -\gamma q \tag{7.7}$$

with initial condition $q(0) = e^{ip_T} - 1$. Solving this, we have

$$q(s) = e^{ip(s)} - 1 = \left(e^{ip_T} - 1 \right) e^{-\gamma s} \tag{7.8}$$

for $s \in [0, t]$. Summing over a_ℓ and b_ℓ (for $\ell = 1, ..., T - 1$) and enforcing constraints on A_ℓ and B_ℓ in similar fashion, we have ODEs

$$\begin{aligned} \dot{A} &= -i \left[k_{21} \left(e^{i(B-A)} - 1 \right) + \alpha_1 \left(e^{ip} - 1 \right) \right] \\ \dot{B} &= -i \left[k_{12} \left(e^{i(A-B)} - 1 \right) + \alpha_2 \left(e^{ip} - 1 \right) \right] . \end{aligned} \tag{7.9}$$

Substitute in our expression for $e^{ip(s)} - 1$, and make the change of variables

$$Y_1(s) := e^{iA(s)} \qquad\qquad Y_2(s) := e^{iB(s)} . \tag{7.10}$$

Now we have

$$\begin{aligned} \dot{Y}_1 &= -k_{21} Y_1 + k_{21} Y_2 + \alpha_1 \left(e^{ip_T} - 1 \right) e^{-\gamma t} \, Y_1 \\ \dot{Y}_2 &= k_{12} Y_1 - k_{12} Y_2 + \alpha_2 \left(e^{ip_T} - 1 \right) e^{-\gamma t} \, Y_2 \end{aligned} \tag{7.11}$$

with initial condition $Y_1(0) = e^{iA_T}$ and $Y_2(0) = e^{iB_T}$.

What remains of our path integral is

$$P(x, a, b, t; x_0, a_0, b_0, 0) = \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \frac{dp_T dA_T dB_T}{(2\pi)^3} e^{-ip_T x + ip_0 x_0 - iA_T a + iA_0 a_0 - iB_T b + iB_0 b_0} \tag{7.12}$$

where

$$\begin{aligned} p_0 &:= p(t) = \left(e^{ip_T} - 1 \right) e^{-\gamma t} \\ A_0 &:= A(t) \\ B_0 &:= B(t) . \end{aligned} \tag{7.13}$$

---

[2]Unfortunately, we must be somewhat cavalier about convergence. These integrals are not technically well-defined without some sort of regulator analogous to the $i\epsilon$ prescription from quantum field theory. But since these solutions will be compared with alternative approaches known to work, this issue is not particularly worrisome.

Let us simplify things in three ways. First, we are interested in the solution marginalized over gene state, so we can sum over a and b, i.e.

$$P(x, t; x_0, 0) = \sum_{a=0}^{\infty} \sum_{b=0}^{\infty} P(x, a, b, t; x_0, a_0, b_0, 0)$$

$$= \int \int \int \frac{dp_T dA_T dB_T}{(2\pi)^3} e^{-ip_T x + ip_0 x_0 + iA_0 a_0 + iB_0 b_0} \sum_{a=0}^{\infty} e^{-iA_T a} \sum_{b=0}^{\infty} e^{-iB_T b} \quad (7.14)$$

$$= \int \frac{dp_T}{2\pi} e^{-ip_T x + ip_0 x_0} \int \frac{dA_T}{2\pi} \frac{e^{iA_0(A_T)a_0}}{1 - e^{-iA_T}} \int \frac{dB_T}{2\pi} \frac{e^{iB_0(B_T)b_0}}{1 - e^{-iB_T}}$$

where we have written $A_0(A_T)$ and $B_0(B_T)$ to emphasize that A_0 and B_0 are functions of A_T and B_T respectively. Interpreting the integrals over A_T and B_T as contour integrals enforces $A_T = B_T = 0$, which affects A_0 and B_0 (via the ODE above) by changing the initial conditions to $Y_1(0) = Y_2(0) = 1$.

Second, for numerical purposes the generating function (the discrete Fourier transform of the probability distribution) is equally acceptable. Hence, we compute

$$\psi(\theta, t) = \sum_{x=0}^{\infty} e^{i\theta x} P(x, t; x_0, 0)$$

$$= \int \frac{dp_T}{2\pi} e^{ip_0 x_0 + iA_0 a_0 + iB_0 b_0} \sum_{x=0}^{\infty} e^{i(\theta - p_T)x} \quad (7.15)$$

$$= \int \frac{dp_T}{2\pi} \frac{e^{ip_0 x_0 + iA_0 a_0 + iB_0 b_0}}{1 - e^{i(\theta - p_T)}} .$$

This integral has the effect of enforcing the constraint $p_T = \theta$, which again affects the ODEs. We have

$$\psi(\theta, t) = \exp\{ip_0 x_0 + iA_0 a_0 + iB_0 b_0\} \quad (7.16)$$

where $p_0 = (e^{i\theta} - 1) e^{-\gamma t}$ and A_0 and B_0 are the solutions to the ODEs above at $s = t$. Finally, we can simplify our answer by considering the solution at steady state, i.e. taking $t \to \infty$. In this limit, our initial gene state does not matter, so we arbitrarily take $a_0 = 1$ and $b_0 = 0$. We also have $p_0 = 0$. Hence,

$$\psi_{ss}(\theta) = Y_1 \quad (7.17)$$

where Y_1 is the long time limit of the solution of

$$\dot{Y}_1 = -k_{21} Y_1 + k_{21} Y_2 + \alpha_1 \left(e^{i\theta} - 1\right) e^{-\gamma s} Y_1$$
$$\dot{Y}_2 = k_{12} Y_1 - k_{12} Y_2 + \alpha_2 \left(e^{i\theta} - 1\right) e^{-\gamma s} Y_2 \quad (7.18)$$
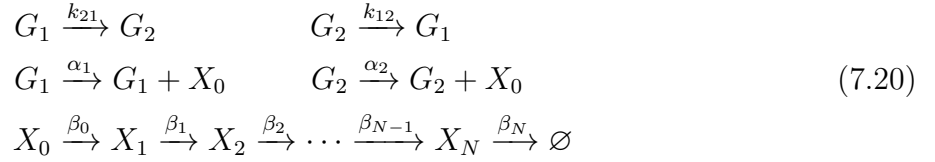
with initial condition $Y_1(0) = Y_2(0) = 1$. Actually, since Y_1 and Y_2 eventually converge to the same value, we can write

$$\psi_{ss}(\theta) = f_1 Y_1(\infty) + f_2 Y_2(\infty) \quad (7.19)$$

228

for any nonnegative coefficients f_1 and f_2 with $f_1 + f_2 = 1$. Although it is possible to construct an infinite series solution of the above, we will leave the solution as-is, because this set of ODEs is likely to offer a more numerically stable implementation than solving them explicitly. If we did proceed, we would either recover one of our results from Part I, or the previously known hypergeometric solution [21, 22].

7.3.2 Solution to multistep splicing plus switching via path integral method

Let X_0 denote unspliced/nascent RNA, X_1 denote RNA that has experienced one splicing step (i.e. one intron has been removed), X_2 denote RNA that has experienced two splicing steps, and so on. Assuming there are N splicing steps, let X_N denote the fully processed/mature RNA. As before, let G_1 and G_2 denote the two possible gene states. Transcription coupled to a switching gene, with multiple downstream splicing steps, can be modeled using the list of chemical reactions

$$
\begin{aligned}
G_1 &\xrightarrow{k_{21}} G_2 & G_2 &\xrightarrow{k_{12}} G_1 \\
G_1 &\xrightarrow{\alpha_1} G_1 + X_0 & G_2 &\xrightarrow{\alpha_2} G_2 + X_0 \\
X_0 &\xrightarrow{\beta_0} X_1 \xrightarrow{\beta_1} X_2 \xrightarrow{\beta_2} \cdots \xrightarrow{\beta_{N-1}} X_N \xrightarrow{\beta_N} \varnothing
\end{aligned}
\tag{7.20}
$$

where k_{21} and k_{12} parameterize the rates of gene switching, α_1 and α_2 parameterize the transcription rate in each gene state, β_i parameterizes the rate of the ith splicing step (for $0 \le i < N$), and β_N parameterizes the mature RNA's degradation rate.

This time, the model has $N+1+2$ species: $N+1$ RNA species ($\mathbf{x} \in \mathbb{N}^{N+1}$), and two gene states ($a, b \in \mathbb{N}$). As before, we are interested in the probability $P := P(\mathbf{x}, a, b, t; \mathbf{x}_0, a_0, b_0, t_0)$ of transitioning from a state with $(\mathbf{x}_0, a_0, b_0)$ at time t_0 to a state with $(\mathbf{x}, a, b)$ at time t.

One can write down a path integral description of this problem similar to the one in Sec. 7.3.1, with the main difference being more notational baggage. In order to distinguish vector elements (i.e. $\mathbf{x} = (x_0, x_1, ..., x_{N+1})$) from values of counts at the ℓ-th time step, we will indicate time step indices with superscripts. The path integral reads:

$$
P = \sum_{a^{(1)},...,a^{(T-1)}} \sum_{b^{(1)},...,b^{(T-1)}} \int \frac{dA^{(1)} dB^{(1)} \cdots dA^{(T)} dB^{(T)}}{(2\pi)^T (2\pi)^T} \prod_{j=0}^{N} \sum_{x_j^{(1)},...,x_j^{(T-1)}} \int \frac{dp_j^{(1)} \cdots dp_j^{(T)}}{(2\pi)^T}
$$

$$
\exp \left\{ \sum_{\ell=1}^{T} -i\mathbf{p}^{(\ell)} \cdot \left[\mathbf{x}^{(\ell)} - \mathbf{x}^{(\ell-1)} \right] - iA^{(\ell)} \left[a^{(\ell)} - a^{(\ell-1)} \right] - iB^{(\ell)} \left[b^{(\ell)} - b^{(\ell-1)} \right] \right.
$$

$$
+ k_{21} a^{(\ell-1)} \Delta t \left[e^{iB^{(\ell)} - iA^{(\ell)}} - 1 \right] + k_{12} b^{(\ell-1)} \Delta t \left[e^{iA^{(\ell)} - iB^{(\ell)}} - 1 \right]
$$

$$
+ \alpha_1 a^{(\ell-1)} \Delta t \left[e^{ip_0^{(\ell)}} - 1 \right] + \alpha_2 b^{(\ell-1)} \Delta t \left[e^{ip_0^{(\ell)}} - 1 \right]
$$

$$
\left. + \sum_{j=0}^{N-1} \beta_j x_j^{(\ell-1)} \Delta t \left[e^{ip_{j+1}^{(\ell)} - ip_j^{(\ell)}} - 1 \right] + \beta_N x_N^{(\ell-1)} \Delta t \left[e^{-ip_N^{(\ell)}} - 1 \right] \right\}
$$

$$
\tag{7.21}
$$

where we have gone from $t_0 = 0$ to t in T steps, with $\Delta t := (t - t_0)/T$. Proceeding as before, we first sum over $x_j^{(\ell)}$ for all $\ell = 1, ..., T-1$ and $j = 0, 1, ..., N$. Then we integrate over $p_j^{(\ell)}$ for all $\ell = 1, ..., T-1$ and $j = 0, 1, ..., N$, again using Cauchy's integral formula to evaluate all of the contour integrals. If we define

$$u_j := e^{ip_j(s)} - 1 \tag{7.22}$$

for $j = 0, 1, ..., N$, we can write the set of backwards ODEs we obtain as

$$\dot{u}_0 = -\beta_0 (u_0 - u_1)$$
$$\vdots$$
$$\dot{u}_j = -\beta_j (u_j - u_{j+1}) \tag{7.23}$$
$$\vdots$$
$$\dot{u}_N = -\beta_N u_N \ .$$

with $u_j(0) = \left(e^{ip_j^{(T)}} - 1\right)$, or equivalently $\mathbf{u}(0) = \left(e^{i\mathbf{p}^{(T)}} - \mathbf{1}\right)$ if we abuse notation slightly. We are only actually interested in solving for $u_0(s)$, since $p_0(s)$ is the only one of the $p_j(s)$ coupled to the a and b variables; when we take the $t \to \infty$ limit, the dependence on the other $u_j(s)$ will disappear.

The result for $u_0(s)$ turns out to be related to the coefficients of the up ladder operators we found in Part I, along with the q_j coefficients. We have

$$u_0(s) = e^{ip_0(s)} - 1 = \sum_{j=0}^{N} q_j \mathbf{v}^{(j)} \cdot \left(e^{i\boldsymbol{\theta}} - \mathbf{1}\right) e^{-\beta_j s} \tag{7.24}$$

where the q_j are defined as

$$q_0 = 1$$
$$q_1 = -\frac{\beta_0}{\beta_1 - \beta_0}$$
$$q_2 = \frac{\beta_0 \beta_1}{(\beta_2 - \beta_0)(\beta_2 - \beta_1)} \tag{7.25}$$
$$q_j = (-1)^j \frac{\beta_0 \cdots \beta_{j-1}}{(\beta_j - \beta_0) \cdots (\beta_j - \beta_{j-1})} \qquad (\ 1 \le j \le N\ )$$

and the vectors $\mathbf{v}^{(k)}$ (for all $k = 0, 1, ..., N$) are defined as

$$v_j^{(k)} = \begin{cases} 0 & j < k \\ 1 & j = k \\ \frac{\beta_k \cdots \beta_{j-1}}{(\beta_{k+1} - \beta_k) \cdots (\beta_j - \beta_k)} & j > k \ . \end{cases} \tag{7.26}$$

Continue as in the previous section: sum over the gene-state-related variables, integrate over their corresponding momenta $A^{(\ell)}$ and $B^{(\ell)}$, and make the three simplifications we used earlier (marginalize over gene states, compute the generating function, take the $t \to \infty$ limit). Finally, we obtain the ODEs

$$
\dot{Y}_1 = -k_{21}Y_1 + k_{21}Y_2 + \alpha_1 \left[\sum_{j=0}^{N} c_j e^{-\beta_j s} \right] Y_1
$$
$$
\dot{Y}_2 = k_{12}Y_1 - k_{12}Y_2 + \alpha_2 \left[\sum_{j=0}^{N} c_j e^{-\beta_j s} \right] Y_2
$$

(7.27)

with $Y_1(0) = Y_2(0) = 1$, with

$$
c_j := q_j \mathbf{v}^{(j)} \cdot \left(e^{i\boldsymbol{\theta}} - \mathbf{1} \right) \ .
$$

(7.28)

As in the one species case, the generating function is given by the long time limit of $Y_1(s)$ (or any weighted sum of the long time limits of Y_1 and Y_2 whose coefficients add to 1, i.e. $f_1 + f_2 = 1$).

7.3.3 Special cases of multistep splicing solution

Let us unpack the solution we derived in the previous section for the two special cases that will interest us most: the case of $N = 1$ (i.e. one splicing step, two species), and $N = 2$ (i.e. two splicing steps, three species). To ease notation slightly, we will write $g_j := e^{i\theta_j}$.

The first few q_j coefficients were given above, and do not depend on N. For $N = 1$, the relevant vectors read

$$
\mathbf{v}^{(0)} = \left(1, \frac{\beta_0}{\beta_1 - \beta_0} \right)
$$
$$
\mathbf{v}^{(1)} = (0, 1) \ .
$$

(7.29)

Hence,

$$
u_0(s) = \left[(g_0 - 1) + \frac{\beta_0}{\beta_1 - \beta_0} (g_1 - 1) \right] e^{-\beta_0 s} - \frac{\beta_0}{\beta_1 - \beta_0} (g_1 - 1) e^{-\beta_1 s}
$$

(7.30)

so that

$$
c_0 := (g_0 - 1) + \frac{\beta_0}{\beta_1 - \beta_0} (g_1 - 1)
$$
$$
c_1 := -\frac{\beta_0}{\beta_1 - \beta_0} (g_1 - 1) \ .
$$

(7.31)

When $N = 2$, we have

$$
\mathbf{v}^{(0)} = \left(1, \frac{\beta_0}{\beta_1 - \beta_0}, \frac{\beta_0 \beta_1}{(\beta_1 - \beta_0)(\beta_2 - \beta_0)} \right)
$$
$$
\mathbf{v}^{(1)} = \left(0, 1, \frac{\beta_1}{\beta_2 - \beta_1} \right)
$$
$$
\mathbf{v}^{(2)} = (0, 0, 1) \ .
$$

(7.32)

231

Hence,

$$
\begin{aligned}
u_0(s) = {} & \left[(g_0 - 1) + \frac{\beta_0}{\beta_1 - \beta_0}(g_1 - 1) + \frac{\beta_0\beta_1}{(\beta_1 - \beta_0)(\beta_2 - \beta_0)}(g_2 - 1) \right] e^{-\beta_0 s} \\
& - \frac{\beta_0}{\beta_1 - \beta_0} \left[(g_1 - 1) + \frac{\beta_1}{\beta_2 - \beta_1}(g_2 - 1) \right] e^{-\beta_1 s} + \frac{\beta_0\beta_1}{(\beta_2 - \beta_0)(\beta_2 - \beta_1)}(g_2 - 1)e^{-\beta_2 s}
\end{aligned}
\tag{7.33}
$$

so that

$$
\begin{aligned}
c_0 &= (g_0 - 1) + \frac{\beta_0}{\beta_1 - \beta_0}(g_1 - 1) + \frac{\beta_0\beta_1}{(\beta_1 - \beta_0)(\beta_2 - \beta_0)}(g_2 - 1) \\
c_1 &= -\frac{\beta_0}{\beta_1 - \beta_0} \left[(g_1 - 1) + \frac{\beta_1}{\beta_2 - \beta_1}(g_2 - 1) \right] \\
c_2 &= \frac{\beta_0\beta_1}{(\beta_2 - \beta_0)(\beta_2 - \beta_1)}(g_2 - 1) \ .
\end{aligned}
\tag{7.34}
$$

7.4   Efficient implementation and validation

In this section, we discuss how to implement the ODE-based approach we derived in the previous section to efficiently compute the steady state solutions of our transcription and splicing models. We validate our results against stochastic simulations, finite state projection, and the hypergeometric solution [21, 22] of the one species problem.

Our basic task is to use the ODEs we derived (Eq. 7.18 in the one species case, and Eq. 7.27 in the general case) to compute values of the generating function on the unit circle, so that we can inverse Fourier transform the result to recover the steady state probability distribution. There are various implementation tricks one can use to speed up computation. Because ODE computations for different values of $\boldsymbol{\theta}$ are independent, we vectorized our ODE integration routine over them so that they are computed in parallel. This prevents the use of adaptive methods, but we found that the standard fourth order Runge Kutta method was fast and stable. Because our probability distribution should be real rather than complex-valued, we can halve runtime by using an inverse real fast Fourier transform (irfft) rather than a standard inverse fast Fourier transform.

As mentioned earlier, we can integrate Y_1 and Y_2 and compute the generating function by using the long time limit of $f_1 Y_1 + f_2 Y_2$ for any nonnegative coefficients f_1 and f_2 with $f_1 + f_2 = 1$. Empirically, we found that using $f_1 = k_{12}/(k_{12} + k_{21})$ and $f_2 = k_{21}/(k_{12} + k_{21})$ lead to convergence much more quickly than other alternatives.

Qualitatively, the results obtained using our ODE approach match results obtained via FSP (Fig. 7.4). As can be seen in Fig. 7.5, the results quantitatively agree with both FSP and stochastic simulations. However, as the state space increases in size, it scales much
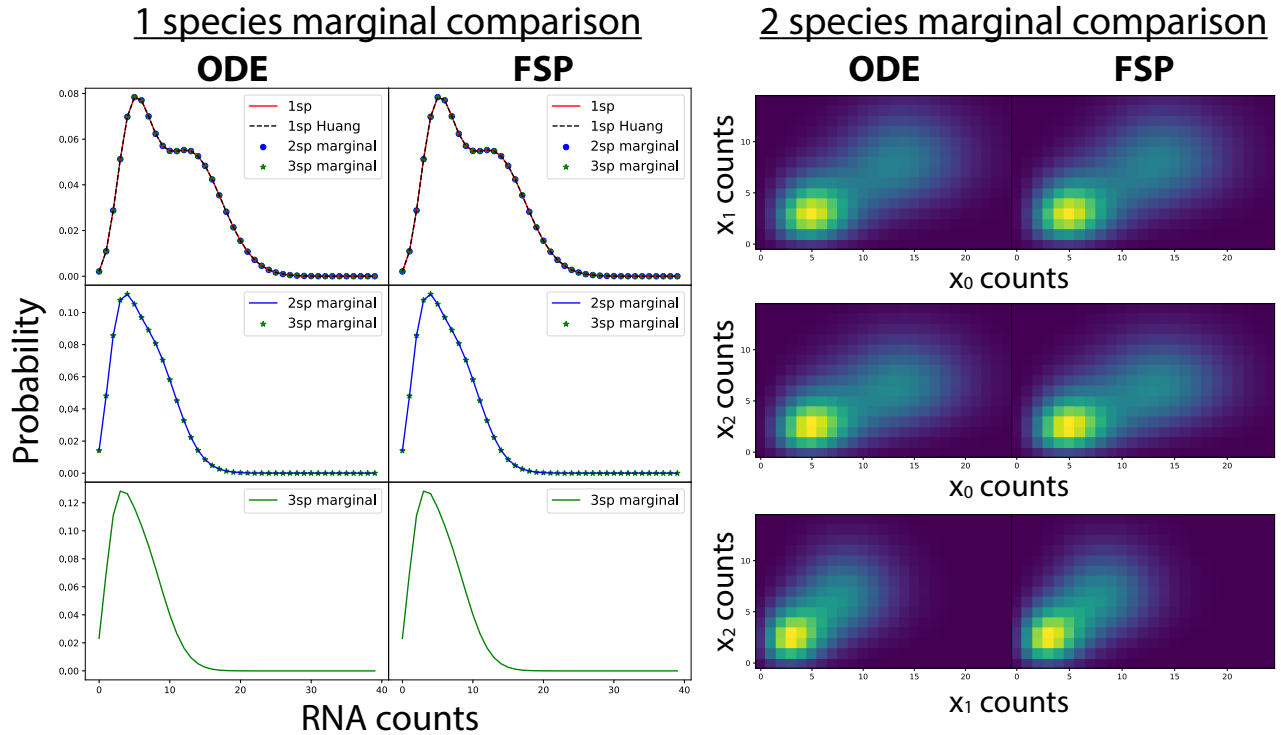
Figure 7.4: Qualitative validation of the ODE approach against FSP. For a representative parameter set, we computed the solution to the three species splicing problem and compared various marginal distributions, finding no significant differences. LEFT: Marginals for the first, second, and third RNA species obtained using both our ODE approach and FSP. RIGHT: Marginals involving two species (x_0, x_1 ; x_0, x_2 ; x_1, x_2) obtained using both our ODE approach and FSP.

better than both competing methods; for some parameter regimes, FSP can take ten times longer to evaluate, and stochastic simulations can take one hundred times longer.

7.5 Parameter inference on simulated data

In this section, we use our CME solver to do fast parameter inference on our models of transcription and splicing (for the one species, two species, and three species models). Because we can only infer non-dimensionalized parameter ratios from steady state data, we fix $\beta_0 = 1$, which is equivalent to working in time units where $\beta_0 = 1$. Because the most common use case of inference on real data will be when one transcription rate is very large and the other is approximately zero, we will fix $\alpha_2 = 0$. This leaves the parameters $\alpha_1, \beta_1, \beta_2, k_{12}, k_{21}$ for
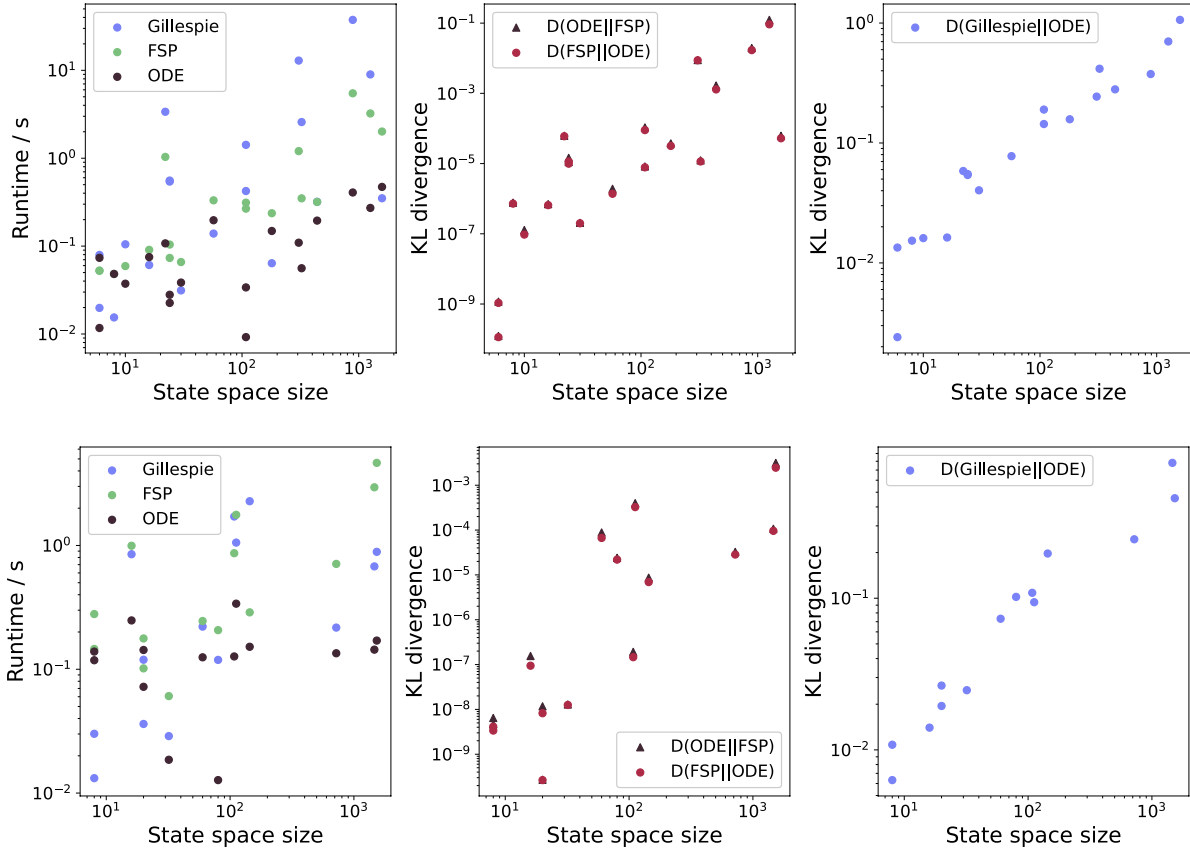
233

Figure 7.5: Performance and accuracy of our ODE-based implementation of the 2 species (1 splicing step) model and 3 species (2 splicing steps) models. The first column shows runtime as a function of state space size for many randomly sampled parameter sets. The second column shows the Kullback-Leibler divergence for those same parameter sets. Where $D(ODE||FSP)$ and $D(FSP||ODE)$ correspond means that the FSP and ODE methods yield equally good approximations of one another. The last column shows KL divergence of Gillespie vs ODE as a function of state space size.

us to infer.

A given fitting procedure began with generating a ground truth parameter set based on randomly sampling from loguniform distributions that spanned three orders of magnitude. Then we simulated data using $N = 1000$ cells using Gillespie's algorithm. In order to fit this data, we used a Kullback-Leibler-divergence-based cost function and a standard gradient descent algorithm. Each cost function evaluation involved using the ODE method to generate a steady state probability distribution. The initial guess was based on a method of moments approach.

Results for the two species model are pictured in Fig. 7.6. The fitted distributions qualitatively agree with the simulated data, and fitted parameters fairly closely match the ground truth parameters. Not all parameters are expected to be equally identifiable. Since
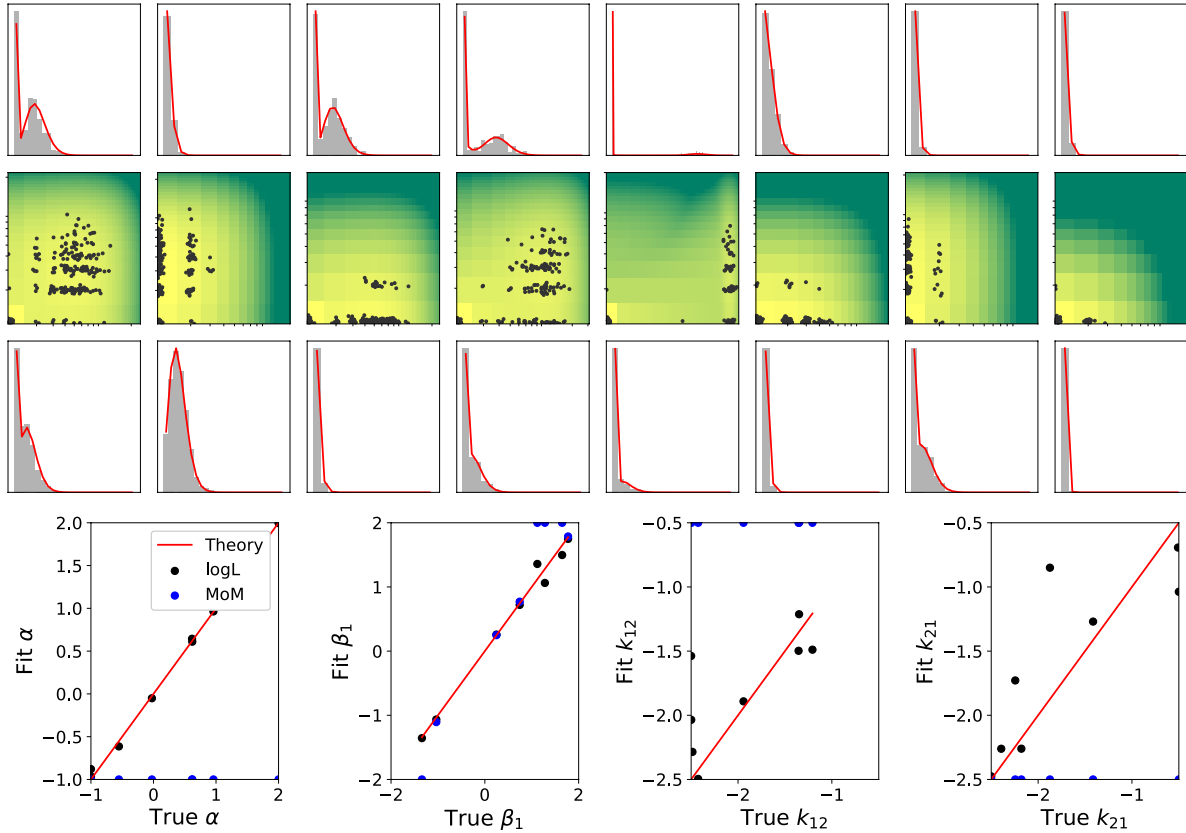
Figure 7.6: Parameter inference on the 2 species (1 splicing step) model given simulated data ($N = 1000$ cells). Top: sample fits. Bottom: inferred parameters vs true parameters. The black dots correspond to the result of inference, while the blue dots correspond to the initial guess obtained via the method of moments.

gene state is usually not observable, from a moments point of view, k_{12} and k_{21} must be extracted from high order moments. This suggests they are harder to identify, and matches what we found.

Results are similar for the three species model (Fig. 7.7), which has a much larger state space. As expected, parameter fits are slightly worse, but still quite close to the ground truth. Overall, the fitting procedure yielded results that were fairly good for both models; however, timing varied greatly between them. Fitting 20 parameter sets took on the order of 15 minutes for the two species model, and on the order of several hours for the three species model. This is mainly due to having to generate on the order of 1000 steady state probability distributions before gradient descent converges.
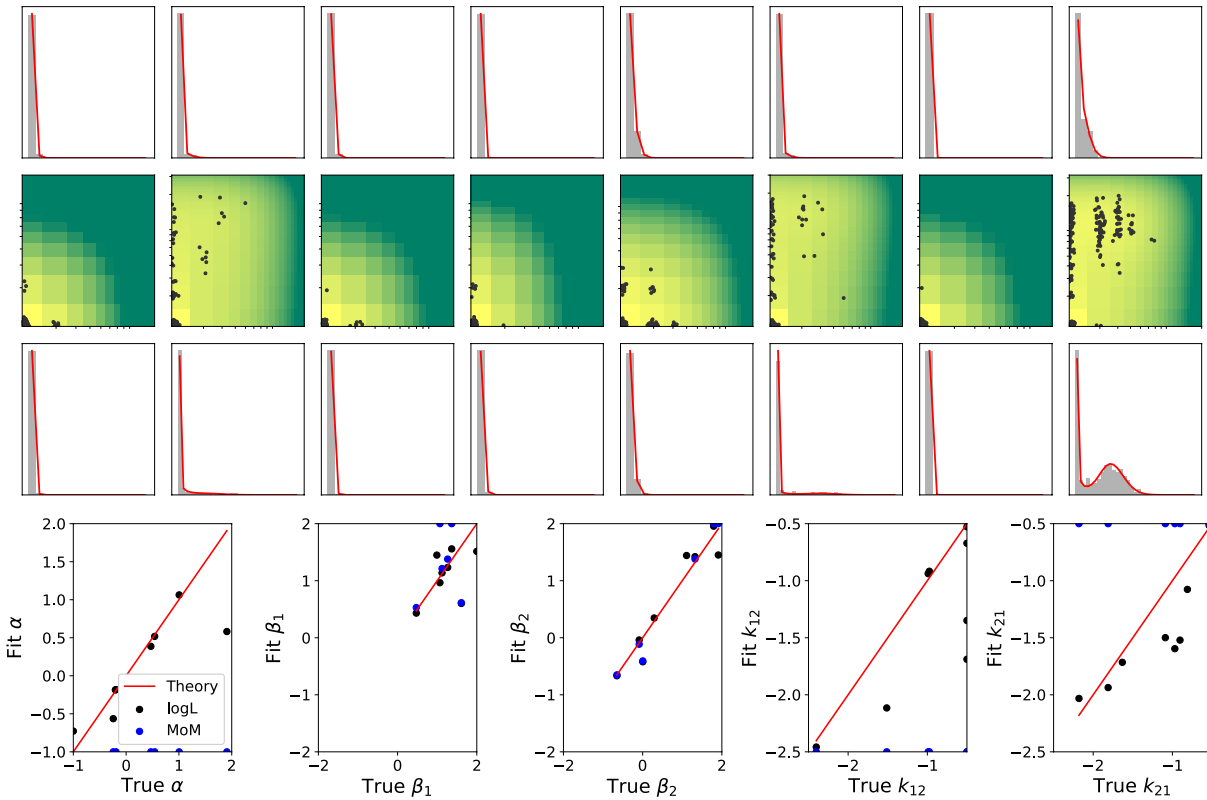
Figure 7.7: Parameter inference on the 3 species (2 splicing steps) model given simulated data ($N = 1000$ cells). Top: sample fits. Bottom: inferred parameters vs true parameters. The black dots correspond to the result of inference, while the blue dots correspond to the initial guess obtained via the method of moments.

7.6 Discussion

Using a state space path integral approach inspired by quantum mechanics, we obtained fast and stable numerical implementations of the steady state solutions of the birth-death-switching and multistep splicing problems, and showed that their performance compares favorably against existing alternatives (e.g. stochastic simulations, other analytic solutions, and FSP). Independently of the derivation of these results, their open source implementations are readily available for biophysicists and quantitative biologists to use for fitting simple models of transcription and splicing dynamics given single cell RNA counts data.

The broad strokes of our path-integral-based technique—deriving ODEs for the steady state generating function, solving them numerically, and then inverse fast Fourier transforming the result—do not sensitively depend on the details of this particular problem. Unlike our results in Part I, this approach seems particularly easy to generalize to more gene states with some arbitrary switching topology. Instead of two coupled ODEs, one would end up with as many ODEs as there are distinct gene states. Whether this kind of extension would be useful remains to be seen, especially since the primary use case of these models is the two state bursty parameter regime. Since one can evaluate these path integrals straightforwardly given first order reactions in general, rather than just the monomolecular reactions we have used here, it may also be possible to adapt this approach to include protein dynamics. On the other hand, the time-dependent terms of the ODEs would no longer be simple sums of exponentials, and their dependence on the $e^{i\theta_j}$ may be harder to determine (or otherwise deal with numerically).

One extension which is certainly possible is generalizing this approach to include splicing topologies where (i) introns can be removed in an arbitrary order and (ii) alternative splicing is allowed. As pointed out in the discussion of Part I, this primarily requires a better understanding of the representation theory of monomolecular reaction networks (that is, determining what replaces the q_j coefficients and $\mathbf{v}^{(k)}$ vectors in our ODEs).

In appraising both this work and the work we did on these problems in Part I, it is interesting to note that there do not seem to be solutions to these problems which are preferable both from a theoretical point of view and a numerical implementation point of view. Rather, our part I solutions are preferable for investigating theoretical properties of the solutions (e.g. computing moments and biologically relevant approximations), while the approach we worked out here is much more suitable for parameter inference.

While we have made some progress on better understanding how to evaluate steady state probability distributions for these problems in a numerically efficient way, we expect to run into additional practical difficulties when fitting real RNA counts data. One generically expects to run into severe identifiability-related problems, e.g. it may be hard to experimentally distinguish RNA with one intron removed from RNA with two introns removed. It may be necessary to compute probability distributions that assume such identifiability issues arise, or else to come up with algorithms for inferring species identity given ambiguous data.

Bibliography

[1] Brian Munsky, Guoliang Li, Zachary R. Fox, Douglas P. Shepherd, and Gregor Neuert. Distribution shapes govern the discovery of predictive models for gene regulation. Proceedings of the National Academy of Sciences, 115(29):7533–7538, 2018.

[2] Sheel Shah, Yodai Takei, Wen Zhou, Eric Lubeck, Jina Yun, Chee-Huat Linus Eng, Noushin Koulena, Christopher Cronin, Christoph Karp, Eric J. Liaw, Mina Amin, and Long Cai. Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell, 174(2):363–376.e16, July 2018.

[3] Heather L. Drexler, Karine Choquet, and L. Stirling Churchman. Splicing Kinetics and Coordination Revealed by Direct Nascent RNA Sequencing through Nanopores. Molecular Cell, 77(5):985–998.e8, March 2020.

[4] Hervé Le Hir, Ajit Nott, and Melissa J. Moore. How introns influence and enhance eukaryotic gene expression. Trends in Biochemical Sciences, 28(4):215 – 220, 2003.

[5] Stefan Stamm, Shani Ben-Ari, Ilona Rafalska, Yesheng Tang, Zhaiyi Zhang, Debra Toiber, T.A. Thanaraj, and Hermona Soreq. Function of alternative splicing. Gene, 344:1 – 20, 2005.

[6] Jennifer C. Long and Javier F. Caceres. The SR protein family of splicing factors: master regulators of gene expression. Biochemical Journal, 417(1):15–27, 12 2008.

[7] Qianliang Wang and Tianshou Zhou. Alternative-splicing-mediated gene expression. Phys. Rev. E, 89:012713, Jan 2014.

[8] Kian Huat Lim, Zhou Han, Hyun Yong Jeon, Jacob Kach, Enxuan Jing, Sebastien Weyn-Vanhentenryck, Mikaela Downs, Anna Corrionero, Raymond Oh, Juergen Scharner, Aditya Venkatesh, Sophina Ji, Gene Liau, Barry Ticho, Huw Nash, and Isabel Aznarez. Antisense oligonucleotide modulation of non-productive alternative splicing upregulates gene expression. Nature Communications, 11(1):3501, Jul 2020.

[9] Jarnail Singh and Richard A Padgett. Rates of in situ transcription and splicing in large human genes. Nature Structural & Molecular Biology, 16(11):1128–1133, November 2009.

[10] Heng Xu, Samuel O. Skinner, Anna Marie Sokac, and Ido Golding. Stochastic Kinetics of Nascent RNA. Physical Review Letters, 117(12):128101, 2016.

[11] Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E. Kastriti, Peter Lönnerberg, Alessandro Furlan, Jean Fan, Lars E. Borm, Zehua Liu, David van Bruggen, Jimin Guo, Xiaoling He, Roger Barker, Erik Sundström, Gonçalo Castelo-Branco, Patrick Cramer, Igor

Adameyko, Sten Linnarsson, and Peter V. Kharchenko. RNA velocity of single cells. Nature, 560(7719):494–498, August 2018.

[12] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. PLOS Computational Biology, 3(10):1–8, 10 2007.

[13] Brian K. Mannakee, Aaron P. Ragsdale, Mark K. Transtrum, and Ryan N. Gutenkunst. Sloppiness and the Geometry of Parameter Space, pages 271–299. Springer International Publishing, Cham, 2016.

[14] Katherine N. Quinn, Heather Wilber, Alex Townsend, and James P. Sethna. Chebyshev approximation and the global geometry of model predictions. Phys. Rev. Lett., 122:158302, Apr 2019.

[15] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of Computational Physics, 22(4):403 – 434, 1976.

[16] Daniel T. Gillespie. Exact stochastic simulation of coupled chemical reactions. The Journal of Physical Chemistry, 81(25):2340–2361, 1977.

[17] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. The Journal of Chemical Physics, 124(4):044104, 2006.

[18] Slaven Peleš, Brian Munsky, and Mustafa Khammash. Reduction and solution of the chemical master equation using time scale separation and finite state projection. The Journal of Chemical Physics, 125(20):204104, 2006.

[19] Zachary Fox, Gregor Neuert, and Brian Munsky. Finite state projection based bounds to compare chemical master equation models using single-cell data. The Journal of Chemical Physics, 145(7):074101, 2016.

[20] John J. Vastola and William R. Holmes. Chemical Langevin equation: A path-integral view of Gillespie's derivation. Phys. Rev. E, 101:032417, Mar 2020.

[21] Srividya Iyer-Biswas, F. Hayot, and C. Jayaprakash. Stochasticity of gene products from transcriptional pulsing. Phys. Rev. E, 79:031911, Mar 2009.

[22] Lifang Huang, Zhanjiang Yuan, Peijiang Liu, and Tianshou Zhou. Effects of promoter leakage on dynamics of gene expression. BMC Systems Biology, 9(1):16, Mar 2015.

Chapter 8

Epilogue

In this thesis, we studied analytic solutions of the chemical master equation—especially using methods inspired by quantum mechanics. We were thoroughly able to understand monomolecular reaction networks, which are in some sense the 'harmonic oscillator' of gene regulation. Fortunately, despite the simplicity of these models, we were able to apply them to learn something about transcription and splicing. We were even able to do this when transcription is coupled to a bursty gene, which yields highly nontrivial stochastic dynamics.

The state space path integral, we have found, is a highly versatile tool for solving chemical master equations, and it has been used to construct some of the most general solutions yet known. Still, there is an unsatisfying fact that may be nagging at the back of our minds: *what about more realistic systems?*

We have not treated systems involving binding (e.g. $A + B \to C$) or translation ($m \to m + p$). Can the tools discussed in this thesis handle those problems? Some preliminary work suggests that the answer is a definite *yes* for systems involving first order reactions like translation. Heuristically, this is because systems involving only first order reactions yield path integrals that can be solved using the 'backwards ODE' trick we have used in a number of places. See Chapter 4 for some discussion of this. The eigenvalues/spectrum of these systems also seems to be linearly spaced, which means a ladder operator approach should work in principle, although the form of these ladder operators (in terms of more basic creation and annihilation operators) may be quite complicated. Generically, it seems that ladder operators for first order reaction networks may involve infinite sums of creation and annihilation operators, unlike what we have dealt with in the case of monomolecular processes.

Allowing binding reactions yields a much more difficult problem. The spectrum generically goes like n^2 instead of n, and ladder operator approaches probably do not work without significant modification. Traditional approaches, like the method of characteristics, do not work. Path integral methods may have potential, since the Gaussian integrals that appear due binding (e.g. $X + X \to X$ would contribute a term like x^2 in the Doi-Peliti action) should be solvable. But there are some technical details that prevent us from having done so already. Our main task here would be to create some analogue of the backwards ODE trick (and hence reduce the computation of many integrals to just solving ODEs), which does not work when there are binding reactions.

I suspect that, in the same way understanding in detail how to solve the birth-death process lead to understanding the solution structure of monomolecular processes, thoroughly studying a toy model involving translation or binding might lead to a general solution. It would be quite theoretically satisfying to construct a solution of the CME for systems involving any combination of monomolecular and bimolecular reactions: as we discussed in

Chapter 2, this is essentially *all* biological systems of interest! However, such a solution would almost certainly be impractical for numerical implementation; we will still need simulations. Even if it exists, it may be highly unstable, and exhibit some analogue of chaos (so that small changes in parameters dramatically change the probability distribution).

On the other hand, such a closed form solution may not exist. Such solutions generically do not exist for interacting quantum field theories; why should they exist in this mathematically analogous case?

If one desires efficient numerical implementations for some practical purpose, like parameter inference, we have found that full CME solutions may not actually be particularly helpful. What *is* helpful is *almost* solving the CME, and leaving its solution (or, more typically, the solution of its generating function) in terms of some ODE which can be solved numerically. This approach combines the advantages of understanding a problem's solution structure with the stability associated with numerically integrating ODEs, and should be explored further. Such approaches may offer promising results when combined with other numerical solution methods, like simulations or finite state projection (for example, one can imagine solving certain subsystems using the ODE approach, and combining the obtained results with simulation results for a different subsystem).

We did not apply our methods to understand real data in this thesis, partly because data involving distributions of partly spliced RNA is presently hard to come by. But the data is coming. Quantitative single cell experiments continue to improve, and there is a real need for theory that will permit us to reckon with their results. Biology has advanced tremendously over the past decade, and we have especially improved our ability to do sensitive high-throughput experiments and develop sophisticated algorithms for analyzing those experiments. But it should be remembered that, as well as experiments and algorithms, we need ideas.

We can measure and categorize as many cells as we like, so that we understand the 'what' questions. *What is happening* inside a cell when it undergoes a cell fate transition? But 'big picture' aims of understanding these systems, like the robust control of cell fate for applications in regenerative medicine, will remain elusive until we have a better handle on different questions. How should we *think* about what happens inside a cell when it undergoes a cell fate transition?

There *is* a need for theory. Exactly what form it should take to be most useful, I do not know. Perhaps the CME does not offer the right foundation for thinking about stochastic gene regulation. But there is a need for theoretical understanding of *some kind*. Something that goes beyond what *is*, and suggests what *could be*. Something that suggests how to *think* and *compute*. Something that tells us what the wrong but useful models of biological systems are.

How should we *think* about one kind of cell turning into another kind of cell? How should we *think* about cellular decision-making? How should we *think* about the relationship between descriptions of these systems at different scales? How should we *think* about emergent features?

For now, I can only ask the questions.

Chapter 9

Postscript: Who's Afraid of Max Delbrück?

Labels say nothing about nature, but they do say something about the people studying it. In practice, whether one calls oneself a physicist, biologist, or something else is thought to have real sociological consequences—for getting academic jobs, getting funded, and getting published, among other things. These questions of academic identity can become particularly conspicuous for those researchers that work on the boundary of multiple disciplines, who may for professional reasons feel the need to assert that they are a member of one tribe over another.

As Martin notes in an account of the recent rise of condensed-matter physics, its success "is a story of categories and why they matter." Identifying condensed-matter physics as a distinct subdiscipline "made a statement about the type of activity physics was supposed to be", and "drew a line between who was a physicist and who wasn't". Even the rebranding of "solid-state physics" to "condensed-matter physics" in the 1960s and 1970s was an effort to emphasize the intellectual value and conceptual unity of the field as much as it was an attempt to accurately describe it [1]. Nowadays, with some flavor of condensed-matter physics being the dominant activity of most working physicists, there is no longer any real dispute regarding whether it counts as legitimate physics.

The story may be repeating again on the interface of physics and biology. In 2020, the National Academies explicitly included "Biological Physics/The Physics of Living Systems" for the first time in its decadal survey of US research [2], a sign that it is increasingly thought of as truly a part of physics, rather than just physics applied to an outside subject. With the confluence of recent breakthroughs in experimental techniques (e.g. CRISPR), the high-throughput collection of multi-omic data, and the developing sophistication of algorithms for 'big data'-style analyses, the field is poised for explosive growth in the twenty-first century.

But there remains significant confusion over the nature of the field, which is in part reflected by the diversity of names in use. Here are a few:

> biological physics, physics of living systems, biophysics, physical biology, living state physics, living matter physics

The oldest of these terms is "biophysics", which was proposed by Karl Pearson in his 1892 book *The Grammar of Science*. On the need to understand "the genesis of the living from the lifeless", he writes [3]:

> "A branch of science is therefore needed dealing with the application of the laws of inorganic phenomena, or Physics, to the development of organic forms. This branch of science which endeavors to show that the facts of *Biology* . . . constitute particular cases of general physical laws has been termed *Ætiology*.

It would perhaps be better to call it *Bio-physics*. This science does not appear to have advanced very far at present, but it not improbably has an important future."

By 1920, the challenge of creating such a field seems to have been met, with Alexander Forbes outlining his view of biophysics in a *Science* article and reporting the development of a biophysics course at Harvard [4]. The next term to appear may be "physical biology", which seems to have been introduced by Walter Porstmann in a 1915 article [5]:

"It will be the function of this new branch of science to investigate biological phenomena as regards their physical aspects, just as Physical Chemistry has treated the physical aspects of chemical phenomena. Because this field has not yet been systematically explored . . . the individual data of *Physical Biology* appear, as yet, as more or less disconnected facts . . . As results gathered in this disconnected fashion accumulate, the need of their unification into a harmonious whole, into a distinct discipline of science, becomes more and more acutely felt."

In 1925, the first textbook on physical biology—Alfred Lotka's (of Lotka-Volterra fame) *Elements of Physical Biology*—appeared. In it, he more than once distinguishes the project of physical biology from that of biophysics, citing Forbes' article and describing biophysics as studying "the physics of individual life processes" (e.g. nerve conduction), and physical biology as the application of physical principles to biology in a broader sense (and seemed to have in mind modeling evolutionary dynamics as an example). "Physical biology would, in this terminology, include biophysics as a subordinate province", noted Lotka [6].

Being particular about labels was by no means exclusive to Lotka. Decades later, after Nicolas Rashevsky spent a lifetime advancing what he called a "mathematical biophysics" analogous to the more familiar mathematical physics, Rashevsky was invited to give a prestigious lecture at the first Gordon Research Conference on "biomathematics", which was to be held in 1965. Recoiling at the term, Rashevsky wrote in a letter to the conference organizer that it was an "etymological monstrosity" that gave the wrong sense of what the field was about [7].

In the late 1960s and early 1970s, many physicists studying biological problems faced "skepticism when presenting . . . at the Biophysical Society Meetings"; these mounting frustrations led directly to the 1973 establishment of the American Physical Society's Division of Biological Physics [8]. One speculates that "biological physics" may have been chosen to distinguish the organization from the culture of the Biophysical Society and "biophysics", which (then and now) has different connotations.

All of this is just to say that labels seem to have mattered historically to many of the workers on the interface of physics and biology. Even now, the connotations of terms like "biological physics" and "biophysics" differ for a variety of historical and sociological reasons. The recently popularity of the "physics of living systems" label within the physics community may be analogous to the "condensed-matter" rebranding of "solid-state physics", its use intending to situate the discipline as an activity within physics rather than outside or adjacent to it.

For the remainder of this essay, let us put aside the business of what this field might be called. Regardless of its name, to what ethos should it aspire? In search of a spirit of biological physics, it may be appropriate to reflect on the diverging experiences of two physicists that turned to biology: Max Delbrück and Nicolas Rashevsky.

The case of Max Delbrück

Max Delbrück was born in Berlin in 1906, and matured amidst the scarcity and social upheaval associated with the first World War. Although he began his university studies interested in astronomy, exciting developments in quantum mechanics and the lamentable state of German astronomy led him to pursue theoretical physics instead [9].

Delbrück completed his PhD thesis on a quantum mechanical treatment of the lithium molecule in 1930, and continued to work on topics related to quantum mechanics until around 1937. He was accepted by the 'insiders' of physics, and wrote papers with luminaries like Lise Meitner and George Gamow. The extent to which he was intimately familiar with the movers and shakers of the day may be illustrated by the time he jokingly served as a butler at a costume party thrown by Schrödinger [10].

But Delbrück's primary research interests were not to remain in physics. Captivated by a lecture given by Bohr in 1932 entitled "Light and Life", Delbrück took very seriously Bohr's suggestion that there may be some manifestation of his complementarity principle in biology. In particular, an organism can be viewed "either as a living organism or as a jumble of molecules", just as an electron can be viewed either as a particle or a wave; perhaps there is some "mutually exclusive feature" of each description analogous to position and momentum in quantum mechanics [9].

While working as a theoretical physicist for Lise Meitner, Delbrück moonlighted as a biologist, and in 1935 published his first biology-related paper. In the 1940s, following a Rockefeller Foundation-supported mutation research fellowship at Caltech, and with some continued support from the foundation, Delbrück's career in biology blossomed [10]. Over the course of the decade, he published a landmark paper with Salvador Luria establishing that Darwinian selection applies to bacteria; helped build the Phage Group, an intellectual school and forebear of much of modern molecular biology; and began the legendary phage course at Cold Spring Harbor Laboratory, intending to "spread the new gospel" [11] of bacteriophage work. For his community building and contributions to genetics, he received a third of the 1969 Nobel Prize in Physiology or Medicine.

During the war, he was an instructor of physics at Vanderbilt University, but had a rudimentary research lab in the biology department. While the arrangement was strange for the time, Delbrück's sterling physics credentials and the quality of his research output yielded no complaints. In 1947, he left Vanderbilt to become a Professor of Biology at Caltech, where he remained for the rest of his career [10].

The case of Nicolas Rashevsky

Nicolas Rashevsky was born in 1899 in Chernigov, a small village in Ukraine. He was well-educated—in literature and languages, for example—and learned voraciously. Already by the age of 19, he had a PhD in theoretical physics from the University of Kiev. Afterwards, Rashevsky was rapidly uprooted by the Russian Revolution, and by 1925 had moved to the United States to work as a theoretical physicist at Westinghouse Electric Company.

Rashevsky did most of his work in theoretical physics in the 1920s, publishing on a variety of fields: quantum mechanics, relativity, photomagnetism, thermionic emission, and colloidal dynamics, to name a few. But a chance encounter at a "social occasion" with a biologist, who described the reason for cell division as a mystery, piqued his interest in biological problems. Soon enough, Rashevsky—reflecting on the absence of a biology analogue to the then-well-established field of mathematical physics—in 1926 began dreaming of a similar 'mathematization' of biology [7, 12].

In the 1930s, Rashevsky concentrated his efforts fully on this project, and with Rockefeller Foundation support found a position at the University of Chicago. He published on a variety of problems, including nerve conduction, cell growth, and development. His first book-length examination of "mathematical biophysics" as he envisioned it, *Mathematical Biophysics: Physicomathematical Foundations of Biology*, was released in 1938 [13].

At the same time, Rashevsky ran into mounting obstacles. Facing a chilly reception in mainstream biology journals, in 1939 he started the Bulletin of Mathematical Biophysics, which housed the majority of his articles for the remainder of his career. Over the years, he fought to maintain funding for the intellectual school he had founded at the University of Chicago (the Committee on Mathematical Biology), for mathematical biology to receive official recognition from the National Institutes of Health, and to maintain the independence of his group from meddlesome university administrators. In many of these battles, he was thwarted, or only partially successful. By the time of his death, he left a mixed scientific legacy [7].

In some ways, the lives of Delbrück and Rashevsky were strikingly parallel. Both grew up against a backdrop of war, both were trained as theoretical physicists and switched to thinking primarily about biological problems in the 1930s, both received significant financial support from the Rockefeller Foundation, and both were community builders and prolific researchers. But Delbrück met mainstream acceptance, while Rashevsky encountered obstacle after obstacle. Why?

Perhaps Delbrück was luckier, or had better connections. But one can readily identify two marked differences in scientific philosophy that may have impacted their trajectories. Firstly, Delbrück thought of himself more as a natural philosopher than a physicist or biologist [14], whereas Rashevsky adhered strictly to the ethos of theoretical physics. Rashevsky was interested in the 'mathematization' of biology not just for its practical utility, but for its own sake [15]. Crudely, one might say that Delbrück pursued questions, while Rashevsky

pursued an identity. Secondly, Delbrück was more willing to meet biologists halfway by doing experiments and assiduously checking his hypotheses against the experimental results of others.

There is something to be said for Rashevsky's iron will, and his sturdy commitment to his scientific vision in spite of all opposition. He was a trailblazer in the establishment of mathematical biology, and his efforts directly led to the founding of the Society for Mathematical Biology [7]. But his dogmatic approach may have been alienating to many biologists.

A play by American playwright Edward Albee famously had the provocative title *Who's Afraid of Virginia Woolf?*. Virginia Woolf was an English novelist whose plays often featured themes of people in British high society living inauthentic lives under many masks. The title of the play, in one interpretation, means: 'Who's afraid to live without illusions?' [16]

Max Delbrück never seemed to worry about whether he was called a physicist or a biologist. In fact, he seemed to delight in using one label or the other as it suited him. He saw himself more akin to a natural philosopher, someone who pursues the ultimate truths about the natural world regardless of what tools or paradigms their answers require. As we continue to pursue biological physics, it may be prudent to emulate that style, and courageously pursue interesting questions whether or not they are properly considered physics at the time—to act as if we live in a world without disciplinary boundaries. Who's afraid of Max Delbrück?

Bibliography

[1] Joseph D. Martin. When condensed-matter physics became king. Physics Today, 72(1):30, 2019.

[2] Biological Physics/Physics of Living Systems: A Decadal Survey. https://www.nationalacademies.org/our-work/biological-physicsphysics-of-living-systems-a-decadal-survey. Accessed: 2020-09-01.

[3] Karl Pearson. The Grammar of Science. A. and C. Black, 1911.

[4] Alexander Forbes. Biophysics. Science, 52(1345):331–332, 1920.

[5] Walter Porstmann. Rundschau. Ein Problem aus der physikalischen Zoologie: Einfluss physikalischer Momente auf die Gestalt der Fische. Prometheus. Illustrierte Wochenschrift über die Fortschritte in Gewerbe, Industrie und Wissenschaft, 26:1317–1319, 1915.

[6] Alfred J. Lotka. Elements of Physical Biology. William & Wilkins Company, 1925.

[7] Maya M. Shmailov. Intellectual Pursuits of Nicolas Rashevsky. Springer, 2016.

[8] Abigail Dove. APS Membership Unit Profile: The Division of Biological Physics. APS News, 27(11), 2018.

[9] Max Delbrück. Interview with Max Delbrück. Caltech Oral Histories, 1979.

[10] Robert T Lagemann and Wendell G Holladay. To Quarks and Quasars: A History of Physics and Astronomy at Vanderbilt University. Vanderbilt University. Dept. of Physics and Astronomy, 2000.

[11] Gunther Stent, James D Watson, and J Cairns. Phage and the Origins of Molecular Biology. Cold Springs Harbor Laboratory of Quantitative Biology, New York, 1966.

[12] Tara H. Abraham. Nicolas Rashevsky's Mathematical Biophysics. Journal of the History of Biology, 37(2):333–385, 2004.

[13] Nicolas Rashevsky et al. Mathematical Biophysics. University of Chicago Press, 1938.

[14] Lily E. Kay. Conceptual models and analytical tools: The biology of physicist Max Delbrück. Journal of the History of Biology, 18(2):207–246, Jun 1985.

[15] Nicolas Rashevsky. Foundations of Mathematical Biophysics. Philosophy of Science, 1(2):176–196, 1934.

[16] William Flanagan. Edward Albee, The Art of Theater No. 4. The Paris Review, 39, 1966.