Reaction-Diffusion Systems in and out of Equilibrium - Methods for Simulation and Inference

Dissertation
zur Erlangung des Grades eines Doktors der Naturwissenschaften

am Fachbereich Physik der Freien Universität Berlin

vorgelegt von

Christoph Fröhner

Berlin, 2019

Erstgutachter Prof. Dr. Frank Noé

Zweitgutachter Prof. Dr. Roland Netz

Tag der Disputation 16.04.2020

Selbstständigkeitserklärung

Name: Fröhner

Vorname: Christoph

Ich erkläre gegenüber der Freien Universität Berlin, dass ich die vorliegende Dissertation selbstständig und ohne Benutzung anderer als der angegebenen Quellen und Hilfsmittel angefertigt habe. Die vorliegende Arbeit ist frei von Plagiaten. Alle Ausführungen, die wörtlich oder inhaltlich aus anderen Schriften entnommen sind, habe ich als solche kenntlich gemacht. Diese Dissertation wurde in gleicher oder ähnlicher Form noch in keinem früheren Promotionsverfahren eingereicht.

Abstract

Reaction-diffusion methods allow treatment of mesoscopic dynamic phenomena of soft condensed matter especially in the context of cellular biology. Macromolecules such as proteins consist of thousands of atoms, in reaction-diffusion models their interaction is described by effective dynamics with much fewer degrees of freedom. Reaction-diffusion methods can be categorized by the spatial and temporal length-scales involved and the amount of molecules, e.g. classical reaction kinetics are macroscopic equations for fast diffusion and many molecules described by average concentrations. The focus of this work however is interacting-particle reaction-dynamics (iPRD), which operates on length scales of few nanometers and time scales of nanoseconds, where proteins can be represented by coarse-grained beads, that interact via effective potentials and undergo reactions upon encounter. In practice these systems are often studied using time-stepping computer simulations. Reactions in such iPRD simulations are discrete events which rapidly interchange beads, e.g. in the scheme A + B$\rightleftharpoons$C the two interacting particles A and B will be replaced by a C complex and vice-versa. Such reactions in combination with the interaction potentials pose two practical problems: *i)* To achieve a well defined state of equilibrium, it is of vital importance that the reaction transitions obey microscopic reversiblity (detailed balance). *ii)* The mean rate of a bimolecular association reaction changes when the particles interact via a pair-potential. In this work the first question is answered both theoretically and algorithmically. Theoretically by formulating the state of equilibrium for a closed iPRD system and the requirements for detailed balance. Algorithmically by implementing the detailed balance reaction scheme in a publicly available simulator ReaDDy 2 for iPRD systems. The second question is answered by deriving concrete formulae for the macroscopic reaction rate as a function of the intrinsic parameters for the Doi reaction model subject to pair interactions. Especially this work addresses two important scenarios: Reversible reactions in a closed container and irreversible bimolecular reactions in the diffusion-influenced regime.

A characteristic of reactions occurring in cellular environments is that the number of species involved in a physiological response is very large. Unveiling the network of necessary reactions is a task that can be addressed by a data-driven approach. In particular, analyzing observation data of such processes can be used to learn the important governing dynamics. This work gives an overview of the inference of dynamical reactive systems for the different reaction-diffusion models. For the case of reaction kinetics a method called Reactive Sparse Identification of Nonlinear Dynamics (Reactive SINDy) is developed that allows to obtain a sparse reaction network out of candidate reactions from time-series observations of molecule concentrations.

Zusammenfassung

Reaktionsdiffusionsverfahren ermöglichen die Behandlung mesoskopischer dynamischer Phänomene von weicher kondensierter Materie, insbesondere im Kontext der Zellbiologie. Makromoleküle wie Proteine bestehen aus Tausenden von Atomen. In Reaktions-Diffusions-Modellen wird ihre Wechselwirkung durch effektive Dynamik mit wesentlich weniger Freiheitsgraden beschrieben. Reaktionsdiffusionsverfahren können nach den beteiligten räumlichen und zeitlichen Längenskalen und der Menge der Moleküle klassifiziert werden. Zum Beispiel ist klassische Reaktionskinetik definiert durch makroskopische Gleichungen im Limit schneller Diffusion und vieler Moleküle. Die Menge der Moleküle wird dann als mittlere Konzentrationen dargestellt. Der Schwerpunkt dieser Arbeit liegt jedoch auf der Reaktionsdynamik interagierender Teilchen (iPRD), die auf Längen- und Zeitskalen von wenigen Nanometern und Nanosekunden arbeitet. Hier können Proteine durch grobkörnige Kügelchen dargestellt werden, die über effektive Potentiale interagieren und Reaktionen bei Kontakt eingehen. In der Praxis werden diese Systeme häufig mithilfe von Computersimulationen untersucht, die die Zeit in endlichen Zeitschritten inkrementieren. Reaktionen in solchen iPRD-Simulationen sind diskrete Ereignisse, welche Teilchen instantan austauschen. Zum Beispiel im Schema $A + B \rightleftharpoons C$ werden die beiden wechselwirkenden Teilchen A und B durch einen C-Komplex ersetzt und umgekehrt. Solche Reaktionen in Kombination mit den Wechselwirkungspotentialen werfen zwei praktische Probleme auf: *i)* Um einen genau definierten Gleichgewichtszustand zu erreichen, ist es von entscheidender Bedeutung, dass die Reaktionsübergänge mikroskopisch reversibel sind (detailliertes Gleichgewicht). *ii)* Die mittlere Geschwindigkeit einer bimolekularen Assoziationsreaktion ändert sich, wenn die Partikel über ein Paarpotential interagieren. In dieser Arbeit wird die erste Frage sowohl theoretisch als auch algorithmisch beantwortet: theoretisch durch die Formulierung des Gleichgewichtszustands für ein geschlossenes iPRD-System und Aufstellen der Anforderungen für detailliertes Gleichgewicht, algorithmisch durch die Implementierung des reversiblen Reaktionsschemas in einer öffentlich verfügbaren Simulationsbibliothek ReaDDy 2 für iPRD-Systeme. Zur Beantwortung der zweiten Frage werden konkrete Formeln für die makroskopische Reaktionsrate in Abhängigkeit von den intrinsischen Parametern für das Doi-Reaktionsmodell abgeleitet, bei denen Paarwechselwirkungen auftreten. Insbesondere werden in dieser Arbeit zwei wichtige Szenarien angesprochen: Reversible Reaktionen in einem geschlossenen Behälter und irreversible bimolekulare Reaktionen im diffusions-beeinflussten Regime.

Ein Merkmal von Reaktionen, die in zellulären Umgebungen auftreten, ist, dass die Anzahl der molekularen Spezies, die an einem physiologischen Phänomen beteiligt sind, sehr groß ist. Das Netzwerk notwendiger Reaktionen aufzudecken, ist eine Aufgabe, die durch einen datengetriebenen Ansatz gelöst werden kann. Insbesondere das Analysieren von Beobachtungsdaten solcher Prozesse kann verwendet werden, um die essentielle bestimmende Dynamik zu lernen. Diese Arbeit gibt einen Überblick über die Inferenz dynamischer reaktiver Systeme für die verschiedenen Reaktionsdiffusionsmodelle. Für den Fall der klassischen Reaktionskinetik wird eine Methode namens Reactive Sparse Identification of Nonlinear Dynamics (Reactive SINDy) entwickelt, die es ermöglicht, aus Zeitreihenbeobachtungen von Molekülkonzentrationen ein spärliches Reaktionsnetzwerk aus Kandidatenreaktionen zu erhalten.

Acknowledgements

I would like to thank Prof. Dr. Frank Noé for the opportunity of doing research in his lab. I want to thank Prof. Dr. Roland Netz whose “Computerphysik” and “Statistical Physics” lectures initially sparked my interest for theoretical statistical physics and simulation methods. I thank all colleagues and friends from the Noé lab, the Biocomputing group, and the ZIB.

I thank my physics teacher, the late Stephan Kirsch for teaching me the basics of the only subject that really ever interested me in school.

I want to thank Sven Hartmann, Sebastian Baum, Hansjochen Köckert, Sarah Meergans, Linda Swierkosz, Christian Böhnke, Michael Hellwig, Tobias Jankowski, Eric Drägerdt, Norbert Kähler, Tom Gibhardt, Christian Vogel, Dominik Hagin, Gabriel Wieland, Andreas Benjamin Christopher Wilde, Cherno Drammeh, Stephan Rath, Nicole Kwasny.

I thank my family for always being supportive in every respect.

Most importantly I thank my loving wife Laura for moral support throughout the years. I thank my daughter Helena whose laughter is the greatest joy I have experienced in life.

List of Publications

The results of this work were published in

- Christoph Fröhner and Frank Noé. "Reversible Interacting-Particle Reaction Dynamics". In: *The Journal of Physical Chemistry B* 122.49 (2018), pp. 11240–11250. DOI: 10.1021/acs.jpcb.8b06981
- Manuel Dibak, Christoph Fröhner, Frank Noé and Felix Höfling. "Diffusion-influenced reaction rates in the presence of pair interactions". In: *The Journal of Chemical Physics* 151.16 (2019), p. 164105. DOI: 10.1063/1.5124728
- Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics". In: *PLoS Computational Biology* 15.2 (2019), e1006830. DOI: 10.1371/journal.pcbi.1006830
- Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "Reactive SINDy: Discovering governing reactions from concentration data". In: *The Journal of chemical physics* 150.2 (2019), p. 025101. DOI: 10.1063/1.5066099

The contribution of the author Christoph Fröhner to these results is clarified in the respective chapters.

Contents

Chapter 1

Introduction

A majority of cellular processes can be described by reactions initiated by macromolecules [Alb+08]: genetic information is stored in long polynucleotide chains and is processed by proteins, kinases catalyze phosphorylation of other functioning proteins [ZL02; Hat+10], binding pockets open and close both spontaneously and externally induced [KM83], membranes form interfaces and separate functionally different compartments of the cell [Lip95], cytoskeletons are constructed and maintained [LA17], hydrogenases catalyze reversible oxidation of hydrogen [Har+18]. All these functions are performed under the consumption of free energy, which means that living organisms are chemical factories that operate far from equilibrium. Understanding these metabolic, and signaling pathways [Bar02] is fundamental not only in medical applications, but also for a wider field of soft matter, e.g. catalytic functions enabled by the self-assembly of macromolecules into functional superstructures [PBV09], or by core-shell nanoreactors [Gal+16; Roa+17].

Many of the mentioned processes occur in crowded environments where a well-mixing cannot be assumed. Thus microscopic insights from an experimental, theoretical and simulation point of view are required to resolve the mechanics of the underlying processes. However the number of involved molecular species and hence the resulting number of inter-dependencies is very large. Additionally the spatial- and temporal scales between the microscopic dynamics and the resulting macroscopic behavior are very different. A hierarchical multi-scalar approach needs to be put to work, that bridges the atomistic physical descriptions and the system's biology. Reaction-diffusion methods are predestined for this mesoscopic task of studying effective, coarse-grained behavior.

Consider the example of the two proteins barnase and barstar [SF93] associating in a cellular environment. Barnase is a bacterial protein which damages the cell's ribonucleic acids (RNA). Usually barnase is inhibited by another protein barstar. Both proteins form an extremely stable complex [Pla+17]. The stationary ratio of finding the system in the bound state versus the unbound state is largely determined by which configuration is energetically more favorable. This also means that e.g. changing the ionic strength of the solvent can enhance or weaken the interaction of the two molecules and thus change the stationary state [Vij+98]. However the association is so rapid, that the process is strongly diffusion-influenced, i.e. the time-dependent rate of association is determined by the frequency of encounter of the two molecules. Also the time-dependent association kinetics of many such molecules in solution depends very strongly on the mobility of the reactants which might be hindered by crowding agents or geometric confinements like cellular organelles. An effective model of the two proteins should take the different time-scales of diffusion and formation of the complex into account.

How does one systematically construct coarse-grained descriptions from atomistic descriptions? Typically one has to find a mapping from the atomistic degrees of freedom x at a given time to a much simpler coordinate $y = \xi(x)$ via a narrowing function ξ, i.e. x is a vector with much more elements than

y. For the example of two molecules A and B reversibly associating in solution $\mathrm{A} + \mathrm{B} \rightleftarrows \mathrm{AB}$, think of y as a reaction coordinate that clearly distinguishes the bound and unbound state. The mapping ξ induces a free energy F (or potential of mean force) via $e^{-\beta F(y)} = \int e^{-\beta U(x)} \delta(\xi(x) - y) \mathrm{d}x$ where $\beta^{-1} = k_B T$ is the thermal energy with the Boltzmann constant k_B and the temperature T, and $U(x)$ is the potential energy landscape in the coordinate x. The definition of F assures that the statistics in y achieve the same stationary ratio of bound versus unbound as the original coordinate x, this is called thermodynamic consistency. If there is a high barrier in the landscape of $F(y)$ that distinguishes bound and unbound state, then we can formulate a much simpler memory-less process on the coordinate y which encompasses all the important and slow dynamics of the original dynamics in x. This is exemplified by Bicout and Szabo [BS08], where a high barrier in the potential of mean force leads to a memory-less overall kinetics in the low dimensional representation. When the diffusion time-scales of encounter are a limiting factor, then the geometry of the environment and the interaction of the molecules have an impact on the time-dependent kinetics. That means to obtain a memory-less process for the association kinetics the diffusion in the unbound state must be resolved spatially, but the formation after the encounter is well described by a Markov jump process into the bound state. This directly leads to a coarse-grained reaction-diffusion model, where the two molecules A and B may be represented as particles that have to encounter via diffusion up to a certain distance and can then undergo the reaction. Such models are called particle-based reaction-diffusion (PBRD) or interacting-particle reaction-dynamics (iPRD). The latter highlights the fact that particles diffuse subject to interaction potentials. Such reaction-diffusion models have much fewer parameters than the original atomistic description. For example a protein which may contain thousands of atoms is effectively described by one or several connected beads. Hence we can study a system of several of these molecules and their interplay, e.g. via computer simulations.

There are several different reaction-diffusion approaches apart from PBRD or iPRD to model molecular systems. Each one is suited well for a particular spatio-temporal scale and a particle-number scale as will be described in Section 1.3.

1.1 Interacting Particle Reaction Dynamics (iPRD)

Interacting Particle Reaction Dynamics (iPRD) governs the time-evolution of diffusion and reactions of coarse-grained beads. Among available reaction-diffusion models iPRD possesses the highest level of detail. The motion of macromolecules in cellular environments is classically governed by Newton's equation of motion in the presence of all other atoms, specifically all solvent atoms. A macromolecule collides very frequently with the solvent atoms. In fact so frequent and short-lived, that its effect is very much stochastic. In the coarse description, the time-scales of interest are much larger than the time-scales of the collisions in the all-atom descriptions. When the macromolecules in aqueous environments are represented as individual beads, the effect of the solvent is well represented by a stochastic force. The motion becomes diffusive. Hence the coarse-grained beads diffuse in three-dimensional space subject to potentials, which accounts for confined geometries like cellular compartments, or pair interactions like steric repulsion or short ranged screened electrostatics. Beads can change their type (or species) via reactions. These can represent chemical reactions like phosphorylation or protein-ligand binding or configurational changes in the macromolecules. Unimolecular reactions of the type $\mathrm{A} \longrightarrow \mathrm{B}$ occur spontaneously while bimolecular reactions like $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C}$ occur only when A and B encounter.

Observables of iPRD system are often not calculated from the governing equations but by simulation of the process [AT17; HFN19; And17; DYK18; KHU19; HFE05; ZW05; ML16]. Many techniques known from molecular dynamics simulation can also be applied to iPRD. The main differences are that transport is usually diffusive in iPRD and chemical reactions are not modeled on an atomic level, but are discrete events that add particles to- or remove particles from the system.

1.1.1 Diffusion

In 1909 Jean Perrin experimentally verified [Per10] Einstein's quantitative theory [Ein05] about the movement of small particles suspended in liquids (e.g. water) which seem static at large scales but are mobile on a microscopic scale. There is a chaotic motion that results from impacts of the particles with the molecules of the solution. The motion of all molecules within a glass of water at room temperature has far too many degrees of freedom to efficiently describe it by a set of equations of motion. But the sheer number of molecules and thus the high frequency of collisions allows to understand the so called Brownian motion – first observed by Robert Brown in 1827 – of the colloid particles as a stochastic process whose time evolution is described by the Langevin equation. The latter is a Newtonian equation of motion for the velocity v of a free particle with mass m subject to a dissipative force and a fluctuating force

$$m\dot{v} = -\gamma v + F_{\mathrm{r}}(t) \tag{1.1}$$

where γ is the friction coefficient and the random force $F_{\mathrm{r}}(t)$ is mean-free and the second moment is delta-correlated

$$\begin{aligned} \langle F_{\mathrm{r}}(t) \rangle &= 0 \\ \langle F_{\mathrm{r}}(t) F_{\mathrm{r}}(t') \rangle &= 2\gamma k_B T \delta(t - t') \end{aligned} \tag{1.2}$$

where the strength of the random force is subject to Einstein's version of the fluctuation-dissipation theorem. In simple terms the value of the random force must compensate for the fact that in thermal equilibrium with temperature T every degree of freedoms quadratic variation is associated with an amount of energy that is $k_B T/2$ (equipartition theorem). The velocity autocorrelation of the process described by Eq. (1.1) decays with a characteristic time m/γ. When the time between consecutive observations of the process is larger than m/γ, the observed acceleration $\dot{v}$ will mostly be zero because the velocity between collisions is constant. In this case the left hand side of Eq. (1.1) vanishes. The above arguments shall hold also when the particle is subject to an external potential $U(x)$ with the force $-\nabla U$, we can formulate a stochastic equation for the position x in the *overdamped limit* $t \gg m/\gamma$

$$\dot{x} = -\nabla U/\gamma + F_{\mathrm{r}}(t)/\gamma \tag{1.3}$$

At this point we introduce the phenomenological diffusion coefficient $D = k_B T/\gamma$. The convenience of using D is that it can be measured as the ratio of the mean squared traveled distance $\langle x(t)^2 \rangle$ and the traveled time t. As such the diffusion coefficient should be understood as an observable rather than an intensive parameter. We also redefine the random term $r(t) \equiv F_{\mathrm{r}}(t)/\gamma$ which is now a random velocity and rewrite the equation into a form which is associated with the term *Brownian dynamics* or *overdamped Langevin dynamics*

$$\dot{x} = -\beta D \nabla U(x) + r(t) \tag{1.4}$$

and the moments of the random velocity directly follow from Eq. (1.2)

$$\begin{aligned} \langle r(t) \rangle &= 0 \\ \langle r(t) r(t') \rangle &= 2D\delta(t - t') \end{aligned} \tag{1.5}$$

From Eq. (1.4) the very popular Brownian Dynamics simulation technique can be derived through a Euler-Maruyama discretization of time. This is also the driving mechanism of diffusive transport in iPRD simulations as described in Chapter 4. More formally such stochastic processes are described as driven by a Wiener process, which is discussed in various textbooks [Ött96; Van92; Gar+85].

1.1.2 Doi's Reaction Model

Since Smoluchowski's description of colloids in 1916 [Smo16] which coagulate on direct contact with a certain distance R, there were two notable variations of microscopic reaction models for bimolecular

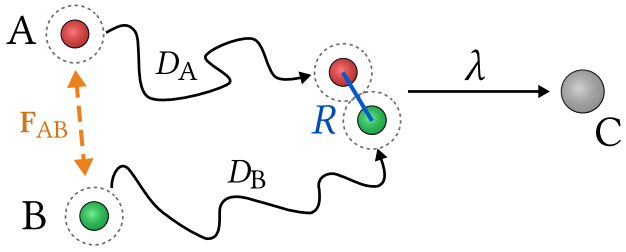

Figure 1.1: Doi reaction model for bimolecular association. Two particles of species A and B approach via diffusion with coefficients D_A and D_B respectively and form an encounter complex when their distance is smaller than the reaction radius R. The encounter complex decays with the frequency λ into the product C.

reactions of the type A + B $\longrightarrow$ C. One is the Collins-Kimball [CK49] model, which modifies the Smoluchowski model by assuming a probability smaller than 1 of two particles reacting upon contact. The second one was mentioned by Teramoto and Shigesada [TS67] and also studied by Doi [Doi75], where a sink function is introduced, which has a constant value (e.g. λ) if the distance of two particles is smaller than R, and vanishes otherwise. This model has been termed Doi model, or $\lambda - \varrho$ model, or volume-reaction model, and is often used in the context of iPRD simulations which integrates time in discrete steps. In such Brownian Dynamics simulations the exact time of the encounter need not be determined. The overall reaction rate is sufficiently approximated when the integration timestep width is smaller than both the intrinsic reaction timescale λ^{-1} and the residence time R^2/D in the reaction volume of radius R due to diffusion with coefficient D. For unimolecular reactions of the type A$\longrightarrow$ B, these reaction events occur spontaneously, i.e. according to a Poisson process. In the case of bimolecular reactions of the type A + B$\longrightarrow$ C there is an additional parameter R. When the distance of A and B is smaller than the *reaction radius*, only then the reaction event can fire. The reaction radius is a predefined constant, which roughly resembles the size of the two reacting molecules. This situation is depicted in Fig. 1.1.

The respective reaction and diffusion time-scales give rise to a scaling parameter, which is useful in distinguishing reaction-limited (well-mixed) and diffusion-limited behavior. The latter is achieved by $\lambda \longrightarrow \infty$, at which point the Doi model becomes equivalent to the Smoluchowski model or the Debye model [Deb42] in the presence of potentials. The scaling parameter is $\sqrt{\lambda/D}R \equiv \kappa R$, and κ^{-1} is a length that can be regarded as the penetration depth of two particles into their respective reaction volume before they decay due to the reaction sink. Alternatively $(\kappa R)^2 = \lambda R^2/D = \lambda\tau_d$ compares the time-scales as described above. Generally $\kappa R \ll 1$ is the reaction-limited case, $\kappa R \gg 1$ is the diffusion-limited case, and $\kappa R \approx 1$ is called diffusion-influenced.

1.1.3 Macromolecular Structure and its Assembly

The dynamic formation of intermolecular bonds can be described in terms of microscopic reaction processes. In combination with force fields for bonded structures this opens up a broad range of application for iPRD simulations. Self-assembly of macromolecules into even larger structures is a driving motif in biology [PB03; Gib+12] and technical applications [Bla04; Rot06]. Structural properties, like the persistence length of polymers [RC03] can then be studied with time-dependent descriptions.

For example cytoskeletons, that provide cells with structural rigidity while being flexible, are self assembling structures. They are formed out of Actin proteins under the consumption of energy. The filaments that Actin forms [PB03] are also slightly helical, a structural property which iPRD simulations achieve via force fields. A more general example of how a helical structure assembles out of substrate

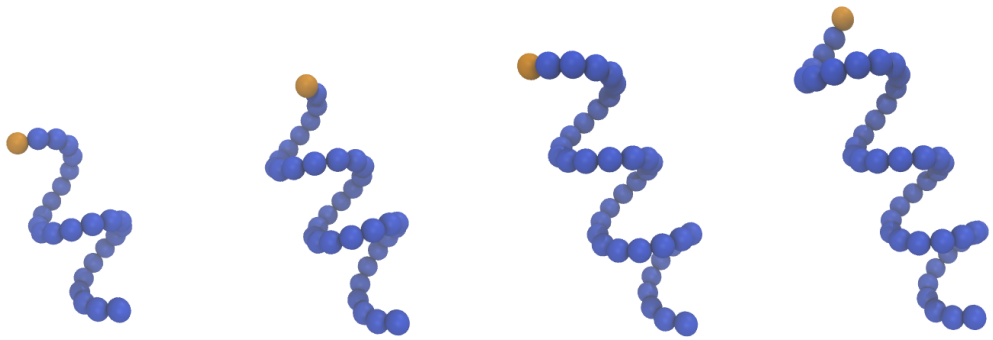

Figure 1.2: Assembly of a linear polymer in helix structure from substrate particles (not shown) in solution. Snapshots from left to right show a time-series over the course of one simulation with ReaDDy 2. The monomers interact via bond-, angle- and dihedral potentials, which defines a helical secondary structure. The head particle (orange) is a reaction site to which freely diffusing substrate particles can bind in a scheme "head + substrate$\longrightarrow$ monomer – head", i.e. the head particle becomes a normal monomer and the substrate particle becomes the head particle.

particles in solution is demonstrated in Fig. 1.2.

The formation of virus capsids is also a self-assembly process [Hag14; Sad16] and is an important step in the maturization process of a virus. Understanding this process in more detail by experiments, theory and simulations would help discover possibilities to influence the process externally. A small example of patchy particles assembling into pentamers, hexamers, and heptamers is shown in Fig. 1.3. In this simple model the patchy particles represent the repetitive protein units of a virus capsid, and the patches fix the location where other patchy particles may be attached. Here we show only a model for early stages of virus capsid assembly, but is easily extended to describe a complete assembly process.

Another possibility to use small repetitive connected units of particles is shown in Fig. 1.4. Here a membrane assembles out of small 3-bead units that represent coarse-grained lipids [CD05]. In this particular example no reactions are needed for the self assembly. The process is purely entropy-driven due to the interaction of the different beads. However reactions can be introduced, e.g. to model local changes to the membrane beads triggered by encounter with other molecules.

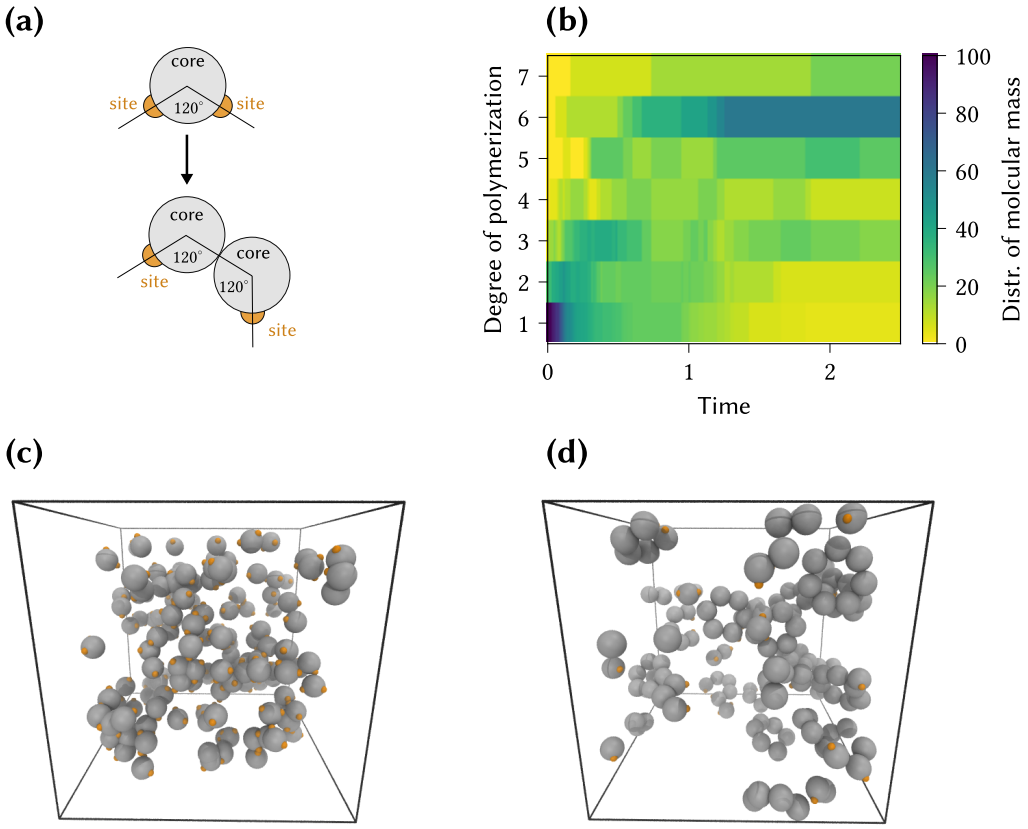

Figure 1.3: Patchy particles assemble into oligomers as an iPRD model for early stages of virus capsid assembly, simulated with ReaDDy 2. **(a)**: One monomer depicts one protein and consists of a "core " and two "site" particles that are in a fixed angle of 120°. The monomer can bind to two other monomers via the reaction sites, two reaction sites perform an association event according to the Doi model. Afterwards the two associated site particles are removed. The formed dimer has two open sites, dihedral potentials assure an alignment of all four particles in a plane. Additional monomers can then bind to the open sites. **(b)**: The distribution (Distr.) of molecular mass (here in numbers of particles) as a function of the degree of polymerization and time for a single simulation run. Initially all particles are present as monomers. In the end there is a mixture of mainly hexamers, as well as some misformed heptamers, pentamers. **(c)**: Snapshot of the simulation box at $t = 0$. **(d)**: Snapshot at $t = 2.5$. Time is in arbitrary units in this example.

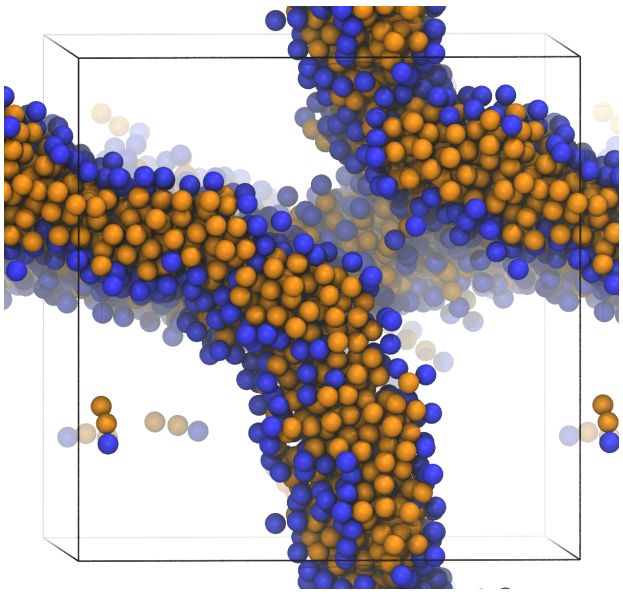

Figure 1.4: Three bead lipid model assembles into a bilayer simulated in ReaDDy 2, inspired by the model of Deserno *et al.* [Rey+07; CD05]. A lipid consists out of a head bead (blue) and two tail beads (orange) which are connected by harmonic springs with length σ and spring constant $60 k_B T/\sigma^2$. Angle potentials between the triplet enforce a 180° configuration with stiffness $10 k_B T$. Non-bonded harmonic repulsion with contact distance 1.05σ and force constant $800 k_B T/\sigma^2$ acts between head and head particles, and between head and tail particles. A non-bonded weak interaction potential built out of harmonic terms that resembles a very soft Lennard-Jones interaction acts between tail and tail particles. This potential has a well with depth of $0.91 k_B T$ and the minimum is at a distance of 1.12σ, any penetration closer is penalized by a harmonic potential with force constant $800 k_B T/\sigma^2$. The range of attraction of tail and tail beads is 2.72σ. Initially 1000 lipids were distributed randomly in a cubic periodic simulation box of fixed edge length 25σ, depicted by black lines. The system was simulated for several days computation time with a time step of $\Delta t = 7.8 \times 10^{-4} \sigma^2/D_0$ where D_0 is the self diffusion coefficient of head and tail beads. The image shows the system at the last frame of the simulation with periodic continuations for visual guidance.

1.2 Equilibrium in Reactive Systems

Reaction-diffusion processes are by definition dynamical processes. Starting in an initial configuration the system's microscopic variables will continue to change over time. However in most applications there exists a stationary state in which the variables do not change anymore. In the case of stochastic dynamics it is the probability distribution that eventually becomes the stationary distribution. Steady states are important because they allow theoretic workers to find simple solutions to an otherwise time-dependent problem. At the same time a steady state can be experimentally probed with much higher resolution than a constantly changing system.

When the system at hand is closed, ergodic, and has microscopic reversibility, then its steady state is also a state of thermodynamic equilibrium. A system being "closed" means that no matter leaves or enters the system from its surrounding. For reactive systems to be closed, all reaction channels must conserve matter in such a sense that no matter enters or leaves the system. Systems in contact with a particle-bath — which ensures that the chemical potential μ is constant — become closed when considering the system of interest and the bath as one large system. If thermodynamic equilibrium is reached the macroscopic behavior can be understood in terms of thermodynamic variables of state.

1.2.1 Thermodynamic Equilibrium for Chemical Reactions

In a system of constant pressure p and constant temperature T the relevant thermodynamic potential is the Gibbs free energy and its change $\mathrm{d}G$ with respect to a change $\mathrm{d}N_i$ in number of particles of species i reads [LJC08; Fer+06]

$$\mathrm{d}G = \sum_i \mu_i \mathrm{d}N_i \tag{1.6}$$

where μ_i is the chemical potential associated with species i. Consider the example $\mathrm{A} + \mathrm{B} \rightleftarrows \mathrm{C}$ and we can reinterpret the $\mathrm{d}N_i$ in terms of the extent of the reaction $\mathrm{d}\xi$ and the stoichiometric coeffcients ν_i, which read 1, 1 and −1 for A, B and C respectively

$$\mathrm{d}N_i = \nu_i \mathrm{d}\xi \quad \longrightarrow \quad \mathrm{d}N_A = \mathrm{d}\xi \quad dN_B = \mathrm{d}\xi \quad dN_C = -\mathrm{d}\xi \tag{1.7}$$

The change in Gibbs free energy then becomes $\mathrm{d}G = (\mu_A + \mu_B - \mu_C)\mathrm{d}\xi$. In equilibrium the Gibbs free energy is minimal so the term

$$\left.\frac{\mathrm{d}G}{\mathrm{d}\xi}\right|_{T,p} = 0 \quad \longrightarrow \quad \mu_A + \mu_B - \mu_C = 0 \tag{1.8}$$

vanishes. The common ansatz for the individual chemical potentials is

$$\mu_i = \mu_i^{\ominus}(p, T) + RT \ln(\{i\}\gamma_i) \tag{1.9}$$

in terms of a standard chemical potential $\mu_i^{\ominus}(p, T)$, the activity $\{i\}$, and the activity coefficient γ_i of the species. Inserting this in Eq. (1.8) yields

$$\Delta G^{\ominus} - RT \ln\left(\frac{\{A\}\{B\}}{\{C\}}\frac{\gamma_A\gamma_B}{\gamma_C}\right) = 0 \tag{1.10}$$

where $\Delta G^{\ominus} = \mu_C^{\ominus} - \mu_A^{\ominus} - \mu_B^{\ominus}$ is the standard change in Gibbs free energy which becomes measurable via the equilibrium constant

$$K = \frac{\{A\}\{B\}}{\{C\}} \tag{1.11}$$

$\Delta G^{\ominus}$ is typically defined under ideal conditions, i.e. $\gamma_i = 1$ by measuring the equilibrium constant.

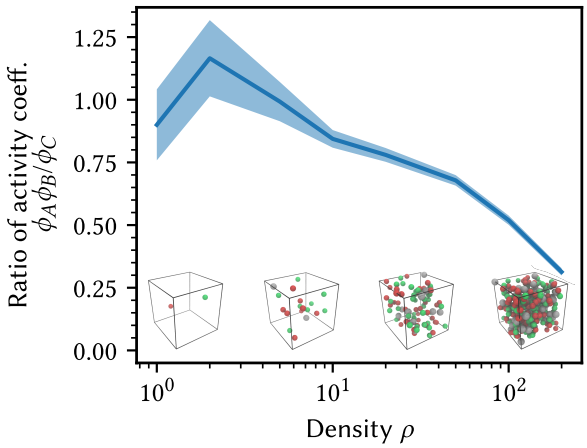

Figure 1.5: Ratio of activity coefficients $\gamma_A\gamma_B/\gamma_C$ as a function of the density $n = (N_A + N_B)/2 + N_C$ of a reactive Lennard-Jones (LJ) suspension simulated in ReaDDy 2. Molecules of species A, B and C interact via a LJ potential with contact distances $\sigma_{ij} = r_i + r_j$ between species i and j where r_i is an assumed radius of the molecule i. Molecules react according to the reversible bimolecular scheme $A + B \rightleftarrows C$ with the reaction radius $R = (26/7)^{1/6}\sigma_{AB}$. The ratio of activity coefficients is calculated as the ratio of activities $\{A\}\{B\}/\{C\}$ of the non-interacting suspension and the LJ suspension.

How do microscopic reaction-diffusion systems in equilibrium relate to these thermodynamic considerations? Let us consider the following experiment: We prepare two reaction-diffusion systems i and ii that perform the reaction $A + B \rightleftarrows C$ in a constant volume V and at constant temperature T. The derivations Eq. (1.6) until Eq. (1.11) can be carried out analogously for the Helmholtz free energy at the expense of making the standard chemical potentials $\mu_i^{\ominus}(V, T)$ dependent on the extensive quantity V. However in this example, we will leave the volume fixed. Both systems i and ii differ in their microscopic behavior, in particular in system ii the molecules interact via a Lennard-Jones interaction, but we prepare them such that under very dilute conditions (i.e. ideal conditions $\gamma_i = 1$) they will tend to the same equilibrium constant K. This equilibrium is simulated with an iPRD model using ReaDDy 2. The result is given in Fig. 1.5. The interaction of the molecules strongly affects the ratio of activities $\gamma_A\gamma_B/\gamma_C$ when the density is increased, in particular the activity coefficients of the dissociated state decreases as the density increases. At low densities the ideal behavior is reproduced. Such data relates the free energies of real solutions in equilibrium and their activities and are at the heart of chemical thermodynamics. Changing activities in dense solutions can e.g. influence the feasibility of metabolic pathways in biological systems, and can lead to a physiological response. To be able to perform thermodynamic studies using reaction-diffusion simulations it is essential that the correct equilibrium state can be sampled.

1.2.2 Equilibrium of a Closed iPRD System

In Section 1.1 we have stated the microscopic rules for a particle-based reaction-diffusion method subject to interaction forces, which is then termed interacting-particle reaction-dynamics (iPRD). When such a system is prepared with an initial state in a constant volume and in contact with a heat bath it will evolve over time. If this system is ergodic and has detailed balance, an ensemble of many copies of this system will reach an equilibrium. Inspired by Masao Doi's work [Doi76] we will formulate this ensemble and its statistical mechanics, which provides a means of calculating observables as ensemble averages.

The *composition* or *state* of the system at a given time is

$$\mathbf{N} = (N_1, \dots N_S) \tag{1.12}$$

which counts how many particles of which species exist and S is the number of species. The most important realization is that the number of compositions that the system can occupy is finite. For this assumption to hold we need the following conditions:

- *Reactions must be reversible.* If there is a reaction that takes the system from composition $\mathbf{N}$ to $\mathbf{M}$, then there must be a reverse reaction which takes the system from $\mathbf{M}$ to $\mathbf{N}$.
- *The system is closed.* Number of particles are only allowed to change in terms of the reversible reactions. This definition of *closed* is less strict than particle conservation. We could alternatively require that the reactions conserve the total mass of the system, and the system is closed with respect to exchange of mass.
- *Reactions into and out of the vacuum state are forbidden.* E.g. the reactions $\emptyset \rightleftarrows$ A are forbidden ($\emptyset$ is the vacuum state, i.e. A particles are spontaneously created and vanish). Note that also the reactions B $\rightleftarrows$ B + A are forbidden because they can be effectively reduced to $\emptyset \rightleftarrows$ A. Because these types of reactions would contradict mass conservation, this rule is just a specification of the "the system is closed" rule.

Since the system is closed, the compositions $\mathbf{N}$ only change due to reactions. If all reactions have a corresponding reversible reaction, the accessible states live on a finite manifold $\mathcal{N}$ which depends on the initial number of particles that live in the system. The number of states in $\mathcal{N}$ is denoted as $|\mathcal{N}| = n$. It generally grows with the number of particles initially in the system, and also with the number of reversible reactions. For example, the system A + B $\rightleftarrows$ C initiated with n A particles and n B particles lives on the manifold of states

$$\mathcal{N} = \{\mathbf{N} = (N_A, N_B, N_C) \,:\, (N_A + N_B)/2 + N_C = n\} \tag{1.13}$$

Each reaction removes one AB pair and adds one C, or adds an AB pair and removes one C.

Each composition $\mathbf{N} \in \mathcal{N}$ has microscopic *configurations* $\mathbf{x}^\mathbf{N}$, where the superscript indicates that the configuration has one component for every particle of every species in the system. Each of these components can be e.g. position and momentum of the particle in $\mathbb{R}^3$, i.e. 6 numbers per particle, then $\mathbf{x}^\mathbf{N} \in \mathbb{R}^{6\|\mathbf{N}\|_1}$ where $\|\mathbf{N}\|_1 = \sum_i^S \mathbf{N}_i$ counts the number of particles in state $\mathbf{N}$. Each state is in thermal equilibrium, i.e. configurations $\mathbf{x}^\mathbf{N}$ are distributed according to

$$p_\mathbf{N}(\mathbf{x}^\mathbf{N}) = \frac{e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})}}{\int e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})} \mathrm{d}\mathbf{x}^\mathbf{N}} \tag{1.14}$$

where $\mathcal{H}_\mathbf{N}$ is the Hamiltonian of system in state $\mathbf{N}$ and $\beta^{-1} = k_B T$.

The global stationary distribution of finding the system in state $\mathbf{N}$ with configuration $\mathbf{x}^\mathbf{N}$ is

$$p(\mathbf{N}, \mathbf{x}^\mathbf{N}) = Z^{-1}(\mathcal{N}, V, T)\, f(\mathbf{N})\, p_\mathbf{N}(\mathbf{x}^\mathbf{N}). \tag{1.15}$$

where f is the stationary number of microstates associated with state $\mathbf{N}$. The number f is related to the *fugacity*, in fact we can interpret it in the form $f(\mathbf{N}) = \exp(\beta k_B T \log f(\mathbf{N}))$, which reveals $-k_B T \log(f(\mathbf{N}))$ to be the free energy of state $\mathbf{N}$ in relation to the other states. $Z(\mathcal{N}, V, T)$ is the partition function, i.e. the total number of microstates in the system of constant $(\mathcal{N}, V, T)$. Normalization of $p(\mathbf{N}, \mathbf{x}^\mathbf{N})$ implies

$$\sum_{\mathbf{N} \in \mathcal{N}} \int p(\mathbf{N}, \mathbf{x}^\mathbf{N}) \mathrm{d}\mathbf{x}^\mathbf{N} = 1. \tag{1.16}$$

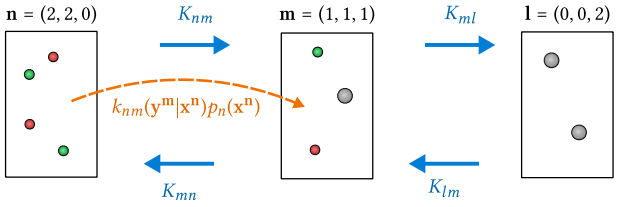

Figure 1.6: Schematic of a closed iPRD phase space which consists of chemical *states* or *compositions* $\mathbf{N}$, $\mathbf{M}$, $\mathbf{L}$ and their corresponding microscopic *configurations* $\mathbf{x}^{\mathbf{N}}$, $\mathbf{y}^{\mathbf{M}}$, $\mathbf{z}^{\mathbf{L}}$ subject to the reaction A + B$\rightleftarrows$C. In this example there are only three reachable states, which are implicitly defined by the number of particles in any of the compositions and the reaction scheme. Configurations $\mathbf{x}^{\mathbf{N}}$ are in local equilibrium with respect to the distribution $p_{\mathbf{N}}(\mathbf{x}^{\mathbf{N}})$. Transitions between states are defined by the microscopic transition rate density $k(\mathbf{y}^{\mathbf{M}}|\mathbf{x}^{\mathbf{N}})$, accompanied by the flux out of state $\mathbf{N}$ into state $\mathbf{M}$ that is $k(\mathbf{y}^{\mathbf{M}}|\mathbf{x}^{\mathbf{N}})p_{\mathbf{N}}(\mathbf{x}^{\mathbf{N}})$. The integral of which over all initial and final configurations yields an element of the transition rate matrix K.

Since each individual state is normalized with respect to configurations, we find the partition function to be the sum of microstates over all compositions

$$Z(\mathcal{N}, V, T) = \sum_{\mathbf{N}\in\mathcal{N}} f(\mathbf{N}) \tag{1.17}$$

Note that f implicitly depends on V, T, the Hamiltonian $\mathcal{H}$, and on the microscopic reactions that allow transitions between compositions.

In essence we have constructed a set of canonical ensembles, which are weighted against each other by f. Since there is one canonical ensemble for each $\mathbf{N}$ in the set $\mathcal{N}$ we may call this an ensemble of constant $(\mathcal{N}, V, T)$ instead of the canonical (N, V, T) ensemble.

Once we have determined the stationary distribution Eq. (1.15) through f and all $p_{\mathbf{N}}$ we can calculate averages of observables. These can either be observables of the state $\mathbf{N}$, the configurations $\mathbf{x}^{\mathbf{N}}$, or both. The latter general case for the observable $g(\mathbf{N}, \mathbf{x}^{\mathbf{N}})$ is described by

$$\langle g \rangle = \sum_{\mathbf{N}\in\mathcal{N}} \int g(\mathbf{N}, \mathbf{x}^{\mathbf{N}}) p(\mathbf{N}, \mathbf{x}^{N}) \mathrm{d}\mathbf{x}^{\mathbf{N}} \tag{1.18}$$

For example in the case A + B$\rightleftarrows$C the equilibrium constant is then measured as $\langle N_A N_B \rangle / \langle N_C \rangle$.

1.2.3 From Microscopic Reactions to Distribution of Compositions

In classical statistical mechanics the probability f to find a substance with a particular composition in the grand canonical ensemble is defined *top-down* upon fixing the chemical potential. Here we will take a *bottom-up* approach, i.e. we formulate microscopic transitions, that will give rise to the stationary distribution $f(\mathbf{N})$ of states.

Therefore we address the microscopic phase space of configurations and equip it with reactions that allow transitions from a composition $\mathbf{N}$ to another composition $\mathbf{M}$. This situation is depicted in Fig. 1.6 for the reaction A + B$\rightleftarrows$C and the number of available compositions $|\mathcal{N}| = n = 3$.

Let $k_r(\mathbf{y}^{\mathbf{M}}|\mathbf{x}^{\mathbf{N}})$ be the microscopic transition rate density to go from configuration $\mathbf{x}^{\mathbf{N}}$ in state $\mathbf{N}$ to configuration $\mathbf{y}^{\mathbf{M}}$ in state $\mathbf{M}$. It may have the following form, which works well with the Doi reaction model (see Section 1.1.2)

$$k_r(\mathbf{y}^{\mathbf{M}}|\mathbf{x}^{\mathbf{N}}) = \lambda_r q(\mathbf{y}^{\mathbf{M}}|\mathbf{x}^{\mathbf{N}}). \tag{1.19}$$

where λ_r is the frequency (also called: probability per unit time, propensity, or microscopic rate constant) of the reaction, and $q(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N})$ is a density that proposes new configurations $\mathbf{y}^\mathbf{M}$ in the new state $\mathbf{M}$ given configuration $\mathbf{x}^\mathbf{N}$ of the old state $\mathbf{N}$. q obeys

$$\int q(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N})\mathrm{d}\mathbf{y}^\mathbf{M} = 1. \tag{1.20}$$

The transition rate density is accompanied by a flux density

$$k_r(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N})p_\mathbf{N}(\mathbf{x}^\mathbf{N}) \tag{1.21}$$

with $p_\mathbf{N}$ from Eq. (1.14).[1] This flux is to be understood as density per $\mathrm{d}\mathbf{x}^\mathbf{N}$ through $p_\mathbf{N}$ and per $\mathrm{d}\mathbf{y}^\mathbf{M}$ through k_r. Integrating this flux over all initial and final states results in the total probability flux from $\mathbf{N}$ to $\mathbf{M}$, which we simply call transition rates

$$K_{\mathbf{NM}} = \sum_{r\in\mathcal{R}(\mathbf{M}|\mathbf{N})} \iint k_r(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N})p_\mathbf{N}(\mathbf{x}^\mathbf{N})\mathrm{d}\mathbf{x}^\mathbf{N}\mathrm{d}\mathbf{y}^\mathbf{M} \tag{1.22}$$

where the summation is executed for all reactions r that lead from composition $\mathbf{N}$ to $\mathbf{M}$ which is indicated by the set $\mathcal{R}(\mathbf{M}|\mathbf{N})$. In most cases this will only have one term because each reaction typically generates a new composition. However, in general there can be multiple reactions with the same products and educts (i.e. they generate the same composition) but with different transition rate densities k_r.

The rate defined in Eq. (1.22) gives rise to a transition rate matrix between the states. Let there be an arbitrary ordering in the set $\mathcal{N}$ with n elements. Then we can enumerate the states $1..n$ and formulate the matrix with elements $K_{\mathbf{NM}} \equiv K_{NM}$

$$K = \begin{pmatrix} K_{11} & K_{12} & \dots & K_{1n} \\ K_{21} & K_{22} & & \vdots \\ \vdots & & \ddots & \vdots \\ K_{n1} & \dots & \dots & K_{nn} \end{pmatrix} \tag{1.23}$$

where we follow the typical convention that

$$K_{ii} = -\sum_{i\neq j} K_{ij} \tag{1.24}$$

Now this matrix certainly defines an ergodic system, because we have constructed the state space to be connected and each reaction is reversible. If the system additionally has detailed balance with respect to the joint global states $(\mathbf{N}, \mathbf{x}^\mathbf{N})$, then there is a stationary vector $\mathbf{f} \in \mathbb{R}_+^n$, which is the vector version of $f(\mathbf{N})$. The stationary vector fulfills the global balance condition

$$\mathbf{f}^\top K = 0 \tag{1.25}$$

which means that in equilibrium all net transition rates into and out of individual states are equal. In other words we are looking for the first eigenvector of K corresponding to eigenvalue 0. From the global balance we can determine $\mathbf{f}$ up to a normalization constant. For a general Markovian process it is customary to remove this ambiguity by assuming that the sum of the $\mathbf{f}_i$ is unity, i.e. $\sum_i \mathbf{f}_i = 1$. However, be reminded that bespoke sum is also the partition function $Z(\mathcal{N}, V, T)$ of the system at hand, see Eq. (1.17).

[1] It is important here to *not* use the global stationary distribution which would yield a probability flux density with respect to the *global* probability, but then microscopic reversibility would imply that these are all equal and the resulting stationary distribution for f would be uniform for all systems studied, which is unreasonable. Here we want to find f under the assumption that a certain amount of probability with respect to the *local* probability flows from one state to another. When f is found, detailed balance with respect to the global probability will assure that each configuration remains in thermal equilibrium.

Fixing it to 1 for every system removes the ability to compare two different ensembles $(\mathcal{N}_1, V_1, T_1)$ and $(\mathcal{N}_2, V_2, T_2)$, since they would have the same free energy $F = -k_B T \log Z$. This means that by defining the microscopic reactions through k_r we cannot determine the overal number of microstates because we are lacking a reference point. Note that the canonical partition function of one composition $\mathbf{N}$ does not represent a meaningful number of microstates anymore in this context because the reactions across the compositions change the "phase space volume" of one microstate. This size is typically given by the Planck constant for classical systems. In our $(\mathcal{N}, V, T)$ ensemble this size differs between compositions, and it is implicitly contained in f, but we cannot estimate it from the transition rates. To establish the missing constraint we would need to perform a thermodynamic integration from an ensemble of known free energy (or number of microstates) into our ensemble at hand, and use that to establish the missing constraint. When measuring observables of one particular system this ambiguity is unproblematic, since Z then only functions as normalization.

In the case of only one pair of reversible reactions which connects the compositions, we can set up the ordering $1..i..n$ such that neighbors i and $i+1$ (likewise i and $i-1$) are connected by the reaction channel. In that case K becomes an unsymmetric tridiagonal matrix

$$K = \begin{pmatrix} K_{11} & K_{12} & & & \\ K_{21} & K_{22} & K_{23} & & \\ & K_{32} & \ddots & \ddots & \\ & & \ddots & \ddots & K_{n-1,n} \\ & & & K_{n,n-1} & K_{nn} \end{pmatrix}.$$

with strictly positive off-diagonals, negative diagonals, and zeros elsewhere. This matrix can be cast into a symmetric tridiagonal matrix by a similarity transformation. A large body of literature is concerned with finding the eigenvectors and -values of such tridiagonal matrices. Furthermore the number of non-zero elements in this matrix scales with n not n^2. In practice, efficient solvers exist for such banded matrices so that we can safely assume a solution for $\mathbf{f}$ can be found within computational boundaries.

1.2.4 Detailed Balance

Detailed balance is also known as microscopic reversibility. For the reactive iPRD system of joint global states $(\mathbf{N}, \mathbf{x}^\mathbf{N})$ it is necessary to keep the configurations $\mathbf{x}^\mathbf{N}$ in each composition $\mathbf{N}$ in local thermal equilibrium with respect to Eq. (1.14). In other words we need to ensure that the net flow of probability between microscopic configurations will leave the global equilibrium Eq. (1.15) unaltered. This form of detailed balance reads

$$p(\mathbf{N}, \mathbf{x}^\mathbf{N}) k_{r+}(\mathbf{y}^\mathbf{M} | \mathbf{x}^\mathbf{N}) = p(\mathbf{M}, \mathbf{y}^\mathbf{M}) k_{r-}(\mathbf{x}^\mathbf{N} | \mathbf{y}^\mathbf{M}) \quad \forall \mathbf{x}^\mathbf{N}, \forall \mathbf{y}^\mathbf{M} \tag{1.26}$$

Note that this requirement is stronger than the global balance from Eq. (1.25). Additionally we require here that each forward reaction $r+$ has exactly one counterpart reaction $r-$, which is also a stronger condition than the plain "every reaction must be reversible", because the detailed balance requires pointwise equality for all $\mathbf{x}^\mathbf{N}$ and for all $\mathbf{y}^\mathbf{M}$. Note also that there can be multiple reaction channels between the two compositions as long as each reaction channel consists of a $(+, -)$ pair for which detailed balance holds.

From Eq. (1.26) one can derive reaction mechanisms that fulfill detailed balance and thus sample the correct thermodynamic equilibrium. In particular one can construct Monte-Carlo schemes for k_r with Metropolis-Hastings acceptance functions, which is done for the special case of the reaction $\mathrm{A} + \mathrm{B} \rightleftarrows \mathrm{C}$ of the isolated pair in Chapter 2, with the simplification that the phase space in the two compositions is the same, i.e. it contains "dummy" or ghost variables for particles that are currently not present in the system. In Section 1.2.5 we demonstrate how the same scheme and global distribution f results from applying the more general theory given above.

1.2.5 Example: Reversible Association of the Isolated Pair

We consider the reaction A + B $\rightleftarrows$ C of the isolated pair, i.e. at all times there is the pair of particles A and B, or one C particle. Also the particles A and B are subject to an interaction potential $U_{AB}(r_{AB})$ which only depends on their distance.

There are only two compositions $\mathbf{N} = (1, 1, 0)$ and $\mathbf{M} = (0, 0, 1)$ with corresponding configurations $\mathbf{x}^\mathbf{N}$ and $\mathbf{y}^\mathbf{M}$. We assume a forward transition density inspired by the Doi reaction model

$$k_+(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N}) = \lambda_{\text{on}} q_+(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N}) = \lambda_{\text{on}} \chi(\mathbf{x}^\mathbf{N}) \delta(\mathbf{x}^\mathbf{N}, \mathbf{y}^\mathbf{M}) \tag{1.27}$$

where λ_{on} is the reaction frequency, χ is a unitless characteristic function that is 1 if A and B are allowed to react and 0 otherwise, and δ is a Dirac delta function which fixes the configuration of the C state given the configuration of the AB state. We want to derive an expression for the backward density with the ansatz

$$k_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) = \lambda_{\text{off}} q_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) = \lambda_{\text{off}} \tilde{q}_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) W_-^{-1}(\mathbf{y}^\mathbf{M}) \tag{1.28}$$

where we need to determine $\tilde{q}_-$ and its corresponding normalization W_-. Note that q_+ is already normalized due to the delta function. For this two state system we know that the stationary distribution f of compositions is proportional to the total transition rate into the composition which we can formulate as the double integral over initial and final configurations from Eq. (1.22)

$$\begin{aligned} f(\mathbf{N}) &\propto \lambda_{\text{on}} \frac{\iint \chi(\mathbf{x}^\mathbf{N}) \delta(\mathbf{x}^\mathbf{N}, \mathbf{y}^\mathbf{M}) e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})} \mathrm{d}\mathbf{x}^\mathbf{N} \mathrm{d}\mathbf{y}^\mathbf{M}}{\int e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})} \mathrm{d}\mathbf{x}^\mathbf{N}} \\ f(\mathbf{M}) &\propto \lambda_{\text{off}} \frac{\iint q_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) e^{-\beta \mathcal{H}_\mathbf{M}(\mathbf{y}^\mathbf{M})} \mathrm{d}\mathbf{y}^\mathbf{M} \mathrm{d}\mathbf{x}^\mathbf{N}}{\int e^{-\beta \mathcal{H}_\mathbf{M}(\mathbf{y}^\mathbf{M})} \mathrm{d}\mathbf{y}^\mathbf{M}}. \end{aligned} \tag{1.29}$$

As discussed in Section 1.2.3 the normalization constant of f is arbitrary when only studying a single system. We plug both k_r and f in the detailed balance equation Eq. (1.26) and solve for q_-

$$\begin{aligned} \tilde{q}_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) W_-^{-1}(\mathbf{y}^\mathbf{M}) = &\frac{e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})}}{e^{-\beta \mathcal{H}_\mathbf{M}(\mathbf{y}^\mathbf{M})}} \chi(\mathbf{x}^\mathbf{N}) \delta(\mathbf{x}^\mathbf{N}, \mathbf{y}^\mathbf{M}) \dots \\ &\dots \times \frac{\iint \tilde{q}_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) W_-^{-1}(\mathbf{y}^\mathbf{M}) e^{-\beta \mathcal{H}_\mathbf{M}(\mathbf{y}^\mathbf{M})} \mathrm{d}\mathbf{y}^\mathbf{M} \mathrm{d}\mathbf{x}^\mathbf{N}}{\iint \chi(\mathbf{x}^\mathbf{N}) \delta(\mathbf{x}^\mathbf{N}, \mathbf{y}^\mathbf{M}) e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})} \mathrm{d}\mathbf{x}^\mathbf{N} \mathrm{d}\mathbf{y}^\mathbf{M}}. \end{aligned} \tag{1.30}$$

Now this expression for q_- depends on its own integral. However it is quite obvious that the following result will fulfill this equation

$$\tilde{q}_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) W_-^{-1}(\mathbf{y}^\mathbf{M}) = \frac{e^{-\beta \mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})}}{e^{-\beta \mathcal{H}_\mathbf{M}(\mathbf{y}^\mathbf{M})}} \chi(\mathbf{x}^\mathbf{N}) \delta(\mathbf{x}^\mathbf{N}, \mathbf{y}^\mathbf{M}). \tag{1.31}$$

This expression lends itself to be cast into a Metropolis-Hastings form, where the factor $\exp\{-\beta[\mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N}) - \mathcal{H}_\mathbf{M}(\mathbf{y}^\mathbf{M})]\}$ is put into an acceptance function and the rest is the proposal density, which is similarly done in Chapter 2. The only thing left to do is find the normalization W_- through $\int q_-(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M}) \mathrm{d}\mathbf{x}^\mathbf{N} = 1$ which yields

$$W_-(\mathbf{y}^\mathbf{M}) = \Gamma_{\text{reac}}^{\text{eff}}, \tag{1.32}$$

which is the effective phase-space reaction volume in the AB state. The specific value of $\Gamma_{\text{reac}}^{\text{eff}}$ depends on the choice of configurations. In Chapter 2 we only consider positional degrees of freedom. Here we shall consider momentum as well

$$\begin{aligned} \mathbf{x}^\mathbf{N} &= (\mathbf{q}_A^\mathbf{N}, \mathbf{q}_B^\mathbf{N}, \mathbf{p}_A^\mathbf{N}, \mathbf{p}_B^\mathbf{N}) \\ \mathbf{y}^\mathbf{M} &= (\mathbf{q}_C^\mathbf{M}, \mathbf{p}_C^\mathbf{M}). \end{aligned} \tag{1.33}$$

For the delta function in q_+ this means that the position of the C particle shall be in the middle between A and B

$$\mathbf{q}_C^{\mathbf{M}} - (\mathbf{q}_A^{\mathbf{N}} + \mathbf{q}_B^{\mathbf{N}})/2 = 0, \tag{1.34}$$

while the momentum of the C particle is distributed among A and B

$$\mathbf{p}_C^{\mathbf{M}} - \mathbf{p}_A^{\mathbf{N}} - \mathbf{p}_B^{\mathbf{N}} = 0.$$

The characteristic function χ which assures that A and B are within reaction radius R then only applies to the positions, momenta are unaffected. Summarizing delta and characteristic function

$$\begin{aligned} \delta(\mathbf{x}^{\mathbf{N}}, \mathbf{y}^{\mathbf{M}}) &= \delta\left(\mathbf{q}_C^{\mathbf{M}} - \frac{\mathbf{q}_A^{\mathbf{N}} + \mathbf{q}_B^{\mathbf{N}}}{2}\right) \delta\left(\mathbf{p}_C^{\mathbf{M}} - \mathbf{p}_A^{\mathbf{N}} + \mathbf{p}_B^{\mathbf{N}}\right) \\ \chi(\mathbf{x}^{\mathbf{N}}) &= \theta(R_{\text{reac}} - \|\mathbf{q}_B^{\mathbf{N}} - \mathbf{q}_A^{\mathbf{N}}\|) \end{aligned} \tag{1.35}$$

where θ is the Heaviside function. We obtain

$$\Gamma_{\text{reac}}^{\text{eff}}(\mathbf{y}^{\mathbf{M}}) = V_{\text{reac}}^{\text{eff}} \sqrt{\frac{8\pi k_B T}{m_A^{-1} + m_B^{-1}}} \exp\left(-\beta \frac{(\mathbf{p}_C^{\mathbf{M}})^2}{2(m_A + m_B)}\right) \tag{1.36}$$

where m_A and m_B are masses of particles A and B respectively. The effective reaction volume $V_{\text{reac}}^{\text{eff}}$ is the same as in Chapter 2

$$V_{\text{reac}}^{\text{eff}} = \int_0^{R_{\text{reac}}} e^{-\beta U_{AB}(r)} 4\pi r^2 \mathrm{d}r \tag{1.37}$$

When only positions are considered as degrees of freedom, one obtains $\Gamma_{\text{reac}}^{\text{eff}}(\mathbf{y}^{\mathbf{M}}) = V_{\text{reac}}^{\text{eff}}$ and for comparison with Chapter 2 we can write down the equilibrium constant as observables [Eq. (1.18)] of the system, which reads

$$K_{\text{eq}} = \frac{\langle N_A N_B \rangle}{\langle N_C \rangle} = \frac{f(\mathbf{N})}{f(\mathbf{M})} = \frac{\lambda_{\text{off}}}{\lambda_{\text{on}}} \frac{V - V_{\text{ex}}}{V_{\text{reac}}^{\text{eff}}} \tag{1.38}$$

which is indeed the equilibrium constant $K_{\text{eq}} = K_d V$ that is prescribed in Chapter 2 via Eq. (2.19).

1.3 Time-Dependent Descriptions of Reaction-Diffusion

In the previous section we have described reactive systems in thermodynamic equilibrium. However many phenomena of living systems e.g. the response of a biological cell to external stimuli, be it physical stress, or lack of nutrition, or the signal for the cell to initiate mitosis, cannot be described only by the statistical mechanics of the system for two reasons:

1. These systems are out of thermal equilibrium because energy and matter enters and leaves these systems at all times. In other words "*In terms of physical chemistry, a closed system has no life.*" [Qia07].

2. Response of a system is by definition a concept which involves time due to causality: we can only observe a response *after* the stimulus has been initiated.

This calls for time-resolved descriptions of the studied systems. Non-equilibrium thermodynamics provides means to study time-dependent processes when the microscopic variables are well described by fields and the assumption of a local equilibrium is justified. Here however we will approach the problem from the point of view of dynamical systems. This allows phenomenological studies of systems even in the absence of equilibrium, microscopic reversibility or ergodicity, and if the systems are open. These descriptions are nevertheless connected to (non-)equilibrium statistical mechanics, e.g. in that the stochastic behavior is often of Langevin type subject to the fluctuation-dissipation theorem. Further the dynamical equations often resemble balance equations of conserved quantities, for example continuity equations for probabilities.

In particular we seek dynamical equations that describe the behavior of reactive systems at different number of particles and spatio-temporal scales. The time-dependent stochastic processes Y_t we consider shall be Markovian.[2] I.e. the processes are memoryless or at least have a memory which is shorter than the smallest timescale that we are interested in. Such dynamics usually give rise to a Master equation which governs the probability distribution $p(y, t)$ of the microscopic variables y as a function of time t

$$\dot{p}(y, t) = Gp(y, t) \tag{1.39}$$

where $\dot{p}$ denotes the time-derivative and G is a *linear* operator with respect to p.

Furthermore we are interested in the effective dynamics when the system size (i.e. all extensive variables) becomes very large. In this limit the fluctuations of the stochastic process become small and we obtain a deterministic *macroscopic equation* [Van92] for the variable y

$$\dot{y}(t) = f(y) \tag{1.40}$$

where y is loosely defined as the mean $y(t) = \langle y \rangle(t) = \int y' p(y', t) \mathrm{d}y'$, $\dot{y}$ is its time derivative, and f determines how the system evolves in time. Note that $f(y)$ is generally non-linear in y.[3]

Ultimately we are interested in reactive systems for which we can identify four scenarios: reaction-diffusion equations (partial differential equations), reaction rate equations (ordinary differential equations), particle-based reaction-diffusion, and chemical master equations. These are arranged in Fig. 1.7 and can be distinguished by two determining factors: the number of particles and the spatio-temporal

[2] Let Nico Van Kampen remind us that "Non-Markov is the rule, Markov is the exception" [Kam98]. I.e. when formulating a Markovian dynamical equation we should have good reason to assume the Markov property. In systems with a strong timescale separation we can be sure that considering only the "slow" coordinates provides a dynamics whose autocorrelation decays exponentially, i.e. there is no memory and the dynamics is Markov.

[3] It should be noted here that the infinitesimal Koopman generator [Klu+19] $\mathcal{K}$ for a system — stochastic or deterministic — governs the time-evolution $\dot{g}(x) = (\mathcal{K}g)(x)$ of *all* functions $g : \mathbb{R}^d \to \mathbb{R}$ defined on the d degrees of freedom. One might consider this the governing theory for all Markovian processes. Also $\mathcal{K}$ is surely linear because it acts on a function-space which makes $\mathcal{K}$ very powerful. One the other hand the operator is infinite dimensional and the problem of solving the dynamical system is shifted to finding an appropriate subspace of functions (i.e. a finite rank approximation) in which the operator can be represented.

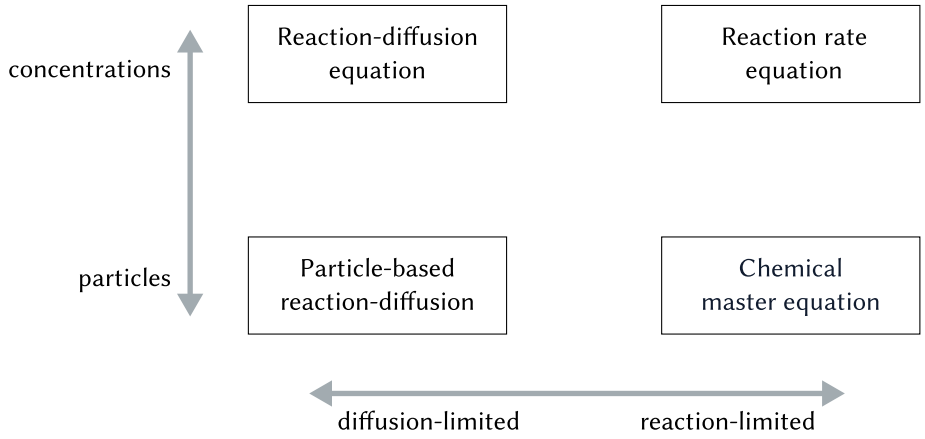

Figure 1.7: Time-dependent descriptions for reactive systems.

scale. The former determines if a problem is well described by an average concentration $c = N/V$ of N molecules in a volume V or requires treatment of individual numbers of molecules. The dynamics of few reacting particles is then a stochastic process and the situation of Eq. (1.39) applies, while concentration-based approaches are deterministic and we obtain a macroscopic equation, see Eq. (1.40). The latter spatio-temporal scale determines if a problem needs spatial resolution which is determined by the relative time-scales of diffusion and reaction. If diffusion is fast and reactions are relatively slow then spatial resolution is not required to formulate a Markovian dynamical equation. When diffusion is slow compared to the reactions the overall kinetics depends on the point and time of encounter. Thus space must be resolved which we can achieve by augmenting the dynamical equation by an additional spatial coordinate x, i.e. we get second stochastic process X_t that shall be Markovian and is coupled to Y_t, $p(y,t)$ becomes $p(y,x,t)$ and $y(t)$ becomes $y(x,t)$. The spatio-temporal scale can also be understood by the amount of non-reactive encounters. For example if two molecules have to encounter 1000 times before associating into a complex then the dynamics is independent of space. If on the other hand they associate at their first encounter the reaction kinetics has a memory of where the molecules started, which we can only get rid of by taking space into the equation.

In the following we will take a route from the reaction rate equations in the upper right of Fig. 1.7 down to the chemical master equation, and then exemplify when spatial resolution is additionally required. In Section 1.1 we have introduced the parameter $\kappa R = \sqrt{\lambda \tau_\mathrm{d}}$ as a direct indicator for the relative diffusion τ_d and reaction λ^{-1} time-scales for the Doi reaction model often used in iPRD simulations and we will demonstrate the number of particles and spatio-temporal scale using iPRD example systems simulated with ReaDDy 2.

Note that here we do not discuss in detail the upper left corner of Fig. 1.7, i.e. space-time partial-differential equations for the concentration of particles. We only note that they are important in biological pattern formation (Turing patterns), or the dynamics of interfaces, e.g. in the Cahn-Hilliard equation. We shall also mention that the overview Fig. 1.7 can be complemented by many descriptions which are partially stochastic or partially space-resolving. [WS17] Also note that in this overview we assume diffusive transport in the absence of velocity fields governed by Navier-Stokes equations.

1.3.1 Systems with Reaction-Limited Behavior

Consider the reaction with index r between species $Z_1, \dots Z_S$, where Z_i is the name of the i-th species of which there are S in total

$$\nu_{r,1} Z_1 + \dots + \nu_{r,s} Z_s \xrightarrow{k_r} \mu_{r,1} Z_1 + \dots + \mu_{r,s} Z_s \tag{1.41}$$

where k_r is the macroscopic rate constant of the r-th reaction, $\nu_{r,\cdot} \in \mathbb{N}$ are the stoichiometric coefficients of the educts of reaction r and the $\mu_{r,\cdot} \in \mathbb{N}$ are corresponding coefficients of the products. When the typical time-scale of diffusion is short compared to the time-scale of reaction the kinetics of the system is *reaction-limited* which is often synonymous with *well-stirred* or *well-mixed*. When the amount of molecules taking part in a reaction-limited process is very high, the dynamics of the system is described in terms of reaction rate equations which are macroscopic equations of the form Eq. (1.40)

$$\dot{\mathbf{y}}(t) = f(\mathbf{y}) = \sum_r k_r \theta_r(\mathbf{y}) \boldsymbol{\eta}_r \tag{1.42}$$

where $\mathbf{y} = ([Z_i], \dots [Z_S]) \in \mathbb{R}_+^S$ is a vector of concentrations with an entry for each molecular species. f specifies the amount of change for each species based on the defined reactions. θ_r is a nonlinear function of the concentrations which counts how much amount of molecular concentration is able to undergo the reaction r. The vector $\boldsymbol{\eta}_r = \boldsymbol{\mu}_r - \boldsymbol{\nu}_r \in \mathbb{Z}^S$ specifies the influence of reaction r on individual species determined by their stoichiometry. For example for bimolecular reactions of type $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C}$, $\theta = y_A y_B$ and $\boldsymbol{\eta} = (-1, -1, +1)$. Note that Eq. (1.42) assumes a linear rate law [ADK18] (there is only one constant for each reaction and it enters the equation linearly) and θ is subject to the law of mass action [GW79] such that the activity of each species is represented by its concentration and stoichiometry

$$\theta_r(\mathbf{y}) = \prod_{i=1}^{S} y_i^{\nu_{r,i}} \tag{1.43}$$

In the general case rate laws can be more complicated.

Systems as described by Eq. (1.42) are fully deterministic but can be oscillatory and even chaotic. Often there is an attractor which is determined by the rate constants. Most notably the concentration is a continuous variable which changes smoothly over time. However when the amount of molecules involved is low then the observed kinetics can vary drastically although we are still in the reaction-limited regime. For example the effect of "transcriptional pulsing" can be observed in the protein-synthesis machinery of procaryotic cells [Chu+06; Elo02]. This effect indicates that stochasticity plays a role in the transcription process of cells because assuming reaction kinetics in the form of Eq. (1.42) will not lead to such behavior.

Let us consider an example. We observe a molecular process of the type $\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{C}$ with association rate constant k_{on} and dissociation rate constant k_{off}, here simulated with an iPRD model in the reaction-limited regime. We set up a fixed number n of AB pairs initially and let the system relax into equilibrium and observe the number of particles as a function of time and the equilibrium constant $K_d = [A][B]/[C]$ in the long-time limit. We perform this experiment multiple times for different number n of initial AB pairs but vary the volume V of the container accordingly such that the concentration n/V is initially always the same. This computer experiment is shown in Fig. 1.8. Part (a) shows the time series of the concentration of A and one observes that the curves roughly follow the same shape, but for small n the process is generally more noisy compared to the limit "RRE" which is given by the reaction rate equations. In particular the case $n = 1$ is essentially a two-state Markov jump process. On the other hand the long time equilibrium constants of all cases are fully described by the dissociation constant $K_d = k_{\mathrm{off}}/k_{\mathrm{on}}$ and the volume V.

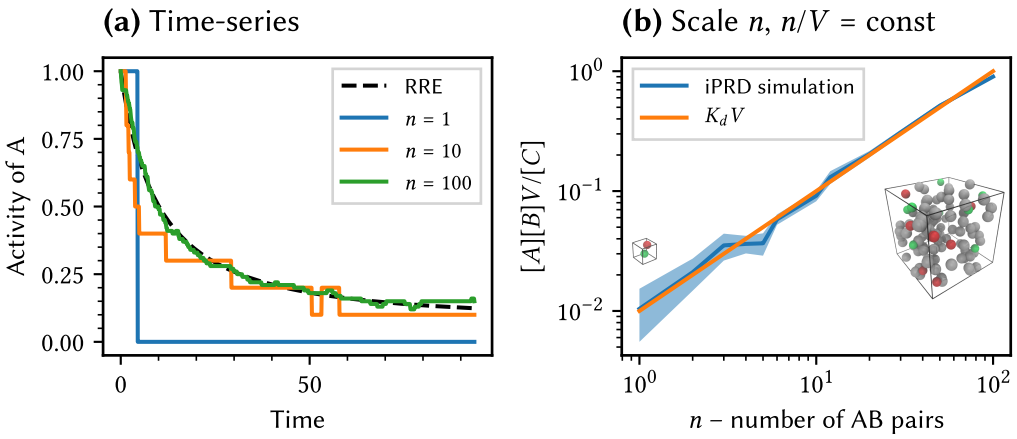

Figure 1.8: Bimolecular reversible reaction A + B$\rightleftarrows$C with fixed association rate constant k_{on} and dissociation constant k_{off} in the reaction-limited regime ($\kappa R = 0.1$) at different number of particles $n = (N_A + N_B)/2 + N_C$ and different volume V such that the density n/V is constant. Dynamics is iPRD simulated with ReaDDy 2. Initially there are n AB pairs placed uniformly in the box, which are then allows to relax while diffusing and reacting. **(a)**: Time-series showing the concentration of A particles. RRE depicts the analytical solution of the reaction rate equations. **(b)**: Equilibrium constant with respect to the volume as a function of the number of initial AB pairs n.

With very few molecules the observed state space becomes discrete, i.e. we count individual molecules with the vector $\mathbf{N} = (N_1, \dots N_S)$ which has an element for every molecular species, the dynamics becomes stochastic. Such a process is called Markov jump process, for reactions it is a one-step process [Van92], i.e. the number of particles only change in predefined quantities at a time, which is defined by the stoichiometry of the considered reactions. The governing equation of the form Eq. (1.39) becomes the chemical master equation [Gar+85; Sun13]

$$\frac{\mathrm{d}p(\mathbf{N}, t)}{\mathrm{d}t} = \sum_r \alpha_r(\mathbf{N} - \boldsymbol{\eta}_r)p(\mathbf{N} - \boldsymbol{\eta}_r, t) - \alpha_r(\mathbf{N})p(\mathbf{N}, t). \tag{1.44}$$

where $\boldsymbol{\eta}_r$ is the change in species number due to reaction r, which is related to the stoichiometry of Eq. (1.41) via $\boldsymbol{\eta}_r = \boldsymbol{\mu}_r - \boldsymbol{\nu}_r$. This equation is a "gain-loss" equation for each composition $\mathbf{N}$. For each reaction r it considers how much probability the state $\mathbf{N}$ gains and how much it loses. These probability fluxes are calculated based on the propensities α, which are frequencies that indicates how likely a reaction is. For a linear rate law, the assumption is that all molecules are in contact and can potentially react. For example the reaction A + B$\longrightarrow$ C then has a propensity $\alpha(\mathbf{N}) = N_A N_B k/V$ where k is the macroscopic rate constant and V the volume of the container, alternatively k/V can be considered a microscopic rate constant. Obeying the law of mass action the propensity of reaction r is then more generally written as [SSG17]

$$\alpha_r(\mathbf{N}) = k_r \prod_{i=1}^{S} \frac{N_i!}{(N_i - \nu_{r,i})! V^{\nu_{r,i}}}. \tag{1.45}$$

This form differs from the simple product in Eq. (1.43) because we have to take into account that we can run out of educt molecules. For example the propensity of A + A$\longrightarrow \cdot$ is proportional to $N_A(N_A - 1)$ which obviously vanishes for $N_A = 1$.

The limit from the chemical master equation Eq. (1.44) to the macroscopic reaction rate equation Eq. (1.42) is the so-called Kurtz limit as it was formalized by Thomas G. Kurtz [Kur78]. This limit essentially assumes $N \longrightarrow \infty$ and $V \longrightarrow \infty$, while N/V remains a constant with vanishing fluctuations.

For practical purposes it shall be noted, that reaction rate equations can usually be integrated numerically as an initial value problem. The chemical master equation however is not easily solved, especially when the state space is unbounded, for example in a birth-death process $\mathrm{A} \rightleftharpoons \emptyset$ the number of A particles is between 0 and ∞. One usually resolves to simulating individual realizations of time-series $\mathbf{N}(t)$ which all start in an initial state $\mathbf{N}(0)$. The time-dependent distribution of many of these trajectories then resembles $p(\mathbf{N}, t)$. The most notable method for simulating these stochastic processes is the Gillespie algorithm [Gil77; ECC09] which is equivalently called Kinetic Monte Carlo (KMC), or Stochastic Simulation Algorithm (SSA). These methods are based on the Markov property such that the exact time of the next reaction event can be drawn from a (self-similar) exponential distribution which only depends on the current state the system is in.

1.3.2 Systems with Spatial Influence

When the reaction-limited assumption breaks down then diffusion time-scale is on a similar order of magnitude or much longer than the reaction time-scale. The number of particles in a container cannot be described by a memoryless process anymore. But making the number of particles dependent on space may resolve the situation and we can formulate another Markovian dynamical equation. If space is divided into many compartments, e.g. separated by membranes, and transport across compartments is slow compared to diffusion within the compartment, then the overall diffusion is well described by a discrete hopping process. In particular such a process is also Poisson, described by a propensity $\alpha(\mathbf{N}_s) = \lambda N_s$ where λ is the frequency of transport out of one compartment into another. This propensity carries the notion of an unimolecular reaction. Formally, reactions and diffusion are then described by a Master equation by broadening the state space to also denote different compartments as different species. The resulting macroscopic equation is thus a reaction rate equation with "reactions" that denote transport from one compartment to another.

If no such compartments and thus no timescale separation exists, the reaction kinetics become time-dependent which leads to non-Markov descriptions and rate laws that deviate from the law of mass action [DK07]. Markovian descriptions then need to describe the movement of individual particles by a diffusion process in continuous space subject to forces. The transport of reactive molecules can have a significant impact on the reaction kinetics, for example due to crowding effects [AT15], or when reactants become sparse such that reactions saturate into the diffusion-limit [RAV14].

Let us consider an example in terms of a simple iPRD model of $\mathrm{A} + \mathrm{A} \longrightarrow \mathrm{B}$ to demonstrate in which way reaction kinetics suffer from hysteresis when the process is *diffusion-influenced*. We set up two systems with the same number of A particles in simulation boxes of the same volume. The two setups differ in the value of $\kappa R = \sqrt{\lambda/D}R$ (0.3 and 30) which relates the microscopic reaction frequency λ and the diffusion time scale R^2/D to explore the spherical reaction volume with radius R and respective diffusion coefficient $D = D_A + D_A$ of A particles. The values of D and λ are then adjusted such that the *mean* association rate constant k is the same for both setups. The procedure to calculate the mean rate for such bimolecular irreversible reactions is described in Chapter 3. Since the mean rate constant is derived under the Markovian assumption by means of first passage times, we expect a deviation in the time-dependent kinetics for the two setups. Initially no particle is in reactive contact with another particle, and we observe the concentration of A particles $a(t)$ as a function of time, which is shown in Fig. 1.9. Not surprisingly we find that the kinetics for the reaction-limited setup $\kappa R = 0.3$ is exactly described by reaction rate equation $\dot{a}(t) = -ka^2$ subject to the law of mass action with the assumed mean rate constant k. Stochastic fluctuations around the macroscopic behavior become apparent when very few particles are left in the box for times $t > 10^2$. The diffusion-influenced kinetics $\kappa R = 30$ deviate from

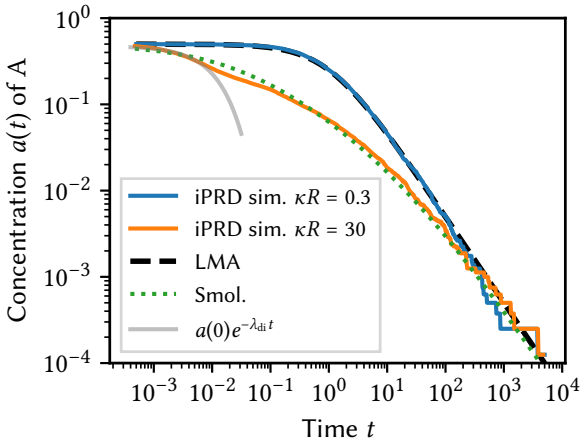

Figure 1.9: Bimolecular irreversible association $A + A \xrightarrow{k} B$ with reaction-limited $\kappa R = 0.3$ and diffusion-influenced behavior $\kappa R = 30$ but with the same *mean* rate constant k. "LMA" is the solution of the reaction rate equations subject to the law of mass action, i.e. $\dot{a} = -ka^2$. "Smol." is the solution of the reaction rate equations with time-dependent rate $k_S(t) = 4\pi DR\left(1 + 4R/\sqrt{\pi Dt}\right)$. Time is given in units of the mean rate and the initial concentration $(ka(0))^{-1}$, which is identical for both processes with $\kappa R = 0.3$ and $\kappa R = 30$. λ_{di} is the microscopic rate constant of the diffusion-influenced process.

the reaction-limited kinetics for the first 6 orders of magnitude in time until it recovers from hysteresis and converges to the case of reaction-rate equations. Note that the case $\kappa R = 30$ is almost "reaction upon contact". This is known as the Smoluchowski limit [Smo16; Smo17] with the time-dependent reaction rate $k_S(t) = 4\pi DR\left(1 + 4R/\sqrt{\pi Dt}\right)$, which is dominated by the encounter rate $4\pi DR$ for large times. The diffusion-influenced case is almost described by this limit, however it displays an initial hysteresis which is well described by the exponential decay $\exp(-\lambda_{\mathrm{di}} t)$ of the first encounter complexes with the intrinsic reaction frequency λ_{di}. It recovers from this hysteresis at about $t \approx 1$, but the reaction proceeds slightly slower than the Smoluchowski limit $\kappa R \longrightarrow \infty$. For long times both scenarios evolve equivalently according to their mean rate, the system has become very dilute such that relative position correlations of particles vanish in between encounters.

Many theories about the non-Markov kinetics of bimolecular reactions have evolved.[LK87b; LK87a; GD96; SL99; GOS01] In these studies the "Brownian Dynamics" simulation approach is often used as a reference, because it directly resolves the space-dependent process of particles diffusing and reacting. In the overview Fig. 1.7 this approach belongs in the lower left corner of particle-based reaction-diffusion methods. In particular iPRD is such a method, and it has been described already in Section 1.1. The governing theory for diffusion-influenced particle-based methods can be constructed in terms of distribution functions [WF73] or quantum field theory [Doi76].

1.3.3 Master equation for a Closed iPRD System

At this point we can formulate a Master equation which governs the time evolution of a closed iPRD system, i.e. particles diffusing and reacting in space. In Section 1.2.2 we have described the stationary distribution $p(\mathbf{N}, \mathbf{x}^{\mathbf{N}})$ of such a system when it is closed and in Section 1.2.3 we have explicitly derived the stationary state from the microscopic reaction transition densities k_r. The latter are already part of the time evolution of the system. The only ingredient we have to add is the drift and diffusion of the

configurations $\mathbf{x}^\mathbf{N}$. Since this is completely orthogonal to the reaction transitions its contribution to the master equation is simply additive. We write $p(\mathbf{N}, \mathbf{x}^\mathbf{N}, t)$ as the time-dependent probability distribution of the system and its time evolution is given by [Doi76]

$$\begin{aligned}\frac{\mathrm{d}p(\mathbf{N}, \mathbf{x}^\mathbf{N}, t)}{\mathrm{d}t} &= \mathcal{L}p(\mathbf{N}, \mathbf{x}^\mathbf{N}, t) \ldots \\ &+ \sum_{\substack{\mathbf{M}\in\mathcal{N} \\ \mathbf{M}\neq\mathbf{N}}} \sum_{r\in\mathcal{R}(\mathbf{M}|\mathbf{N})} \int \left[k_{r,+}(\mathbf{x}^\mathbf{N}|\mathbf{y}^\mathbf{M})p(\mathbf{M}, \mathbf{y}^\mathbf{M}, t) - k_{r,-}(\mathbf{y}^\mathbf{M}|\mathbf{x}^\mathbf{N})p(\mathbf{N}, \mathbf{x}^\mathbf{N}, t)\right] \mathrm{d}\mathbf{y}^\mathbf{M}\end{aligned} \tag{1.46}$$

where $\mathcal{L}$ is the Fokker-Planck or Kolmogorov forward operator [Ris89] which acts only on the configurational part $\mathbf{x}^\mathbf{N}$ of the distribution p. In full it reads

$$\mathcal{L} = -\sum_{i=1}^{\|\mathbf{N}\|_1} \frac{\partial}{\partial x_i} D_i^{(1)}(\mathbf{x}^\mathbf{N}) + \sum_{i=1}^{\|\mathbf{N}\|_1} \sum_{j=1}^{\|\mathbf{N}\|_1} \frac{\partial^2}{\partial x_i \partial x_j} D_{ij}^{(2)}(\mathbf{x}^\mathbf{N}) \tag{1.47}$$

where $D^{(\cdot)}$ are the Kramers-Moyal expansion coefficients, in particular $D^{(1)}$ is the drift vector, which is a force if the particles are subject to a potential and when x are only the positions of particles. $D^{(2)}$ is the diffusion matrix, which simplifies to a constant diagonal matrix if all particles diffuse independently with a certain diffusion coefficient. The second part of Eq. (1.46) is a typical gain-loss term found in any (chemical) one-step Master equation, except that here we have to integrate over the source configurations $\mathbf{y}^\mathbf{M}$ for the "gain" term (target configurations for "loss" term) which belong to another composition $\mathbf{M}$. The summations are performed over all other compositions $\mathbf{M}$ and over all reaction channels r (consisting of the pair of reversible reactions $r, +$ and $r, -$) that connect the compositions $\mathbf{N}$ and $\mathbf{M}$ denoted by the set $\mathcal{R}(\mathbf{M}|\mathbf{N})$. The normalization of p applies for all times analogously to the stationary case

$$\sum_{\mathbf{N}\in\mathcal{N}} \int p(\mathbf{N}, \mathbf{x}^\mathbf{N}, t)\mathrm{d}\mathbf{x}^\mathbf{N} = 1 \quad \forall t \tag{1.48}$$

We should convince ourselves that the stationary solution we have given in Section 1.2.2 is indeed the stationary solution of Eq. (1.46)

$$\frac{\mathrm{d}p(\mathbf{N}, \mathbf{x}^\mathbf{N}, t)}{\mathrm{d}t} = 0 \tag{1.49}$$

Therefore we invoke detailed balance for the reactive transitions [Eq. (1.26)], which immediately cancels the reaction term of Eq. (1.46). It remains to show that

$$\mathcal{L}p(\mathbf{N}, \mathbf{x}^\mathbf{N}, \infty) = \mathcal{L}\left\{Z^{-1}(\mathcal{N}, V, T)\, f(\mathbf{N}) \frac{e^{-\beta\mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})}}{\int e^{-\beta\mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})}\mathrm{d}\mathbf{x}^\mathbf{N}}\right\} \propto \mathcal{L}e^{-\beta\mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})} = 0 \quad \forall\mathbf{N} \tag{1.50}$$

where we have made use of the fact that f, Z, and $\int e^{-\beta\mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})}\mathrm{d}\mathbf{x}^\mathbf{N}$ are mere non-vanishing constants for the operator $\mathcal{L}$. The remaining condition

$$\mathcal{L}e^{-\beta\mathcal{H}_\mathbf{N}(\mathbf{x}^\mathbf{N})} = 0 \quad \forall\mathbf{N} \tag{1.51}$$

implies that $\mathcal{H}$ is the generalized potential for the Fokker-Planck equation. A sufficient condition for stationarity of the Fokker-Planck equation is detailed balance, which has implications for $D^{(1)}$, $D^{(2)}$ and $\mathcal{H}$, see [Ris89] chapter 6.4. These conditions are *for example* fulfilled by

$$D_i^{(1)}(\mathbf{x}^\mathbf{N}) = -\frac{\partial\mathcal{H}_\mathbf{N}}{\partial x_i} \quad \text{and} \quad D_{ij}^{(2)} = \delta_{ij}D \tag{1.52}$$

when all $\mathbf{x}$ are symmetric under time reversal, e.g. positions. We conclude that Eq. (1.15) is indeed the stationary distribution of the proposed Master equation Eq. (1.46) if reactions across- and diffusion within the compositions fulfill detailed balance separately.

1.4 Identification of Reactive Dynamical Systems

The study of reactive systems in an experimental setup poses challenges because the number of molecular species involved in a process might be very large [Con+99; ZL02]. The governing equations provide models for understanding an experimental evidence, but the equations may be solved only for small systems. If the experiment however provides time-resolved measurement data in large quantities we can try to estimate which model best described the observations. The field of system identification can be considered a branch of machine-learning and is concerned with distilling information from data about the underlying system. We may take any equation described in Section 1.3 and interpret it as a model and the rate constants, diffusion coefficients become the parameters w. There are now two questions that one tries to answer:

1. Which model is best suited to describe the given observation data?
2. Given the model and data. What is the optimal set of parameters that best describes the data?

We can understand the second task in terms of Bayesian statistics, in particular we seek to maximize the *posterior* probability $p(w|\mathbf{x})$ with respect to the parameters w given the data $\mathbf{x}$. Bayes theorem states that this probability can be rewritten such that

$$p(w|\mathbf{x}) = p(\mathbf{x}|w)p(w)/p(\mathbf{x}) \propto p(\mathbf{x}|w)p(w) \tag{1.53}$$

which means that the posterior probability $p(w|\mathbf{x})$ we want to maximize is related to the *likelihood* $p(\mathbf{x}|w)$ that given a set of parameters w one obtains the observation $\mathbf{x}$. If we cannot evaluate the likelihood directly we also need a method to generate observations from given parameters. This task may be achieved by a parameterized function f_w. We can then compare the true observations to what f_w predicted. Additionally the *prior* distribution $p(w)$ enters this equation, which encompasses a prior belief about the parameters, e.g. "All parameters in w must be small." which is a distribution that favors small values. The second question above is then answered by maximizing the likelihood subject to the prior.

The parameters w, the choice of likelihood, the prior belief (and optionally the function f_w of generating observations from a set of parameters) can be summarized as the *model*. In short the model is set of assumptions about the problem at hand.

As an example, consider linear regression: We have observed values $\mathbf{x} = (\mathbf{y}, \mathbf{z}) = \{y_i, z_i\}_{i=0}^N$ where y_i are control values set by the experimentalist (e.g. temperature of a gaseous substance) and z_i are the corresponding obtained measurements (e.g. pressure of said substance) of the system. We wish to learn how the system relates the y_i and z_i. We make the assumption that the relation is linear, which gives us a very simple method to generate a prediction $\tilde{z}$ from y via $f_w(y) = wy$. We also make the assumption that the likelihood in this case is maximized by minimizing the residual error between our estimation $\tilde{z} = f_w(y)$ and the true value z. We also have no prior knowledge of the parameter w, so we accept any value that minimizes the error. The optimal value then reads

$$\hat{w} = \operatorname{argmin}_w \|\mathbf{z} - w\mathbf{y}\|_2^2 \tag{1.54}$$

Now solving this for given data is a numerical exercise, but the insight into the system is encapsulated in the assumptions. In the specific example of relating pressure and temperature of gases the achievement was in finding out that they relate linearly which is called Gay-Lussac's law.

This leads to the first and more complicated question: Which model is best suited? In other words: Which are the best assumptions to answer a certain question about the system? The discrepancy between the real system and our assumptions is called the *approximation error*, while Eq. (1.54) only minimizes the *estimation error*.

One possibility to systematically counteract the approximation error is to parameterize the prior belief $p_\alpha(w)$ with a hyperparameter α. Then α is a gradual switch between different models/assumptions.

Eventually this parameter will find its way into the optimization problem when we maximize the likelihood subject to the prior $p_\alpha(w)$. However we cannot optimize for α in the same optimization, because it would not contribute to minimizing the approximation error but only the estimation error. However we can optimize for α by training two models with different α_1 and α_2 on the same data $\mathbf{x}$, which results in two sets of parameters after optimization w_1 and w_2. Now we need to compare the performance of the two models (α_1, w_1) and (α_2, w_2), e.g. by measuring their estimation error with respect to new data $\mathbf{y}$. The new data is measured independent from $\mathbf{x}$, but is measured under identical conditions, i.e. both are samples from the same data distribution $p(\mathbf{x})$. A popular variant of the just described method to quantify the performance of a model is called *cross validation*.

We will proceed to present some methods that aid in identifying reactive systems with assumptions drawn from the descriptions that were presented in Section 1.3.

Sparse identification of nonlinear dynamics (SINDy) [BPK15] is a general framework to infer macroscopic equations from data by an ansatz that is a linear combination of non-linear candidate functions. For reaction rate equations we may interpret Eq. (1.42) as the observable, i.e. the rate of change of the concentrations. SINDy was applied to find the scalar activities that enter each species' temporal change [Man+16]. Alternatively we can formulate a set of vector valued candidate reactions parameterized with rate constants $\mathbf{w}$. Minimizing the estimation error is then achieved by minimizing the difference between the observed change of concentrations and the predicted change of concentrations from our candidate reactions. Minimizing the approximation error is achieved by a prior that favors solutions in which a lot of reaction rates w_i are zero with the goal in mind to find the sparsest set of reactions that led to the observations. This prior enters the likelihood by a 1-norm of the parameters $\|\mathbf{w}\|_1$ with a multiplicative hyperparameter α. Finding α is then achieved by cross validation. This application is called Reactive SINDy and is described in detail in Chapter 5.

When the reaction kinetics is in the limit of few molecules, the observables are samples of a stochastic Poisson process which is governed by the Chemical Master Equation, see Section 1.3.1. With the Gillespie algorithm one can generate samples of the observed process. But since the process is stochastic we cannot use these samples to construct an estimation error. Instead one formulates the likelihood of transitions directly [SSG17]. If all reaction events can be observed this yields a estimate for the best fit of the involved reactions. The state space of compositions $\mathbf{N}$ however has a special structure under certain conditions, which yields a robust estimation [Cha+13] of the rates even when many reaction events happen in between observations.

Inference of processes involving reaction *and* diffusion can be achieved by the same methods as for the chemical master equation, when the diffusion is described by a discrete hopping between well-mixed compartments [Dew+10]. When the observed process is a diffusion process in continuous space driven by a Wiener process one might infer the drift vector and diffusion matrix using the Kramers-Moyal formulae [Ris89]. The method of stochastic SINDy [BNC18] then formulates a linear ansatz of nonlinear candidate functions for drift and diffusion, which is solved by a stepwise sparse regressor to find the simplest solution possible. The full reaction-diffusion dynamics can be understood as several switchable diffusion processes, whose switching behavior is given a Poisson process. Inference of such dynamics is often avoided due to the technicalities of treating the problem in Fock space. Relating the stochastic dynamics to Cox processes however results in an inference method [SGS16].

Another approach for the inference of stochastic dynamics is based on the Perron-Frobenious theorem applied to dynamical operators, that govern the time evolution (e.g. in Eq. (1.39)). The main motif is that the slow and thus important dynamics are determined by the dominant eigenmodes of the dynamical operator. The eigenvalues correspond to timescales and the first eigenvalue is the largest, its eigenvector corresponds to the stationary distribution of the system. All other eigenmodes have smaller timescales. When there is a timescale gap at the n-th eigenvalue, one may describe the possibly very high dimensional system with only the first n slow modes. This opens up a field for efficient dimension reduction [Pér+13; Mar+18; WN17], system identification [BPN14; Klu+19; NR19], and coarse-graining [Wan+19; NBC19].

References

[ADK18] Peter Atkins, Julio De Paula, and James Keeler. *Atkins' physical chemistry*. 8th ed. Oxford: Oxford University Press, 2018. ISBN: 0198700725.

[Alb+08] Bruce Alberts et al. *Molecular biology of the cell*. 5th ed. New York: Garland Science, 2008. ISBN: 9780815344322.

[And17] Steven S. Andrews. "Smoldyn: Particle-based simulation with rule-based modeling, improved molecular interaction and a library interface". In: *Bioinformatics* 33.5 (2017), pp. 710–717. DOI: 10.1093/bioinformatics/btw700.

[AT15] Liliana Angeles-Martinez and Constantinos Theodoropoulos. "The Influence of Crowding Conditions on the Thermodynamic Feasibility of Metabolic Pathways". In: *Biophysical Journal* 109.11 (2015), pp. 2394–2405. DOI: 10.1016/j.bpj.2015.09.030.

[AT17] Satya N. V. Arjunan and Koichi Takahashi. "Multi-algorithm particle simulations with Spatiocyte". In: *Protein Function Prediction*. Springer, 2017, pp. 219–236. DOI: 10.1007/978-1-4939-7015-5_16.

[Bar02] S. Barolo. "Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling". In: *Genes & Development* 16.10 (2002), pp. 1167–1181. DOI: 10.1101/gad.976502.

[Bla04] W. J. Blau. "MATERIALS SCIENCE: Designer Nanotubes by Molecular Self-Assembly". In: *Science* 304.5676 (2004), pp. 1457–1458. DOI: 10.1126/science.1099568.

[BNC18] Lorenzo Boninsegna, Feliks Nüske, and Cecilia Clementi. "Sparse learning of stochastic dynamical equations". In: *The Journal of Chemical Physics* 148.24 (2018), p. 241723. DOI: 10.1063/1.5018409.

[BPK15] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 113.15 (2015), pp. 3932–3937. DOI: 10.1073/pnas.1517384113.

[BPN14] Gregory R. Bowman, Vijay S. Pande, and Frank Noé, eds. *An Introduction to Markov State Models and Their Application to Long Timescale Molecular Simulation*. Springer Netherlands, 2014. DOI: 10.1007/978-94-007-7606-7.

[BS08] D. J. Bicout and Attila Szabo. "Entropic barriers, transition states, funnels, and exponential protein folding kinetics: A simple model". In: *Protein Science* 9.3 (2008), pp. 452–465. DOI: 10.1110/ps.9.3.452.

[CD05] Ira R. Cooke and Markus Deserno. "Solvent-free model for self-assembling fluid bilayer membranes: Stabilization of the fluid phase based on broad attractive tail potentials". In: *Journal of Chemical Physics* 123.22 (2005). DOI: 10.1063/1.2135785.

[Cha+13] I. Chattopadhyay et al. "Inverse Gillespie for inferring stochastic reaction mechanisms from intermittent samples". In: *Proceedings of the National Academy of Sciences* 110.32 (2013), pp. 12990–12995. DOI: 10.1073/pnas.1214559110.

[Chu+06] Jonathan R Chubb et al. "Transcriptional Pulsing of a Developmental Gene". In: *Current Biology* 16.10 (2006), pp. 1018–1025. DOI: 10.1016/j.cub.2006.03.092.

[CK49] Frank C. Collins and George E. Kimball. "Diffusion-Controlled Reactions in Liquid Solutions". In: *Industrial and Engineering Chemistry Research* 41.11 (1949), pp. 2551–2553. DOI: 10.1021/ie50479a040.

[Con+99] Peijun Cong et al. “High-Throughput Synthesis and Screening of Combinatorial Heterogeneous Catalyst Libraries”. In: *Angewandte Chemie International Edition* 38.4 (1999), pp. 483–488. DOI: 10.1002/(SICI)1521-3773(19990215)38:4<483::AID-ANIE483>3.0.CO;2-\#.

[Deb42] P. Debye. “Reaction Rates in Ionic Solutions”. In: *Transactions of The Electrochemical Society* 82.1 (1942), p. 265. DOI: 10.1149/1.3071413.

[Dew+10] Michael A Dewar et al. “Parameter estimation and inference for stochastic reaction-diffusion systems: application to morphogenesis in D. melanogaster”. In: *BMC Systems Biology* 4.1 (2010), p. 21. DOI: 10.1186/1752-0509-4-21.

[DK07] Alexander B. Doktorov and Alexey A. Kipriyanov. “Deviation from the kinetic law of mass action for reactions induced by binary encounters in liquid solutions”. In: *Journal of Physics Condensed Matter* 19.6 (2007). DOI: 10.1088/0953-8984/19/6/065136.

[Doi75] Masao Doi. “Theory of diffusion-controlled reactions between non-simple molecules. I”. In: *Chemical Physics* ll (1975), pp. 107–113. DOI: 10.1016/0301-0104(75)80043-7.

[Doi76] Masao Doi. “Second quantization representation for classical many-particle system”. In: *Journal of Physics A: Mathematical and General* 9.9 (1976), pp. 1465–1477. DOI: 10.1088/0305-4470/9/9/008.

[DYK18] Aleksandar Donev, Chiao-yu Yang, and Changho Kim. “Efficient reactive Brownian dynamics”. In: *The Journal of Chemical Physics* 148.3 (2018), p. 034103. DOI: 10.1063/1.5009464.

[ECC09] Radek Erban, Jonathan Chapman, and S Jonathan Chapman. “Stochastic modelling of reaction-diffusion processes: algorithms for bimolecular reactions.” In: *Physical biology* 6.4 (2009), p. 046001. DOI: 10.1088/1478-3975/6/4/046001.

[Ein05] Albert Einstein. “Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen”. In: *Annalen der Physik* 322.8 (1905), pp. 549–560. DOI: 10.1002/andp.19053220806.

[Elo02] M. B. Elowitz. “Stochastic Gene Expression in a Single Cell”. In: *Science* 297.5584 (2002), pp. 1183–1186. DOI: 10.1126/science.1070919.

[Fer+06] Antonio Fernández-Ramos et al. “Modeling the Kinetics of Bimolecular Reactions”. In: *Chemical Reviews* 106.11 (2006), pp. 4518–4584. DOI: 10.1021/cr050205w.

[Gal+16] Marta Galanti et al. “Reaction rate of a composite core–shell nanoreactor with multiple nanocatalysts”. In: *Physical Chemistry Chemical Physics* 18.30 (2016), pp. 20758–20767. DOI: 10.1039/C6CP01179A.

[Gar+85] Crispin W. Gardiner et al. *Handbook of stochastic methods*. Vol. 3. Springer Berlin, 1985. DOI: 10.1007/978-3-540-70713-4.

[GD96] Irina V. Gopich and Alexander B. Doktorov. “Kinetics of diffusion-influenced reversible reaction A + B $\rightleftharpoons$ C in solutions”. In: *The Journal of Chemical Physics* 105.6 (1996), pp. 2320–2332. DOI: 10.1063/1.472189.

[Gib+12] Thomas Gibaud et al. “Reconfigurable self-assembly through chiral control of interfacial tension”. In: *Nature* 481.7381 (2012), pp. 348–351. DOI: 10.1038/nature10769.

[Gil77] Daniel T. Gillespie. “Exact stochastic simulation of coupled chemical reactions”. In: *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361. DOI: 10.1021/j100540a008.

[GOS01] Irina V. Gopich, Alexander A. Ovchinnikov, and Attila Szabo. “Long-time tails in the kinetics of reversible bimolecular reactions”. In: *Physical Review Letters* 86.5 (2001), pp. 922–925. DOI: 10.1103/PhysRevLett.86.922.

[GW79] Cato M Guldberg and P Waage. "Über die chemische Affinität". In: *Journal für praktische chemie* 19.1 (1879), pp. 69–114.

[Hag14] Michael F. Hagan. "Modeling Viral Capsid Assembly". In: (2014), pp. 1–68. DOI: 10.1002/9781118755815.ch01.

[Har+18] Sven Hartmann et al. "O2-Tolerant H2 Activation by an Isolated Large Subunit of a [NiFe] Hydrogenase". In: *Biochemistry* 57.36 (2018), pp. 5339–5349. DOI: 10.1021/acs.biochem.8b00760.

[Hat+10] Georgia Hatzivassiliou et al. "RAF inhibitors prime wild-type RAF to activate the MAPK pathway and enhance growth". In: *Nature* 464.7287 (2010), pp. 431–435. DOI: 10.1038/nature08833.

[HFE05] Johan Hattne, David Fange, and Johan Elf. "Stochastic reaction-diffusion simulation with MesoRD". In: *Bioinformatics* 21.12 (2005), pp. 2923–2924. DOI: 10.1093/bioinformatics/bti431.

[HFN19] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics". In: *PLoS Computational Biology* 15.2 (2019), e1006830. DOI: 10.1371/journal.pcbi.1006830.

[Kam98] Nicolaas G. van Kampen. "Remarks on non-Markov processes". In: *Brazilian Journal of Physics* 28.2 (1998), pp. 90–96.

[KHU19] Till Köster, Philipp Henning, and Adelinde M. Uhrmacher. "Potential based, spatial simulation of dynamically nested particles". In: *BMC Bioinformatics* 20.1 (2019). DOI: 10.1186/s12859-019-3092-y.

[Klu+19] Stefan Klus et al. "Data-driven approximation of the Koopman generator: Model reduction, system identification, and control". In: *arXiv preprint arXiv:1909.10638* (2019).

[KM83] M. Karplus and J. A. McCammon. "Dynamics of Proteins: Elements and Function". In: *Annual Review of Biochemistry* 52.1 (1983), pp. 263–300. DOI: 10.1146/annurev.bi.52.070183.001403.

[Kur78] Thomas G. Kurtz. "Strong approximation theorems for density dependent Markov chains". In: *Stochastic Processes and their Applications* 6.3 (1978), pp. 223–240. DOI: 10.1016/0304-4149(78)90020-0.

[LA17] Jan Löwe and Linda A. Amos, eds. *Prokaryotic Cytoskeletons*. Springer International Publishing, 2017. DOI: 10.1007/978-3-319-53047-5.

[Lip95] R Lipowsky. *Structure and dynamics of membranes*. Amsterdam New York: Elsevier Science, 1995. ISBN: 0444819754.

[LJC08] G. Lebon, D. Jou, and J. Casas-Vázquez. *Understanding Non-equilibrium Thermodynamics*. Springer Berlin Heidelberg, 2008. DOI: 10.1007/978-3-540-74252-4.

[LK87a] Sangyoub Lee and Martin Karplus. "Kinetics of diffusion-influenced bimolecular reactions in solution. II. Effects of the gating mode and orientation-dependent reactivity". In: *The Journal of Chemical Physics* 86.4 (1987), pp. 1904–1921. DOI: 10.1063/1.452757.

[LK87b] Sangyoub Lee and Martin Karplus. "Kinetics of diffusion-influenced bimolecular reactions in solution. I. General formalism and relaxation kinetics of fast reversible reactions". In: *The Journal of Chemical Physics* 86.4 (1987), pp. 1883–1903. DOI: 10.1063/1.452140.

[Man+16] Niall M. Mangan et al. "Inferring biological networks by sparse identification of nonlinear dynamics". In: *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 2.1 (2016), pp. 52–63. DOI: 10.1109/TMBMC.2016.2633265.

[Mar+18] Andreas Mardt et al. “VAMPnets for deep learning of molecular kinetics”. In: *Nature Communications* 9.1 (2018), pp. 1–11. DOI: 10.1038/s41467-017-02388-1.

[ML16] Paul J. Michalski and Leslie M. Loew. “SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume”. In: *Biophysical journal* 110.3 (2016), pp. 523–529. DOI: 10.1016/j.bpj.2015.12.026.

[NBC19] Feliks Nüske, Lorenzo Boninsegna, and Cecilia Clementi. “Coarse-graining Molecular Systems by Spectral Matching”. In: *arXiv preprint arXiv:1904.07177* (2019).

[NR19] Frank Noé and Edina Rosta. “Special Topic: Markov Models of Molecular Kinetics”. In: *arXiv preprint arXiv:1911.00774* (2019).

[Ött96] Hans Christian Öttinger. “Stochastic Processes, Polymer Dynamics, and Fluid Mechanics”. In: *Stochastic Processes in Polymeric Fluids*. Springer, 1996, pp. 1–15.

[PB03] Thomas D. Pollard and Gary G. Borisy. “Cellular motility driven by assembly and disassembly of actin filaments”. In: *Cell* 112.4 (2003), pp. 453–465. DOI: 10.1016/S0092-8674(03)00120-X.

[PBV09] Vivek Polshettiwar, Babita Baruwati, and Rajender S. Varma. “Self-Assembly of Metal Oxides into Three-Dimensional Nanostructures: Synthesis and Application in Catalysis”. In: *ACS Nano* 3.3 (2009), pp. 728–736. DOI: 10.1021/nn800903p.

[Pér+13] Guillermo Pérez-Hernández et al. “Identification of slow molecular order parameters for Markov model construction”. In: *The Journal of Chemical Physics* 139.1 (2013), p. 015102. DOI: 10.1063/1.4811489.

[Per10] Jean Perrin. *Brownian Movement and Molecular Reality*. London: Taylor and Francis, 1910.

[Pla+17] Nuria Plattner et al. “Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling”. In: *Nature Chemistry* 9.10 (2017), pp. 1005–1011. DOI: 10.1038/nchem.2785.

[Qia07] Hong Qian. “Phosphorylation Energy Hypothesis: Open Chemical Systems and Their Biological Functions”. In: *Annual Review of Physical Chemistry* 58.1 (2007), pp. 113–142. DOI: 10.1146/annurev.physchem.58.032806.104550.

[RAV14] Arnulf Rosspeintner, Gonzalo Angulo, and Eric Vauthey. “Bimolecular Photoinduced Electron Transfer Beyond the Diffusion Limit: The Rehm–Weller Experiment Revisited with Femtosecond Time Resolution”. In: *Journal of the American Chemical Society* 136.5 (2014), pp. 2026–2032. DOI: 10.1021/ja4118279.

[RC03] Michael Rubinstein and Ralph H. Colby. *Polymer physics*. Vol. 23. Oxford University Press New York, 2003. ISBN: 9780198520597.

[Rey+07] Benedict J. Reynwar et al. “Aggregation and vesiculation of membrane proteins by curvature-mediated interactions”. In: *Nature* 447.7143 (2007), pp. 461–464. DOI: 10.1038/nature05840.

[Ris89] Hannes Risken. *The Fokker-Planck Equation*. Vol. 18. Springer Series in Synergetics. Berlin, Heidelberg: Springer, 1989. DOI: 10.1007/978-3-642-61544-3.

[Roa+17] Rafael Roa et al. “Catalyzed Bimolecular Reactions in Responsive Nanoreactors”. In: *ACS Catalysis* 7.9 (2017), pp. 5604–5611. DOI: 10.1021/acscatal.7b01701.

[Rot06] Paul W. K. Rothemund. “Folding DNA to create nanoscale shapes and patterns”. In: *Nature* 440.7082 (2006), p. 297. DOI: 10.1038/nature04586.

[Sad16] S. Kashif Sadiq. “Reaction–diffusion basis of retroviral infectivity”. In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2080 (2016), p. 20160148. DOI: 10.1098/rsta.2016.0148.

[SF93] Gideon Schreiber and Alan R. Fersht. "Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering". In: *Biochemistry* 32.19 (1993), pp. 5145–5150. DOI: 10.1021/bi00070a025.

[SGS16] David Schnoerr, Ramon Grima, and Guido Sanguinetti. "Cox process representation and inference for stochastic reaction–diffusion processes". In: *Nature Communications* 7.1 (2016). DOI: 10.1038/ncomms11729.

[SL99] Jaeyoung Sung and Sangyoub Lee. "Nonequilibrium distribution function formalism for diffusion-influenced bimolecular reactions: Beyond the superposition approximation". In: *The Journal of Chemical Physics* 111.3 (1999), pp. 796–803. DOI: 10.1063/1.479367.

[Smo16] Marian von Smoluchowski. "Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen". In: *Zeitschrift fur Physikalische Chemie* 92U.1 (1916), p. 129. DOI: 10.1515/zpch-1918-9209.

[Smo17] M. v. Smoluchowski. "Grundriß der Koagulationskinetik kolloider Lösungen". In: *Kolloid-Zeitschrift* 21.3 (1917), pp. 98–104. DOI: 10.1007/BF01427232.

[SSG17] David Schnoerr, Guido Sanguinetti, and Ramon Grima. "Approximation and inference methods for stochastic biochemical kinetics—a tutorial review". In: *Journal of Physics A: Mathematical and Theoretical* 50.9 (2017), p. 093001. DOI: 10.1088/1751-8121/aa54d9.

[Sun13] Vikram Sunkara. "Analysis and Numerics of the Chemical Master Equation". In: (2013).

[TS67] Ei Teramoto and Nanako Shigesada. "Theory of bimolecular reaction processes in liquids". In: *Progress of Theoretical Physics* 37.1 (1967), pp. 29–51. DOI: 10.1143/PTP.37.29.

[Van92] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*. 1st ed. Elsevier, 1992. ISBN: 9780444529657.

[Vij+98] Muthusamy Vijayakumar et al. "Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar1". In: *Journal of molecular biology* 278.5 (1998), pp. 1015–1024. DOI: 10.1006/jmbi.1998.1747.

[Wan+19] Jiang Wang et al. "Machine learning of coarse-grained molecular dynamics force fields". In: *ACS central science* (2019). DOI: 10.1021/acscentsci.8b00913.

[WF73] Gerald Wilemski and Marshall Fixman. "General theory of diffusion-controlled reactions". In: *The Journal of Chemical Physics* 58.9 (1973), pp. 4009–4019. DOI: 10.1063/1.1679757.

[WN17] Hao Wu and Frank Noé. "Variational approach for learning Markov processes from time series data". In: *arXiv preprint arXiv:1707.04659* (2017).

[WS17] Stefanie Winkelmann and Christof Schütte. "Hybrid models for chemical reaction networks: Multiscale theory and application to gene regulatory systems". In: *The Journal of Chemical Physics* 147.11 (2017), p. 114115. DOI: 10.1063/1.4986560.

[ZL02] Wei Zhang and Hui Tu Liu. "MAPK signal pathways in the regulation of cell proliferation in mammalian cells". In: *Cell research* 12.1 (2002), p. 9. DOI: 10.1038/sj.cr.7290105.

[ZW05] Jeroen S. van Zon and Pieter Rein ten Wolde. "Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics". In: *Physical Review Letters* 94.12 (2005), p. 128103. DOI: 10.1103/PhysRevLett.94.128103.

Chapter 2

Detailed balance in particle based reactions

The results of this chapter have been published in the following paper:

> Christoph Fröhner (CF) and Frank Noé. "Reversible Interacting-Particle Reaction Dynamics".
> In: *The Journal of Physical Chemistry B* 122.49 (2018), pp. 11240–11250.
> DOI: 10.1021/acs.jpcb.8b06981,
>
> to be obtained via
> http://pubs.acs.org/articlesonrequest/AOR-6EEMXSMurU29aSc4Zg6v

The contributions of the authors were as follows: Both Frank Noé and CF conceived the project, laid out the theory and wrote the paper. CF implemented the algorithms, performed the simulations, analyzed and visualized the data. Frank Noé supervised the work.

Summary

Interacting-Particle Reaction Dynamics (iPRD) simulates the spatiotemporal evolution of particles that experience interaction forces and can react with one another. The combination of interaction forces and reactions enable a wide range of complex reactive systems in biology and chemistry, but give rise to new questions such as how to evolve the dynamical equations in a computationally efficient and statistically correct manner. Here we consider reversible reactions such as A + B $\rightleftarrows$ C with interacting particles and derive expressions for the microscopic iPRD simulation parameters such that desired values for the equilibrium constant and the dissociation rate are obtained in the dilute limit. We then introduce a Monte-Carlo algorithm that ensures detailed balance in the iPRD time-evolution (iPRD-DB). iPRD-DB guarantees the correct thermodynamics at all concentrations and maintains the desired kinetics in the dilute limit, where chemical rates are well-defined and kinetic measurement experiments usually operate. We show that in dense particle systems, the incorporation of detailed balance is essential to obtain physically realistic solutions. iPRD-DB is implemented in ReaDDy 2 (https://readdy.github.io).

2.1 Introduction

Particle based reaction diffusion (PBRD) dynamics is a detailed model for simulating the spatiotemporal evolution of reactive particles [EC09; ZW05a; HS14; And17]. Resolving the trajectories of every reactive particle is important in applications where the reactants cannot be assumed to be spatially well-mixed [ACH16; Alb+16] or always sufficiently abundant to be described by a continuous concentration [Elo02; Bha04] – e.g., consider many cases of cellular signalling and reactions in nontrivial architectures [DM09; PB03; Sad16]. A common implementation of PBRD is to propagate particle positions with overdamped Langevin dynamics (Brownian motion) in discrete time steps, and execute discrete reaction events such as $A + B \longrightarrow C$ with a certain probability when two reactive particles A and B are close in space. When the system is sufficiently dilute, such simulations can be sped up by exploiting solutions of the one- or two-particle diffusion equation [ZW05a; ZW05b; TTW10; Opp+06; Don+10].

A recent extension of PBRD is the interacting-Particle Reaction Dynamics (iPRD) method [SN13; Sch+14; Bie+15], in which particles are additionally subject to interaction forces. Alternatively, iPRD could be characterized as a form of coarse-grained Molecular Dynamics (MD) simulation with reactions between particles. Particle interaction forces are useful to model order and structure on mesoscopic lengthscales, such as the space-exclusion in dense particle systems [SN13; HF13], the restriction of diffusing particles to arbitrarily-shaped membranes [SN13; Gun+15; Sch+14], the large-scale structure of polymers [HFN19] and membranes [SWN18], and the clustering of attractive proteins [Ull+15]. The combination of interaction forces and reactions allow an even wider range of complex reactive systems in biology and chemistry to be modeled, such as the dynamics of phototransduction that involve protein diffusion in particle-dense photoreceptor membranes [Sch+14], the effect of transmembrane protein oligomers on these dynamics [Gun+15], the recruitment of proteins to endosomes [Pos+13; Sch+17], and the assembly, diffusion, and dissociation of polymers [HFN19]. The idea of combining PBRD with particle interaction forces is also found in MD-GFRD [VBW15; Vij+17], where the close particle interactions are simulated by MD and the reaction-diffusion model is used to derive an efficient way to propagate particles while they are not-interacting. In contrast, in iPRD particle interactions and reactions occur simultaneously, with the idea that reaction events are a suitable way to coarse-grain complicated events such as protein-protein binding, whose kinetics might be obtained from Markov State Models of all-atom MD simulations [Dib+18]. MD-GFRD simulations can be used to speed up iPRD simulations when the system is sufficiently dilute [VBW15; Vij+17; SN17], and with free-propagator reweighting, this speedup can also be obtained in the regime where particles are interacting [JH14].

An open question is: What is the statistically correct way to model the dynamical evolution of simultaneously interacting and reacting particles? Specifically, we consider reversible reactions, such as $A + B \rightleftarrows C$, as they are found in nature, but also in technological applications. Examples include reversible protein-drug binding [Sco+16; Pau+17], reversible protein-protein association that can now be simulated at atomistic detail [Pla+17], and metal ion deposition to / removal from electrodes in batteries that are driving charging and discharging [AT08; Boy+06]. To derive a statistically correct simulation scheme of $A + B \rightleftarrows C$ via iPRD, we need to answer the following questions:

1. Which bimolecular reaction scheme should be used, i.e. under which conditions will two particles A and B fuse into a C particle?

2. How do we choose the microscopic parameters of this reaction scheme such that the iPRD simulation samples the macroscopic kinetic quantities that have been obtained from experiments or more detailed MD simulations?

3. When executing $A + B \longrightarrow C$ or $C \longrightarrow A + B$, where should the product particles be placed, such that the simulation obeys detailed balance?

The answers to these three questions are coupled.

Question 1: For the sake of analytical computations, the best-studied reaction scheme is the Smoluchowski model where diffusing particles react instantly when they establish contact, defined by a reaction distance R [Smo16]. The Collins-Kimball model [CK49] reduces the probability of reacting upon contact to a finite value ≤ 1. Reversible reactions in the Collins-Kimball model are discussed in [RQ16], for interacting particles of isolated pairs an analytical description is found in [AS90]. In iPRD simulations we instead use the Doi model [TS67; Doi76]:

$$\mathrm{A} + \mathrm{B} \rightleftarrows \mathrm{AB} \rightleftarrows \mathrm{C}$$

Here two particles A and B form a reactive complex AB when their distance is less or equal to R. This process is simulated by the dynamical model that propagates particles (e.g. overdamped Langevin equation). Whenever a reactive complex AB exists, it can decay to a C particle with a microscopic rate constant λ. The reverse process happens with a microscopic rate constant k_{off}. The Doi model is well compatible with a finite-time-stepping simulation scheme, where the formation of AB can be easily checked in every time-step as part of the particle neighbor list update.

Question 2: When using the Doi model, how should the parameters in this model be chosen? The dissociation rate constant k_{off} can be directly obtained from kinetic experiments or all-atom MD simulations with accelerated sampling methods [Pla+17; Pau+17; DB14]. For the Doi model where A and B encounter from a long distance via normal diffusion without interaction forces, the association parameters R and λ can be computed from an equation derived in [Doi75; EC09]. When A and B interact, such a result can still be obtained numerically [Dib+19]. In Sec. 2.2 we develop a theory for the $\mathrm{A}+\mathrm{B} \rightleftarrows \mathrm{C}$ reaction of an isolated pair, that is independent of the diffusion coefficient D. This enables to choose λ for given dissociation rate constant k_{off}, interaction radius R, and A – B interaction potential such that the iPRD simulation will produce a desired equilibrium constant and association rate constant at low particle concentrations, as they are typically found in experiments measuring these constants.

Question 3: Time-reversible processes evolving in thermodynamic equilibrium obey detailed balance [Van92]. For example, consider that we have system with one particle A and B each at positions $\mathbf{x}_A, \mathbf{x}_B$ and we perform the forward reaction to a system with one particle C at position $\mathbf{x}_C$. Detailed balance implies that the equilibrium probability of being in the A, B system at $\mathbf{x}_A, \mathbf{x}_B$ times the forward reaction rate must be equal to the equilibrium probability of being in the C system at $\mathbf{x}_C$ times the backward reaction rate, and this must be true for all system configurations. Vice versa, enforcing detailed balance is a technically convenient way to automatically achieve a desired equilibrium distribution. It implies a relationship between forward and backward reaction rates and also that the reaction scheme that allows for a forward reaction $\mathbf{x}_A, \mathbf{x}_B \longrightarrow \mathbf{x}_C$ must also allow for the reverse reaction, and vice versa. For non-interacting PBRD, a detailed balance scheme was first introduced in [MW08]. Other schemes have been developed more recently [KS14; DYK18]. In Sec. 2.3, we develop a general detailed-balance scheme for iPRD (iPRD-DB). The scheme includes a Metropolis-Hastings [Met+53; Has70] acceptance step that ensures the resulting dynamics fulfill detailed balance for abitrary configurations of interacting particles. In the dilute limit (one A and B particle pair reacting to a single C particle and back), the proposal steps are designed such that they are always accepted and the desired equilibrium association and dissociation rate constants are obtained. When the so-parametrized particles enter a dense phase, the kinetics and equilibria will naturally change, but do so in a physically realistic manner. In particular, we show that in a dense particle system where the reaction $\mathrm{A}+\mathrm{B} \rightleftarrows \mathrm{C}$ involves a change in effective particle volume, the iPRD-DB scheme leads to a solution that is consistent with Le Chatelier's principle, while a regular Doi scheme that ignores detailed balance produces unphysical solutions.

The implementation of the iPRD-DB scheme is included in the ReaDDy 2 software package [HFN19].

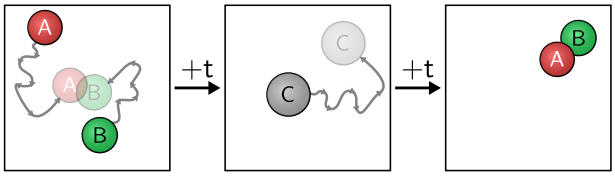

Figure 2.1: Schematic time evolution of a reaction-diffusion system of an isolated pair subject to the reaction A + B $\rightleftarrows$ C with the Doi model. Particles A and B diffuse and can form a complex particle C when they are closer than a certain reaction radius, here depicted as the sum of the radii of the two particles. The complex particle C diffuses as well and can dissociate into A and B again. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

2.2 Bimolecular reaction in equilibrium

We consider a system of molecules with three species, in which molecules A and B reversibly form a complex C. We want to simulate particle dynamics involving such reactions with iPRD, where particles interact with a potential when they are close, and a certain microscopic reaction scheme is employed, see Fig. 2.1. This section answers the question how the microscopic parameters of this reaction scheme need to be chosen such that the equilibrium constant and the dissociation rate measurable in a bulk experiments will be reproduced. This result will be used in the next section as part of designing a scheme obeying detailed balance.

2.2.1 Macroscopic rate model

The macroscopic reaction dynamics is described by the scheme

$$\mathrm{A} + \mathrm{B} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \mathrm{C}, \tag{2.1}$$

where k_{on} is a macroscopic bimolecular association rate constant, measured in units of per time and per concentration, while k_{off} is the dissociation rate constant, measured in units of per time. These are related to the macroscopic dissociation constant K_d, measured in units of concentration:

$$K_d = \frac{k_{\mathrm{off}}}{k_{\mathrm{on}}}. \tag{2.2}$$

We assume that both the association- and the dissociation process obey a linear rate law [ADK18], according to the law of mass action (LMA). We define the effective association rate K_{on}

$$K_{\mathrm{on}} = k_{\mathrm{on}} V^{-1}, \tag{2.3}$$

which is the frequency of association per AB complex. Likewise we define the effective dissociation rate K_{off}

$$K_{\mathrm{off}} = k_{\mathrm{off}}, \tag{2.4}$$

which is the frequency of dissociation per C molecule. We denote π_i as the stationary probability of state i. The ratio of stationary probabilities π_{AB}/π_C is given by the ratio of effective rates in equilibrium, where the number of association events per time is equal to the number of dissociation events per time

$$\frac{\pi_{AB}}{\pi_C} = \frac{K_{\mathrm{off}}}{K_{\mathrm{on}}} = \frac{[A]_{\mathrm{eq}}[B]_{\mathrm{eq}}}{[C]_{\mathrm{eq}}} V = K_d V. \tag{2.5}$$

2.2.2 Microscopic distribution

For the following we will assume that there is only either one pair of A and B particles or one C particle which live inside the volume V. The vectors $\mathbf{x} \in \mathbb{R}^9$, contain the euclidean positions for three particles. Individual positions are denoted by $\mathbf{x}_a$, $\mathbf{x}_b$, and $\mathbf{x}_c$ for particles A, B, and C respectively. Additionally there is a phase $i \in \{AB, C\}$, where AB is the dissociated phase and C is the associated phase. The joint distribution for states $x_i = (\mathbf{x}, i) \in \mathbb{R}^9 \times \{AB, C\}$ of finding the system in phase i and particle positions $\mathbf{x}$ is

$$p(x_i) = \begin{cases} \pi_{AB}\, p_{AB}(\mathbf{x}) & \text{for } i = AB \\ \pi_C\, p_C(\mathbf{x}) & \text{for } i = C \end{cases} \tag{2.6}$$

Note that in phase AB there is still a position for the C particle, such that the dimension of the microscopic phase space is equal for both phases. The same occurs for the positions of A and B in the phase C. In both cases, the residual variables have no effect. In phase space integrals these will be accounted for by a volume factor. Hence all phase space integrals use the measure $\mathrm{d}\mathbf{x} = \mathrm{d}\mathbf{x}_a \mathrm{d}\mathbf{x}_b \mathrm{d}\mathbf{x}_c$, where each $\mathrm{d}\mathbf{x}_j$ has units of volume. Introducing a Fock space for treatment of changing number of particles is circumvented by considering at most three particles - the isolated pair and the complex - and having the non existing particles contribute a constant factor to the partition function.

In phase AB the two particles A and B are subject to an interaction potential $U(\mathbf{x}) = U(|\mathbf{x}_b - \mathbf{x}_a|) = U(r)$ depending only on the distance $r = |\mathbf{x}_b - \mathbf{x}_a|$ of A and B. The potential is cut off at R_{int}, i.e. $U(r) = 0$, if $r > R_{\text{int}}$. The stationary distribution of positions $\mathbf{x}$ in phase AB is

$$p_{AB}(\mathbf{x}) = Z_{AB}^{-1} \exp(-\beta U(r)) \quad \text{with} \quad r = |\mathbf{x}_b - \mathbf{x}_a|$$

where $\beta^{-1} = k_B T$ is the thermal energy of the system which is coupled to a heat bath with temperature T and the normalization constant can be computed as follows (see Appendix A),

$$Z_{AB} = V^2 (V - V_{\text{ex}}) \tag{2.7}$$

$$V_{\text{ex}} = V_{\text{int}} - V_{\text{int}}^{\text{eff}} \tag{2.8}$$

$$V_{\text{int}} = \frac{4}{3} \pi R_{\text{int}}^3 \tag{2.9}$$

$$V_{\text{int}}^{\text{eff}} = \int_0^{R_{\text{int}}} e^{-\beta U(r)} 4\pi r^2 \mathrm{d}r, \tag{2.10}$$

where V_{int} is the interaction volume of the reactive particles, $V_{\text{int}}^{\text{eff}}$ the effective accessible volume due to particle interaction and V_{ex} is the reduction of the accessible volume.

In phase C the stationary distribution of positions $\mathbf{y}$ is

$$p_C(\mathbf{y}) = Z_C^{-1}$$

with the partition function

$$Z_C = \int \mathrm{d}\mathbf{y} = \iiint \mathrm{d}\mathbf{y}_a \mathrm{d}\mathbf{y}_b \mathrm{d}\mathbf{y}_c = V^3.$$

2.2.3 Doi reaction model

The microscopic reaction model is defined by the association rate function $\lambda^+(\mathbf{x})$ and the dissociation rate function $\lambda^-(\mathbf{y})$. The former describes the probability per unit time with which two particles A and B can react when the system is in phase AB and depends on positions $\mathbf{x}$. The latter describes the probability per unit time with which a C particle dissociates into A and B when the system is in phase C. We assume

that $\lambda^+(\mathbf{x})$ is radially symmetric, i.e. it only depends on $r = |\mathbf{x}_b - \mathbf{x}_a|$. Any microscopic reaction model, described by $\lambda^+(\mathbf{x})$ will result in an effective association rate $K_{\text{on}}^{\text{micro}}$ which reads

$$K_{\text{on}}^{\text{micro}} = \int \lambda^+(\mathbf{x}) p_{AB}(\mathbf{x}) \mathrm{d}\mathbf{x} \tag{2.11}$$

For $\lambda^+(\mathbf{x})$ and $\lambda^-(\mathbf{y})$ we use the Doi reaction model as depicted in Fig. 2.1, i.e. the microscopic association reaction rate function is a constant λ_{on}, when particles A and B are closer than the reaction radius R_{reac}

$$\lambda^+(\mathbf{x}) = \lambda_{\text{on}} \chi_{\text{reac}}(r) \quad \text{with} \quad r = |\mathbf{x}_b - \mathbf{x}_a|, \tag{2.12}$$

where $\chi_{\text{reac}}(r)$ indicates that A and B are within reactive distance

$$\chi_{\text{reac}}(r) = \begin{cases} 1, & \text{if } r < R_{\text{reac}} \\ 0, & \text{otherwise.} \end{cases} \tag{2.13}$$

The microscopic dissociation rate function is constant and chosen equal to the macroscopic dissociation rate constant

$$\lambda^-(\mathbf{y}) = k_{\text{off}}. \tag{2.14}$$

We evaluate the effective microscopic association rate (2.11) for the Doi reaction model (2.12) and obtain

$$\begin{aligned} K_{\text{on}}^{\text{micro}} &= \lambda_{\text{on}} Z_{AB}^{-1} V^2 \int_0^{R_{\text{reac}}} e^{-\beta U(r)} 4\pi r^2 \mathrm{d}r \\ &= \lambda_{\text{on}} \frac{V_{\text{reac}}^{\text{eff}}}{V - V_{\text{ex}}} \end{aligned} \tag{2.15}$$

where the effective reaction volume $V_{\text{reac}}^{\text{eff}}$ takes a similar form as the effective interaction volume , but with another radius R_{reac}

$$V_{\text{reac}}^{\text{eff}} = \int_0^{R_{\text{reac}}} e^{-\beta U(r)} 4\pi r^2 \mathrm{d}r. \tag{2.16}$$

2.2.4 Computing the microscopic association rate constant that reproduces the macroscopic equilibrium

For the following we will assume a given dissociation constant K_d and a given dissociation rate constant k_{off}. Using Eqs. (2.3, 2.2) we state the effective association rate according to the law of mass action

$$K_{\text{on}} = \frac{k_{\text{off}}}{K_d V}. \tag{2.17}$$

We require that the micro- and macroscopic effective rates match

$$K_{\text{on}}^{\text{micro}} \stackrel{!}{=} K_{\text{on}} \tag{2.18}$$

and find the restrictions on the microscopic reaction model. This results in a choice for the microscopic association rate constant λ_{on}, that will yield the desired equilibrium as in Eq. (2.5). We will call this specific value $\tilde{\lambda}_{\text{on}}$

$$\tilde{\lambda}_{\text{on}} = \frac{k_{\text{off}}}{K_d V} \frac{V - V_{\text{ex}}}{V_{\text{reac}}^{\text{eff}}}. \tag{2.19}$$

The relation of this expression to other diffusion influenced rate calculations is discussed in Appendix B.

2.3 Interacting-Particle Reaction Dynamics with Detailed Balance

Transition rates k^+ and k^- of association (+) and dissociation (−) respectively between states x_{AB} and y_C, with stationary probability distributions p defined in Eq. (2.6) shall obey detailed balance

$$p(x_{AB})k^+(\mathbf{y}|\mathbf{x}) = p(y_C)k^-(\mathbf{x}|\mathbf{y}). \tag{2.20}$$

We split the transition rates k into proposal rate and acceptance probability

$$\begin{aligned} k^+(\mathbf{y}|\mathbf{x}) &= \lambda^+(\mathbf{x})q^+(\mathbf{y}|\mathbf{x})\ \alpha^+(\mathbf{y}|\mathbf{x}) && \text{association} \\ k^-(\mathbf{x}|\mathbf{y}) &= \underbrace{\lambda^-(\mathbf{y})q^-(\mathbf{x}|\mathbf{y})}_{\text{proposal}}\ \underbrace{\alpha^-(\mathbf{x}|\mathbf{y})}_{\text{acceptance}} && \text{dissociation} \end{aligned} \tag{2.21}$$

where $\lambda^+(\mathbf{x})$ is the absolute rate of proposing a transition A + B ⟶ C when in particle configuration $\mathbf{x}$. $q^+(\mathbf{y}|\mathbf{x})$ is the normalized density to propose the positions $\mathbf{y}$, given that the positions were $\mathbf{x}$. $\alpha^+(\mathbf{y}|\mathbf{x})$ is the absolute probability of accepting the proposed positions. Similarly $\lambda^-(\mathbf{y})$ is the absolute rate of proposing a transition C ⟶ A + B, $q^-(\mathbf{x}|\mathbf{y})$ is the according proposal density and $\alpha^-(\mathbf{x}|\mathbf{y})$ the absolute probability of accepting the proposal. All q and α satisfy

$$\int q^i(\mathbf{y}|\mathbf{x})\mathrm{d}\mathbf{y} = 1 \quad \text{and} \quad \alpha^i(\mathbf{y}|\mathbf{x}) \leqslant 1 \quad \text{for } i \in \{+,-\}$$

2.3.1 Derive the backward proposal from the forward proposal

We assume the association proposal density q^+ as given, and want to derive the dissociation proposal density q^- and both α^+ and α^- subject to detailed balance. Therefore we include all terms that depend on the particle positions into the reverse proposal density q^-, such that most terms in Eq. (2.20) cancel and acceptances α^+ and α^- become independent of the particle positions of the dissociated phase. The reverse proposal density reads

$$q^-(\mathbf{x}|\mathbf{y}) = Q(\mathbf{y})^{-1}q^+(\mathbf{y}|\mathbf{x})\frac{\lambda^+(\mathbf{x})}{\lambda^-(\mathbf{y})}\frac{p_{AB}(\mathbf{x})}{p_C(\mathbf{y})} \tag{2.22}$$

with the normalization function $Q(\mathbf{y})$ such that

$$Q(\mathbf{y}) = \frac{1}{\lambda^-(\mathbf{y})p_C(\mathbf{y})}\int q^+(\mathbf{y}|\mathbf{x})\lambda^+(\mathbf{x})p_{AB}(\mathbf{x})\mathrm{d}\mathbf{x} \tag{2.23}$$

Note that Q must depend on $\mathbf{y}$ to fulfil the normalization $\forall\mathbf{y}$ (in the Doi model it will reduce to a constant). Inserting Eqs. (2.22, 2.21) into Eq. (2.20), the detailed balance condition reduces to

$$\frac{\alpha^+(\mathbf{y}|\mathbf{x})}{\alpha^-(\mathbf{x}|\mathbf{y})} = \frac{\pi_C}{\pi_{AB}}\frac{1}{Q(\mathbf{y})} \tag{2.24}$$

Reminding that $\alpha \leqslant 1$ naturally leads to using the Metropolis-Hastings [Has70; Met+53] acceptance function

$$\begin{aligned} \alpha^+(\mathbf{y}|\mathbf{x}) &= \min\left\{1, \frac{\pi_C}{\pi_{AB}\,Q(\mathbf{y})}\right\} \\ \alpha^-(\mathbf{x}|\mathbf{y}) &= \min\left\{1, \frac{\pi_{AB}\,Q(\mathbf{y})}{\pi_C}\right\} \end{aligned} \tag{2.25}$$

which fulfills the given detailed balance condition (2.24). For a practical implementation one needs to know both proposal densities $q^+(\mathbf{y}|\mathbf{x})$ and $q^-(\mathbf{x}|\mathbf{y})$, and the corresponding acceptance probabilities $\alpha^+(\mathbf{y}|\mathbf{x})$ and $\alpha^-(\mathbf{x}|\mathbf{y})$.

2.3.2 Apply DB to Doi model

Assuming the Doi model (2.12, 2.14), we state the association proposal density $q^+(\mathbf{y}|\mathbf{x})$ and derive the dissociation proposal density $q^-(\mathbf{x}|\mathbf{y})$ (2.22). The normalized association proposal density reads

$$q^+(\mathbf{y}|\mathbf{x}) = V^{-2}\delta\left(\mathbf{y}_c - \frac{\mathbf{x}_a + \mathbf{x}_b}{2}\right) \tag{2.26}$$

where the Dirac delta function $\delta(\cdot)$ assures that the C particle's proposed position $\mathbf{y}_c$ is in the middle between the A and B particles from the initial positions $\mathbf{x}$. The volume term V^{-2} is required for normalization, due to the measure $\mathrm{d}\mathbf{y} = \mathrm{d}\mathbf{y}_a\mathrm{d}\mathbf{y}_b\mathrm{d}\mathbf{y}_c$. Additionally the volume term can be understood as a uniform placement of A and B in the final positions $\mathbf{y}$. Since A and B are not considered in the associated state, it is irrelevant where they are. Hence Eq. (2.26) fulfills $\int q^+(\mathbf{y}|\mathbf{x})\mathrm{d}\mathbf{y} = 1$. The normalization Q of the dissociation proposal density from Eq. (2.23) can be evaluated and reduces to a constant (see Appendix C)

$$Q = \frac{\lambda_{\mathrm{on}}}{k_{\mathrm{off}}}\frac{V_{\mathrm{reac}}^{\mathrm{eff}}}{V - V_{\mathrm{ex}}}. \tag{2.27}$$

The dissociation proposal density (2.22) then becomes

$$\begin{aligned} q^-(\mathbf{x}|\mathbf{y}) = \left(V V_{\mathrm{reac}}^{\mathrm{eff}}\right)^{-1}\delta\left(\mathbf{y}_c - \frac{\mathbf{x}_a + \mathbf{x}_b}{2}\right)\ldots \\ \times \chi_{\mathrm{reac}}(r)e^{-\beta U(r)}, \end{aligned} \tag{2.28}$$

with $r = |\mathbf{x}_b - \mathbf{x}_a|$. This density can be read as: given a C particle at position $\mathbf{y}_c$, positions $\mathbf{x}_a$ and $\mathbf{x}_b$ of particles A and B are restricted to radial shells concentric around $\mathbf{y}_c$ due to the delta function. These shells must not be larger than the reaction radius due to the indicator function. The distance is additionally weighted with the Boltzmann factor of the interaction potential U.

Using the normalization constant Q from Eq. (2.27) the acceptance probabilities from Eq. (2.25) are directly obtained. Using the microscopic association rate given in Eq. (2.19) results in an acceptance probability of unity in both directions

$$\alpha^+(\mathbf{y}|\mathbf{x}) = \alpha^-(\mathbf{x}|\mathbf{y}) = 1 \quad \text{for } \lambda_{\mathrm{on}} = \tilde{\lambda}_{\mathrm{on}} \text{ from (2.19)}.$$

2.3.3 Generalize for other types of reactions

The presented Metropolis-Hastings Monte Carlo method can be performed for other types of reversible reactions, namely reversible conversion reactions of the type

$$\underbrace{\mathrm{A}}_{\mathbf{x}} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \underbrace{\mathrm{B}}_{\mathbf{y}} \quad \text{with } \lambda^+(\mathbf{x}), \lambda^-(\mathbf{y}) \tag{2.29}$$

as well as reversible enzymatic reactions of the type

$$\underbrace{\mathrm{A} + \mathrm{C}}_{\mathbf{x}} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \underbrace{\mathrm{B} + \mathrm{C}}_{\mathbf{y}} \quad \text{with } \lambda^+(\mathbf{x}), \lambda^-(\mathbf{y}) \text{ and } R_{\mathrm{reac}}, \tag{2.30}$$

with macroscopic forward and backward rates k_{on} and k_{off}. For those two reactions we can also construct a microscopic probability density for positions $\mathbf{x}$ and $\mathbf{y}$ for the dilute case in the fashion of Eq. (2.6). Here the microscopic phase space only has positions for A and B particles, the C particle in reaction (2.30) can be placed at the origin without loss of generality. The reaction functions $\lambda^+(\mathbf{x})$ and $\lambda^-(\mathbf{y})$ for the

	$\underbrace{\mathrm{A}+\mathrm{B}}_{\mathbf{x}} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \underbrace{\mathrm{C}}_{\mathbf{y}}$	$\underbrace{A}_{\mathbf{x}} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \underbrace{\mathrm{B}}_{\mathbf{y}}$	$\underbrace{\mathrm{A}+\mathrm{C}}_{\mathbf{x}} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftarrows}} \underbrace{\mathrm{B}+\mathrm{C}}_{\mathbf{y}}$
$\lambda^+(\mathbf{x})$	$\lambda_{\mathrm{on}}\chi_{\mathrm{reac}}(\mathbf{x})$	λ_{on}	$\lambda_{\mathrm{on}}\chi_{\mathrm{reac}}(\mathbf{x})$
$\lambda^-(\mathbf{y})$	λ_{off}	λ_{off}	$\lambda_{\mathrm{off}}\chi_{\mathrm{reac}}(\mathbf{y})$
$q^+(\mathbf{y}\|\mathbf{x})$	$V^{-2}\delta\left(\mathbf{y}_c - \frac{\mathbf{x}_a+\mathbf{x}_b}{2}\right)$	$\delta(\mathbf{y}-\mathbf{x})$	$\delta(\mathbf{y}-\mathbf{x})$
$q^-(\mathbf{x}\|\mathbf{y})$	$\left(V V_{\mathrm{reac}}^{\mathrm{eff}}\right)^{-1}\delta\left(\mathbf{y}_c - \frac{\mathbf{x}_a+\mathbf{x}_b}{2}\right)$ $\ldots \times \chi_{\mathrm{reac}}(\mathbf{x}) e^{-\beta U_{AB}(\mathbf{x})}$	$\delta(\mathbf{x}-\mathbf{y})$	$\delta(\mathbf{x}-\mathbf{y})$
$f^+(\mathbf{y}\|\mathbf{x})$	$e^{-\beta(E(\mathbf{y})-[E(\mathbf{x})-U_{AB}(\mathbf{x})])}$	$e^{-\beta(E(\mathbf{y})-E(\mathbf{x}))}$	$\frac{V_{\mathrm{reac,A}}^{\mathrm{eff}}}{V_{\mathrm{reac,B}}^{\mathrm{eff}}} e^{-\beta(E(\mathbf{y})-E(\mathbf{x}))}$
$f^-(\mathbf{x}\|\mathbf{y})$	$e^{-\beta([E(\mathbf{x})-U_{AB}(\mathbf{x})]-E(\mathbf{y}))}$	$e^{-\beta(E(\mathbf{x})-E(\mathbf{y}))}$	$\frac{V_{\mathrm{reac,B}}^{\mathrm{eff}}}{V_{\mathrm{reac,A}}^{\mathrm{eff}}} e^{-\beta(E(\mathbf{x})-E(\mathbf{y}))}$
constraints	$k_{\mathrm{on}} = \lambda_{\mathrm{on}} V \frac{V_{\mathrm{reac}}^{\mathrm{eff}}}{V-V_{\mathrm{ex}}}$ $k_{\mathrm{off}} = \lambda_{\mathrm{off}}$	$k_{\mathrm{on}} = \lambda_{\mathrm{on}}$ $k_{\mathrm{off}} = \lambda_{\mathrm{off}}$	$k_{\mathrm{on}} = \lambda_{\mathrm{on}} V \frac{V_{\mathrm{reac,A}}^{\mathrm{eff}}}{V-V_{\mathrm{ex,A}}}$ $k_{\mathrm{off}} = \lambda_{\mathrm{off}} V \frac{V_{\mathrm{reac,B}}^{\mathrm{eff}}}{V-V_{\mathrm{ex,B}}}$

Table 2.1: Summary of the iPRD-DB quantities for three different kinds of reversible reactions: reversible association (see Sec. 2.2.2), reversible unimolecular conversion, and reversible bimolecular enzymatic reaction (see Sec. 2.3.3). Quantities are: absolute proposal rates λ, proposal densities q, and acceptance probabilities $\alpha = \min\{1, f\}$, as described in Sec. 2.3. Superscript + and – denote the "on" and "off" process respectively, corresponding to the definition of the reaction. $\mathbf{x}$ and $\mathbf{y}$ are the microscopic positions of particles. Constraints describe for which microscopic parameters the acceptance probabilities will be unity in the dilute limit. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

conversion reaction (2.29) are constants λ_{on} and λ_{off} respectively. For the enzymatic reaction (2.30) both reaction functions are additionally multiplied with an indicator function depending on the reaction radius R_{reac}. As in Sec. 2.2.4 we can compute the microscopic rate constants λ that reproduce the macroscopic kinetics in the dilute limit. In the case of the enzymatic reaction (2.30), there appear excluded volumes $V_{\mathrm{ex,A}}$, $V_{\mathrm{ex,B}}$ and effective reaction volumes $V_{\mathrm{reac,A}}^{\mathrm{eff}}$, $V_{\mathrm{reac,B}}^{\mathrm{eff}}$. These are defined analogously to the volumes V_{ex}, see Eq. (2.8), and $V_{\mathrm{reac}}^{\mathrm{eff}}$, see Eq. (2.16), with the difference that $V_{\mathrm{ex,A}}$ and $V_{\mathrm{reac,A}}^{\mathrm{eff}}$ are calculated based on the interaction potential of A and C, and $V_{\mathrm{ex,B}}$ and $V_{\mathrm{reac,B}}^{\mathrm{eff}}$ are calculated based on the interaction potential of B and C. To assure detailed balance we make the same ansatz for transition rates as in Eq. (2.21). The proposal densities q are constructed much simpler, because in these types of reactions no new positions must be generated, i.e. the q are delta functions. However during the species conversion, molecules might be subject to potentials with respect to educt and product states. We gather the change of potential energy during the reaction in the variable ΔE. We summarize all of these findings in Tab. 2.1.

2.4 Results

We have proposed a method of executing reversible reactions according to detailed balance. It can be used to perform reactions in a stochastic reaction-diffusion simulation. A schematic implementation is shown in the pseudo code Alg. 1.

In order to illustrate our method, we perform many-particle simulations with molecular species A, B and C engaging in the reversible association reaction shown in Eq. (2.1). The simulation is performed using

Algorithm 1: Reaction diffusion algorithm for n integration steps with time step size τ Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

```
initialize list of particles/system state p
repeat
    f ← calculate forces for state p
    p ← propagate diffusion subject to f and τ
    L ← list of possible reaction events in p
    while L not empty do
        select next event l from L
        u1 ← RANDOM-UNIFORM
        λ ← microscopic rate constant of l
        if u1 < 1 − exp(λτ) then
            E1 ← calculate energy of state p
            p ← propose event l according to density q
            E2 ← calculate energy of state p
            a ← acceptance for l and energies E1 and E2
            u2 ← RANDOM-UNIFORM
            if u2 < a then
                accepted, keep the state p
            else
                p ← revert the event l
        remove l out of L
        remove all events out L, that shared particles with event l
until n steps performed
```

overdamped Langevin dynamics in the particle interaction potential with a fixed time-step integrator. The potential $U(r)$ between the particles A and B is chosen as a harmonic repulsion with cutoff R_{int} and force constant κ, that only depends on the distance $r = |\mathbf{x}_b - \mathbf{x}_a|$ between A and B

$$U(r) = \begin{cases} \frac{1}{2}\kappa(r - R_{\text{int}})^2, & \text{if } r < R_{\text{int}} \\ 0, & \text{otherwise} \end{cases} \tag{2.31}$$

For this choice of potential the effective interaction volume from Eq. (2.10) yields an expression containing errorfunctions. In general the effective interaction volume can be determined numerically.

During one time step of length τ, we first integrate the diffusive motion of particles and then perform the reactions. The boundaries of the system are periodic, obeying the minimum image convention and wrapping positions upon crossing the border.

In the reaction step all possible reaction events are determined, this depends on the considered reactions, reaction radii and the current particle configuration. Then the list of reaction events is processed. An event is selected from the list. The event will be proposed with absolute probability $p = 1 - \exp(-\lambda\tau)$ depending on the microscopic rate constant λ of the associated type of reaction. The event is performed, generating another particle configuration drawn from the proposal densities in Eqs. (2.26, 2.28). From the change in potential energy and the type of reaction the acceptance probability a is calculated. If the event is accepted the new configuration is kept. If the event is rejected the old configuration has to be restored. Then the processed event is removed from the list of events. Additionally any event is removed that would propose an event with the same particles as the processed one, since these might not exist anymore.

The total probability of performing a particular event is pa. If a is chosen according to Eq. (2.25) and Eq. (2.27) and the proposal density of the dissociation reaction includes the Boltzmann factor as in Eq. (2.28), we will refer to this as the proposed **DB** reaction scheme. We refer to the **Doi** reaction scheme if $a = 1$, regardless of the energy difference, and if the proposal density does not include the Boltzmann factor of the interaction potential of the reactants.

2.4.1 Dilute limit

We validate Alg. 1 by performing it on the system of particles A, B and C. These particles are subject to the reaction (2.1) and a harmonic repulsion potential as in Eq. (4.6). At any point in time there is either the C particle or two particles A and B, i.e. there is only one instance of each molecule species. Thus these simulations are in the dilute limit. The only interactions occur between the A and B particle.

Validation of reaction kinetics

We show that the proposed detailed balance reaction scheme always yields the desired macroscopic equilibrium distribution π_{AB}/π_C from Eq. (2.5). Additionally we demonstrate under which circumstances the simulated effective on- and off-rates, K_{on} and K_{off}, will match those given by Eq. (2.3) and Eq. (2.4). The results are seen in Fig. 2.2. The simulation parameters are given in Table 2.2.

Fig. 2.2a shows that for very low λ_{on}, the effective association rate K_{on} cannot exceed a certain value because the proposal frequency is limited and K_{off} is in turn diminished by rejection of dissociation events in order to reproduce the desired equilibrium constant $\pi_{AB}/\pi_C = K_d V$. For very high λ_{on}, association events will be rejected, thus limiting K_{on} to the LMA value, while dissociation events are executed with frequency $K_{\text{off}} = k_{\text{off}}$. The transition between these two regimes is where $\lambda_{\text{on}} = \tilde{\lambda}_{\text{on}}$ as in Eq. (2.19). Fig. 2.2b shows that, when one uses the appropriate association rate constant from Eq. (2.19), one can reproduce the expected reaction kinetics for varying K_d.

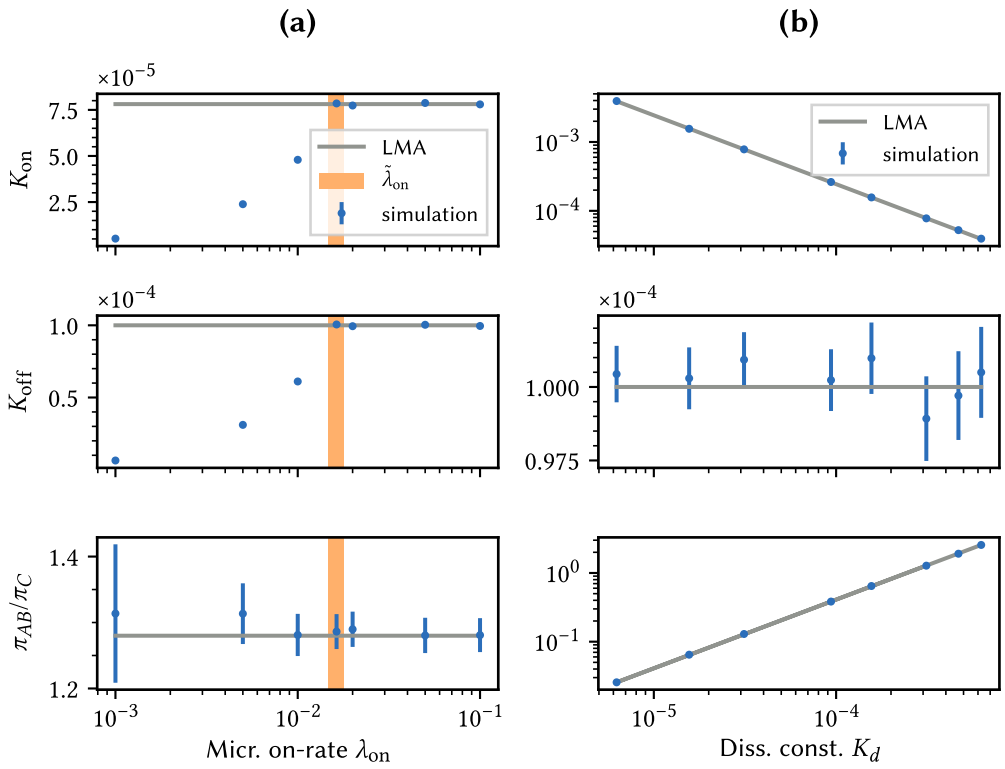

Figure 2.2: Validation of the proposed detailed balance reaction scheme in dilute systems by stochastic particle-based reaction-diffusion simulations (see Alg. 1). Shown are observables of the macroscopic reaction kinetics: the effective association rate K_{on}, the effective dissociation rate K_{off} and the equilibrium constant π_{AB}/π_C. Reference values (law of mass action - LMA) for K_{on}, K_{off} and π_{AB}/π_C correspond to macroscopic behavior described in Sec. 2.2. See simulation parameters in Tab. 2.2. **(a)** Microscopic association rate constant λ_{on} is varied. $\tilde{\lambda}_{on}$ corresponds to Eq. (2.19). **(b)** The given dissociation constant K_d is varied. The microscopic association rate constant is $\lambda_{on} = \tilde{\lambda}_{on}(K_d)$. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

Quantity	Symbol	Value
Dissociation constant	K_d	3.125×10^{-4}
Dissociation rate constant	k_{off}	10^{-4}
Volume	V	$16 \times 16 \times 16$
Diffusion constant of each particle	D	5
Reaction radius	R_{reac}	2
Interaction radius	R_{int}	2
Force constant	κ	5
Time step length		
in Fig. 2.2	τ_1	10^{-4}
in Fig. 2.3	τ_2	1.25×10^{-5}
Number of integration steps		
in Fig. 2.2	m_1	3×10^{10}
in Fig. 2.3	m_2	4.8×10^{11}

Table 2.2: Unitless parameters used in the simulations of dilute systems, see Fig. 2.2 and 2.3. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

Microscopic reversibility

We now demonstrate that the proposed DB reaction scheme (Alg. 1) indeed produces trajectories in thermodynamic equilibrium, while the naive Doi scheme leads to periodic cycles in phase space, corresponding to an unintended nonequilibrium scenario. To this end, we distinguish three substates of the dissociated state, defined by the inter-particle distance r of particles A and B, and the reaction radius R. We define states 1-4 as follows:

1. The complex state, C
2. A and B are very close $r \leqslant \frac{3}{4}R$
3. A and B are still in reactive range $\frac{3}{4}R < r \leqslant R$
4. A and B are not within reactive range $r > R$

Using again a reversibly reacting system with a single A, B pair or a single C complex, we determine the stationary distribution π for this definition of states, and the transition rates K connecting them. A process that fulfils detailed balance must yield

$$\pi_i K_{ij} = \pi_j K_{ji} \tag{2.32}$$

for all pairs of states i, j. We measure π and K from simulations and compare the Doi reaction scheme and the proposed DB reaction scheme in the presence of a harmonic repulsion potential between A and B. In this comparison all system parameters are identical, only the reaction mechanism differs. Results are presented in Fig. 2.3 and simulation parameters are given in Tab. 2.2.

From Fig. 2.3 it is evident that for the present case of interacting particles, the naive Doi reaction scheme produces a cyclic probability flux that violates DB. In the proposed DB reaction scheme, this is not the case and all given probability fluxes obey Eq. (2.32).

Note that for both reaction schemes, there occurs a unidirectional transition $4 \longrightarrow 1$ due to the time splitting we employ during one simulation step (first the diffusion step and then the reaction step). This artificial transition is a result of the time-step discretization error and not related to the DB scheme. It

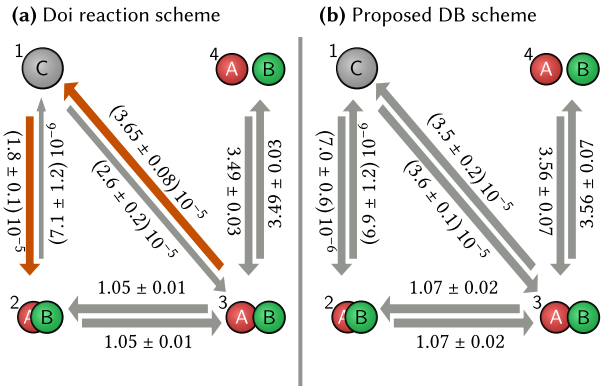

Figure 2.3: Probability fluxes between associated and dissociated states measured from particle-based reaction-diffusion simulations (see Alg. 1) in the dilute limit. Compared are the Doi reaction scheme and the proposed detailed balance reaction scheme (DB). Definitions of the states 1-4 are given in Sec. 2.4.1. Arrows depict transitions between these states as observed in the simulations. The width of the arrows encodes the probability flux $\pi_i K_{ij}$, also given as numeric values measured from multiple independent simulations giving rise to the standard error of the mean. The widths of two adjacent arrows are normalized with respect to each other (not globally). See parameters in Tab. 2.2. **(a)** Doi reaction scheme. The probability fluxes for the transitions $1 \longrightarrow 2$ and $1 \longrightarrow 3$ are imbalanced compared to their respective counterparts, resulting in a circular flux of probability. **(b)** Detailed balance reaction scheme. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

Quantity	Symbol	Value
Dissociation constant	K_d	2×10^{-2}
Dissociation rate constant	k_{off}	10^{-3}
Volume	V	$20 \times 20 \times 20$
Particle radii		
case $r_A^3 + r_B^3 < r_C^3$	(r_A, r_B, r_C)	$(1, 1, 1.4)$
case $r_A^3 + r_B^3 > r_C^3$	(r_A, r_B, r_C)	$(1, 1, 1.1)$
Diffusion constants per radius		
for species $i \in \{A, B, C\}$	D/r_i	5
Interaction radius for pair		
of species $(i, j) \forall i, j \in \{A, B, C\}$	$R_{\text{int}}(i, j)$	$r_i + r_j$
Reaction radius	R_{reac}	2
Force constant	κ	10
Time step length	τ	5×10^{-4}
Time steps until equilibrated		
dilute system with $n = 50$	m_{dilute}	1.2×10^8
dense system with $n = 900$	m_{dense}	9×10^6

Table 2.3: Unitless parameters used in the simulations of dense systems, see Fig. 2.4 and 2.5. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

occurs with an absolute rate of less than 10^{-6}, all other transitions have $K_{ij} > 10^{-5} \forall (i, j) \neq (4, 1)$. Thus its probability flux is not shown here.

2.4.2 System of many particles

Finally, we study how a dense mixture of interacting particles behaves when the DB algorithm is employed, and we compare this behavior with the naive Doi algorithm and what is expected from physical intuition. The Algorithm 1 is performed for a system of many A, B and C particles confined to the volume V with periodic boundaries. In this scenario we assign physical radii r_A, r_B, and r_C to the particles. Particles are subject to harmonic repulsion potentials (4.6) acting between all pairs of species A, B, and C, where the interaction radius is chosen as the sum of the particles' radii. See parameters in Tab. 2.3. Particles are subject to the reaction (2.1). Employing the DB reaction scheme introduced in Sec. 2.3 can therefore result in rejected Monte-Carlo moves, which will affect the thermodynamics and kinetics of the simulation system in the dense limit.

In Sec. 2.2.2 and following we had assumed that phase space consists of only three particles A, B and C. In the case of many possible reactants one is presented with multiple possible reaction events. For one particular event we will use the proposal densities from Eqs. (2.26, 2.28) to treat the particles taking part in the event. All other particles will be considered static excess objects. This means that the microscopic distributions from Eq. (2.6) gain another Boltzmann factor from interactions with the excess particles. Note that the partition functions Z_{AB} and Z_C will differ from their "dilute" values. In Sec. 2.3.2 we have seen that a particular choice of parameters leads to the prefactor in the acceptance becoming unity. Hence, the advantage of such a Markov Chain Monte Carlo algorithm is that one does not need to know constant factors of the stationary distribution to draw samples from said distribution. Along these lines we construct an acceptance function for the many particle case, that includes a Boltzmann factor of the energy difference and a prefactor of unity, assuming that internal reaction parameters correspond to a certain but unknown macroscopic equilibrium. We will use the association rate constant derived in

Eq. (2.19). Obviously this equilibrium will differ from the one in Eq. (2.5). But one can guarantee detailed balance never the less.

The change of potential energy is $\Delta\varepsilon$. It does not include the interaction between A and B as this is already accounted for by the proposal probabilities q^+ and q^-. We may write $\Delta\varepsilon$ as the total change of potential energy ΔE minus the interaction energy U_{AB}. We formulate the acceptance for the many particle case:

$$\begin{aligned} \alpha^+(\mathbf{y}|\mathbf{x}) &= \min\{1, \exp(-\beta\Delta\varepsilon^+)\} \\ \alpha^-(\mathbf{x}|\mathbf{y}) &= \min\{1, \exp(-\beta\Delta\varepsilon^-)\} \end{aligned} \tag{2.33}$$

where the changes of energies are given by

$$\begin{aligned} \Delta\varepsilon^+ &= E(\mathbf{y}) - [E(\mathbf{x}) - U_{AB}(\mathbf{x})] \\ \Delta\varepsilon^- &= [E(\mathbf{x}) - U_{AB}(\mathbf{x})] - E(\mathbf{y}). \end{aligned} \tag{2.34}$$

We set up the system with a certain number of A and B particles and no C particles. We control the quantity $n = (N_A + N_B)/2 + N_C$ which is conserved during a simulation. The system equilibrates without the reaction, we then switch the reaction on and let the system equilibrate again.

We compute three observables in the equilibrated state, i.e. when observables are stable and converged: the equilibrium constant $\pi_{AB}/\pi_C = V[A][B]/[C]$, the total potential energy of the system U in units of k_BT and the pressure P in units of $V^{-1}k_BT$. The pressure is measured from evaluating the virial term of acting forces as described in [AT87]. Individual reactions are integrated with either the proposed DB scheme or the Doi reaction scheme.

Fig. 2.4a shows the results for the case when an association reaction of A and B increases the total volume occupied by particles such that $r_A^3 + r_B^3 < r_C^3$. The associated state is energetically less favourable. In the dilute limit both methods Doi and DB reproduce the macroscopic equilibrium population $\pi_{AB}/\pi_C = K_dV$. For increasing number of particles both methods differ significantly. The Doi reaction scheme favours the energetically higher associated configuration C. The Doi scheme produces an equilibrium constant of roughly $\pi_{AB}/\pi_C \approx 80$ for the highest density simulated. The DB scheme adjusts the effective association probability by rejecting association events. This results in a steady state, where almost no C particles exist with an equilibrium constant exceeding $\pi_{AB}/\pi_C > 3 \times 10^3$. For all $n > 50$, the DB scheme finds a steady state of lower energy and lower pressure compared to the Doi scheme. Fig. 2.5a and b show representative simulation snapshots of the steady states for Doi and DB scheme.

Fig. 2.4b shows the case when a C particle occupies less volume than A and B combined such that $r_A^3 + r_B^3 > r_C^3$, which could correspond to two proteins A and B, which only fully fold in a bound state. In the dilute case both methods Doi and DB reproduce the same behaviour in all three observables. For increasing number of particles the Doi method produces a similar steady state population as in Fig. 2.4a where the C state is favoured. The DB scheme produces states favouring the C state even stronger thus reducing the system's potential energy and pressure compared to the Doi scheme. Fig. 2.5c and d show representative simulation snapshots of the steady states for Doi and DB scheme.

2.5 Conclusion

We have derived an algorithm to perform iPRD simulations of molecules undergoing reversible reactions of the form A + B $\rightleftarrows$ C according to detailed balance. This method is called iPRD-DB.

Detailed balance guarantees that simulations of an isolated system generate samples according to thermodynamic equilibrium. We have shown that in a dense reactive mixture of particles, that exhibit volume exclusion due to pair-wise potentials, the steady state of the system simulated with iPRD-DB is in agreement with Henri Le Chatelier's principle [ADK18], i.e. that the achieved steady state concentrations strongly depend on the interaction of molecules. Biochemical pathways often show switch-like behavior,

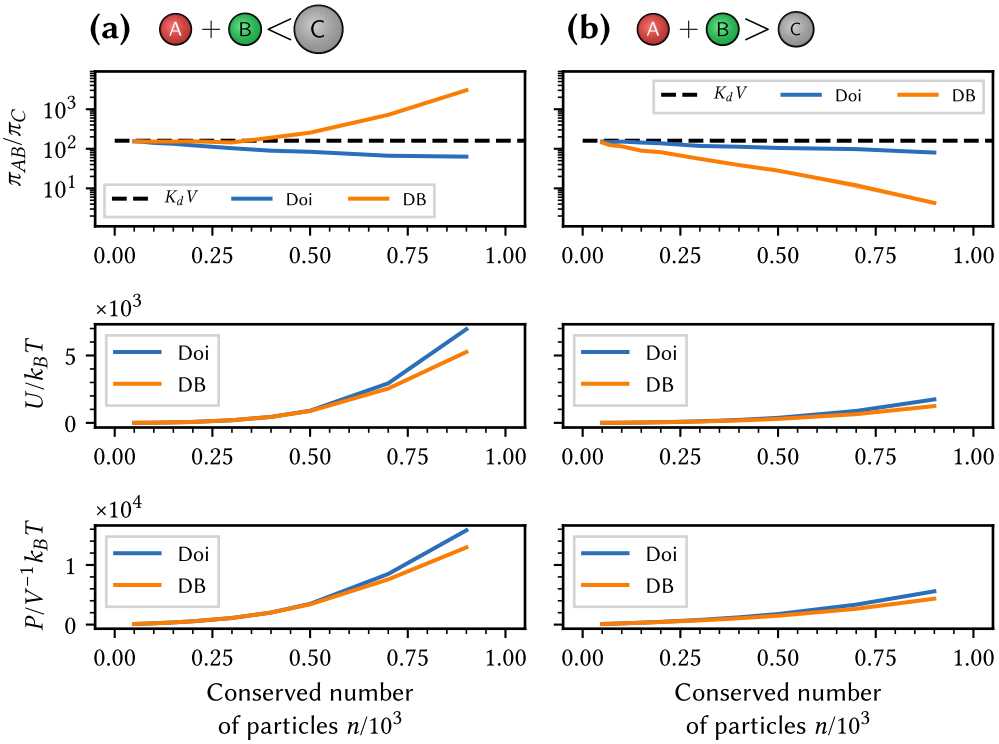

Figure 2.4: Steady state observables measured in particle-based reaction-diffusion simulations with multiple particles. The quantity $n = (N_A + N_B)/2 + N_C$ is conserved during a simulation. Shown are ensemble- and time-averaged values of the equilibrium constant $\pi_{AB}/\pi_C = V[A][B]/[C]$, the potential energy U in units of k_BT, the pressure P in units of $V^{-1}k_BT$. Compared are the two reaction schemes Doi and DB, see Sec. 2.4. See simulation parameters in Tab. 2.3 **(a)** An association reaction of A and B increases the total volume occupied by particles such that $r_A^3 + r_B^3 < r_C^3$. **(b)** The C particle occupies less volume than A and B combined such that $r_A^3 + r_B^3 > r_C^3$. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

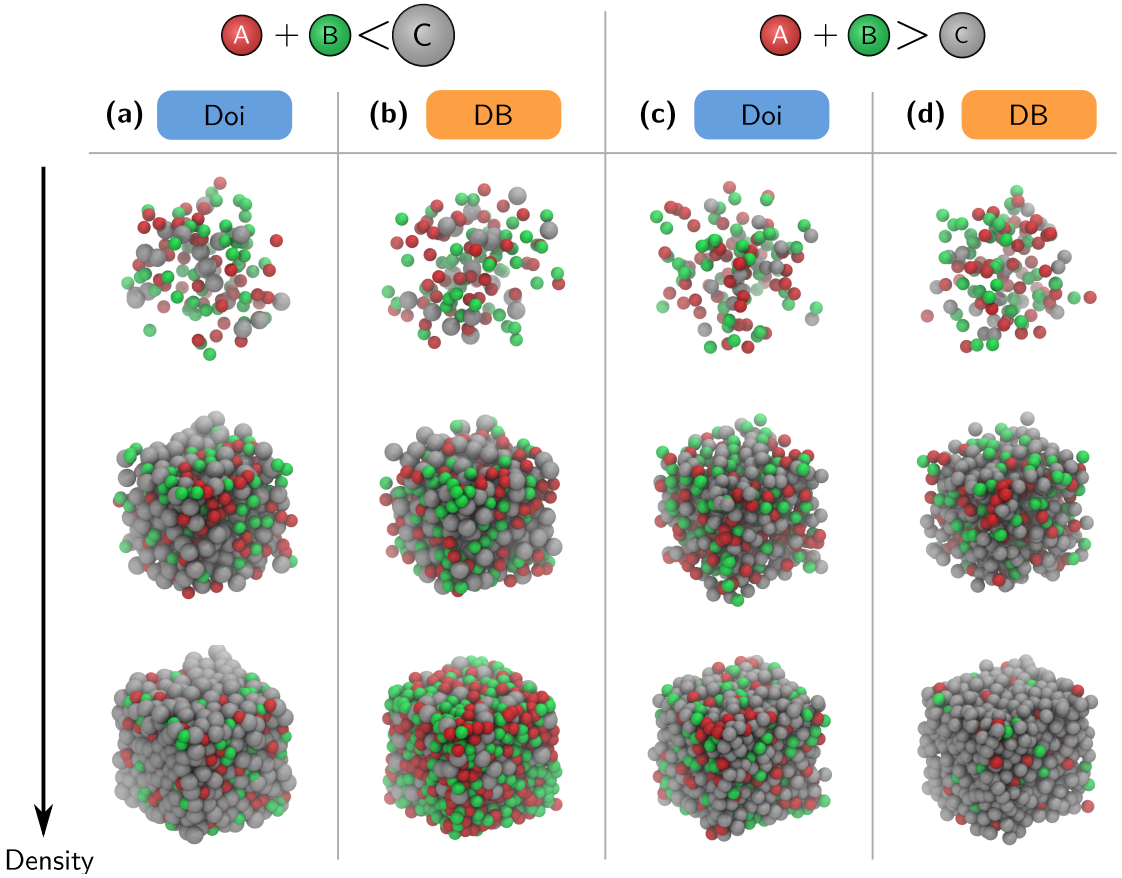

Figure 2.5: Steady state configurations of particle-based reaction-diffusion simulations subject to the reaction A + B $\rightleftarrows$ C for different densities in terms of the number of particles n initially in the system. Compared are the two reaction schemes Doi and DB, see Sec. 2.4 at different particle radii respectively. See simulation parameters in Tab. 2.3. **(a)** The associated state occupies more volume than the dissociated state, reactions are handled with the Doi scheme. **(b)** The associated state occupies more volume than the dissociated state, reactions are handled with the DB scheme **(c)** The associated state occupies less volume than the dissociated state, reactions are handled with the Doi scheme. **(d)** The associated state occupies less volume than the dissociated state, reactions are handled with the DB scheme. Reprinted with permission from *The Journal of Physical Chemistry B* "Reversible Interacting-Particle Reaction Dynamics", Fröhner and Noé, 2018. Copyright 2018 American Chemical Society.

and are thus sensitive to such changes in concentrations of agents [NB94; HN00; Mar+17]. Sampling the correct equilibrium is crucial when simulating such processes.

The iPRD-DB method can be generalized for other types of reactions, such as a reversible change of molecule species A $\rightleftarrows$ B, or a reversible enzymatic reaction A + C $\rightleftarrows$ B + C, which describes a Michaelis-Menten experiment when the backwards rate becomes very small.

Furthermore the iPRD-DB method is accompanied by an equation for the microsopic rate constant λ that assures the correct macroscopic reaction kinetics. This equation, see Eq. (2.19), relates the macroscopic kinetic parameters K_d and k_{off} in a dilute environment with the microscopic iPRD model parameters: microscopic rate constant λ, reaction radius R, and force parameters that determine the excluded volume V_{ex}. Thus, it provides a choice for λ, which in the iPRD-DB algorithm functions as the absolute proposal rate. For this choice the acceptance probability reduces to the Boltzmann factor describing the change of energy with respect to educt and product states. We also provide proposal densities such that the acceptance becomes unity in the dilute case.

Having measured K_d and k_{off} in an *in vitro* scenario, a microscopic iPRD model can be constructed subject to Eq. (2.19) and can then be analyzed numerically to gain insights about the *in vivo* process, where molecules may occur in very low copy numbers and diffuse anomalously due to complex geometries, making experimental measurements cumbersome in this regime. Note that the expression relating K_d and k_{off} with λ and R is independent of the diffusion coefficient D, i.e. an iPRD model can be adjusted to resemble the in vivo effective diffusion, which may, e.g. be obtained from flourescence correlation spectroscopy experiments [Tho02].

An open question is what the analytical reference chemical equilibrium is when going to dense particle mixtures.

Appendix

A. Normalization constant Z_{AB}

The normalization is

$$\begin{aligned} Z_{AB} &= \int e^{-\beta U(\mathbf{x})} \mathrm{d}\mathbf{x} \\ &= \int \mathrm{d}\mathbf{x}_c \iint e^{-\beta U(\mathbf{x}_b - \mathbf{x}_a)} \mathrm{d}\mathbf{x}_a \mathrm{d}\mathbf{x}_b \\ &= V\,(I_1 + I_2) \end{aligned}$$

If there are no external potentials present, the latter integral factorizes

$$\begin{aligned} I_2 &= \iint_{|\mathbf{x}_b - \mathbf{x}_a| > R_{\text{int}}} \mathrm{d}\mathbf{x}_a \mathrm{d}\mathbf{x}_b \\ &= \int \left(\int_{|\mathbf{x}_b - \mathbf{x}_a| > R_{\text{int}}} \mathrm{d}\mathbf{x}_b \right) \mathrm{d}\mathbf{x}_a \\ &= (V - V_{\text{int}}) \int \mathrm{d}\mathbf{x}_a = (V - V_{\text{int}})\, V \end{aligned}$$

where V_{int} is the interaction volume, that only depends on the cut-off distance of the potential R_{in}, not the potential itself. Since the potential U only depends on the relative position $\mathbf{x}_b - \mathbf{x}_a$, one can fix the

position of one particle without changing the value of the integral I_1

$$\begin{aligned} I_1 &= \iint_{|\mathbf{x}_b-\mathbf{x}_a|\leqslant R_{\text{int}}} e^{-\beta U(\mathbf{x}_b-\mathbf{x}_a)} \mathrm{d}\mathbf{x}_a \mathrm{d}\mathbf{x}_b \\ &= \int \left(\int_{|\mathbf{x}_b-\mathbf{x}_a|\leqslant R_{\text{int}}} e^{-\beta U(\mathbf{x}_b-\mathbf{x}_a)} \mathrm{d}\mathbf{x}_b \right) \mathrm{d}\mathbf{x}_a \\ &= V_{\text{int}}^{\text{eff}} \int \mathrm{d}\mathbf{x}_a = V_{\text{int}}^{\text{eff}}\ V \end{aligned}$$

The effective accessible volume inside the interaction radius is given by:

$$V_{\text{int}}^{\text{eff}} = V_{\text{int}} - V_{\text{ex}},$$

which defines the excluded volume V_{ex} due to interaction

B. Relation to diffusion-influenced rate constant derivations

To understand Eq. (2.19) we formulate the association rate constant for our problem using Eq. (2.2)

$$k_{\text{on}} = \tilde{\lambda}_{\text{on}} V \frac{V_{\text{reac}}^{\text{eff}}}{V - V_{\text{ex}}}. \tag{2.35}$$

This rate is linearly dependent on the effective reaction volume from Eq. (2.16), i.e. if one increases the repulsion force between particles A and B the association rate will decrease. One further notices that the diffusion of particles is not considered in this equation, since we assume they are at all times distributed according to Eq. (2.6). This is true only because of the reversible reaction that the isolated pair is subject to. The diffusion approach of A and B need not be considered here. It is therefore crucial in an algorithm to generate samples from the stationary distribution we assumed.

At this point we can establish a connection with other treatments of diffusion influenced reaction rates. The formula derived by Doi [Doi75] describes the association rate constant for particles approaching each other via diffusion from the far-field. It includes the relative diffusion constant of the two particles D and reads

$$k_{\text{on,Doi}} = 4\pi D R \left(1 - \sqrt{\frac{D}{\lambda_{\text{on}} R^2}} \tanh\left(\sqrt{\frac{\lambda_{\text{on}} R^2}{D}} \right) \right)$$

Assuming the fast diffusion limit of this yields [EC09]

$$\lambda \ll \frac{D}{R^2} \quad \longrightarrow \quad k_{\text{on,Doi}} \approx \lambda_{\text{on}} \frac{4}{3} \pi R^3. \tag{2.36}$$

If we on the other hand assume the large volume limit of the expression from Eq. (2.35) we arrive at

$$R^3 \ll V \quad \longrightarrow \quad k_{\text{on}} = \tilde{\lambda}_{\text{on}} V_{\text{reac}}^{\text{eff}}. \tag{2.37}$$

Comparing Eqs. (2.36,2.37) we see that they match if the term $4\pi R^3/3$ is identified as the effective reaction volume without potentials.

C. Normalization of dissociation proposal density

Additionally we need $Q(\mathbf{y})$ from 2.23

$$\begin{aligned} Q(\mathbf{y}) &= \frac{\lambda_{\text{on}} V}{k_{\text{off}} Z_{AB}} \iiint \delta\left(\mathbf{y}_c - \frac{\mathbf{x}_a + \mathbf{x}_b}{2}\right) \chi_{\text{reac}}(\mathbf{x}) \dots \\ &\quad \times e^{-\beta U(|\mathbf{x}_b - \mathbf{x}_a|)} \mathrm{d}\mathbf{x}_a \mathrm{d}\mathbf{x}_b \mathrm{d}\mathbf{x}_c \\ &= \frac{\lambda_{\text{on}} V^2}{k_{\text{off}} Z_{AB}} \iint_{|\mathbf{x}_b - \mathbf{x}_a| \leqslant R} \delta\left(\mathbf{y}_c - \frac{\mathbf{x}_a + \mathbf{x}_b}{2}\right) \dots \\ &\quad \times e^{-\beta U(|\mathbf{x}_b - \mathbf{x}_a|)} \mathrm{d}\mathbf{x}_a \mathrm{d}\mathbf{x}_b \end{aligned}$$

The delta function can be reformulated in relative coordinates of A and B, that have to placed symmetric around $\mathbf{y}_c$. This eliminates another integral, which yields 1, due to the delta function. The only remaining degree of freedom is the distance of A and B, which results in an integral, that is identical to the effective reaction volume $V_{\text{reac}}^{\text{eff}}$ from Eq. (2.16).

References

[ACH16] Shailesh R. Agarwal, Colleen E. Clancy, and Robert D. Harvey. "Mechanisms Restricting Diffusion of Intracellular cAMP". In: *Scientific Reports* 6.1 (2016), p. 19577. DOI: 10.1038/srep19577.

[ADK18] Peter Atkins, Julio De Paula, and James Keeler. *Atkins' physical chemistry*. 8th ed. Oxford: Oxford University Press, 2018. ISBN: 0198700725.

[Alb+16] D. Albrecht et al. "Nanoscopic compartmentalization of membrane protein motion at the axon initial segment". In: *Journal of Cell Biology* 215.1 (2016), pp. 37–46. DOI: 10.1083/jcb.201603108.

[And17] Steven S. Andrews. "Smoldyn: Particle-based simulation with rule-based modeling, improved molecular interaction and a library interface". In: *Bioinformatics* 33.5 (2017), pp. 710–717. DOI: 10.1093/bioinformatics/btw700.

[AS90] Noam Agmon and Attila Szabo. "Theory of reversible diffusion-influenced reactions". In: *The Journal of Chemical Physics* 92.9 (1990), p. 5270. DOI: 10.1063/1.458533.

[AT08] M. Armand and J. M. Tarascon. "Building better batteries". In: *Nature* 451.7179 (2008), pp. 652–657. DOI: 10.1038/451652a.

[AT87] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. New York: Oxford University Press, 1987. DOI: 10.1093/oso/9780198803195.001.0001.

[Bha04] Upinder S. Bhalla. "Signaling in Small Subcellular Volumes. I. Stochastic and Diffusion Effects on Individual Pathways". In: *Biophysical Journal* 87.2 (2004), pp. 733–744. DOI: 10.1529/biophysj.104.040469.

[Bie+15] Johann Biedermann et al. "ReaDDyMM: Fast interacting particle reaction-diffusion simulations using graphical processing units." In: *Biophysical journal* 108.3 (2015), pp. 457–61. DOI: 10.1016/j.bpj.2014.12.025.

[Boy+06] S. Boyanov et al. "FeP: Another Attractive Anode for the Li-Ion Battery Enlisting a Reversible Two-Step Insertion/Conversion Process". In: *Chemistry of Materials* 18.15 (2006), pp. 3531–3538. DOI: 10.1021/cm060433m.

[CK49] Frank C. Collins and George E. Kimball. "Diffusion-controlled reaction rates". In: *Journal of Colloid Science* 4.4 (1949), pp. 425–437. DOI: 10.1016/0095-8522(49)90023-9.

[DB14] Weina Du and Peter G. Bolhuis. "Sampling the equilibrium kinetic network of Trp-cage in explicit solvent". In: *The Journal of Chemical Physics* 140.19 (2014), p. 195102. DOI: 10.1063/1.4874299.

[Dib+18] Manuel Dibak et al. "MSM/RD: Coupling Markov state models of molecular kinetics with reaction-diffusion simulations". In: *The Journal of Chemical Physics* 148.21 (2018), p. 214107. DOI: 10.1063/1.5020294.

[Dib+19] Manuel Dibak et al. "Diffusion-influenced reaction rates in the presence of pair interactions". In: *The Journal of Chemical Physics* 151.16 (2019), p. 164105. DOI: 10.1063/1.5124728.

[DM09] Gary J. Doherty and Harvey T. McMahon. "Mechanisms of Endocytosis". In: *Annual Review of Biochemistry* 78.1 (2009), pp. 857–902. DOI: 10.1146/annurev.biochem.78.081307.110540.

[Doi75] Masao Doi. "Theory of diffusion-controlled reactions between non-simple molecules. I". In: *Chemical Physics* ll (1975), pp. 107–113. DOI: 10.1016/0301-0104(75)80043-7.

[Doi76] Masao Doi. "Stochastic theory of diffusion-controlled reaction". In: *Journal of Physics A* 9.9 (1976), p. 1479. DOI: 10.1088/0305-4470/9/9/009.

[Don+10] Aleksandar Donev et al. "A First-Passage Kinetic Monte Carlo algorithm for complex diffusion-reaction systems". In: *Journal of Computational Physics* 229.9 (2010), pp. 3214–3236. DOI: 10.1016/j.jcp.2009.12.038.

[DYK18] Aleksandar Donev, Chiao-yu Yang, and Changho Kim. "Efficient reactive Brownian dynamics". In: *The Journal of Chemical Physics* 148.3 (2018), p. 034103. DOI: 10.1063/1.5009464.

[EC09] Radek Erban and S Jonathan Chapman. "Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions". In: *Physical Biology* 6.4 (2009), p. 046001. DOI: 10.1088/1478-3975/6/4/046001.

[Elo02] M. B. Elowitz. "Stochastic Gene Expression in a Single Cell". In: *Science* 297.5584 (2002), pp. 1183–1186. DOI: 10.1126/science.1070919.

[FN18] Christoph Fröhner and Frank Noé. "Reversible Interacting-Particle Reaction Dynamics". In: *The Journal of Physical Chemistry B* 122.49 (2018), pp. 11240–11250. DOI: 10.1021/acs.jpcb.8b06981.

[Gun+15] Monika Gunkel et al. "Higher-Order Architecture of Rhodopsin in Intact Photoreceptors and Its Implication for Phototransduction Kinetics". In: *Structure* 23.4 (2015), pp. 628–638. DOI: 10.1016/j.str.2015.01.015.

[Has70] W. K. Hastings. "Monte carlo sampling methods using Markov chains and their applications". In: *Biometrika* 57.1 (1970), pp. 97–109. DOI: 10.1093/biomet/57.1.97.

[HF13] Felix Höfling and Thomas Franosch. "Anomalous transport in the crowded world of biological cells". In: *arXiv* 76.4 (2013), p. 046602. DOI: 10.1088/0034-4885/76/4/046602.

[HFN19] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics". In: *PLoS Computational Biology* 15.2 (2019), e1006830. DOI: 10.1371/journal.pcbi.1006830.

[HN00] Alan Hall and Catherine D. Nobes. "Rho GTPases: molecular switches that control the organization and dynamics of the actin cytoskeleton". In: *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 355.1399 (2000). Ed. by L. Wolpert and J. C. Smith, pp. 965–970. DOI: 10.1098/rstb.2000.0632.

[HS14] Max Hoffmann and Ulrich S. Schwarz. "Oscillations of Min-proteins in micropatterned environments: a three-dimensional particle-based stochastic simulation approach". In: *Soft Matter* 10.14 (2014), pp. 2388–2396. DOI: 10.1039/c3sm52251b.

[JH14] Margaret E. Johnson and Gerhard Hummer. "Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems". In: *Physical Review X* 4.3 (2014), p. 031037. DOI: 10.1103/PhysRevX.4.031037.

[KS14] Heinrich C. R. Klein and Ulrich S. Schwarz. "Studying protein assembly with reversible Brownian dynamics of patchy particles". In: *The Journal of Chemical Physics* 140.18 (2014), p. 184112. DOI: 10.1063/1.4873708.

[Mar+17] Andrea L. Marat et al. "mTORC1 activity repression by late endosomal phosphatidylinositol 3,4-bisphosphate". In: *Science* 356.6341 (2017), pp. 968–972. DOI: 10.1126/science.aaf8310.

[Met+53] Nicholas Metropolis et al. "Equation of State Calculations by Fast Computing Machines". In: *The Journal of Chemical Physics* 21.6 (1953), pp. 1087–1092. DOI: 10.1063/1.1699114.

[MW08] Marco J. Morelli and Pieter Rein ten Wolde. "Reaction Brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network". In: *The Journal of Chemical Physics* 129.5 (2008), p. 054112. DOI: 10.1063/1.2958287.

[NB94] Claude Nuoffer and William E. Balch. "GTPases: multifunctional molecular switches regulating vesicular traffic." In: *Annual Review of Biochemistry* 63.1 (1994), pp. 949–990. DOI: 10.1146/annurev.bi.63.070194.004505.

[Opp+06] Tomas Opplestrup et al. "First-passage Monte Carlo algorithm: Diffusion without all the hops". In: *Physical Review Letters* 97.23 (2006), pp. 10–13. DOI: 10.1103/PhysRevLett.97.230602.

[Pau+17] Fabian Paul et al. "Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations". In: *Nature Communications* 8.1 (2017), p. 1095. DOI: 10.1038/s41467-017-01163-6.

[PB03] Thomas D. Pollard and Gary G. Borisy. "Cellular motility driven by assembly and disassembly of actin filaments". In: *Cell* 112.4 (2003), pp. 453–465. DOI: 10.1016/S0092-8674(03)00120-X.

[Pla+17] Nuria Plattner et al. "Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling". In: *Nature Chemistry* 9.10 (2017), pp. 1005–1011. DOI: 10.1038/nchem.2785.

[Pos+13] York Posor et al. "Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate". In: *Nature* 499.7457 (2013), pp. 233–237. DOI: 10.1038/nature12360.

[RQ16] Mauricio J. del Razo and Hong Qian. "A discrete stochastic formulation for reversible bimolecular reactions via diffusion encounter". In: *Communications in Mathematical Sciences* 14.6 (2016), pp. 1741–1772. DOI: 10.4310/CMS.2016.v14.n6.a13.

[Sad16] S. Kashif Sadiq. "Reaction–diffusion basis of retroviral infectivity". In: *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374.2080 (2016), p. 20160148. DOI: 10.1098/rsta.2016.0148.

[Sch+14] Johannes Schöneberg et al. "Explicit Spatiotemporal Simulation of Receptor-G Protein Coupling in Rod Cell Disk Membranes". In: *Biophysical Journal* 107.5 (2014), pp. 1042–1053. DOI: 10.1016/j.bpj.2014.05.050.

[Sch+17] Johannes Schöneberg et al. “Lipid-mediated PX-BAR domain recruitment couples local membrane constriction to endocytic vesicle fission”. In: *Nature Communications* 8.May (2017), p. 15873. DOI: 10.1038/ncomms15873.

[Sco+16] Duncan E. Scott et al. “Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge”. In: *Nature Reviews Drug Discovery* 15.8 (2016), pp. 533–550. DOI: 10.1038/nrd.2016.29.

[Smo16] Marian von Smoluchowski. “Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen”. In: *Zeitschrift fur Physikalische Chemie* 92U.1 (1916), p. 129. DOI: 10.1515/zpch-1918-9209.

[SN13] Johannes Schöneberg and Frank Noé. “ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments.” In: *PLoS One* 8.9 (2013), e74261. DOI: 10.1371/journal.pone.0074261.

[SN17] Luigi Sbailò and Frank Noé. “An efficient multi-scale Green’s function reaction dynamics scheme”. In: *The Journal of Chemical Physics* 147.18 (2017), p. 184106. DOI: 10.1063/1.5010190.

[SWN18] Mohsen Sadeghi, Thomas R. Weikl, and Frank Noé. “Particle-based membrane model for mesoscopic simulation of cellular dynamics”. In: *The Journal of Chemical Physics* 148.4 (2018), p. 044901. DOI: 10.1063/1.5009107.

[Tho02] Nancy L. Thompson. “Fluorescence Correlation Spectroscopy”. In: *Topics in Fluorescence Spectroscopy*. Springer, 2002, pp. 337–378. DOI: 10.1007/0-306-47057-8_6.

[TS67] Ei Teramoto and Nanako Shigesada. “Theory of bimolecular reaction processes in liquids”. In: *Progress of Theoretical Physics* 37.1 (1967), pp. 29–51. DOI: 10.1143/PTP.37.29.

[TTW10] Koichi Takahashi, Sorin Tanase-Nicola, and Pieter Rein ten Wolde. “Spatio-temporal correlations can drastically change the response of a MAPK pathway”. In: *Proceedings of the National Academy of Sciences* 107.6 (2010), pp. 2473–2478. DOI: 10.1073/pnas.0906885107.

[Ull+15] Alexander Ullrich et al. “Dynamical Organization of Syntaxin-1A at the Presynaptic Active Zone”. In: *PLOS Computational Biology* 11.9 (2015). Ed. by Kim T. Blackwell, e1004407. DOI: 10.1371/journal.pcbi.1004407.

[Van92] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*. 1st ed. Elsevier, 1992. ISBN: 9780444529657.

[VBW15] Adithya Vijaykumar, Peter G. Bolhuis, and Pieter Rein ten Wolde. “Combining molecular dynamics with mesoscopic Green’s function reaction dynamics simulations”. In: *The Journal of Chemical Physics* 143.21 (2015), p. 214102. DOI: 10.1063/1.4936254.

[Vij+17] Adithya Vijaykumar et al. “Multiscale simulations of anisotropic particles combining molecular dynamics and Green’s function reaction dynamics”. In: *The Journal of Chemical Physics* 146.11 (2017), p. 114106. DOI: 10.1063/1.4977515.

[ZW05a] Jeroen S. van Zon and Pieter Rein ten Wolde. “Green’s-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space”. In: *The Journal of Chemical Physics* 123.23 (2005), p. 234910. DOI: 10.1063/1.2137716.

[ZW05b] Jeroen S. van Zon and Pieter Rein ten Wolde. “Simulating biochemical networks at the particle level and in time and space: Green’s function reaction dynamics”. In: *Physical Review Letters* 94.12 (2005), p. 128103. DOI: 10.1103/PhysRevLett.94.128103.

Chapter 3

Diffusion-influenced reaction rates in the presence of pair interactions

The results of this chapter have been published in the following paper:

> Manuel Dibak, Christoph Fröhner (CF), Frank Noé and Felix Höfling. "Diffusion-influenced reaction rates in the presence of pair interactions". In: *The Journal of Chemical Physics* 151.16 (2019), p. 164105. DOI: 10.1063/1.5124728

Parts of the text and illustrations have been adopted unchanged in this document. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

Manuel Dibak and CF contributed equally to this work. In particular the contributions of the authors were as follows: Felix Höfling, Manuel Dibak and CF conceived the project and laid out the theory. Manuel Dibak implemented the semi-analytical solution. CF performed the iPRD simulations and analyzed the resulting data. Manuel Dibak and CF visualized the data. All contributors wrote the paper.

Summary

The kinetics of bimolecular reactions in solution depends, among other factors, on intermolecular forces such as steric repulsion or electrostatic interaction. Microscopically, a pair of molecules first has to meet by diffusion before the reaction can take place. In this work, we establish an extension of Doi's volume reaction model to molecules interacting via pair potentials, which is a key ingredient for interacting-particle-based reaction–diffusion (iPRD) simulations. As a central result, we relate model parameters and macroscopic reaction rate constants in this situation. We solve the corresponding reaction–diffusion equation in the steady state and derive semi-analytical expressions for the reaction rate constant and the local concentration profiles. Our results apply to the full spectrum from well-mixed to diffusion-limited kinetics. For limiting cases, we give explicit formulas, and we provide a computationally inexpensive numerical scheme for the general case, including the intermediate, diffusion-influenced regime. The obtained rate constants decompose uniquely into encounter and formation rates, and we discuss the effect of the potential on both subprocesses, exemplified for a soft harmonic repulsion and a Lennard-Jones potential. The analysis is complemented by extensive stochastic iPRD simulations, and we find excellent agreement with the theoretical predictions.

3.1 Introduction

A microscopic view on bimolecular chemical reactions in solution is essential for our understanding of many biological processes and technological applications; recent examples include, most prominently, protein functioning via complex formation [Sco+16; Pla+17], ligand binding [Hou10; Pau+17], and oligomerisation [BSS14; Sch+17], and on the other hand, catalysis in nanoreactors [Her+12; Gal+16] or ion deposition in batteries [Zho+17; AT08]. Such reactions are often strongly influenced by diffusion of at least one reactant, even more if transport occurs in a heterogeneous environment such as the interior of cells or on cellular membranes [MM06; ZRM08; HF13; Wei14].

In eukaryotes, the intracellular space is densely crowded by macromolecules, meandered by filamental networks, and compartmentalized by extended organelles, typically rendering diffusion at small scales anomalous [Eto+18; Wit+19; Ban+16; SW16; Kus+05; MJC16; Alb+16; Hor+10]. Different modelling strategies have been advised to account for such situations [SG18]: spatio-temporal master equations exploit metastability of diffusion between compartments [WS16], and crowding has been incorporated into the reaction–diffusion master equation on a mesoscale level [ELM18]. In particle-based Brownian dynamics simulations, crowding is implemented frequently as explicit excluded volume via hard or short-range repulsions [Rid+08; KY10; Dor+10; GYB10; TT14; EK15], which can give rise to complex-shaped structures on a cascade of scales [Höf+08; Spa+16; Sch+15; PF19].

Stochastic particle–based reaction diffusion simulations have become increasingly popular in the past decade [MW08; EC09; JH14; Sch+14; SUN14; VBW15; And17; ML16; AT17; SWN18; And18]. Such simulation methods and frameworks evolve the reaction–diffusion processes microscopically and have experienced advancements both in accuracy and computational performance [DYK18; FN18; Dib+18; SN17; SS19]. A recent development is interacting particle reaction dynamics (iPRD) [SN13; Bie+15; HFN19] that allows general interaction potentials on the reactive particles, for example, steric repulsion or electrostatic forces. Such interaction potentials may represent free energy landscapes computed from molecular dynamics (MD) simulations [BSF11; Xu+19; Wu+16].

A bimolecular reaction, $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{X}$, of two molecules A and B in solution occurs as a two-step process: encounter of the two reacting molecules by diffusion, followed by the formation of the product X, which abbreviates, for example, a complex C or the result $\mathrm{A}^* + \mathrm{B}$ of a catalytic reaction. Statistical independence of the durations of both steps suggests that the total reaction rate constant k is the harmonic mean [SS82; SSS80] of an encounter rate k_e and a formation rate k_f:

$$k^{-1} = k_\mathrm{e}^{-1} + k_\mathrm{f}^{-1}. \tag{3.1}$$

The formation rate depends on the detailed chemistry of the reaction process, often pictured as surmounting an activation barrier, whereas the encounter rate is determined by spatial diffusion of the molecules and subject to crowding conditions [KY10; Dor+10; GYB10; EK15], interaction potentials [Deb42], and confining geometries [GMO18]. A *diffusion-influenced* reaction refers to the not uncommon situation that both rates in Eq. (3.1) are of comparable magnitude and both steps are relevant for the overall kinetics [Bha04].

A commonly used reaction scheme in iPRD is Doi's volume reaction model [TS67; Doi75a; Doi75b; Doi76], where a reaction can occur with a microscopic rate λ if molecule centres are within a reaction radius R. Here, we extend this scheme by a pair interaction and relate the model parameters λ and R to the macroscopic reaction rate and its components for encounter and formation, see Eq. (3.1). Inversion of such a relation would allow the calibration of the microscopic model to match experimental rates. We obtain insights into the specific contributions of attractive and repulsive interactions to the reaction kinetics, and we highlight the importance of the local concentration of molecules in the reaction zone, which may differ drastically from the equilibrium distribution.

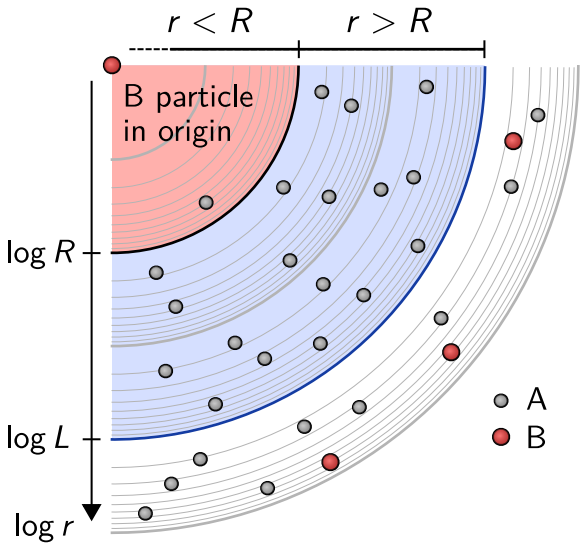

Figure 3.1: System of reactive molecules. Molecules of species A diffuse in space and can react with B molecules if their distance r is smaller than the reaction radius R. If B particles are scarce, a reasonable assumption is that there is no competition between them and one can treat only one of them within a spherical domain of radius $L \gg R$. For the analytical treatment, $L \longrightarrow \infty$, whereas for numerical methods and simulations L is finite. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

3.2 Microscopic model

Microscopic theories for bimolecular reactions date back to Smoluchowski [Smo16] in 1917, who proposed and analysed a model for coagulation of sphere-like molecules in solution that react instantaneously upon contact. Later, Debye [Deb42] amended the model by electrostatic interactions between the reactants, with notable repercussions on the binding rate. Collins and Kimball [CK49b; CK49a] refined Smoluchowski's model by introducing a finite rate at which molecules would react on contact. This model has been widely studied in the literature [SS82; SSS80; AS90; RQN18], however, the singular nature of the reaction surface has drawbacks in computer simulations as the exact time of encounter is not resolved in a time-stepping algorithm. An alternative scheme was suggested by Teramoto and Shigesada [TS67] and further characterized by Doi [Doi75a; Doi75b; Doi76], which permits the reaction of two molecules with a microscopic rate λ, referred to as *propensity* [Gil07], as long as the reactants are within a reaction radius R. This model is often referred to as the *volume reaction model* or *Doi model* and is in the focus of the present study.

Following Smoluchowski [Smo16], we consider a solution of substances A and B, that undergo the reaction

$$\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{A}^* + \mathrm{B}, \tag{3.2}$$

for which the product A^* of the reaction falls out of scope, such that we do not need to consider it. The concentrations c_A and c_B of A and B molecules, respectively, are assumed to be both so dilute that interactions between like molecules can safely be ignored. (Otherwise, the reaction kinetics would non-trivially depend on c_A and c_B and the reaction rate would not be a well-defined constant.) Further, the concentration of B molecules is assumed to be much smaller than that of A, $c_B \ll c_A$, i.e., A molecules are

abundant relative to Bs and there is no competition for reactants between the B molecules. Equivalently, substance B is highly diluted, and the problem can be rephrased as that of a single B molecule surrounded by A molecules in a large, yet finite volume V. It is convenient to switch to the reference frame of the B molecule, and we will choose a spherical volume V of radius L; see Fig. 3.1 for an illustration. In a finite amount of time and for sufficiently large V, the B molecule absorbs only a negligible fraction of As so that we can assume a quasi-steady state with the concentration c_A being constant at the boundary ∂V of the volume.

As microscopic reaction model, we use the Teramoto–Shigesada–Doi model [TS67; Doi75a; Doi75b; Doi76], in which A and B molecules diffuse in space with diffusion constants D_A and D_B, respectively, forming a reactive complex whenever an A is separated from a B by less than the reaction distance R. This reactive complex undergoes reaction (3.2) with a microscopic rate constant or *propensity* λ, thus effectively removing A molecules from the system with a frequency K. More precisely, given a reactive complex, reaction events are triggered by a Poisson clock with parameter λ. The throughput or velocity of reaction (3.2) is then given by

$$\frac{\mathrm{d}c_{A^*}}{\mathrm{d}t} = K c_B \,, \tag{3.3}$$

where c_{A^*} is the overall concentration of the reaction product A^*.

Similarly to Debye's work [Deb42], and as commonly done in iPRD simulations [SN13], our focus here is on situations where A and B molecules interact physically with each other according to an isotropic pair potential $U(\mathbf{r}) = U(|\mathbf{r}|)$; the vector $\mathbf{r}$ denotes the separation of an AB pair. The average concentration field $p(\mathbf{r}, t)$ of A molecules and the corresponding flux (density) $\mathbf{j}(\mathbf{r}, t)$ are then governed by the reaction–diffusion equation

$$\partial_t p(\mathbf{r}, t) = -\nabla \cdot \mathbf{j}(\mathbf{r}, t) - a(\mathbf{r})\, p(\mathbf{r}, t) \,, \tag{3.4a}$$

$$\mathbf{j}(\mathbf{r}, t) := -D\, \mathrm{e}^{-\beta U(\mathbf{r})} \nabla \left[\mathrm{e}^{\beta U(\mathbf{r})} p(\mathbf{r}, t) \right] , \tag{3.4b}$$

with the reaction propensity $a(\mathbf{r}) \geqslant 0$ and $D = D_A + D_B$ the relative diffusion constant of the particles; $\beta = 1/k_\mathrm{B} T$ denotes the inverse of the thermal energy scale as usual. Within the Doi model, the propensity $a(\mathbf{r})$ is implemented in terms of the Heaviside step function, $a(\mathbf{r}) = \lambda\, \theta(R - |\mathbf{r}|)$ such that the B molecule appears as a spherical reactive sink of radius R.

By isotropy of the setup, the steady flux $\mathbf{j}(\mathbf{r})$ of A molecules has only a radial component $j(r)$ that is a function only of the distance $r = |\mathbf{r}|$ to the B molecule. It determines the reaction frequency K through the surface integral

$$K = -\int_{|\mathbf{r}|=R} \mathbf{j}(\mathbf{r}) \cdot \mathbf{n}\, \mathrm{d}\sigma = -4\pi R^2 j(R) \,, \tag{3.5}$$

with the surface normal $\mathbf{n}$ pointing outwards; the minus sign arises due to the fact that particles flow from the boundary to the sink at the origin, $j(r) < 0$. On the other hand, the law of mass action yields the reaction rate equation

$$\frac{\mathrm{d}c_{A^*}}{\mathrm{d}t} = k c_A c_B \,, \tag{3.6}$$

in terms of the macroscopic association rate constant k. Comparing to Eq. (3.3), the latter is related to the microscopic frequency K by $k = K/c_A$, and the reaction rate constant follows as

$$k = \frac{4\pi R^2 |j(R)|}{c_A} \,. \tag{3.7}$$

The goal of the following sections is to calculate the flux profile $j(r)$ of the quasi-steady state and thus the macroscopic rate k, focussing on their dependences on the microscopic reaction parameters, λ and R, and on the pair potential $U(r)$ between A and B molecules. Note that there is no interaction amongst A molecules.

3.3 Solution strategy and classical limiting cases

In this section, we work out the general solution strategy for the reaction–diffusion equations, Eq. (3.4), and obtain analytical solution to important subproblems, which resemble a number of classical results. The stationary solutions $p(\mathbf{r})$ obeys $\partial_t p(\mathbf{r}) = 0$, and thus Eq. (3.4a) reduces to

$$\nabla \cdot \mathbf{j}(\mathbf{r}) = -a(\mathbf{r})\, p(\mathbf{r}). \tag{3.8}$$

According to the quasi-steady state assumption, $p(\mathbf{r})$ further satisfies the Dirichlet boundary condition

$$p(\mathbf{r}) = c_A\,, \quad \mathbf{r} \in \partial V\,. \tag{3.9}$$

Restricting to isotropic potentials, we switch to a single radial coordinate, $r = |\mathbf{r}|$, with the convention that the flux $j(r) = \mathbf{j}(\mathbf{r}) \cdot \mathbf{r}/r$ points outwards:

$$\frac{1}{r^2}\,\partial_r r^2 j(r) = -\lambda\,\theta(R - r)\,p(r) \tag{3.10}$$

with

$$j(r) = -D\mathrm{e}^{-\beta U(r)}\partial_r\left[\mathrm{e}^{\beta U(r)}p(r)\right]. \tag{3.11}$$

In this case and for an infinitely large volume V, Eq. (3.9) simplifies to $p(r \longrightarrow \infty) = c_A$.

To complete the boundary value problem for $p(r)$, we need to specify also the behaviour at the coordinate origin, which is not obvious due to the interaction potential. The total flux through a ball B_ε of radius ε centred at $\mathbf{r} = 0$ obeys:

$$\int_{\partial B_\varepsilon} \mathbf{j}(\mathbf{r}) \cdot \mathbf{n}\,\mathrm{d}\sigma = -\int_{B_\varepsilon} a(\mathbf{r})\,p(\mathbf{r})\,\mathrm{d}^3 r\,, \tag{3.12}$$

invoking Gauss' theorem and inserting Eq. (3.8). Continuity of the solution $p(\mathbf{r})$ together with our choice for $a(\mathbf{r})$ yields $4\pi\varepsilon^2\, j(\varepsilon) \simeq -\lambda p(0) \cdot 4\pi\varepsilon^3/3$, and thus

$$j(0) = 0\,. \tag{3.13}$$

It implies a Robin boundary condition for the concentration profile,

$$\lim_{r\longrightarrow 0}\left[\beta U'(r)\,p(r) + \partial_r p(r)\right] = 0, \tag{3.14}$$

which is satisfied by a Boltzmann distribution (scaled by a constant factor):

$$p(r) \sim \exp(-\beta U(r))\,, \quad r \longrightarrow 0, \tag{3.15}$$

capturing the r-dependence asymptotically.

Note that the preceding derivation does not apply for potentials $U(r)$ that diverge as $r \longrightarrow 0$. In this case, the current $\mathbf{j}(\mathbf{r})$ is not defined at the origin, $\mathbf{r} = 0$, and, strictly speaking, this point must be excluded from the integration domain B_ε, which forbids the application of Gauss' theorem. Yet, the extension of Eq. (3.15) to diverging potentials, $U(r \longrightarrow 0) = +\infty$, is motivated physically as it is improbable that any A molecule reaches the centre of the reaction volume: an upper bound on $p(r)$ is given by the equilibrium distribution, describing the non-reacting case. In particular, $p(\mathbf{r})$ is continuous in $\mathbf{r} = 0$ and so is $\nabla \cdot \mathbf{j}(r)$ by Eq. (3.8), justifying the use of Gauss' theorem *a posteriori*.

Eventually, the step-like reaction propensity in Eq. (3.10) suggests to split the domain at the reaction boundary, $r = R$, and to find separate solutions $p_{\gtrless}$ and $j_{\gtrless}$ in both subdomains, $r \gtrless R$. By inspection of

the r.h.s. of Eqs. (3.10) and (3.11), the flux $j(r)$ is finite and continuous at this interface, which implies that $p(r)$ is continuously differentiable at $r = R$. This provides us with the interface conditions

$$p_>(R) = p_<(R)\,, \tag{3.16}$$

$$j_>(R) = j_<(R) = -K/4\pi R^2, \tag{3.17}$$

making use of Eq. (3.5) in the last step. Matching the solutions of both subdomains will thus yield the sought-after reaction frequency K.

3.3.1 Outer solution

In the outer domain (>), where $R \leqslant r < \infty$, Eq. (3.10) reduces to an equation for the flux alone, $\partial_r r^2 j_>(r) = 0$. Integration from the lower boundary, Eq. (3.17), to some $r > R$ yields:

$$j_>(r) = -\frac{K}{4\pi r^2}\,, \tag{3.18}$$

with unknown rate K. The functional dependence on r is readily understood by the fact that, in the absence of reactions, the integral flux through spheres of radius r is constant (Gauss' theorem). In particular, the solution is compatible with the no-flux condition, $j_>(r \longrightarrow \infty) = 0$, which is implied by the upper boundary, $p_>(r \longrightarrow \infty) = c_A$, together with the vanishing force, $-\nabla U(r \longrightarrow \infty) = 0$, and using Eq. (3.11).

Next, we calculate the concentration profile $p_>(r)$ from Eqs. (3.9) and (3.11). Introducing

$$g(r) := \mathrm{e}^{\beta U(r)} r^{-2} \tag{3.19}$$

for brevity, one finds $(K/4\pi D)\, g(r) = \partial_r \left[\mathrm{e}^{\beta U(r)} p_>(r)\right]$, and after integration over $[r, \infty)$:

$$p_>(r) = \mathrm{e}^{-\beta U(r)} \left[c_A - \frac{K}{4\pi D} \int_r^\infty g(s)\mathrm{d}s\right]\,, \tag{3.20}$$

which is Debye's classical result [Deb42]. If the interaction potential is not present ($U = 0$), this reduces to the familiar solution of the Dirichlet–Laplace problem:

$$p_>(r) = c_A - \frac{K}{4\pi D}\frac{1}{r}\,. \tag{3.21}$$

For diffusion-limited reactions, that is when product formation is fast and $k_\mathrm{f} \ll k_\mathrm{e}$ in Eq. (3.1), particles almost surely react on the surface of the reaction volume and the concentration inside vanishes: $p_<(r) = 0$ for $r \leqslant R$. Then by continuity of $p(r)$ at the interface of the subdomains, Eq. (3.20) is amended by $p_>(R) = 0$ and can be solved for K. This yields the Debye reaction rate constant $k = K/c_A$, which we identify as the encounter rate k_e in the presence of a pair potential:

$$k_\mathrm{e} = 4\pi D \Big/ \int_R^\infty g(s)\mathrm{d}s\,. \tag{3.22}$$

The corresponding concentration profile is given by Eq. (3.20) and reads

$$p_>(r) = c_A \mathrm{e}^{-\beta U(r)} \int_R^r g(s)\,\mathrm{d}s \Big/ \int_R^\infty g(s)\,\mathrm{d}s. \tag{3.23}$$

In particular, $p_>(r)$ is independent of the diffusion constant D. For $U(r) = 0$, these results recover Smoluchowski's rate constant [Smo17] $k = 4\pi DR$ and the profile $p_>(r) = c_A(1 - R/r)$.

3.3.2 Inner solution without potential

In the absence of an interaction potential, Eqs. (3.10) and (3.11) simplify drastically and the concentration inside $p_<(r)$ the reaction volume, $0 \leqslant r \leqslant R$, obeys the Helmholtz equation

$$\left(\partial_r^2 + \frac{2}{r}\partial_r - \kappa^2\right) p_<(r) = 0 \tag{3.24}$$

with the inverse length $\kappa := \sqrt{\lambda/D}$, describing the penetration depth into the reactive domain. The flux takes the form $j_<(r) = -D\partial_r p_<(r)$, which turns the boundary conditions for the flux, Eqs. (3.13) and (3.17), into von Neumann conditions for the concentration, $p_<'(0) = 0$ and $p_<'(R) = K/4\pi DR^2$. Equation (3.24) is equivalent to $\left(\partial_r^2 - \kappa^2\right)[rp_<(r)] = 0$, and the boundary value problem is solved by [EC09]

$$p_<(r) = \gamma \frac{\sinh(\kappa r)}{\kappa r} \tag{3.25}$$

with the constant γ fixed by the upper boundary; in particular, γ is proportional to the reaction frequency K. Matching inner and outer solutions for $p(r)$, Eqs. (3.21) and (3.25), at the interface, $r = R$, leads to $\gamma = c_A/\cosh(\kappa R)$, and Doi's result for the reaction rate constant [Doi75a; EC09] follows:

$$k = 4\pi DR\left[1 - \frac{\tanh(\kappa R)}{\kappa R}\right]. \tag{3.26}$$

The solution naturally decomposes as in Eq. (3.1) into Smoluchowski's encounter rate $k_e = 4\pi DR$, see Eq. (3.22), and a formation rate

$$k_f = 4\pi DR[\kappa R \coth(\kappa R) - 1], \tag{3.27}$$

with $\coth(x) = 1/\tanh(x)$. In the fast-diffusion limit, $\kappa R \ll 1$, i.e., when the reaction propensity λ is low, the formation rate $k_f \simeq (4\pi/3)R^3\lambda$ is simply the product of the reaction volume $V_R = (4\pi/3)R^3$ and the propensity, reflecting well-mixed conditions inside the reaction volume ($p_<(r) = const$). For fast reactions, $\kappa R \gg 1$, we obtain $k_f \simeq 4\pi R^2\kappa^{-1}\lambda$, which we interpret as reactions being restricted to a volume $4\pi R^2\kappa^{-1}$, that is a thin shell of radius R and width κ^{-1}.

3.4 Reaction rates and spatial distributions in the presence of an interaction potential

For the general solution to the reaction–diffusion problem, Eqs. (3.10) and (3.11), in the presence of an interaction potential, it remains to find a solution inside the reaction radius (inner domain) and to match it with Eq. (3.20). As boundary condition we use $j_<(0) = 0$, Eq. (3.13), and solve for the current $j_<(r)$ first.

3.4.1 Constant potential inside the reaction volume

As a preliminary to the general discussion, we consider the analytically accessible situation that the interaction potential is constant within the reaction volume, i.e., $U(r) = U(R)$ for $r \leqslant R$. This may be useful in modelling reactions in electrolytes while neglecting excluded volume effects. Then the inner solution equals the non-interacting case, Eq. (3.25), and can be matched with Eq. (3.20) to find the reaction rate constant

$$k = 4\pi D\left(\frac{R\,g(R)}{\kappa R\coth(\kappa R) - 1} + \int_R^\infty g(r)\,\mathrm{d}r\right)^{-1}. \tag{3.28}$$

In particular, the encounter rate k_e is equal to Debye's result, Eq. (3.22), whereas the formation rate is suppressed by a factor $R^2 g(R) = \mathrm{e}^{\beta U(R)}$ relative to the non-interacting value, Eq. (3.27), and the total rate is the harmonic mean of both, Eq. (3.1).

3.4.2 Solution for arbitrary potentials

We proceed along the lines of the potential-free case, Section 3.3.2, and solve Eqs. (3.10) and (3.11) inside the reaction volume, $0 \leqslant r \leqslant R$, subject to the boundary conditions Eqs. (3.13) and (3.17). Applying the differential operator $\mathrm{e}^{-\beta U(r)}\partial_r \mathrm{e}^{\beta U(r)}$ on both sides of Eq. (3.10) and identifying the flux on the right hand side, one finds the following Dirichlet problem for the dimensionless function $\psi(r) := -4\pi r^2 j_<(r)/K$:

$$\psi''(r) + \left(\beta U'(r) - \frac{2}{r}\right)\psi'(r) - \kappa^2\psi(r) = 0\,, \tag{3.29a}$$

$$\psi(0) = 0\,, \quad \text{and} \quad \psi(R) = 1\,. \tag{3.29b}$$

In the absence of an explicit solution, we use the method of finite differences[Smi85] to compute, in particular, the derivative on the reaction boundary, $\psi'(R)$. The latter determines the concentration on the boundary via Eq. (3.10):

$$p_<(R) = \psi'(R)K/4\pi R^2\lambda\,. \tag{3.30}$$

Eventually, the reaction frequency K is obtained by matching inner and outer solutions for the concentration, Eq. (3.16). Employing the numerical value for $\psi'(R)$ and our previous result, Eq. (3.20), we have

$$\frac{K}{4\pi R^2\lambda}\,\psi'(R) = \mathrm{e}^{-\beta U(R)}\left[c_A - \frac{K}{4\pi D}\int_R^\infty g(s)\,\mathrm{d}s\right]. \tag{3.31}$$

Solving for $K = k/c_A$, yields an exact, closed expression for the macroscopic rate constant k, which is one of our main results:

$$k = 4\pi D\left[\int_R^\infty g(s)\mathrm{d}s + \frac{g(R)\,\psi'(R)}{\kappa^2}\right]^{-1}; \tag{3.32}$$

the pair potential enters through the function $g(r) := \mathrm{e}^{\beta U(r)}r^{-2}$. The result naturally displays the decomposition of Eq. (3.1), and we identify the formation rate as

$$k_\mathrm{f} = \frac{4\pi\lambda}{g(R)\,\psi'(R)}\,, \tag{3.33}$$

which appears to be proportional to the reaction propensity λ; in fact, the value of $\psi'(R)$, as given by Eqs. (3.29), indirectly depends on λ as well. Noteworthy, the diffusion-limited encounter rate k_e is the same as for the Debye problem, see Eq. (3.22), and the classical result, $k = k_\mathrm{e}$, is recovered in the limit of instantaneous reactions, $\lambda \longrightarrow \infty$, i.e., for vanishing k_f^{-1}.

An alternative expression for the formation rate k_f in terms of the concentration $p(R)$ is obtained by substituting $\psi'(R)$ using Eq. (3.30) and $K = kc_A$, which yields $k_\mathrm{f} = kc_A\mathrm{e}^{-\beta U(R)}/p(R)$. Employing the decomposition of the total rate k [Eq. (3.1)] and solving for k_f, one finds

$$k_\mathrm{f} = k_\mathrm{e}\left[\frac{c_A\mathrm{e}^{-\beta U(R)}}{p(R)} - 1\right]. \tag{3.34}$$

Interestingly, the formation rate is fully specified by the encounter rate k_e and the concentration at the reaction boundary relative to its equilibrium value. However, the computation of $p(R)$ requires the full solution of the reaction–diffusion problem.

The concentration profile $p(r)$ follows from integration of Eq. (3.11) in terms of $\psi(r)$ and using continuity, Eq. (3.16), to eliminate $p_<(R)$ to find

$$p(r) = c_A\mathrm{e}^{-\beta U(r)}\left[1 - \frac{k}{4\pi D}\int_r^\infty g(s)\psi(s)\mathrm{d}s\right], \tag{3.35}$$

with the convention $\psi(r) = 1$ for $r > R$. Alternatively the density profile can also be found by Eq. (3.10), from the solution $\psi(r)$ as $p_<(r) = \psi'(r)K/4\pi r^2\lambda$. However, we observed the numerical integration in Eq. (3.35) to yield smaller errors.

3.4.3 Perturbative solution for slow reactions

Slow reactions, $\lambda \ll DR^2$, corresponding to a well-mixed reaction volume, are described by a large penetration depth $\kappa^{-1} \gg R$. This suggests to expand the concentration profile $p_<(r)$ in the small parameter $\kappa R \ll 1$, introducing functions $p_0, p_1, \ldots$:

$$p_<(r) = p_0(r) + (\kappa R)^2 p_1(r) + O\big((\kappa R)^4\big)\,; \tag{3.36}$$

here, we neglect terms of order $(\kappa R)^4$. Corresponding fluxes $j_0(r), j_1(r), \ldots$ are defined by virtue of Eq. (3.11). Inserting the expansion into Eq. (3.10) for $r \leqslant R$ and sorting by powers of $\kappa^2 = \lambda/D$, one finds that the 0^{th} order is satisfied by the equilibrium distribution in the absence of reactions:

$$p_0(r) = c_A \mathrm{e}^{-\beta U(r)}\,, \tag{3.37}$$

which is accompanied by a vanishing flux, $j_0(r) \equiv 0$, due to detailed balance. The flux $j_1(r)$ at order $(\kappa R)^2$ obeys

$$\frac{1}{r^2}\,\partial_r r^2 j_1(r) = -\kappa^2 D\, p_0(r)\,, \tag{3.38}$$

which can be integrated to yield

$$j_1(r) = -\frac{\kappa^2 D c_A}{r^2}\int_0^r \mathrm{e}^{-\beta U(s)} s^2\, ds \tag{3.39}$$

for $0 \leqslant r \leqslant R$, where we used the boundary condition $j(0) = 0$ [Eq. (3.13)]. With this, the reaction rate constant k follows from Eq. (3.7) straightforwardly:

$$k = \kappa^2 D \int_0^R \mathrm{e}^{-\beta U(r)}\, 4\pi r^2 \mathrm{d}r + O\big((\kappa R)^4\big)\,. \tag{3.40}$$

It allows for a simple interpretation valid for slow reactions: the macroscopic rate $k \simeq \lambda V_{\text{eff}}$ is the product of the reaction propensity λ and an effectively accessible reaction volume [FN18],

$$V_{\text{eff}} = \int_{|\mathbf{r}|\leqslant R} \mathrm{e}^{-\beta U(\mathbf{r})}\, \mathrm{d}^3 r\,. \tag{3.41}$$

3.4.4 Numerical details

The computation of the reaction rate [Eq. (3.32)] for arbitrary potentials and reaction parameters requires the numerical solution of the boundary-value problem, Eq. (3.29), and of the integral, Eq. (3.22). We checked our numerical implementation by comparing to the analytically exactly tractable, albeit peculiar case of a logarithmic potential,

$$U(r) = \begin{cases} -2k_B T \log(r/R), & r < R \\ 0, & \text{otherwise.} \end{cases} \tag{3.42}$$

With this, $g(r) = R^{-2}\,\theta(R - r)$ is a step function, and the coefficient $\beta U'(r) - 2/r$ in Eq. (3.29a) reduces to $-4/r$. The differential equation can be solved using computer algebra, yielding $\psi'(R)$ and the reaction rate according to Eq. (3.32) as

$$k = 2\pi D R \left\{3 - \frac{(\kappa R)^2}{(\kappa R)^2 - 2[\kappa R \coth(\kappa R) - 1]}\right\}. \tag{3.43}$$

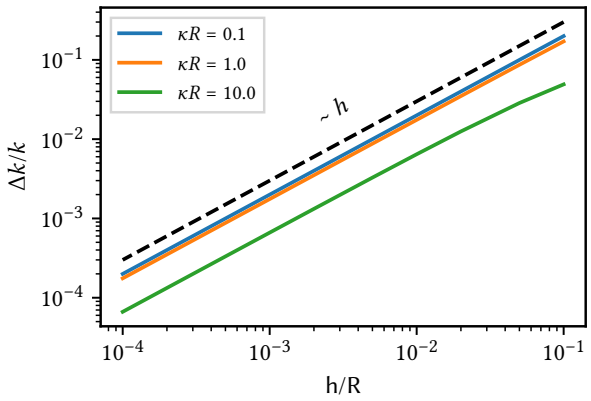

Figure 3.2: Relative error $\Delta k/k$ of the reaction rate constant k of the numerical solution [Eq. (3.32)] with respect to the analytical solution [Eq. (3.43)] for a diverging potential [Eq. (3.42)]. The numerical result is obtained for different discretisation widths h given in units of the reaction radius R and for different reactivities κR. The dashed line depicts a linear scaling, $\Delta k/k \sim h$. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

The Debye rate was computed via the adaptive quadrature routines from QUADPACK. For numerical solutions to Eq. (3.29), we used the method of finite differences [Smi85] by discretising the domain $[0, R]$ into N sub-intervals of equal size $h := R/N$. Let us note that at the outer most grid points, $r = 0$ and $r = R$, Eq. (3.29a) does not require evaluation if central differences are used to compute $\psi'(r)$ and $\psi''(r)$ from $\psi(r)$. For a range of values of κR, we computed the error Δk between the numerical and the analytical results for the rate, see Fig. 3.2. The relative error $\Delta k/k$ scales approximately linearly with h and decreases with increasing κR. For the worst case studied, $\kappa R = 0.1$, we conclude that an accuracy better than 10^{-3} is reached by choosing a grid spacing of $h = 10^{-4}R$, which is still well feasible in terms of computational costs. This value of h is used for all subsequent calculations.

Finally, we have checked that all terms in Eq. (3.29a) are bounded. In particular, we argue that the term $[\beta U'(r) - 2/r]\psi'(r)$ vanishes in the limit $r \longrightarrow 0$. The expression is proportional to $[\beta U'(r) - 2/r]r^2 p(r)$ after re-substituting $\psi(r)$ and using Eq. (3.10). Further, we anticipate that the concentration profile is bounded from above by the equilibrium distribution, $p(r) \leqslant c_A \mathrm{e}^{-\beta U(r)}$, as reactions can only lower the concentration in the reaction volume, see Fig. 3.8. With this, $(2/r)\, r^2 p(r) \longrightarrow 0$ and $|\beta U'(r)\, p(r)| \leqslant c_A \left|\partial_r \mathrm{e}^{-\beta U(r)}\right|$, and it remains to show that $\left|\partial_r \mathrm{e}^{-\beta U(r)}\right| \xrightarrow{r \to 0} 0$. This is fulfilled by certain logarithmic potentials, such as in Eq. (3.42), and by algebraically diverging potentials, $\beta U(r \longrightarrow 0) \simeq a r^{-m}$ with $a, m > 0$. In the latter case, putting $y := r^{-m}$ we have $\left|\partial_r \mathrm{e}^{-\beta U(r)}\right| \simeq a m\, y^{(m+1)/m} \mathrm{e}^{-a y} \longrightarrow 0$ as $y \longrightarrow \infty$.

3.5 iPRD simulations

Complementary to the preceding theoretical analysis, we have performed extensive simulations of the microscopic reaction–diffusion dynamics in the steady state. We "measure" the absolute reaction rate k of the reaction (3.2) and the radial distribution function $p(r)$ of A molecules relative to a B molecule.

Quantity	Symbol	Value	Unit
Propensity of reaction (3.2)	λ	varies	τ_d^{-1}
Soft repulsion strength	b	40	k_BT/R^2
Soft repulsion range	r_0	1	R
LJ interaction strength	ε	1	k_BT
LJ interaction range	σ	$(26/7)^{-1/6}$	R
LJ cutoff radius	r_c	2.5	R
Integration time step	Δt	10^{-4}	τ_d
Radius of simulation domain	L	10	R
Width of factory shell	ΔL	5	R
Number of factory particles	N_f	1.5×10^4	1
Propensity to create A	f_+	0.01	τ_d^{-1}
Propensity to absorb A	f_-	0.01	τ_d^{-1}

Table 3.1: Parameters used in the particle simulations. Basic units of length, time, and energy are R, $\tau_\mathrm{d} := R^2/D$, and k_BT, respectively. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

3.5.1 Simulation setup and protocol

Stochastic simulations of the interacting particle-based reaction–diffusion dynamics (iPRD) are performed with the software ReaDDy 2 [HFN19; SN13], which integrates the motion of particles and reactions between them explicitly in three-dimensional space. In ReaDDy, time is discretised into steps of fixed size Δt. A single step consists of first integrating the Brownian motion of molecules via the Euler–Maruyama scheme and then handling reaction events according to the Doi model (Section 3.2). After each step, one can evaluate observables, such as the positions of particles or the number of reactions that occurred.

The simulation setup is constructed spherically symmetric around a single B molecule in the coordinate origin, as depicted in Fig. 3.1. In particular, we use a spherical domain of finite radius L, which will be filled with A molecules such that at the boundary, $r = L$, the concentration $p(L)$ of A molecules matches a given constant. Within the whole domain, A particles diffuse subject to the interaction potential $U(r)$, whereas the B molecule is fixed in space; here, we restrict ourselves to potentials that are cut off at a distance $r_c < L$. The conversion reaction (3.2) takes place with reaction propensity λ inside the sphere with $r \leqslant R$. We have run a large number of simulations for varying propensity λ and different potentials $U(r)$, see below. Simulation units were chosen such that distances are measured in terms of the reaction radius R, energies in terms of the thermal energy k_BT, and times in terms of the combination $\tau_\mathrm{d} := R^2/D$, which is proportional to the time to explore the reaction volume by diffusion. The parameters used are listed in Table 3.1; in particular, a time step $\Delta t = 10^{-4}\tau_\mathrm{d}$ was used throughout production runs. The chosen time step is sufficiently small to be suitable for the Lennard-Jones potential, which generally calls for much smaller integration steps than the harmonic repulsion due to an increased stiffness. For a given set of parameters and a desired accuracy goal, an optimal time step can be found by systematic comparison of a sequence of simulations with our analytical results.

Aiming at the simulation of a stationary reaction kinetics, we coat the domain by a *factory shell*, with radial coordinates in $r \in [L, L + \Delta L]$, that yields a constant supply of A molecules. Adjacent to the shell, for $r \geqslant L + \Delta L$, an external harmonic potential is added that prevents A molecules from escaping and thereby closing the simulation domain. The factory shell contains N_f factory (F) particles, which are fixed in space at random positions according to a uniform distribution. F particles create and absorb A

molecules through the reversible reaction

$$\mathrm{F} \underset{f_-}{\overset{f_+}{\rightleftharpoons}} \mathrm{F} + \mathrm{A}\,. \tag{3.44}$$

The forward reaction has propensity f_+ and is of fission type: a new A molecule is placed at a random distance $d \in [0, R_f]$ from the active F particle. The backward reaction is of fusion type, by which an A molecule is absorbed with propensity f_- if it is closer than R_f to an F particle. Due to the fact that the number of F particles is conserved, the factory reactions (3.44) are pseudo-unimolecular, i.e. they can be reduced to

$$\mathrm{A} \rightleftharpoons \varnothing, \tag{3.45}$$

which leads to a steady-state concentration $p(L)$ of As. The latter depends also on the outflux $K = 4\pi L^2 |j(L)|$ of A molecules, which can diffuse freely into and out of this shell and migrate towards the origin due to the reaction of interest, Eq. (3.2). Lacking an *a priori* knowledge of the concentration $p(L)$ and the concentration c_A in the far field ($r \to \infty$), we run simulations with a certain set of parameters N_f, f_+, f_-, and R_f and estimate the resulting value of c_A accurately from the observed steady-state profile $p(r)$. Specifically, we fit the solution $p(r) = c_A - K/4\pi D r$ [Eq. (3.21)], to the data for $p(r)$ in the range $\max(R, r_c) \leqslant r \leqslant L$, where both interactions and reactions are absent and A molecules diffuse freely. This yields the extrapolated concentration at far distances, $p(r \to \infty) = c_A$. Note that the reaction frequency K is directly available from the simulation by counting reaction events.

The above procedure relies on the fact that shifting the upper boundary from infinity to $r = L$ merely shifts the concentration $p(r)$ by an additive constant, leaving the integral flux through spheres of radius r unchanged, provided that r is outside of the interaction range. This is a consequence of Gauss's theorem, see also Eq. (3.11). Therefore, simulation results with a finite volume can be mapped exactly to the infinite case upon using the effective far-field concentration c_A as determined above.

A data production cycle starts with uniformly distributing A molecules in the factory shell with a concentration that roughly anticipates the expected c_A. This initial state is relaxed by evolving the reaction–diffusion dynamics for a time span of $t_{\mathrm{eq}} = 300\tau_{\mathrm{d}}$, by executing 3×10^5 integration steps with a coarser time step size of $\Delta t = 10^{-3}\tau_{\mathrm{d}}$. Equilibration is verified by observing that the number of A particles does not vary significantly. The time step is then decreased to $\Delta t = 10^{-4}\tau_{\mathrm{d}}$ and the system equilibrated for another time span of $30\tau_{\mathrm{d}}$. During the subsequent production run of length similar to t_{eq}, we record the two main observables:

1. the concentration profile $p(r)$ as the radial distribution function (RDF) of A molecules relative to the B molecule in the centre, and

2. the number of reactions (3.2) that were performed in each integration step, yielding the reaction frequency K and thus the macroscopic reaction rate constant $k = K/c_A$.

Observing the RDF in the case without a reaction and comparing it against the Boltzmann distribution is used to verify the time step.

One such simulation procedure took roughly 512 hours on a single CPU. Simulations were run for 3 different potentials and 5 different propensities, for each combination statistical averages over 13 independent realisations were taken, altogether yielding 195 simulations that were run in parallel. The cumulative CPU time amounts to 100,000 hours.

3.5.2 Pair potentials

In the following, we consider two different isotropic pair potentials for the interaction between A and B molecules, and we compare to the non-interacting case ($U = 0$). The employed potentials are visualized

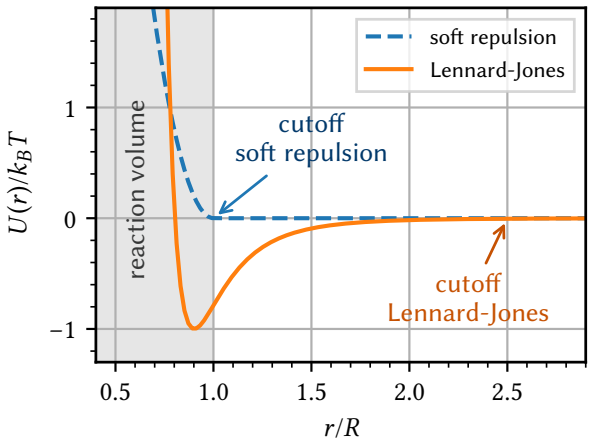

Figure 3.3: Pair potentials $U(r)$ used in our study of the steady-state reaction kinetics [Eqs. (3.47) and (4.6)] for the parameters given in Table 3.1. The separation r of molecule centres is given in units of the reaction radius R, and the potential energy U is given in terms of the thermal energy k_BT; the shaded region marks the reaction sphere in which reaction (3.2) can occur. Arrows indicate the location of the interaction cutoffs. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

in Fig. 3.3, and all relevant parameters are given in Table 3.1. The first potential describes an ultra-soft steric repulsion, which is common for macromolecules such as polymer rings [Poi+15]. For simplicity, we assume that A and B molecules repel each other only when their centres are within a cutoff radius r_0, and we use a harmonic form:

$$U(r) = \frac{1}{2}b(r - r_0)^2\,, \quad r \leqslant r_0\,, \tag{3.46}$$

and $U(r) = 0$ otherwise; here, $b > 0$ is a harmonic spring constant chosen to be stiff, $br_0 \gg k_BT$, and we set the cutoff equal to the reaction radius, $r_0 = R$.

The second potential is a commonly truncated form of the Lennard-Jones (LJ) potential, which combines a strong steric repulsion of nearly overlapping molecules with a short-range attraction due to van der Waals forces:

$$U(r) = 4\varepsilon\left[(\sigma/r)^{12} - (\sigma/r)^6\right]\theta(r_c - r), \tag{3.47}$$

with σ and $\varepsilon > 0$ being a length and an energy, respectively, that set the range and the strength of the interaction. The value of ε is also the depth of the potential well at $r = \sigma$. Here we choose σ such that the potential minimum lies *within* the reaction volume, specifically, the inflection point of $U(r)$ is set at the boundary, $R = (26/7)^{1/6}\sigma \approx 1.24\sigma$. The attractive part of the interaction is truncated at $r_c = 2.5R$.

3.6 Results and discussion

3.6.1 Macroscopic rates

Simulation results for the reaction rate constant k as a function of the propensity $\lambda = \kappa^2 D$ are shown in Fig. 3.4 for the above potentials. They are compared to the theoretical predictions from the reaction–diffusion problem, Eqs. (3.4), as follows: For the non-interacting case ($U = 0$), the exact solution is available in closed form, Eq. (3.26). For the soft repulsion and the LJ potential, the solution is available

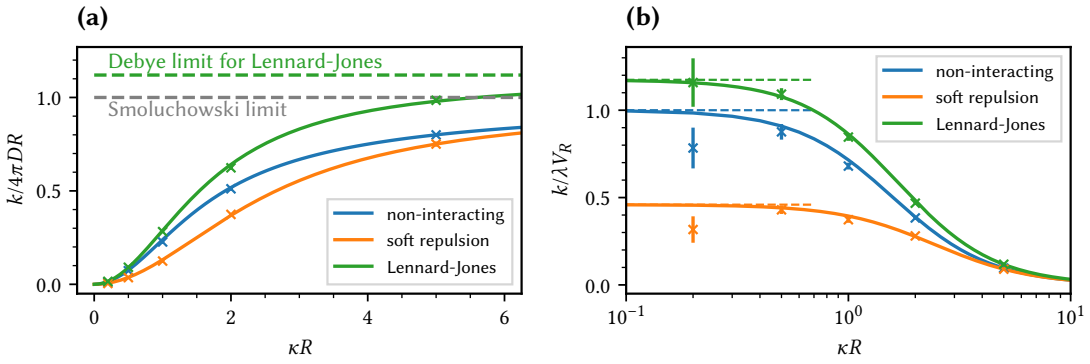

Figure 3.4: **(a):** Macroscopic rate constant k as a function of the reactivity κR with the inverse penetration depth $\kappa = \sqrt{\lambda/D}$ and the reaction radius R for different pair potentials $U(r)$. Data are given relative to the Smoluchowski rate constant $4\pi DR$ (grey dashed line) in terms of the relative diffusion constant $D = D_A + D_B$ and the reaction radius R. Symbols are results of interacting particle-based stochastic simulations of the reaction–diffusion process (iPRD simulations). Solid lines show theoretical predictions obtained from exact expressions [non-interacting case, Eq. (3.26)] or quasi-analytic solutions [soft harmonic repulsion and LJ potential, Eq. (3.32)] of the reaction–diffusion problem, Eqs. (3.4). The green dashed line indicates the Debye limit, Eq. (3.22), for the LJ potential. **(b):** Macroscopic rate constant k as a function of the reactivity κR normalized by the perturbative solution $k^{(0)} \simeq \lambda V_R$ of the non-interacting case for slow reactions [Eq. (3.40)]. Dashed lines indicate the ratios of the accessible to the total reaction volume V_{eff}/V_R for each potential [Eq. (3.41)], which is the prediction of perturbation theory. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

only in quasi-analytic form, Eq. (3.32), i.e., the final expressions for k are explicit in terms of a numerical quadrature as in the Debye problem and the numerical solution to a one-dimensional boundary value problem in the interior of the reaction sphere, see Section 3.4.4. As dimensionless control parameter we choose the combination $\kappa R = R\sqrt{\lambda/D}$, which distinguishes the reaction- and diffusion-limited regimes, $\kappa R \ll 1$ and $\kappa R \gg 1$, respectively. Equivalently, $(\kappa R)^2 = \lambda\tau_\mathrm{d}$ controls the reaction propensity relative to the diffusion time $\tau_\mathrm{d} = R^2/D$.

For all choices of the potential, the agreement between theory and simulations is excellent, see Fig. 3.4a. In all three cases, the reaction rate k increases monotonically with the reaction propensity λ and saturates at Debye's result, Eq. (3.22), for a diffusion-limited reaction ($\kappa R \longrightarrow \infty$). In this limit, the reaction occurs almost surely upon first contact and details inside of the reaction volume become irrelevant, the formation rate diverges, $k_\mathrm{f} \longrightarrow \infty$. Note that for the truncated soft repulsion, Eq. (4.6), the limiting value equals the Smoluchowski rate as the potential is zero in the outer domain. For slow reactions, $\kappa R \ll 1$, the initial increase of k depends quadratically on κR and it coincides with the prediction $k \simeq \lambda V_\mathrm{eff}$ of perturbation theory, Eq. (3.40). This regime is better visualised by normalising k with the perturbation result for the non-interacting case, $k^{(0)} = \lambda V_R$, where $V_R = (4\pi/3)R^3$, see Fig. 3.4b. From the limit $\kappa R \longrightarrow 0$ it is evident that also the constant of proportionality V_eff as calculated from Eq. (3.41) matches very well with the numerical results. For $\kappa R = 0.2$ noticeable relative deviations are seen in the simulation data, indicating that the slow-reaction regime is challenging to explore by the particle-based approaches such as iPRD. The figure shows further that the perturbation solution deviates by no less than 10% from the full solution for $\kappa R \lesssim 0.5$.

How is the reaction rate constant k changed due to the presence of the investigated potentials? A repulsion within the reaction volume slows down the reaction relative to the non-interacting case, which we attribute to the greatly diminished accessible reaction volume (Fig. 3.4, soft repulsion). The effect is most pronounced for slow reactions, which are most sensitive to a reduction of the actual penetration depth relative to its value κ^{-1} of the free case. Evaluating Eq. (3.41) for the specific harmonic repulsion used here, V_eff and thus k are reduced by a factor of ≈ 2.2 relative to the non-interacting case.

An attractive interaction between A and B molecules, on the other hand, is expected to enhance the encounter rate k_e and thus to speed up the overall reaction. Already the short-ranged well of the truncated LJ potential, Eq. (3.47), suffices to increase k_e by 12% with respect to the free case, Eq. (3.22). Noting that only the part of the potential outside of the reaction volume, $r > R$, contributes to k_e, we can test the dependence on the attraction by varying the interaction range σ at fixed R, see Fig. 3.5. The encounter rate becomes maximal at $\sigma = R$, i.e., when the integral in Eq. (3.22) is taken over the full domain where the potential is negative, $U(r) < U(r \longrightarrow \infty)$.

The ramifications of the potential on the formation rate k_f are more subtle: the strongly repulsive part of the LJ potential should lead to a decrease as the accessible reaction volume is diminished. Concomitantly, the potential well induces an enrichment of A molecules at the boundary of the reaction volume, which would increase k_f. The combination of both can lead to a non-monotonic dependence of the formation rate on the position of the reaction boundary relative to the potential well, which indeed we observe in the numerical solutions to Eq. (3.33), see Fig. 3.5. The position of the maximum in k_f depends on κR and shifts towards larger σ/R for higher reaction propensity. For the parameters given in Table 3.1, the effectively accessible reaction volume is *increased* by $\approx 17\%$ over the free volume V_R (Fig. 3.4b), and for all κR the overall rate constant k is larger than for non-interacting molecules.

By the Markov property of the microscopic reaction–diffusion process, the total reaction rate constant k is the harmonic mean of the partial rates for encounter and formation, Eq. (3.1), and thus, k is bounded from above by the smaller rate: $k \leqslant \min(k_\mathrm{e}, k_\mathrm{f})$. The relative importance of both processes depends on the rescaled reaction propensity κR, which is nicely seen from Fig. 3.6 for the Lennard-Jones potential with $\sigma/R = 0.1$ and $\varepsilon/k_B T = 13$. One reads off that the formation and diffusion-limited regimes, where the other contribution can safely be neglected, are delimited by $\kappa R \lesssim 10^{-1}$ and $\kappa R \gtrsim 10^{1}$, respectively. Inbetween, there is a wide window of propensities, where both processes enter the overall rate constant. Here, an

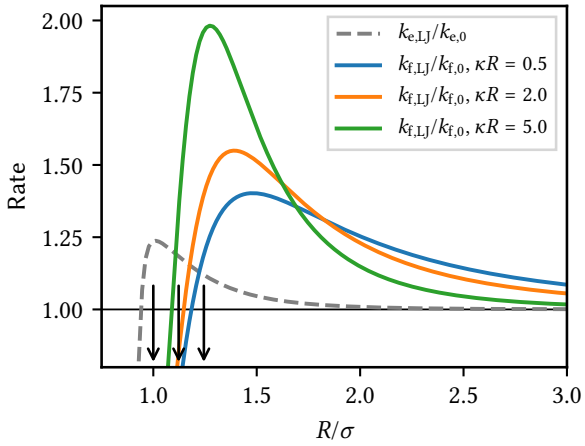

Figure 3.5: Dependence of the partial reaction rates k_e and k_f on the attractive part of the LJ potential with depth $\varepsilon/k_BT = 1$, which is tested by varying the interaction range σ for fixed reaction radius R. The rates are normalised by their values for the non-interacting case, Eqs. (3.22) and (3.27). Black arrows indicate the zero crossing, the minimum, and the inflection point of the Lennard-Jones potential. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

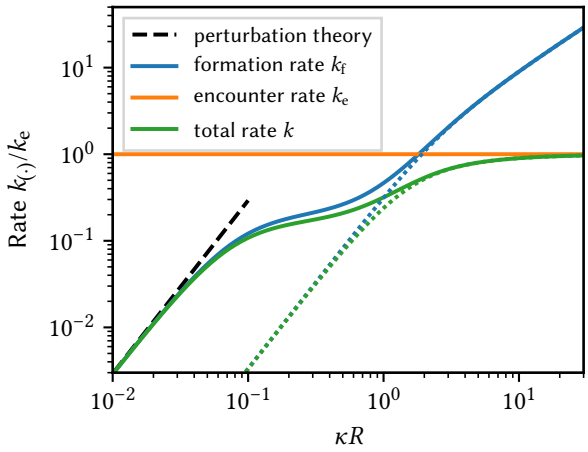

Figure 3.6: Encounter, formation and total rate constants as a function of the reactivity κR by changing the propensity $\lambda = \kappa^2 D$ for a Lennard-Jones potential with energy $\varepsilon/k_BT = 13$ and reaction radius $\sigma/R = 0.1$. The dashed line shows the perturbative solution where $k \propto \kappa^2$. Dotted lines represent the non-interacting case [Eqs. (3.26) and (3.27)], normalized by the corresponding encounter rate, $4\pi DR$. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

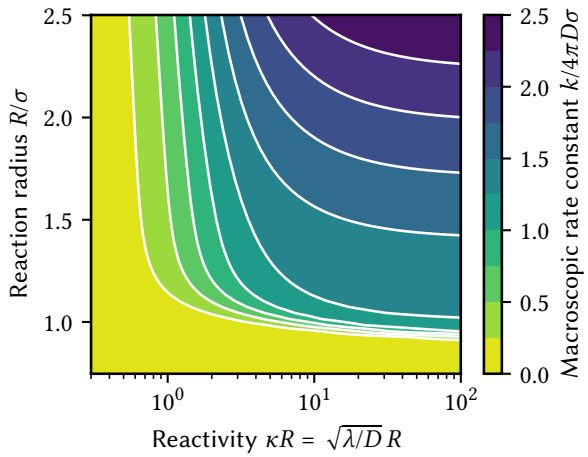

Figure 3.7: The macroscopic rate constant k in the presence of a Lennard–Jones potential with particle diameter σ and energy depth that is equal to the thermal energy $\varepsilon = k_B T$. Here k is a function of the unit–less reactivity $\kappa R = \sqrt{\lambda/D}\,R$ and a function of the reaction radius R, with the microscopic rate constant λ, relative diffusion constant D. k is given in units of $4\pi D\sigma$, which is the encounter rate up to particle diameter if no reaction and potential would be present. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

enhanced availability of reactants due to the deep potential well compensates a slower reaction propensity so that the formation rate displays an approximately plateau-like behaviour for $0.1 \lesssim \kappa R \lesssim 0.5$. For sufficiently fast reactions, the accumulation disappears and k_f starts increasing again towards its large κR behaviour, $k_f \sim \kappa R$, which resembles the potential-free case as reactions are confined to a thin shell near $r = R$. Note that k_f is a monotonic function of κR, which follows from Eq. (3.34) and anticipating the monotonic decrease of $p(R)$ as κR increases, see Fig. 3.8.

Motivated by the practical question how to choose the model parameters λ and R for given reaction rate k and diffusivity D and given interaction potential, we have scrutinized further the dependence of k on both the propensity κR and the reaction radius R/σ, exemplified for the Lennard-Jones potential (Fig. 3.7). For slow reactions, $\kappa R \lesssim 1$, the rate constant k is insensitive to the reaction radius. In the diffusion-limited regime, $\kappa R \gtrsim 10$, the rate constant k mainly depends on the reaction radius R/σ and is insensitive to the value of κR. Inbetween, $1 \lesssim \kappa R \lesssim 10$, both parameters must be adjusted carefully. From physical considerations, the reaction radius R should be comparable to the molecular radius σ, which delimits the freedom in the choice of λ.

3.6.2 Concentration profiles

Simulation results for the concentration profile $p(r)$, more precisely, the radial distribution of A molecules relative to Bs, are shown in Fig. 3.8 for three different propensities λ, expressed in terms of $\kappa = \sqrt{\lambda/D}$, and for the different interactions considered above. The data are compared to the theoretical predictions developed in Sections 3.3 and 3.4, and the quantitative agreement is very good for all cases studied. Thus, the iPRD simulations corroborate our theoretical analysis and the numerical results, which in turn are used to validate the implementation of the simulation algorithm.

For the non-interacting case (Fig. 3.8a), we have closed analytic expressions for $p(r)$ inside and outside

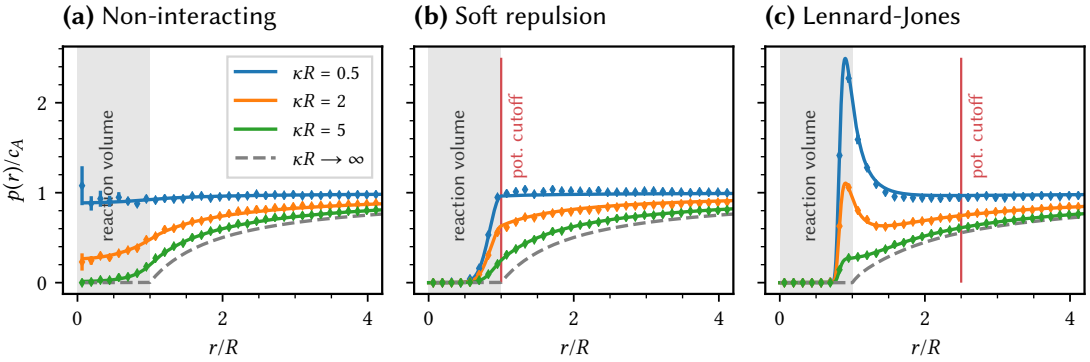

Figure 3.8: Radial distribution $p(r)$ of A molecules around a B molecule for different reaction propensities λ, here expressed by $\kappa = \sqrt{\lambda/D}$. The panels show results for (a) the non-interacting case, (b) the soft harmonic repulsion [Eq. (4.6)], and (c) a truncated LJ potential [Eq. (3.47)]. Data points are results from iPRD simulations, and solid lines theoretical predictions from Eqs. (3.21) and (3.25) for the non-interacting case and from numerical solutions to Eqs. (3.20) and (3.29) otherwise. Grey dashed lines represent the limit $\kappa R \longrightarrow \infty$ of almost sure reactions upon contact [Eq. (3.23)]. Grey shaded areas mark the interior of the reaction volume ($r \leqslant R$), and vertical lines indicate the respective positions r_c of the potential cutoffs. Reprinted from *The Journal of Chemical Physics* "Diffusion-influenced reaction rates in the presence of pair interactions", Dibak et al., 2019, with the permission of AIP Publishing.

of the reaction volume, Eqs. (3.25) and (3.21), respectively. For the soft repulsive and the LJ potentials [Eqs. (3.47) and (4.6)], profiles in the outer domain are obtained from Eq. (3.20) by a quadrature, and in the inner domain from the numerical solution for $\psi'(r)$ of the boundary value problem, Eq. (3.29). At distances $r > r_c$, where neither a reaction can occur nor a potential is present, the constant flux implies for the profile, $p(r) = c_A(1 - k/4\pi Dr)$, see Eq. (3.21).

For slow reactions, $\kappa R \ll 1$, the concentration profile at leading order in κR is expected to equal the equilibrium distribution, $p_0(r) = c_A e^{-\beta U(r)}$, subject to the specific boundary condition $p(r \longrightarrow \infty) = c_A$ [Eq. (3.37)]. Indeed, for $\kappa R = 0.5$ both the numerical and simulation results for $p(r)$ are hardly distinguishable from $p_0(r)$ in all three cases studied, see Fig. 3.8; for $U = 0$ it holds $p_0(r) = c_A$ everywhere. Upon increasing κR, the concentration is decreasing uniformly and, in the limit of an instantaneous product formation, $\kappa R \longrightarrow \infty$, the profile $p(r)$ vanishes inside the reaction volume and approaches Debye's solution, Eq. (3.23), outside as expected. For the non-interacting case and the soft repulsive potential, the latter simplifies to Smoluchowski's result, $p(r) = c_A(1 - R/r)$ for $r \geqslant R$; for the truncated LJ potential used here, the differences are small and hardly seen in the graph (Fig. 3.8c). Summarising, the equilibrium distribution and Debye's solution constitute upper and lower bounds on $p(r)$.

After having understood these limits, we will discuss the consequences of the interaction potential on the profiles in more detail. Adding a soft repulsion within the reaction volume to mimic an excluded volume largely reduces the probability of finding a particle inside the reaction volume (Fig. 3.8b) and thus suppresses the product formation rate k_f (see also Fig. 3.4b). Yet, the effect is more pronounced for slow reactions as the interior of the reaction volume becomes less and less accessible upon increasing κR, and we conclude that the repulsion is particularly relevant for slow reactions. The attractive well of the LJ potential on the other hand induces an enrichment of A molecules near the reaction boundary, which is more developed for smaller κR (Fig. 3.8c).

3.7 Conclusion

We have studied the reaction kinetics of a bimolecular association process $A + B \longrightarrow X$ in the steady state for molecules that diffuse in space and interact through an isotropic pair potential $U(r)$. Within Doi's volume reaction model, we have calculated the reaction rate constant k and the distribution function $p(r)$ of AB pairs as a function of the microscopic reaction propensity λ. The explicit dependence of the model on λ enables us to systematically probe the kinetics from the well-mixed to the diffusion-limited regime. The transition between the regimes is conveniently captured by the dimensionless quantity $R\sqrt{\lambda/D}$, which we abbreviate as the *reactivity* κR of an AB pair; the length κ^{-1} describes how far molecule centres can penetrate the reaction volume of radius R before they react and $D := D_A + D_B$ is the relative diffusion constant. Specifically, our approach bridges between the two well-studied cases $\kappa R \ll 1$ (reaction-limited or well-mixed) and $\kappa R \gg 1$ (diffusion-limited or fast-reaction limit). Similarly, $\lambda\tau_d = (\kappa R)^2$ can be used to classify these regimes, however in terms of the residence time $\tau_d = R^2/D$ in the reaction volume (as obtained for non-interacting molecules).

Over the entire spectrum of κR values and for arbitrary pair potentials, our analytical result for the reaction rate constant displays the Markovian decomposition $k^{-1} = k_e^{-1} + k_f^{-1}$ into encounter k_e and formation k_f rates [Eq. (3.32)]. Thereby, k_e is always given by Debye's result Eq. (3.22). Interestingly, k_f can be expressed in terms of k_e and the substrate concentration $p(R)$ at the reaction boundary, see Eq. (3.34), the latter being non-trivial to calculate. The well-mixed limit is dominated by the formation rate k_f and can be solved by perturbation theory (see Section 3.4.3), which yields $k = \lambda V_{\text{eff}}$ in terms of the effectively accessible reaction volume V_{eff}. In the absence of a potential, V_{eff} simplifies to the volume of the reactive sphere $V_R = (4\pi/3)R^3$. On the other hand, the diffusion limit is dominated by the encounter rate k_e: a reaction occurs almost surely upon entering the reaction volume. Our expression for k reproduces the Smoluchowski encounter rate $4\pi DR$ in the absence of potentials and Debye's result [Deb42], when particles diffuse subject to an interaction potential $U(r)$.

In the application-relevant diffusion-influenced regime (see Section 3.4), where k_e is of comparable magnitude as k_f, we obtained semi-analytical expressions for the rate k and the local concentration $p(r)$ that require numerical evaluation [Eqs. (3.32) and (3.35)]. Practically, one has to solve a one-dimensional boundary value problem for the reaction–diffusion equation inside the reaction volume and to compute an integral over the domain outside the reaction volume; the computational costs of both tasks are negligible. We tested our numerical scheme against explicit analytic solutions for a logarithmically repulsive potential. A closed expression for the rate k is given for general potentials outside in the case that molecules do not interact if their centres are within the reaction volume [Eq. (3.28)]; this may be useful to model, e.g., reactions in electrolytes while neglecting excluded volume.

We have studied the detailed dependence of the rate k on the reactivity parameter κR for two different potentials: a soft harmonic repulsion inside the reaction volume, and a truncated Lennard-Jones potential combining excluded volume and attraction. Our numerical results for the rate k and the concentration $p(r)$ show excellent agreement with extensive stochastic particle-based reaction-diffusion simulations. We draw the following physical conclusions:

1. A purely repulsive potential decreases both partial rates, k_e and k_f, and so also the overall rate constant k compared to the non-interacting case.

2. An attraction speeds up the reaction generally. Outside the reaction volume, it increases the encounter rate k_e; here, the sign of $U(r) - U(r \longrightarrow \infty)$ matters, which points at an energetic origin. For the formation rate k_f, the force $-U'(r)$ inside the reaction volume and the value $p(R)$ on the boundary enter.

3. For mixed situations as for the LJ potential, both contributions, k_e and k_f, are non-monotonic in the position of the reaction boundary (Fig. 3.5) and can lead to non-trivial dependencies of the total rate k on the model parameters λ and R (Fig. 3.6).

Concluding, we have established a microscopic simulation model that extends Doi's volume reaction model to interacting molecules. This model is at the core of iPRD simulations, which permit treatment of spatially resolved reaction processes in cells and nanotechnology at different levels of coarse graining. The obtained relation between k and the parameters λ, R facilitates the development of quantitative iPRD models based on experimental values of the macroscopic rate k. The interaction potential $U(r)$, can either be chosen *ad hoc* based on physical insight or determined as the potential of mean force in atomistic simulations [BSF11; Xu+19; Wu+16]. The freedom to choose an interaction potential within the reaction volume offers the opportunity to implement coarse-grained simulations that switch between representations of bound complexes using either explicit potential wells and barriers or stochastic reactions. The present study focuses on the dilute limit, which serves as a well-defined starting point for the investigation of concentration and crowding effects on the reaction rate and the distribution of molecules.

References

[Alb+16] D. Albrecht et al. "Nanoscopic compartmentalization of membrane protein motion at the axon initial segment". In: *Journal of Cell Biology* 215.1 (2016), pp. 37–46. DOI: 10.1083/jcb.201603108.

[And17] Steven S. Andrews. "Smoldyn: Particle-based simulation with rule-based modeling, improved molecular interaction and a library interface". In: *Bioinformatics* 33.5 (2017), pp. 710–717. DOI: 10.1093/bioinformatics/btw700.

[And18] Steven S. Andrews. "Particle-Based Stochastic Simulators". In: *Encyclopedia of Computational Neuroscience*. Ed. by Dieter Jaeger and Ranu Jung. New York, NY: Springer, 2018, pp. 1–5. DOI: 10.1007/978-1-4614-7320-6_191-2.

[AS90] Noam Agmon and Attila Szabo. "Theory of reversible diffusion-influenced reactions". In: *The Journal of Chemical Physics* 92.9 (1990), p. 5270. DOI: 10.1063/1.458533.

[AT08] M. Armand and J. M. Tarascon. "Building better batteries". In: *Nature* 451.7179 (2008), pp. 652–657. DOI: 10.1038/451652a.

[AT17] Satya N. V. Arjunan and Koichi Takahashi. "Multi-algorithm particle simulations with Spatiocyte". In: *Protein Function Prediction*. Springer, 2017, pp. 219–236. DOI: 10.1007/978-1-4939-7015-5_16.

[Ban+16] Daniel S. Banks et al. "Characterizing anomalous diffusion in crowded polymer solutions and gels over five decades in time with variable-lengthscale fluorescence correlation spectroscopy". In: *Soft Matter* 12 (2016), pp. 4190–4203. DOI: 10.1039/c5sm01213a.

[Bha04] Upinder S. Bhalla. "Signaling in Small Subcellular Volumes. I. Stochastic and Diffusion Effects on Individual Pathways". In: *Biophysical Journal* 87.2 (2004), pp. 733–744. DOI: 10.1529/biophysj.104.040469.

[Bie+15] Johann Biedermann et al. "ReaDDyMM: Fast interacting particle reaction-diffusion simulations using graphical processing units." In: *Biophysical journal* 108.3 (2015), pp. 457–61. DOI: 10.1016/j.bpj.2014.12.025.

[BSF11] Ignasi Buch, S. Kashif Sadiq, and Gianni De Fabritiis. "Optimized potential of mean force calculations for standard binding free energies". In: *Journal of Chemical Theory and Computation* 7.6 (2011), pp. 1765–1772. DOI: 10.1021/ct2000638.

[BSS14] Jacqueline Burré, Manu Sharma, and Thomas C. Südhof. "α-Synuclein assembles into higher-order multimers upon membrane binding to promote SNARE complex formation". In: *Proceedings of the National Academy of Sciences of the United States of America* 111.40 (2014), E4274–E4283. DOI: 10.1073/pnas.1416598111.

[CK49a] Frank C. Collins and George E. Kimball. "Diffusion-controlled reaction rates". In: *Journal of Colloid Science* 4.4 (1949), pp. 425–437. DOI: 10.1016/0095-8522(49)90023-9.

[CK49b] Frank C. Collins and George E. Kimball. "Diffusion-Controlled Reactions in Liquid Solutions". In: *Industrial and Engineering Chemistry Research* 41.11 (1949), pp. 2551–2553. DOI: 10.1021/ie50479a040.

[Deb42] P. Debye. "Reaction Rates in Ionic Solutions". In: *Transactions of The Electrochemical Society* 82.1 (1942), p. 265. DOI: 10.1149/1.3071413.

[Dib+18] Manuel Dibak et al. "MSM/RD: Coupling Markov state models of molecular kinetics with reaction-diffusion simulations". In: *The Journal of Chemical Physics* 148.21 (2018), p. 214107. DOI: 10.1063/1.5020294.

[Dib+19] Manuel Dibak et al. "Diffusion-influenced reaction rates in the presence of pair interactions". In: *The Journal of Chemical Physics* 151.16 (2019), p. 164105. DOI: 10.1063/1.5124728.

[Doi75a] Masao Doi. "Theory of diffusion-controlled reactions between non-simple molecules. I". In: *Chemical Physics* ll (1975), pp. 107–113. DOI: 10.1016/0301-0104(75)80043-7.

[Doi75b] Masao Doi. "Theory of diffusion-controlled reactions between non-simple molecules. II". In: *Chemical Physics* 3 (1975), pp. 115–121. DOI: 10.1016/0301-0104(75)80044-9.

[Doi76] Masao Doi. "Stochastic theory of diffusion-controlled reaction". In: *Journal of Physics A* 9.9 (1976), p. 1479. DOI: 10.1088/0305-4470/9/9/009.

[Dor+10] N. Dorsaz et al. "Diffusion-limited reactions in crowded environments". In: *Physical Review Letters* 105.12 (2010), p. 120601. DOI: 10.1103/PhysRevLett.105.120601.

[DYK18] Aleksandar Donev, Chiao-yu Yang, and Changho Kim. "Efficient reactive Brownian dynamics". In: *The Journal of Chemical Physics* 148.3 (2018), p. 034103. DOI: 10.1063/1.5009464.

[EC09] Radek Erban and S Jonathan Chapman. "Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions". In: *Physical Biology* 6.4 (2009), p. 046001. DOI: 10.1088/1478-3975/6/4/046001.

[EK15] Carlos Echeverria and Raymond Kapral. "Enzyme kinetics and transport in a system crowded by mobile macromolecules". In: *Physical Chemistry Chemical Physics* 17 (2015), pp. 29243–29250. DOI: 10.1039/C5CP05056A.

[ELM18] Stefan Engblom, Per Lötstedt, and Lina Meinecke. "Mesoscopic modeling of random walk and reactions in crowded media". In: *Physical Review E* 98.3 (2018), p. 033304. DOI: 10.1103/PhysRevE.98.033304.

[Eto+18] Fred Etoc et al. "Non-specific interactions govern cytosolic diffusion of nanosized objects in mammalian cells". In: *Nature Materials* 17.8 (2018), pp. 740–746. DOI: 10.1038/s41563-018-0120-7.

[FN18] Christoph Fröhner and Frank Noé. "Reversible Interacting-Particle Reaction Dynamics". In: *The Journal of Physical Chemistry B* 122.49 (2018), pp. 11240–11250. DOI: 10.1021/acs.jpcb.8b06981.

[Gal+16] Marta Galanti et al. "Reaction rate of a composite core–shell nanoreactor with multiple nanocatalysts". In: *Physical Chemistry Chemical Physics* 18.30 (2016), pp. 20758–20767. DOI: 10.1039/C6CP01179A.

[Gil07] Daniel T. Gillespie. "Stochastic Simulation of Chemical Kinetics". In: *Annual Review of Physical Chemistry* 58.1 (2007), pp. 35–55. DOI: 10.1146/annurev.physchem.58.032806.104637.

[GMO18] Denis S. Grebenkov, Ralf Metzler, and Gleb Oshanin. "Strong defocusing of molecular reaction times results from an interplay of geometry and reaction control". In: *Communications Chemistry* 1.1 (2018), p. 96. DOI: 10.1038/s42004-018-0096-x.

[GYB10] R. Grima, S. N. Yaliraki, and M. Barahona. "Crowding-Induced Anisotropic Transport Modulates Reaction Kinetics in Nanoscale Porous Media". In: *Journal of Physical Chemistry B* 114.16 (2010), pp. 5380–5385. DOI: 10.1021/jp9025865.

[Her+12] Pablo Hervés et al. "Catalysis by metallic nanoparticles in aqueous solution: model reactions". In: *Chemical Society Reviews* 41 (2012), pp. 5577–5587. DOI: 10.1039/C2CS35029G.

[HF13] Felix Höfling and Thomas Franosch. "Anomalous transport in the crowded world of biological cells". In: *arXiv* 76.4 (2013), p. 046602. DOI: 10.1088/0034-4885/76/4/046602.

[HFN19] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics". In: *PLoS Computational Biology* 15.2 (2019), e1006830. DOI: 10.1371/journal.pcbi.1006830.

[Höf+08] Felix Höfling et al. "Critical dynamics of ballistic and Brownian particles in a heterogeneous environment". In: *Journal of Chemical Physics* 128.16, 164517 (2008), p. 164517. DOI: 10.1063/1.2901170.

[Hor+10] Margaret R. Horton et al. "Development of anomalous diffusion among crowding proteins". In: *Soft Matter* 6 (2010), pp. 2648–2656. DOI: 10.1039/B924149C.

[Hou10] Miles D. Houslay. "Underpinning compartmentalised cAMP signalling through targeted cAMP breakdown". In: *Trends in Biochemical Sciences* 35.2 (2010), pp. 91–100. DOI: 10.1016/j.tibs.2009.09.007.

[JH14] Margaret E. Johnson and Gerhard Hummer. "Free-Propagator Reweighting Integrator for Single-Particle Dynamics in Reaction-Diffusion Models of Heterogeneous Protein-Protein Interaction Systems". In: *Physical Review X* 4.3 (2014), p. 031037. DOI: 10.1103/PhysRevX.4.031037.

[Kus+05] Akihiro Kusumi et al. "Paradigm Shift of the Plasma Membrane Concecpt from the two-dimensional Continuum Fluid to the Partitioned Fluid: High-Speed Single-Molecule Tracking of Membrane Molecules". In: *Annual Review of Biophysics and Biomolecular Structure* 34.1 (2005), pp. 351–378. DOI: 10.1146/annurev.biophys.34.040204.144637.

[KY10] Jun Soo Kim and Arun Yethiraj. "Crowding Effects on Association Reactions at Membranes". In: *Biophysical Journal* 98.6 (2010), pp. 951–958. DOI: 10.1016/j.bpj.2009.11.022.

[MJC16] R. Metzler, J.-H. Jeon, and A.G. Cherstvy. "Non-Brownian diffusion in lipid membranes: Experiments and simulations". In: *Biochimica et Biophysica Acta (BBA) - Biomembranes* 1858.10 (2016), pp. 2451–2467. DOI: 10.1016/j.bbamem.2016.01.022.

[ML16] Paul J. Michalski and Leslie M. Loew. "SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume". In: *Biophysical journal* 110.3 (2016), pp. 523–529. DOI: 10.1016/j.bpj.2015.12.026.

[MM06] Eurico Melo and Jorge Martins. "Kinetics of bimolecular reactions in model bilayers and biological membranes. A critical review". In: *Biophysical Chemistry* 123.2-3 (2006), pp. 77–94. DOI: 10.1016/j.bpc.2006.05.003.

[MW08] Marco J. Morelli and Pieter Rein ten Wolde. "Reaction Brownian dynamics and the effect of spatial fluctuations on the gain of a push-pull network". In: *The Journal of Chemical Physics* 129.5 (2008), p. 054112. DOI: 10.1063/1.2958287.

[Pau+17] Fabian Paul et al. "Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations". In: *Nature Communications* 8.1 (2017), p. 1095. DOI: 10.1038/s41467-017-01163-6.

[PF19] Charlotte F. Petersen and Thomas Franosch. "Anomalous transport in the soft-sphere Lorentz model". In: *Soft Matter* 15 (2019), pp. 3906–3913. DOI: 10.1039/C9SM00442D.

[Pla+17] Nuria Plattner et al. "Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling". In: *Nature Chemistry* 9.10 (2017), pp. 1005–1011. DOI: 10.1038/nchem.2785.

[Poi+15] Peter Poier et al. "An Anisotropic Effective Model for the Simulation of Semiflexible Ring Polymers". In: *Macromolecules* 48.14 (2015), pp. 4983–4997. DOI: 10.1021/acs.macromol.5b00603.

[Rid+08] Douglas Ridgway et al. "Coarse-Grained Molecular Simulation of Diffusion and Reaction Kinetics in a Crowded Virtual Cytoplasm". In: *Biophysical Journal* 94.10 (2008), pp. 3748–3759. DOI: 10.1529/biophysj.107.116053.

[RQN18] Mauricio J. del Razo, Hong Qian, and Frank Noé. "Grand canonical diffusion-influenced reactions: A stochastic theory with applications to multiscale reaction-diffusion simulations". In: *Journal of Chemical Physics* 149.4 (2018), p. 044102. DOI: 10.1063/1.5037060.

[Sch+14] Johannes Schöneberg et al. "Explicit Spatiotemporal Simulation of Receptor-G Protein Coupling in Rod Cell Disk Membranes". In: *Biophysical Journal* 107.5 (2014), pp. 1042–1053. DOI: 10.1016/j.bpj.2014.05.050.

[Sch+15] Simon K. Schnyder et al. "Rounding of the localization transition in model porous media". In: *Soft Matter* 11.4 (2015), pp. 701–711. DOI: 10.1039/C4SM02334J.

[Sch+17] Johannes Schöneberg et al. "Lipid-mediated PX-BAR domain recruitment couples local membrane constriction to endocytic vesicle fission". In: *Nature Communications* 8.May (2017), p. 15873. DOI: 10.1038/ncomms15873.

[Sco+16] Duncan E. Scott et al. "Small molecules, big targets: Drug discovery faces the protein-protein interaction challenge". In: *Nature Reviews Drug Discovery* 15.8 (2016), pp. 533–550. DOI: 10.1038/nrd.2016.29.

[SG18] Stephen Smith and Ramon Grima. "Spatial Stochastic Intracellular Kinetics: A Review of Modelling Approaches". In: *Bulletin of Mathematical Biology* 81.8 (2018), pp. 2960–3009. DOI: 10.1007/s11538-018-0443-1.

[Smi85] Gordon D. Smith. *Numerical solution of partial differential equations: finite difference methods*. Oxford University Press, 1985. ISBN: 9780198596509.

[Smo16] Marian von Smoluchowski. "Versuch einer mathematischen Theorie der Koagulationskinetik kolloider Lösungen". In: *Zeitschrift fur Physikalische Chemie* 92U.1 (1916), p. 129. DOI: 10.1515/zpch-1918-9209.

[Smo17] M. v. Smoluchowski. "Grundriß der Koagulationskinetik kolloider Lösungen". In: *Kolloid-Zeitschrift* 21.3 (1917), pp. 98–104. DOI: 10.1007/BF01427232.

[SN13] Johannes Schöneberg and Frank Noé. "ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments." In: *PLoS One* 8.9 (2013), e74261. DOI: 10.1371/journal.pone.0074261.

[SN17] Luigi Sbailò and Frank Noé. "An efficient multi-scale Green's function reaction dynamics scheme". In: *The Journal of Chemical Physics* 147.18 (2017), p. 184106. DOI: 10.1063/1.5010190.

[Spa+16] Markus Spanner et al. "Splitting of the Universality Class of Anomalous Transport in Crowded Media". In: *Physical Review Letters* 116 (2016), p. 060601. DOI: 10.1103/PhysRevLett.116.060601.

[SS19] Luigi Sbailò and Luigi Delle Site. "On the formalization of asynchronous first passage algorithms". In: *Journal of Chemical Physics* 150.13 (2019), p. 134106. DOI: 10.1063/1.5083147.

[SS82] David Shoup and Attila Szabo. "Role of diffusion in ligand binding to macromolecules and cell-bound receptors." In: *Biophysical Journal* 40.1 (1982), p. 33. DOI: 10.1016/S0006-3495(82)84455-X.

[SSS80] Attila Szabo, Klaus Schulten, and Zan Schulten. "First passage time approach to diffusion controlled reactions". In: *Journal of Chemical Physics* 72.8 (1980), pp. 4350–4357. DOI: 10.1063/1.439715.

[SUN14] Johannes Schöneberg, Alexander Ullrich, and Frank Noé. "Simulation tools for particle-based reaction-diffusion dynamics in continuous space". In: *BMC Biophysics* 7.1 (2014), p. 11. DOI: 10.1186/s13628-014-0011-5.

[SW16] Olivia Stiehl and Matthias Weiss. "Heterogeneity of crowded cellular fluids on the meso- and nanoscale". In: *Soft Matter* 12.47 (2016), pp. 9413–9416. DOI: 10.1039/c6sm01436d.

[SWN18] Mohsen Sadeghi, Thomas R. Weikl, and Frank Noé. "Particle-based membrane model for mesoscopic simulation of cellular dynamics". In: *The Journal of Chemical Physics* 148.4 (2018), p. 044901. DOI: 10.1063/1.5009107.

[TS67] Ei Teramoto and Nanako Shigesada. "Theory of bimolecular reaction processes in liquids". In: *Progress of Theoretical Physics* 37.1 (1967), pp. 29–51. DOI: 10.1143/PTP.37.29.

[TT14] Fabio Trovato and Valentina Tozzini. "Diffusion within the Cytoplasm: A Mesoscale Model of Interacting Macromolecules". In: *Biophysical Journal* 107.11 (2014), pp. 2579–2591. DOI: 10.1016/j.bpj.2014.09.043.

[VBW15] Adithya Vijaykumar, Peter G. Bolhuis, and Pieter Rein ten Wolde. "Combining molecular dynamics with mesoscopic Green's function reaction dynamics simulations". In: *The Journal of Chemical Physics* 143.21 (2015), p. 214102. DOI: 10.1063/1.4936254.

[Wei14] Matthias Weiss. "Crowding, Diffusion, and Biochemical Reactions". In: *New Models of the Cell Nucleus: Crowding, Entropic Forces, Phase Separation, and Fractals*. Ed. by Ronald Hancock and Kwang W. Jeon. Vol. 307. Int. Rev. Cell Mol. Biol. Academic Press, 2014. Chap. 11, pp. 383–417. DOI: 10.1016/B978-0-12-800046-5.00011-4.

[Wit+19] Patrick Witzel et al. "Heterogeneities Shape Passive Intracellular Transport". In: *Biophysical Journal* 117.2 (2019), pp. 203–213. DOI: 10.1016/j.bpj.2019.06.009.

[WS16] Stefanie Winkelmann and Christof Schütte. "The spatiotemporal master equation: Approximation of reaction-diffusion dynamics via Markov state modeling". In: *The Journal of Chemical Physics* 145.21 (2016), p. 214107. DOI: 10.1063/1.4971163.

[Wu+16] Hao Wu et al. "Multiensemble Markov models of molecular thermodynamics and kinetics". In: *Proceedings of the National Academy of Sciences of the United States of America* 113.23 (2016), E3221–E3230. DOI: 10.1073/pnas.1525092113.

[Xu+19] Xiao Xu et al. “Interaction of Proteins with Polyelectrolytes: Comparison of Theory to Experiment”. In: *Langmuir* 35.16 (2019), pp. 5373–5391. DOI: 10.1021/acs.langmuir.8b01802.

[Zho+17] Limin Zhou et al. “Recent Developments on and Prospects for Electrode Materials with Hierarchical Structures for Lithium-Ion Batteries”. In: *Advanced Energy Materials* 8.6 (2017), p. 1701415. DOI: 10.1002/aenm.201701415.

[ZRM08] Huan-Xiang Zhou, Germán Rivas, and Allen P. Minton. “Macromolecular Crowding and Confinement: Biochemical, Biophysical, and Potential Physiological Consequences”. In: *Annual Review of Biophysics* 37.1 (2008), pp. 375–397. DOI: 10.1146/annurev.biophys.37.032807.125817.

Chapter 4

ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics

The results of this chapter have been published in the following paper:

> Moritz Hoffmann, Christoph Fröhner (CF), and Frank Noé. "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics". In: *PLoS Computational Biology* 15.2 (2019), e1006830. DOI: 10.1371/journal.pcbi.1006830

The contributions of the authors were as follows: Moritz Hoffmann, CF and Frank Noé conceived the project. Moritz Hoffmann implemented a majority of ReaDDy 2. CF implemented parts of ReaDDy 2. CF laid out the validation scenarios contained in the given paper. Moritz Hoffmann and CF analyzed and visualized the data. All contributors wrote the paper.

Summary

Interacting-particle reaction dynamics (iPRD) combines the simulation of dynamical trajectories of interacting particles as in molecular dynamics (MD) simulations with reaction kinetics, in which particles appear, disappear, or change their type and interactions based on a set of reaction rules. This combination facilitates the simulation of reaction kinetics in crowded environments, involving complex molecular geometries such as polymers, and employing complex reaction mechanisms such as breaking and fusion of polymers. iPRD simulations are ideal to simulate the detailed spatiotemporal reaction mechanism in complex and dense environments, such as in signalling processes at cellular membranes, or in nano- to microscale chemical reactors. Here we introduce the iPRD software ReaDDy 2, which provides a Python interface in which the simulation environment, particle interactions and reaction rules can be conveniently defined and the simulation can be run, stored and analyzed. A C++ interface is available to enable deeper and more flexible interactions with the framework. The main computational work of ReaDDy 2 is done in hardware-specific simulation kernels. While the version introduced here provides single- and multi-threading CPU kernels, the architecture is ready to implement GPU and multi-node

kernels. We demonstrate the efficiency and validity of ReaDDy 2 using several benchmark examples. ReaDDy 2 is available at the https://readdy.github.io/ website.

4.1 Introduction

The physiological response of biological cells to stimuli can be a many-stage process. A widely studied example is the MAPK pathway [XGG96; TTW10]. Many of such signaling pathways incorporate G-protein coupled receptors (GPCR) [Trz+12] and cyclic adenosine monophosphate (cAMP) [BB02]. These are related to various diseases [TD11; Dra+09; Abr+04]. An extracellular stimulus can activate the membrane bound GPCRs and lead to localized synthesis of cAMP as second messengers. Their transport through the cell is diffusive, however due to the geometry of cellular compartments cAMP molecules are non-uniformly distributed [ACH16; Hou10]. Their presence needs to be resolved in space and time to understand their function.

Particle-based reaction dynamics (PBRD) simulations [EC09; ZW05b; DYK18] are amongst the most detailed approaches to model reaction kinetics computationally as they simulate each reactive molecule as a particle and therefore can be used as a tool to investigate systems of the aforementioned kind. Reactions can occur when reactive particles are in proximity, resembling the physical process. PBRD is suitable when the spatial distribution of molecules does not equilibrate rapidly and must therefore be resolved, and some reactants are locally scarce, such that their discrete number must be kept track of [HS14; Bha04]. There is a wide range of simulation tools for PBRD [SUN14], including Smoldyn [And17], MCell [Ker+08], Cell++ [San+06], eGFRD [TTW10], mesoRD [HFE05], spatiocyte [AT10], SpringSaLaD [ML16], and SRSim [Gru+10]. A simulation tool that takes the molecular structure into account is SDA [Mar+15; GW98; GW97]. A recent review of particle-based stochastic simulators can be found in [And18]. Alternatively when the spatial resolution is of less importance, one can apply more efficient tools like Lattice Microbes [RSL13; Rob+11; Rob+09] which generates realizations of the reaction-diffusion master equation (RDME) [Isa13; Isa09]. In case of large copy numbers of particles it can make sense to think of them in terms of concentrations and build hybrid models [Fra+13].

PBRD simulations usually contain purely reactive particles that are not subject to interaction forces, e.g., to model space exclusion with repulsive interactions or clustering with attractive interactions. On the other hand, molecular dynamics (MD) simulations are designed to model particle dynamics including complex interactions between the particles or particles and an external field. The particles in MD simulations are often atoms or groups of atoms and higher-order structures such as molecules are represented by topology graphs that define the bonding structure between particles and thus, together with a MD force field, imply which pair, triplets and quadruplets or particles interact by means of bond, angle and torsion potentials. While reactive force fields [HKD07; KW11; Dui+01] include reactivity on the chemistry scale, and soft matter MD simulation tools include breakable bonds [Arn+13; Lim+06], current MD models and simulation packages do not incorporate generic particle reactions.

Interacting-particle reaction dynamics (iPRD) was introduced in [SN13] to combine the benefits of PBRD and MD simulations by modeling particle-based reaction dynamics while enabling full-blown interactions between particles as well as particles and the environment. Available simulation tools that are capable of special cases of iPRD simulations are, e.g., the MD packages LAMMPS [Auh+03; Pli95] which is capable of forming and breaking bonds dynamically and ESPResSo [Arn+13; Lim+06] which additionally has an implementation of catalytic reactions. In comparison to the iPRD simulator ReaDDy [SN13], these do not support full iPRD and are built and optimized for particle numbers that stay roughly constant. Comparing iPRD and PBRD, the interaction potentials in iPRD can be used to induce structure on mesoscopic length scales, e.g., volume-exclusion in crowded systems [HF13; SN13], clustering of weakly interacting macromolecules [Ull+15], restriction of diffusing particles to arbitrarily-shaped membranes [SN13; Gun+15; SUN14]. Furthermore it allows to study the large-scale structure of

oligomers [KS14], polymers and membranes [SWN18]. When not only considering interactions but also reactions, a wide range of reactive biochemical systems are in the scope of the model. For example, the reaction dynamics of photoreceptor proteins in crowded membranes [Sch+14] including cooperative effects of transmembrane protein oligomers [Gun+15] have been investigated. Another example is endocytosis, in which different proteins interact in very specific geometries [Pos+13; Sch+17]. The simulation tool Cytosim [NF07] is another software package that can be used to investigate mesoscopic biochemical systems, specifically geared towards the simulation of the cytoskeleton.

The price of resolving these details is that the computation is dominated by computing particle-particle interaction forces. Although non-interacting particles can be propagated quickly by exploiting solutions of the diffusion equation [SN17; ZW05a; ZW05b; Don+10], interacting particles are propagated with small time-steps [VBW15; Vij+17], restricting the accessible simulation timescales whenever parts of the system are dense. As this computational expense is not entirely avoidable when the particle interactions present in iPRD are needed to model the process of interest realistically, it is important to have a simulation package that can fully exploit the computational resources.

Here we introduce the iPRD simulation framework ReaDDy 2, which is significantly faster, more flexible, and more conveniently usable than its predecessor ReaDDy 1 [SN13; Bie+15]. Specifically, ReaDDy 2 includes the following new features:

- **Computational efficiency and flexibility**: ReaDDy 2 defines computing kernels which perform the computationally most costly operations and are optimized for a given computing environment. The current version provides a single-CPU kernel that is four to ten times (depending on system size) faster than ReaDDy 1, and a multi-CPU kernel that scales with 80% efficiency to number of physical CPU cores for large particle systems (Section 4.3.2). Kernels for GPUs or parallel multi-node kernels can be readily implemented with relatively little additional programming work (Section 4.3).

- **Python user interface**: ReaDDy 2 can be installed via the conda package manager and used as a regular python package. The python interface provides the user with functionality to compose the simulation system, define particle interactions, reactions and parameters, as well as run, store and analyze simulations.

- **C++ user interface**: ReaDDy 2 is mainly implemented in C++. Developers interested in extending the functionality of ReaDDy 2 in a way that interferes with the compute kernels, e.g., by adding new particle dynamics or reaction schemes, can do that via the C++ user interface.

- **Reversible reaction dynamics**: ReaDDy 2 can treat reversible iPRD reactions by using steps that obey detailed balance, as described in [FN18] (iPRD-DB), and thus ensure correct thermodynamic behavior for such reactions (Section 4.4.1).

- **Topologies**: We enable building complex multi-particle structures, such as polymers, by defining topology graphs (briefly: topologies, see Section 4.2.4). As in MD simulations, topologies are an efficient way to encode which bonded interactions (bond, angle and torsion terms) should act between groups of particles in the same topology. Note that particles in topologies can still be reactive. For example, it is possible to define reactions that involve breaking or fusing polymers (Section 4.4.4).

- **Potentials and boundaries**: Furthermore, the range of by default supported interaction potentials has been broadened, now including harmonic repulsion, a harmonic interaction potential with a potential well, Lennard–Jones interaction, and screened electrostatics. The simulation volume can also be equipped with partially or fully periodic boundary conditions.

This chapter summarizes the features of ReaDDy 2 and the demonstrates its efficiency and validity of ReaDDy 2 using several benchmarks and reactive particle systems. With few exceptions, we limit our description to the general features that are not likely to become outdated in future versions. Please see https://readdy.github.io/ for more details, tutorials and sample code.

4.2 interacting-Particle Reaction Dynamics (iPRD)

The ReaDDy 2 simulation system consists of particles interacting by potentials and reactions (Fig. 4.1) at a temperature T. Such a simulation system is confined to a box with repulsive or periodic boundaries. A boundary always has to be either periodic or be equipped with repulsive walls so that particles cannot diffuse away arbitrarily. To simulate iPRD dynamics in complex architectures, such as cellular membrane environments with specific shapes, additional potentials can be defined that confine the particle to a sub-volume of the simulation box (Section 4.2.2).

4.2.1 Interacting particle dynamics

ReaDDy 2 provides a developer interface to flexibly design models of how particle dynamics are propagated in time. The default model, however, is overdamped Langevin dynamics with isotropic diffusion as this is the most commonly used PBRD and iPRD model. In these dynamics a particle i moves according to the stochastic differential equation:

$$\frac{\mathrm{d}\mathbf{x}_i(t)}{\mathrm{d}t} = -\frac{D_i(T)}{k_B T}\mathbf{f}_i(t) + \sqrt{2D_i(T)}\boldsymbol{\xi}_i(t), \tag{4.1}$$

where $\mathbf{x}_i(t) \in \mathbb{R}^3$ contains the particle position at time t, $D_i(T)$ is its diffusion coefficient, k_B is the Boltzmann constant, and T the system temperature. The particle moves according to the deterministic force $\mathbf{f}_i$ and the stochastic velocity $\sqrt{2D_i(T)}\boldsymbol{\xi}_i$ in which $\boldsymbol{\xi}_i$ are independent, Gaussian distributed random variables with moments

$$\langle \boldsymbol{\xi}_i(t)\rangle = \mathbf{0}, \quad \langle \boldsymbol{\xi}_i(t)\boldsymbol{\xi}_i^\top(t')\rangle = \mathbf{I}\delta(t-t'),$$

where $\mathbf{I}$ is the identity matrix. The stochastic terms $\boldsymbol{\xi}_i$ and $\boldsymbol{\xi}_j$ are uncorrelated for particles $i \neq j$. In ReaDDy 2 the default assumption is that the diffusion coefficients $D_i(T)$ are given for the simulation temperature T. Additionally, we offer the option to define diffusion coefficients for a reference temperature $T_0 = 293K$ and then generate the diffusion coefficients at the simulation temperature T by employing the Einstein-Smoluchowski model for particle diffusion in liquids [Von06; Ein05]:

$$D_i(T) = D_i(T_0)\frac{T}{T_0}.$$

This way, simulations at different temperatures are convenient while only having to specify one diffusion constant. Using this model, the dynamics are

$$\frac{\mathrm{d}\mathbf{x}_i(t)}{\mathrm{d}t} = -\frac{D_i(T_0)}{k_B T_0}\mathbf{f}_i(t) + \sqrt{2D_i(T_0)\frac{T}{T_0}}\boldsymbol{\xi}_i(t). \tag{4.2}$$

This means that the mobility is preserved if the temperature changes and Eq. (4.1) is recovered for $T = T_0$.

The simplest integration scheme for Eqs. (4.1) and (4.2) is Euler-Maruyama, according to which the particle positions evolve as:

$$\mathbf{x}_i(t+\tau) = \mathbf{x}_i(t) - \tau\frac{D_i(T_0)}{k_B T_0}\mathbf{f}_i(t) + \sqrt{2D_i(T_0)\frac{T}{T_0}\tau}\boldsymbol{\eta}_t. \tag{4.3}$$

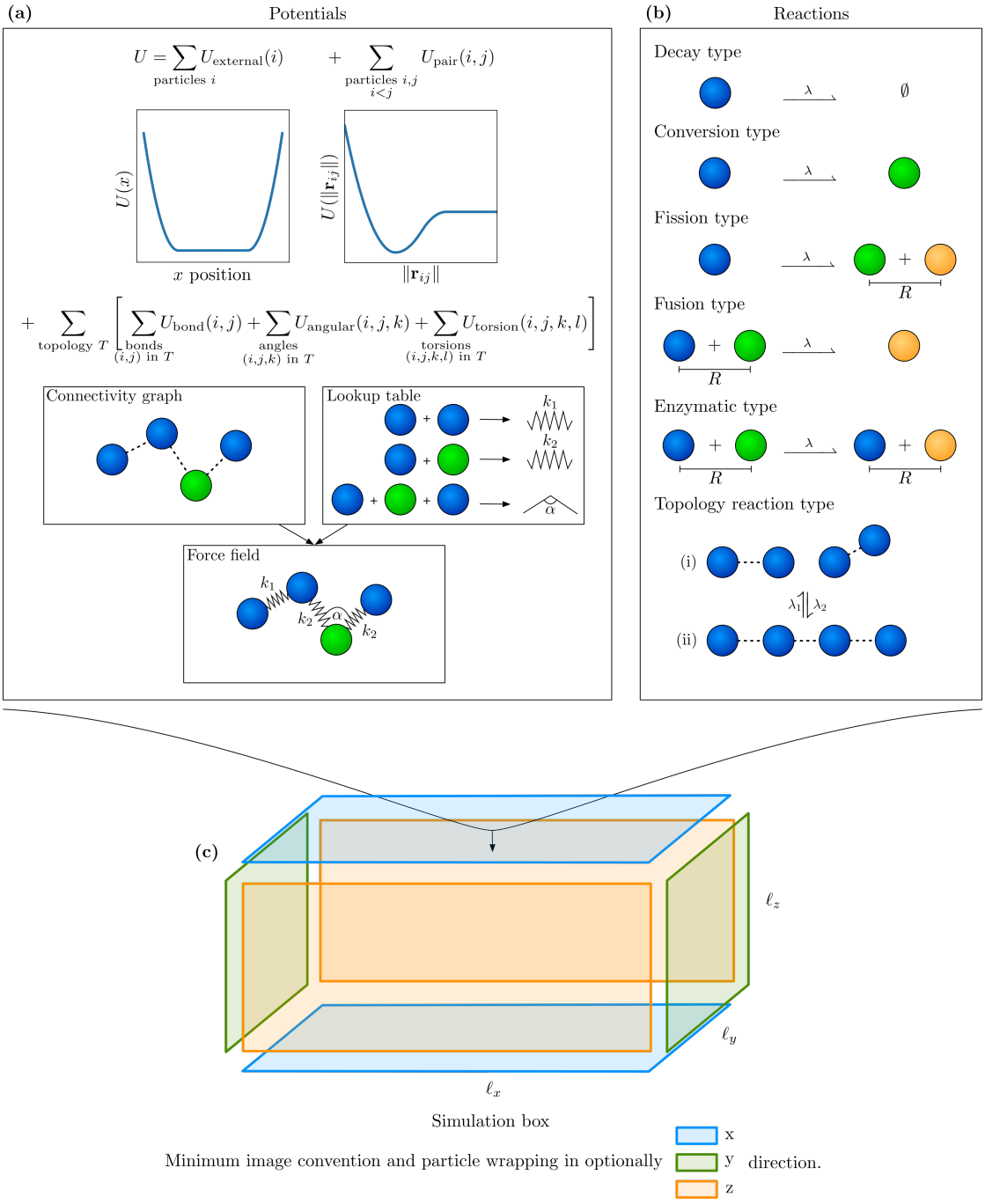

Figure 4.1: **The simulation model.** **(a)** Potentials: Particles are subject to position-dependent external potentials, such as boundary potentials or external fields and interaction potentials involving two, three or four particles. As in MD force fields, bonded potentials are defined within particle groups called "topologies" whose bonding structure is defined by a connectivity graph. **(b)** Reactions: Most reactions are unimolecular or bimolecular particle reactions. Topology reactions act on the connectivity graphs and particle types and therefore change the particle bonding structure. **(c)** Simulation box: The simulation box with edge lengths ℓ_x, ℓ_y, and ℓ_z. It can optionally be periodic in a combination of x, y, and z directions, applying the minimum image convention.

Where $\tau > 0$ is a finite time step size and $\boldsymbol{\eta}_t \sim \mathcal{N}(0,1)$ is a normally-distributed random variable. The diffusion constant D_i effects the magnitude of the random displacement. The particles' positions are loosely bound to a cuboid simulation box with edge lengths ℓ_x, ℓ_y, ℓ_z (Fig. 4.1). If a boundary is non-periodic it is equipped with a repulsive wall given by the potential

$$V_{\text{wall}}(\mathbf{x}_i^{(j)}) = \frac{1}{2} k d(\mathbf{x}_i^{(j)}, W_j)^2 \tag{4.4}$$

acting on every component j of the single particle position $\mathbf{x}_i$, where k is the force constant, $W = \prod_{j=1}^{3} W_j = \prod_{j=1}^{3} [x^{(j)}_{\text{origin}}, x^{(j)}_{\text{origin}} + x^{(j)}_{\text{extent}}]$ the cuboid in which there is no repulsion contribution of the potential, and $d(\cdot, W_j) := \inf\{d(\cdot, w) : w \in W_j\}$ the shortest distance to the set W_j. The cuboid can be larger than the simulation box in the periodic directions. In non-periodic directions there must be at least one repulsive wall for which this is not the case.

Due to the soft nature of the walls particles still can leave the simulation box in non-periodic directions. In that case they are no longer subject to pairwise interactions and bimolecular reactions however still are subject to the force of the wall pulling them back into the box.

Other types of dynamical models and other integration schemes can be implemented in ReaDDy 2 via its C++ interface. For example, non-overdamped dynamics, anisotropic diffusion [Vij+17; Mun+09], hydrodynamic interactions [EM78] or employing the MD-GFRD scheme to make large steps for noninteracting particles will all affect the dynamical model and can be realized by writing suitable plugins.

4.2.2 Potentials

The deterministic forces are given by the gradient of a many-body potential energy U (Fig. 4.1a):

$$\mathbf{f}_i = \nabla_i \left(\sum_i U_{\text{ext}}(\mathbf{x}_i) + \sum_{i \neq j} U_{\text{pair}}(\mathbf{x}_i, \mathbf{x}_j) + \sum_{i \neq j \neq k} U_{\text{triple}}(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k) + \dots \right)$$

The potentials are defined by the user. ReaDDy 2 provides a selection of standard potential terms, additional custom potentials can be defined via the C++ interface and then included into a Python simulation script.

External potentials only depend on the absolute position of each particle. They can be used, e.g., to form softly repulsive walls Eq. (4.4) and spheres, or to attach particles to a surface, for example to model membrane proteins. Furthermore the standard potential terms enable the user to simulate particles inside spheres and exclude particles from a spherical volume. The mentioned potential terms can also be combined to achieve more complex geometrical structures. Pair potentials generally depend on the particle distance and can be used, e.g., to model space exclusion at short distances.

A fundamental restriction of ReaDDy 2 interaction potentials is that they have a finite range and can therefore be cut off. This means that, e.g., full electrostatics is not supported but screened electrostatic interactions are implemented (Section 4.4.5). Additionally a harmonic repulsion potential, a weak interaction potential made out of three harmonic terms, and Lennard–Jones interaction are incorporated.

ReaDDy 2 has a special way of treating interaction potentials between bonded particles. Topologies define graphs of particles that are bonded and imply which particle pairs interact via bond constraints, which triples interact via angle constraints, and which quadruplets interact via a torsions potential. See Section 4.2.4 for details.

4.2.3 Reactions

Reactions are discrete events, that can change particle types, add, and remove particles (Fig. 4.1b). Each reaction is associated with a microscopic rate constant $\lambda > 0$ which has units of inverse time and

represents the probability per unit time of the reaction occurring. The integration time-steps used in ReaDDy 2 should be significantly smaller than the inverse of the largest reaction rate, and we therefore compute discrete reaction probabilities by:

$$p(\lambda;\tau) = 1 - e^{-\lambda\tau}. \tag{4.5}$$

In the software it is checked whether the time step τ is smaller than the inverse reaction rate up to a threshold factor of 10, otherwise a warning is displayed as discretization errors might become too large. In general, ReaDDy 2 reactions involve either one or two reactants. At any time step, a particle that is subject to an unary reaction will react with probability $p(\lambda;\tau)$. If there are two products, they are placed within a sphere of specified radius R_u around the educt's position $\mathbf{x}_0$. This is achieved by randomly selecting an orientation $\mathbf{n} \in \mathbb{R}^3$, distance $d \leqslant R_\mathrm{u}$, and weights $w_1 \geqslant 0, w_2 \geqslant 0$, s.t. $w_1 + w_2 = 1$. The products are placed at $\mathbf{x}_1 = \mathbf{x}_0 + d w_1 \mathbf{n}$ and $\mathbf{x}_2 = \mathbf{x}_0 - d w_2 \mathbf{n}$. Per default, $w_1 = w_2 = 0.5$ and the distances d are drawn such that the distribution is uniform with respect to the volume of the sphere. When it is necessary to produce new particles, we suggest to define a producing particle A and use the unary reaction $\mathrm{A} \rightharpoonup \mathrm{A} + \mathrm{B}$ with corresponding placement weights $w_1 = 0$, $w_2 = 1$ so that the A particle stays at its position.

The basic binary reaction scheme is the Doi scheme [Doi76; TS67] in which a reactive complex is defined by two reactive particles being in a distance of R_b or less, where R_b is a parameter, e.g., see Fig. 4.1b Fusion or Enzymatic reaction. The reactive complex then forms with probability $p(\lambda;\tau)$ while the particles are within distance.

Optionally ReaDDy 2 can simulate reversible reactions using the reversible iPRD-DB scheme developed in [FN18]. This scheme employs a Metropolis-Hastings algorithm that ensures the reversible reaction steps to be made according to thermodynamic equilibrium by accounting for the system's energy in the educt and product states.

4.2.4 Topologies

Topologies are a way to group particles into superstructures. For example, large-scale molecules can be represented by a set of particles corresponding to molecular domains assembled into a topology. A topology also has a set of potential energy terms such as bond, angle, and torsion terms associated. The specific potential terms are implied by finding all paths of length two, three, and four in the topology connectivity graph. The sequence of particle types associated to these paths then is used to gather the potential term specifics, e.g., force constant, equilibrium length or angle, from a lookup table (Fig. 4.1a).

Reactions are not only possible between particles, but also between a topology and a particle (Fig. 4.1b) or two topologies. In order to define such reactions, one can register topology types and then specify the consequences of the reaction on the topology's connectivity graph. We distinguish between global and local topology reactions.

Global topology reactions are triggered analogously to unary reactions, i.e., they can occur at any time with a fixed rate and probability as given in Eq. (4.5). Any edge in the graph can be removed and added. Moreover, any particle type as well as the topology type can be changed, which may result in significant changes in the potential energy. If the reaction causes the graph to split into two or more components, these components are subsequently treated as separate topologies that inherit the educt's topology type and therefore also the topology reactions associated with it. Such a reaction is the topology analogue of a particle fission reaction.

A local topology reaction is triggered analogously to binary reactions with probability $p(\lambda;\tau)$ if the distance between two particles is smaller than the reaction radius. At least one of the two particles needs to be part of a topology with a specific type. The product of the reaction is then either yielded by the formation of an edge and/or a change of particle and topology types. In contrast to global reactions only certain changes to particle types and graphs can occur:

- Two topologies can fuse, i.e., an additional edge is introduced between the vertices corresponding to the two particles that triggered the reaction.
- A topology and a free particle can fuse by formation of an edge between the vertex of the topology's particle and a newly introduced vertex for the free particle.
- Two topologies can react in an enzymatic fashion, i.e., particle types of the triggering particles and topology types can be changed.
- Two topologies and a free particle can react in an enzymatic fashion analogously.

In all of these cases the involved triggering particles' types and topology types can be changed.

4.2.5 Simulation setup and boundary conditions

Once the potentials, the reactions (Fig. 4.1a,b), and a temperature T have been defined, a corresponding simulation can be set up. A simulation box can be periodic or partially periodic, see Fig. 4.1c. Periodicity in a certain direction means that with respect to that direction particle wrapping and the minimum image convention are applied. Non-periodic directions require a harmonically repelling wall as given in Eq. (4.4).

In order to define the initial condition, particles and particle complexes are added explicitly by specifying their 3D position and type. A simulation can now be started by providing a time step size τ and a number of integration steps.

4.3 Design and Implementation

ReaDDy 2 is mainly written in C++ and has Python bindings making usage, configuration, and extension easy while still being able to provide high performance. To encourage usage and extension of the software, it is Open Source and licensed under the BSD-3 license. It therefore can not only be used in other Open Source projects without them requiring to have a similar license, but also in a commercial context.

4.3.1 Design

The software consists of three parts. The user-visible toplevel part is the python user interface, see Fig. 4.2a. It is a language binding of the C++ user interface (Fig. 4.2b) and has additional convenience functionality. The workflow consists out of three steps:

1. The user is creating a *readdy.ReactionDiffusionSystem*, including information about temperature, simulation box size, periodicity, particle species, reactions, topologies, and physical units. Per default the configurational parameters are interpreted in a unit set well suited for cytosolic environments (lengths in nm, time in ns, and energy in kJ/mol), e.g., particles representing proteins in solution. The initial condition, i.e., the positions of particles, is not yet specified.
2. The system can generate one or many instances of *readdy.Simulation*, in which particles and particle complexes can be added at certain positions. When instantiating the simulation object, a compute kernel needs to be selected, in order to specify how the simulation will be run (e.g., single-core or multi-core implementation). Additionally, observables to be monitored during the simulation are registered, e.g., particle positions, forces, or the total energy. A simulation is started by entering a time step size $\tau > 0$ in units of time and a number of integration steps that the system should be propagated.

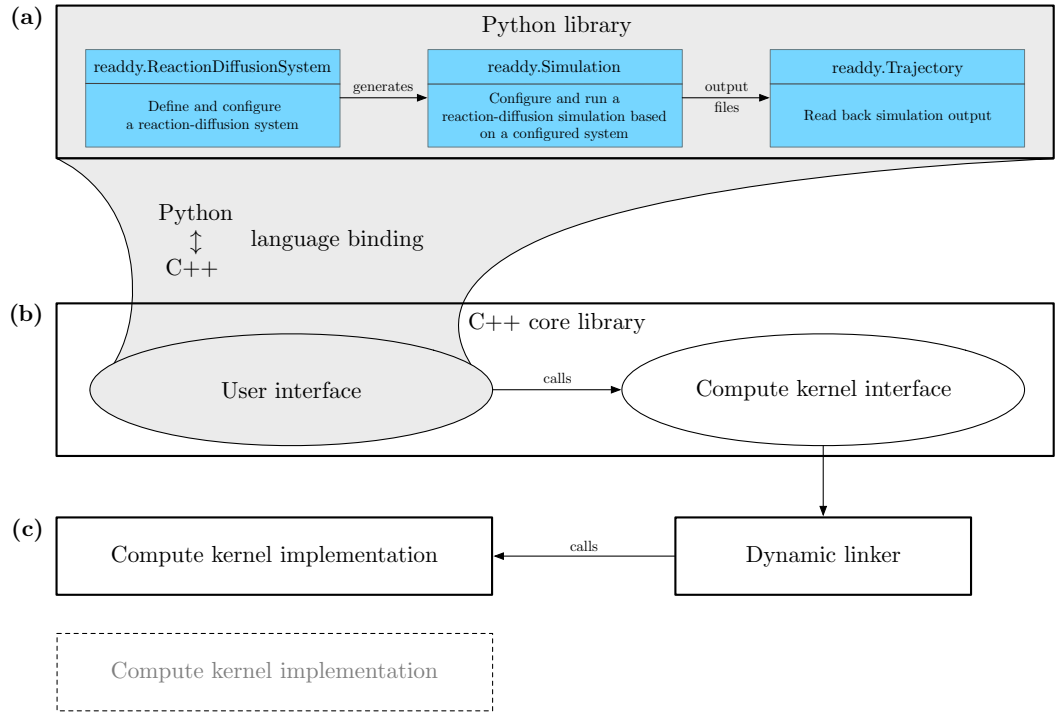

Figure 4.2: **The software structure.** **(a)** Python user interface: Provides a Python binding to the "C++ user interface" with some additional convenience functionality. The user creates a "readdy.ReactionDiffusionSystem" and defines particle species, reactions, and potentials. From a configured system, a "readdy.Simulation" object is generated, which can be used to run a simulation of the system given an initial placement of particles. **(b)** C++ core library: The core library serves as an adapter between the actual implementation of the algorithms in a compute kernel and the user interface. **(c)** Compute kernel implementation: Implements the compute kernel interface and contains the core simulation algorithms. Different compute kernel implementations support different hard- or software environments, such as serial and parallel CPU implementations. The compute kernel is chosen when the "readdy.Simulation" object is generated and then linked dynamically in order to provide optimal implementations for different computing environments under the same user interface.

3. When a simulation has been performed, the observables' outputs have been recorded into a file. The file's contents can be loaded again into a *readdy.Trajectory* object that can be used to produce trajectories compatible with the VMD molecular viewer [HDS96].

Running a simulation based on the *readdy.Simulation* object invokes a simulation loop. The default simulation loop is given in Alg. 2. Individual steps of the loop can be omitted. This enables the user to, e.g., perform pure PBRD simulations by skipping the calculation of forces. Performing a step in the algorithm leads to a call to the compute kernel interface, see Fig. 4.2b. Depending on the selected compute kernel the call is then dispatched to the actual implementation. Compute kernel implementations (Fig. 4.2c) are dynamically loaded at runtime from a plugin directory. This modularity allows ReaDDy 2 to run across many platforms although not every computing kernel may run on a given platform, such as a CUDA-enabled computing kernel. ReaDDy version 2 includes two iPRD computing kernels: a single threaded default computing kernel, and a dynamically-loaded shared-memory parallel kernel.

The computing kernels contain implementations for the single steps of the simulation loop. Currently,

Algorithm 2: ReaDDy 2 default simulation loop. Each of the calls are dispatched to the compute kernel, see Fig. 4.2. Furthermore, the user can decide to switch off certain calls in the simulation loop while configuring the simulation.

```
Initialize compute kernel;
if has output file then
    Write simulation setup;
Set up neighbor list;
Compute forces;
Evaluate observables;
while continue simulating do
    Call integrator;
    Update neighbor list;
    Perform reactions;
    Perform topology reactions;
    Update neighbor list;
    Calculate forces;
    Evaluate observables;
Tear down compute kernel;
```

integrator and reaction handler are exchangeable by user-written C++ extensions. Hence, there is flexibility considering what is actually performed during one step of the algorithm or even what kind of underlying model is applied.

In comparison to the predecessor ReaDDy 1, the software is a complete rewrite and extension. The functionality of the Brownian dynamics integrator has been preserved, however the reaction handlers can behave slightly differently. In particular, if during an integration step a reaction conflict occurs, i.e., there are at least two reaction events which involve the same educt particles, only one of these events can be processed. One possibility of choosing the to-be processed event is the so-called "UncontrolledApproximation", which draws the next reaction event uniformly from all events and prunes conflicting events. Another possibility is drawing the next reaction event from all events weighted by their respective reaction probability. Since this approach is loosely based on the reaction order in the Gillespie SSA, this reaction handler is named "Gillespie" in ReaDDy 2.

With respect to the microscopic evaluation of a reaction event, the ReaDDy 1 implementation places product particles of fission type reactions at a fixed distance, which is handled more flexibly in the current implementation, see Section 4.2.3.

4.3.2 Performance

To benchmark ReaDDy 2, we use a reactive system with three particle species A, B, and C introduced in [SN13] with periodic boundaries instead of softly repelling ones. The simulation temperature is set to $T = 293\,\mathrm{K}$ and the diffusion coefficients are given by $D_\mathrm{A} = 143.1\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, $D_\mathrm{B} = 71.6\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$, and $D_\mathrm{C} = 68.82\,\mu\mathrm{m}^2\,\mathrm{s}^{-1}$. Particles of these types are subject to the two reactions $\mathrm{A} + \mathrm{B} \longrightarrow \mathrm{C}$ with microscopic association rate constant $\lambda_\mathrm{on} = 10^{-3}\,\mathrm{ns}^{-1}$ and reaction radius $R_1 = 4.5\,\mathrm{nm}$, and $\mathrm{C} \longrightarrow \mathrm{A} + \mathrm{B}$ with microscopic dissociation rate constant $\lambda_\mathrm{off} = 5 \times 10^{-5}\,\mathrm{ns}^{-1}$ and dissociation radius $R_2 = R_1$. Particles

are subject to an harmonic repulsion interaction potential which reads

$$U(r) = \begin{cases} \frac{\kappa}{2}(r-\sigma)^2 & \text{, for } r \leqslant \sigma, \\ 0 & \text{, otherwise,} \end{cases} \tag{4.6}$$

where σ is the distance at which particles start to interact and $\kappa = 10\,\mathrm{kJ\,mol^{-1}\,nm^{-2}}$ is the force constant. The interaction distance σ is defined as sum of radii associated to the particles' types, in this case $r_A = 1.5\,\mathrm{nm}$, $r_B = 3\,\mathrm{nm}$, and $r_C = 3.12\,\mathrm{nm}$. All particles are contained in a cubic box with periodic boundaries. The edge length is chosen such that the initial number density of all particles is $\varrho_{tot} = 3141\,\mathrm{nm^{-3}}$. This total density is distributed over the species, such that the initial density of A is $\varrho_A = \varrho_{tot}/4$, the initial density of B is $\varrho_B = \varrho_{tot}/4$, and the initial density of C is $\varrho_C = \varrho_{tot}/2$. For the chosen microscopic rates these densities roughly resemble the steady-state of the system. The performance is measured over a simulation timespan of 300 ns which is much shorter than the equilibration time of this system. Thus the overall number of particles does not vary significantly during measurement and we obtain the computation time at constant density.

In the following the benchmark results are presented. A comparison between the sequential reference compute kernel, the parallel implementation, and the previous Java-based ReaDDy 1 [SN13] is made with respect to their performance when varying the number of particles in the system keeping the density constant. Since the particle numbers fluctuate the comparison is based on the average computation time per particle and per integration step (Fig. 4.3). The sequential kernel scales linearly with the number of particles, whereas the parallelized implementation comes with an overhead that depends on the number of threads. The previous Java-based implementation does not scale linearly for large particle numbers, probably owing to Java's garbage collection. The parallel implementation starts to be more efficient than the sequential kernel given sufficiently many particles.

Fig. 4.4 shows the strong scaling behavior of the parallel kernel, i.e. the speedup and efficiency for a fixed number of particles as a function of the used number of threads. For sufficiently large particle numbers, the kernel scales linear with the number of physical cores and an efficiency of around 80%. In hyperthreading mode, it then continues to scale linear with the number of virtual cores with an efficiency of about 55–60%.

The number of steps per day for a selection of particle numbers and kernel implementation is displayed in Table 4.1. For a system with 13,000 particles and a time step size of $\tau = 1\,\mathrm{ns}$ (e.g., membrane proteins [SN13]), a total of 17 ms simulation time per day can be collected on a six-core machine (Fig. 4.3 for details). The current ReaDDy kernels are thus suited for the detailed simulation of processes in the millisecond- to second timescale, which include many processes in sensory signalling and signal transduction at cellular membranes.

Table 4.1: **Number of steps per day for the benchmark system.**

Approximate number of particles	Steps per day sequential kernel	Peak performance steps per day parallel kernel	Number of threads
250	2.8×10^8	2.6×10^8	4
1000	7.9×10^7	1.2×10^8	7
13000	5.6×10^6	1.7×10^7	11
40000	1.8×10^6	6.3×10^6	11

Number of time steps per day for benchmark system of Section 4.3.2 using the machine described in Fig. 4.3. In case of the parallelized implementation the peak performance with respect to the number of threads is shown.

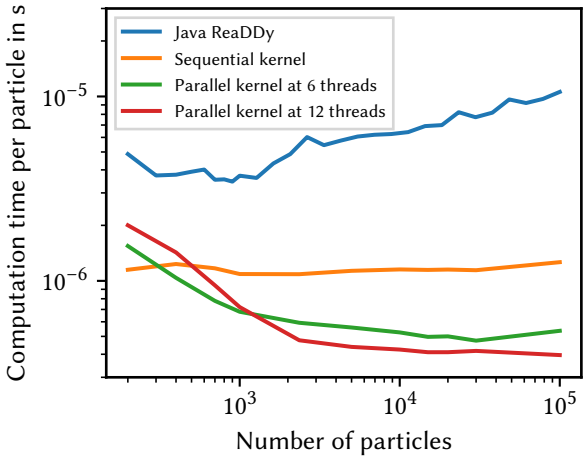

Figure 4.3: **Performance comparison.** Average computation time per particle and integration step for the benchmark system of Section 4.3.2 using a machine with an Intel Core i7 6850K processor, i.e., six physical cores at 3.8 GHz, and 32 GB DDR4 RAM at 2.4 GHz (dual channel). The number of particles is varied, but the particle density is kept constant. The sequential kernel (orange) has a constant per-particle CPU cost independent of the particle number. For large particle numbers, the parallel kernels are a certain factor faster (see scaling plot Fig. 4.4). For small particle numbers of a few hundred the sequential kernel is more efficient. ReaDDy 2 is significantly faster and scales much better than the previous Java-based ReaDDy 1 [SN13].

4.4 Results

In the following, several aspects of the model applied in ReaDDy 2 are validated and demonstrated by considering different application scenarios and comparing the results to analytically obtained results, simulations from other packages, or literature data.

4.4.1 Reaction kinetics and detailed balance

We simulate the time evolution of particle concentrations of the benchmark system described in Section 4.3.2. In contrast to the benchmarks, the considered system initially only contains A and B particles at equal numbers. It then relaxes to its equilibrium mixture of A, B, and C particles (Fig. 4.5). Since the number of A and B molecules remain equal by construction, only the concentrations of A and C are shown.

In addition we compare the solutions with and without harmonic repulsion potentials Eq. (4.6) between all particles, as well as two different methods for executing the reactions: The Doi reaction scheme as described in Section 4.2.3 and the detailed-balance reaction scheme iPRD–DB described in [FN18].

In contrast to Section 4.3.2, we construct a macroscopic reference system with rate constants $k_{\text{on}} = 3.82 \times 10^{-1}\,\text{nm}^3\text{s}^{-1}$ and $k_{\text{off}} = 5 \times 10^{-5}\,\text{s}^{-1}$ resembling a cellular system. The microscopic reaction rate constants λ_{on} and λ_{off} are then chosen with respect to the reference system taking interaction potentials

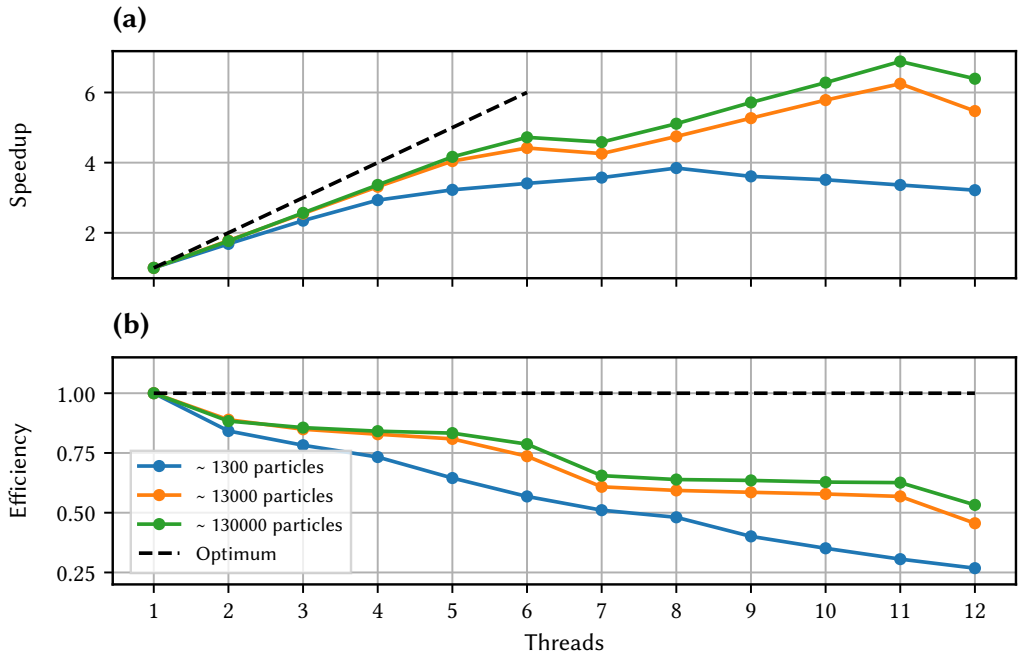

Figure 4.4: **Speedup and efficiency.** Parallel speedup and efficiency of the benchmark system of Section 4.3.2 as a function of the number of cores using the machine described in Fig. 4.3. **(a)** Speedup with different numbers of cores compared to one core. Optimally one would like to have a speedup that behaves like the identity (black dashed line). **(b)** Efficiency is the speedup divided by the number of threads, i.e., how efficiently the available cores were used.

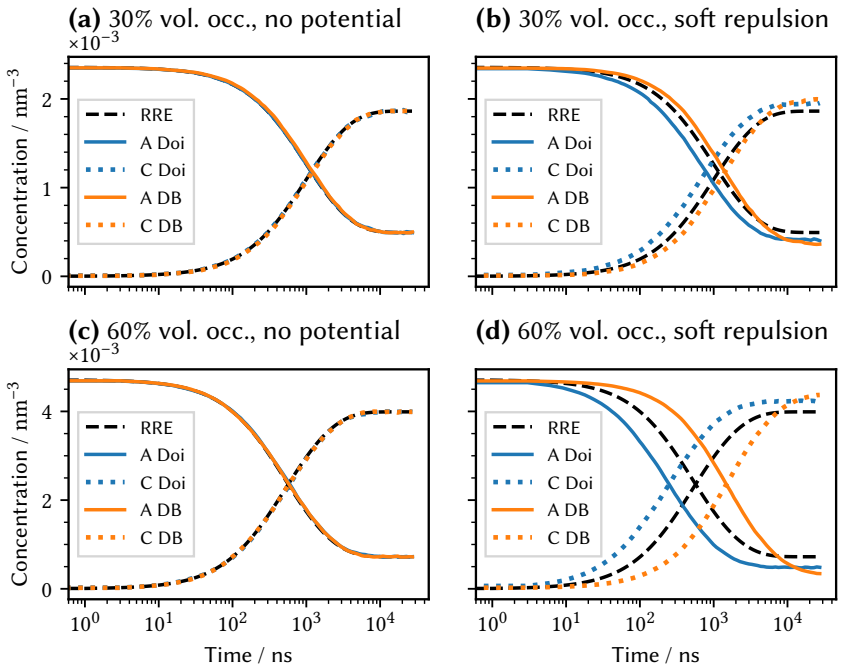

Figure 4.5: **Reaction kinetics and detailed balance.** Concentration time series of a the reaction-diffusion system introduced in Section 4.3.2 with the reversible reaction A + B $\leftrightharpoons$ C. Compared are cases with and without harmonic repulsion Eq. (4.6). Additionally we compare two different reaction mechanisms, the Doi reaction scheme and the detailed balance (iPRD-DB) method for reversible reactions. **(a)** 30% volume occupation and no interaction potentials. **(b)** 30% volume occupation with harmonic repulsion between all particles. **(c)** 60% volume occupation and no interaction potentials. **(d)** 60% volume occupation with harmonic repulsion between all particles.

between A and B into account. In particular,

$$\lambda_{\text{off}} = k_{\text{off}}, \tag{4.7}$$

$$\lambda_{\text{on}} = \frac{k_{\text{on}}}{V_{\text{eff}}}, \tag{4.8}$$

where $V_{\text{eff}} = \int_0^R \exp(-\beta U) 4\pi r^2 \mathrm{d}r$ is the accessible reaction volume, R the reaction radius, β the inverse thermal energy, and U the pair potential. The harmonic repulsion potential reduces V_{eff} with respect to the volume of the reactive sphere. The expression Eq. (4.8) originates from an approximation for k_{on} in a sufficiently well-mixed (i.e., reaction–limited) and sufficiently diluted system. The derivation can be found in [FN18] based on calculating the total association rate constant k_{on} for an isolated pair of A and B particles. In this case one obtains $\lambda_{\text{off}} = 5 \times 10^{-5}\,\text{ns}^{-1}$ for the microsopic dissociation rate constant. The microscopic association rate constant reads $\lambda_{\text{on}} = 10^{-3}\,\text{ns}^{-1}$ for the noninteracting system and $\lambda_{\text{on}} = 2.89 \times 10^{-3}\,\text{ns}^{-1}$ for the interacting system. Note that for non-reversible binary reactions without interaction potentials the formula provided by [Doi75; EC09] describes the relation between λ and k for slow diffusion encounter. In the case of non-reversible binary reactions with interaction potentials and slow diffusion encounter such a relation can still be numerically computed [Dib+19].

Using the macroscopic rate constants k_{on} and k_{off}, a solution can be calculated for the mass-action reaction rate equations (RRE). This solution serves as a reference for the noninteracting system (no potentials), because the system parameters put the reaction kinetics in the mass-action limit.

In the noninteracting system, the ReaDDy solution and the RRE solution indeed agree (Fig. 4.5a,c). In the case of interacting particles, see Fig. 4.5b, d, an exact reference is unknown. We observe deviations from the RRE solution that become more pronounced with increasing particle densities. A difference between the two reaction schemes can also be seen. The Doi reaction scheme shows faster equilibration compared to RRE for increasing density, whereas the iPRD-DB scheme shows slower equilibration, as it has a chance to reject individual reaction events based on the change in potential energy. Thus an increased density leads to more rejected events, consistent with the physical intuition that equilibration in a dense system should be slowed down. Furthermore the equilibrated states differ depending on the reaction scheme, showing a dependence on the particle density. For denser systems the iPRD-DB scheme favors fewer A and B particles than the Doi scheme, consistent with the density-dependent equilibria described in [FN18].

4.4.2 Diffusion

Next we simulate and validate the diffusive behavior of noninteracting particle systems and the subdiffusive behavior of dense interacting particle systems. The simulation box contains particles with diffusion coefficient D_0 and is equipped with softly repelling walls, in order to introduce finite size effects. The observations are carried out with and without interaction potential. In the case without interaction potential we compare with an analytical solution and the case of an interaction potential is compared to the literature.

Length x is given in units of σ, time t is given in units of σ^2/D_0, and energy is given in units of $k_B T$. The cubic box has an edge length of $\ell \approx 28\sigma$.

The noninteracting particle simulation has a mean-squared displacement of particles in agreement with the analytic solution given by Fick's law for diffusion in three dimensions

$$\langle (\mathbf{x}_t - \mathbf{x}_0)^2 \rangle = 6 D_0 t, \tag{4.9}$$

where $\mathbf{x}$ is the position of a particle and t is time (Fig. 4.6). For long timescales $t \geqslant 10^1$ transport is obstructed by walls, resulting in finite size saturation.

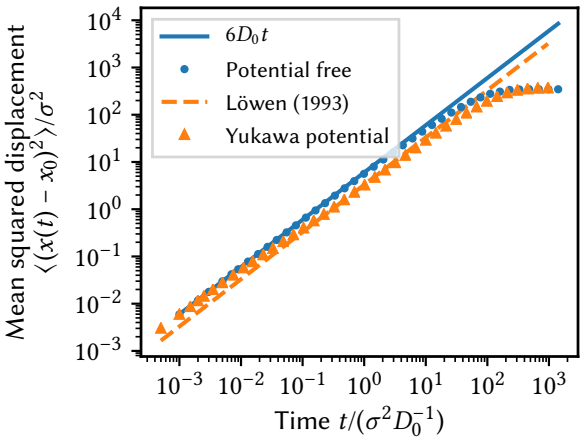

Figure 4.6: **Diffusion in crowded environments.** Mean squared displacement as a function of time. Multiple particles are diffusing with intrinsic diffusion coefficient D_0 in a cubic box with harmonically repelling walls. Triangles were obtained by using the Yukawa repulsion potential Eq. (4.10) between all particles. The dashed line represents an effective diffusion coefficient from the literature [LS93] for the same Yukawa repulsion potential.

Fig. 4.6 also shows that more complex transport can be modeled, as, e.g., found in crowded systems. Particles interact via the Yukawa potential [Yuk35]

$$U(r) = \begin{cases} U_0 \sigma \exp\left(-\lambda \frac{r-\sigma}{\sigma}\right)/r & \text{, for } r \leqslant r_c, \\ 0 & \text{, otherwise,} \end{cases} \tag{4.10}$$

where $U_0 = k_B T$ is a repulsion energy, σ is the length scale, $\lambda = 8$ is the screening parameter, and $r_c = 2.5\sigma$ the cutoff radius.

The particle density is $n\sigma^3 = 0.6$ with n being the number density. In such a particle system, the mean-squared displacement differs significantly from the analytical result for free diffusion after an initial time $t \geq 10^{-2}$ in which particles travel their mean free path length with diffusion constant D_0. At intermediate timescales $t \in [10^{-2}, 10^{-1})$, particle transport is subdiffusive due to crowding. At long timescales, $t \in [10^{-1}, 10^{1})$, the particles are again diffusive with an effective diffusion coefficient D that is reduced to reflect the effective mobility in the crowded systems. We compare this to an effective diffusion coefficient obtained by Brownian dynamics simulations from Löwen and Szamel [LS93] and find that they qualitatively agree. For large timescales $t \geqslant 10^1$ finite size saturation can explicitly be observed as almost every particle has been repelled at least once by the boundaries.

To quantitatively compare the long-time effective diffusion coefficient D, we set up 1100 particles in a periodic box without repelling walls with the edge length chosen to give the densired density $n\sigma^3 = 0.6$. The cutoff of the potential Eq. (4.10) is set to $r_c = 5\sigma$, where $U(r_c) < 10^{-14} k_B T$. The particle suspension is equilibrated for at least $t_{\text{eq}} \geqslant 3$ with a time-step size of $\tau = 10^{-5}$. We observe the mean squared displacement until $t_{\text{obs}} = 4.5$ and measure the diffusion coefficient as the slope of a linear function for $t \in [4, 4.5)$. We obtain $D/D_0 = 0.54 \pm 0.01$, which agrees with the reference value [LS93] $D^{\text{ref}}/D_0 = 0.55 \pm 0.01$.

4.4.3 Thermodynamic equilibrium properties

We validate that ReaDDy 2's integration of equations of motion yields the correct thermodynamics of a Lennard-Jones colloidal fluid in an (N, V, T) ensemble. To this end, we simulate a system of N particles confined to a periodic box with volume V at temperature T. The results and comparisons with other simulation frameworks and analytical results are shown in Table 4.2. The particles interact via the Lennard-Jones potential

$$U(r) = 4\varepsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^{6} \right],$$

with ε being the depth of the potential well and σ the diameter of particles. The potential is cut off at $r_C = 4\sigma$ and shifted to avoid a discontinuity. The rescaled temperature is $T^* = k_B T \varepsilon^{-1} = 3$. We perform simulations of the equilibrated Lennard-Jones system for 10^6 integration steps with rescaled time step size $\tau^* = 10^{-4}$. Time units are σ^2/D and are determined by the self-diffusion coefficient D of the particles. We measure the rescaled pressure $P^* = P\sigma^3\varepsilon^{-1}$ by estimating the virial term from forces acting in the system as described in [AT87]. Additionally we measure the rescaled potential energy per particle $u^* = UN^{-1}\varepsilon^{-1}$. Both pressure and potential energy are calculated every 100th time step. This sampling gives rise to the mean and its error of the mean given for the ReaDDy 2 results in Table 4.2. Comparing HALMD [CH11] and ReaDDy 2, the latter shows larger energy and pressure in the third decimal place for the lower density $\varrho^* = 0.3$. For the higher density $\varrho^* = 0.6$ pressure differs in the first decimal place and energy in the second. This can be explained by ReaDDy 2 using an Euler scheme Eq. (4.3) to integrate motion of particles, which has a discretization error of first order in the time step size $\mathcal{O}(\tau)$. On the other hand HALMD uses a Velocity-Verlet method [Swo+82], which has a discretization error of second order in the time step size $\mathcal{O}(\tau^2)$.

Table 4.2: **Thermodynamic equilibrium properties of a Lennard–Jones colloidal fluid in a (N, V, T) ensemble.**

	density ϱ^*	pressure P^*	energy u^*
ReaDDy 2	0.3	1.0253 ± 0.0004	−1.6704 ± 0.0003
HALMD [CH11]	0.3	1.0234 ± 0.0003	−1.6731 ± 0.0004
Johnson et al. [JZG93]	0.3	1.023 ± 0.002	−1.673 ± 0.002
Ayadim et al. [AOA09]	0.3	1.0245	−1.6717
ReaDDy 2	0.6	3.711 ± 0.002	−3.2043 ± 0.0004
HALMD [CH11]	0.6	3.6976 ± 0.0008	−3.2121 ± 0.0002
Johnson et al. [JZG93]	0.6	3.69 ± 0.01	−3.212 ± 0.003
Ayadim et al. [AOA09]	0.6	3.7165	−3.2065

Results of the ReaDDy 2 framework are compared to other simulation frameworks and analytical results for validation.

4.4.4 Topology reactions

We illustrate ReaDDy 2's ability to model complex reactions between multi-particle complexes, called "topology reactions". We model polymers as linear chains of beads, held together by harmonic bonds and stiffened by harmonic angle potentials.

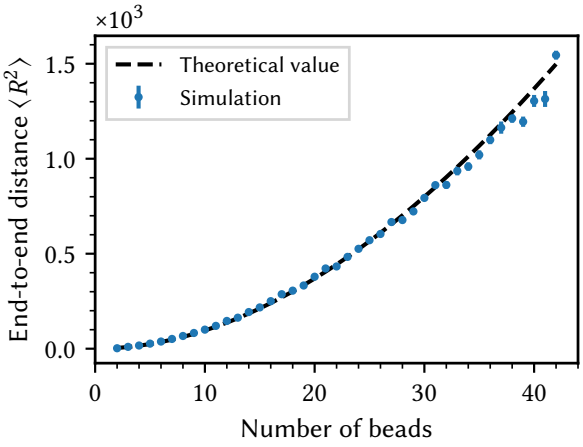

Figure 4.7: **Mean-squared end-to-end distance of worm-like chains.** The theoretical mean-squared end-to-end distance of worm-like chains as a function of number of beads Eq. (4.11) is compared to simulation data obtained from linear chains of beads as described in Section 4.2.4. Error bars depict errors over the mean from multiple measurements.

When considering just one worm-like chain with a certain amount of beads n, its equilibrium mean-squared end-to-end distance should behave like [RC03]

$$\langle R^2 \rangle = 2 l_p R_{\max} - 2 l_p^2 \left(1 - \exp\left(-\frac{R_{\max}}{l_p} \right) \right), \tag{4.11}$$

where $l_p = 4lk(k_B T)^{-1}$ is the persistence length, $R_{\max} = (n-1)l$ the chain contour length, l the bond length, and k the force constant of the harmonic angles. In order to verify that the considered chain model obeys the mechanics of a worm-like chain, the theoretical mean-squared end-to-end distance Eq. (4.11) can be compared to observations from simulations, see Fig. 4.7. For each fixed number of beads, an isolated chain was relaxed into an equilibrium state without performing topology reactions, yielding a squared end-to-end distance at the end of the simulation. This experiment was repeated 51 times. From the figure it can be observed that there is good agreement between the theoretical and measured mean-squared end-to-end distances.

In a system with many of these chains, we introduce two different particle types for the beads. Either they are head particles and located at the ends of a polymer chain or they are core particles and located between the head particles, as shown in Fig. 4.8a, c in blue and orange, respectively.

We impose two different topology reactions in the system with many chains (Fig. 4.8a):

1. Association: Two nearby head particles (distance $\leq R$) can connect with rate λ_1. The topology is changed by adding an edge between the connected particles, resulting in the addition of one bond and two angle potentials. Additionally, the particle types of the two connected particles change from “head” to “core”.

2. Dissociation: A chain with n particles can dissociate with microscopic rate $n\lambda_2$, such that longer chains have a higher probability to dissociate than shorter chains. When a dissociation occurs, a random edge between two core particles is removed. The particle types of the respective core particles are changed to “head”. As a result, the graph decays into two connected components which subsequently are treated as autonomous topology instances.

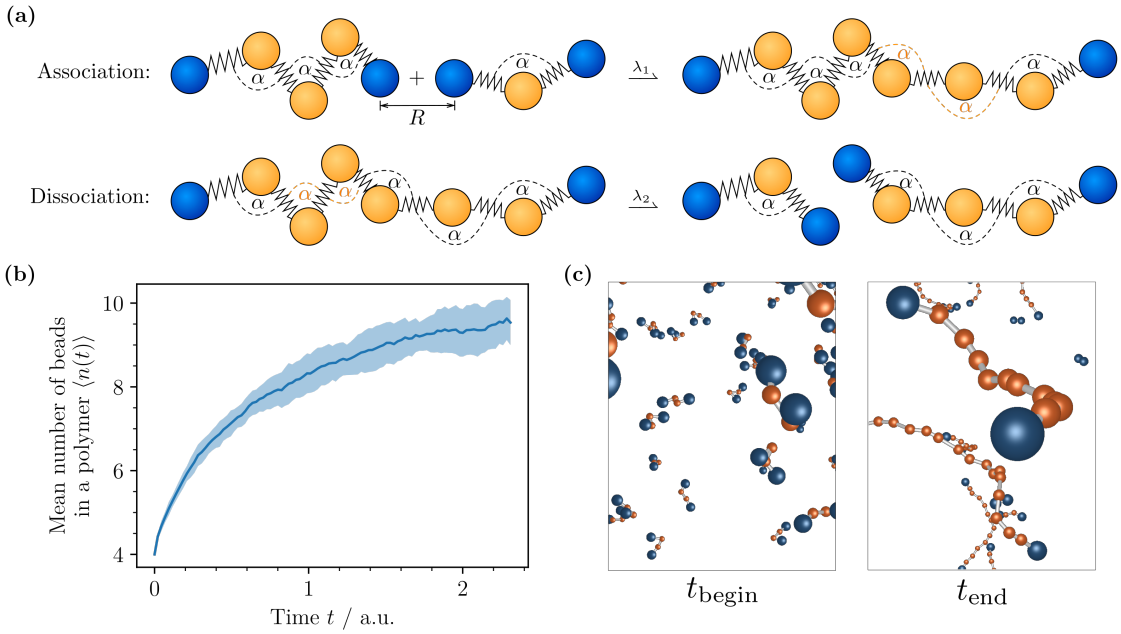

Figure 4.8: **Topology reactions example.** Illustrative simulation of polymer assembly/disassembly using topology reactions. **(a)** Sketch of the involved topology reactions. *Association*: When two ends of different topologies come closer than R, there is a rate λ_1 that an edge is formed. *Dissociation*: The inverse of association with a rate λ_2 and a randomly drawn edge that is removed. **(b)** The number of beads in a polymer $\langle n(t) \rangle$ over time averaged over 15 realizations. **(c)** Two representative particle configurations showing the initial state and the end state at time t_{begin} and t_{end}, respectively.

The temporal evolution of the average length of polymer chains is depicted in Fig. 4.8b. The simulation was performed 15 times with an initial configuration of 500 polymers containing four beads each. After sufficient time $\langle n(t) \rangle$ reaches an equilibrium value. Over the course of the simulation the polymers diffuse and form longer polymers. This can also be observed from the two snapshots shown in Fig. 4.8c, depicting a representative initial configuration at t_{begin} and a representative configuration at the end of the simulation at time. In that case, there are polymers of many different lengths.

4.4.5 Nontrivial bimolecular association kinetics at high concentrations

This section studies a biologically inspired system with three macromolecules A, B, and C, that resemble, e.g., proteins in cytosol. The macromolecules A and B can form complexes C that also can dissociate back into their original components, i.e., we introduce reactions

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{C}. \tag{4.12}$$

This form of interaction has been studied for proteins bovine serum albumin and hen egg white lysozyme in coarse-grained atomistic detail in [MMW14] and for barnase and barstar in [Pla+17]. Here, we consider the case where the association reaction of Eq. (4.12) does not preserve volume, i.e., the complex C is more compact.

The presence of ions in aqueous solutions has effects on protein interactions [RAL78], therefore we assume the reversibly associating macromolecules to be weakly charged and thus subject to the Debye-Hückel interaction potential [DH23] including an additional repulsion term

$$U_{s_1 s_2}(r) = q_{s_1} q_{s_2} \frac{e^2}{\varepsilon_0 \varepsilon_r} \frac{\exp(-\kappa r)}{r} + U_r \left(\frac{\sigma_{s_1 s_2}}{r} \right)^{12}, \tag{4.13}$$

where $s_1, s_2 \in \{\mathrm{A}, \mathrm{B}, \mathrm{C}\}$, q are partial charges associated with the macromolecules, e is the elementary charge, ε_0 is the vacuum permittivity, ε_r is the relative permittivity of an aqueous solution, κ is the screening parameter that describes shielding due to ions in the solution, U_r is the repulsion energy, and $\sigma_{s_1 s_2} = r_{s_1} + r_{s_2}$ is the sum of two particle radii. Here, we do not take hydration effects into account.

We investigate the equilibrium constant $K = [\mathrm{A}][\mathrm{B}]/[\mathrm{C}]$ for different number densities $n = (N_A + N_B)/2 + N_C$. In case of a reversibly associating fluid described by the law of mass action, the equilibrium constant is given by $K = k_{\text{off}}/k_{\text{on}}$, where k_{on} is the macroscopic association rate constant of Eq. (4.12) and k_{off} the respective dissociation rate constant. In a well-mixed (i.e., reaction-limited) and sufficiently diluted system, k_{on} can be approximated as in Section 4.4.1. However, for a diffusion-influenced process which we consider here, k_{on} is typically understood as a harmonic mean of encounter and formation rates [SS82; SSS80; GH80; NH79], i.e., $k_{\text{on}}^{-1} = k_{\text{enc}}^{-1} + k_{\text{form}}^{-1}$. At low densities, only two-body interactions between A and B determine the on-rate constant, in this limit, k_{on} can be evaluated numerically as a function of the microscopic association rate constant λ_{on} in the presence of the interaction potential, based on solving the Smoluchowski diffusion equation with a sink term that accounts for the volume reaction model, see [Dib+19]. Furthermore, in dense reversibly associating fluids, many-body interactions have an influence on k_{on}, in particular due to competition for reactants, clustering, volume exclusion, and caging [NH79].

Thus, it is challenging to find a consistent analytical description over multiple orders of magnitudes in density. In contrast, we perform an empirical evaluation by simulations as shown in Fig. 4.9. To this end, we set up 6 simulations for different $n \in [2 \times 10^1, 1.5 \times 10^4]$ in a constant volume which then are allowed to relax into an equilibrium state subject to detailed-balance and yield a measurement $K(n)$. The exact simulation parameters can be found in Table 4.3. The reference value for the dilute case is given by $K_{\text{dilute}} = k_{\text{off}}/k_{\text{on}}^{\text{dilute}}$, where $k_{\text{off}} = \lambda_{\text{off}}$ and $k_{\text{on}}^{\text{dilute}}$ is a function of the microscopic association rate constant λ_{on} as well as the interaction potential (Eq. (4.13)) and is numerically computed as described in [Dib+19].

Table 4.3: **Parameters of density-dependent reaction kinetics.**

Quantity	Symbol	Value	Unit
Thermal energy	$k_B T$	2.49	$\mathrm{kJ\,mol^{-1}}$
Volume	V	100^3	$\mathrm{nm^3}$
Radius A	r_A	1	nm
Radius B	r_B	0.8	nm
Radius C	r_C	1	nm
Diffusion coeff. A	D_A	0.01	$\mathrm{nm^2\,ns^{-1}}$
Diffusion coeff. B	D_B	0.0125	$\mathrm{nm^2\,ns^{-1}}$
Diffusion coeff. C	D_C	0.01	$\mathrm{nm^2\,ns^{-1}}$
Charge A	q_A	1.3	–
Charge B	q_B	−1	–
Charge C	q_C	0	–
Screening parameter	κ	3.82	$\mathrm{nm^{-1}}$
Debye-Hückel prefactor	$e^2 \varepsilon_0^{-1} \varepsilon_r^{-1}$	2349	$\mathrm{kJ\,nm\,mol^{-1}}$
Repulsion energy	U_r	1.	$\mathrm{kJ\,mol^{-1}}$
Cutoff radius	r_{cutoff}	4.7	nm
Reaction radius	R	2.	nm
Equilibrium constant	K_{dilute}	6.16×10^{-5}	$\mathrm{nm^{-3}}$
Macroscopic rate constant	k_{on}	0.11	$\mathrm{nm^3\,ns^{-1}}$
Macroscopic rate constant	k_{off}	6.58×10^{-6}	$\mathrm{ns^{-1}}$
Microscopic rate constant	λ_{on}	5.61×10^{-3}	$\mathrm{ns^{-1}}$
Microscopic rate constant	λ_{off}	6.58×10^{-6}	$\mathrm{ns^{-1}}$
Timestep	τ	0.1	ns

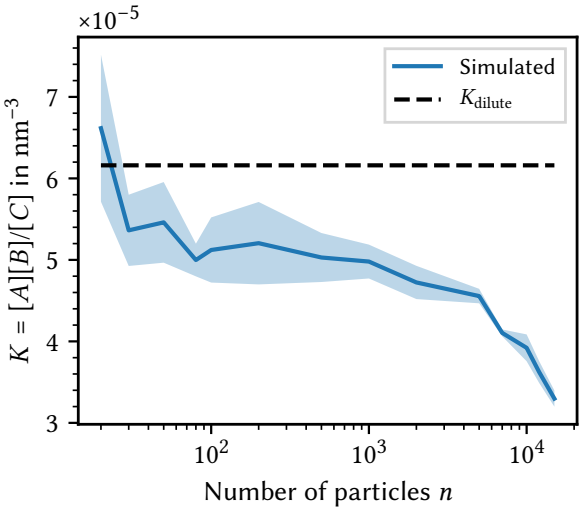

Figure 4.9: **Equilibrium constant transition from dilute to dense systems.** The equilibrium constant K is obtained by simulation for different choices of the number of particles $n = (N_A + N_B)/2 + N_C$ which corresponds to a density due to constant volume of the simulation box and compared to an analytically obtained equilibrium constant of a dilute system (dashed line). The number of particles n remains constant during the course of a simulation. The shaded areas are standard deviations from the recorded data.

We show that the reference value K_{dilute} is recovered by the simulation for low densities. For increasing densities more complex behavior can be observed. In particular, there is a drop in the value of K for $n \gtrsim 10^2$ which then is followed by a roughly stable regime up to $n \approx 5 \times 10^3$. For even higher densities, the equilibrium state is dominated by the complexes C likely due to finite size of the simulation volume. This drop in the equilibrium constant is in accord with Le Chatelier's principle [ADK18], i.e., the system prefers the state of lower free energy.

4.5 Availability and Future Directions

We have described the iPRD simulation framework ReaDDy 2 for combined particle interaction dynamics and reaction kinetics, which permits to conduct highly realistic simulations of signal transduction in crowded cellular environments or chemical nanoreactors with complex geometries. ReaDDy 2 follows up upon and significantly extends the simulation package ReaDDy 1. ReaDDy 2 is significantly faster than its predecessor, it can be easily installed as a Python conda package, and it can be flexibly used and reconfigured via its Python interface.

In comparison to molecular dynamics software packages, ReaDDy 2 does not include long range interactions. The software comes with a set of default interaction potentials. These include, e.g., harmonic repulsion which can model steric repulsion, Lennard–Jones interaction, and screened electrostatics which provide a way to model charged interaction at short ranges. Furthermore, ReaDDy allows for implementation of any short-ranged potential via a C++ interface. It is possible to implement and subsequently use hydration models which are short-ranged [SSN12; HD03] in the ReaDDy 2 framework. Hydrodynamic interactions are currently not included. They can be added by, e.g., providing an appropriate integrator which represents these interactions by a particle pairwise friction tensor [EM78].

Currently all pair potentials implemented in ReaDDy 2 are isotropic, however anisotropic interactions can be emulated by using particle complexes, in particular allowing for patchy particles. If the particles and interactions should be anisotropic themselves, a new computation kernel or appropriate integrator can be implemented into the framework via the C++ interface.

We have conducted a set of numerical studies, showing that ReaDDy 2 produces quantitatively accurate results where references from analytical solutions or other simulation packages were available, and physically meaningful results where reference solutions were not available.

For a quick and easy start into simulating and developing with ReaDDy 2 step by step tutorials, sample code, and further details are available online (https://readdy.github.io/). The software itself is Open Source and available under a permissive licence in order to enable a broad group of people to run simulations without forcing them to make their own work public.

ReaDDy 2 has been designed to be easily extensible. Planned extensions include simulation kernels for specialized hardware platforms, such as graphics processors and highly parallel HPC environments. Also planned is a MD-GFRD integrator [SN17] to speed up computations in dilute systems, and a particle-based membrane model as described in [SWN18] that reproduces mechanical properties of cellular membranes.

In its current state, membranes can be modeled in terms of external forces, i.e., constraining particles onto two-dimensional surfaces. As these constraints only apply to selected particle types, it is possible to, e.g., grow polymers against a static membrane, where one end is anchored.

References

[Abr+04] Rinat Abramovitch et al. "A Pivotal Role of Cyclic AMP-Responsive Element Binding Protein in Tumor Progression". In: *Cancer Research* 64.4 (2004), pp. 1338–1346. DOI: 10.1158/0008-5472.CAN-03-2089.

[ACH16] Shailesh R. Agarwal, Colleen E. Clancy, and Robert D. Harvey. "Mechanisms Restricting Diffusion of Intracellular cAMP". In: *Scientific Reports* 6.1 (2016), p. 19577. DOI: 10.1038/srep19577.

[ADK18] Peter Atkins, Julio De Paula, and James Keeler. *Atkins' physical chemistry*. 8th ed. Oxford: Oxford University Press, 2018. ISBN: 0198700725.

[And17] Steven S. Andrews. "Smoldyn: Particle-based simulation with rule-based modeling, improved molecular interaction and a library interface". In: *Bioinformatics* 33.5 (2017), pp. 710–717. DOI: 10.1093/bioinformatics/btw700.

[And18] Steven S. Andrews. "Particle-Based Stochastic Simulators". In: *Encyclopedia of Computational Neuroscience*. Ed. by Dieter Jaeger and Ranu Jung. New York, NY: Springer, 2018, pp. 1–5. DOI: 10.1007/978-1-4614-7320-6_191-2.

[AOA09] A. Ayadim, M. Oettel, and S. Amokrane. "Optimum free energy in the reference functional approach for the integral equations theory". In: *Journal of Physics Condensed Matter* 21.11 (2009). DOI: 10.1088/0953-8984/21/11/115103.

[Arn+13] A. Arnold et al. "ESPResSo 3.1 — Molecular Dynamics Software for Coarse-Grained Models". In: *Meshfree Methods for Partial Differential Equations VI*. Ed. by M. Griebel and M. A. Schweitzer. Vol. 89. Lecture Notes in Computational Science and Engineering. Springer, 2013, pp. 1–23. DOI: 10.1007/978-3-642-32979-1_1.

[AT10] Satya Nanda Vel Arjunan and Masaru Tomita. "A new multicompartmental reaction-diffusion modeling method links transient membrane attachment of E. coli MinE to E-ring formation". In: *Systems and Synthetic Biology* 4.1 (2010), pp. 35–53. DOI: 10.1007/s11693-009-9047-2.

[AT87] Michael P. Allen and Dominic J. Tildesley. *Computer Simulation of Liquids*. New York: Oxford University Press, 1987. DOI: 10.1093/oso/9780198803195.001.0001.

[Auh+03] Rolf Auhl et al. "Equilibration of long chain polymer melts in computer simulations". In: *The Journal of Chemical Physics* 119.24 (2003), pp. 12718–12728. DOI: 10.1063/1.1628670.

[BB02] Joseph A. Beavo and Laurence L. Brunton. "Cyclic nucleotide research — still expanding after half a century". In: *Nature Reviews Molecular Cell Biology* 3.9 (2002), pp. 710–717. DOI: 10.1038/nrm911.

[Bha04] Upinder S. Bhalla. "Signaling in Small Subcellular Volumes. I. Stochastic and Diffusion Effects on Individual Pathways". In: *Biophysical Journal* 87.2 (2004), pp. 733–744. DOI: 10.1529/biophysj.104.040469.

[Bie+15] Johann Biedermann et al. "ReaDDyMM: Fast interacting particle reaction-diffusion simulations using graphical processing units." In: *Biophysical journal* 108.3 (2015), pp. 457–61. DOI: 10.1016/j.bpj.2014.12.025.

[CH11] Peter H Colberg and Felix Höfling. "Highly accelerated simulations of glassy dynamics using GPUs: Caveats on limited floating-point precision". In: *Computer Physics Communications* 182.5 (2011), pp. 1120–1129. DOI: 10.1016/j.cpc.2011.01.009.

[DH23] Peter Debye and Eirich Hückel. "Zur Theorie der Elektrolyte. I. Gefrierpunktserniedrigung und verwandte Erscheinungen". In: *Physikalische Zeitschrift* 24.9 (1923), pp. 185–206.

[Dib+19] Manuel Dibak et al. "Diffusion-influenced reaction rates in the presence of pair interactions". In: *The Journal of Chemical Physics* 151.16 (2019), p. 164105. DOI: 10.1063/1.5124728.

[Doi75] Masao Doi. "Theory of diffusion-controlled reactions between non-simple molecules. I". In: *Chemical Physics* ll (1975), pp. 107–113. DOI: 10.1016/0301-0104(75)80043-7.

[Doi76] Masao Doi. "Stochastic theory of diffusion-controlled reaction". In: *Journal of Physics A* 9.9 (1976), p. 1479. DOI: 10.1088/0305-4470/9/9/009.

[Don+10] Aleksandar Donev et al. "A First-Passage Kinetic Monte Carlo algorithm for complex diffusion-reaction systems". In: *Journal of Computational Physics* 229.9 (2010), pp. 3214–3236. DOI: 10.1016/j.jcp.2009.12.038.

[Dra+09] Duska Dragun et al. "Autoimmune mediated G-protein receptor activation in cardiovascular and renal pathologies". In: *Thrombosis and Haemostasis* 101.4 (2009), pp. 643–648. DOI: 10.1160/TH08-10-0710.

[Dui+01] Adri C. T. van Duin et al. "ReaxFF: A Reactive Force Field for Hydrocarbons". In: *The Journal of Physical Chemistry A* 105.41 (2001), pp. 9396–9409. DOI: 10.1021/jp004368u.

[DYK18] Aleksandar Donev, Chiao-yu Yang, and Changho Kim. "Efficient reactive Brownian dynamics". In: *The Journal of Chemical Physics* 148.3 (2018), p. 034103. DOI: 10.1063/1.5009464.

[EC09] Radek Erban and S Jonathan Chapman. "Stochastic modelling of reaction–diffusion processes: algorithms for bimolecular reactions". In: *Physical Biology* 6.4 (2009), p. 046001. DOI: 10.1088/1478-3975/6/4/046001.

[Ein05] Albert Einstein. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen". In: *Annalen der Physik* 322.8 (1905), pp. 549–560. DOI: 10.1002/andp.19053220806.

[EM78] Donald L. Ermak and J. A. McCammon. "Brownian dynamics with hydrodynamic interactions". In: *The Journal of Chemical Physics* 69.4 (1978), pp. 1352–1360. DOI: 10.1063/1.436761.

[FN18] Christoph Fröhner and Frank Noé. "Reversible Interacting-Particle Reaction Dynamics". In: *The Journal of Physical Chemistry B* 122.49 (2018), pp. 11240–11250. DOI: 10.1021/acs.jpcb.8b06981.

[Fra+13] Benjamin Franz et al. "Multiscale reaction-diffusion algorithms: PDE-assisted Brownian dynamics". In: *SIAM Journal on Applied Mathematics* 73.3 (2013), pp. 1224–1247. DOI: 10.1137/120882469.

[GH80] Richard F. Grote and James T. Hynes. "The stable states picture of chemical reactions. II. Rate constants for condensed and gas phase reaction models". In: *The Journal of Chemical Physics* 73.6 (1980), pp. 2715–2732. DOI: 10.1063/1.440485.

[Gru+10] Gerd Gruenert et al. "Rule-based spatial modeling with diffusing, geometrically constrained molecules". In: *BMC bioinformatics* 11.1 (2010), p. 307. DOI: 10.1186/1471-2105-11-307.

[Gun+15] Monika Gunkel et al. "Higher-Order Architecture of Rhodopsin in Intact Photoreceptors and Its Implication for Phototransduction Kinetics". In: *Structure* 23.4 (2015), pp. 628–638. DOI: 10.1016/j.str.2015.01.015.

[GW97] Razif R. Gabdoulline and Rebecca C. Wade. "Simulation of the diffusional association of barnase and barstar". In: *Biophysical journal* 72.5 (1997), pp. 1917–1929. DOI: 10.1016/S0006-3495(97)78838-6.

[GW98] Razif R. Gabdoulline and Rebecca C. Wade. "Brownian dynamics simulation of protein–protein diffusional encounter". In: *Methods* 14.3 (1998), pp. 329–341. DOI: 10.1006/meth.1998.0588.

[HD03] Bertil Halle and Monika Davidovic. "Biomolecular hydration: from water dynamics to hydrodynamics". In: *Proceedings of the National Academy of Sciences* 100.21 (2003), pp. 12135–12140. DOI: 10.1073/pnas.2033320100.

[HDS96] William Humphrey, Andrew Dalke, and Klaus Schulten. "VMD: Visual molecular dynamics". In: *Journal of Molecular Graphics* 14.1 (1996), pp. 33–38. DOI: 10.1016/0263-7855(96)00018-5.

[HF13] Felix Höfling and Thomas Franosch. "Anomalous transport in the crowded world of biological cells". In: *arXiv* 76.4 (2013), p. 046602. DOI: 10.1088/0034-4885/76/4/046602.

[HFE05] Johan Hattne, David Fange, and Johan Elf. "Stochastic reaction-diffusion simulation with MesoRD". In: *Bioinformatics* 21.12 (2005), pp. 2923–2924. DOI: 10.1093/bioinformatics/bti431.

[HKD07] Detlef W.M. Hofmann, Liudmila Kuleshova, and Bruno D'Aguanno. "A new reactive potential for the molecular dynamics simulation of liquid water". In: *Chemical Physics Letters* 448.1-3 (2007), pp. 138–143. DOI: 10.1016/j.cplett.2007.09.063.

[Hou10] Miles D. Houslay. "Underpinning compartmentalised cAMP signalling through targeted cAMP breakdown". In: *Trends in Biochemical Sciences* 35.2 (2010), pp. 91–100. DOI: 10.1016/j.tibs.2009.09.007.

[HS14] Max Hoffmann and Ulrich S. Schwarz. "Oscillations of Min-proteins in micropatterned environments: a three-dimensional particle-based stochastic simulation approach". In: *Soft Matter* 10.14 (2014), pp. 2388–2396. DOI: 10.1039/c3sm52251b.

[Isa09] Samuel A. Isaacson. "The Reaction-Diffusion Master Equation as an Asymptotic Approximation of Diffusion to a Small Target". In: *SIAM Journal on Applied Mathematics* 70.1 (2009), pp. 77–111. DOI: 10.1137/070705039.

[Isa13] Samuel A. Isaacson. "A convergent reaction-diffusion master equation". In: *The Journal of Chemical Physics* 139.5 (2013), p. 054101. DOI: 10.1063/1.4816377.

[JZG93] J. Karl Johnson, John A. Zollweg, and Keith E. Gubbins. "The Lennard-Jones equation of state revisited". In: *Molecular Physics* 78.3 (1993), pp. 591–618. DOI: 10.1080/00268979300100411.

[Ker+08] Rex A. Kerr et al. "Fast Monte Carlo Simulation Methods for Biological Reaction-Diffusion Systems in Solution and on Surfaces". In: *SIAM Journal on Scientific Computing* 30.6 (2008), pp. 3126–3149. DOI: 10.1137/070692017.

[KS14] Heinrich C. R. Klein and Ulrich S. Schwarz. "Studying protein assembly with reversible Brownian dynamics of patchy particles". In: *The Journal of Chemical Physics* 140.18 (2014), p. 184112. DOI: 10.1063/1.4873708.

[KW11] Shina C. L. Kamerlin and Arieh Warshel. "The empirical valence bond model: theory and applications". In: *Wiley Interdisciplinary Reviews: Computational Molecular Science* 1.1 (2011), pp. 30–45. DOI: 10.1002/wcms.10.

[Lim+06] H.J. Limbach et al. "ESPResSo – An Extensible Simulation Package for Research on Soft Matter Systems". In: *Computer Physics Communications* 174.9 (2006), pp. 704–727. DOI: 10.1016/j.cpc.2005.10.005.

[LS93] Hartmut Löwen and Grzegorz Szamel. "Long-time self-diffusion coefficient in colloidal suspensions: theory versus simulation". In: *Journal of Physics: Condensed Matter* 5.15 (1993), pp. 2295–2306. DOI: 10.1088/0953-8984/5/15/003.

[Mar+15] Michael Martinez et al. "SDA 7: A modular and parallel implementation of the simulation of diffusional association software". In: *Journal of Computational Chemistry* 36.21 (2015), pp. 1631–1645. DOI: 10.1002/jcc.23971.

[ML16] Paul J. Michalski and Leslie M. Loew. "SpringSaLaD: a spatial, particle-based biochemical simulation platform with excluded volume". In: *Biophysical journal* 110.3 (2016), pp. 523–529. DOI: 10.1016/j.bpj.2015.12.026.

[MMW14] Paolo Mereghetti, Michael Martinez, and Rebecca C. Wade. "Long range Debye-Hückel correction for computation of grid-based electrostatic forces between biomacromolecules". In: *BMC Biophysics* 7.1 (2014), p. 4. DOI: 10.1186/2046-1682-7-4.

[Mun+09] T. Munk et al. "Effective Perrin theory for the anisotropic diffusion of a strongly hindered rod". In: *EPL (Europhysics Letters)* 85.3 (2009), p. 30003. DOI: 10.1209/0295-5075/85/30003.

[NF07] Francois Nedelec and Dietrich Foethke. "Collective Langevin dynamics of flexible cytoskeletal fibers". In: *New Journal of Physics* 9.11 (2007), pp. 427–427. DOI: 10.1088/1367-2630/9/11/427.

[NH79] Scott H. Northrup and James T. Hynes. "Short range caging effects for reactions in solution. I. Reaction rate constants and short range caging picture". In: *The Journal of Chemical Physics* 71.2 (1979), pp. 871–883. DOI: 10.1063/1.438378.

[Pla+17] Nuria Plattner et al. "Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling". In: *Nature Chemistry* 9.10 (2017), pp. 1005–1011. DOI: 10.1038/nchem.2785.

[Pli95] Steve Plimpton. "Fast Parallel Algorithms for Short-Range Molecular Dynamics". In: *Journal of Computational Physics* 117.1 (1995), pp. 1–19. DOI: 10.1006/jcph.1995.1039.

[Pos+13] York Posor et al. "Spatiotemporal control of endocytosis by phosphatidylinositol-3,4-bisphosphate". In: *Nature* 499.7457 (2013), pp. 233–237. DOI: 10.1038/nature12360.

[RAL78] M. Thomas Record, Charles F. Anderson, and Timothy M. Lohman. "Thermodynamic analysis of ion effects on the binding and conformational equilibria of proteins and nucleic acids: the roles of ion association or release, screening, and ion effects on water activity". In: *Quarterly Reviews of Biophysics* 11.2 (1978), pp. 103–178. DOI: 10.1017/S003358350000202X.

[RC03] Michael Rubinstein and Ralph H. Colby. *Polymer physics*. Vol. 23. Oxford University Press New York, 2003. ISBN: 9780198520597.

[Rob+09] Elijah Roberts et al. "Long time-scale simulations of in vivo diffusion using GPU hardware". In: *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*. IEEE. IEEE, 2009, pp. 1–8. DOI: 10.1109/IPDPS.2009.5160930.

[Rob+11] Elijah Roberts et al. "Noise Contributions in an Inducible Genetic Switch: A Whole-Cell Simulation Study". In: *PLoS computational biology* 7.3 (2011). Ed. by Peter Csermely, e1002010. DOI: 10.1371/journal.pcbi.1002010.

[RSL13] Elijah Roberts, John E. Stone, and Zaida Luthey-Schulten. "Lattice microbes: High-performance stochastic simulation method for the reaction-diffusion master equation". In: *Journal of Computational Chemistry* 34.3 (2013), pp. 245–255. DOI: 10.1002/jcc.23130.

[San+06] Chris Sanford et al. "Cell++—simulating biochemical pathways". In: *Bioinformatics* 22.23 (2006), pp. 2918–2925. DOI: 10.1093/bioinformatics/btl497.

[Sch+14] Johannes Schöneberg et al. "Explicit Spatiotemporal Simulation of Receptor-G Protein Coupling in Rod Cell Disk Membranes". In: *Biophysical Journal* 107.5 (2014), pp. 1042–1053. DOI: 10.1016/j.bpj.2014.05.050.

[Sch+17] Johannes Schöneberg et al. "Lipid-mediated PX-BAR domain recruitment couples local membrane constriction to endocytic vesicle fission". In: *Nature Communications* 8.May (2017), p. 15873. DOI: 10.1038/ncomms15873.

[SN13] Johannes Schöneberg and Frank Noé. "ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments." In: *PLoS One* 8.9 (2013), e74261. DOI: 10.1371/journal.pone.0074261.

[SN17] Luigi Sbailò and Frank Noé. "An efficient multi-scale Green's function reaction dynamics scheme". In: *The Journal of Chemical Physics* 147.18 (2017), p. 184106. DOI: 10.1063/1.5010190.

[SS82] David Shoup and Attila Szabo. "Role of diffusion in ligand binding to macromolecules and cell-bound receptors." In: *Biophysical Journal* 40.1 (1982), p. 33. DOI: 10.1016/S0006-3495(82)84455-X.

[SSN12] E. Schneck, F. Sedlmeier, and R. R. Netz. "Hydration repulsion between biomembranes results from an interplay of dehydration and depolarization". In: *Proceedings of the National Academy of Sciences* 109.36 (2012), pp. 14405–14409. DOI: 10.1073/pnas.1205811109.

[SSS80] Attila Szabo, Klaus Schulten, and Zan Schulten. "First passage time approach to diffusion controlled reactions". In: *Journal of Chemical Physics* 72.8 (1980), pp. 4350–4357. DOI: 10.1063/1.439715.

[SUN14] Johannes Schöneberg, Alexander Ullrich, and Frank Noé. "Simulation tools for particle-based reaction-diffusion dynamics in continuous space". In: *BMC Biophysics* 7.1 (2014), p. 11. DOI: 10.1186/s13628-014-0011-5.

[SWN18] Mohsen Sadeghi, Thomas R. Weikl, and Frank Noé. "Particle-based membrane model for mesoscopic simulation of cellular dynamics". In: *The Journal of Chemical Physics* 148.4 (2018), p. 044901. DOI: 10.1063/1.5009107.

[Swo+82] William C. Swope et al. “A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters”. In: *The Journal of Chemical Physics* 76.1 (1982), pp. 637–649. DOI: 10.1063/1.442716.

[TD11] Amantha Thathiah and Bart De Strooper. “The role of G protein-coupled receptors in the pathology of Alzheimer’s disease”. In: *Nature Reviews Neuroscience* 12.2 (2011), pp. 73–87. DOI: 10.1038/nrn2977.

[Trz+12] B. Trzaskowski et al. “Action of Molecular Switches in GPCRs - Theoretical and Experimental Studies”. In: *Current Medicinal Chemistry* 19.8 (2012), pp. 1090–1109. DOI: 10.2174/092986712799320556.

[TS67] Ei Teramoto and Nanako Shigesada. “Theory of bimolecular reaction processes in liquids”. In: *Progress of Theoretical Physics* 37.1 (1967), pp. 29–51. DOI: 10.1143/PTP.37.29.

[TTW10] Koichi Takahashi, Sorin Tanase-Nicola, and Pieter Rein ten Wolde. “Spatio-temporal correlations can drastically change the response of a MAPK pathway”. In: *Proceedings of the National Academy of Sciences* 107.6 (2010), pp. 2473–2478. DOI: 10.1073/pnas.0906885107.

[Ull+15] Alexander Ullrich et al. “Dynamical Organization of Syntaxin-1A at the Presynaptic Active Zone”. In: *PLOS Computational Biology* 11.9 (2015). Ed. by Kim T. Blackwell, e1004407. DOI: 10.1371/journal.pcbi.1004407.

[VBW15] Adithya Vijaykumar, Peter G. Bolhuis, and Pieter Rein ten Wolde. “Combining molecular dynamics with mesoscopic Green’s function reaction dynamics simulations”. In: *The Journal of Chemical Physics* 143.21 (2015), p. 214102. DOI: 10.1063/1.4936254.

[Vij+17] Adithya Vijaykumar et al. “Multiscale simulations of anisotropic particles combining molecular dynamics and Green’s function reaction dynamics”. In: *The Journal of Chemical Physics* 146.11 (2017), p. 114106. DOI: 10.1063/1.4977515.

[Von06] Marian Von Smoluchowski. “Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen”. In: *Annalen der Physik* 326.14 (1906), pp. 756–780. DOI: 10.1002/andp.19063261405.

[XGG96] Jun Xing, David D. Ginty, and Michael E. Greenberg. “Coupling of the RAS-MAPK Pathway to Gene Activation by RSK2, a Growth Factor-Regulated CREB Kinase”. In: *Science* 273.5277 (1996), pp. 959–963. DOI: 10.1126/science.273.5277.959.

[Yuk35] Hideki Yukawa. “On the interaction of elementary particles. I”. In: *Proceedings of the Physico-Mathematical Society of Japan. 3rd Series* 17 (1935), pp. 48–57. DOI: 10.11429/ppmsj1919.17.0_48.

[ZW05a] Jeroen S. van Zon and Pieter Rein ten Wolde. “Green’s-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space”. In: *The Journal of Chemical Physics* 123.23 (2005), p. 234910. DOI: 10.1063/1.2137716.

[ZW05b] Jeroen S. van Zon and Pieter Rein ten Wolde. “Simulating biochemical networks at the particle level and in time and space: Green’s function reaction dynamics”. In: *Physical Review Letters* 94.12 (2005), p. 128103. DOI: 10.1103/PhysRevLett.94.128103.

Chapter 5

Reactive SINDy: Discovering governing reactions from concentration data

The results of this chapter have been published in the following paper:

> Moritz Hoffmann, Christoph Fröhner (CF), and Frank Noé. "Reactive SINDy: Discovering governing reactions from concentration data". In: *The Journal of Chemical Physics* 150.2 (2019), p. 025101. 10.1063/1.5066099

Parts of the text and illustrations have been adopted unchanged in this document. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

Moritz Hoffmann and CF contributed equally to this work. In particular the contributions were as follows: Moritz Hoffmann, CF and Frank Noé conceived the project and laid out the theory. Moritz Hoffmann and CF set up the software pipeline for generation of training data, the minimization procedure and cross validation. Moritz Hoffmann took care of the results in the low-noise limit (Section 5.3.1). CF ran the cross validation for noisy measurements (Sections 5.3.2 and 5.3.3) on the compute cluster. CF applied the Reactive SINDy method to the predator-prey and MAPK example (Sections 5.3.4 and 5.3.5). Moritz Hoffmann and CF analyzed and visualized all resulting data. All contributors wrote the paper.

Summary

The inner workings of a biological cell or a chemical reactor can be rationalized by the network of reactions, whose structure reveals the most important functional mechanisms. For complex systems, these reaction networks are not known a priori and cannot be efficiently computed with *ab initio* methods, therefore an important approach goal is to estimate effective reaction networks from observations, such as time series of the main species. Reaction networks estimated with standard machine learning techniques such as least-squares regression may fit the observations, but will typically contain spurious reactions. Here we extend the sparse identification of nonlinear dynamics (SINDy) method to vector-valued ansatz functions, each describing a particular reaction process. The resulting sparse tensor regression method "reactive SINDy" is able to estimate a parsimonious reaction network. We illustrate that a gene regulation network can be correctly estimated from observed time series.

5.1 Introduction

Mapping out the reaction networks behind biological processes, such as gene regulation in cancer [Abr+04], is paramount to understanding the mechanisms of life and disease. A well-known example of gene regulation is the lactose operon whose crystal structure was resolved in [Lew+96] and dynamics were modeled in [YM03]. The system's "combinatorial control" in E. coli cells was quantitatively investigated in [Kuh+07], in particular studying repression and activation effects. These gene regulatory effects often appear in complex networks [She+02] and there exist databases resolving these for certain types of cells, e.g., E. coli cells [Gam+16] and yeast cells [Lee02]. Another example where mapping the active reactions is important is that of chemical reactors [Roa+17], where understanding which reactions are accessible for a given set of educts and reaction conditions is important to design synthesis pathways [Con+99; KR05].

The traditional approach to determine a reaction network is to propose the structure of the network based on chemical insight and subsequently fit the parameters given available data [Sch+14]. To decipher complex reaction environments such as biological cells, it would be desirable to have a data-driven approach that can answer the question which reactions are underlying a given observation, e.g., the time series of a set of reactants. However, in sufficiently complex reaction environments the number of reactive species and possible reactions is practically unlimited – as an illustration, consider vast amount of possible isomerizations and post-translational modifications for a single protein molecule. Therefore, the more specific formulation is "given observations of a set of chemical species, what is the *minimal set* of reactions necessary to explain their time evolution?". This formulation calls for a machine learning method that can infer the reaction network underlying the observation data.

Knowledge about the reaction network can be applied to parameterize other numerical methods to further investigate the processes at hand. Such methods include particle-based approaches derived from the chemical master equation [Gil77; WS16; WS17; Isa09], as well as highly detailed but parameter-rich methods such as particle-based or interacting-particle reaction dynamics [SN13; HFN19b; FN18; DYK18; And17; ZW05a; ZW05b] capable of fully resolving molecule positions in space and time – see [SUN14; And18] for recent reviews.

Existing methods to infer regulatory networks include ARCANE [Mar+06] that uses experimental essay data and information theory, as well as the likelihood approach presented in [Tia+07] that takes the stochasticity of observed reactant time series into account.

The method presented in this work can identify underlying complex reaction networks from concentration time series by following the law of parsimony, i.e., by inducing sparsity in the resulting reaction network. This promotes the interpretability of the model and avoids overfitting. We formulate the problem as data-driven identification of a dynamical system, which renders the method consistent with and an extension of the framework of sparse identification of nonlinear dynamics (SINDy) [BPK15]. Specifically, the problem of identifying a reaction network from time traces of reactant concentrations can be solved by finding a linear combination from a library of candidate nonlinear functions (ansatz functions) that each corresponds to a reaction acting on a set of reactants. With this formulation, the reaction rates can be determined *via* regression. Sparsity is induced by equipping the regression algorithms with a sparsity inducing regularization. SINDy was investigated, generalized, and applied in many different ways, e.g., including control [BPK16] (SINDYc), in the context of partial differential equations [Rud+17], updating already existing models [Qua+18] (abrupt-SINDy), and looking into convergence properties [ZS18].

We extend and apply SINDy to the case of learning reaction networks from non-equilibrium concentration data. Similar approaches make use of SINDy but do not resolve specific reactions [Man+16], use weak formulations to avoid numerical temporal derivatives [PT17], or use compressive sensing and sparse Bayesian learning [Pan+12].

Our extension of the original SINDy method mostly involves estimating parameters which are coupled across the equations of the arising dynamical system. In the context of learning reaction networks this

means that we look for specific reactions and their rate constants that might have lead to the observations instead of net flux across species. We demonstrate the algorithm on a gene regulatory network in three different scenarios of measurement: When there is no noise in the data we can find, given sufficient amounts of data, all relevant processes of the ground truth. If there is noise in the data we converge to the correct reaction network and rates with decreasing levels of noise. The third scenario generalizes the method to two measurements with different initial conditions, also converging to the correct model with decreasing levels of noise.

We additionally demonstrate the algorithm on time series data of the mitogen activated protein kinases (MAPK) pathway as an example for a bimodal system and on time series data of the Lotka–Volterra system which describes oscillatory predator-prey dynamics subject to social friction. In both systems reactive SINDy recovers the generating reaction network whereas non-sparse estimation detects many spurious processes.

5.2 Reactive SINDy: Sparse learning of reaction kinetics

We are observing the concentrations of S chemical species in time t:

$$\mathbf{x}(t) = \begin{pmatrix} x_1(t) \\ \vdots \\ x_S(t) \end{pmatrix} \in \mathbb{R}^S. \tag{5.1}$$

We assume that their dynamics are governed by classical reaction-rate equations subject to the law of mass action. A general expression for the change of concentration of reactant s as a result of order-0 reactions (creation), order-1 reactions (transitions of other species into s, transitions of s into other species, or annihilation), order-2 reactions (production or consumption of s by the encounter of two species), etc, is given by:

$$\dot{x}_s = \sum_i \beta^{(i)}_{s,0} + \sum_i \beta^{(i)}_{s,1} x_i + \sum_{i,j} \beta^{(i,j)}_{s,2} x_i x_j + \dots \tag{5.2}$$

where the $\beta^{(\dots)}_{s,k}$-values are constants belonging to the reactions of order k. These rate constants however can incorporate several underlying reactions at once. For example, the two reactions

$$s_1 \xrightarrow{\xi_1} s_2 \tag{5.3}$$

$$s_1 \xrightarrow{\xi_2} s_3 \tag{5.4}$$

both contribute to $\dot{x}_1 = \beta^{(1)}_{1,1} x_1 = -(\xi_1 + \xi_2) x_1$. To disentangle (5.2) into single reactions, we choose a library of R possible ansatz reactions that each represent a single reaction:

$$\mathbf{y}_r(\mathbf{x}(t)) = \begin{pmatrix} y_{r,1}(\mathbf{x}(t)) \\ \vdots \\ y_{r,S}(\mathbf{x}(t)) \end{pmatrix}, \quad r = 1, \dots, R. \tag{5.5}$$

With this ansatz, the reaction dynamics (5.2) becomes a set of linear equations with unknown parameters ξ_r that represent the sought macroscopic rate constants:

$$\dot{\mathbf{x}}_i(t) = \sum_{r=1}^{R} y_{r,i}(\mathbf{x}(t)) \xi_r, \quad i = 1, \dots, S, \tag{5.6}$$

where ξ_r are the to-be estimated macroscopic rate constants. The two reactions in the previous example (5.3-5.4) would be modeled by the functions

$$y_1(\mathbf{x}) = (-x_1, x_1, 0)^\top,$$
$$y_2(\mathbf{x}) = (-x_1, 0, x_1)^\top,$$

illustrating that the values of the coefficients ξ_1 and ξ_2 can be used to decide whether a single reaction is present and to what degree.

Now suppose we have measured the concentration vector (5.1) at T time points $t_1 < \cdots < t_T$. We represent these data as a matrix

$$\mathbf{X} = \begin{pmatrix} \mathbf{x}(t_1) & \mathbf{x}(t_2) & \cdots & \mathbf{x}(t_T) \end{pmatrix}^\top \in \mathbb{R}^{T\times S}. \tag{5.7}$$

Given this matrix, a library $\Theta : \mathbb{R}^{T\times S} \to \mathbb{R}^{T\times S\times R}$, $\mathbf{X} \mapsto \begin{pmatrix}\theta_1(\mathbf{X}) & \theta_2(\mathbf{X}) & \cdots & \theta_R(\mathbf{X})\end{pmatrix}$ of R candidate (ansatz) reactions can be proposed with corresponding reaction functions

$$\theta_r(\mathbf{X}) = \begin{pmatrix} \mathbf{y}_r(\mathbf{X}_{1*})^\top \\ \vdots \\ \mathbf{y}_r(\mathbf{X}_{T*})^\top \end{pmatrix} \in \mathbb{R}^{T\times S}, \quad r = 1, \dots, R, \tag{5.8}$$

where $\mathbf{X}_{i*}$ denotes the i-th row in $\mathbf{X}$. Applying the concentration trajectory to the library yields $\Theta(\mathbf{X}) \in \mathbb{R}^{T\times S\times R}$.

The goal is to find coefficients $\Xi = \begin{pmatrix}\xi_1 & \xi_2 & \cdots & \xi_R\end{pmatrix}^\top$, so that

$$\dot{\mathbf{X}} = \Theta(\mathbf{X})\Xi = \sum_{r=1}^{R} \theta_r(\mathbf{X})\xi_r. \tag{5.9}$$

In particular, the system is linear in the coefficients Ξ, which makes regression tools such as elastic net regularization [ZH05] applicable. To this end, one can consider the regularized minimization problem (reactive SINDy):

$$\hat{\Xi} = \arg\min_{\Xi} \left(\frac{1}{2T} \left\| \dot{\mathbf{X}} - \Theta(\mathbf{X})\Xi \right\|_F^2 + \alpha\lambda\|\Xi\|_1 + \alpha(1-\lambda)\|\Xi\|_2^2 \right) \quad \text{subject to } \Xi \geqslant 0. \tag{5.10}$$

Here, $\|\cdot\|_F$ denotes the Frobenius norm, $\lambda \in [0, 1]$ is a hyperparameter that interpolates linearly between LASSO [Tib96; HTF09] and Ridge [HK70] methods, and $\alpha \geqslant 0$ is a hyperparameter that, depending on λ, can induce sparsity and give preference to smaller solutions in the L_1 or L_2 sense. For $\alpha = 0$ the minimization problem reduces to standard least-squares (LSQ) with the constraint $\Xi \geqslant 0$. Reactive SINDy (5.10) is therefore a generalization of the SINDy method to vector-valued ansatz functions.

Since only the concentration data $\mathbf{X}$ is available but not its temporal derivative, $\dot{\mathbf{X}}$ is approximated numerically by second order finite differences with the exception of boundary data. Once the pair $(\mathbf{X}, \dot{\mathbf{X}})$ is obtained, the problem becomes invariant under temporal reordering. Hence, when presented with multiple trajectories the data matrices $\mathbf{X}_i$ and $\dot{\mathbf{X}}_i$ can simply be concatenated.

In order to solve (5.10) the numerical sequential least-squares minimizer SLSQP [Kra88] is applied via the software package SciPy [JOP+01]. Code related to this chapter can be found under `https://github.com/readdy/readdy_learn`.

5.3 Results

We demonstrate the method by estimating the reactions of a gene-regulatory network from time series of concentrations of the involved molecules. Let $S := \{A, B, C\}$ be a set of three species of proteins which

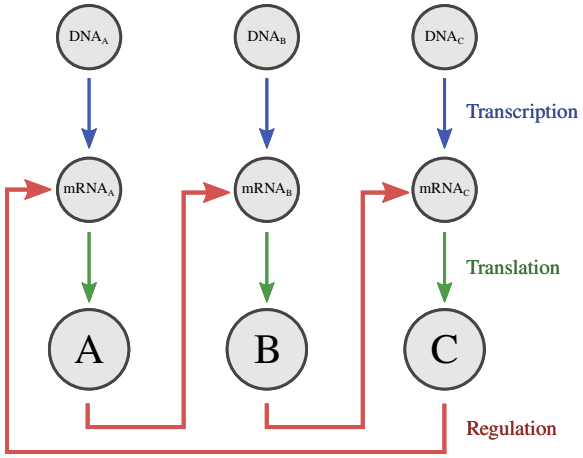

Figure 5.1: The regulation network example described in Sec. 5.3. Each circle depicts a species, each arrow corresponds to one reaction. Blue arrows denote transcription from DNA to mRNA, green arrows denote translation from mRNA to protein, and red arrows denote the regulatory network. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

are being translated each from their respective mRNA molecule. Each mRNA in turn has a corresponding DNA which it is transcribed from. The proteins and mRNA molecules decay over time whereas the DNA concentration remains constant. The network contains reactions of the following form [TO01]

$$\begin{aligned}
\mathrm{DNA}_i &\longrightarrow \mathrm{DNA}_i + \mathrm{mRNA}_i && \text{(transcription)},\\
\mathrm{mRNA}_i &\longrightarrow \mathrm{mRNA}_i + i && \text{(translation)},\\
\mathrm{mRNA}_i &\longrightarrow \emptyset && \text{(decay of mRNA)},\\
i &\longrightarrow \emptyset && \text{(decay of protein)},\\
i + \mathrm{mRNA}_j &\longrightarrow i && \text{(regulation of } j \in S\text{)},
\end{aligned}$$

for each of the species $i \in S$. These reactions model a regulation of species j by virtue of the fact that the transcription product inhibits the transcription processes. In our example proteins of type A regulate the $\mathrm{mRNA_B}$ molecules, proteins of type B regulate the $\mathrm{mRNA_C}$ molecules and proteins of type C regulate the $\mathrm{mRNA_A}$ molecules (Fig. 5.1). Using this reaction model, time series of concentrations are generated using the rates given in Tab 5.2 under the initial condition described in Tab 5.1a, which were chosen so that all the reactions in the reaction model significantly contribute to the temporal evolution of the system's concentrations. The generation samples the integrated equations equidistantly with a discrete time step of $\tau = 3 \cdot 10^{-3}$ yielding 667 frames which amounts to a cumulative time of roughly $T = 2$.

The proposed estimation method is applied to analyze these time series of concentrations in order to recover the underlying reaction network from data. To this end we use the library of ansatz functions given in Tab. 5.2, which contains a large number of possible reactions, only few of which are actually part of the model.

	DNA_A	$mRNA_A$	A	DNA_B	$mRNA_B$	B	DNA_C	$mRNA_C$	C
(a)	1	2	0	1	0	3	1	0	0
(b)	1	1.5	0	1	0	2	1	0	1

Table 5.1: Initial conditions **(a)** and **(b)** used to generate concentration time series. Reaction rates can be found in Tab. 5.2. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

5.3.1 Learning the reaction network in the low-noise limit

We first demonstrate that the true reaction network can be reconstructed when using a finite amount of observation data without additional measurement noise, i.e., the observations are reflecting the true molecule concentrations at any given time point. The minimization problem (5.10) is solved using the concentration time series shown in Fig. 5.1b.

We first set the hyperparameter $\alpha = 0$ in the minimization problem (5.10), which results in constrained least-squares regression without any of the regularization terms. In this case we estimate a reaction network that can reproduce the observations almost exactly (Fig. 5.2). However, the result is mechanistically wrong as the sparsity pattern does not match the reaction network used to generate the data. On the one hand many spurious reactions are estimated that were not in the true reaction scheme and would lead to wrong conclusions about the mechanism, such as $A + A \longrightarrow A$ and $A + C \longrightarrow C$. More dramatically, the reaction responsible for the decay of A particles is completely ignored (Fig. 5.3).

Next, we sought sparse solutions by using $\alpha > 0$ and additionally eliminating reactions with rate constants smaller than a cutoff value κ. For a suitable choice of hyperparameters $\alpha \approx 1.91 \cdot 10^{-7}$, $\lambda = 1$, and $\kappa = 0.22$, a sparse solution is obtained that finds the correct reaction scheme and also recovers the decay reaction (Fig. 5.3).

The value of the cutoff κ was determined by comparing the magnitude of estimated rates and finding a gap, see Fig. 5.8. The hyperparameter pair (α, λ) was obtained by a grid search and evaluating the difference $\|\hat{\Xi}_{\alpha,\lambda} - \Xi\|_1$, where $\hat{\Xi}_{\alpha,\lambda}$ is the estimated model under a particular hyperparameter choice and Ξ is the ground truth. If the ground truth is unknown, a hyperparameter pair can be estimated by utilizing cross-validation as in the following sections.

5.3.2 Learning the reaction network from data with stochastic noise

In contrast to Sec. 5.3.1, we now employ data that includes measurement noise. Such noise can originate from uncertainties in the experimental setup or from shot noise in single- or few-molecule measurements. In gene regulatory networks such noise is commonly observed when only few copy numbers of mRNA are present [Gol+05; Ber78; Elo02]. In order to simulate noise from few copies of molecules, the system of Sec. 5.3 with initial conditions as given in Tab. 5.1a is integrated using the Gillespie stochastic simulation algorithm (SSA) [Gil76; Gil77]. In the limit of many particles and realizations, the Gillespie SSA converges to the integrated reaction-rate equations subject to the law of mass action. As our model is based on exactly these dynamics, the initial condition's concentrations are interpreted in terms of hundreds of particles. Each realization is then transformed back to a time series of concentrations. We define the noise level as the mean-squared deviation of the concentration time series from the integrated reaction-rate equations. Data with different noise levels are prepared by averaging multiple realizations of the time series obtained by the Gillespie SSA.

It can be observed that decreasing levels of noise lead to fewer spurious reactions when applying reactive SINDy (5.10), see Fig. 5.4a. Also the estimation error $\|\xi - \hat{\xi}\|_1$ with respect to the ground truth ξ

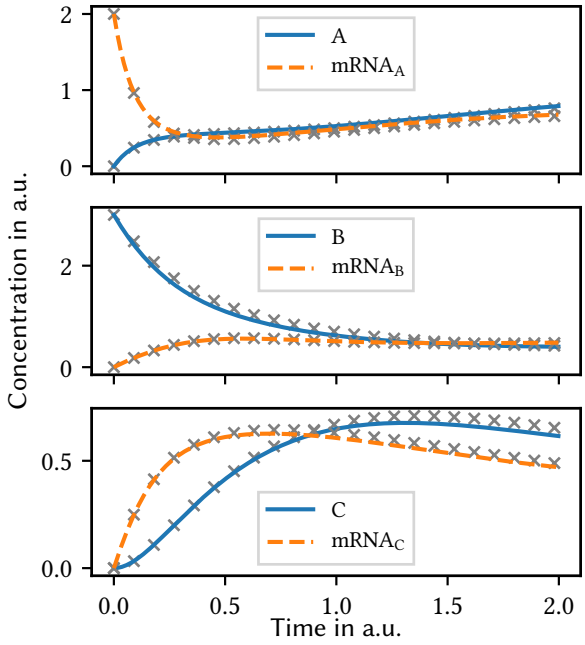

Figure 5.2: Concentration time series generated from integrating the reaction network shown in Fig. 5.1a. The initial condition prescribes positive concentration values only for B protein and $mRNA_A$ species (Tab. 5.1a). This initial condition is used in the subsequent sections for further analysis. Gray dots depict concentration time series yielded from the LSQ rates estimated in Sec 5.3.1. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

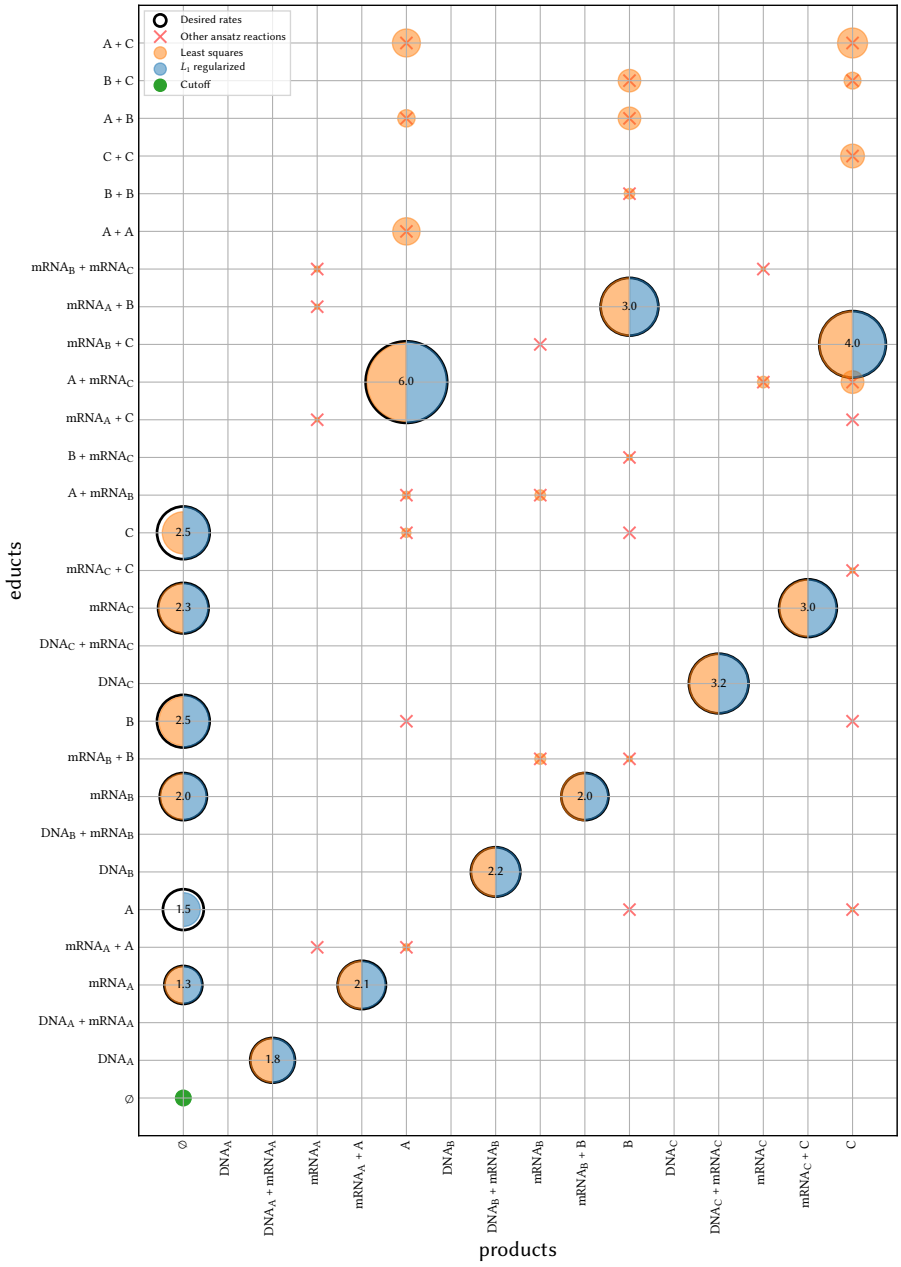

Figure 5.3: Estimated reaction rates in the system described in Sec. 5.3.1. The y and x axes contain reaction educts and products, respectively. A circle at position (i, j) represents a reaction $i \rightharpoonup j$ whose rate has a linear relation with the area of the circle. The black outlines denote the reactions with which the system was generated and contain the respective rate value. Red crosses denote reactions that were used as additional ansatz reactions. Blue circles are estimated by LSQ and orange circles depict rates which were obtained by solving the minimization problem (5.10). The latter rates are subject to a cutoff $\kappa = 0.22$ corresponding to the green circle's area under which a sparse solution with the correct processes can be recovered. If a certain rate was estimated in both cases, two wedges instead of one circle are displayed. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

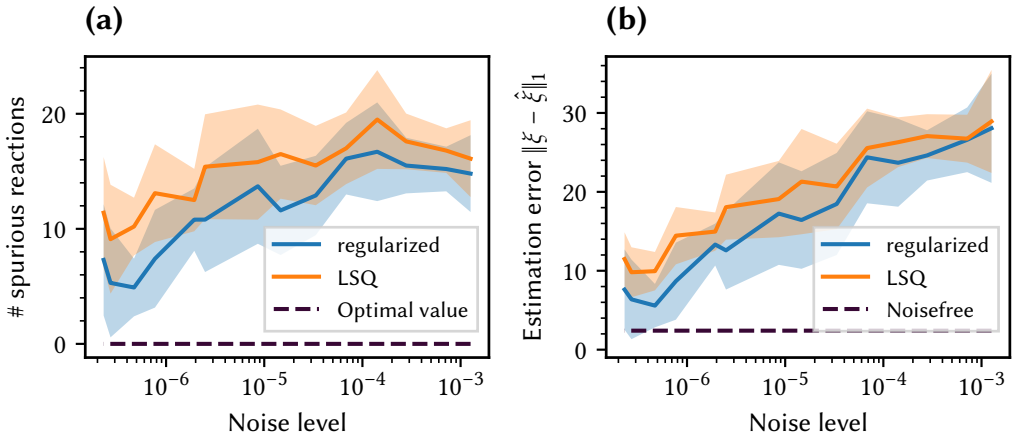

Figure 5.4: Convergence of the estimation error when estimating the system described in Sec. 5.3.1 with varying levels of noise by application of reactive SINDy (5.10) with and without regularization in blue and orange, respectively. The procedure was independently repeated 10 times with different realizations giving rise to the mean and standard deviation depicted by solid lines and shaded areas, respectively. **(a)**: The number of detected spurious reactions up to the cutoff value introduced in Sec. 5.3.1 over different levels of noise. **(b)**: The estimation error given by the mean absolute error between the generating reaction rates ξ and the estimated reaction rates $\hat{\xi}$ over different levels of noise. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

decreases with decreasing levels of noise (Fig. 5.4b). In both cases, the regularized method with a suitable hyperparameter pair (α, λ) performs better than LSQ.

The hyperparameters (α, λ) are obtained by shuffling the data and performing a 10-fold cross validation.

5.3.3 Learning the reaction network from multiple initial conditions

Preparing the experiment that generates the data in different initial conditions can help identifying the true reaction mechanisms as a more diverse dataset makes it easier to confirm or exclude the participation of specific reactions. This section extends the analysis of Sec. 5.3.2 to two initial conditions, where the first initial condition is identical to the one used previously and the second initial condition is given in Tab. 5.1b.

The corresponding time series are depicted in Fig. 5.5a. The gray graph corresponds to a sample trajectory generated by the Gillespie SSA. For both initial conditions the same time step of $\tau = 3 \cdot 10^{-3}$ has been applied, amounting to $2 \cdot 667 = 1334$ frames. Once the data matrices

$$\mathbf{X}_1 = \begin{pmatrix} \boldsymbol{x}_1(t_1) & \cdots & \boldsymbol{x}_1(t_{667}) \end{pmatrix}, \; \mathbf{X}_2 = \begin{pmatrix} \boldsymbol{x}_2(t_1) & \cdots & \boldsymbol{x}_2(t_{667}) \end{pmatrix}$$

and the corresponding derivatives $\dot{\mathbf{X}}_1, \dot{\mathbf{X}}_2$ have been obtained, the frames are concatenated so that

$$\mathbf{X} = \begin{pmatrix} \boldsymbol{x}_1(t_1) & \cdots & \boldsymbol{x}_1(t_{667}) & \boldsymbol{x}_2(t_1) & \cdots & \boldsymbol{x}_2(t_{667}) \end{pmatrix},$$

analogously for $\dot{\mathbf{X}}$.

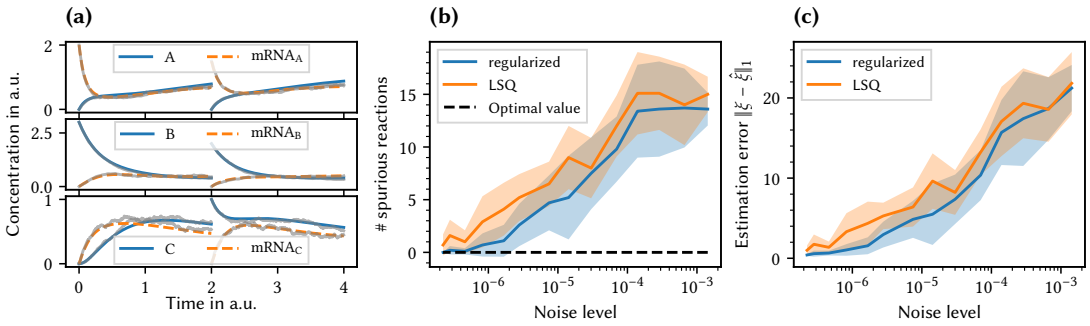

Figure 5.5: Convergence of estimation error of reaction schemes from noisy gene-regulation data starting from two different initial conditions under decreasing levels of noise. The minimization problem (5.10) was solved for $\alpha = 0$ (LSQ) and with regularization. This was repeated 10 times on different sets of observation data generated by Gillespie SSA, giving rise to mean and standard deviation (solid lines and shaded areas, respectively). **(a)**: Concentration time series corresponding to the initial conditions, generated by integrating the reaction-rate equations. The first initial condition is identical to the one used in Sec. 5.3.1 and Sec. 5.3.2. The second initial condition (Tab. 5.1b) prescribes positive initial concentrations for mRNA$_A$, B, and C species. The gray graphs are sample realizations of integration using the Gillespie SSA. **(b)**,**(c)**: Analogously to Fig. 5.4 with the difference that 20-fold cross validation was used for hyperparameter estimation. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

Similarly to Sec. 5.3.2, decreasing levels of noise lead to fewer spurious reactions (Fig. 5.5b) and a smaller L_1 distance to the ground truth (Fig. 5.5c). Again applying the optimization problem with a suitable set of parameters $(\alpha, \lambda, \kappa)$ performs better than LSQ. Compared to the previous section the convergence is better due to twice as much available data. At noise levels of smaller than roughly 10^{-6} the model can reliably be recovered when using the regularized method.

The hyperparameters (α, λ) are obtained by shuffling the data and performing a 20-fold cross validation.

5.3.4 Application to MAPK cascade

The reactive SINDy method is applied to the mitogen activated protein kinases (MAPK) pathway [XGG96] which is an important regulatory mechanism of biological cells to respond to stimuli and is involved in proliferation, differentiation, inflammation, and apoptosis [ZL02]. Single-cell MAPK kinetics can be observed experimentally [Ryu+18]. Mathematically MAPK kinetics are often modelled using reaction rate equations [KCG05; Ort+05] which enables analysis using reactive SINDy.

Generally a MAPK pathway consists of multiple stages of kinases that are either inactive or active, denoted by "*". Their activation occurs due to phosphorylation catalyzed by the upstream kinase of the previous stage, dephosphorylation is catalyzed by phosphatases. When the kinase is active it can activate other downstream kinases of the next stage. The initial activation is often due to an external stimulus. The response of the whole cascade is the amount of activated substrate after the final stage, typically measured as a function of the initial stimulus.

Here the MAPK pathway is modeled with three stages of kinases MAPK, MAPKK, and MAPKKK. The initial stimulus is called S and the final substrate to be activated is a transcription factor TF. The

ground truth reaction network consists of activation/phosphorylation reactions

$$\begin{aligned}
\mathrm{S} + \mathrm{MAPKKK} &\longrightarrow \mathrm{S} + \mathrm{MAPKKK}* \\
\mathrm{MAPKKK}* + \mathrm{MAPKK} &\longrightarrow \mathrm{MAPKKK}* + \mathrm{MAPKK}* \\
\mathrm{MAPKK}* + \mathrm{MAPK} &\longrightarrow \mathrm{MAPKK}* + \mathrm{MAPK}* \\
\mathrm{MAPK}* + \mathrm{TF} &\longrightarrow \mathrm{MAPK}* + \mathrm{TF}*
\end{aligned}$$

and deactivation/dephosphorylation reactions

$$\begin{aligned}
\mathrm{MAPKKK}* &\longrightarrow \mathrm{MAPKKK} \\
\mathrm{MAPKK}* &\longrightarrow \mathrm{MAPKK} \\
\mathrm{MAPK}* &\longrightarrow \mathrm{MAPK} \\
\mathrm{TF}* &\longrightarrow \mathrm{TF}.
\end{aligned}$$

For simplicity we assume phosphatase to be abundant such that deactivations effectively become first order reactions. The external stimulus S is not consumed such that time integration of these reactions yields a steady state in which the response, i.e., the concentration [TF $*$] can be measured as a function of the stimulus concentration [S]. Using the rate constants given in Tab. 5.3 we obtain the response curve given in Fig. 5.6a.

We generate concentration time series data of the MAPK reactions above at three different initial conditions, each differing in the amount of stimulus [S]. The response yielded by the chosen initial conditions is marked in Fig. 5.6a by vertical dashed lines. The concatenated time series is a dataset of 300 frames in total. We use the library Θ of ansatz reactions Tab. 5.3. The hyperparameter $\alpha = 6.6 \times 10^{-9}$ was determined by shuffling the data and performing 15-fold cross validation. The estimated rate constants were obtained by solving the minimization problem (5.10) with $\lambda = 1$. The results are given in Fig. 5.6b. Least-squares estimation detects 5 of the 8 reaction processes that belong to the ground truth model. However it also detects 12 spurious reaction processes (θ_{18} - θ_{29}). Reactive SINDy estimation detects all reactions of the ground truth, two processes (θ_4 and θ_8) show deviations in rate constants. Generally reactive SINDy yields a sparse model which allows further simplification of the reaction network by dropping out reaction processes that lie beneath a certain cutoff. In this case for example a cutoff of $\kappa = 0.25$ would directly recover the ground truth reaction network. Quantitatively, one may consider the L_1 norm of the relative distance of estimated rate constants $\hat{\xi}_r$ to the non-zero rate constants of the ground truth ξ_r

$$\sum_{r=1}^{8} \left| (\hat{\xi}_r - \xi_r)/\xi_r \right|$$

which yields 167% error for least-squares and 21% error for reactive SINDy.

5.3.5 Application to Lotka–Volterra system

As biological pathways often exhibit oscillatory behavior [IZ08] which can stem from positive or negative feedback loops [Shi+09] we apply reactive SINDy to an idealized oscillatory system, namely the Lotka–Volterra system. The predator-prey dynamics of two species A (prey) and B (predator) is defined by the

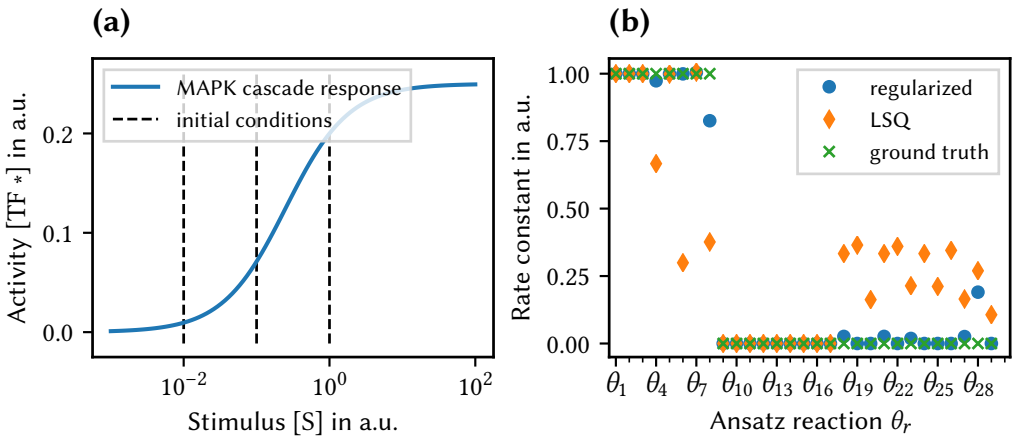

Figure 5.6: Application of reactive SINDy to the MAPK pathway system. **(a)** The response curve of the MAPK cascade as a function of external stimulus given as a constant concentration [S]. The activity is the steady state concentration of activated transcription factors [TF *]. Dashed lines show the values of [S] at which concentration time series data was generated. **(b)** Estimated rate coefficients of candidate reactions (see Tab. 5.3) after application of reactive SINDy (regularized) to the time series data. Least-squares estimation (LSQ) and the ground truth model for comparison. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

reaction network

$$\begin{aligned} \mathrm{A} &\longrightarrow \mathrm{A} + \mathrm{A} && \text{(prey growth)},\\ \mathrm{A} + \mathrm{B} &\longrightarrow \mathrm{B} + \mathrm{B} && \text{(predator eats prey)},\\ \mathrm{B} &\longrightarrow \emptyset && \text{(predator decay)},\\ \mathrm{A} + \mathrm{A} &\longrightarrow \emptyset && \text{(prey friction)},\\ \mathrm{B} + \mathrm{B} &\longrightarrow \emptyset && \text{(predator friction)}. \end{aligned}$$

From this model we generated concentration time series data with 200 frames which is displayed in Fig. 5.7a. The library of ansatz reactions Θ is given in Tab. 5.4. The hyperparameter $\alpha = 2.7 \times 10^{-7}$ was determined by shuffling the data and performing 5-fold cross validation. The estimated rate constants were obtained by solving the minimization problem (5.10) with $\lambda = 1$. The results are depicted in Fig. 5.7b. Least-squares estimation detects all reactions of the ground truth model but also two spurious processes (θ_6 and θ_7) with a higher rate than the first two underlying processes (θ_1 and θ_2). Reactive SINDy recovers the true reaction network with minor deviations in rate constants. As in Sec. 5.3.4, considering the L_1 norm of the relative distance to the ground truth for non-zero rate constants

$$\sum_{r=1}^{5} \left| (\hat{\xi}_r - \xi_r)/\xi_r \right|$$

yields 75% error for least-squares and 7% error for reactive SINDy.

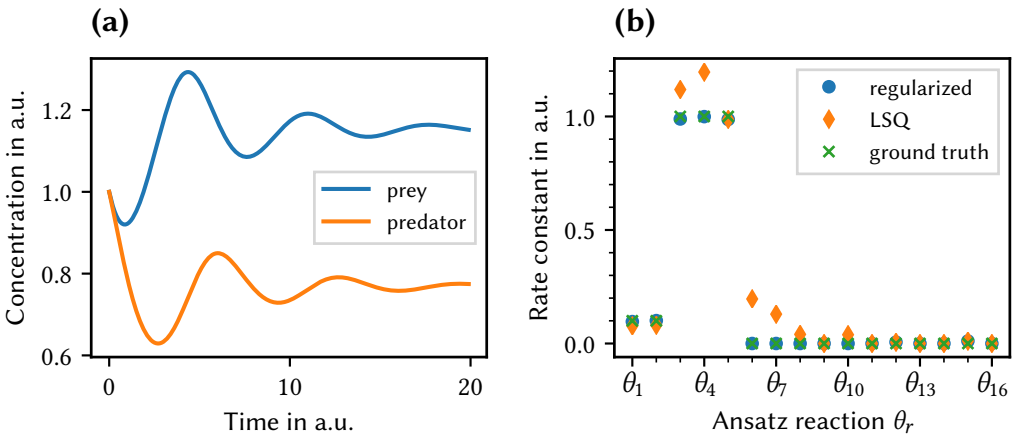

Figure 5.7: Application of reactive SINDy to the Lotka–Volterra system with social friction. **(a)** Concentration data as a function of time for predator and prey species. **(b)** Estimated rate coefficients of candidate reactions (see Tab. 5.4) after application of reactive SINDy (regularized) to the time series data. Least-squares estimation (LSQ) estimation and the ground truth model for comparison. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

5.4 Conclusion

In this work we have extended the SINDy method to reactive SINDy, not only parsimoniously detecting potentially nonlinear terms in a dynamical system from noisy data, but also yielding, in this case, a sparse set of rates with respect to generating reactions (5.8). Mathematically this has been achieved by permitting vector-valued basis functions and obtaining a tensor linear regression problem. We have applied this method on data generated from a gene regulation network, a MAPK pathway, and a Lotka–Volterra system and could successfully recover the underlying reaction networks.

The studies of Sec. 5.3.2 and Sec. 5.3.3 have shown that the applied regularization terms can mitigate noise up to a certain degree compared to the unregularized method, so that identification of the reaction network is more robust and closer to the ground truth. Potentially, this method could be used to identify reaction networks from time series measurements even if the initial conditions are not always exactly identical, as was demonstrated in Sec. 5.3.3.

One apparent limitation is that the method can only be applied if the data stems from the equilibration phase, as the concentration-based approach has derivatives equal zero in the equilibrium, which precludes the reaction dynamics to be recovered. Thus, in the case of oscillatory systems the reaction network can be recovered robustly.

In future work, we will consider the identification of reaction schemes from instantaneous fluctuations of particle numbers in equilibrium.

References

[Abr+04] Rinat Abramovitch et al. "A Pivotal Role of Cyclic AMP-Responsive Element Binding Protein in Tumor Progression". In: *Cancer Research* 64.4 (2004), pp. 1338–1346. DOI: 10.1158/0008-5472.CAN-03-2089.

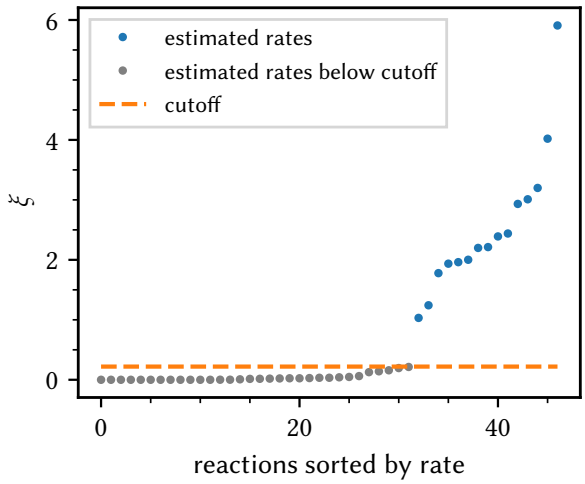

Figure 5.8: Reaction rates sorted by their magnitude to determine the cutoff $\kappa = 0.22$ of Sec. 5.3.1. The rates were estimated using the regularized minimization problem. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

[And17] Steven S. Andrews. "Smoldyn: Particle-based simulation with rule-based modeling, improved molecular interaction and a library interface". In: *Bioinformatics* 33.5 (2017), pp. 710–717. DOI: 10.1093/bioinformatics/btw700.

[And18] Steven S. Andrews. "Particle-Based Stochastic Simulators". In: *Encyclopedia of Computational Neuroscience*. Ed. by Dieter Jaeger and Ranu Jung. New York, NY: Springer, 2018, pp. 1–5. DOI: 10.1007/978-1-4614-7320-6_191-2.

[Ber78] Otto G. Berg. "A model for the statistical fluctuations of proteins numbers in a microbial population". In: *Journal of Theoretical Biology* 71.4 (1978), pp. 587–603. DOI: 10.1016/0022-5193(78)90326-0.

[BPK15] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Discovering governing equations from data by sparse identification of nonlinear dynamical systems". In: *Proceedings of the National Academy of Sciences* 113.15 (2015), pp. 3932–3937. DOI: 10.1073/pnas.1517384113.

[BPK16] Steven L. Brunton, Joshua L. Proctor, and J. Nathan Kutz. "Sparse identification of nonlinear dynamics with control (SINDYc)". In: *IFAC-PapersOnLine* 49.18 (2016), pp. 710–715. DOI: 10.1016/j.ifacol.2016.10.249.

[Con+99] Peijun Cong et al. "High-Throughput Synthesis and Screening of Combinatorial Heterogeneous Catalyst Libraries". In: *Angewandte Chemie International Edition* 38.4 (1999), pp. 483–488. DOI: 10.1002/(SICI)1521-3773(19990215)38:4<483::AID-ANIE483>3.0.CO;2-\#.

[DYK18] Aleksandar Donev, Chiao-yu Yang, and Changho Kim. "Efficient reactive Brownian dynamics". In: *The Journal of Chemical Physics* 148.3 (2018), p. 034103. DOI: 10.1063/1.5009464.

[Elo02] M. B. Elowitz. "Stochastic Gene Expression in a Single Cell". In: *Science* 297.5584 (2002), pp. 1183–1186. DOI: 10.1126/science.1070919.

[FN18] Christoph Fröhner and Frank Noé. "Reversible Interacting-Particle Reaction Dynamics". In: *The Journal of Physical Chemistry B* 122.49 (2018), pp. 11240–11250. DOI: 10.1021/acs.jpcb.8b06981.

[Gam+16] Socorro Gama-Castro et al. "RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond". In: *Nucleic Acids Research* 44.D1 (2016), pp. D133–D143. DOI: 10.1093/nar/gkv1156.

[Gil76] Daniel T. Gillespie. "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions". In: *Journal of Computational Physics* 22.4 (1976), pp. 403–434. DOI: 10.1016/0021-9991(76)90041-3.

[Gil77] Daniel T. Gillespie. "Exact stochastic simulation of coupled chemical reactions". In: *The Journal of Physical Chemistry* 81.25 (1977), pp. 2340–2361. DOI: 10.1021/j100540a008.

[Gol+05] Ido Golding et al. "Real-time kinetics of gene activity in individual bacteria". In: *Cell* 123.6 (2005), pp. 1025–1036. DOI: 10.1016/j.cell.2005.09.031.

[HFN19a] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "Reactive SINDy: Discovering governing reactions from concentration data". In: *The Journal of chemical physics* 150.2 (2019), p. 025101. DOI: 10.1063/1.5066099.

[HFN19b] Moritz Hoffmann, Christoph Fröhner, and Frank Noé. "ReaDDy 2: Fast and flexible software framework for interacting-particle reaction dynamics". In: *PLoS Computational Biology* 15.2 (2019), e1006830. DOI: 10.1371/journal.pcbi.1006830.

[HK70] A. E. Hoerl and R. W. Kennard. "Ridge Regression: Biased Estimation for Nonorthogonal Problems". In: *Technometrics* 12 (1970), pp. 55–67. DOI: 10.1080/00401706.1970.10488634.

[HTF09] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer New York, 2009. DOI: 10.1007/978-0-387-84858-7_2.

[Isa09] Samuel A. Isaacson. "The Reaction-Diffusion Master Equation as an Asymptotic Approximation of Diffusion to a Small Target". In: *SIAM Journal on Applied Mathematics* 70.1 (2009), pp. 77–111. DOI: 10.1137/070705039.

[IZ08] Jameel Iqbal and Mone Zaidi. "TNF-induced MAP kinase activation oscillates in time". In: *Biochemical and biophysical research communications* 371.4 (2008), pp. 906–911. DOI: 10.1016/j.bbrc.2008.03.113.

[JOP+01] Eric Jones, Travis Oliphant, Pearu Peterson, et al. *SciPy: Open source scientific tools for Python*. 2001–.

[KCG05] Walter Kolch, Muffy Calder, and David Gilbert. "When kinases meet mathematics: the systems biology of MAPK signalling". In: *FEBS letters* 579.8 (2005), pp. 1891–1895. DOI: 10.1016/j.febslet.2005.02.002.

[KR05] Lioubov Kiwi-Minsker and Albert Renken. "Microstructured reactors for catalytic reactions". In: *Catalysis today* 110.1-2 (2005), pp. 2–14. DOI: 10.1016/j.cattod.2005.09.011.

[Kra88] Dieter Kraft. "A software package for sequential quadratic programming". In: *Technical Report DFVLR-FB 88-28, Institut für Dynamik der Flugsysteme, Oberpfaffenhofen* (1988).

[Kuh+07] T. Kuhlman et al. "Combinatorial transcriptional control of the lactose operon of Escherichia coli". In: *Proceedings of the National Academy of Sciences* 104.14 (2007), pp. 6043–6048. DOI: 10.1073/pnas.0606717104.

[Lee02] T. I. Lee. "Transcriptional Regulatory Networks in Saccharomyces cerevisiae". In: *Science* 298.5594 (2002), pp. 799–804. DOI: 10.1126/science.1075090.

[Lew+96] Mitchell Lewis et al. "Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer". In: *Science* 271.5253 (1996), pp. 1247–1254. DOI: 10.1126/science.271.5253.1247.

[Man+16] Niall M. Mangan et al. "Inferring biological networks by sparse identification of nonlinear dynamics". In: *IEEE Transactions on Molecular, Biological and Multi-Scale Communications* 2.1 (2016), pp. 52–63. DOI: 10.1109/TMBMC.2016.2633265.

[Mar+06] Adam A. Margolin et al. "ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context". In: *BMC Bioinformatics* 7.Suppl 1 (2006), S7. DOI: 10.1186/1471-2105-7-S1-S7.

[Ort+05] Richard J. Orton et al. "Computational modelling of the receptor-tyrosine-kinase-activated MAPK pathway". In: *Biochemical Journal* 392.2 (2005), pp. 249–261. DOI: 10.1042/BJ20050908.

[Pan+12] Wei Pan et al. "Reconstruction of arbitrary biochemical reaction networks: A compressive sensing approach". In: *2012 IEEE 51st IEEE Conference on Decision and Control (CDC)*. IEEE, 2012, pp. 2334–2339. DOI: 10.1109/CDC.2012.6426216.

[PT17] Yannis Pantazis and Ioannis Tsamardinos. "A unified approach for sparse dynamical system inference from temporal measurements". In: *Bioinformatics* 35.18 (2017). Ed. by Oliver Stegle, pp. 3387–3396. DOI: 10.1093/bioinformatics/btz065.

[Qua+18] Markus Quade et al. "Sparse identification of nonlinear dynamics for rapid model recovery". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 28.6 (2018), p. 063116. DOI: 10.1063/1.5027470.

[Roa+17] Rafael Roa et al. "Catalyzed Bimolecular Reactions in Responsive Nanoreactors". In: *ACS Catalysis* 7.9 (2017), pp. 5604–5611. DOI: 10.1021/acscatal.7b01701.

[Rud+17] Samuel H. Rudy et al. "Data-driven discovery of partial differential equations". In: *Science Advances* 3.4 (2017), e1602614. DOI: 10.1126/sciadv.1602614.

[Ryu+18] Hyunryul Ryu et al. "Integrated Platform for Monitoring Single-cell MAPK Kinetics in Computer-controlled Temporal Stimulations". In: *Scientific reports* 8.1 (2018), p. 11126. DOI: 10.1038/s41598-018-28873-1.

[Sch+14] Johannes Schöneberg et al. "Explicit Spatiotemporal Simulation of Receptor-G Protein Coupling in Rod Cell Disk Membranes". In: *Biophysical Journal* 107.5 (2014), pp. 1042–1053. DOI: 10.1016/j.bpj.2014.05.050.

[She+02] Shai S. Shen-Orr et al. "Network motifs in the transcriptional regulation network of Escherichia coli". In: *Nature Genetics* 31.1 (2002), pp. 64–68. DOI: 10.1038/ng881.

[Shi+09] Sung-Young Shin et al. "Positive-and negative-feedback regulations coordinate the dynamic behavior of the Ras-Raf-MEK-ERK signal transduction pathway". In: *Journal of cell science* 122.3 (2009), pp. 425–435. DOI: 10.1242/jcs.036319.

[SN13] Johannes Schöneberg and Frank Noé. "ReaDDy–a software for particle-based reaction-diffusion dynamics in crowded cellular environments." In: *PLoS One* 8.9 (2013), e74261. DOI: 10.1371/journal.pone.0074261.

[SUN14] Johannes Schöneberg, Alexander Ullrich, and Frank Noé. "Simulation tools for particle-based reaction-diffusion dynamics in continuous space". In: *BMC Biophysics* 7.1 (2014), p. 11. DOI: 10.1186/s13628-014-0011-5.

[Tia+07] Tianhai Tian et al. "Simulated maximum likelihood method for estimating kinetic rates in gene expression". In: *Bioinformatics* 23.1 (2007), pp. 84–91. DOI: 10.1093/bioinformatics/btl552.

[Tib96] Robert Tibshirani. *Regression Selection and Shrinkage via the Lasso*. 1996. DOI: 10.2307/2346178.

[TO01] Mukund Thattai and A. van Oudenaarden. "Intrinsic noise in gene regulatory networks". In: *Proceedings of the National Academy of Sciences* 98.15 (2001), pp. 8614–8619. DOI: 10.1073/pnas.151588598.

[WS16] Stefanie Winkelmann and Christof Schütte. "The spatiotemporal master equation: Approximation of reaction-diffusion dynamics via Markov state modeling". In: *The Journal of Chemical Physics* 145.21 (2016), p. 214107. DOI: 10.1063/1.4971163.

[WS17] Stefanie Winkelmann and Christof Schütte. "Hybrid models for chemical reaction networks: Multiscale theory and application to gene regulatory systems". In: *The Journal of Chemical Physics* 147.11 (2017), p. 114115. DOI: 10.1063/1.4986560.

[XGG96] Jun Xing, David D. Ginty, and Michael E. Greenberg. "Coupling of the RAS-MAPK Pathway to Gene Activation by RSK2, a Growth Factor-Regulated CREB Kinase". In: *Science* 273.5277 (1996), pp. 959–963. DOI: 10.1126/science.273.5277.959.

[YM03] Necmettin Yildirim and Michael C. Mackey. "Feedback Regulation in the Lactose Operon: A Mathematical Modeling Study and Comparison with Experimental Data". In: *Biophysical Journal* 84.5 (2003), pp. 2841–2851. DOI: 10.1016/S0006-3495(03)70013-7.

[ZH05] Hui Zou and Trevor Hastie. "Regularization and variable selection via the elastic-net". In: *Journal of the Royal Statistical Society* 67.2 (2005), pp. 301–320. DOI: 10.1111/j.1467-9868.2005.00503.x.

[ZL02] Wei Zhang and Hui Tu Liu. "MAPK signal pathways in the regulation of cell proliferation in mammalian cells". In: *Cell research* 12.1 (2002), p. 9. DOI: 10.1038/sj.cr.7290105.

[ZS18] Linan Zhang and Hayden Schaeffer. "On the Convergence of the SINDy Algorithm". In: *Multiscale Modeling & Simulation* 17.3 (2018), pp. 948–972. DOI: 10.1137/18m1189828.

[ZW05a] Jeroen S. van Zon and Pieter Rein ten Wolde. "Green's-function reaction dynamics: a particle-based approach for simulating biochemical networks in time and space". In: *The Journal of Chemical Physics* 123.23 (2005), p. 234910. DOI: 10.1063/1.2137716.

[ZW05b] Jeroen S. van Zon and Pieter Rein ten Wolde. "Simulating biochemical networks at the particle level and in time and space: Green's function reaction dynamics". In: *Physical Review Letters* 94.12 (2005), p. 128103. DOI: 10.1103/PhysRevLett.94.128103.

Reaction			rate	description
DNA_A	$\rightharpoonup$	$DNA_A + mRNA_A$	$k_1 = 1.8$	transcription of $mRNA_A$
$mRNA_A$	$\rightharpoonup$	$mRNA_A + A$	$k_2 = 2.1$	translation of A proteins
$mRNA_A$	$\rightharpoonup$	$\emptyset$	$k_3 = 1.3$	$mRNA_A$ decay
A	$\rightharpoonup$	$\emptyset$	$k_4 = 1.5$	decay of A proteins
DNA_B	$\rightharpoonup$	$DNA_B + mRNA_B$	$k_5 = 2.2$	transcription of $mRNA_B$
$mRNA_B$	$\rightharpoonup$	$mRNA_B + B$	$k_6 = 2.0$	translation of B proteins
$mRNA_B$	$\rightharpoonup$	$\emptyset$	$k_7 = 2.0$	$mRNA_B$ decay
B	$\rightharpoonup$	$\emptyset$	$k_8 = 2.5$	decay of B proteins
DNA_C	$\rightharpoonup$	$DNA_C + mRNA_C$	$k_9 = 3.2$	transcription of $mRNA_C$
$mRNA_C$	$\rightharpoonup$	$mRNA_C + C$	$k_{10} = 3.0$	translation of C proteins
$mRNA_C$	$\rightharpoonup$	$\emptyset$	$k_{11} = 2.3$	$mRNA_C$ decay
C	$\rightharpoonup$	$\emptyset$	$k_{12} = 2.5$	decay of C proteins
$mRNA_A + A$	$\rightharpoonup$	A	$k_{13} = 0$	self regulation of A proteins
$mRNA_B + B$	$\rightharpoonup$	B	$k_{14} = 0$	self regulation of B proteins
$mRNA_C + C$	$\rightharpoonup$	C	$k_{15} = 0$	self regulation of C proteins
$mRNA_B + A$	$\rightharpoonup$	A	$k_{16} = 0$	regulation of $mRNA_B$
$mRNA_C + B$	$\rightharpoonup$	B	$k_{17} = 0$	regulation of $mRNA_C$
$mRNA_A + C$	$\rightharpoonup$	C	$k_{18} = 0$	regulation of $mRNA_A$
$mRNA_C + A$	$\rightharpoonup$	A	$k_{16} = 6.0$	regulation of $mRNA_C$
$mRNA_B + C$	$\rightharpoonup$	C	$k_{17} = 4.0$	regulation of $mRNA_B$
$mRNA_A + B$	$\rightharpoonup$	B	$k_{18} = 3.0$	regulation of $mRNA_A$
$mRNA_A + A$	$\rightharpoonup$	$mRNA_A$	$k_{19} = 0$	artificial fusion
$mRNA_B + B$	$\rightharpoonup$	$mRNA_B$	$k_{20} = 0$	artificial fusion
$mRNA_A + B$	$\rightharpoonup$	$mRNA_A$	$k_{21} = 0$	artificial fusion
$mRNA_B + C$	$\rightharpoonup$	$mRNA_B$	$k_{22} = 0$	artificial fusion
$mRNA_C + A$	$\rightharpoonup$	$mRNA_C$	$k_{23} = 0$	artificial fusion
$mRNA_A + C$	$\rightharpoonup$	$mRNA_A$	$k_{24} = 0$	artificial fusion
$mRNA_B + A$	$\rightharpoonup$	$mRNA_B$	$k_{25} = 0$	artificial fusion
A + A	$\rightharpoonup$	A	$k_{26} = 0$	A regulates A
B + B	$\rightharpoonup$	B	$k_{27} = 0$	B regulates B
C + C	$\rightharpoonup$	C	$k_{28} = 0$	C regulates C
B + A	$\rightharpoonup$	A	$k_{29} = 0$	artificial fusion
C + B	$\rightharpoonup$	B	$k_{30} = 0$	artificial fusion
A + C	$\rightharpoonup$	C	$k_{31} = 0$	artificial fusion
C + A	$\rightharpoonup$	A	$k_{32} = 0$	artificial fusion
B + C	$\rightharpoonup$	C	$k_{33} = 0$	artificial fusion
A + B	$\rightharpoonup$	B	$k_{34} = 0$	artificial fusion
A	$\rightharpoonup$	B	$k_{35} = 0$	artificial conversion
B	$\rightharpoonup$	C	$k_{36} = 0$	artificial conversion
C	$\rightharpoonup$	A	$k_{37} = 0$	artificial conversion
A	$\rightharpoonup$	C	$k_{38} = 0$	artificial conversion
C	$\rightharpoonup$	B	$k_{39} = 0$	artificial conversion
B	$\rightharpoonup$	A	$k_{40} = 0$	artificial conversion
$mRNA_B + mRNA_C$	$\rightharpoonup$	$mRNA_A$	$k_{41} = 0$	artificial fusion
$mRNA_C + mRNA_B$	$\rightharpoonup$	$mRNA_C$	$k_{42} = 0$	artificial fusion
$mRNA_C + A$	$\rightharpoonup$	C	$k_{43} = 0$	artificial fusion

Table 5.2: Full set of ansatz reactions Θ used in Sec. 5.3 for the gene-regulatory network. The given rate constants define the ground truth reaction model. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

Reaction			rate	description
S + MAPKKK	$\longrightarrow$	S + MAPKKK $*$	$k_1 = 1$	external stimulus activates MAPKKK
MAPKKK $*$	$\longrightarrow$	MAPKKK	$k_2 = 1$	dephosphorylation
MAPKKK $*$ + MAPKK	$\longrightarrow$	MAPKKK $*$ + MAPKK $*$	$k_3 = 1$	phosphorylation of MAPKK
MAPKK $*$	$\longrightarrow$	MAPKK	$k_4 = 1$	dephosphorylation
MAPKK $*$ + MAPK	$\longrightarrow$	MAPKK $*$ + MAPK $*$	$k_5 = 1$	phosphorylation of MAPK
MAPK $*$	$\longrightarrow$	MAPK	$k_6 = 1$	dephosphorylation
MAPK $*$ + TF	$\longrightarrow$	MAPK $*$ + TF $*$	$k_7 = 1$	phosphorylation of transcription factor
TF $*$	$\longrightarrow$	TF	$k_8 = 1$	dephosphorylation
MAPKKK + MAPKK	$\longrightarrow$	MAPKKK + MAPKK $*$	$k_9 = 0$	artificial reaction
MAPKKK + MAPK	$\longrightarrow$	MAPK $*$	$k_{10} = 0$	artificial reaction
MAPKKK + TF	$\longrightarrow$	MAPKKK + TF $*$	$k_{11} = 0$	artificial reaction
MAPKKK $*$ + MAPK	$\longrightarrow$	MAPKKK $*$ + MAPK $*$	$k_{12} = 0$	artificial reaction
MAPKKK $*$ + TF	$\longrightarrow$	MAPKKK $*$ + TF $*$	$k_{13} = 0$	artificial reaction
MAPKK + TF	$\longrightarrow$	MAPKK + TF $*$	$k_{14} = 0$	artificial reaction
MAPKK $*$ + TF	$\longrightarrow$	MAPKK $*$ + TF $*$	$k_{15} = 0$	artificial reaction
MAPK + TF	$\longrightarrow$	MAPK + TF $*$	$k_{16} = 0$	artificial reaction
MAPKK + MAPK	$\longrightarrow$	MAPKK + MAPK $*$	$k_{17} = 0$	artificial reaction
MAPKKK + MAPKK $*$	$\longrightarrow$	MAPKKK + MAPKK	$k_{18} = 0$	artificial reaction
MAPKKK + MAPK $*$	$\longrightarrow$	MAPKKK + MAPK	$k_{19} = 0$	artificial reaction
MAPKKK + TF $*$	$\longrightarrow$	MAPKKK + TF	$k_{20} = 0$	artificial reaction
MAPKKK $*$ + MAPKK $*$	$\longrightarrow$	MAPKKK $*$ + MAPKK	$k_{21} = 0$	artificial reaction
MAPKKK $*$ + MAPK $*$	$\longrightarrow$	MAPKKK $*$ + MAPK	$k_{22} = 0$	artificial reaction
MAPKKK $*$ + TF $*$	$\longrightarrow$	MAPKKK $*$ + TF	$k_{23} = 0$	artificial reaction
MAPKK + MAPK $*$	$\longrightarrow$	MAPKK + MAPK	$k_{24} = 0$	artificial reaction
MAPKK + TF $*$	$\longrightarrow$	MAPKK + TF	$k_{25} = 0$	artificial reaction
MAPKK $*$ + MAPK $*$	$\longrightarrow$	MAPKK $*$ + MAPK	$k_{26} = 0$	artificial reaction
MAPKK $*$ + TF $*$	$\longrightarrow$	MAPKK $*$ + TF	$k_{27} = 0$	artificial reaction
MAPK + TF $*$	$\longrightarrow$	MAPK + TF	$k_{28} = 0$	artificial reaction
MAPK $*$ + TF $*$	$\longrightarrow$	MAPK $*$ + TF	$k_{29} = 0$	artificial reaction

Table 5.3: Full set of ansatz reactions Θ used in Sec. 5.3.4 for the MAPK system. The given rate constants define the ground truth reaction model. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.

Reaction			rate	description
A + A	$\longrightarrow$	$\emptyset$	$k_1 = 0.1$	social friction of prey
B + B	$\longrightarrow$	$\emptyset$	$k_2 = 0.1$	social friction of predator
A	$\longrightarrow$	A + A	$k_3 = 1$	prey growth
A + B	$\longrightarrow$	B + B	$k_4 = 1$	predator eats prey
B	$\longrightarrow$	$\emptyset$	$k_5 = 1$	predator decays
A + B	$\longrightarrow$	A + A	$k_6 = 0$	artificial reaction
A	$\longrightarrow$	$\emptyset$	$k_7 = 0$	artificial reaction
B + B	$\longrightarrow$	B	$k_8 = 0$	artificial reaction
B	$\longrightarrow$	B + B	$k_9 = 0$	artificial reaction
A + A	$\longrightarrow$	A	$k_{10} = 0$	artificial reaction
A + B	$\longrightarrow$	A	$k_{11} = 0$	artificial reaction
A + B	$\longrightarrow$	B	$k_{12} = 0$	artificial reaction
A + A	$\longrightarrow$	B	$k_{13} = 0$	artificial reaction
A	$\longrightarrow$	B	$k_{14} = 0$	artificial reaction
B	$\longrightarrow$	A	$k_{15} = 0$	artificial reaction
A	$\longrightarrow$	B + B	$k_{16} = 0$	artificial reaction

Table 5.4: Full set of ansatz reactions Θ used in Sec. 5.3.5 for the Lotka–Volterra system. The given rate constants define the ground truth reaction model. Reprinted from *The Journal of chemical physics* "Reactive SINDy: Discovering governing reactions from concentration data", Hoffmann, Fröhner, and Noé, 2019, with the permission of AIP Publishing.