UC Berkeley
UC Berkeley Electronic Theses and Dissertations

Title
Variational sampling and optimal design of rare nonequilibrium molecular dynamics

Permalink
https://escholarship.org/uc/item/3gm562gw

Author
Das, Avishek

Publication Date
2022

Peer reviewed|Thesis/dissertation

Variational sampling and optimal design of rare nonequilibrium molecular dynamics

by

Avishek Das

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David T. Limmer, Chair
Professor Phillip L. Geissler
Professor Ehud Altman

Spring 2022

Variational sampling and optimal design of rare nonequilibrium molecular dynamics

Abstract

Variational sampling and optimal design of rare nonequilibrium molecular dynamics

by

Avishek Das

Doctor of Philosophy in Chemistry

University of California, Berkeley

Professor David T. Limmer, Chair

Nonequilibrium driving can independently tune the structure and dynamics of molecular and colloidal systems, resulting in novel assembly and transport phenomena that are not found in equilibrium. Exploring nonequilibrium dynamics through numerical simulations is crucial towards understanding the working mechanisms of functional inorganic or biological materials. However, behavior of dynamical observables in such systems at experimental timescales is often dominated by statistically rare fluctuations that are poorly sampled in simulated trajectories. Rare event sampling algorithms that assume a Boltzmann distribution of configurations and detailed balance for dynamics cannot be applied out of thermal equilibrium, thus limiting available techniques that can efficiently sample nonequilibrium rare events.

In this thesis we have developed novel variational algorithms for the sampling and design of rare nonequilibrium molecular dynamics trajectories by application of an optimized driving force. This approach relies on the equivalence of a trajectory ensemble conditioned on a rare event to occur to an ensemble driven with the optimal force where the rare event occurs typically. For systems with many interacting degrees of freedom, we numerically learn the optimal force within arbitrary basis sets by statistically estimating explicit gradients of trajectory probability. This method allows us to efficiently compute large deviation functions of dynamical observables in nonequilibrium steady-states, and to automate the inverse design of self-assembling colloids and molecular machines for desired structure and dynamics. We have finally augmented our approach with reinforcement learning techniques, resulting in a new paradigm to efficiently compute nonequilibrium reaction rates, Variational Path Sampling. Our approach of using optimized forces to improve sampling of nonequilibrium trajectories is versatile and can give access to rare reactive fluctuations and dynamical phases that cannot be sampled otherwise.

To

Ashok and Didi

Contents

Acknowledgments

The past five years have been simultaneously the most challenging and happiest in my life so far. I owe my identity as a person and a researcher today to the patience, love and support from many, many people who believed in me, inspired me and dreamed for me. I am grateful to be able to thank them here.

First and foremost, I will thank my adviser David Limmer without whom this dissertation would not exist. I have learnt most of my 'Chandlerian statistical mechanics' from his brilliant instruction and from teaching and working with him. His endless patience and willingness to help with hands-on research, deep technical intuition and foresight, scientific creativity, and relentless enthusiasm about all things chemistry have made working with him extraordinary fun. David has actively encouraged my exploration of truly fundamental questions in statistical mechanics by shielding me from non-academic logistical constraints and giving me a lot of independence, for which I am thankful. He has diligently taught me how to systematize the process of abstract scientific inquiry into careful experimentation, analysis, inference and publication, in short, how to be a practicing scientist. I very much look forward to many more years of enjoyable scientific discussions with David in the future.

Many other scientists and students at Berkeley and at other universities contributed greatly to my development as a researcher. I am thankful to my collaborators Naomi Ginsberg, Dmitri Talapin and Sam Teitelbaum for our many interesting conversations about designing and detecting the self-assembly of charged nanocrystals. I have learnt an incredible amount and had a lot of fun from discussing this problem with Christian Tanner, James Utterback, Vivian Wall, Igor Coropceanu and Josh Portner. I have also greatly enjoyed exploring the connections between reinforcement learning and importance sampling in my collaboration with Juan Garrahan and Dominic Rose. The wealth of knowledge I have gained from them on the statistical physics of machine learning paved the way for some of the most crucial developments in this dissertation. I am also grateful to Robert Jack and Hugo Touchette for encouragingly supporting my research on large deviation sampling in its early stages, engaging with it over the years and providing many valuable comments all through. Similarly I am indebted to Phillip Geissler and Eric Neuscamman for discussions on the physics of trajectory sampling and variational algorithms, and to Phill and Ehud Altman for a critical reading of this dissertation. Not only have I intellectually and scientifically benefited from all my interactions with these leading researchers, but they have been inspirational role-models to me on how to practice fundamental science.

Getting to this stage today has taken many leaps of faith from past mentors who trusted and indulged my academic potential. I would like to thank my parents first for bringing me up in an intellectual and inquisitive environment and for nurturing in me a love for books and knowledge. I am thankful for the memorable courses by my undergraduate professors E. D. Jemmis, Vijay Shenoy and D. D. Sarma, all of whom showed me how fascinating the study of 'more is different' in the chemistry and physics of condensed matter could be. I also owe my sincere gratitude to Prof. Gopalan Rajaraman from IIT Bombay, with whom I did my first theoretical chemistry project and fell in love with the power and reductionist

host and housemate during my time here in Berkeley, and for being such an inspirational role-model for identifying what the important things in life are. I thank my comrades in our academic student workers' unions UAW-2865 and UAW-SRU for having ensured strong workplace contracts that have given me the luxury of being a fulfilled researcher in the first place. And I am thankful for the many wonderful queer friends I have made in the bay area, who have had enormous roles in shaping me as the human being I am today.

Finally, I would like to thank the most important people in my life, who have showed me the meaning of unconditional love. My sister Jayita and brother-in-law Kaushik have been my pillars of support through the many arduous times we have together gone through. Words fail to describe how lucky I feel to have had them as supportive family. And I thank Ashok, my partner and my best friend, for being the sweetest, funniest, most extraordinary person I have ever known and for being there for me every single day through grad school. Some of my happiest moments in the past few years I have spent with him, either virtually or in-person. I am incredibly excited to see both of us grow together in the coming years.

All of the above still feel like massive understatements on the amount of help and support I have actually had from people. Kindness, love and inspiration from friends and strangers have sustained me through graduate school that resulted in this dissertation. I sincerely hope I am able to pay it forward.

Chapter 1

Introduction

1.1 Nonequilibrium rare events in physics and chemistry

The past few decades have seen rapid growth in the capability of computer simulations to address questions about the microscopic dynamics of diverse chemical and biological systems across multiple length and time scales. Computer simulations have been used to elucidate reaction pathways for physical processes like dissolution of a solute [1], for deciphering reaction mechanisms of atomic and molecular rearrangements in solvents [2], to understand energy, charge and mass transport through electrochemical cells, microfluidic devices and crystalline and amorphous solids [3, 4, 5, 6, 7], and to explore assembly and ordering phenomena in inorganic and biological materials [8, 9, 10]. Even in classical limits where the effects of quantum coherence are negligible, dynamics in a condensed phase is often collective, involving interactions of a given system of interest with thermal, mechanical or chemical baths that can inject noise and memory into the system's dynamics [11]. The complexity of such dynamics at the nanoscale makes any analytical treatment prohibitive, and hence conclusions about reaction mechanisms, transport properties and inverse design of materials are often drawn from computer simulations. Nanoscale dynamical information obtained from computer simulations are also extensively used to build coarse-grained effective models of materials, allowing a reduced description at longer length and time scales [12]. But inferring such high-level description from the microscopic dynamics of molecules faces the crucial challenge of sampling rare dynamical events.

Rare events in molecular simulations often give rise to complicated nonlinear dynamics that are of broad interest [13]. Molecular dynamics simulations involve integrating a known equation of motion, such as Newton's equations, in the presence of many degrees of freedom. For temporal integration we choose a quadrature timestep that is usually smaller than the fastest modes of motion in the system, that of atomic translations and molecular rotations in the sub-picosecond scale. However, condensed phase environments stabilize multiple metastable states in molecules, and crucial dynamics like the folding of a protein from

its secondary to tertiary structure, or the nucleation of a stable phase from a metastable one, occur at timescales many orders of magnitude longer than molecular motion [14, 15]. This is not due to the progress along these reaction coordinates being inherently slow, but arises from the presence of kinetic bottlenecks with narrow funnels of entry that are only rarely encountered. As a result, these dynamics are only observed in molecular simulations with a low probability even though they drastically alter the properties of the system on experimental timescales. Developing special numerical techniques to sample such rare but interesting events with correct statistical likelihood in molecular dynamics simulations has occupied much of the efforts of statistical physicists working in physical chemistry problems. A majority of these techniques assume the system to be in thermal equilibrium, which simplifies the fluctuations of observables due to the configurational distribution being Boltzmann. This however limits their applicability to driven systems that are ubiquitously encountered and exploited in functional chemistry and biology.

A system kept away from thermal equilibrium by hindering its relaxation or through a continuous supply of energy is subject to fewer physical constraints than one evolving within an equilibrium state. As a consequence, the application of external forces or the internal consumption of energy can produce structures and responses without equilibrium equivalent [16, 17, 18]. Driving a system dissipatively can decouple its dynamical constraints from its structure, and thus stabilize otherwise transient states, or generate dynamical fluxes between otherwise non-reactive structures. Nonequilibrium driving can assist molecular motors to transduce energy and move ions against thermodynamic gradients [19], stabilize clusters of motile bacteria without attractive interactions [20], and overcome thermodynamic limits on the selectivity of assembly and control [21, 22]. A majority of manufacturing processes in industrial chemistry, synthesis protocols for nanomaterials, geophysics and geochemistry of climate fluctuations, and biological reactions that govern living systems occur in nonequilibrium conditions [23, 24, 25, 26]. Advances in the theory and modeling of nonequilibrium steady-states, borrowing tools from probability theory and stochastic calculus [19, 27, 28], have resulted in an increased interest in trying to understand the behavior in systems out of equilibrium and leverage their versatility to design new functional materials [20, 29, 30, 31, 22, 32, 33]. Thermodynamic partition function formalism has been generalized to ensembles of trajectories via large deviation theory [34] and transport theories have been developed to quantify linear and higher-order response functions to nonequilibrium perturbations [35, 36]. The second law of thermodynamics has also been generalized to far-from-equilibrium dynamics through fluctuation theorems that relate fluctuations of nonequilibrium structure and dynamics to thermodynamic quantities like entropy or work [37, 38, 39], and thermodynamic uncertainty relations that always bound current fluctuations [40, 41].

Despite progress in these theoretical directions, quantifying emergent nonequilibrium behavior with computer simulations is currently hampered by the lack of robust tools to sample the rare fluctuations required to estimate response functions, overcome kinetic bottlenecks, and reach the timescales of experimental relevance. Commonly used numerical methods for sampling nonequilibrium dynamics rely on stratifying trajectory ensembles to build up rare events stepwise, but do not apply additional importance sampling forces that can gener-

ate inherently uncorrelated statistical realizations. In the present work we have developed variational techniques to sample and design rare dynamical fluctuations for a generic class of stochastic systems that violate detailed balance. Our approach relies on optimizing an additional driving force that reproduces the unbiased rare fluctuations of the unperturbed system. Our numerical methods have allowed us to learn this force in complex molecular and colloidal systems within arbitrary basis sets and efficiently extract the statistics and mechanism of nonequilibrium rare events. Chapter 1 briefly discusses the trajectory ensemble formalism to quantify rare events in transient and steady-state trajectories, and the backgrounds of large deviations and reaction rate theory that our formalism is based on. In Chapter 2 we develop a variational algorithm for the use of optimized control forces to sample rare current and activity fluctuations of particles in a nonequilibrium steady-state on a one-dimensional model potential. Chapter 3 applies this algorithm towards the automated discovery of inverse design principles for DNA-labeled colloids in a shear flow. In Chapter 4 we then develop more efficient optimization algorithms based on reinforcement learning tools to solve finite and infinite duration variational problems. Finally, Chapter 5 applies this approach to develop a new algorithm to compute reaction rates in far-from-equilibrium systems like active matter, that we call Variational Path Sampling (VPS).

1.2 Langevin dynamics and trajectory ensembles

We aim to temporally integrate the equations of motion at molecular and colloidal scales to simulate rare dynamics of systems of interest. Conventional molecular dynamics in an isolated system is described by Newton's equations of motion that are integrated with symplectic schemes that conserve energy, like the velocity Verlet algorithm [42]. However, driving a system with external forces performs work on it, and in order to relax into a steady-state, the excess energy must be discarded by the system as dissipation into a bath. Driven systems are thus often modeled as open systems where energy is not conserved, and where there is on average a finite dissipative flow between the system and the bath, resulting in nonequilibrium. Moreover, dynamics in a condensed phase can often be simplified by interpreting the interesting part of the system to be open and in contact with the rest of the system as a thermal bath, which only affects the interesting part through frictional forces and Brownian noise. This dynamics could be interpreted as a method to thermostat the system at a fixed temperature, or as an approximation to the noisy diffusive dynamics encountered by molecules and colloids suspended in liquid solvents or in contact with a large number of thermalized vibrational excitations. We will describe our nonequilibrium variational algorithms only in the context of this Langevin dynamics. Generalization to other kinds of stochastic dynamics is straightforward as long as every trajectory has a closed form nonzero probability.

We consider dynamics described by a Langevin equation of the form [11],

$$\dot{\mathbf{r}}_i(t) = \mathbf{v}_i(t) \tag{1.1}$$

$$m_i\dot{\mathbf{v}}_i(t) = \mathbf{F}_i[\mathbf{r}^N(t)] - \gamma_i\mathbf{v}_i(t) + \boldsymbol{\eta}_i(t) \tag{1.2}$$

where $\dot{\mathbf{r}}_i$ and $\dot{\mathbf{v}}_i$ are the rate of change of the i-th particle's position and velocity, γ_i is the corresponding friction coefficient, and $\mathbf{F}_i[\mathbf{r}^N(t)]$ is the sum of all conservative and nonconservative forces exerted on the i-th particle that depends on the full configuration of the N-particle system, $\mathbf{r}^N$. The final term, $\boldsymbol{\eta}_i(t)$, represents a thermal bath-induced white-noise, as a Gaussian random number with zero mean, $\langle \eta_{i\alpha}(t) \rangle = 0$, and delta-correlated in time,

$$\langle \eta_{i\alpha}(t)\eta_{j\beta}(t') \rangle = 2\gamma_i k_{\mathrm{B}}T\delta_{ij}\delta_{\alpha\beta}\delta(t-t') \tag{1.3}$$

for component (α, β), with $k_{\mathrm{B}}T$ being Boltzmann's constant times the temperature. The relation between the variance of the Gaussian white noise with the friction coefficient of the bath derives from the second fluctuation-dissipation theorem. Given the approximate diameter of a particle is σ and it is suspended in the bath with friction γ, the time taken to diffuse its own length is given by $t^* = \sigma^2/D$, where $D_i = k_{\mathrm{B}}T/\gamma$ is the diffusion constant. In most cases we will consider the high-friction limit of the Langevin equation, formally taken as $m/\gamma t^* \to 0$, where the particles lose their inertia and instead are governed by the overdamped Langevin equation

$$\gamma_i\dot{\mathbf{r}}_i(t) = \mathbf{F}_i[\mathbf{r}^N(t)] + \boldsymbol{\eta}_i(t) \tag{1.4}$$

The overdamped equation can model the dynamics of a wide range of colloidal systems including nanocrystals, proteins and polymers in a condensed phase with a sufficiently large bath. On the other hand, describing inertial dynamics like vibrations and hydrodynamic effects require using the full Langevin equation in an underdamped regime. We note that the underdamped equation could be considered formally to be a generalized overdamped equation for all dynamical coordinates, in the joint phase space of position and velocity, with the latter being in contact with a bath at zero friction. Hence we will do most discussions of numerical techniques developed here in the context of the overdamped Langevin equations without loss of generality. Similarly, generalizations to systems with state-dependent diffusion matrices are straightforward from our formalism.

The Langevin equation is simulated by discretizing time into integration timesteps. Given $\dot{\mathbf{r}}(t)dt$ is a stochastic differential, care must be taken with respect to discretizing time-integrals and connecting them to physical observables. We will follow in most cases the Ito formalism, where the integration of the dynamics is interpreted as

$$\gamma_i\mathbf{r}_i(t+\Delta t) = \gamma_i\mathbf{r}_i(t) + \mathbf{F}_i[\mathbf{r}^N(t)]\Delta t + \sqrt{2\gamma_i k_{\mathrm{B}}T\Delta t}\boldsymbol{\mathcal{N}}_i^t \tag{1.5}$$

where $\boldsymbol{\mathcal{N}}_i^t$ is a vector of independent Gaussian random numbers with zero mean and unit variance. Here each step taken in $\mathbf{r}$ is fully defined by the forces evaluated at the previous

value of $\mathbf{r}$. This should be contrasted with the Stratonovich convention,

$$\gamma_i \mathbf{r}_i(t + \Delta t) = \gamma_i \mathbf{r}_i(t) + \mathbf{F}_i \left[\frac{\mathbf{r}^N(t) + \mathbf{r}^N(t + \Delta t)}{2} \right] \Delta t + \sqrt{2\gamma_i k_\mathrm{B} T \Delta t} \boldsymbol{\mathcal{N}}_i^t, \qquad (1.6)$$

where the dynamical equation has to be solved self-consistently at every timestep. Choosing either of the two conventions results in different dynamics if the pre-factor of the Gaussian noise depends on the configurations $\mathbf{r}$. The Stratonovich formalism makes differential calculus easier as the chain rule of differentiation holds unchanged, while the Ito formalism simplifies integral calculus as the change of $\mathbf{r}(t)$ in every step is uncorrelated from the noise $\mathcal{N}_i^t$. In our work we will use the Ito formalism for integrating the Langevin equation as well as integrating all stochastic integrals for physical observables. With this convention, the time evolution of the probability distribution of the configurations $\{\mathbf{r}^N\}$, denoted as $P(\mathbf{r}^N, t)$, is described in forward direction by the Fokker-Planck equation [43]

$$\frac{\partial P(\mathbf{r}^N, t)}{\partial t} = \sum_i \left[-\boldsymbol{\nabla}_i \cdot \left(\mathbf{F}_i[\mathbf{r}^N(t)] P(\mathbf{r}^N, t) \right) + D_i \boldsymbol{\nabla}_i^2 P(\mathbf{r}^N, t) \right] \qquad (1.7)$$

This equation can be derived from the Langevin dynamics by a Kramers-Moyal expansion, a moment expansion of the transition probabilities due to the stochastic noise [43]. The Fokker-Planck equation, also known as the forward Kolmogorov equation, along with boundary conditions describes the forward time evolution of the full probability distribution of the stochastic dynamics arising from Eq. 1.4. In principle, the Fokker-Planck equation formally predicts the probability of all rare dynamics that the Langevin dynamics can generate. In practice, it is often not possible to obtain analytical or even numerical solutions from the Fokker-Planck equation when the system size is large. In many-body systems with identical particles, the Fokker-Planck equation is often hierarchically coarse-grained to describe effective evolution of one or two-body densities [44].

Dynamical events of interest simulated with the Langevin equation also manifest in the ensemble of trajectories. We will denote a trajectory as a stochastic object $\mathbf{X}(t_f)$ defined as a sequence of configurations generated from a Langevin dynamics of duration t_f, $\mathbf{X}(t_f) \equiv \{\mathbf{r}^N(0), \mathbf{r}^N(\Delta t), \mathbf{r}^N(2\Delta t), \ldots, \mathbf{r}^N(t_f)\}$. An ensemble generated by Langevin trajectories with all possible statistical realizations of the Gaussian noise will be referred to as the trajectory ensemble associated with the dynamics. For any particular trajectory with a pre-specified initial condition, every step has been taken with a Gaussian probability $\exp(-\mathcal{N}_i^2/2)/\sqrt{2\pi}$. We write the total trajectory probability in the fine discretization limit as [45]

$$P[\mathbf{X}] \propto p(\mathbf{r}^N(0)) \exp \left(\int_0^{t_f} dt \sum_i \frac{(\gamma_i \dot{\mathbf{r}}_i - \mathbf{F}_i)^2}{4\gamma_i k_\mathrm{B} T} \right) \qquad (1.8)$$

where the normalization constant is formally infinite. The exponent is referred to as the stochastic action associated with each trajectory, in analogy with the path-integral description of Schrödinger's wave mechanics. All three levels of descriptions of the stochastic dynamics of our system of interest, through configurations, densities and trajectories as in

Equations 1.4, 1.7 and 1.8, are fully equivalent, and we will often draw connections between the three to fully characterize the physical and chemical mechanisms of rare fluctuations. However, the fact that the full trajectory probability has a closed form expression in terms of the applied forces, even in cases where the forces are nonconservative and the dynamics is out of equilibrium, makes our approach of reweighting trajectories with an optimal control force viable. Hence we will be using trajectory ensembles constructed from molecular dynamics simulations to apply our variational algorithms.

1.3   Steady-state rare events: large deviation theory

The statistics of persistent rare events in nonequilibrium steady-states can be analyzed using large deviation theory. Large deviation theory is a branch of probability dealing with the exponential scaling of large fluctuations of the sample mean or empirical mean of a large number of stochastic observables. This can be thought of as a generalization of the central limit theorem in probability towards describing large fluctuations of sample means, or *large deviations*. The central components in large deviation theory are the rate function, which describes the log-likelihood of the sample mean, and the scaled cumulant generating function for its probability distribution [46]. These two are jointly referred to as large deviation functions and, in the case of the probability distribution being convex and differentiable, are connected through the Legendre-Fenchel transform, a generalization of the Legendre transform used for thermodynamic free energies. In recent years, the classical theory for constructing thermodynamic free energies from statistical mechanical partition functions has been reinterpreted with the language of large deviation theory, with the thermodynamic system size being the large deviation limit [34]. This has allowed interpreting steady-state trajectory ensembles to also be in the (temporal) large deviation limit, and has provided large deviation functions as dynamical free energies for specific observables of interest, encoding stability and response in arbitrarily far-from-equilibrium steady-states. In this section we will review the large deviation formalism for trajectory ensembles and its use to obtain statistics of rare events.

For ease of notation we will denote all coordinates of the system, $\mathbf{r}^N(t)$, sometimes simply as $\mathbf{r}$. For a specific trajectory, $\mathbf{X}(t_f) = \{\mathbf{r}(0), ..., \mathbf{r}(t_f)\}$ spanning an observation time, t_f, we are interested in fluctuations of time-averaged dynamical observables A_{t_f} of the form

$$A_{t_f}[\mathbf{X}(t_f)] = \frac{1}{t_f} \int_0^{t_f} dt \, f[\mathbf{r}(t)] + \frac{1}{t_f} \int_0^{t_f} dt \, \mathbf{g}[\mathbf{r}(t)] \cdot \dot{\mathbf{r}}(t) \tag{1.9}$$

where f is a scalar function and $\mathbf{g}$ is a vector function with components, g_i, with the second term being evaluated in the Ito sense [47]. Many physically relevant observables like the particle density, particle current, and entropy production can all be expressed in this trajectory averaged form. We will be interested in the statistics of rare time-extensive fluctuations, the large deviations, of A_{t_f} in the long time limit, $t_f \to \infty$.
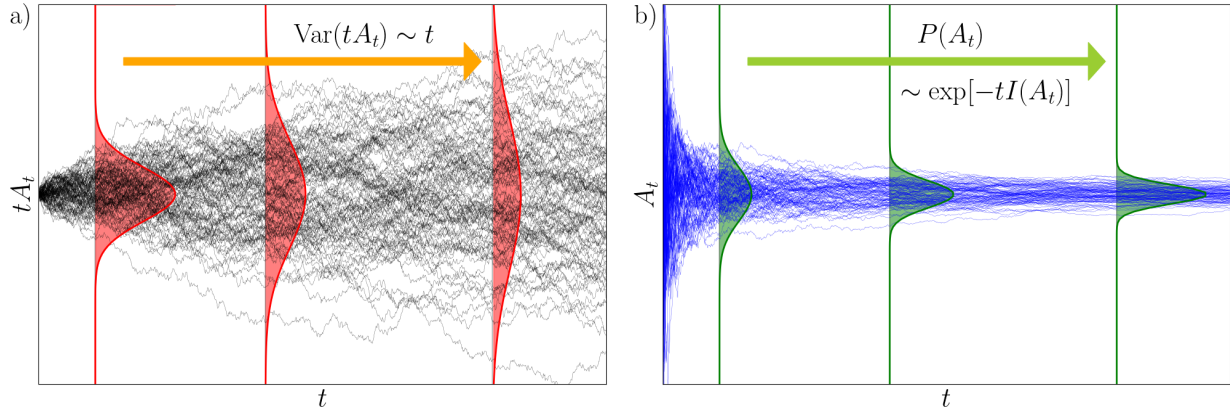
Figure 1.1: Large deviation principle for observable A_t arising from Brownian diffusion. a) Trajectories of the time-integrated observable tA_t starting from the same point diverge in time with a variance proportional to t over times longer than intrinsic correlation times of the system. c) Trajectories of the time-averaged analogue A_t of the same observable. The probabilities concentrate around the mean value with asymptotic scaling as exponential in time multiplied by the negative of the rate function $I(A_t)$.

We assume that in the long time limit, the probability distribution of A_{t_f} satisfies a large deviation principle, with a rate function, or log likelihood, $I(A)$, defined by [34]

$$I(A) = -\lim_{t_f \to \infty} \frac{1}{t_f} \ln\langle \delta(A - A_{t_f}[\mathbf{X}(t_f)])\rangle \qquad (1.10)$$

where the angular brackets denote a trajectory average

$$\langle \delta(A - A_{t_f}[\mathbf{X}(t_f)])\rangle = \int D[\mathbf{X}(t_f)]\delta(A - A_{t_f}[\mathbf{X}(t_f)])P[\mathbf{X}(t_f)] \qquad (1.11)$$

and $P[\mathbf{X}(t_f)]$ denotes the path probability associated with trajectory $\mathbf{X}(t_f)$. We will consider thermalized systems that have exponentially decaying correlation functions and thus are expected to obey the large deviation principle. The large deviation principle for time-integrated and time-averaged observables in the long time limit is illustrated in Fig. 1.1.

The long time behavior of A_{t_f} can also be characterized by its scaled cumulant generating function (SCGF), defined as

$$\Psi(s) = \lim_{t_f \to \infty} \frac{1}{t_f} \ln\left\langle e^{-st_f A_{t_f}} \right\rangle \qquad (1.12)$$

where s is a counting parameter conjugate to A_{t_f}, and denotes the extent of *biasing* or *tilting* on the typical value of A_{t_f}. Larger positive or negative values of s probe rarer fluctuations.

This is clear by noting that the derivatives of $\Psi(s)$ report on the cumulants of A_{t_f}, such as

$$\frac{d\Psi(s)}{ds} = -\langle A_{t_f} \rangle, \text{ and } \qquad \frac{d^2\Psi(s)}{ds^2} = t_f \langle (\delta A_{t_f})^2 \rangle \qquad (1.13)$$

where $\delta A = A - \langle A \rangle$ are the fluctuations of A. We refer to the rate function, $I(A)$, and the SCGF, $\Psi(s)$, collectively as the large deviation functions. When the rate function is convex, it can be obtained from the SCGF using a Legendre-Fenchel transform

$$I(A) = \inf_s[-sA - \Psi(s)] \qquad (1.14)$$

where inf refers to an infimum taken over all possible values of s. Together, the large deviation functions encode stability of dynamical phases in nonequilibrium systems [48, 49, 50, 51, 52], characterize complex dynamical behavior [53, 54, 55, 56, 57] and describe linear and nonlinear multivariate response of nanoscale systems to nonequilibrium perturbations [58, 36, 59, 60, 61]. The SCGF defined in the trajectory ensemble picture can also be derived from the eigenspectrum of a tilted backward Fokker-Planck propagator defined as

$$L_s = -sf + \sum_i \left[\frac{s}{2}(\boldsymbol{\nabla}_i \cdot \mathbf{g}_i)\mathbf{g}_i + \mathbf{F}_i \cdot (\boldsymbol{\nabla}_i - s\mathbf{g}_i) \right.$$
$$\left. + \gamma_i k_{\mathrm{B}} T(\boldsymbol{\nabla}_i^2 - s(\boldsymbol{\nabla}_i \cdot \mathbf{g}_i) - 2s\mathbf{g}_i \cdot \boldsymbol{\nabla}_i + s^2\mathbf{g}_i^2) \right] \qquad (1.15)$$

where we have suppressed the arguments of $\mathbf{F}_i$, $\mathbf{g}_i$, and f for compactness. This operator can be derived similar to the Fokker-Planck equation from the Kramers-Moyal expansion for infinitesimal propagation of $\langle \exp(-st_f A_f) \rangle$ with increasing t_f [62]. We have adopted the form of the operator used in [63] towards the Ito definition of A_{t_f} rather than Stratonovich, by adding the first term in the square brackets in Eq. 1.15. This operator satisfies an eigenvalue equation

$$L_s\phi_s(\mathbf{r}^N) = \Psi(s)\phi_s(\mathbf{r}^N) \qquad (1.16)$$

where $\Psi(s)$ and $\phi_s(\mathbf{r}^N)$ are respectively the largest real eigenvalue and corresponding right eigenvector of L_s, which follows from the Perron-Frobenius theorem and the long time limit of the SCGF. Constructing and solving this eigenvalue equation in a chosen functional basis for $\phi_s(\mathbf{r}^N)$ is an option to obtain the SCGF $\Psi(s)$, however, the dimension of the solution space scales exponentially with the system size, rendering this approach impractical in many-body systems.

1.4 Barrier crossing events: reaction rate theory

Interesting chemical events in liquid and solid phase involving transduction of energy from one class of dynamical modes in a molecule to another often happen when a trajectory goes over a free energy barrier that separates long-lived metastable states. Rate constants of such rare events in equilibrium depend on the free energy cost of being at the top of

the barrier as well as the dynamical flux contributing to such a barrier crossing. One of the fundamental approaches to quantify reaction rates in equilibrium systems is transition state theory that assumes configurations at the top of the barrier to be in quasi-equilibrium with the rest of the system, thus decoupling the dynamics from the Boltzmann energetics. In reality, this approximation leads to an overestimation of the rate, as many trajectories that cross the barrier still recross back to the reactant basin in a correlated fashion to the configurational statistics. Thus a full treatment for exact reaction rate computation in even equilibrium system requires a trajectory ensemble formalism [13, 64]. Additionally, in nonequilibrium reactions, such as the nucleation of a liquid phase in repulsive motile particles [65], in the shear induced unfolding of proteins in blood vessels [66] and in rare switching events in biochemical networks [67], the existence of a free energy as a state-function is not guaranteed, and dynamics is inherently coupled to structural fluctuations. In the limit of low temperature compared to the barrier-height, Freidlin-Wentzell theory approximates barrier-crossing trajectories to be comprised of a long sequence of infinitesimal noise fluctuations along a progress coordinate, thus mapping the fluctuation into a large deviation and yielding corresponding rate estimates [68, 69]. However, for systems with diffusive degrees of freedom at ambient temperature, the overlap between the problem of sampling short barrier-crossing trajectories of finite duration and that of sampling large deviations in steady-state trajectories of infinite duration remains unexplored and the tools used in either disciplines are mostly disparate.

In this section we will discuss how to compute reaction rates from ensembles of trajectories. When we want to sample reactive trajectories that cross a free energy barrier transiently in time but otherwise follow typical steady-state statistics, either in or out of equilibrium, we want to restrict our trajectory ensemble to have a finite duration t_f, but initialized from an ensemble of steady-state configurations $\{\mathbf{r}^N(0)\}$. We will denote the initial and final basins in configuration space as A and B, and the rate constant of going from A to B as k. In case of the reaction being rare on simulation timescales, t_f usually is in the range $\tau^{\ddagger} < t_f << 1/k$, where $\tau^{\dagger}$ is the timescale of relaxation within the initial metastable basin or that of the typical transition path duration, while $1/k$ is the typical waiting time for the reaction to happen spontaneously. This separation of timescales is usually associated with a large gap in the eigenspectrum of the Fokker-Planck operator governing the dynamics in the system. We have schematically illustrated this for a general order parameter $\mathbf{q}$ in Fig. 1.2(a). We also define indicator functions $h_A(t)$ and $h_B(t)$ at any given time t as

$$h_{A(B)}[\mathbf{r}^N(t)] = \begin{cases} 1 & \text{if } \mathbf{r}^N(t) \in A(B) \\ 0 & \text{else.} \end{cases} \tag{1.17}$$

The rate constant k can be computed as the reactive flux of trajectories into B conditioned on starting from A,

$$k = \frac{d}{dt_f}\langle h_B(t_f)|h_A(0)\rangle = \frac{d}{dt_f}\frac{\langle h_A(0)h_B(t_f)\rangle}{\langle h_A \rangle} \tag{1.18}$$
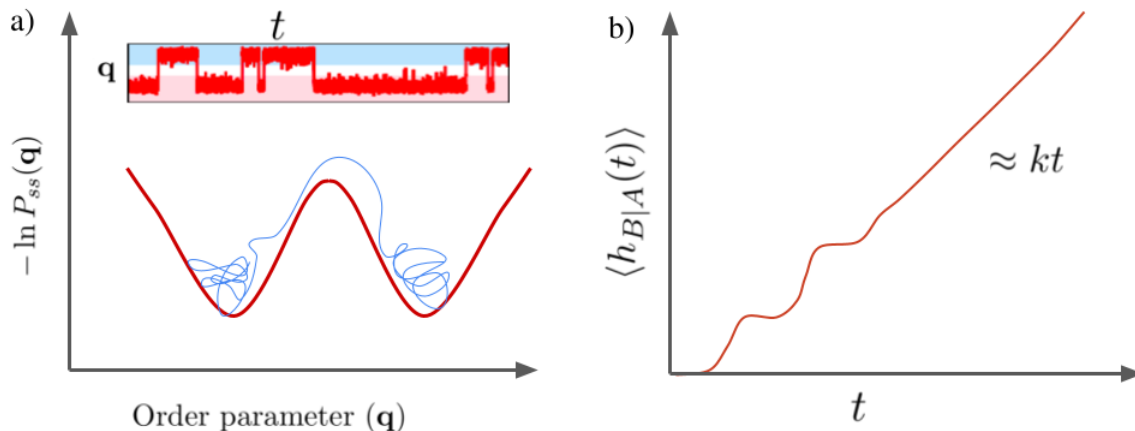
Figure 1.2: Reaction rate computation from side-side correlation function. a) Separation of timescales between the relaxation time within each metastable well and the typical waiting time for barrier-crossing. Blue line is a schematic illustration of the proportion of time spent by a barrier-crossing trajectory in the wells and on the barrier. *(Inset)* Time series of order parameter **q**. b) Linear regime of the side-side correlation function, with the slope given by the reaction rate.

where the averages are computed over trajectories $\mathbf{X}(t_f)$ and the average in the denominator is time-independent in a steady-state. Eq. 1.18 for the reaction rate yields a time-independent constant for all times that fall between the separated molecular and reactive timescales, and so the side-side correlation function $\langle h_B(t)|h_A(0)\rangle$ is linear in t in that regime. When the separation of timescales is large, the offset to the correlation function from the initial nonlinear portion is small enough, and the rate constant can be computed from a simple time-scaling [70], as

$$k \approx \frac{\langle h_B(t_f)|h_A(0)\rangle}{t_f} = \frac{1}{t_f}\frac{Z_{AB}(t_f)}{Z_A} \tag{1.19}$$

where $Z_{AB}(t)$ and Z_A are defined as trajectory partition functions $Z_{AB}(t) = \langle h_A(0)h_B(t)\rangle$ and $Z_A = \langle h_A \rangle$. This linear scaling regime is schematically demonstrated in Fig. 1.2 (b). Reaction rate constants can thus be computed by counting the number of reactive trajectories proportional to all trajectories.

The side-side correlation is the central observable of interest for all sampling techniques for reactive trajectories. Sometimes the computation of the side-side correlation is done in two steps,

$$k_{AB} = \langle \dot{h}_B(t)|h_A(0)\rangle = \frac{\langle \dot{h}_{A+}(\delta t)h_A(0)\rangle}{\langle h_A \rangle}\langle h_B(t)|h_{A+}(\delta t)\rangle \tag{1.20}$$

where the first term in the product is the *escape flux* from the basin A towards B, with $\dot{h}(\delta t)$ denoting the time-derivative of the indicator $h(\delta t)$ when some infinitesimal time δt has elapsed after last being in A, and the second term in the product is the probability of reaching B given a trajectory has escaped towards B but is still in the vicinity of A. The escape flux and the conditional reactive probabilties can be evaluated separately and then combined to produce the reactive flux, *i.e.*, the reaction rate.

The growth statistics of the side-side correlation produces only the rate constant, but does not by itself contain information about the reaction mechanism. A microscopic quantity that encodes the reaction mechanism is the committor function, defined as the probability of reaching target state B after some transient time $\tau^{\ddagger}$ when started from system configurations $\{\mathbf{r}^N\}$, as

$$p(\mathbf{r}^N, \tau^{\ddagger}) = \frac{\langle \delta(\mathbf{r}^N(0) - \mathbf{r}^N)h_B(\tau^{\ddagger}) \rangle}{\langle \delta(\mathbf{r}^N(0) - \mathbf{r}^N) \rangle} \tag{1.21}$$

where angular brackets again mean a trajectory average [70, 64]. We see that the committor function is just a generalization of a side-side correlation function for starting from all coordinates instead of from just the initial metastable basin. The transient time $\tau^{\ddagger}$ is comparable to the typical time for natural transition paths, *i.e.*, the time taken between the last escape from A and the first subsequent entry into B. Since trajectories started from most regions in configuration space do not react within this short time, the committor will be mostly either 0 or 1, but for only a few special configurations that connect A to B along the correct reaction coordinate, the committor will smoothly change in value, taking a value of 0.5 at the exact set of transition states. Thus the committor function encodes the intrinsic mechanism of the rare event by identifying the dynamical bottleneck. Regions of configuration space having different values of the committor function are said to be partitioned by a separatrix, another name for an iso-committor surface. The direction of the gradient of the committor function, comprising of normals to consecutive separatrix surfaces, can be identified as the reaction coordinate, which is the natural coordinate on which the reaction proceeds. Knowing the reaction coordinate will equip us to tune the exact reaction mechanism microscopically and thus a key goal of enhanced sampling algorithms is to learn a committor surface from harvested reactive trajectories.

1.5 Existing sampling techniques for rare nonequilibrium trajectories

Enhanced sampling methods within equilibrium ensembles are standard tools that enable the determination of phase diagrams and the calculation of rates of rare events, through the evaluation of equilibrium free energies and rare dynamics that obeys detailed balance. The methods typically rely on either enhancing the sampling of the tails of a Boltzmann distribution by adding a suitable umbrella potential [42, 71], or by resampling parts of configurational

and trajectory ensembles to construct rare fluctuations [72], or both [73]. Nonequilibrium systems pose additional complications over equilibrium ones in terms of sampling both rare structures and dynamics. Nonequilibrium structural configurations are not Boltzmann distributed and the existence of free energies as a functional of only density-like variables is not guaranteed, so methods analogous to umbrella sampling in configuration space have not been possible. Though some recent approaches discuss candidate theories for quantifying the amplification of reaction rates with nonequilibrium driving, they feature bounds instead of exact reweighting relations, and involve dynamical observables like heat dissipation rather than thermodynamic state-functions like energy or entropy [74, 75]. Nonequilibrium dynamics on the other hand breaks detailed balance, and it is not possible to construct a rare barrier-crossing trajectory by propagating a transition state configuration forward and backward in time. Instead, most nonequilibrium importance sampling methods for barrier-crossing trajectories start with a distribution of initial configurations and push the trajectory ensemble up the barrier in a forward ratchet-like fashion, performing nonequilibrium *meta-work* in the space of the trajectories [76]. We have discussed below some of these trajectory sampling methods for nonequilibrium large deviations and reactive events.

The dynamical analogues of equilibrium free energies for nonequilibrium steady-states are large deviation functions that describe the likelihood of fluctuations of time-averaged observables and encode the response and stability of the system. The aim of dynamical importance sampling methods for nonequilibrium steady-states is thus to estimate large deviation functions from a trajectory ensemble. As large deviation functions like the SCGF contain exponentially rare statistics, the tails of the probability distribution of the observable of interest must be sampled in order to have an accurate estimate. Efforts to compute large deviation functions in systems with many degrees of freedom have largely been restricted to Monte Carlo based approaches like cloning [77, 78] and list-based algorithms [79]. Most current algorithms scale exponentially in computational effort the further the rare fluctuation is from the mean behavior, as apart from stratification or population dynamics, most do not employ additional importance sampling [80, 81, 82, 83]. Recent work adding control forces to importance sample trajectory based Monte Carlo has demonstrated that even an approximate force can greatly improve the efficiency of Monte Carlo methods in estimating large deviation functions [84, 85, 86]. Consequently there has been much work to find approximate control forces analytically or through empirical arguments in both lattice-based and continuous systems [87, 88, 89] and several iterative effective force optimization techniques have been proposed with varying levels of generality or accuracy [90, 91, 92, 93]. The control forces in general can have many-body components in interacting particle systems [94, 89], can be long-ranged in systems with dynamical phase transitions [95], and can stabilize otherwise metastable states [96]. However, many-body control forces have been derived only from analytical approximations in simple limits, and efficient numerical techniques to learn many-body optimal forces in systems of chemical relevance is currently lacking.

In contrast, nonequilibrium rate calculation has been entirely reliant on trajectory stratification and Monte-Carlo resampling based approaches. Most such techniques build on the classic Weighted Ensemble method introduced by Huber and Kim [97] and subsequently re-

fined and applied to equilibrium biological problems like protein folding and protein-ligand binding by Zuckerman, Chong and others [98]. The original Weighted Ensemble algorithm is applicable for nonequilibrium dynamics but not very efficient for generating reactive trajectories. The Transition Path Sampling algorithm [70, 99], widely used to compute rates of molecular isomerization and dissociation in solvents at equilibrium, can be applied to nonequilibrium systems by either restricting the set of moves that require detailed balance [81], or by directly accounting for the change in nonequilibrium path action by noise sampling [100]. The most widely used method for computing reaction rates in stochastic nonequilibrium systems is Forward Flux Sampling, that starts with the forward flux of escaping trajectories from the reactant basin and generates reactive trajectories stepwise by dividing the progress coordinate into multiple bins [101, 102]. A related approach is nonequilibrium umbrella sampling, that keeps track of the entry and exit flux of bins in configuration space in order to output the nonequilibrium steady-state distribution [103, 104]. The application of control forces to enhance the sampling of reactive events has been extremely limited and been confined only to equilibrium systems. Recent works by Rotskoff and Vanden-Eijnden recast the rare event sampling as a variational learning of an equilibrium committor surface [105, 106], however, connections have not been made between the committor function and the optimal force. Other approaches that use an importance sampling potential have explicitly used Boltzmann statistics, and have only seen moderate improvements in sampling statistics on adding suboptimal control potentials [107, 108]. However, the exact optimal control force is formally known to generate uncorrelated statistical realizations of rare trajectories directly [109], thus making the rate computation *direct* without need of additional importance sampling, in arbitrarily far-from-equilibrium systems. Hence there exists a dearth of efficient numerical methods to compute the optimal control force within conveniently chosen basis sets for evaluating nonequilibrium reactive events in complex molecular and colloidal systems. It is this gap that our variational techniques are able to address.

Chapter 2

Variational optimization of steady-state trajectories

In this chapter we will develop a variational algorithm to optimize a control force within a chosen basis to compute large deviation functions. For Markovian systems, there exists an optimal control force, which is the unique additional force having the smallest contribution to the path ensemble measure that can be added to the system to make a rare fluctuation typical [63, 110]. This optimal control force satisfies several variational identities [111]. By deriving such a variational principle and explicit forms for the gradients required to optimize it, we develop an algorithm that approximates the control force sufficiently well so as to make quantitatively accurate estimates of the likelihood of rare events within nonequilibrium steady-states. In this way, we generalize previous work on variational control of single particle systems to interacting, continuous force systems, bypassing the need for exponentially scaling Monte Carlo sampling. Our algorithm is similar in strategy to the recent use of thermodynamic variational principles to compute equilibrium free energies [71], and to the Rayleigh-Ritz variational principle that others have used to nonperturbatively compute effective forces far from equilibrium [112]. The variational principle that underlies our algorithm is related to minimum-entropy production principles [113, 114]. and the Donsker-Varadhan formula in Markov Stochastic processes [115]. While our variational estimate of the large deviation function is subject to errors associated with the representation of the control force, we derive exact corrections that can be evaluated straightforwardly. In the first two systems studied, these corrections are easy to evaluate, as our control forces are sufficiently close to the optimal control forces to make these corrections perturbatively small. However, in cases where the corrections are large, we show that using optimized control forces in conjunction with standard Monte Carlo algorithms can increase the statistical efficiency in the estimation of large deviation functions by orders of magnitude. In this way, our algorithm is similar to the use of variationally optimized wavefunctions for quantum Diffusion Monte Carlo calculations [116].[1]

---

[1]Most of the content of this chapter was originally part of the publication [117].

2.1 Variational principle from Girsanov reweighting

Computing either of the large deviation functions of A_{t_f} requires sampling exponentially rare fluctuations. These rare fluctuations can in principle be made to occur more frequently by introducing an additional control force into the system as a means of importance sampling. In the presence of the additional force, $\boldsymbol{\lambda}(\mathbf{r}^N)$, over the original force, $\mathbf{F}(\mathbf{r^N})$, the computation of the SCGF can done by changing the path ensemble measure,

$$\Psi(s) = \lim_{t_f \to \infty} \frac{1}{t_f} \ln \int D[\mathbf{X}(t_f)] e^{-s t_f A_{t_f}} \frac{P[\mathbf{X}(t_f)]}{P_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)]} P_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)]$$
$$= \lim_{t_f \to \infty} \frac{1}{t_f} \ln \left\langle e^{O_{\boldsymbol{\lambda}}} \right\rangle_{\boldsymbol{\lambda}} \tag{2.1}$$

where $\langle \cdot \rangle_{\boldsymbol{\lambda}}$ denotes an average in the controlled path ensemble with path probabilities $P_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)]$, and $O_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)]$ can be derived from the difference in Onsager-Machlup path-actions $\Delta U_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)]$, [45]

$$O_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)] = -s t_f A_{t_f} + \int_0^{t_f} dt \sum_i \frac{\boldsymbol{\lambda}_i^2 - 2\boldsymbol{\lambda}_i \cdot (\gamma_i \dot{\mathbf{r}}_i - \mathbf{F}_i)}{4\gamma_i k_{\mathrm{B}} T} \equiv -s t_f A_{t_f} - \Delta U_{\boldsymbol{\lambda}} \tag{2.2}$$

interpreted in the Ito sense. Changing the force for such a Gaussian process does not change the normalization constant associated with the path ensemble in the long time limit where boundary terms from the initial and final configurations can be ignored.

Expanding Eq. 2.1 in terms of its cumulants, and using Jensen's inequality, we find a variational expression for the SCGF,

$$\Psi(s) \geq \lim_{t_f \to \infty} \frac{1}{t_f} \langle O_{\boldsymbol{\lambda}} \rangle_{\boldsymbol{\lambda}} \tag{2.3}$$

in terms of the mean of $O_{\boldsymbol{\lambda}}[\mathbf{X}(t_f)]$, averaged within the controlled path ensemble. This expression is identical to previous work by Chetrite and Touchette that was derived using the contraction principle [111]. To see this, here we note that we can write $\langle O_{\boldsymbol{\lambda}} \rangle_{\boldsymbol{\lambda}}$ in two ways,

$$\langle O_{\boldsymbol{\lambda}} \rangle_{\boldsymbol{\lambda}} = \left\langle -s t_f A_{t_f} + \int_0^{t_f} dt \sum_i \frac{\boldsymbol{\lambda}_i^2 - 2\boldsymbol{\lambda}_i \cdot (\gamma_i \dot{\mathbf{r}}_i - \mathbf{F}_i)}{4\gamma_i k_{\mathrm{B}} T} \right\rangle_{\boldsymbol{\lambda}} \tag{2.4}$$

$$= \left\langle -s t_f A_{t_f} - \int_0^{t_f} dt \sum_i \frac{\boldsymbol{\lambda}_i^2}{4\gamma_i k_{\mathrm{B}} T} \right\rangle_{\boldsymbol{\lambda}} \tag{2.5}$$

where we derive the second expression from the first by using the driven Langevin equation $\gamma_i \dot{\mathbf{r}}_i = \mathbf{F}_i + \boldsymbol{\lambda}_i + \boldsymbol{\eta}_i$ and that, for an Ito dynamics, $\langle \boldsymbol{\lambda}_i(t) \cdot \boldsymbol{\eta}_i(t) \rangle = 0$ because of causality.

Among the forces that make the rare value of the observable statistically typical, the one closest to the original force is the optimal force that realizes the supremum of the inequality.

This many-body function can be approximated within a chosen ansatz with variationally
optimizable parameters $\{\chi\}$. In the limit that $\{\chi\}$ represents all possible functional forms
of the many-body force, this ansatz becomes exact [111], so that

$$\Psi(s) = \sup_{\{\chi_1,\chi_2,\ldots\}} \lim_{t_f \to \infty} \frac{1}{t_f} \left\langle O_{\boldsymbol{\lambda}[\{\chi\}]} \right\rangle_{\boldsymbol{\lambda}[\{\chi\}]} \tag{2.6}$$

where the optimal coefficients $\{\chi\}$ will in general depend on s.

The existence of a control force that saturates the supremum in Eq. (2.6) follows again
from the largest real eigenvalue $\Psi(s)$ and the corresponding right eigenvector $\phi_s(\mathbf{r^N})$, of the
generator of the SCGF in Eq. 1.15, that directly gives the value of the SCGF,. The optimal
force $\boldsymbol{\lambda}_s^*$ that solves Eq. 2.6 is related to ϕ_s through a Hopf-Cole transform [118, 63, 111]
defined as

$$\boldsymbol{\lambda}_{s,i}^* = \mathbf{F}_i + 2\gamma_i k_{\mathrm{B}} T (-s\mathbf{g} + \boldsymbol{\nabla}_i \ln \phi_s) \tag{2.7}$$

and the controlled dynamics associated with this optimal force can be obtained from a
generalized Doob transform of L_s [63, 119]. That $\boldsymbol{\lambda}_s^*$ saturates the variational inequality 2.3
can be seen by writing the eigenvalue equation as

$$\Psi(s) = \phi_s^{-1} L_s \phi_s, \tag{2.8}$$

and multiplying both sides by a steady-state configurational probability distribution asso-
ciated with a generic control force, $\rho_{\boldsymbol{\lambda}}(\mathbf{r}^N)$. If we then integrate over all coordinates with
appropriate boundary conditions, it can be shown that the variational surface is convex
around the supremum [113],

$$\Psi(s) = \lim_{t_f \to \infty} \frac{1}{t_f} \left\langle O_{\boldsymbol{\lambda}} + \int_0^{t_f} dt \sum_i \frac{(\boldsymbol{\lambda} - \boldsymbol{\lambda}^*)^2}{4\gamma_i k_{\mathrm{B}} T} \right\rangle_{\boldsymbol{\lambda}} \tag{2.9}$$

However, though the variationally surface is locally convex, it may not be globally convex.
Following the same derivation as above but with starting from an eigenvalue equation for a
general (non-dominant) eigenvalue, the corresponding eigenvector, and its Hopf-Cole trans-
form, it can be shown that the SCGF estimator in Eq. 2.6 can give the exact value of the
real part of any eigenvalue of the tilted generator, for a physically meaningful force. This
means that at dynamical crossovers when two eigenvalues approach each other and cross as
a function of s, the variational expression is almost (infinitesimally) satisfied by two distinct
forces, making the variational surface either slowly varying or bimodal.

For an interacting many-body system, the dominant eigenvector is a many body state,
and therefore the optimal control force is many-bodied. Generally, we will assume that the
control force is well approximated by a low rank ansatz such as obtained from a low order
many body expansion.

Obtaining the SCGF from directly diagonalizing the tilted generator in many-body sys-
tems is prohibitively expensive due to the size of the multi-dimensional state space over

which L_s is defined. There have been recent advances to approximate this state space using Matrix Product States for lattice based models [120, 121]. However, for continuous space systems with many particles, it is expected that Eq. 2.6 will present a physically motivated way to formulate approximate solutions to the eigenvalue problem and to the computation of $\Psi(s)$, and subsequently, $I(A)$. It is worth noting that the constrained optimization of a variational expression analogous to (2.6) can also be directly used to compute $I(A)$ [111], with a straightforward extension of the algorithm described below.

2.2 Optimization algorithm with explicit gradients

In order to optimize Eq. 2.6 by gradient descent, we need to calculate derivatives of $\langle O_\lambda[\mathbf{X}(t_f)]\rangle_\lambda$ with respect to the variational parameters $\{\chi\}$ in the limit of a large t_f. Using these explicitly calculated gradients in the optimization algorithm can reduce the noise and numerical instabilities associated with finite difference schemes, that are generally used to empirically estimate the gradients from the optimization trajectory through the parameter space. The explicit gradients that we use have the form of expectation values in the controlled ensemble,

$$
\lim_{t_f \to \infty} \frac{1}{t_f} \frac{\partial}{\partial \chi} \langle O_\lambda \rangle_\lambda
$$
$$
= \lim_{t_f \to \infty} \frac{1}{t_f} \left[\left\langle \frac{\delta O_\lambda[\mathbf{X}]}{\delta \boldsymbol{\lambda}} \cdot \frac{\partial \boldsymbol{\lambda}}{\partial \chi} \right\rangle_\lambda + \left\langle O_\lambda[\mathbf{X}] \frac{\partial \ln P_\lambda[\mathbf{X}]}{\partial \chi} \right\rangle_\lambda \right] \tag{2.10}
$$

where χ is any of the optimizable parameters specifying the control force. The first term is a simple trajectory average and is straightforward to compute. In a steady-state, we replace the time-extensive integral divided by the trajectory time with only the instantaneous contribution,

$$
\lim_{t_f \to \infty} \frac{1}{t_f} \left\langle \frac{\delta O_\lambda[\mathbf{X}]}{\delta \boldsymbol{\lambda}} \cdot \frac{\partial \boldsymbol{\lambda}}{\partial \chi} \right\rangle_\lambda = \left\langle \frac{\delta \dot{O}_\lambda}{\delta \boldsymbol{\lambda}} \cdot \frac{\partial \boldsymbol{\lambda}}{\partial \chi} \right\rangle_\lambda \tag{2.11}
$$

where Eqns. (1.9) and (2.2) have been used to write $O_\lambda[\mathbf{X}]$ as a time integral of $\dot{O}_\lambda(t)$.

The second term is a two-time correlation function that is expected to have a high variance if directly computed in this form in a steady-state. Explicit functional forms of $\partial \ln P_\lambda[\mathbf{X}]/\partial \chi$ can be derived from the normalized path probabilities as,

$$
\left\langle O_\lambda[\mathbf{X}] \frac{\partial \ln P_\lambda[\mathbf{X}]}{\partial \chi} \right\rangle_\lambda = \left\langle \int_0^{t_f} dt\, \dot{O}_\lambda(t) \int_0^{t_f} dt' \sum_i \frac{\boldsymbol{\eta}_i(t')}{2\gamma_i k_B T} \cdot \frac{\partial \boldsymbol{\lambda}_i(t')}{\partial \chi} \right\rangle_\lambda
$$
$$
- \left\langle \int_0^{t_f} dt\, \dot{O}_\lambda(t) \right\rangle_\lambda \left\langle \int_0^{t_f} dt' \sum_i \frac{\boldsymbol{\eta}_i(t')}{2\gamma_i k_B T} \cdot \frac{\partial \boldsymbol{\lambda}_i(t')}{\partial \chi} \right\rangle_\lambda . \tag{2.12}
$$

The second term comes from the derivative of the normalization constant for $P_{\boldsymbol{\lambda}}[\mathbf{X}]$, and
it is formally zero as reweighting with a changed force doesn't change the normalization
constants for the Gaussian white noise. We see this explicitly as $\langle \boldsymbol{\eta}_i(t') \cdot \partial \boldsymbol{\lambda}_i(t')/\partial \chi \rangle = 0$
due to causality where the product is Ito. Nevertheless, including this term during empirical
estimation of the gradient reduces its variance, working as a noisy baseline subtraction.

The averages in Eq. 2.12 can be computed by propagating additional coordinates $y_\chi(t)$
associated with each variational parameter χ as

$$y_\chi(0) = 0, \quad \dot{y}_\chi(t) = \sum_i \frac{\boldsymbol{\eta}_i(t)}{2\gamma_i k_{\mathrm{B}} T} \cdot \frac{\partial \boldsymbol{\lambda}_i(t)}{\partial \chi} \tag{2.13}$$

where the sum has been performed over all dynamical coordinates of the system, and its
fluctuation is defined as $\delta \dot{y}_\chi(t) = \dot{y}_\chi(t) - \langle \dot{y}_\chi(t) \rangle_{\boldsymbol{\lambda}}$. These fictitious coordinates are known in
the literature as Malliavin weights [122] and have previously been used to calculate parameter
sensitivity of steady-state distributions in Langevin systems [123]. Provided these averages
are evaluated in the steady-state generated by the control force, $\gamma_i \dot{\mathbf{r}}_i = \mathbf{F}_i + \boldsymbol{\lambda}_i + \boldsymbol{\eta}_i$, we
can invoke time-translational invariance and note that only past noise history correlates
with the observable to simplify Eq. 2.10. We start by defining the fluctuations in $\dot{O}_{\boldsymbol{\lambda}}(t)$ as
$\delta \dot{O}_{\boldsymbol{\lambda}}(t) = \dot{O}_{\boldsymbol{\lambda}}(t) - \langle \dot{O}_{\boldsymbol{\lambda}}(t) \rangle_{\boldsymbol{\lambda}}$ and rewriting the right-hand side of Eq. 2.10 as

$$\int_0^{t_f} dt \int_0^{t_f} dt' \langle \delta \dot{O}_{\boldsymbol{\lambda}}(t) \delta \dot{y}_\chi(t') \rangle_{\boldsymbol{\lambda}}$$

$$= \int_{-t_f}^0 d\tau \int_{-\tau}^{t_f} dt' \langle \delta \dot{O}_{\boldsymbol{\lambda}}(\tau + t') \delta \dot{y}_\chi(t') \rangle_{\boldsymbol{\lambda}} + \int_0^{t_f} d\tau \int_0^{t_f - \tau} dt' \langle \delta \dot{O}_{\boldsymbol{\lambda}}(\tau + t') \delta \dot{y}_\chi(t') \rangle_{\boldsymbol{\lambda}} \tag{2.14}$$

$$= \int_{-t_f}^0 d\tau \int_{-\tau}^{t_f} dt' \langle \delta \dot{O}_{\boldsymbol{\lambda}}(\tau) \delta \dot{y}_\chi(0) \rangle_{\boldsymbol{\lambda}} + \int_0^{t_f} d\tau \int_0^{t_f - \tau} dt' \langle \delta \dot{O}_{\boldsymbol{\lambda}}(\tau) \delta \dot{y}_\chi(0) \rangle_{\boldsymbol{\lambda}} \tag{2.15}$$

where at first we have replaced $\tau = t - t'$ and changed the limits of integration appropriately,
and then used time-translation invariance in the nonequilibrium steady-state. We now note
that $\langle \delta \dot{O}_{\boldsymbol{\lambda}}(\tau) \delta \dot{y}_\chi(0) \rangle_{\boldsymbol{\lambda}}$ is zero for all $\tau < 0$, as the Malliavin weight $\dot{y}_\chi(0)$ is proportional to
$\boldsymbol{\eta}_i(0)$ which is causally independent from all past values of $\dot{O}_{\boldsymbol{\lambda}}$. Hence the first term vanishes
and the second term simplifies to,

$$\int_0^{t_f} dt \int_0^{t_f} dt' \langle \delta \dot{O}_{\boldsymbol{\lambda}}(t) \delta \dot{y}_\chi(t') \rangle_{\boldsymbol{\lambda}} = \int_0^{t_f} (t_f - \tau) d\tau \langle \delta \dot{O}_{\boldsymbol{\lambda}}(\tau) \delta \dot{y}_\chi(0) \rangle_{\boldsymbol{\lambda}} \tag{2.16}$$

Dividing both sides by t_f, taking the limit of $t_f \to \infty$ and assuming a rapidly decaying
correlation function in the steady-state gives the second term in Eq. 2.10 as

$$\lim_{t_f \to \infty} \frac{1}{t_f} \left\langle O_{\boldsymbol{\lambda}}[\mathbf{X}] \frac{\partial \ln P_{\boldsymbol{\lambda}}[\mathbf{X}]}{\partial \chi} \right\rangle_{\boldsymbol{\lambda}} = \int_0^\infty dt \left\langle \delta \dot{y}_\chi(0) \delta \dot{O}_{\boldsymbol{\lambda}}(t) \right\rangle_{\boldsymbol{\lambda}} \tag{2.17}$$

where the gradient is proportional to an integrated time-correlation function. This is an
example of a generalized fluctuation-dissipation formula [124].

Putting together the two contributions,

$$\lim_{t_f \to \infty} \frac{1}{t_f} \frac{\partial}{\partial \chi} \langle O_{\boldsymbol{\lambda}} \rangle_{\boldsymbol{\lambda}} = \left\langle \frac{\delta \dot{O}_{\boldsymbol{\lambda}}[\mathbf{X}]}{\delta \boldsymbol{\lambda}} \cdot \frac{\partial \boldsymbol{\lambda}}{\partial \chi} \right\rangle_{\boldsymbol{\lambda}} + \int_0^{\infty} dt \left\langle \delta \dot{y}_{\chi}(0) \delta \dot{O}_{\boldsymbol{\lambda}}(t) \right\rangle_{\boldsymbol{\lambda}}, \qquad (2.18)$$

we arrive at an explicit form for the gradient of our SCGF estimate with respect to the variational parameters that can be estimated as time-averages from a straightforward molecular dynamics trajectory with the control forces. In practice, we will take the integral over the time correlation function in Eq. (2.18) up to a time Δt. The exact form we use to evaluate this correlation function on-the-fly is obtained from Eq. 2.17 by translating the correlation time-lag $[0, t]$ to $[-t, 0]$, replacing the dummy variable t by $-t$ and carrying out the integration explicitly, as

$$\int_0^{\infty} dt \left\langle \delta \dot{y}_{\chi}(0) \delta \dot{O}_{\boldsymbol{\lambda}}(t) \right\rangle_{\boldsymbol{\lambda}} = \int_0^{\infty} dt \left\langle \delta \dot{y}_{\chi}(-t) \delta \dot{O}_{\boldsymbol{\lambda}}(0) \right\rangle_{\boldsymbol{\lambda}} = \int_{-\infty}^{0} dt \left\langle \delta \dot{y}_{\chi}(t) \delta \dot{O}_{\boldsymbol{\lambda}}(0) \right\rangle_{\boldsymbol{\lambda}} \quad (2.19)$$

$$\approx \left\langle \delta y_{\chi}(0+) \delta \dot{O}_{\boldsymbol{\lambda}}(0) \right\rangle_{\boldsymbol{\lambda}} - \left\langle \delta y_{\chi}(-\Delta t) \delta \dot{O}_{\boldsymbol{\lambda}}(0) \right\rangle_{\boldsymbol{\lambda}} \qquad (2.20)$$

Here the time $0+$ denotes the next infinitesimal timestep after time 0, and Δt is large enough such that the estimator is accurate upto a desired tolerance. We propagate the dynamics of y_{χ} and keep a history upto lag-time Δt, such that we can statically correlate with $\dot{O}_{\boldsymbol{\lambda}}$ at every timestep. The baseline subtraction at Δt and the subtraction of the averages for computing the correlations from joint expectations helps keep the variance of the gradient well-behaved, even in a steady-state where $y_{\chi}(t)$ is a random walk with a time-extensive variance [122].

We also note here that when $\langle O_{\boldsymbol{\lambda}} \rangle_{\boldsymbol{\lambda}}$ is represented as Eq. 2.4, the first term in the gradient in Eq. 2.18 is zero. This is seen readily by inserting Eq. 2.4 into Eq. 2.18 to get the first term as

$$\left\langle \frac{\delta \dot{O}_{\boldsymbol{\lambda}}[\mathbf{X}]}{\delta \boldsymbol{\lambda}} \cdot \frac{\partial \boldsymbol{\lambda}}{\partial \chi} \right\rangle_{\boldsymbol{\lambda}} = \left\langle \int_0^{t_f} dt \sum_i \frac{\boldsymbol{\lambda}_i - \gamma_i \dot{\mathbf{r}}_i + \mathbf{F}_i}{2\gamma_i k_{\mathrm{B}}T} \cdot \frac{\partial \boldsymbol{\lambda}}{\partial \chi} \right\rangle_{\boldsymbol{\lambda}}$$

$$= \left\langle \int_0^{t_f} dt \sum_i -\frac{\boldsymbol{\eta}_i(t)}{2\gamma_i k_{\mathrm{B}}T} \cdot \frac{\partial \boldsymbol{\lambda}(t)}{\partial \chi} \right\rangle_{\boldsymbol{\lambda}} = 0 \qquad (2.21)$$

due to causality. It can be shown that this term is also proportional to the derivative of the normalization constant for the trajectory probabilities with respect to changing the force, and this is zero [125]. However, if one uses the form of $\langle O_{\boldsymbol{\lambda}} \rangle_{\boldsymbol{\lambda}}$ from Eq. 2.5, the first term in Eq. 2.18 is not zero and must be independently computed as a static average.

When the explicit gradients are averaged over a long steady-state trajectory at every optimization step, the noise in the gradients become small compared to their means and they can be directly used to perform variants of gradient descent. We perform an iterative optimization in the parameter space spanned by $\{\chi\}$ in order to estimate the SCGF. We use

an algorithm called Nesterov's Accelerated Gradient Descent [126, 127] which shows super-linear convergence. This algorithm updates a conjugate momentum p^χ in parameter space to accelerate optimization with inertia, as well as prevents inertial oscillations with a predictive step. The conjugate momentum is a sum of the scaled gradient with scaling parameter α^χ, and the scaled momentum from the previous optimization step with scaling parameter ν^χ. These parameters can be chosen at the start of the optimization or can be tuned with iterations to improve convergence. We stop either a fixed number of optimization steps or when the norm of the force gradients are less than a tolerance value. The optimization algorithm is given in pseudocode form in Algorithm 1.

This algorithm converges to a local optimum in the parameter space, which can be different from the global optimum when the variational surface or its projection on a truncated basis space is not convex. The convergence can also be significantly slow at values of s near a crossover point or a phase transition. In the event that we converge to a local optimum, we incur a systematic error in the SCGF that can be corrected with a cumulant expansion.

2.3   Corrections of systematic errors

In general, the ansatz specified by the parameters $\{\chi\}$ will not form a complete basis for a many body system. This is because generically, the dominant eigenvector of Eq. 1.16 is a many-body state, containing exponentially many parameters, and not expected to be exactly expressible with a low rank form. Because of this, the variationally converged SCGF $\widetilde{\Psi}(s)$ obtained from Eq. 2.6 with an optimized force $\widetilde{\boldsymbol{\lambda}}$ will have a systematic error. This error, and errors associated with convergence to a local maximum, can both be corrected in principle by computing the remaining terms of the cumulant expansion

$$\Psi(s) = \widetilde{\Psi}(s) + \lim_{t_f \to \infty} \frac{1}{t_f} \sum_{\ell=2}^{\infty} \frac{\kappa_\ell}{\ell!} \tag{2.22}$$

where $\{\kappa_\ell\}$ are the second and higher cumulants in the expansion of $\ln\langle \exp(O_{\widetilde{\boldsymbol{\lambda}}}) \rangle_{\widetilde{\boldsymbol{\lambda}}}$ and the force $\widetilde{\boldsymbol{\lambda}}$ is the solution of the variational problem in the approximate and incomplete ansatz. If the ansatz used to express the control force, $\widetilde{\boldsymbol{\lambda}}$, is close enough to the optimal force obtained from the Doob transform, the correction terms are small in magnitude and the series will converge quickly. This will occur when the trajectory distribution generated by the controlled dynamics has significant overlap with the tilted distribution of the original dynamics.

In cases where the ansatz is poor and many cumulants are needed, brute force convergence of the correction will be difficult. In such cases, control forces can be used as guiding functions for estimating the SCGF through Monte Carlo based approaches like the cloning algorithm. In the cloning algorithm, an ensemble of N_w trajectories generated from the ordinary path probabilities $P[\mathbf{X}(t_f)]$ are branched with corresponding weights of $\exp(-st_f A_{t_f})$. However, under the controlled dynamics, following Eq. 2.1, the weighted path probabilities can be

---

Algorithm 1 Nesterov's Accelerated Gradient Descent with Malliavin weights

---

1: **inputs** Variational parameters for a general force $\boldsymbol{\lambda}_\chi(\mathbf{r}^N, t)$

2: **parameters** Force gradient and momentum learning rates α^χ, ν^χ; total optimization steps I; trajectory length t_f consisting of J timesteps of duration δt each; number of trajectories N

3: **initialize** choose initial weights χ, define iteration variables i and j, force gradients and conjugate momenta $\delta_{\boldsymbol{\lambda}}^\chi$ and p^χ, define functional form for stepwise differential increments (rewards) ξ to the loss-function $st_f A_{t_f} + \Delta U_{\boldsymbol{\lambda}}$

4: $i \leftarrow 0$

5: $p^\chi \leftarrow 0$

6: **repeat**

7:     Generate trajectory $[\mathbf{X}(t)]$ with first-order Euler propagation in presence of the additional force $\boldsymbol{\lambda}[\{\chi + \nu^\chi p^\chi\}]$ and wait till it relaxes into a nonequilibrium steady-state. Configurations, times, noises (with variance $2\gamma k_{\mathrm{B}} T \Delta t$), Malliavin weights, and rewards are denoted by $\mathbf{r}_j^N, t_j, \boldsymbol{\eta}_j, y_\chi(t_j)$ and $\xi(t_j) = \xi_j$ respectively.

8:     $j \leftarrow 0$

9:     $\delta_{\boldsymbol{\lambda}}^\chi \leftarrow 0$

10:     $y_\chi(t_0) \leftarrow 0$

11:     **repeat**

12:         $\dot{y}_\chi(t_j) \leftarrow \boldsymbol{\eta}_j \cdot \nabla_\chi \boldsymbol{\lambda}_\chi(\mathbf{r}_j^N, t_j)/2k_{\mathrm{B}} T \Delta t$

13:         $y_\chi(t_{j+1}) \leftarrow y_\chi(t_j) + \Delta t \dot{y}_\chi(t_j)$

14:         $\delta_{\boldsymbol{\lambda}}^\chi \leftarrow \delta_{\boldsymbol{\lambda}}^\chi + \xi_j y_\chi(t_{j+1})$

15:         $j \leftarrow j + 1$

16:     **until** $j = J$

17:     $\delta_{\boldsymbol{\lambda}}^\chi \leftarrow \delta_{\boldsymbol{\lambda}}^\chi/J$

18:     average $\delta_{\boldsymbol{\lambda}}^\chi$ over N trajectories to get $\overline{\delta}_{\boldsymbol{\lambda}}^\chi$

19:     $p^\chi \leftarrow \nu^\chi p^\chi - \alpha^\chi \overline{\delta}_{\boldsymbol{\lambda}}^\chi$

20:     $\chi \leftarrow \chi + p^\chi$

21:     $i \leftarrow i + 1$

22: **until** $i = I$

---

written as [84, 89]

$$P_s[\mathbf{X}(t_f)] \propto e^{-st_f A_{t_f}} P[\mathbf{X}] = e^{O_{\boldsymbol{\lambda}}[\mathbf{X}]} P_{\boldsymbol{\lambda}}[\mathbf{X}] \tag{2.23}$$

where system evolution under an approximate controlled dynamics is nonconservative and must be accompanied by branching steps with weights given by $\exp(O_{\boldsymbol{\lambda}})$. An estimate of the SCGF is then obtained from the normalization constant of this weight, so that in the limit of large N_w,

$$\Psi(s) = \frac{1}{t_f} \ln \frac{1}{N_w} \sum_{j=1}^{N_w} e^{O_{\boldsymbol{\lambda}}[\mathbf{X}_j]} \tag{2.24}$$

where $O_{\boldsymbol{\lambda}}[\mathbf{X}_j]$ denotes the time-integrated observable for the walker labelled as j.

When the variationally optimized $\widetilde{\boldsymbol{\lambda}}$ is used to generate trajectories and to compute the branching probabilities, the efficiency of the cloning algorithm is improved as the control force samples the rare fluctuations in the observable. When $\widetilde{\boldsymbol{\lambda}}$ is actually the optimal force $\boldsymbol{\lambda}_s^*$ derived from the Doob transform, all trajectories achieve the rare fluctuation as typical behavior, and the weight of each trajectory becomes a constant. In this situation no trajectories are killed in the branching step of the cloning algorithm, and the sampling is statistically optimal [77]. However, even with an approximate ansatz the variationally optimized force slows down the rate of death of uncorrelated trajectories with increasing t_f, as demonstrated in Section 2.5.

The variational algorithm along with the cumulant-correction has improved scaling properties compared to the cloning algorithm. By adopting an approximate ansatz for the many-body force containing, for example, one-body and two-body terms, for a system of identical particles we can exploit their permutation symmetry and optimize a single one-body and two-body force. Hence the variational algorithm scales linearly with the system size, the computational cost arising only from the propagation of trajectories of interacting particles. This is in contrast to the cloning algorithm, which has an exponential scaling for observables that are system size extensive [81]. Also, while the cloning algorithm scales exponentially with s, the variational algorithm depends on the bias only through the complexity of the optimal force and scales linearly with the number of variational parameters required to approximate the force. Hence in cases that the dominant part of the optimal force can be simply expressed within the choice of the ansatz, the computational cost for the algorithm to converge does not increase with s. This indicates a resummation of the exponential bias through the modification of the control force. Neither does the algorithm scale with increasing observation time t_f, as the $t_f \to \infty$ limit has already been incorporated in the algorithm. Lastly, this algorithm can be parallelized trivially by distributing the computation of the expectation values at each step of the iteration to independent trajectories on independent processors.

To study the accuracy and efficiency of our variational algorithm to compute the SCGF and the optimal force, in the next sections we apply it to three different continuous time and space systems. The first is a benchmark system where we can test our algorithm against a numerically exact result. This model consists of a driven underdamped particle in a

one-dimensional periodic potential, for which we have studied rare fluctuations of the total current. The second system is comprised of multiple repulsive overdamped particles, where we have focused on the fluctuations of the total activity, which measures how much the particles explore configuration space. In this system, we demonstrate the ability of our algorithm to compute the optimal control force even through singular changes in the SCGF across a dynamical phase transition. The third system consists of multiple repulsive particles in a periodic potential having a wavelength equal to the particle diameter and an external nonequilibrium force, such that it is a Brownian analogue of the Asymmetric Simple Exclusion Process (BASEP). We study its current SCGF and find that many-body effective forces are necessary to generate rare backward current fluctuations at high packing fractions.

2.4 Current fluctuations in non-interacting overdamped/underdamped system

An underdamped particle being driven on a periodic potential by a constant external force is a simple system with two dynamical coordinates, position and velocity, that can exhibit non-trivial nonequilibrium properties due to competing ballistic and diffusive modes of transport [128, 129]. Large deviation functions for current fluctuations in this model can be obtained by numerically exact diagonalizations of the tilted generator, and the controlled ensemble can show diverse behavior in different parameter regimes [130]. We consider this model to benchmark our variational optimization algorithm.

Specifically, we consider an underdamped particle of mass m moving in a one-dimenional periodic box of length $L = 2\pi$. The forces acting on the particle are derived from a cosine potential, $V(x) = V_0 \cos(x)$, where V_0 is the magnitude of the potential, and include a constant external driving force, F_{ext}. For the particle in contact with a bath of temperature, T, and friction coefficient, γ, the equations of motion for the position, x, and velocity, v, are

$$\dot{x} = v$$
$$m\dot{v} = F(x) - \gamma v + \eta \tag{2.25}$$

where $F(x) = -V'(x) + F_{\text{ext}}$ and $\eta(t)$ is a Gaussian white noise with

$$\langle \eta(t) \rangle = 0 \qquad \langle \eta(t)\eta(t') \rangle = 2\gamma k_{\text{B}} T \delta(t - t') \tag{2.26}$$

where k_{B} is Boltzmann's constant. These equations of motion have the form of Eqs. (1.1) and (1.3) with two dynamical coordinates and a vanishing noise in position [130]. We investigate the statistics of the time-averaged current flowing through the system,

$$J_{t_f} = \frac{1}{t_f} \int_0^{t_f} dt \, v(t) \tag{2.27}$$

which measures the total displacement of the particle. The SCGF for current is given by

$$\Psi(s) = \sup_{u(x,v)} \frac{1}{t_f} \left\langle \int_0^{t_f} dt \left(-sv + \frac{u^2 - F^2 - 2\gamma v(u - F)}{4\gamma k_B T} - \frac{m\dot{v}(u - F)}{2\gamma k_B T} \right) \right\rangle_u \tag{2.28}$$

where the path average is obtained from the controlled dynamics

$$m\dot{v} = u(x,v) - \gamma v + \eta \tag{2.29}$$

and the optimal force is in general a function of both position and velocity. We expand the additional force $\lambda(x,v) = u(x,v) - F(x)$ in an ansatz

$$u(x,v) = F(x) + \sum_{p=-M_1}^{M_1} \sum_{q=0}^{M_2} c_{p,q} e^{ipx} v^q \tag{2.30}$$

where c_{pq} are parameters that can be optimized variationally subject to $c^*_{-p,q} = c_{p,q}$, and the number of position and velocity basis functions are $(2M_1 + 1)$ and $(M_2 + 1)$ respectively. The basis is complete in the limit of large M_1 and M_2. Note that this force incorporates the periodicity of x and also allows the external nonequilibrium driving, which is the $p = q = 0$ term, to be optimized. In the high friction limit, the dynamics becomes overdamped and in that limit the optimal force becomes a function of just the particle position. For small friction, inertia is important and the general form of the optimal force must be considered. We note that this velocity-dependent drift function is a *force* only in a generalized sense.

The SCGF and the optimized control force obtained from the variational algorithm can be compared to numerically exact results obtained by solving the eigenvalue equation for the tilted generator given by [130]

$$L_s = v\frac{\partial}{\partial x} - \frac{1}{m}[v\gamma - F(x)]\frac{\partial}{\partial v} + \frac{\gamma k_B T}{m^2}\frac{\partial^2}{\partial v^2} - sv \tag{2.31}$$

as in Eq. (1.16). The exact control force is obtained using the right eigenvector $\phi_s(x,v)$ corresponding to the largest real eigenvalue, as

$$u(x,v) = F(x) + \frac{2\gamma k_B T}{m}\frac{\partial \ln \phi_s(x,v)}{\partial v} \tag{2.32}$$

where numerical diagonalization of L_s can be performed by expressing the right and left eigenvectors over a position-velocity grid and representing the differential operators in L_λ using a second order finite difference scheme. The boundary conditions are periodic in the position grid and reflective in the velocity grid, so that only forward (backward) difference at the minimum (maximum) velocity grid point is used to represent the differential operator.

We have computed the cumulant-corrected large deviation functions in this system and have compared them to the numerically exact results. We have worked with $k_B T = 1$ and $\gamma = 1$. These parameters along with the length of the box $L = 2\pi$ let us define our natural
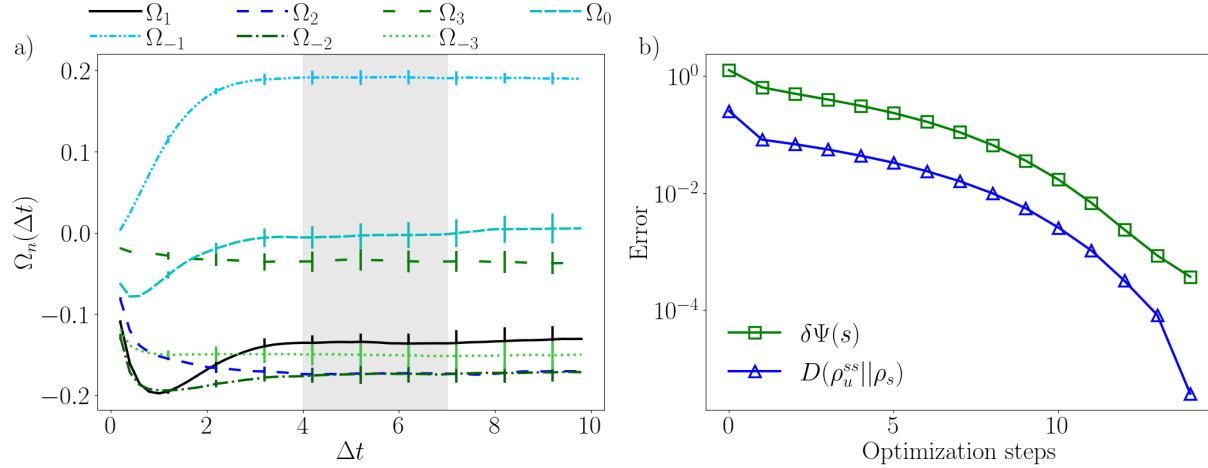
Figure 2.1: a) Convergence of $\Omega_n(\Delta t)$ for $s = -0.5$, for different n. Shaded region represents optimal choice of Δt for gradient descent. b) Simultaneous convergence of SCGF and biased density for $s = -0.5$.

time unit as $t^* = 4\pi^2 k_B T / \gamma L^2$. All observables have been reported in dimensionless units following these definitions. We have done our computations at two values of mass, $m/\gamma t^* = 1$ and $m/\gamma t^* \to 0$. We have also chosen $V_0 = 2$ and $F_{ext} = 1$. The numerically exact result was obtained with a grid of 140×50 points in the position-velocity space. The position points span all of the box and the velocity points are centered at $(F_{ext} - 2sk_B T)/\gamma$ corresponding to the mean velocity in the $V_0 \to 0$ limit. For all the simulations, the timestep was chosen to be 0.001 natural time units. For $m/\gamma t^* \to 0$, an Euler scheme was used to integrate the overdamped equation of motion, while for $m/\gamma t^* = 1$, a velocity Verlet scheme was used [42].

For each iterative step during the optimization, a trajectory of duration 10^4 units was simulated. During the first half of each trajectory, the system was allowed to come to a steady-state, and the time-averaged gradients were computed only with the second half of the trajectories. For implementing Eq. 2.18, we integrated the correlation function up to $\Delta t = 100$. The size of the basis was $M_1 = 3, M_2 = 1$ for $m/\gamma t^* = 1$ and $M_1 = 3, M_2 = 0$ for $m/\gamma t^* \to 0$, the overdamped limit. The optimization parameters used for the gradient descent were $\mu = 0.5, \nu = 0.2$. Near $s = 0$, all c_{pq} were initialized at zero, and subsequent optimizations with increasing magnitude of s were initialized from a previously optimized set of c_{pq} taken from the nearest value of s. In the overdamped limit, an accurate estimate of the SCGF could be obtained with just the variational optimization, with the cumulant correction merely a confirmation of the optimal control forces being correct. However for $m/\gamma t^* = 1$, the variational SCGF had to be corrected with cumulants computed with an observation time $t_f = 100$ and a total trajectory length 10^5 units.

The choice of a finite integration limit Δt to compute the integral in Eq. 2.17 depends on both the intrinsic correlation times of the system and the timescale of the variance of the
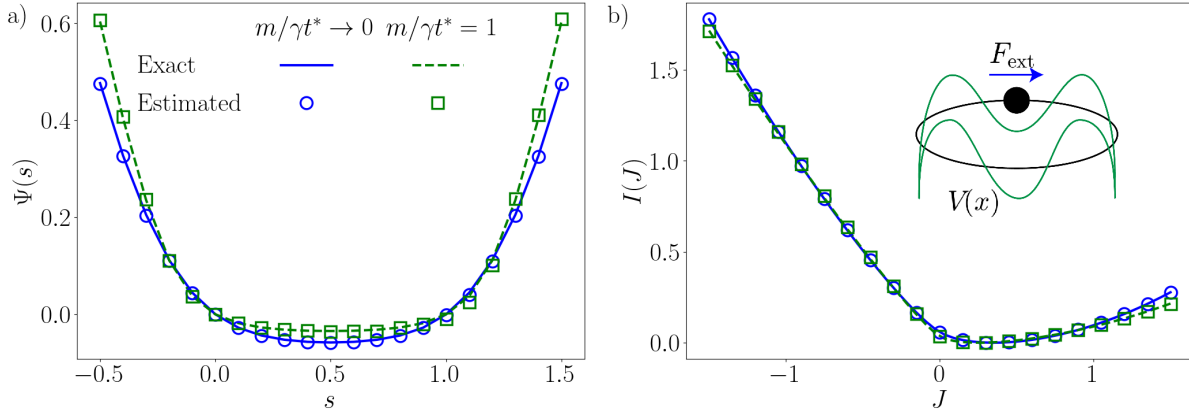
Figure 2.2: Large deviation functions for current fluctuations in a driven underdamped system in a periodic potential. a) SCGF for $m/\gamma t^* = 1$ with $M_1 = 3, M_2 = 1$ and for $m/\gamma t^* \to 0$ with $M_1 = 3, M_2 = 0$. b) Rate functions obtained by a numerical Legendre-Fenchel transform of the SCGFs. The legend is the same as that used in a). (*Inset*) Schematic diagram of the simulated system.

integrated correlation function to diverge. To illustrate this, we plot

$$\Omega_n(\Delta t) = \int_0^{\Delta t} dt \left\langle \delta \dot{y}_n(0) \delta \dot{O}_{\boldsymbol{\lambda}}(t) \right\rangle_{\boldsymbol{\lambda}} \tag{2.33}$$

for this system in the $m/\gamma t^* \to 0$ limit. The ansatz can be written in this limit as

$$u(x) = F(x) + c_0 + \sum_{n=1}^{3} [c_n \cos x + c_{-n} \sin x] \tag{2.34}$$

and for Fig. 2.1, we have chosen c_n parameter values randomly between -1 and 1, with $s = -0.5$. We see that even though the correlation function converges for large Δt, the error in the computed gradient increases steadily. For all the results in this paper, Δt was chosen to balance between these two effects so that the computed gradients suffers from no systematic error and minimum statistical error.

Using these numerically estimated gradients, the accelerated gradient descent algorithm converges superlinearly, and in Fig. 2.1 we have plotted the decrease of the systematic error $\delta\Psi(s)$ in the current SCGF estimate with optimization steps in the limit $m/\gamma t^* \to 0$. We also show the simultaneous convergence of the controlled ensemble steady-state density $\rho_u^{ss}(x)$ to the true biased steady-state density $\rho_s(x) \propto \Theta_s(x)\phi_s(x)$ where Θ_s and ϕ_s are the dominant left and right eigenvectors of the tilted generator (2.31). We demonstrate this by plotting the relative entropy of the two,

$$D(\rho_u^{ss}||\rho_s) = \int dx \, \rho_u^{ss}(x) \log\left(\frac{\rho_u^{ss}(x)}{\rho_s(x)}\right) \tag{2.35}$$
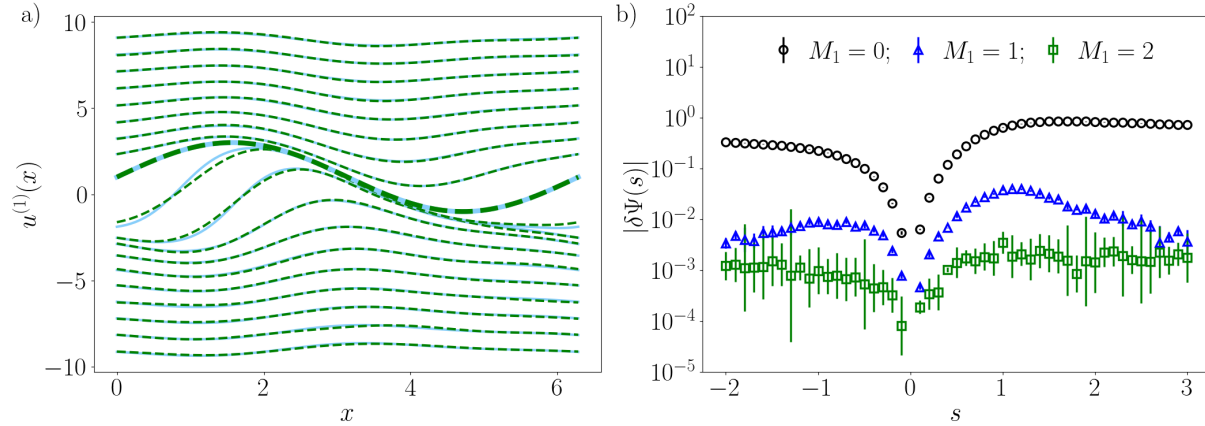
Figure 2.3: Overdamped limit, $m/\gamma t^* \to 0$, of the driven particle on a periodic potential. a) Optimized control forces (dashed lines) overlaid on the exact control force (solid lines). The thick curve is for $s = 0$ and the curves above (below) are for s in intervals of $+0.5(-0.5)$. b) Basis size errors in the variational estimate of $\Psi(s)$, where the deviation $\delta\Psi(s) = \widetilde{\Psi}(s) - \Psi(s)$ is the difference between the finite basis result $\widetilde{\Psi}(s)$ from the exact SCGF.

which shows that even as only the current is being optimized to have a nontypical value, the entire trajectory ensemble simultaneously converges to the exact biased ensemble.

Following this procedure, we obtain estimates of SCGFs that are in quantitative agreement with the numerically exact results throughout the range of s considered, as shown in Fig. 2.2(a). We have also calculated the rate functions for the current, Fig. 2.2(b), in these two parameter regimes by a numerical Legendre-Fenchel transform of the SCGFs.

The SCGFs in Fig. 2.2(a) both have a *locked* region where the current changes slowly with s, and an *unlocked* region for larger magnitudes of s. Due to the time-reversal properties of L_s, the SCGF shows a Gallavotti-Cohen symmetry [131]

$$\psi(\lambda) = \psi(-F_{\text{ext}}/k_{\text{B}}T - \lambda) \tag{2.36}$$

which is clear through the reflection symmetry about $s = 0.5$ of the SCGF in Fig. 2.2(a). Analogously, the rate function obeys a fluctuation theorem symmetry

$$I(J) = I(-J) + F_{\text{ext}}J/k_{\text{B}}T \tag{2.37}$$

indicating the exponentially rare probability of a current in the direction opposite to the applied force.

Figure 2.3(a) shows the position-dependent optimal forces obtained in the overdamped limit, $u^{(1)}(x)$, overlaid on the numerically exact answers obtained from diagonalization [132], for multiple values of s. In the limit of $|s| \to \infty$, the optimal forces approach the free-diffusion limit, where the majority of contribution comes from a constant nonequilibrium driving.
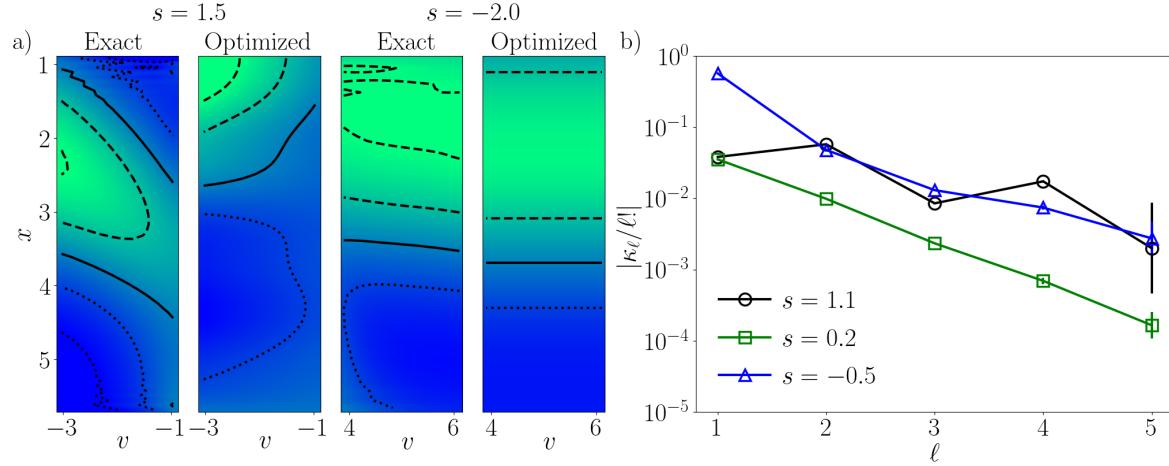
Figure 2.4: Underdamped system, $m/\gamma t^* = 1$, of the driven particle on a periodic potential. a) (L-R) Exact and optimized control forces, $u(x, v)$, for $s = 1.5$, with the solid contour at $u(x, v) = -2$, and the dashed (dotted) contours being at differences of $+1$ (-1). Exact and optimized control forces, $u(x, v)$, for $s = -2$ with the solid contour at $u(x, v) = 5$ and the dashed (dotted) contours being at differences of $+1$ (-1). b) Convergence of the cumulant expansion for representative values of s.

When $|s|$ is of the order $|F_{\text{ext}}|/k_{\text{B}}T$, the forces have a non-trivial position dependence. This is manifested in the size of the basis-set, M_1, required to obtain the optimal control force accurately. Figure 2.3(b) shows the effect of finite basis size on the error made in estimating $\Psi(s)$. Increasing M_1 reduces the error and ultimately the ansatz becomes exact when M_1 is large. The error decreases when going to larger $|s|$ as the forces are easier to represent using the first few basis functions. The error bars were computed from 5 independent estimates of the SCGF using independent trajectories.

For the $m/\gamma t^* = 1$ system, inertial effects are important and the optimal force depends on both position and velocity, and the optimal force has a complicated functional dependency that is difficult to represent using a small number of basis functions. Using a truncated basis to represent the control force leads to a systematic error in the SCGF estimate obtained using Eq. 2.28 that can be corrected using the cumulant expansion in Eq. (2.22). Figure 2.4(a) shows the approximate forces obtained from the variational optimization compared to the numerically exact results. When s is near the Gallavotti-Cohen symmetry point, the average current is small and the optimal control force is a complicated function of both v and x. Within our ansatz, the optimized $u(x, v)$ does not reproduce the exact form of the optimal control force. Nevertheless, these approximate forces recover the majority of the SCGF, so that the cumulant expansion converges for all tested s points. Figure 2.4(a) also contains the optimal force at a larger positive s, where the forces lose their velocity dependence and simplify towards the free-diffusion limit. In this limit, position based forces are sufficient to

recover the SCGF quantitatively.

Figure 2.4(b) shows the convergence of the consecutive terms of the cumulant expansion in Eq. 2.22 for different values of s. κ_1, the first cumulant, is identical to $\widetilde{\Psi}(s)$, the variational estimate. Error bars were calculated using 5 independent trajectories for the estimation of the cumulants. Even though our basis is small and approximate, the cumulants computed from a single trajectory have decreasing amplitudes for various values s, showing that the variational force is accurate enough to approach the force derived from the Doob transform. We note that the sign of the cumulants need not be positive, and therefore the variational structure in the estimate of $\Psi(s)$ holds only for the first cumulant. Further, the magnitude of the terms in the cumulant expansion need not be strictly decreasing. Figure 2.4(b) includes an example of a nonmonotonic convergence for $s = 1.1$. Moreover, the sign of the error of the approximate SCGF at a given truncation of the cumulant expansion can change resulting in the cancellation of errors of two oppositely signed cumulant corrections and an accidental near agreement of the exact SCGF. We have found that by considering the convergence of the consecutive terms of the cumulant expansion we can reliably determine the accuracy of the approximate SCGF.

2.5   Activity fluctuations in interacting overdamped system

To study how this algorithm performs in an interacting system, we consider the fluctuations of the activity in a system of overdamped repulsive particles on a line. In both lattice and continuum models of volume excluding particles in one dimension, it has been reported that there are two characteristically distinct types of activity fluctuations, with a dynamical phase transition separating them [133]. For rare large negative values of the activity, such systems spontaneously phase separate into macroscopically sized clusters, whereas for rare small values of the activity, they form a hyperuniform phase in which long-wavelength density fluctuations are suppressed. This behavior emerges as a singularity in the SCGF and a closing of the gap in the eigenspectrum of the tilted operator, which in the hydrodynamic scaling limit is predicted to occur with a critical point at $s_c \to 0^+$ [133, 89]. This system is thus suitable to test the effectiveness of the variational algorithm in computing rare fluctuations that are collective in origin.

Specifically, we study the fluctuations of dynamical activity in a system of N overdamped repulsive particles in a one-dimensional periodic box of length L. The equation of motion is

$$\gamma \dot{x}_i = F_i(\mathbf{x}) + \eta_i \tag{2.38}$$

where $F_i(\mathbf{x})$ is the total force felt by the i-th particle,

$$F_i(\mathbf{x}) = -\frac{\partial}{\partial x_i} \sum_{j \neq i} V_{\mathrm{WCA}}(x_{ij}) \tag{2.39}$$

where $x_{ij} = x_i - x_j$ and the force is derived from a WCA pair potential

$$V_{\text{WCA}}(r) = \left[4\epsilon \left(\frac{\sigma^{12}}{r^{12}} - \frac{\sigma^6}{r^6} \right) + \epsilon \right], \quad r < 2^{1/6}\sigma \tag{2.40}$$

$$= 0, \quad r \geq 2^{1/6}\sigma$$

with characteristic energy, ϵ, and length scale, σ. The Gaussian white noise, η_i, is specified by

$$\langle \eta_i(t) \rangle = 0, \qquad \langle \eta_i(t)\eta_j(t') \rangle = 2\gamma k_{\text{B}}T\delta_{ij}\delta(t - t') \tag{2.41}$$

We work with $k_{\text{B}}T = 0.5$, $\gamma = 1$ and $\sigma = 1$. As before, we define our unit of time for this system as $2k_{\text{B}}T/\gamma\sigma^2$ and we have reported all observables in dimensionless units. Additionally, we set $\epsilon = 1$ and consider a density of $\rho = N\sigma/L = 0.5$, so that the box is half-filled.

We study a measure of activity derived from the probability that the particles stay in the same state in a short time interval [134]. This form of the activity,

$$K_{t_f} = \frac{1}{t_f} \int_0^\tau dt \sum_i \left(\frac{F_i^2}{4\gamma k_{\text{B}}T} + \frac{1}{2\gamma}\frac{\partial F_i}{\partial x_i} \right) \tag{2.42}$$

is also a part of the time-symmetric component of the path-action [135], and its long time statistics are similar to other commonly used metrics that count the total number of hops for particles on a lattice [136, 137]. Using Ito's Lemma to simplify the last term in Eq. 2.2, the variational expression for the SCGF becomes

$$\Psi(s) = \sup_{\mathbf{u}(x_1,x_2,...,x_N)} \frac{1}{t_f} \left\langle \int_0^{t_f} dt \sum_i \left[-s \left(\frac{F_i^2}{4\gamma k_{\text{B}}T} + \frac{1}{2\gamma}\frac{\partial F_i}{\partial x_i} \right) \right.\right.$$
$$\left.\left. + \frac{u_i^2 - F_i^2}{4\gamma k_{\text{B}}T} + \frac{1}{2\gamma}\frac{\partial(u_i - F_i)}{\partial x_i} \right] \right\rangle_{\mathbf{u}} \tag{2.43}$$

where in addition to the force, we require the gradient of both the original and the control force.

For this system, the optimal control force $\mathbf{u}(\mathbf{x})$ is in general long-range and many-bodied. Previous work on related one-dimensional systems have shown long-range repulsive interactions stabilizing the hyperuniform state for values of activity small in magnitude [95], and long-range attractive forces acting on the surface of particle clusters that emerge in rare large negative fluctuations of the activity [89]. For our variational ansatz, we have approximated the many-body force as a sum of long-range pairwise interactions. Pair forces are the lowest rank approximation to this system due to its translational invariance. From the Hopf-Cole transform, optimization of a pair force is analogous to optimization of a two-body Jastrow function as used in variational quantum Monte Carlo [138].

To represent the control force, we expand it in a basis of Laguerre polynomials L_p with coefficients c_p as

$$u_i = \sum_{j \neq i} \left[-\frac{\partial}{\partial x_i}V_{\text{WCA}}(x_{ij}) + \sum_{p=1}^{M_3} c_p L_p(\tilde{x}_{ij})e^{-\tilde{x}_{ij}/2}\frac{x_{ij}}{|x_{ij}|} \right] \tag{2.44}$$
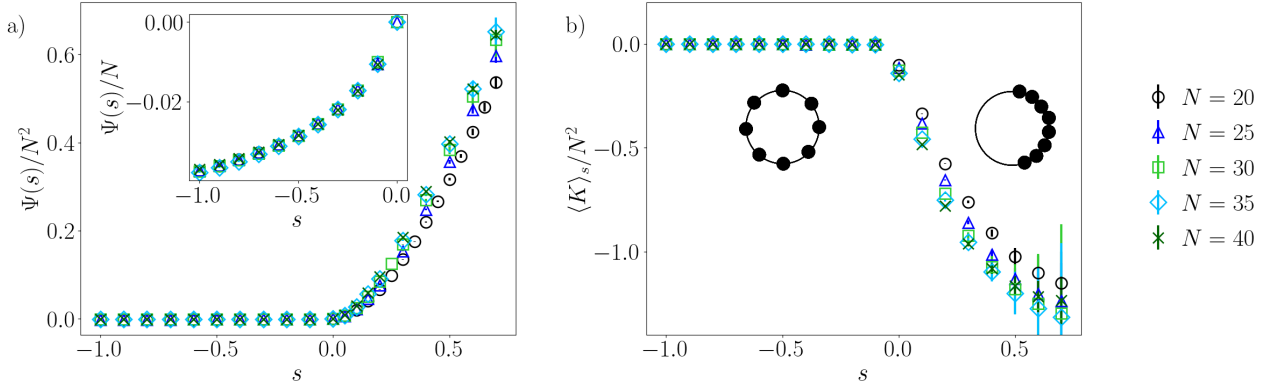
Figure 2.5: Size-scaling of activity fluctuations of repulsive particles on a line. a) $O(N^2)$ scaling of $\Psi(s)$ in the phase-separated state. (*Inset*) $O(N)$ scaling in the hyperuniform state. b) Change in mean activity across the dynamical phase transition. (*Inset*) Schematic representation of the phase-separated (left) and hyperuniform (right) states.

where $\tilde{x}_{ij} = \alpha - \beta|x_{ij}|$ is a linear transformation on the distance between particles i and j. The parameters α and β can be adjusted to set a scale and a cutoff for where the force smoothly decays to zero, and M_3 determines the size of the basis. The basis is complete for all possible two-body forces in the limit of large M_3. The exponential factor makes the basis functions orthogonal and aids in the convergence of the optimization. We have used $M_3 = 10$ for all of our results. We have fixed $\beta = 2/L$, and optimized $\{c_p\}$ and α with starting values of 0 and $L/2$ respectively. In each iteration of the optimization, a trajectory of length 2×10^4 time units is simulated, the first half again reserved for equilibration and the second half being used to compute the gradients. For computing the integrated correlation function in Eq. 2.18, we have used $\Delta t = 200$ units. After obtaining the optimized control force in this ansatz, we use it to compute the unbiased SCGF using a cumulant expansion as before, with an observation time $t_f = 10$ and a total trajectory length of 5×10^4 units. Across the range of s considered, we find convergence using the first three cumulants to correct the variational result. The SCGF obtained from this cumulant expansion is identical to results obtained using a guided cloning algorithm that has been described later in this section.

In Figure 2.5(a) we have plotted the size scaled SCGF, and the mean activity, for positive and negative values of s. For $s < 0$, we find the system in a hyperuniform state, where all particles are pushed apart from each other and long-range density fluctuations are suppressed [133]. The SCGF is size-extensive in this range of s. For $s \gg 0$ the particles phase separate, forming a single cluster. In the region where s is positive but small, there is a phase transition to this clustered state accompanied by an inflection point in the mean activity, shown in Fig. 2.5(b), obtained from taking the numerical derivative of the SCGF, $\langle K \rangle_s = \Psi'(s)$. The extensive scaling regime has been explored systematically in a related model and found to
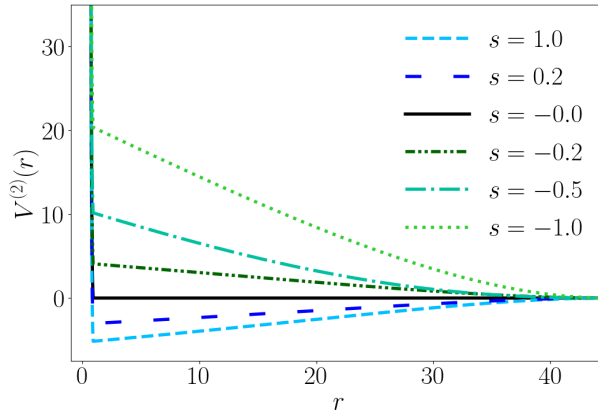
Figure 2.6: Optimal pair-potential for positive and negative s for $N = 40$.

agree well with predictions from macroscopical fluctuation theory [89]. In our studies, we find it limited to $0 < s < 0.02$. For large positive values of s, the cluster is a highly compressed solid with system-spanning correlations that result in the SCGF scaling super-extensively. In this regime of the SCGF, the typical force is on the order of $\sqrt{N}$, and can continue to increase with increasing s because of the soft repulsion of the WCA potential. Inspection of the distribution of mean squared forces reveals that the cluster is not homogeneous, but most compressed in its interior with lower density near the edges, with a system size independent profile. The phase transition from a disordered state to a clustered state is in accord with previous observations in related systems, and result in diverging correlation times rendering the precise study of the critical point difficult [89, 133]. We therefore focus our attention on the two phases on either side of that transition. Error bars were obtained from independent statistics from 3 distinct trajectories.

Figure 2.6 shows the effective pair-potential, $V^{(2)}(r)$, derived from the optimal control force at different values of s, for $N = 40$, obtained by the numerical integration of the control force. The potential is long-ranged and repulsive in the hyperuniform phase, and long-ranged and attractive in the clustered phase. The long-range potential leads to the observed size scaling in Fig. 2.5, because it imposes infinite range correlations. We also observe that the depth of the attractive potential for increasingly positive values of s tends to saturate, while the magnitude of the repulsive potential for increasingly negative s does not. This difference arises from the steeply rising WCA forces that can achieve more negative values of $\langle K \rangle_s$ with just a slight decrease in the nearest neighbor distance in the controlled system. In the hyperuniform phase, achieving the rarer values of activity implies an exponentially small number of collisions between the particles, which leads to an increasing repulsive control force. These optimal control forces derived from the variational ansatz do not contain many-body components unlike analytically derived approximate forces [89], yet they achieve the same phenomenology of phase separation and hyperuniformity described previously.
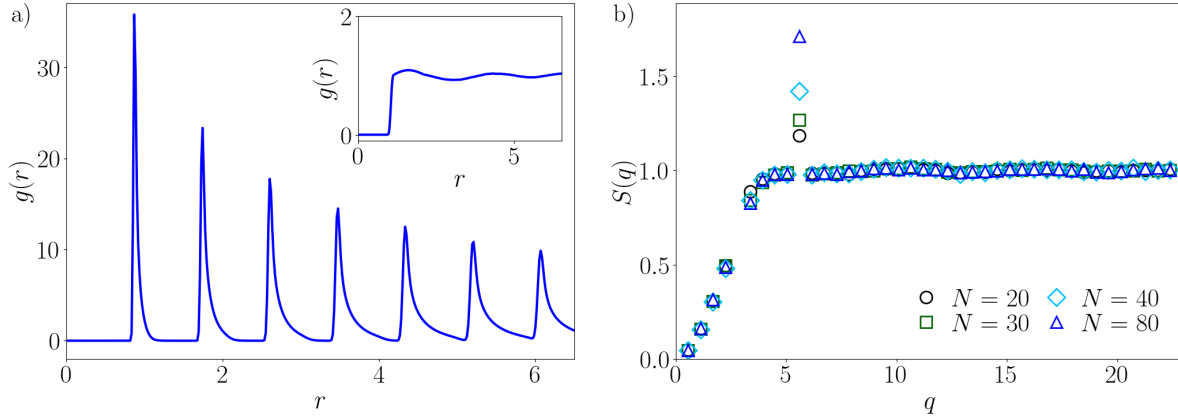
Figure 2.7: Characterization of the two dynamical phases for $N = 80$. a) Pair distribution functions within the phase separated, $s = 0.1$, and (*inset*) hyperuniform, $s = -0.1$, states. b) Structure factor for various system sizes in the hyperuniform state, $s = -1$.

Figure 2.7(a) characterizes the steady-state radial distribution function $g(r)$,

$$\rho g(r) = N \langle \delta(r - |x_{12}|) \rangle_{\mathbf{u}} \tag{2.45}$$

obtained in these phases, for a system size of $N = 80$, where x_{12} denotes the interparticle distance between each distinct pair of particles. In the phase-separated state, the particles form a solid cluster that has sharp peaks in $g(r)$ at intervals of σ. In the hyperuniform phase, the particles are repelled away from each other and $g(r)$ has little structure aside from the volume-exclusion. We also characterize the structure of the hyperuniform state through the structure factor, $S(q)$, as a function of the wavenumber q, obtained from

$$S(q) = \frac{1}{L} \left\langle \left| \sum_{j=1}^{N} e^{-iqx_j} \right|^2 \right\rangle_{\mathbf{u}} \tag{2.46}$$

where the averages are computed in the ensemble with the control force. A linear increase of $S(q)$ from zero at small q is a signature of the suppression of long-wavelength density fluctuations in the hyperuniform phase, which we confirm in Fig. 2.7(b). The spike at $q = 2\pi/2^{1/6}\sigma$ results from $2^{1/6}\sigma$ being the distance of closest approach of the repulsive particles without experiencing a force.

Under large positive activity bias, we find that the overdamped repulsive particles form a highly compressed cluster. This cluster is described by system-spanning correlations. Shown in Fig. 2.8 is the size-scaled profile for the first term of the collective activity (2.42), $\langle F_i^2 \rangle_s / 4\gamma k_{\mathrm{B}} T$, with respect to a size-scaled particle index $N_i = i - (N+1)/2$. The particles are indexed from one end of the cluster to the other, such that the center of the cluster is indexed at $N_i = 0$. The compressed cluster does not break apart during the duration of the
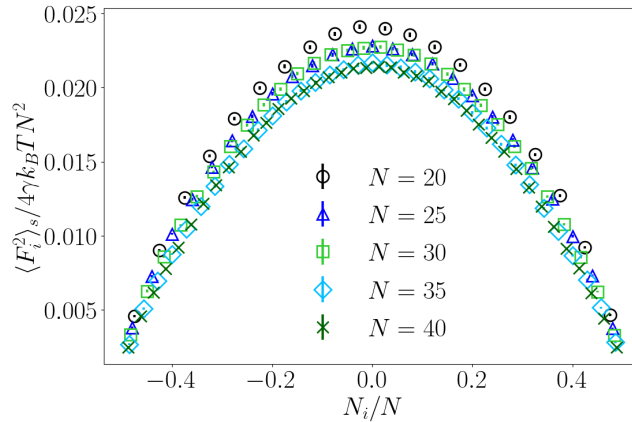
Figure 2.8: Size-scaling of the mean squared force profile within the cluster for $s = 0.1$.

trajectories observed, so that large $|N_i|$ unambiguously refers to particles close to the surface of the cluster. The total mean activity $\langle K \rangle_s$ is proportional to the total mean squared force appearing in the first term, such that the profile of the second term in the definition looks analogous only with an opposite sign [137]. The $O(N^2)$ scaling of the mean squared force and its size-invariant parabolic profile explains the super-extensive SCGF scaling and the system spanning correlations in this s regime.

While we have not investigated the phase transition directly, the disparate behavior of either side of the dynamical phase transition provides a useful test of our ability to obtain control forces, as the structure and dynamics of the system in the phase separated and hyperuniform states are very different. Despite their differences in both regimes, we are able to obtain control forces that are near enough to the optimal force to converge the large deviation functions using a brute force evaluation of the remaining cumulant expansion. Nevertheless, we expect this strategy may fail in general, in which case a more robust means of estimating the remaining contribution must be employed. To explore such alternatives, we apply these control forces as guiding functions within the cloning algorithm [84]. To quantify the statistical benefit from the control forces, we start with a trajectory ensemble of $N_w = 32000$ walkers and monitor the decay rate in the number of uncorrelated walkers, N_c, with and without the control forces. The number of uncorrelated walkers is defined as those with a distinct history, having not been previously merged into an existing walker. Figures 2.9(a) and (b) show the statistics of the walkers with respect to observation time, with and without the control forces, in a system with 20 particles, and branching steps taken every 0.5 time units. We have plotted $f_s^c(t) = N_c(t)/N_w(t)$, where t is the observation time, to represent the growth of correlation in the trajectory ensemble.

In the clustered state, incorporating the control forces improves the number of uncorrelated walkers by multiple orders of magnitude. For larger positive s, an unbiased estimate of the SCGF can be obtained only when the variational control forces are used. The improve-
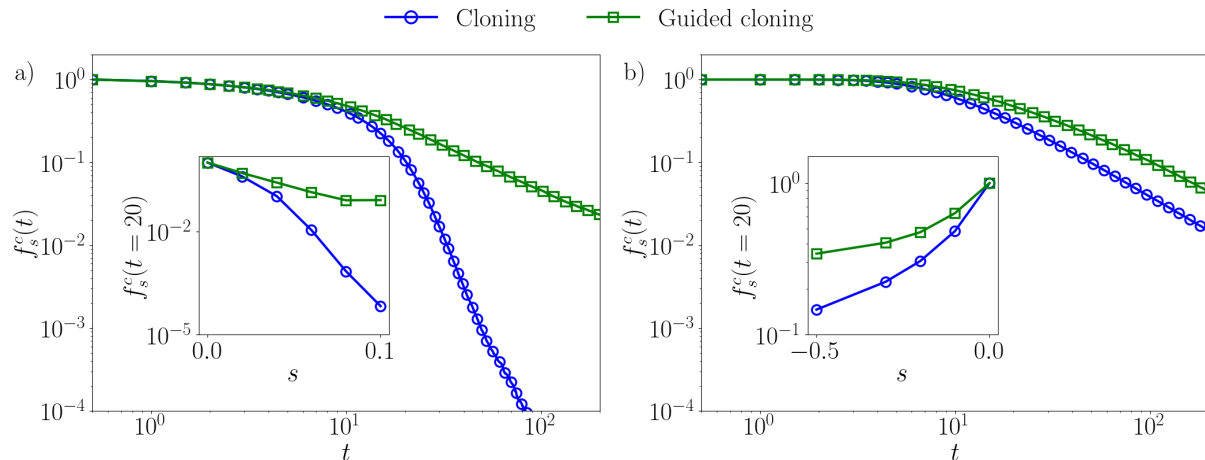
Figure 2.9: Improvement of walker statistics of the cloning algorithm using approximate control forces as guiding functions in an N=20 system, represented by $f_s^c(t) = N_c(t)/N_w(t)$, after an observation time t. Blue circles are without a guiding force and green squares are with the variationally optimized guiding force. Decay of the fraction of uncorrelated walkers with increasing observation time in a) the phase-separated state ($s = 0.04$) and b) the hyperuniform state ($s = -0.2$). (*Insets*) Decay of the fraction of uncorrelated walkers after $t = 20$ as a function of s in a) the phase-separated state ($s = 0.04$) and b) the hyperuniform state ($s = -0.2$).

ment in the statistics of the walkers increases for more positive s because the magnitude of the SCGF grows rapidly, and therefore the weight carried by the branching step increases. We see this effect in the inset, where we show the fraction of uncorrelated walkers left after an observation time and how it varies with s [84].

The decay of the walkers depends on the overlap between the tilted trajectory ensemble and that generated from the controlled dynamics. Slower decay will result when the control dynamics generates a trajectory ensemble that is close, in this sense, to the tilted trajectory ensemble. This behavior is analogous to other approximate guiding function based importance sampling, such as that arrived by iterative feedback [86] or analytical approximation [89]. These effects are seen in the hyperuniform phase as well, albeit the decay of walkers in the ordinary cloning algorithm is less drastic, and so is the improvement by incorporating the guiding forces. The improvement in statistical efficiency upon including the optimized forces is not restricted to the cloning algorithm, and could be analogously adopted within transition path sampling [89] or forward flux sampling [102].

2.6 Current fluctuations in interacting overdamped system

The Asymmetric Simple Exclusion Process (ASEP) has been extensive studied as a model for many-body systems in nonequilibrium statistical physics [139]. ASEP is defined on a one-dimensional lattice where each site can accommodate at most one particle, and the particles hop forward or backward to neighboring sites with asymmetric rates without overtaking each other. Boundary-driven nonequilibrium steady-states in ASEP have been well-characterized to contain several dynamical phases that arise from both local density and collective current order-parameters [140, 141]. Conditioning trajectory ensembles on large deviations of currents have been previously shown to induce long-range effective forces that decay logarithmically [95, 142]. Specialized tools using matrix product states and other tensor network techniques have also recently been developed to sample rare current fluctuations and characterize the dynamical phase behavior in ASEP and its generalization to two dimensions [121, 143].

Despite being defined only on a discrete lattice in one spatial dimension, ASEP has been widely used to study the phenomenology of the nonequilibrium stochastic transport of biomolecular cargo inside the cell [144]. The movement of ribosomes on RNA during protein synthesis [145], the active traffic of load-carrying molecular motors on microtubules [146] and the single-file transport of ions and small molecules through ion-channels [147] have been mapped into ASEP to characterize their dynamical regimes and their energetic efficiency. The mapping between ASEP and these systems comprising of continuous space hopping between a sequence of metastable wells becomes accurate in the limit that barrier heights are large compared to thermal energy. However, rare current fluctuations are often collective and involve correlated hops involving neighboring multiple particles, for which the introduction of infinitesimal fluctuations in inter-particle distances may change the phase behavior. Indeed, dynamical phases featuring current reversals have been previously observed in lattice models but not found in its continuous space analogue in a model for stochastic pumps [148, 149]. To better understand the continuous space generalization of ASEP and to characterize its driven phase behavior, Brownian ASEP (BASEP) has been recently introduced as a model for single-file transport [150, 151]. BASEP consists of hard volume-excluding spheres diffusing on a one-dimensional periodic potential driven by a constant external force. This model features rich structural phenomogy depending on the flux of incoming and outgoing particles, average density and the commensurability of particle diameters with the wavelength of the periodic potential. But the rare current fluctuations, response and the dynamical phase behavior of BASEP trajectory ensembles have not been studied and compared against those of ASEP to determine the latter's applicability in studying real space dynamics.

In this section we have used our variational algorithm to construct the large deviation functions for currents in a soft-sphere analogue of the BASEP. We have generalized our model we used in Section 2.4 for describing multiple interacting particles in a periodic potential.

We consider N diffusive particles governed by the overdamped Langevin equation,

$$\gamma \dot{x}_i = F_i(x^N) + \eta_i \tag{2.47}$$

in a one-dimensional box of length $2N\pi/\rho$, where ρ is the packing fraction of the system, F_i is the total force on the i-th particle, and η_i is a Gaussian white noise with zero mean and variance $2\gamma k_{\mathrm{B}}T$. Each particle experiences a periodic potential $V(x_i) = V_0 \cos(x_i)$, a constant external driving force F_{ext}, and a pairwise WCA repulsion $V_{\mathrm{WCA}}(x_i - x_j)$ from all the other particles, similar to the previous section in Eq. 2.40. We consider only the case $\sigma = 2\pi$, such that each potential well can accommodate at most one particle without high energetic cost. We have defined a collective current J over a trajectory $\mathbf{X}(t_f)$ of duration t_f as

$$J = \frac{1}{t_f} \int_0^{t_f} dt \sum_i \dot{x}_i dt \tag{2.48}$$

The collective current can thus be also interpreted as the drift in the center of mass of the particles. A peculiarity of working with a continuous description of space is that if $V_0 = 0$, momentum is well-defined and conserved modulo the effect of the thermal noise. The current in the $V_0 = 0$ limit is exactly as if each particle is being driven independently by the external force F_{ext}. This means that the effect of having a high density of particles only becomes manifest in the statistics of average collective current through a non-zero external potential energy function. When V_0 is non-zero and comparable to $k_{\mathrm{B}}T$, at high ρ non-zero current is generated from sequential correlated hops from collections of adjacent particles. Such current-carrying mechanism is shown in Fig. 2.10(a) via the collection of particle coordinates $\{x_i\}$ as a function of time t at parameters $\rho = 0.5$, $k_{\mathrm{B}}T = 0.5$, $V_0 = 1$ and $F_{\mathrm{ext}} = 1.5$.

For a system of $N = 100$ particles, we compute the SCGF $\Psi(s)$ associated with biasing this time-averaged current, by applying our variational algorithm to compute an optimal force. As basis functions we choose a combination of one-body periodic functions and two-body pairwise functions, $u_i = F_i(x) + u_i^{(1)}(x) + \sum_{j \neq i} u_i^{(2)}(x_i - x_j)$, that are described by Equations 2.34 and 2.44. We study the convergence of the variational estimate for the SCGF at three values of packing fraction, a dilute limit, an intermediate regime and a dense limit where every well is occupied. We have learnt this force with $M_1 = 5$ and $M_3 = 5$ basis functions for bias s in the range of $[-2, 5]$. We have worked with units $k_{\mathrm{B}}T = 0.5$, $\gamma = 1$ and our natural time unit as $t^* = 2\rho^2 k_{\mathrm{B}}T/\gamma N^2$ which evaluates to the same timescale defined in Section 2.4. All observables have been reported in dimensionless units following these definitions. We have also chosen $V_0 = 1$ and the external driving force $F_{\mathrm{ext}} = 1$. For all the simulations, the timestep was chosen to be 0.001 natural time units and an Euler scheme was used to integrate the overdamped equation of motion. For each iterative step during the optimization, a trajectory of duration 5×10^7 units was simulated. Similar to our computations with the single overdamped particle, the system was allowed to come to a steady-state for the first half of the trajectory, and the time-averaged gradients were
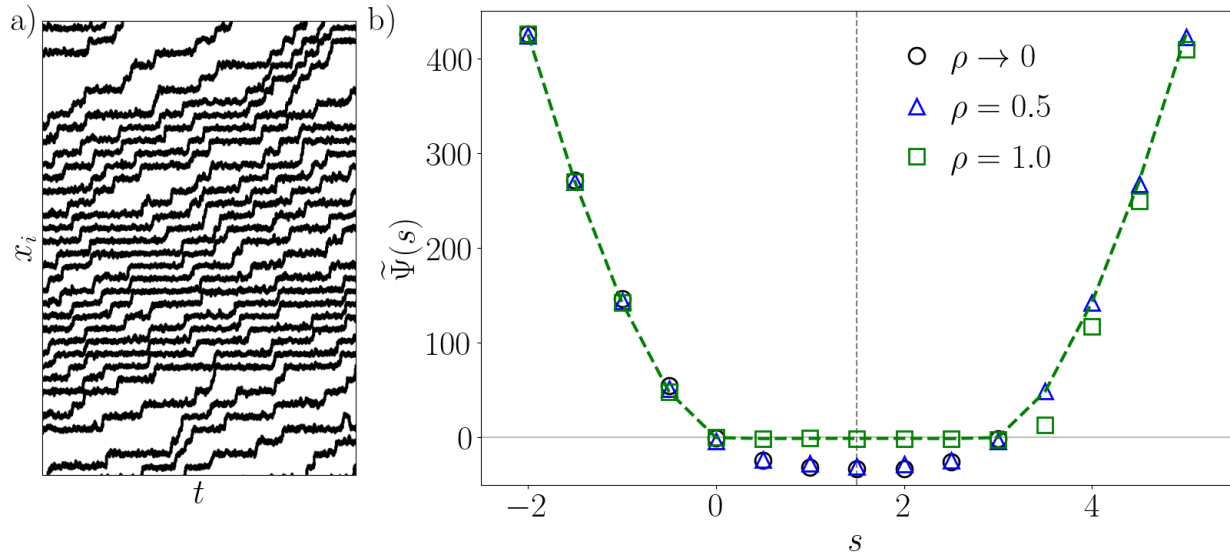
Figure 2.10: a) Snapshot of the many-particle trajectory ensemble at half-filling, demonstrating correlated sequential hops that lead to non-zero current. b) Variational SCGF estimators for $\rho = 0$, 0.5 and 1.0. Dotted vertical line represents Gallavotti-Cohen symmetry point at $s = 1.5$. Green dashed line represents the segment of SCGF for $\rho = 1$, $s < 1.5$ reflected to the region $s > 1.5$.

computed only with the second half of the trajectories. For implementing Eq. 2.18, we integrated the correlation function up to $\Delta t = 100$. We chose optimization learning rates $\mu = 0.1, \nu = 0.0$.

In Fig. 2.10(b) we have shown the variational SCGF estimates obtained for our system at three different densities. We find that the SCGF follows the similar locked and unlocked regimes we had seen for the single particle case, which is identical to the dilute limit. The SCGF does not change much from the dilute limit to $\rho = 0.5$, and only changes appreciably near full packing At a high density, the biased current value in the locked part of the SCGF reduces from the low-density limit. This is manifested as a sharp change at $s = 0$, which is arising from a diverging second cumulant of current, the latter being proportional to the second derivative of the SCGF. For any density, the SCGF should respect a Gallavotti-Cohen reflection symmetry about the line $s = 1.5$. This is respected for small and intermediate densities, but when the box is fully packed, we see this symmetry not being bveyed by the variational estimator, which implies that the optimal force cannot be adequately represented within our basis set ansatz. Nevertheless, given the variational principle and that the symmetry must hold for the correct $\Psi(s)$, we get an estimate of the error being made by reflecting one-half of the SCGF on the other half. This construction is shown as the green dashed line

in Fig. 2.10(b).

Despite the percentage error being small, we could not correct the variational estimate with a converging cumulant expansion or a guided cloning procedure with statistics from upto 10^9 timesteps. This is due to the fluctuations in the system being extremely rare, and also hints at the converged forces being quite far from the actual optimal force in the system. Looking closer into the one-body and two-body components of our optimized force, we find that for all values of s the two-body component optimizes to a negligible value compared to the one-body parts. This suggests that a pairwise two-body force ansatz is a poor approximation for the many-body nature of the true optimal force. Improved representations of many-body effective forces through neural networks [152] might be promising as better basis sets for optimal forces generating rare current fluctuations in this system.

2.7 Conclusion

We have developed a variational algorithm to compute optimal control forces for Langevin models driven into nonequilibrium steady-states. We have used the control forces to sample rare fluctuations in time integrated dynamical observables like current and activity, in order to compute large deviation functions, and shown that they can be used to improve the efficiency of the cloning algorithm. Our variational algorithm, along with the correction of the systematic error with the cumulant expansion, has improved scaling properties compared to trajectory ensemble methods, and can be useful in dealing with many-particle chemical or biological systems.

Though we worked with Langevin models of structureless particles, the algorithm is straightforward to generalize to higher dimensions, where optimal control forces might have significant rotational components. It can also be extended to lattice models, where the rate matrix has to be expressed in a variational ansatz. A system modeled by a different stochastic equation of motion, like that employing an Andersen thermostat [42] or quantum trajectory-based approaches [153, 154], can also be treated through this algorithm by changing only the functional forms of the path-actions provided a Doob transformation exists.

The versatility of the variational algorithm allows for its use with different force ansatzes. In the activity-biased system, using a low-rank approximation for a many-body optimal control force was sufficiently accurate. However in cases where the control force is not expressible in a simple functional form or even as a many-body expansion, machine learning using artificial neural networks could be used to approximate it. The variational algorithm relies on evaluating functional derivatives of the force with respect to the parameters, which can be automated with autodifferentiation algorithms [155], as has already been demonstrated in equilibrium free energy calculations [156]. The use of techniques developed in this paper can aid the formulation of such optimization algorithms in the future. Additionally, this algorithm can be used for model reduction in high-dimensional systems [157], and hence to extend Variational Force-Matching and Ultra Coarse Graining algorithms [158, 159, 160] out

of equilibrium, so that biomolecular and other soft matter systems can be simulated over large length and time scales with effective forces in nonequilibrium steady-states.

Lastly, this framework of solving the optimal forces can tackle inverse-design problems out of equilibrium. Various inverse-design algorithms have been proposed that can obtain optimal forces to rationalize materials design with targeted properties and to guide directed self-assembly of smaller objects [161, 162]. Our variational algorithm can be used to obtain optimal forces suitable for targeted assembly or tailored particle distributions when nonequilibrium driving forces are present, and hence can be used to characterize and predict dynamical phases in new functional materials.

Chapter 3

Inverse design of nonequilibrium colloidal self-assembly

The self-assembly of soft and biological matter out of equilibrium can result in novel structures and dynamical responses not constrained by thermodynamic considerations [163, 164, 165, 166, 167, 33]. The microscopic violation of detailed balance in such systems can be used to design a wide range of functional materials with enhanced thermomechanical, optoelectronic or drug-delivery properties [168, 169, 170]. Predictive inverse design to drive the assembly of target dissipative structures requires a dynamical description of the system [171, 172, 173, 48, 174]. In this chapter we apply our variational algorithm to automate the discovery of inverse design principles for colloidal self assembly in a nonequilibrium steady-state in molecular dynamics simulations. The algorithm uses a variational principle arising from rare dynamical fluctuations of the system in a trajectory ensemble, and optimizes the yield of target clusters or their rate of formation with statistically estimated explicit gradients in the design parameter space. We demonstrate the performance of this algorithm by obtaining optimal design principles for the self-assembly of DNA-labeled colloids [175] driven out of equilibrium by a shear flow. We expect that the ability to uncover general optimal inverse design principles away from equilibrium will enable bottom-up synthesis of new materials and elucidate the processes encoding structure in biological contexts [176, 177, 18].

Self-assembly of nanoscale building blocks is increasingly used to engineer functional materials with novel properties arising from their complex nanostructures. Colloidal systems offer a versatile paradigm for inverse design towards a desired target structure due to the independent tunability of shape, valency and assembly environment [178]. In thermodynamic equilibrium, stabilizing a target structure amounts to lowering its free energy, which is typically achieved by increasing both the interaction strength and specificity. This principle has been exploited to achieve the self-assembly of a variety of clusters and superlattices from colloids and nanocrystals with crystal facets decorated with organic ligands or DNA [179, 180, 8]. However, employing these principles in practice requires mitigating dynamical effects like slow coarsening and kinetic trapping [173, 172]. Optimal forces for self-assembly must achieve a trade-off between slow relaxation at high interaction strengths, and slow

growth at high interaction specificity [181, 182]. Self-assembly of colloids and biomolecules in nonequilibrium steady-states provide a route to decouple kinetics from stability and mitigate this tradeoff. Directive self-assembly has been achieved by driving the system with a constant supply of chemical fuel, or by applying external fields [183, 184, 185, 186]. However, the design of such systems must confront the continuous supply of energy necessary to prevent the system from relaxing to equilibrium. Existing computational methods to discover inverse design principles for nonequilibrium self assembly are limited due to the configurational probability not following the Boltzmann distribution and the corresponding variational structure afforded by the free energy no longer being valid under such dissipative conditions.

Recent advances in the theoretical treatment of the stochastic thermodynamics of nonequilibrium steady-states have made possible a trajectory ensemble description of self-assembly, treating structure and dynamics on an equal statistical footing [27, 19]. This has enabled basic principles governing assembly away from equilibrium to be formulated [22, 74]. In this work we develop a perspective and accompanying numerical technique based on these insights. Rather than considering the probability of observing a state and tuning its associated free energy, we consider the likelihood that a trajectory forms a specific structure as quantified by a stochastic action, and how that action is changed by modifying the intermolecular and applied forces. We show that fluctuations around a nonequilibrium steady-state encode the susceptibility of a system to assemble, in a manner analogous to a fluctuation-dissipation relationship. Further, optimal forces that assemble a target structure or maximize reactive flux while minimizing the change to the stochastic action satisfy a variational principle [63, 111]. We extend and apply our optimization algorithm to solve this variational expression and compute the optimal control force to sample rare dynamical phases [117]. We show that this algorithm can be used to solve the inverse design problem, deciphering how rare fluctuations encode stability away from equilibrium.

We outline below an inverse design algorithm for the self-assembly of sheared DNA-coated colloids into different target nanoclusters. The algorithm is based on a variational principle relating rare fluctuations in an ensemble of trajectories conditioned on evolving a target structure, to effective forces achieving the target as the typical dynamical state. Working with a trajectory ensemble, where the probability distribution is known, circumvents the difficulty of not knowing the distribution of configurations within a nonequilibrium steady-state. To solve the variational problem, we have used generalized response relations for the gradients of the steady-state trajectory probability to a change in the inter-particle and externally driven forces.[1]

---

[1]Most of the content of this chapter was originally part of the publication [187].

3.1 Coarse-grained model for DNA-labeled colloids

For concreteness, we consider a model of N colloidal particles in a cubic box of length L, evolving with an overdamped Langevin equation of the form,

$$\gamma\dot{\mathbf{r}}_i = \mathbf{u}_i + \boldsymbol{\eta}_i \tag{3.1}$$

where $\dot{\mathbf{r}}_i$ are time derivatives of the coordinates of the i-th particle and $\mathbf{u}_i$ are the forces acting on it. The friction coefficient of the colloids with the thermal bath is denoted γ and $\boldsymbol{\eta}_i$ are Gaussian white noise that satisfy

$$\langle\boldsymbol{\eta}_i(t)\rangle = 0\,, \quad \langle\boldsymbol{\eta}_i(t)\boldsymbol{\eta}_j(t')\rangle = 2\gamma k_{\mathrm{B}}T\mathbf{I}_3\delta_{ij}\delta(t-t') \tag{3.2}$$

where $\mathbf{I}_3$ is the 3×3 identity matrix and $k_{\mathrm{B}}T$ is Boltzmann's constant times the temperature. The angular brackets denote an averaging operation over the random noise distribution. As we consider dynamics in the presence of a shear flow, we use Lees Edwards boundary conditions [188].

We use an ansatz of DNA-labeled spherical isotropic colloids as programmable building blocks for self-assembly. The interaction between these colloids, mediated by the DNA molecules attached to their surface, consists of a volume-exclusion repulsion and a short-range attraction [189]. The effective interaction strengths and the pairwise specificity can be independently tuned by varying the sequences of the grafted DNA molecules. During self-assembly, the short-range forces generate a competition between local and global order that leads to frustration and unique phase behavior and dynamical effects [190, 191]. This system has been computationally and experimentally demonstrated to form finite nanoclusters with specific target structures [192, 193]. The high-dimensional design space has the possibility to offer multiple pathways to stabilize any cluster out of the many nearly degenerate states formed without the specificity of the DNA-mediated attraction. To illustrate the performance of the variational algorithm, we consider the nonequilibrium self-assembly of 21 such rigid and nonrigid clusters, some examples of which are demonstrated in Figure 3.1a.

We examine the self-assembly of these colloidal particles under a constant linear shear flow. Shear flows are known phenomenologically to alter the stability of compact and extended colloidal structures [194, 195]. A recent paradigm of colloidal assembly being increasingly explored is that in a microfluidic device, where the confining walls generate a strong shear on the assembling clusters [196, 197]. This system offers a canonical nonequilibrium setting to explore inverse design principles. Taken together, the forces acting on the i-th colloid are

$$\mathbf{u}_i = \mathbf{f}_i^{\mathrm{S}}(\mathbf{r}_i) - \boldsymbol{\nabla}_i\sum_{j\neq i} V(r_{ij}) \tag{3.3}$$

where $V(r) = V_{\mathrm{WCA}}(r) + V_{\mathrm{Morse}}(r)$ and $V_{\mathrm{WCA}}(r_{ij})$ is a WCA pair potential representing the volume exclusion interactions, and $V_{\mathrm{Morse}}(r_{ij})$ denotes the DNA-mediated short-range pairwise attraction. The force due to a shear flow, $\mathbf{f}_i^{\mathrm{S}}(\mathbf{r}_i)$, has the form

$$\mathbf{f}_i^{\mathrm{S}}(\mathbf{r}_i) = fz_i\hat{\mathbf{x}} \tag{3.4}$$
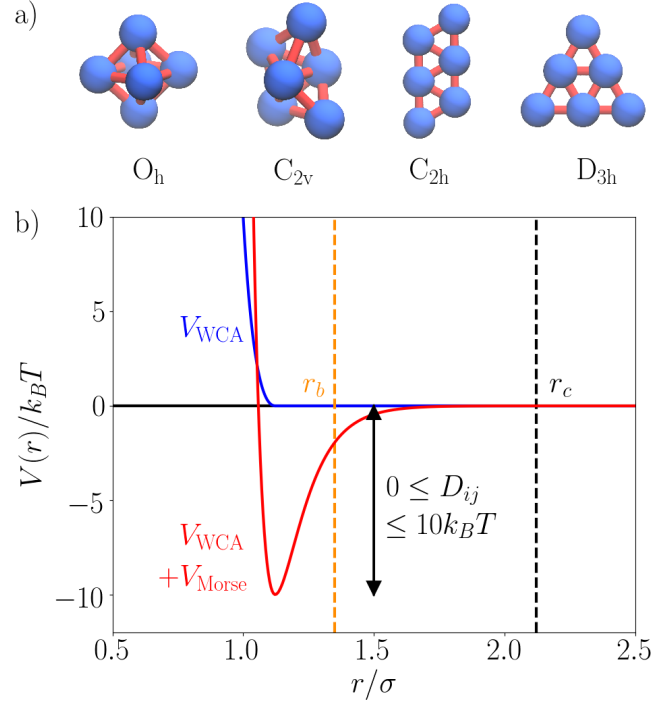
Figure 3.1: Model details for self-assembly of DNA-labeled colloids. a) Examples of finite
rigid and nonrigid nanoclusters for which we have studied design principles, along with the
corresponding point groups for molecular symmetries denoted underneath. b) Graphical
forms of the potential energy functions, the WCA potential (blue) and the WCA and Morse
potential combined (red) for $D_{ij} = 10k_{\mathrm{B}}T$. The orange dashed line denotes the bond cutoff
$r_b = 1.35\sigma$ and the black dashed line denotes the potential cutoff $r_c = 2.12\sigma$.

which has magnitude f and generates a constant gradient of the x component of the velocity
along the z direction. The WCA pair potential has the functional form

$$V_{\mathrm{WCA}}(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^{6} \right] + \epsilon, \quad r_{ij} < 2^{1/6}\sigma$$

$$= 0, \quad r_{ij} \geq 2^{1/6}\sigma \tag{3.5}$$

with particle diameter σ and energy scale ϵ. The attractive Morse potential has the functional
form

$$V_{\mathrm{Morse}}(r_{ij}) = D_{ij} \left(e^{-2\alpha(r_{ij}-2^{1/6}\sigma)} - 2e^{-\alpha(r_{ij}-2^{1/6}\sigma)} \right) \tag{3.6}$$

where D_{ij} is the magnitude of the bond energy and α determines its width.

We work in units of $k_{\mathrm{B}}T = 1$, $\gamma = 1$ and $\sigma = 1$. The natural time scale with these
units is $t_0 = \gamma\sigma^2/k_{\mathrm{B}}T$ and times are expressed in these units throughout. We set $\epsilon = 10k_{\mathrm{B}}T$

and $\alpha^{-1} = \sigma/10$. The attractive energy scale D_{ij} and the shear flow rate f are tuned as variational parameters to induce self-assembly. They have been restricted to vary within the range $0 \leq D_{ij} \leq 10k_{\mathrm{B}}T$ and $0 \leq f \leq 50k_{\mathrm{B}}T/\sigma^2$ to avoid large relaxation times and to stay within the overdamped regime. Figure 3.1b shows the potentials for the inter-particle interactions. The Morse potential and its force have both been truncated and shifted, using the Shifted Forces approximation [198], to decay smoothly to zero at $r_c = 2.12\sigma$.

In order to avoid finite size effects in the formation of small clusters, we study a low packing fraction of $\phi = 0.01$. We use a first order Euler discretization for the equation of motion in Eq. 3.1. Since the potentials in Eq. 3.3 are narrow and short-range, we have to use a small timestep of $5 \times 10^{-5}t_0$ in order to sample the potentials accurately. We have used trajectories of length ranging from $\tau/t_0 = 2.5 \times 10^3$ to 10^4.

3.2   Automated inverse design algorithm

Variational principle

In order to uncover design principles for self-assembly, we consider the task of finding the set of forces that fulfill the condition of assembling a target structure as the typical state of the system in the long time limit. Such tasks in stochastic dynamics are generalizations of Brownian bridges and known to have unique solutions [111]. They have played an important role recently in the application of large deviation theory to physical systems driven far from equilibrium [86, 81, 93, 199, 117].

We start by defining an observable A_τ as a time averaged indicator function for a target cluster,

$$A_\tau[\mathbf{r}^N(t)] = \frac{1}{\tau} \int_0^\tau \mathbf{1}[\mathbf{r}^N(t)] \, dt \tag{3.7}$$

where $\mathbf{1}[\mathbf{r}^N(t)] = 1$ for a configuration satisfying a geometric criterion consistent with a target cluster and $\mathbf{1}[\mathbf{r}^N(t)] = 0$ otherwise, for each time t along a trajectory $\mathbf{r}^N(t)$ of total duration τ. The average value of the observable quantifies the yield of the target cluster. For all colloidal clusters considered, $\mathbf{1}$ is computed by constructing a bond-connectivity matrix. A cutoff of $r_b = 1.35\sigma$ has been used to define a bond between two particles. Indicator functions for rigid target clusters are then uniquely determined by permutation-invariant measures of this connectivity matrix [200]. For nonrigid target clusters, along with the bond-connectivity matrix, we additionally consider measures of the geometry of the cluster for defining the indicator function.

Rather than considering trajectories conditioned on a particular value of A_τ directly, which is numerically cumbersome, we work within an ensemble equivalent representation [118]. Using a counting parameter λ, we can statistically bias a system towards a particular value of A_τ within a nonequilibrium steady state. The cumulant generating function $\Psi(s)$ is the partition function associated with the trajectory ensemble under the statistical action

of s,

$$\Psi(s) = \lim_{\tau \to \infty} \frac{1}{\tau} \ln \left\langle e^{-s\tau A_\tau} \right\rangle_0 \tag{3.8}$$

where the angular brackets denote a path average over trajectory probability $P_0[\mathbf{r}^N(t)]$, as

$$\left\langle e^{-s\tau A_\tau} \right\rangle_0 = \int D[\mathbf{r}^N(t)] \exp\left(-s\tau A_\tau[\mathbf{r}^N(t)]\right) P_0[\mathbf{r}^N(t)] \tag{3.9}$$

The subscript 0 refers to the average being computed in a reference ensemble where the particles do not typically show the desired self-assembly behavior. For this reference system we have chosen an equilibrium ensemble of colloids interacting only with the WCA repulsive forces, i.e., $D_{ij} = f = 0$, such that $\mathbf{u}_i = -\boldsymbol{\nabla}_i \sum_{j \neq i} V_{\mathrm{WCA}}(r_{ij})$ which is denoted as $\mathbf{F}_i^{\mathrm{WCA}}(\mathbf{r}^N)$.

When the optimizable parameters are tuned to vary $\mathbf{u}$, the trajectory probability $P_0[\mathbf{r}^N(t)]$ changes to $P_{\mathbf{u}}[\mathbf{r}^N(t)]$. The cumulant generating function can be estimated in the modified ensemble as,

$$\Psi(s) = \lim_{\tau \to \infty} \frac{1}{\tau} \ln \int D[\mathbf{r}^N(t)] e^{-s\tau A_\tau} \frac{P_0[\mathbf{r}^N(t)]}{P_{\mathbf{u}}[\mathbf{r}^N(t)]} P_{\mathbf{u}}[\mathbf{r}^N(t)]$$

$$= \lim_{\tau \to \infty} \frac{1}{\tau} \ln \left\langle e^{-s\tau A_\tau + \Delta S[\mathbf{u}]} \right\rangle_{\mathbf{u}} \tag{3.10}$$

where the functional form of the relative action $\Delta S[\mathbf{u}]$ can be derived from Onsager-Machlup theory [45],

$$\Delta S[\mathbf{u}] = S[\mathbf{u}] - S[\mathbf{F}^{\mathrm{WCA}}]$$

$$= \int_0^\tau \sum_{i=1}^N \frac{(\dot{\mathbf{r}}_i - \mathbf{u}_i)^2 - (\dot{\mathbf{r}}_i - \mathbf{F}_i^{\mathrm{WCA}})^2}{4\gamma k_{\mathrm{B}} T} dt \tag{3.11}$$

with the integral being computed in the Ito sense. This change of measure analogous to a Girsanov transform [63] relates the original likelihood of self-assembly in the reference ensemble to the ensemble under the control force.

Since the exponential is a convex function, we apply Jensen's inequality to Eq. 3.10

$$\Psi(s) \geq \lim_{\tau \to \infty} \frac{1}{\tau} \langle -s\tau A_\tau + \Delta S[\mathbf{u}] \rangle_{\mathbf{u}} \tag{3.12}$$

to obtain a variational expression for the cumulant generating function. In the long time limit for $\tau \to \infty$, we can replace trajectory averages with static averages and simplify the relative action using the equation of motion for $\dot{\mathbf{r}}_i$. Hence we arrive at our final variational expression,

$$\Psi(s) \geq \langle \Omega[\mathbf{u}] \rangle_{\mathbf{u}} = \left\langle -s\mathbf{1}(\mathbf{r}^N) - \sum_{i=1}^N \frac{(\mathbf{u}_i - \mathbf{F}_i^{\mathrm{WCA}})^2}{4\gamma k_{\mathrm{B}} T} \right\rangle_{\mathbf{u}} \tag{3.13}$$

where $\langle \Omega[\mathbf{u}] \rangle_{\mathbf{u}}$ is the target function to optimize.

For a bounded observable like the indicator function, a large negative value of s enforces the desired conditioning. The problem of saturating the variational inequality is known to have a unique solution when $\mathbf{u}(\mathbf{r}^N)$ can take all possible functional forms, the optimal force being a generalization of Doob's h-transform [111]. Optimizing $\langle \Omega[\mathbf{u}] \rangle_\mathbf{u}$ will lead to a set of many-body forces for assembling target clusters in high yield. In practice the use of only one-body and two-body forces in Eqs. 3.3-3.6 need not saturate the inequality. The second term in the variational expression is associated with the Kullback-Leibler divergence between the reference and conditioned trajectory ensembles. This term enforces the smallest excess force out of all possible control forces, and thus acts as a regularizer in the optimization process. While the solution to Eq. 3.13 uniquely selects the force that shows fluctuations closest to rare fluctuations in the original ensemble, it is not a unique inverse design criterion and alternatives can in principle be constructed [201]. However, the optimization scheme that we construct in the next section can be generally extended to other functional forms of regularizers.

Stochastic gradient descent

To numerically optimize Eq. 3.13, we derive explicit gradients of the variational estimator $\langle \Omega[\mathbf{u}] \rangle_\mathbf{u}$, using an algorithm that we have previously employed to estimate large deviation functions in nonequilibrium steady-states [117]. The general form of the gradient with respect to any variational parameter $c \in \{D_{ij}, f\}$ is

$$\frac{\partial \langle \Omega[\mathbf{u}] \rangle_\mathbf{u}}{\partial c} = \left\langle \frac{\partial \Omega[\mathbf{u}]}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial c} \right\rangle_\mathbf{u} - \int_0^\infty \left\langle \delta\Omega(t) \delta \left(\frac{\partial \dot{S}[\mathbf{u}]}{\partial \mathbf{u}} \frac{\partial \mathbf{u}}{\partial c} \right)(0) \right\rangle_\mathbf{u} dt \qquad (3.14)$$

where $\dot{S}[\mathbf{u}]$ is the time derivative of the action $S[\mathbf{u}]$. Equation 3.14 is a generalized fluctuation-dissipation relation for a nonequilibrium response in the design parameter space. For computational purposes, we approximate the gradient expression by integrating the correlation function in the second term up to a fixed large time interval $\Delta t = 5t_0$. Due to the small density, we have to use a low variance estimate for the explicit gradient in Eq. 3.14 for the specific case of optimizing the shear flow rate f as described later.

The corresponding variational algorithm consists of a stochastic gradient descent optimization [202] for a large negative value of s in Eq. 3.13. Starting from an initial point in parameter space $\{D_{ij}, f\}$, we simulate the dynamics of the system using Eq. 3.1, and after relaxation into a steady state, statistically estimate the explicit gradient of the variational estimator, Eq. 3.14. We perform stochastic gradient descent updating all variational parameters c at every step of the optimization, with the update rule at the n-th step being

$$c_{n+1} = c_n + \alpha_n \frac{\partial \langle \Omega[\mathbf{u}] \rangle_\mathbf{u}}{\partial c} \bigg|_{c_n} \qquad (3.15)$$

where the stochastic gradients are evaluated within a steady-state with the current value of the parameters, and α_n is the learning rate for any of the c parameters in the n-th optimiza-
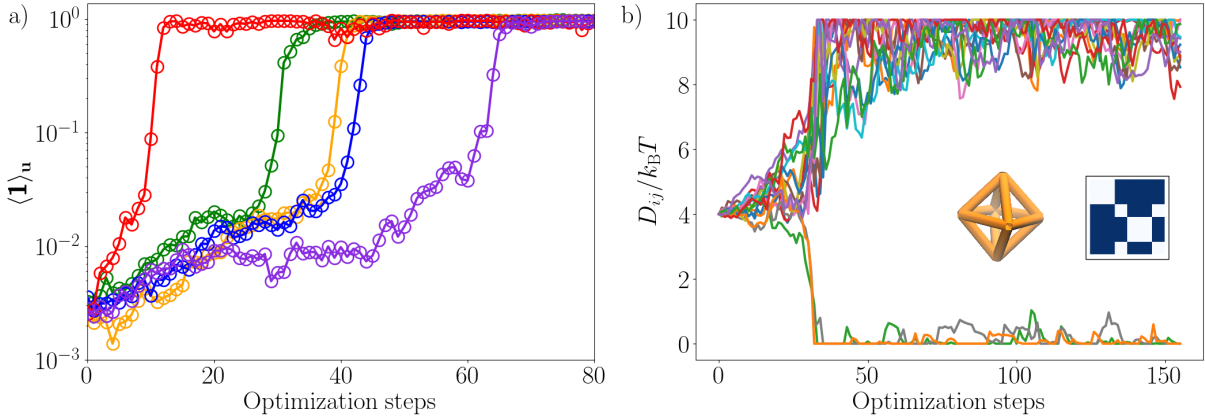
Figure 3.2: Optimization procedure for an O_h cluster. a) The convergence of the yield with increasing number of optimization steps. Different colors represent varying learning rates. b) Convergence of D_{ij} for the green yield curve in (a). (*Inset*) Bond structure and the corresponding MA D_{ij} matrix for the O_h cluster. Blue and white elements in the matrix denote bonds with $D_{ij} = 10k_BT$ and $D_{ij} = 0$, respectively.

tion step. The level of noisy fluctuations in each parameter during the optimization process can be tuned independently through the corresponding learning rates. If the variational surface changes sharply, the learning rate has to be decreased with increasing n to anneal to the optimal solution basin. The learning rates have also been chosen individually for each example such that in each optimization step, the rate of change of D_{ij}/k_BT is in the range $[0.1, 0.5]$ and that of $f\sigma^2/k_BT$ is in the range $[1, 5]$.

Convergence and choice of s

To illustrate the performance of the optimization algorithm, we study the assembly of 6 particles into an octahedral (O_h) target cluster. An octahedron is not the highest yield cluster formed in a system of 6 hard sphere colloids with infinitely short-range attractions [192], and is formed in only 6% yield with strong, nonspecific interactions. Figure 3.2a shows the yield as a function of optimization steps with different learning rates and different trajectory noise histories. For all these examples, the yield is optimized over multiple orders of magnitude with the final converged value being close to 100%. This change of the order parameter over several orders of magnitude arises from the observable being defined as the probability of forming the target cluster, and in cases where the change is more drastic, would necessitate the use of two different learning rates in Eq. 3.15. At a constant learning rate, the learning curves show two distinct regions, such that a gradual rise in yield is followed by
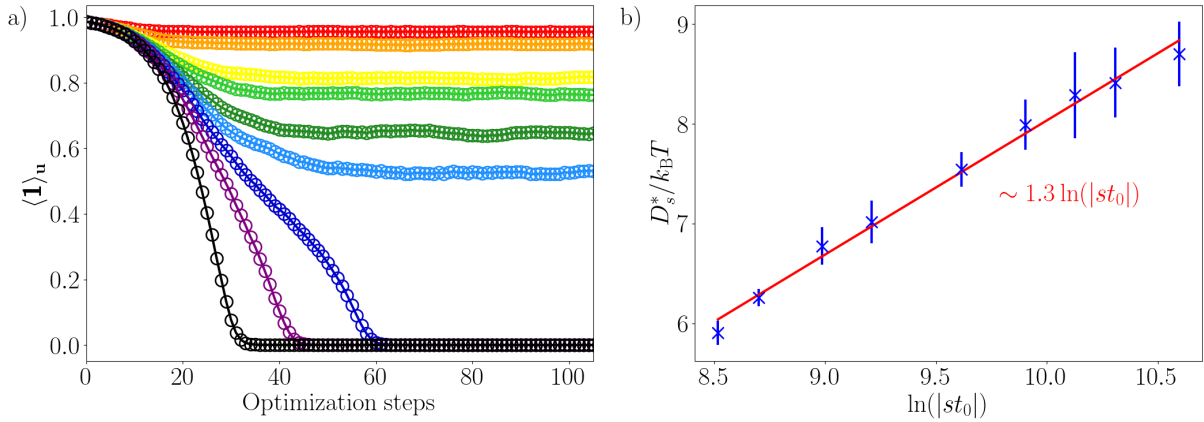
Figure 3.3: Stability of the high-yield solution with s. a) Convergence of the yield starting from the $10k_{\mathrm{B}}T$ MA solution. Different colors represent varying values of st_0 in the range $[-4 \times 10^4, -10^3]$, with the converged value of the yield increasing monotonically with $|s|$. b) Blue crosses with errorbars are the converged bond energy of the MA solution at varying values of s. Red line is a linear fit.

a rapid convergence to the saturation value.

Figure 3.2b shows the convergence of D_{ij} for one of the optimization runs. For 6 particles there are 15 distinct interactions, all of which are optimized. The starting point is a non-specific attraction $D_{ij} = 4k_{\mathrm{B}}T$ for all ij pairs. The optimization curve shows two regions, an initial spreading of the D_{ij} values followed by a rapid permutation symmetry breaking and a clear segregation of the 15 interactions into 12 attractive and 3 repulsive parameters. The 12 attractive interactions are all statistically equal, as are the 3 repulsive interactions. The attractive interactions correspond to the 12 bonds in the connectivity matrix for the octahedron. The symmetry breaking is spontaneous and is aided by the initial noisy fluctuations during optimization. Different noise histories in the trajectory lead to a symmetry breaking for which different sets of D_{ij} parameters become attractive or repulsive. For the finite clusters considered, this symmetry breaking is general. We refer to the specific D_{ij} solutions for the optimal yield of a target cluster as an *alphabet*, and the particular D_{ij} in which there is a pairwise attractive interaction for every contact in the target structure as a *Maximal Alphabet* (MA). This strategy has been previously shown to be effective in the self-assembly of short-range interacting colloids into small clusters [175, 193].

For the octahedral cluster, we have studied the stability of the MA solution for varying values of s. We find a bimodal structure of the variational surface, with the algorithm converging to either a MA solution or a trivial solution $\mathbf{u} = \mathbf{F}^{\mathrm{WCA}}$, depending on the value of s. Figure 3.3a demonstrates the convergence of the octahedral yield with different values of st_0 varying in the range $[-4 \times 10^4, -10^3]$, starting from an MA solution with $10k_{\mathrm{B}}T$ bond

Figure 3.4: Design principles for rigid clusters. a) Bond structure of clusters (i-ix), along with their corresponding point groups, optimal yields and their converged *alphabets* on a color scheme indicated by the colorbar at the top. b) Yield as a function of a nonspecific attraction $D_{ij} = D$ for fixed shear $f = 0$. c) Yield as a function of the shear rate f for fixed optimal *alphabets*. The colors of the points in b) and c) correspond to the colors of the clusters in a).

energies and a yield of 100%. For moderate but decreasing values of $|s|$, the algorithm remains stable in the MA solution, but with monotonically decreasing yields and bond energies. For less negative values of s than a critical value of $s_c = -5 \times 10^3/t_0$, the MA solution becomes unstable and the algorithm finds the $\mathbf{u} = \mathbf{F}^{\text{WCA}}$ solution. Rather than optimizing the yield in Eq. 3.13, at small values of s the second term is optimized. For some moderate s values, the MA basin is only a local optimum and the crossover behavior shows s-dependent hysteresis. We find this bistablity of the variational surface to be generic.

Asymptotic dependence of solution on s

We analytically solve for the asymptotic dependence of the bond energy D as a function of the bias s, for the formation of one bond, independent from the dynamics of the other bonds. We can represent this simplified system as an equilibrium, two state Markov model with the rates of transition between the bonded and unbonded states k_b and k_u, respectively. These rates are determined by their mean $\nu = (k_u + k_b)/2$ and their ratio $k_u/k_b = \exp(\Delta F/k_B T)$,

where ΔF is the free energy difference of the two states.

The cumulant generating function associated with fluctuations in the bonded state satifies a eigenvalue equation of the form,

$$W_s r_s = \Psi(s) r_s \tag{3.16}$$

where r_s is a right eigenvecture, and operator W_s is given by

$$W_s = \begin{pmatrix} -k_u - s & k_u \\ k_b & -k_b \end{pmatrix} \tag{3.17}$$

which is equal to the adjoint of the transition rate matrix when $s = 0$. For the optimal rates that generate the statistics equivalent to this rate matrix, we need the Doob's transform of the matrix [63]. For this purpose we diagonalize the matrix to find the eigenvector corresponding to the dominant eigenvalue as $(r_1, 1)$ where

$$
\begin{aligned}
r_1 = \frac{e^{-\frac{\Delta F}{k_{\mathrm{B}} T}}}{4\nu} &\left[(-s + 2\nu) + e^{\frac{\Delta F}{k_{\mathrm{B}} T}}(-s - 2\nu) \right. \\
&\left. - \sqrt{e^{2\frac{\Delta F}{k_{\mathrm{B}} T}}(-s - 2\nu)^2 + (-s + 2\nu)^2 + 2e^{\frac{\Delta F}{k_{\mathrm{B}} T}}(s^2 + 4\nu^2)} \right]
\end{aligned} \tag{3.18}
$$

and k_u and k_b have been rewritten with ΔF and ν.

The modified rates that generate the optimal dynamics are given by $\tilde{k}_u = k_u / r_1 t_0$ and $\tilde{k}_b = k_b r_1 t_0$. The modified free energy difference corresponding to these rates is $\Delta \tilde{F} = k_{\mathrm{B}} T \ln(\tilde{k}_u / \tilde{k}_b)$. Using Eq. (3.18), in the $s \to -\infty$ limit the optimal free energy goes as

$$\Delta \tilde{F} \sim \Delta F - 2 k_{\mathrm{B}} T \ln(|s t_0|) \tag{3.19}$$

In this limit, the free energy is dominated by the negative of the bond-formation energy D, and hence the latter is asymptotically $D \sim 2 k_{\mathrm{B}} T \ln(|s t_0|)$.

Figure 3.3b shows the dependence of the average optimized bond energies of the converged MA solutions from the previous figure, as a function of varying s. This illustrates the depth of the MA basin in the parameter space. In the limit that each bond is formed independently, D_s^* is expected to asymptotically vary as $\sim 2 k_{\mathrm{B}} T \ln(|s t_0|)$. Over the range of s considered, we find a logarithmic dependence but with a different coefficient, $\sim 1.3 \ln(|s t_0|)$. This suggests that the free energy is approximately pairwise additive.

In the limit of large negative s, which in practice is chosen such that the estimate of the first term in Eq. 3.13 is at least an order of magnitude larger than the negative second term, the variational algorithm can be used to automate the discovery of optimal forces for the self-assembly of clusters of arbitrary shapes and sizes, in a nonequilibrium steady-state. The optimal forces stabilize the target clusters in an arbitrary ensemble without accounting for the dynamics of transient relaxation towards its steady-state. For fixed number of tunable parameters, the computational cost scales linearly with system size since the only bottleneck
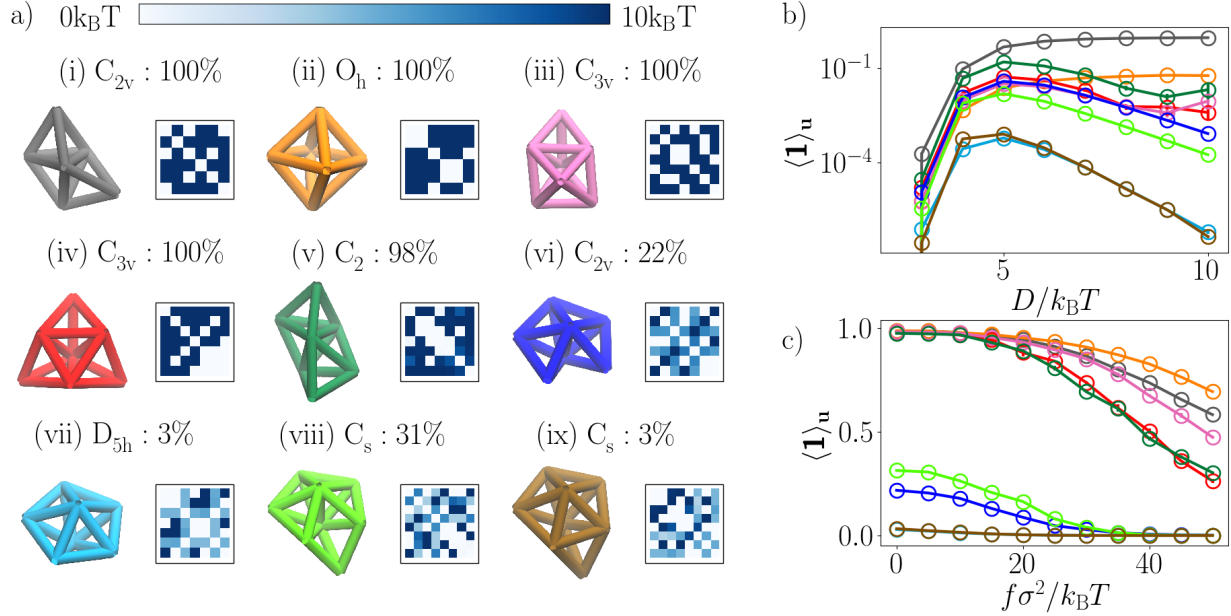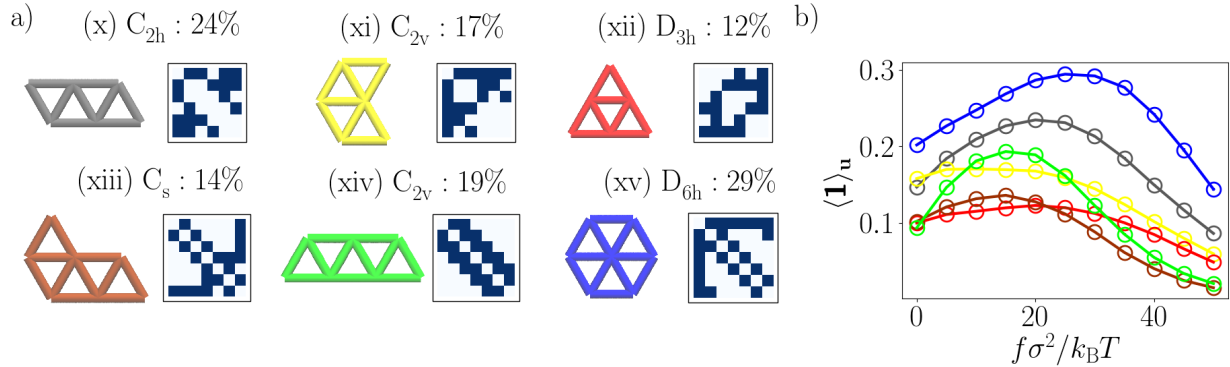
Figure 3.5: Design principles for nonrigid clusters. a) Bond structure of clusters (x-xv), along with their corresponding point groups, optimal yields and their converged *alphabets* on the same color scheme that was used in Fig. 3.4. b) Yield as a function of the shear rate for D_{ij} fixed at the optimal *alphabet*. The colors of the points correspond to the clusters in a).

is propagating a steady-state trajectory long enough to compute statistically converged gradients. The algorithm also scales linearly with the number of variational parameters, but with a small proportionality constant as all the gradients are estimated from the same trajectory. The use of the statistically estimated gradients significantly lowers the computational cost in contrast to numerically estimating the gradients from finite difference techniques by propagating multiple trajectories at different points in the parameter space. We next use our variational algorithm to study and rationalize the optimal design principles for a collection of rigid and nonrigid clusters.

3.3 Design principles for yields of rigid and nonrigid clusters

We have investigated the formation of small low-energy rigid and nonrigid clusters with 6,7 or 8 particles. We discover distinct design principles of these clusters and rationalize our findings by analyzing the response function of yield to the shear flow rate. We also demonstrate that the variational algorithm can obtain high yield optimal solutions even with constraints imposed on the total number of experimentally realizable design parameters. The design principles we obtain are expected to be general for the nonequilibrium self-assembly of short-range interacting colloids in a sheared steady-state.

Rigid clusters

We study the formation of a family of rigid clusters that are all known to be the lowest energy structures for systems of hard sphere colloids with identical infinitely short-range attractions. These clusters have previously been systematically enumerated and tabulated [203, 204], and their free-energy landscapes have been theoretically and experimentally studied [192]. These finite clusters are the colloidal analogs to small molecules and have been shown to be involved in the controlled seeding and growth of polycrystalline phases and kinetically arrested gels [205, 206].

For each of these clusters, we optimize $\{D_{ij}, f\}$ to extremize the yield within this force ansatz. In the limit that the attractive interactions between the particles were infinitely short-range, there would be no internal low-energy distortion modes and the bond-connectivity matrix would correspond to a unique geometry. For our optimization, the indicator function refers to the corresponding bond-connectivity matrix conditions being satisfied. Figure. 3.4a summarizes the design principles discovered for these clusters. The point groups for the symmetries of each of these clusters have been indicated along with the highest yields obtained. For chiral C_n clusters, the yields are the racemic yield.

For each of these clusters, the corresponding optimal *alphabet* discovered by the variational algorithm has also been indicated. We find that for clusters (i-iv), the optimal solution for D_{ij} is the MA. For the chiral C_2 cluster (v), the optimal *alphabet* is closely related to the MA but has a higher symmetry and is equivalent to a smaller 3×3 *alphabet*, while having the same yield. For clusters (vi-ix), all of which contain a radial 5-fold motif, the optimal yields are much less than 100%, and the optimal *alphabets* are not MA. The reason is the competition with structures with higher number of bonds. These lower energy structures would be geometrically unfeasible for infinitely short-range attractions, however in our model the short-range bonds have nonzero vibrations, which is sufficient to lead to the formation of the extra bonds. Unlike MA, the optimal *alphabet* discovered by the variational algorithm penalizes the formation of these lower energy competing structures.

Figures 3.4b and 3.4c show two slices through the optimization landscape in the parameter space of $\{D_{ij}, f\}$. Figure 3.4b is a diagonal slice through D_{ij}, such that all $D_{ij} = D$, while fixing $f = 0$. We find that for the two 6 particle clusters (i) and (ii), there is a monotonic increase in yield with increasing D. This suggests that even when the attractive forces are not infinitely short-range, both of these clusters are energetically the most stable configurations, and the only competing structures are higher in energy. When $D = 10k_BT$, the C_{2v} cluster (i) is formed with a yield of 93% compared to the 6% yield of the O_h cluster (ii), which is consistent with the stabilization due to the rotational entropy in the former [192]. All the other clusters (iii-ix) in Fig. 3.4b show a turnover in yield with increasing D. This is due to competing low energy structures that are formed at large enough nonspecific D. We expect this design principle of a turnover in yield with increasing nonspecific attractions to be general for larger clusters, since most larger clusters built from short-range interacting particles will contain the radial 5-fold motif [204]. The value of the attraction D at the yield turnover is determined by a competition between the energetic stabilization of the lower

energy structure and the destabilization of structures with missing bonds.

Figure 3.4c shows a slice through the optimization landscape with varying the shear rate f, while fixing D_{ij} at their optimal value found by the variational algorithm. We see that for this class of rigid clusters, the yield monotonically decreases with increasing shear rate. We expect this feature to be general for rigid clusters, since rigid clusters need no additional geometric stabilization that can be provided by shear, which only energetically destabilizes the bonds in the cluster. This perspective is confirmed in Sec. 3.3 using a linear response theory.

Nonrigid clusters

For our ansatz of short-range interacting colloids, clusters with N particles but fewer than $3N - 6$ bonds in total, and fewer than 3 bonds for every particle, are not minimally rigid in that they have zero energy deformation modes [203, 204]. These clusters are not formed in high yield as stable ground state structures in equilibrium. We have used the variational algorithm to uncover optimal nonequilibrium design principles for a family of such nonrigid clusters. The clusters we have considered belong to a family of planar two-dimensional structures known as polyiamonds. They have been shown to self-assemble from colloids in the presence of a spatial heterogeneity, like in hydrodynamically driven assembly of sedimenting colloids in the presence of a substrate [207, 208, 209]. Within our control force ansatz, we investigate whether the shear flow planes are sufficient to stabilize these clusters.

For the optimization process, the indicator function for the cluster yield has been defined using both the bond connectivity matrix and the flatness of the clusters, as discussed in the next subsection. Figure 3.5a shows the optimal design principles obtained for clusters (x-xiv). The variational algorithm converges on the MA solution for the D_{ij} parameters for all of these clusters. The yields, however, are not 100% due to contribution from competing buckled configurations where the polyiamonds fold over the triangular faces to form tetrahedral motifs. Moreover, with the MA fixed, the optimal yield is at a non-zero shear flow. Figure 3.5b shows the yield as a function of the shear rate for fixed optimal alphabets. The location of the turnover in yield depends on the competition between geometric stabilization of the planar structure from the shear flow lines and energetic destabilization of the bonds. This design principle of planar two-dimensional clusters being stabilized by a shear flow appears to be general, and stands in contrast to the rigid clusters which are strictly destabilized by shear.

Effect of shear on geometry of nonrigid clusters

Shear flow enhances the yield of nonrigid clusters (x-xv) by stabilizing a planar geometry and suppressing buckling modes. Here we have computed the probability distribution $P(\overline{\cos\theta})$ of the average flatness, defined as $\overline{\cos\theta} = [\cos(\theta_{ABC}) + \cos(\theta_{DEF})]/2$, conditioned on the correct bond connectivity matrix for the cluster being satisfied. The angles, θ_{ABC} and θ_{DEF} are defined in Fig. 3.6. The flatness is -1 for a perfectly planar geometry, but increases due
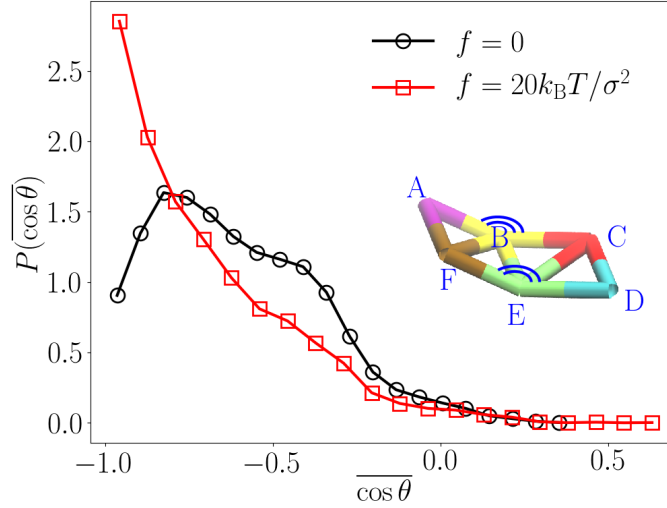
Figure 3.6: Flatness distribution for the geometry of the C_{2h} cluster (x), keeping the D_{ij} fixed at the optimal solution and changing the shear rate f. (*Inset*) The flatness is measured through the angles θ_{ABC} and θ_{DEF}.

to buckling and bending of the nonrigid cluster. For defining the indicator function for the nonrigid cluster, we used a flatness cutoff of $\overline{\cos\theta} \leq -0.8$. We have looked at a population where the bond-connectivity matrix condition is satisfied but the flatness is unconditioned. This is illustrated in Fig. 3.6 for the C_{2h} cluster (x). We have kept the D_{ij} forces fixed at the optimized *alphabet*, and plotted the distribution of flatness at two values of shear, at equilibrium with $f = 0$ and also at $f = 20k_BT/\sigma^2$ which is close to the optimal value for highest yield. We find that the planar geometry is a transient state at equilibrium, with the most probable states corresponding to the buckling of one or both of the angles θ_{ABC} and θ_{DEF}. The shear flow destabilizes the buckled conformations and stabilizes the planar state instead, so that at $f = 20k_BT/\sigma^2$ the most probable conformation is the correct planar geometry.

Estimating explicit gradients in the low-density limit

Using Eq. 3.14, the second term in the explicit gradient with respect to the shear flow rate f is

$$
\left\langle \delta\Omega(t)\delta\left(\frac{\partial\dot{S}[\mathbf{u}]}{\partial\mathbf{u}}\frac{\partial\mathbf{u}}{\partial f}\right)(0)\right\rangle_{\mathbf{u}}
$$
$$
= \frac{1}{2\gamma k_BT}\left\langle \delta\Omega(t)\delta\left(\sum_{i=1}^{N}\eta_i^x z_i\right)(0)\right\rangle_{\mathbf{u}} \tag{3.20}
$$
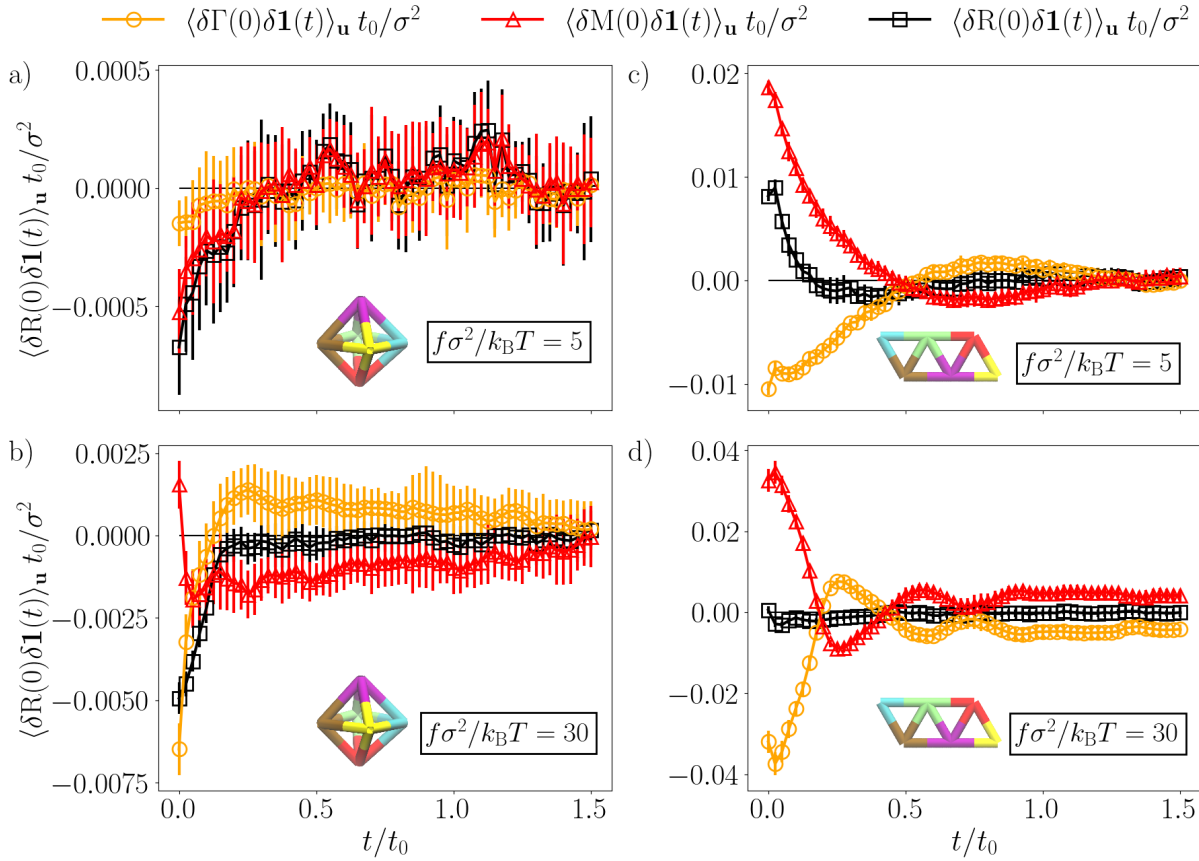
Figure 3.7: Response of yield to shear for a rigid and a nonrigid cluster, with D_{ij} fixed at the corresponding optimal *alphabets*, and shear fixed at $f = 5k_{\mathrm{B}}T/\sigma^2$. a) Total correlation function (black squares) for an O_{h} cluster (ii), and its torque (orange circles) and virial stress (red triangles) parts. b) The same correlation functions for the $C_{2\mathrm{h}}$ cluster (x).

where we have simplified the stochastic action using the equation of motion. Since z_i appears in the expression independently for each particle, and the density of the particles is vanishingly small, the z-diffusion timescale of the cluster diverges, and the correlation function takes a long time to converge. Thus any gradient estimate we obtain by integrating the correlation function to a finite time Δt will contain a systematic error. In order to obtain an unbiased gradient, we recognize that in the large λ limit we are working in, the major part of $\Omega(t)$ is from $\mathbf{1}(t)$, which by our definition depends only on the internal coordinates of the particle, and due to the spatial translation symmetry in our system, is decoupled from the center-of-mass diffusion. This decoupling is directly expressed by a regrouping of terms in

the sum over particles,

$$s \left\langle \delta \mathbf{1}(t) \delta \left(\sum_{i=1}^{N} \eta_i^x z_i \right) (0) \right\rangle_{\mathbf{u}}$$

$$= \frac{s}{N} \left\langle \delta \mathbf{1}(t) \delta \left(\left[\sum_{i=1}^{N} \eta_i^x \right] \left[\sum_{i=1}^{N} z_i \right] \right) (0) \right\rangle_{\mathbf{u}}$$

$$+ \frac{s}{N} \left\langle \delta \mathbf{1}(t) \delta \left(\sum_{\substack{i,j=1 \\ i>j}}^{N} (\eta_i^x - \eta_j^x)(z_i - z_j) \right) (0) \right\rangle_{\mathbf{u}} \tag{3.21}$$

where in the first term the z coordinate of the center of mass has been explicitly factored out. We identify that the indicator function does not correlate with the center-of-mass motion and so the first term is 0. We use only the second term to approximately evaluate the gradients of Eq. 3.13 with respect to the shear flow rate f.

Response of structure to shear

The origin of the response of yield to shear flow is related to the relaxation dynamics of order parameter fluctuations in the unperturbed system. The response coefficients for rigid and nonrigid clusters can be understood using a generalized linear response theory [210, 36]. Keeping the D_{ij} parameters fixed, the linear response of the yield to a change in shear flow rate can be decomposed into two terms,

$$\frac{\partial \langle \mathbf{1} \rangle_{\mathbf{u}}}{\partial f} = \frac{1}{k_{\mathrm{B}}T} \int_0^{\infty} dt \; \langle \delta \Gamma(0) \delta \mathbf{1}(t) \rangle_{\mathbf{u}}$$

$$+ \frac{1}{k_{\mathrm{B}}T} \int_0^{\infty} dt \; \langle \delta \mathrm{M}(0) \delta \mathbf{1}(t) \rangle_{\mathbf{u}} \tag{3.22}$$

where we have used the low variance estimator for the correlation function in 3.3. Here, Γ is related to the dynamical torque acting on the cluster,

$$\Gamma = \frac{1}{2\gamma N} \sum_{i>j} [(\gamma \dot{x}_i - f z_i) - (\gamma \dot{x}_j - f z_j)](z_i - z_j) \tag{3.23}$$

and M refers to the virial stress on the cluster due to internal forces,

$$\mathrm{M} = \frac{1}{2\gamma N} \sum_{i>j} [F_i^x - F_j^x](z_i - z_j) \tag{3.24}$$

where $\mathbf{F}_i = -\boldsymbol{\nabla}_i \sum_{j \neq i} [V_{\mathrm{WCA}}(r_{ij}) + V_{\mathrm{Morse}}(r_{ij})]$ is the conservative force acting on the i-th particle. Γ and M are time-reversal asymmetric and symmetric parts respectively [211] of

the full stochastic action gradient, $R = \Gamma + M$. Decomposing in this form is necessary to preserve the deconvolution of the center-of-mass motion from the internal coordinates of the cluster [212].

This linear response function is 0 at equilibrium due to the spatial parity symmetry of the system. Hence we have characterized the different components of this correlation function at a small value of shear $f = 5k_BT/\sigma^2$, for the rigid cluster (ii) and the nonrigid cluster (x), fixing D_{ij} to their corresponding MA interactions. The results are shown in Fig. 3.7. We find that the component coming from the virial stress has opposite signs at small times for the rigid and the nonrigid cluster, which accounts for the opposite signs of the gradient of the yield. The tumbling motion of the clusters in a shear flow couples positively with the internal order parameter in the case of a nonrigid cluster and leads to an increase in yield with increasing shear flow rate at small values of shear. The shear flow planes function as a spatial heterogeneity that is generally a precondition for stabilizing these planar clusters during self-assembly. At large shear, the yields of both rigid and nonrigid clusters are decreased with increasing shear due to the larger anti-correlation between the torque and the indicator function.

These design principles in and out-of-equilibrium are general in their scope of applicability for small clusters formed by DNA-coated colloids. Nevertheless, a key limitation of this approach is the linear system size scaling of the number of different kinds of DNA-labeled colloids required in order to assemble a cluster, evident in the corresponding quadratic scaling of the number of variational parameters. We have addressed this limitation in the next section.

Smaller alphabets

Engineering a system with an extensive number of specific interactions is difficult, even with DNA-coated colloids. It is advantageous in this regard to uncover *alphabet*s that code for the minimal sufficient interactions to stabilize a target structure, in such a way that does not increase with increasing system size. For example, polymers and crystals are macroscopic structures that can be assembled with a finite number of specific interactions, as both only require a repeating microscopic number of components to be stabilized, either a sequence of monomers or a unit cell. For clusters that do not have a clear repeating unit, discovering optimal design principles with smaller *alphabets* is nontrivial.

We have studied this problem by considering the 6 different low-symmetry 8-particle clusters (xvi-xxi) shown in Fig. 3.8. For each of these clusters, there is no direct way to partition the interactions into 2 or 4 classes based on their bonding environment or symmetry. We have used the variational algorithm to optimize the yield of each of these clusters, with a 2 particle, 4 particle and 8 particle *alphabet*, in which D_{ij} has 3, 10 and 28 independent variational parameters respectively. Clusters (xvi-xix) have near 100% yields for the full sized *alphabet*. Clusters (xx) and (xxi) compete with higher bonded clusters containing the radial 5-fold motif, and so have an optimal yield of lower than 100% even with an 8 particle *alphabet*. Nevertheless, for all the clusters, a 4 particle *alphabet* can give quite large
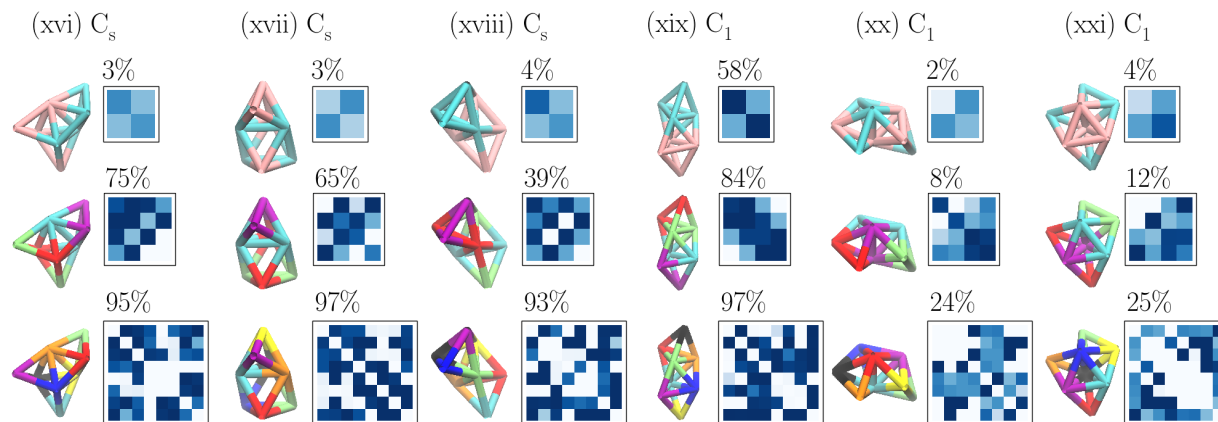
Figure 3.8: Design principles for low symmetry clusters (xvi-xxi) with smaller alphabets. Each column corresponds to the point group of each cluster, its optimal yields and forces with 2×2, 4×4, and 8×8 sized *alphabets* in the three rows. The color scheme for the bonded structure refers to the optimal partitioning of 8 particles within 2, 4 or 8 labels, and the visualization of the D_{ij} matrices follow the same color scheme as in Fig. 3.4.

yields in comparison to the maximum possible yield. The variational algorithm identifies the optimal way to partition the groups of interactions of these clusters despite the lack of clear symmetry.

For the C_1 cluster (xix), even a 2 particle *alphabet* has a high yield, despite not having any exact two-fold symmetry. In this case, the algorithm has recognized a near-symmetry in the cluster and has partitioned it into 2 groups. The symmetry of these letters is close to the symmetrical alphabet that was discovered by the algorithm in a related C_2 cluster (v) in Fig. 3.4. We expect this potential to discover optimal design principles for large clusters with a small number of groups, to be promising towards the self-assembly of experimentally realizable systems with practical constraints on the limits of bottom-up design.

3.4 Design principles for yields of microphase separation

In the previous sections we discussed optimization of the yield of colloidal nanoclusters under a shear flow. The design principles we found however are only valid in systems consisting of particles of only one set of the corresponding alphabet. In cases where there are many copies of the alphabet, it is an entirely new paradigm to search for a set of design principles that produce many copies of identically sized nanoclusters. Formation of such a phase in a self-assembled system, identified by the spontaneous emergence of a microscopic cluster size scale even in the presence of a thermodynamic number of particles, is referred to as
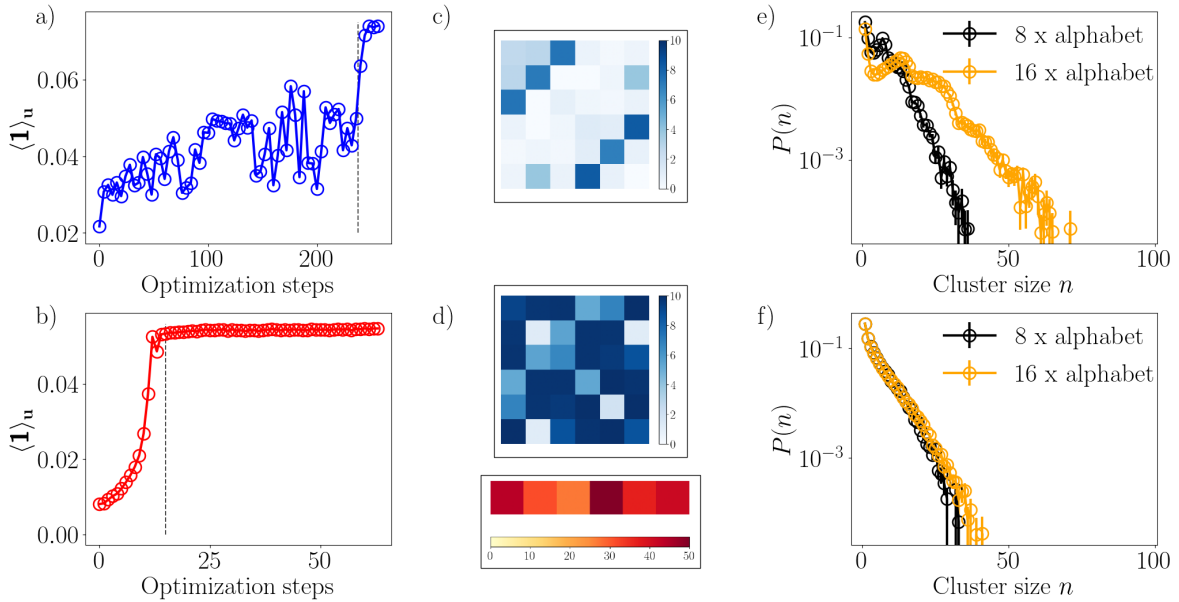
Figure 3.9: Optimized design principles for microphase separation into 6-particle clusters. (a) and (b) are optimization curves for shear driving and active driving respectively. Dashed black lines denote the optimization step where the learning rates are lowered in order to anneal into the final solution well. (c) and (d) are optimized D_{ij} and v_i alphabets for (a) and (b) respectively. The shear in (a) optimizes to $50k_BT/\sigma^2$, the maximum allowed value. (e) and (f) denote cluster size distributions for the optimized system with $N = 48$ and a test system with the converged forces but with $N = 96$.

microphase separation or self-limited assembly. Such a phase will compete with a typical gas phase or bulk condensed phase, where the cluster size distribution either has a most probable value at 1 or $\mathcal{O}(N)$, with N being the thermodynamic system size. A system with self-limited assembly will show a peak at a microscopic size in the cluster size distribution, with microscopic being defined as the peak position being $\mathcal{O}(1)$, invariant to changing the system size in the thermodynamic limit. Apart from the cluster size distribution, peaks in structure factor corresponding to a recurring microscopic lengthscale can also be used as a suitable order parameter to detect a microphase [213].

Biological systems widely use self-limited assembly to regulate volumes of cells, vesicles, and viral capsids [214, 215], efficiently package genetic material in nuclei [216], and tune mechanoelastic properties of cytoskeletal bundles [217]. The mechanism for stabilizing these structures are current areas of research. Stabilizing a microphase against bulk condensation requires either conservative forces or dissipative driving to compete against surface tension. In equilibrium systems, multiple mechanisms of self-limited assembly via short-range at-

tractive and long-range repulsive interactions, through geometric frustration, and through
kinetic control, have been well studied [213, 218, 219]. Inverse design algorithms for design-
ing microphase separation by optimizing a Boltzmann distribution have also been previously
studied [220]. But the role of nonequilibrium mechanical or chemical control of microphase
separation, such as is ubiquitous in active processes in cells, is not well understood.

In this section we apply the variational algorithm to design DNA-labeled colloidal par-
ticles for microphase separation under nonequilibrium driving. Our observable to optimize
is a time averaged indicator function, as defined in Eq. 3.7, for the formation of 6-particle
clusters of any geometry and connectivity. As basis set we choose a 6×6 alphabet, where
now the nonzero diagonal elements denote pairwise interaction strength between multiple
copies of identical particles. We simulated with overdamped Langevin dynamics a periodic
box of 8 copies of the alphabet, with a total of $N = 48$ particles at a packing fraction of 0.02.
The particles feel pairwise WCA repulsion given by Eq. 3.5, and a Morse potential attrac-
tion with tunable amplitudes given by Eq. 3.6, representing a coarse-grained DNA-mediated
effective force. We optimize the yield of 6-particle clusters under two different dissipative
driving conditions: a shear flow with a tunable rate given by Eq. 3.4, and alternately by
making the colloidal particles active with a self-propulsion force given by

$$\mathbf{f}_i^A = v_i \hat{\mathbf{p}}_i \tag{3.25}$$

where v_i is an optimizable magnitude of self-propulsion, and $\hat{\mathbf{p}}_i$ is a three-dimensional po-
larization vector that diffuses as [221]

$$\dot{\hat{p}}_i = \sqrt{2D_r}(\hat{\mathbf{p}}_i \times \mathbf{\Gamma}_i) \tag{3.26}$$

Here $D_r = 3k_\mathrm{B}T/\gamma\sigma^2$ is the rotational diffusivity and $\mathbf{\Gamma}_i$ are unit variance Gaussian white
noise that satisfy

$$\langle \mathbf{\Gamma}_i(t) \rangle = 0 \qquad \langle \mathbf{\Gamma}_i(t)\mathbf{\Gamma}_j(t^{'}) \rangle = 2\mathbf{I}_3\delta_{ij}\delta(t - t^{'}) \tag{3.27}$$

We optimize self-propulsion v_i also within a 6-particle alphabet. We choose the biasing
parameter s in Eq. 3.13 to be -10^{10} as the second term in the cost-function scales with the
system size and is larger than in previous sections.

Figure 3.9 shows the results of the optimization for microphase separation with shear
driving and active driving. (a) and (b) are learning curves showing the increasing yield of
6-particle clusters with optimization steps. Similar to previous sections, we have had to lower
the learning rate in order to anneal smoothly into the solution well. The optimized values of
the yield are $5 - 7\%$, compared with an initial value of $1 - 2\%$. Subfigures (c) and (d) show
the optimized alphabets in the respective systems. In the shear-driven system, the shear flow
rate optimizes to a value of $50k_\mathrm{B}T/\sigma^2$ which is the maximum we allow during optimization.
The optimized D_{ij} alphabet is relatively sparse, coding for a few strong attractions. This
implies same particles from multiple alphabets bonding together to increase probability of a
6-particle cluster. This is confirmed by a size-scaling analysis of the cluster-size distribution

profile. For $N = 48$, the distribution is peaked at 1 and shows a second peak at a cluster size of 6, similar to the expected distribution in a microphase separated system, however, on doubling the system size, the second peak moves to approximately double the desired cluster size. This confirms that the alphabet codes for multiple copies of the same kind of particle assembling into 6-particle cluster, and that this is not true microphase separation. Attempts at eliminating these finite size effects by optimizing larger systems directly did not result in a second peak in the cluster-size distribution profile.

The optimization run for active driving converges smoother than the previous case, however, the algorithm finds a large degenerate class of solutions where either all particles experience strong attraction and strong activity, or they are both weak, with many solutions in between. The high attraction and activity solution has been shown with colormaps in Fig. 3.9(d). This solution does not result in a microphase, as has been shown in the cluster-size distribution profile in subfigure (f). The distribution is peaked at 1 and it is invariant to system size, thus being typical behavior for a gaseous phase. For optimal forces with weaker attractions, the optimized activities required to melt the clusters into a gas are also weak. The variational algorithm finds this same class of gas-phase solutions regardless of choice of starting points in parameter space.

Hence we see that shear driving or active driving are unable to generate microphase separation in coarse-grained models of DNA-labeled colloids. This results from the inability of these mechanically driven dissipative modes to couple to specific probability fluxes that increase the probability of a microphase separated state. In the next section we extend our algorithm to directly optimize reactive fluxes and currents for single nanoclusters to isomerize between different shapes.

3.5 Design principles for reactive flux in steady-states

Self-assembly is a dynamic process that involves stabilizing a target structure as well as increasing the kinetic flux into the target [172]. Fluxionality of a target structure is important in order to correct structural errors during its assembly. Further, any functional structure like an enzyme or a molecular motor must be not only stable but also able to disassemle and reassemble in a dynamic nonequilibrium steady-state. Functional molecules are able to couple their internal reactive modes to external mechanical or chemical dissipative forces, thus breaking detailed balance in specific reaction coordinates that are coupled to their output work. Thus, optimizing self-assembly for optimizing the probability flux in and out of a target structure is the path to designing functional materials. Previous work on optimizing the reactivity of self-assembled nanoclusters have used equilibrium approximations to look at single transient realizations of reactive barrier crossing [222, 162], however, the design of long-lived probability fluxes and currents in and out of functional structures in nonequilibrium steady-states remain unexplored.

In this section, we use our variational algorithm to optimize reactive fluxes and currents flowing between 6-particle clusters of C_{2v} and O_h geometries in a sheared steady-state of
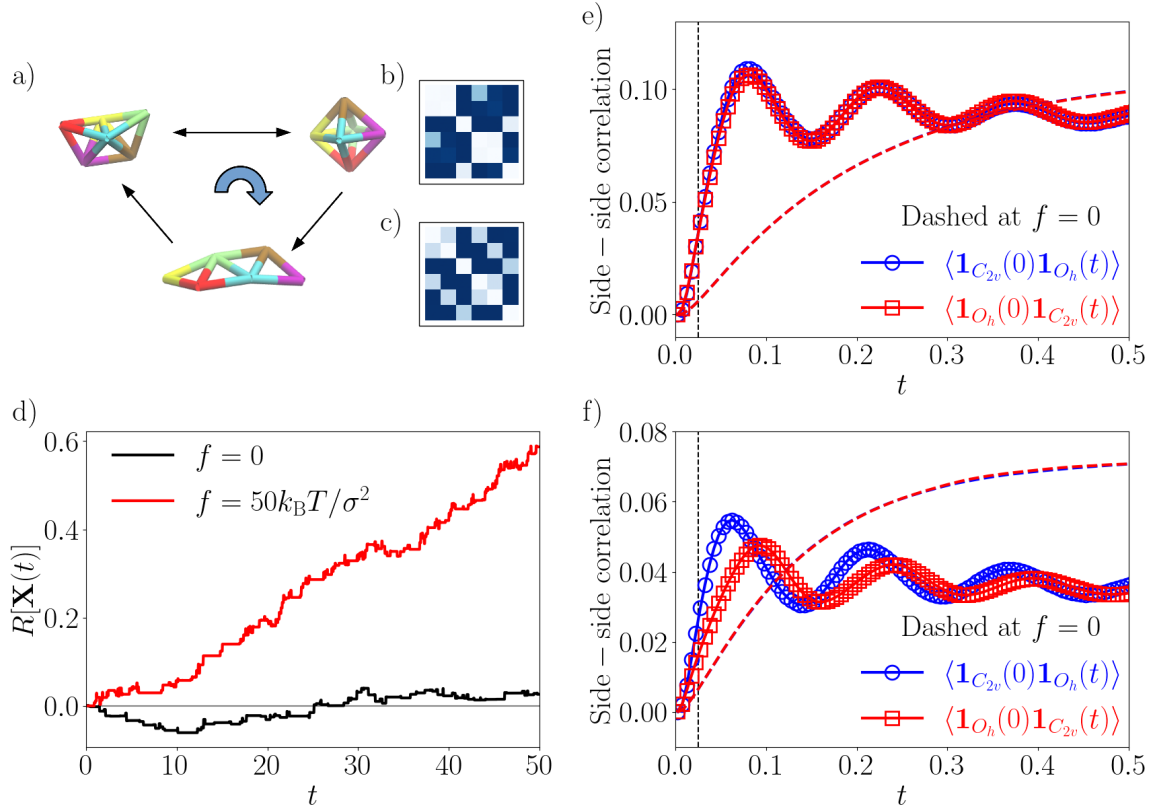
Figure 3.10: Optimal design principles for high reactive fluxes and currents. (a) shows the three major states the sheared colloids go through to manifest high fluxes and currents. (b) and (c) are optimized D_{ij} alphabets for high reactive fluxes between C_{2v} and O_h, and for high reactive current from C_{2v} to O_h respectively. d) Integrated reactive displacement corresponding to the high-current solution, at zero and nonzero shear flow rate. (e) and (f) are the steady-state averaged time-dependent reactive fluxes in both directions for the high flux and high current solutions respectively, at zero (dashed blue and red lines) and nonzero (blue and red symbols) shear flow rates. The vertical dashed black lines denote the delay δt used in the optimization.

$N = 6$ particles. Our flux observable is defined as

$$Q_{\delta t}^{C_{2v} \to O_h} = \frac{1}{\tau} \int_0^\tau \mathbf{1}_{C_{2v}}(t)\mathbf{1}_{O_h}(t + \delta t)dt \tag{3.28}$$

for a fixed delay time δt. Similarly, we define the reverse flux $Q_{\delta t}^{O_h \to C_{2v}}$, and currents $J_{\delta t}^{C_{2v} \to O_h}$

and $J_{\delta t}^{O_h \to C_{2v}}$ as the differences of the fluxes,

$$J_{\delta t}^{C_{2v} \to O_h} = Q_{\delta t}^{C_{2v} \to O_h} - Q_{\delta t}^{O_h \to C_{2v}} \tag{3.29}$$

The delay-time is chosen to be comparable to the typical transition path time needed for individual reactive events to happen, as observed from pre-optimization simulations. This definition of probability current is associated with a nonzero entropy production and is nonzero on average only in a nonequilibrium steady-state. Directly optimizing such probability currents will design the system to couple the specified reaction coordinate maximally to the ambient mechanical dissipatively driven mode.

We choose the same parameters for overdamped Langevin dynamics simulations and optimization as we had chosen in Section 3.2. We have chosen the time delay to be $\delta t = 0.025 t_0$, comparable to observed transition path timescales from a nonspecific alphabet simulation of $D_{ij} = 5 k_B T$ for all pairwise attractions, at equilibrium. We find that both the unidirectional fluxes and the currents are maximized when the shear flow rate is the largest we allow, $f = 50 k_B T / \sigma^2$. Designing the system for high fluxes and currents lead to the prevalence of three major states as shown in Fig. 3.10(a). The high flux optimization runs on either direction yield the same optimal alphabet, Fig, 3.10(b), that is an intermediate between the respective maximal alphabet solutions for C_{2v} and O_h. However, the optimal alphabet for high currents from C_{2v} to O_h resemble neither of those two, but codes for the octahedral cluster to sometimes unfold into the nonrigid C_{2h} cluster shown in subfigure (a), and, keeping the correct order of bonds, refold into a C_{2v} cluster. These irreversible hops are what causes the symmetry of the bidirectional reactive fluxes to break and contributes to positive current round the loop. We have shown in Fig. 3.10(d) the total displacement due to this current,

$$R[\mathbf{X}(t)] = \int_0^t J_{\delta t}^{C_{2v} \to O_h}(t) dt \tag{3.30}$$

that increases to nonzero values with time at the driven nonequilibrium steady-state, as opposed to an equilibrium system having the same conservative forces as the optimized alphabet. These nonzero displacements are used in enzymes and molecular motors to denote the amount of irreversible work done and the emergence of these currents denotes that the molecule has been successfully designed to harvest mechanical nonequilibrium driving to produce usable work [223]. In contrast, optimization runs for the reverse current $J_{\delta t}^{O_h \to C_{2v}}$ were unsuccessful to find any set of forces that led to any positive value. The optimal value of the reverse current was found to be zero arising from either maintenance of the bidirectional reactive flux symmetry or from the nonequilibrium driving being too small to generate any fluxes at all. This selectivity in the ability of only specific reaction coordinates to positively couple to the shear driving is remarkable and could arise from the comparative mechanical deformation modes and radii of gyration of the C_{2v} and O_h clusters, as that is how they couple to a shear flow.

We have explicitly shown the steady-state averaged time-dependent reactive fluxes in the optimized solutions for high flux and high current in Figures 3.10(e) and (f) respectively The

vertical black dashed line denotes the choice of δt, and we see that the side-side correlation functions there have gone into the linear regimes but haven't started saturating yet. The correlation functions in both cases have periodic oscillations corresponding to the fast tumbling motion of the cluster in the sheared steady-state. At the long time limit $t \to \infty$, the reactive fluxes in both directions become identical and time-independent due to the loss of correlation. In subfigure (e), we see that the reactive flux has been maximized in both directions while approximately preserving their symmetry. The shear flow has an active role in this high flux, as clear from the same correlation functions with the same optimized alphabet but at $f = 0$. Subfigure (f) similarly shows the reactive fluxes for the optimized high-current forces. Here we see that the algorithm has found a way to maximally break the symmetry between the two reactive fluxes at the given value of δt, by delaying the reverse reactive flux compared to the forward. This breakdown of detailed balance in this prespecified reaction coordinate is evidently achieved by the only nonconservative force in the system, the shear flow, as demonstrated by the restoration of the symmetry by putting the shear flow rate to 0 even with the optimized pairwise forces.

We have thus shown that the variational algorithm can be used to tune reactive fluxes and currents in a nonequilibrium steady-state. In a sheared system this amounts to maximizing the coupling between mechanical driving and chemical degrees of freedom for the colloids, thus harvesting the dissipative driving to break detailed balance in select modes. The *optimizability* of the currents itself depends on the ability of the custers in the relevant direction to be mechanically affected by a shear flow. This approach opens a promising avenue for application of the variational algorithm to directly design functional materials which may convert dissipation into usable work. In the next section we look at such a molecular machine and directly optimize observables related to its performance.

3.6   Optimal design principles for molecular machines

Molecular machines are functional molecules that consume chemical energy at a nonequilibrium steady-state to perform useful mechanical or chemical work. The operation of molecular machines far from thermal equilibrium enables them to overcome thermodynamic bounds on performance as well as produce a finite power. The operation of molecular machines to produce desired levels of power at high efficiency has been extensively studied with the tools of stochastic thermodynamics and fluctuation theorems[19], with geometric near-equilibrium response theory [201, 224] and with large deviation theory for efficiency fluctuations [225]. The nonequilibrium operation of molecular machines have been analyzed in the framework of free energy and information transduction between multiple nonequilibrium reservoirs and the role of tight internal coupling for efficient operation of molecular machines has been recognized [226]. Nevertheless, most design approaches for molecular dynamics models of machines use phenomenological surveys to find high power and high efficiency solutions, instead of an automated method that can discover novel solutions. In this section we extend our variational algorithm for the discovery of high power and high efficiency regimes
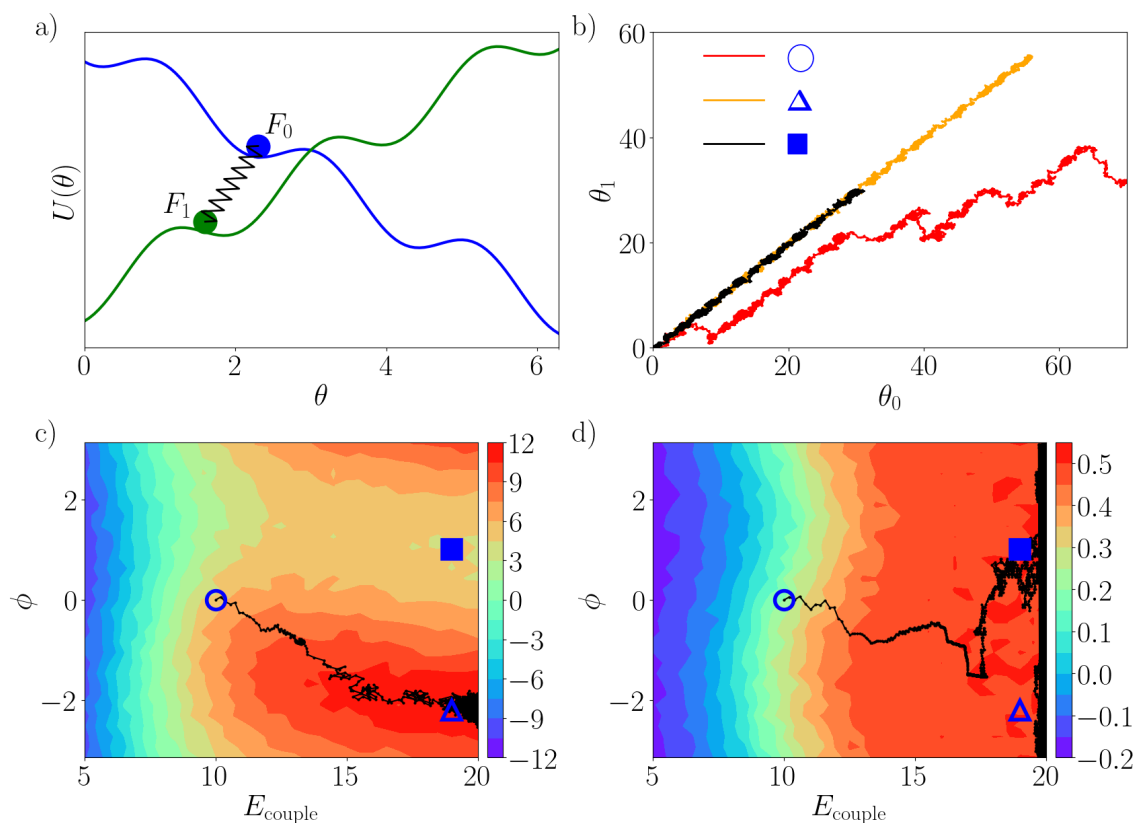
Figure 3.11: Optimal design principles for a model for ATP synthase. a) A schematic representation of the two-particle system, where the F_0 particle (blue) climbs down a potential ramp driven by μ_{H+}, and the coupled F_1 particle is pulled up a ramp doing work against μ_{ATP}. b) Representative statistical realization of the two-dimensional trajectories for a pre-optimization starting point (red), at a set of parameters with maximum efficiency but not high power (black), and a set of parameters with both maximum power and efficiency (orange). (c) and (d) are the optimization landscapes for power and efficiency respectively. Black lines are the trajectories taken by the variational algorithm through the parameter space in either of the two cases. The square, triangle and circle denote the set of parameters corresponding to (b).

that have been previously phenomenologically studied, in a coarde-grained model for the molecular motor ATP synthase.

The $F_0 - F_1$ ATP synthase is vital towards most of the ATP production from aerobic respiration in mitochondria. This motor resides across membranes in mitochondria and transforms cross-membrane proton motive force into energetic phosphate bonds to make ATP. The motor consists of two subunits F_0 and F_1 each having a three-fold symmetry. The

F_0 subunit is rotated by protons moving across the membrane due to proton motive force μ_{H+}, which in turn drives the F_1 subunit to rotate and produce μ_{ATP}, a chemical potential leading to ATP synthesis from ADP and inorganic phosphate. We model the rotation of the F_0 and F_1 subunits as overdamped Langevin dynamics of two coupled particles in one-dimensional periodic cosine potentials [227]. This model is schematically illustrated in Fig. 3.11(a). The coordinates of the particles, θ_0 and θ_1, diffuse in the potential energy landscape

$$V(\theta_0, \theta_1) = -\frac{1}{2}E_0 \cos n_o(\theta_0 - \phi) - \frac{1}{2}E_1 \cos n_1\theta_1 - \frac{1}{2}E_{\text{couple}} \cos(\theta_0 - \theta_1) \tag{3.31}$$

where E_0 and E_1 are barrier heights in the periodic potential, ϕ is a phase offset between θ_0 and θ_1 and E_{couple} is the coupling energy of the two particles. We allow θ_0 and θ_1 to be in the range $[0, 2\pi]$ with periodic boundaries. Aside from the conservative forces, the F_0 particle is driven by a constant external force μ_{H+} and the F_1 particle by a constant force μ_{ATP}. Choice of these two driving forces as $\mu_{H+} > 0$, $\mu_{\text{ATP}} < 0$ and $|\mu_{H+}| > |\mu_{\text{ATP}}|$ pushes the system out of equilibrium and allows the machine to do work against μ_{ATP}. Similar to Ref. [227], we choose $n_0 = n_1 = 3$ and $E_0 = E_1$. With this model, the input and output powers and the efficiency associated with each trajectory are defined by

$$\mathcal{P}_{H+}[\mathbf{X}(t)] = \frac{2\pi\mu_{H+}}{\tau} \int_0^\tau \dot{\theta}_0 dt \tag{3.32}$$

$$\mathcal{P}_{\text{ATP}}[\mathbf{X}(t)] = -\frac{2\pi\mu_{\text{ATP}}}{\tau} \int_0^\tau \dot{\theta}_1 dt \tag{3.33}$$

$$\eta = \frac{\mathcal{P}_{\text{ATP}}}{\mathcal{P}_{H+}} \tag{3.34}$$

with the efficiency being bounded by $-\mu_{\text{ATP}}/\mu_{H+}$. The parameter regimes for optimal performance in terms of these metrics in this model has been studied in detail in Ref. [227].

We use our variational algorithm to find optimal choice of the parameters used in Eq. 3.31 for independently maximizing power and efficiency. For a cost-function, we retain only the first term in Eq. 3.13 with the power or efficiency as observables, the choice of s thus being rendered to that of the learning rate. Optimization of a trajectory-averaged output power is exactly similar to the approach used for yields and reactive fluxes, and for gradients of trajectory averaged efficiency with respect to variational parameters $\{c\}$, we use the chain rule

$$\frac{\partial\eta}{\partial c} = \frac{\mathcal{P}_{H+}\partial_c\mathcal{P}_{\text{ATP}} - \mathcal{P}_{\text{ATP}}\partial_c\mathcal{P}_{H+}}{\mathcal{P}_{H+}^2} \tag{3.35}$$

We hold constant $E_0 = E_1 = 2k_{\text{B}}T$, $\mu_{H+} = 4k_{\text{B}}T$ and $\mu_{\text{ATP}} = 2k_{\text{B}}T$ to avoid trivial optimization of power with increased chemical potentials and decreasing barrier heights and that of efficiency with changing ratio of chemical potentials. Our variational parameters ϕ and E_{couple} are allowed to vary in $[-\pi, \pi]$ and $[5k_{\text{B}}T, 20k_{\text{B}}T]$ respectively. For simulating the dynamics at each optimization step and estimating gradients, we use a timestep of 10^{-5}, an

integration time Δt for Malliavin weights as 10, and a trajectory length of 500, all in units of the natural diffusive timescale of the system with friction coefficients and diffusion constant both being unity.

We initialize the power and efficiency optimization runs from $\phi = 0, E_{\text{couple}} = 10k_{\text{B}}T$ and had to lower learning rates in two steps to anneal to the solution well. The results of the optimization runs through the two-dimensional parameter landscape for power and efficiency are shown in Fig. 3.11(c) and (d) respectively, with single statistical realizations of equal durations of unoptimized, high-efficiency and high-power trajectories shown in Fig. 3.11(b). We see that unoptimized trajectories have generally correlated motion between θ_0 and θ_1 albeit with sudden failure modes that lead to loss of power and efficiency. In contrast, high power is uniquely achieved at a high value of E_{couple} around a phase shift of -2.2, which is similar to results in Ref. [227]. The high power trajectories also have high efficiency, as shown by the highly correlated motion of θ_1 versus θ_0 with a high slope. There is however a large near-degenerate regime of parameters, at high E_{couple}, where near-maximum efficiency close to the theoretical bound is achieved irrespective of the phase shift. This is the regime of tight coupling between input and output coordinates where any energy leakage is minimal. The algorithm finds this degenerate solution basin and diffuses out in it, as the noise in the stochastically evaluated gradients becomes comparable in magnitude to the gradient of the efficiency landscape. There is therefore a class of solutions at high coupling strength and positive phase shift that have high efficiency but low output power. One such statistical realization is shown in Fig. 3.11(b), where the two coordinates have tightly efficient correlations but the total amount of displacement for the same amount of elapsed time is less than the high-power solution. Further analysis in terms of the exact mechanism of successful and failed energy transduction segments in these trajectories is required in order to generalize these design principles to machines with more degrees of freedom that can dissipate and reroute energy.

3.7   Conclusion

We have developed an inverse design algorithm for the self-assembly of colloidal clusters in a nonequilibrium steady-state. The formalism exploits a variational structure originating from large deviation techniques for importance sampling in trajectory ensembles. The algorithm optimizes the yield of clusters of arbitrary shapes, sizes and geometry by tuning control forces within an arbitrarily chosen ansatz, with statistically estimated explicit gradients. We have demonstrated the performance of the algorithm using an ansatz of DNA-labeled colloidal clusters self-assembled in a shear flow, and have obtained design rules for different families of rigid and nonrigid clusters. This algorithm scales linearly both in system size and in the number of optimizable parameters in the force ansatz, but its performance is independent of the specific order parameter chosen for the self-assembly process. For example, the choice of a locally defined structural order parameter such as the density or the degree of crystallinity as the optimized observable can produce design principles for the assembly of extended

dispersed or periodic structures out of equilibrium. Similarly, dynamical order parameters like the instantaneous flux between two stable states and performance metrics of molecular machines like output power and efficiency can also be optimized using the same variational procedure in a suitable trajectory ensemble. Hence this algorithm can be used to tune both structural and dynamical properties of clusters in a nonequilibrium steady-state to produce dynamical phases having no equilibrium analogs.

This variational algorithm differs from other available inverse design algorithms for soft matter in and out of equilibrium. The equivalent variational structure in configuration space for systems in thermal equilibrium, where the potential energy function is optimizable and explicit gradients can be statistically estimated by autodifferentiation, has been used extensively as the basis of both importance sampling and inverse design algorithms [71, 156, 228]. This configuration space approach with explicit gradients is not feasible in nonequilibrium systems due to the probability measure being non-Boltzmann. Out of equilibrium, there have been theoretical approaches to rationalizing design principles in one or two component systems [22, 229, 21].

In the absence of a closed form expression for the configuration space measure, machine learning algorithms have been previously used to identity optimal design principles by tracking an order parameter during or at the end of finite-duration trajectories [230, 231, 232, 233]. Machine learning or neuroevolution based approaches are equivalent to estimating numerical gradients in the design space using finite-difference methods, and have similar convergence properties as explicit gradient based methods in the limit of small optimization steps [234]. However, since our multidimensional statistical gradient estimates are obtained using information from the same trajectory, our explicit gradient based method is expected to be advantageous in a high-dimensional design space as typically encountered in the bottom-up design of soft materials.

Finally, a class of design algorithms employ a trajectory ensemble based approach to statistically estimate explicit gradients of the dynamical response in colloidal systems [162, 222] and are formally closest to our approach. These algorithms probe the transient dynamical response of self-assembly trajectories, which however do not predict the long-time properties of the self-assembled cluster in a nonequilibrium steady-state. In our algorithm, we have explicitly evaluated the long-time steady-state limit and arrived at the novel fluctuation-dissipation relation in design space in Eq. 3.14. This will enable our explicit-gradient based method to be directly used to optimize both structural and dynamic properties of driven phases of soft matter, and automate the discovery of new functional materials.

Data Availability

Chapter 4

Reinforcement Learning of rare diffusive dynamics

We have so far developed an optimization technique based on stochastic gradient descent to optimize control forces to sample and design rare but important trajectories. The performance of our variational algorithm relies on getting a low-variance estimate of the explicit gradients of the cost function in parameter space. In the steady-state problems we have studied, we have seen a large improvement in statistical efficiency of sampling compared to direct population dynamics approaches, allowing us to probe novel regimes of rare fluctuations. However, the accelerated gradient descent method only incorporates memory through an inertia vector, while discarding prior information about the parts of the cost-function landscape that it visited before. Additionally, generalization of the optimization algorithm to finite duration trajectories amplifies the problem of noisy gradients, as in finite duration the gradients have to explicitly computed by a temporal double integral of a correlation function, instead of a single steady-state integral. This necessitates development of more advanced learning algorithms where auxiliary functions can reduce the variance of gradients and store memory of the cost-function landscape. In this chapter, borrowing notions from reinforcement learning [236], we have developed a method to generate rare dynamical trajectories directly through the optimization of an auxiliary dynamics that generates an ensemble of trajectories with the correct relative statistical weights. Within this ensemble of trajectories, a variational estimate of the likelihood of the rare event is obtainable from a simple expectation value. We have built on the work in Refs. [117, 125] and past literature on reinforcement learning for continuous time processes [237, 238, 239, 240, 241, 242, 243, 244].

   The techniques of reinforcement learning aim to learn the best decisions to make in each state in order to achieve some goal. Algorithms developed in this context have led to many significant advancements in recent years across tasks requiring an intelligent agent to interact with an environment, such as in gameplay [245, 246, 247] and robotics [248, 249, 250], with a variety of recent applications in physics [251, 252, 253, 254, 255, 256, 257, 258]. However, many of these situations are framed as discrete time problems, with relatively little work

done in stochastic continuous time control [237, 238]. For diffusive processes and importance sampling molecular dynamics, we formulate a reinforcement learning procedure to learn the correct force to influence the probability of choosing each next state. From this perspective, we take a policy gradient based approach [259, 260, 238, 248, 249], learning a generative model for the evolution of the state. The optimized force found is such that rare events are made typical while staying close to the original force, providing a dynamics that can aid in efficiently sampling the targeted trajectory ensemble.

A key advantage of the reinforcement learning techniques we develop is the use of an additional learning process for a function which guides the optimization of the dynamics, a so-called value function [261], which describes how relevant each state is to the rare events of interest. This value function substantially reduces the variance in estimates of the gradient of the parameters specifying a force, allowing for the use of less data in each optimization step and subsequently more complex approximations to the auxiliary dynamics. We show how this approach can be successfully applied to both finite time problems in which the dynamics is constrained to guarantee the occurrence of some rare transition like a barrier crossing, and to time-homogeneous problems where we are interested in the statistics of time-integrated observables in the long time limit as characterized by its large deviation function.[1]

4.1   Trajectory ensemble Formalism

We consider systems evolving with a diffusive dynamics over time t of a configuration $\mathbf{x}$. These configurations evolve according to a force vector $\mathbf{F}(\mathbf{x}, t)$ and noise vector of equal dimension $\mathbf{W}$ with associated constant noise matrix $\mathbb{G}$ invertible within the stochastically evolving subspace, represented by the stochastic differential equation (SDE)

$$d\mathbf{x} = \mathbf{F}(\mathbf{x}, t)dt + \mathbb{G} \cdot d\mathbf{W}, \tag{4.1}$$

where the noise $\mathbf{W}$ follows a Wiener process, with increments $d\mathbf{W}$ drawn from a Gaussian with zero mean and dt variance. Throughout we will work in dimensionless variables that imply unit energy scales and mobilities. The requirement of $\mathbb{G}$ being invertible within the stochastic subspace may in principle be relaxed, however in that case there may be multiple noise vectors corresponding to the same change of state, making the evaluation of transition probabilities necessary for our optimization approach difficult. We will follow the Ito convention for ease of notation and implementation with standard numerical integrators. Throughout, we do not assume in Eq. 4.1 that the force is gradient or that the noise obeys a detailed balance, and thus our approach is generally applicable to equilibrium as well as nonequilibrium dynamics.

We aim to probe rare fluctuations in trajectory observables. Here we consider trajectories, $\mathbf{X}_{0,T}$, defined as the sequence of configurations over an observation time T, though generalizations of fluctuating observation times are possible [263] . Generally, we will consider

---

[1]Most of the content of this chapter was originally part of the publication [262].

observables that are functions of time-integrated variables over the trajectory,

$$O\left[\mathbf{X}_{0,T}\right] = \int_0^T dt\, A[\mathbf{x}_t, t] + \mathbf{B}[\mathbf{x}_t, t] \cdot \dot{\mathbf{x}}(t), \tag{4.2}$$

where the first term is a state dependent observable, while the second term depends on a stochastic increment, with both $A[\mathbf{x}_t, t]$ and $\mathbf{B}[\mathbf{x}_t, t]$ being state dependent. However, we will also consider cases in which $A[\mathbf{x}_t, t]$ is a function of a single time in order to impose end point conditioning. Expectations of functions of such observables are defined through path integrals of the form

$$\langle f\left(O\left[\mathbf{X}_{t,t'}\right]\right)\rangle_p = \int D\mathbf{X}_{t,t'} d\mathbf{x}_t\, P\left[\mathbf{X}_{t,t'}\right] f\left(O\left[\mathbf{X}_{t,t'}\right]\right), \tag{4.3}$$

where $P\left[\mathbf{X}_{t,t'}\right]$ is the total probability of a trajectory decomposable into $P\left[\mathbf{X}_{t,t'}\right] = p\left[\mathbf{X}_{t,t'}|\mathbf{x}_t\right] \rho(\mathbf{x}_t)$ where $p\left[\mathbf{X}_{t,t'}|\mathbf{x}_t\right]$ is the transition probability conditioned on starting in configuration $\mathbf{x}_t$ with initial probability $\rho(\mathbf{x}_t)$.

Probabilities for trajectories between times t and t' starting at $\mathbf{x}_t$ are defined by

$$p\left[\mathbf{X}_{t,t'}|\mathbf{x}_t\right] \propto \exp\left\{-\tfrac{1}{2}\int_t^{t'} dt''\, |\mathbb{G}^{-1} \cdot (\dot{\mathbf{x}} - \mathbf{F})|^2\right\} \tag{4.4}$$

where we suppressed the arguments of $\mathbf{x}_t$ and $\mathbf{F}[\mathbf{x}_t, t]$ for shorthand. This is the standard Onsager-Machlup form for the diffusive dynamics considered here [264]. The measure over paths between times t and t' starting from position $\mathbf{x}_t$ is defined such that

$$\int D\mathbf{X}_{t,t'} p\left[\mathbf{X}_{t,t'}|\mathbf{x}_t\right] = 1 \tag{4.5}$$

where the transition probability is normalized when integrated over all trajectories. These path probabilities satisfy

$$p\left[\mathbf{X}_{t,t''}|\mathbf{x}_t\right] = p\left[\mathbf{X}_{t',t''}|\mathbf{x}_{t'}\right] p\left[\mathbf{X}_{t,t'}|\mathbf{x}_t\right] \tag{4.6}$$

and

$$D\mathbf{X}_{t,t''} = D\mathbf{X}_{t',t''} D\mathbf{X}_{t,t'} \tag{4.7}$$

due to the Markovian noise in Eq. 4.1.

Trajectories sampled with $P\left[\mathbf{X}_{0,T}\right]$ will be dominated by the most typical values of $O\left[\mathbf{X}_{0,T}\right]$. We will encode the rare trajectories with atypical values of $O\left[\mathbf{X}_{0,T}\right]$ by reweighting the original trajectory ensemble defined by Eq. 4.4, multiplying each trajectory by an observable dependent factor. Such reweightings occur naturally in statistical studies of rare events and are isomorphic to extended ensemble approaches in equilibrium configurational problems. The ensemble of events we are interested in is constructed by weighting the probability of trajectories in the original dynamics by an exponentially positive number,

$$P_s\left[\mathbf{X}_{0,T}\right] = e^{-sO\left[\mathbf{X}_{0,T}\right] - \lambda(s,T)} P\left[\mathbf{X}_{0,T}\right], \tag{4.8}$$

where $P_s[\mathbf{X}_{0,T}]$ is denoted as a tilted path ensemble, biased by a statistical field s in such a way to promote rare fluctuations in $O[\mathbf{X}_{0,T}]$. The quantity $\lambda(s,T)$ normalizes the tilted distribution, and is identifiable as a cumulant generating function (CGF)

$$\lambda(s,T) = \ln Z(s,T) = \ln \left\langle e^{-sO[\mathbf{X}_{0,T}]} \right\rangle_p, \tag{4.9}$$

and equal to the logarithm of the tilted path partition function $Z(s,T)$. The reweighted path ensemble generally defines a new transition probability $p_s[\mathbf{X}_{t,t'}|\mathbf{x}_t]$ and initial condition. The evaluation of $\lambda(s,T)$ is a common objective in studies of diffusive systems as it describes the statistics of $O[\mathbf{X}_{0,T}]$. Contributions to $\lambda(s,T)$ or $P_s[\mathbf{X}_{0,T}]$ are dominated by trajectories with large or small values of $O[\mathbf{X}_{0,T}]$, depending on the sign of s. The exponential bias, $\exp(-sO[\mathbf{X}_{0,T}])$, can also be constructed to function as a filter based on fulfilling specific criteria. In such cases $P_s[\mathbf{X}_{0,T}]$ is identified as the probability that a trajectory fulfills a specific conditioning, and its ensemble a corresponding conditioned path ensemble. Common examples are Brownian bridges [109, 265, 266], where trajectories are conditioned to end at $\mathbf{x}_T = \mathbf{x}'$, in which $O[\mathbf{X}_{0,T}]$ is 1 if $\mathbf{x}_T = \mathbf{x}'$ and is 0 otherwise, and s is taken sufficiently negative that only those trajectories for which the constraint is satisfied have significant weight.

4.2 Gradient optimization for finite time constrained dynamics

Our aim is to find a dynamics which generates trajectories with probability as close to the reweighted trajectories ensemble as possible. For the diffusive dynamics considered here, this is exactly achievable in principle through a so-called generalized Doob transformation [267, 142, 51, 63, 110, 268]. The generalized Doob transformation defines a modified dynamics with an added drift force that is generally time dependent but with an identical noise as in the original SDE. However, constructing this transformation is often not possible in practice, as it requires diagonalizing a modified Fokker-Planck operator which in interacting systems is exponentially complex [269]. Here we aim to parametrize a drift force with tunable parameters θ to approximate the generalized Doob transform. With the modified force, $\mathbf{F}_\theta(\mathbf{x},t)$, we have a modified SDE

$$d\mathbf{x} = \mathbf{F}_\theta(\mathbf{x},t)dt + \mathbb{G}\,d\mathbf{W}, \tag{4.10}$$

with corresponding trajectory probabilities

$$p_\theta[\mathbf{X}_{t,t'}|\mathbf{x}_t] \quad \propto \exp\left\{-\tfrac{1}{2}\int_t^{t'} dt'' \left|\mathbb{G}^{-1}\cdot(\dot{\mathbf{x}} - \mathbf{F}_\theta)\right|^2\right\} \tag{4.11}$$

which still satisfy the Markovian properties of the original dynamics and the same normalization constant. See Ref. [125] for a discussion of problems in which the optimal dynamics is required to be non-Markovian, in the context of discrete time Markov processes.

We seek to learn a set of parameters θ to minimize the Kullback-Leibler (KL) divergence between the modified dynamics and the reweighted trajectory ensemble defined by Eq. 4.8. The KL divergence is defined as

$$D_{\mathrm{KL}}(p_\theta|p_s) \;=\; \left\langle \ln\left(\frac{p_\theta[\mathbf{X}_{0,T}|\mathbf{x}_0]\rho(\mathbf{x}_0)}{p_s[\mathbf{X}_{0,T}|\mathbf{x}_0]\rho(\mathbf{x}_0)} \right) \right\rangle_{p_\theta}, \tag{4.12}$$

where the expectation is taken with respect to the parametrized dynamics. This quantity is a measure of the similarity between the modified and reweighted trajectory ensembles. Achieving a zero value when p_θ is given by the generalized Doob transform, the KL divergence has a unique minimum when this Doob transformed dynamics is contained within the space of parametrized dynamics, providing a variational estimate of the CGF. We note that this definition of the KL divergence differs from much of the literature considering optimization of a parametrized diffusive dynamics [270, 271, 272, 273], where the parametrized dynamics p_θ and target dynamics p_s appear in an opposite way. In principle the initial distribution should also be parametrized, as it will be modified by the reweighting, however depending on the space of distributions chosen these can be hard to sample. We drop this modification for simplicity.

Low variance gradient estimation

In order to optimize the force, $\mathbf{F}_\theta$, we follow techniques introduced in the reinforcement learning literature [236, 274, 275, 276, 248, 277]. Substituting the parametrized and reweighted trajectory probabilities into the KL divergence, we may rewrite it as an average over a parameter dependent time-integrated observable

$$D_{\mathrm{KL}}(p_\theta|p_s) \;=\; -\left\langle R\left[\mathbf{X}_{0,T}\right]\right\rangle_{p_\theta} + \lambda(s,T) \tag{4.13}$$

where in the language of reinforcement learning we define a return, $R\left[\mathbf{X}_{0,T}\right]$, as

$$R\left[\mathbf{X}_{0,T}\right] \;=\; -sO\left[\mathbf{X}_{0,T}\right] - \ln\left(\frac{p_\theta\left[\mathbf{X}_{0,T}|\mathbf{x}_0\right]}{p\left[\mathbf{X}_{0,T}|\mathbf{x}_0\right]} \right) \tag{4.14}$$

with the negative of the average of the second term measuring the KL divergence, $D_{\mathrm{KL}}(p_\theta|p)$, between the parametrized dynamics and the original dynamics. This return is analogous to a regularized form of reinforcement learning [275, 277] similar to that considered in maximum-entropy reinforcement learning [248, 249, 276]. When evaluated at the generalized Doob transform the KL divergence vanishes and the return evaluates to the CGF. Away from the Doob transform, the positivity of the KL divergence results in the return variationally bounding the CGF from below [111].

We aim to minimize the KL divergence through stochastic gradient descent in the parameter space. For this we need the gradient of $D_{\mathrm{KL}}(p_\theta|p_s)$ with respect to θ,

$$\nabla_\theta D_{\mathrm{KL}}(p_\theta|p_s) \;=\; -\;\left\langle R\left[\mathbf{X}_{0,T}\right] \nabla_\theta \ln p_\theta\left[\mathbf{X}_{0,T}|\mathbf{x}_0\right]\right\rangle_{p_\theta} \tag{4.15}$$

where we note

$$\langle \nabla_\theta R \left[\mathbf{X}_{0,T} \right] \rangle_{p_\theta} = 0 \tag{4.16}$$

due to conservation of probability [125]. The factor multiplying the return is commonly referred to as the Malliavin weight in the stochastic analysis literature [122], and corresponds to a particular case of the eligibility traces found in reinforcement learning [261, 236, 278, 279, 280], which we denote as $y_\theta(T) = \nabla_\theta \ln p_\theta \left[\mathbf{X}_{0,T} | \mathbf{x}_0 \right]$. It can be rewritten by substituting the path probability,

$$y_\theta(t'') - y_\theta(t') = \int_{t'}^{t''} dt \, \dot{y}_\theta(t), \tag{4.17}$$

where

$$\dot{y}_\theta(t) = \left[\mathbb{G}^{-1} \cdot (\dot{\mathbf{x}}(t) - \mathbf{F}_\theta(t)) \right] \cdot \left[\mathbb{G}^{-1} \cdot \nabla_\theta \mathbf{F}_\theta(t) \right] \tag{4.18}$$

is the integrand of the Malliavin weight.

Were we to stop at Eq. 4.15, we would proceed to optimize a generative model (the diffusive dynamics with our parameterized force) of the trajectories using a score-function based approach, similar to standard unsupervised learning. However, following the methods of reinforcement learning, we can use a combination of the Markovianity of the generative model and other variance reduction techniques to produce a gradient estimator which is much more efficient to estimate. To begin with, we can simplify Eq. 4.15 by noting that due to Markovianity, the Malliavin weight only correlates with the return in the future, and we can rewrite the gradient as

$$
\begin{aligned}
\nabla_\theta D_{\mathrm{KL}}(p_\theta | p_s) &= - \left\langle \int_0^T dt R \left[\mathbf{X}_{t^-,T} \right] \dot{y}_\theta(t) \right\rangle_{p_\theta} \\
&= \chi_{\mathrm{MCR}}(\theta, T), \tag{4.19}
\end{aligned}
$$

where we used t^- as a shorthand for $t - \epsilon$ for some small positive ϵ. We refer to the optimization of the modified dynamics using this formulation of the gradient as χ_{MCR}, as it is analogous to the Monte-Carlo returns (MCR), or REINFORCE [281, 282] policy gradient algorithm in reinforcement learning. In the long observation time limit, employing this gradient in stochastic optimization reduces to previous variational Monte Carlo procedures [117].

This estimator of the gradient is non-optimal for two reasons. First, it requires evaluation of a two time correlation function. In steady state, stationarity can be invoked to eliminate one of those integrals, however under finite time conditioning this simplification is not possible. Second, it has a high variance and requires significant averaging to converge accurate gradients. This is because both the Malliavin weight and the return undergo a random walk with linearly increasing variance [122]. Building on the analogies with the reinforcement learning formalism we define a value function as a path average of the return,

$$V(\mathbf{x}, t) = \langle R \left[\mathbf{X}_{t,T} \right] \rangle_{p_\theta, \mathbf{x}} . \tag{4.20}$$

conditioned on starting at the position and time, $\mathbf{x}_t = \mathbf{x}$. Introduced into the gradients of $D_{\mathrm{KL}}(p_\theta|p_s)$ in distinct ways, the value functions can be used to tame both problems of the naive MCR gradient estimate.

First, we introduce a value function as a baseline that depends only on the state at the time t in order to reduce the variance of the gradient. We note that $\dot{y}_\theta(t)$ is linear in the noise and thus averages to zero when multiplied by a function of the state at or before t. Defining a temporal difference error

$$\delta\left[\mathbf{X}_{t^-,T}, t\right] = R\left[\mathbf{X}_{t^-,T}\right] - V\left(\mathbf{x}_t, t\right), \tag{4.21}$$

we write the dynamical gradient as

$$
\begin{aligned}
\nabla_\theta D_{\mathrm{KL}}(p_\theta|p_s) &= -\left\langle \int_0^T dt\, \delta\left[\mathbf{X}_{t^-,T}, t\right] \dot{y}_\theta(t) \right\rangle_{p_\theta} \\
&= \chi_{\mathrm{MCVB}}(\theta, T)
\end{aligned}
\tag{4.22}
$$

where we have formally subtracted zero. We refer to this gradient estimator as χ_{MCVB}, for Monte Carlo Value Baseline (MCVB) [236]. The subtraction of the state point dependent value function reduces the variance of the gradient by accounting for the mean uncorrelated part of each return between t^- and T with $\dot{y}_\theta(t)$, focusing on how this return differs from the average behaviour encoded by the value function.

Second, we introduce a value function that encodes an estimate of the return in the future in order to further reduce the variance and also the complications associated with estimating the two-time correlation function. We can replace part of the return by a value function that is conditioned at some τ, such that $t^- < \tau < T$,

$$
\begin{aligned}
\langle R\left[\mathbf{X}_{t^-,T}\right] \dot{y}_\theta(t) \rangle &= \langle V\left(\mathbf{x}_{t+\tau}, t+\tau\right) \dot{y}_\theta(t) \rangle \\
&\quad + \langle R\left[\mathbf{X}_{t^-,t+\tau}\right] \dot{y}_\theta(t) \rangle
\end{aligned}
\tag{4.23}
$$

where we set the value function to zero for $V(\mathbf{x}, t)$ with $t > T$. Combining this value function form of the kernel of the gradient with the value baseline, we define another temporal difference error

$$
\begin{aligned}
&\delta'\left[\mathbf{X}_{t^-,t+\tau}, t\right] \\
&= V\left(\mathbf{x}_{t+\tau}, t+\tau\right) + R\left[\mathbf{X}_{t^-,t+\tau}\right] - V\left(\mathbf{x}_t, t\right),
\end{aligned}
\tag{4.24}
$$

and we arrive at a distinct formulation of the gradient

$$
\begin{aligned}
\nabla_\theta D_{\mathrm{KL}}(p_\theta|p_s) &= -\left\langle \int_0^T dt\, \delta'\left[\mathbf{X}_{t^-,t+\tau}, t\right] \dot{y}_\theta(t) \right\rangle_{p_\theta} \\
&= \chi_{\mathrm{AC}}(\theta, T)
\end{aligned}
\tag{4.25}
$$

which we denote $\chi_{\mathrm{AC}}(\theta, T)$ for actor-critic gradient (AC) estimator, for the analogous algorithm in reinforcement learning [236, 248]. Here the value function is seen as criticizing the

transitions generated by the dynamics, i.e. the actor. Variance reduction of gradient estimates is therefore achieved by replacing potentially noisy return samples with the average behaviour expected in the future of the $\mathbf{x}_{t+\tau}$ state. In Sec. 4.4, we will compare the accuracy and statistical efficiency of these three gradient estimators: MCR, MCVB, and AC. Before that we discuss how the value functions are simultaneously parametrized and learnt along side the modified force.

Parametrizing value functions

While the gradient expressions are exact and the use of value functions expected to facilitate their convergence, using them requires knowledge of the exact value function for the modified dynamics, a formidable task in complex problems. In order to make their use tractable, we optimize a representation of the value function in addition to the modified force. Specifically, we introduce a parametrization of the value function denoted V_ψ. To optimize this approximation we note that the value functions satisfy a self-consistency equation called the Bellman equation [283]

$$V(\mathbf{x}, t) = \langle V(\mathbf{x}_{t+\tau}, t+\tau) + R[\mathbf{X}_{t,t+\tau}] \rangle_{p_\theta, \mathbf{x}}, \tag{4.26}$$

which has a unique solution for a given dynamics and return (as defined by the tilting observable and the dynamics via Eq. 4.14). We aim to minimize the error in this equation, thus optimizing our parametrized value towards this unique solution. Our approach is to minimize the squared difference between the two sides of Eq. 4.26 with the true value function replaced by the parametrized value function, and apply gradient descent to it. Such an approach is the subject of gradient temporal difference methods [284, 285, 286], but produces a gradient estimate which is difficult to evaluate, containing products of expectations which require independent samples. A part of the resultant gradient is however simpler to compute. We derive it by substituting only the right hand side of Eq. 4.26 with our parametrized value function to provide a fixed target for the left and defining a corresponding error function based on the squared difference. To construct a loss, we integrate these errors along each trajectory, and average them over the trajectory ensemble. This results in a loss function $L(\psi, \psi_i)$, that we take as a function of two weights, ψ and ψ_i,

$$\begin{aligned} L(\psi, \psi_i) = \\ \frac{1}{2} \Bigg\langle \int_0^T dt \Big\{ \langle V_{\psi_i}(\mathbf{x}_{t+\tau}, t+\tau) + R[\mathbf{X}_{t,t+\tau}] \rangle_{p_\theta, \mathbf{x}} \\ - V_\psi(\mathbf{x}_t, t) \Big\}^2 \Bigg\rangle_{p_\theta}, \end{aligned} \tag{4.27}$$

where the weight ψ_i is the weights after update i, used to provide the fixed target estimate towards which we want to move the functional of ψ. The derivative is then taken with respect

to ψ, before setting $\psi = \psi_i$ to find the gradient of this loss for the current parameters. Such an approach is referred to as semi-gradient in the reinforcement learning literature,[236] used to achieve the majority of state-of-the-art reinforcement learning results, and proves stable provided the data used to estimate the gradient is sampled using a dynamics which is close to p_θ as we intend to do. As mentioned above, alternative methods which additionally consider the variation of the target with ψ can be found in the RL literature, allowing for the use of data sampled from an alternative dynamics, utilized via importance sampling [284, 285, 286].

Writing an approximate temporal difference for the value function parametrization, within MCVB

$$\delta_\psi \left[\mathbf{X}_{t^-,T}, t \right] = R \left[\mathbf{X}_{t^-,T} \right] - V_\psi \left(\mathbf{x}_t, t \right), \tag{4.28}$$

or for AC

$$
\begin{aligned}
&\delta'_\psi \left[\mathbf{X}_{t^-,t+\tau}, t \right] \\
&= V_\psi \left(\mathbf{x}_{t+\tau}, t + \tau \right) + R \left[\mathbf{X}_{t^-,t+\tau} \right] - V_\psi \left(\mathbf{x}_t, t \right),
\end{aligned} \tag{4.29}
$$

we have gradients of the form

$$
\begin{aligned}
&\nabla_\psi L(\psi, \psi_i)|_{\psi=\psi_i} \\
&= -\left\langle \int_0^T dt \, \delta_{\psi_i} \left[\mathbf{X}_{t^-,T}, t \right] \nabla_\psi V_\psi \left(\mathbf{x}_t, t \right)|_{\psi=\psi_i} \right\rangle_{p_\theta},
\end{aligned} \tag{4.30}
$$

for the loss function from the value function parametrization, where for the AC algorithm δ_{ψ_i} is replaced with δ'_{ψ_i}. Given this value function approximation, we can approximate the gradient of the KL divergence by replacing the exact temporal difference with these approximate temporal differences. We then use the same trajectories to estimate the force and value function gradients and simultaneously learn both. For the MCVB algorithm, an approximate value function does not bias the gradients as the future return that correlates with the Malliavin weight stays intact and the expectation of the Malliavin weight is identically 0. However, for the AC algorithm, an approximate value function can introduce a bias into gradients as it replaces the average of the future return, which it may not accurately represent.

Employing gradients with or without value functions, we can construct a stochastic descent algorithm to optimize the modified forces which can be used to estimate the likelihoods of rare events and the trajectories by which they emerge. The algorithms require the evaluation of the forces, value function, their parametric gradients and noises over the course of simulating trajectories. Ensembles of trajectories can then be used to construct an empirical estimate of the gradient via computing the Malliavin weights, returns, and the temporal difference. These empirical estimates then iterate the two weights with respective learning rates α^θ and α^ψ for the force and value function respectively. The resultant algorithm is

outlined in pseudocode below in Alg. 2. Detailed versions of the individual algorithms with computationally efficient on-the-fly implementations for simulating trajectories with discrete timesteps are presented in the next section.

Algorithm 2 Gradient optimization using finite time trajectories

---

1: **inputs** dynamical approximation $F_\theta(\mathbf{x}, t)$, value approximation $V_\psi(\mathbf{x}, t)$
2: **parameters** learning rates α^θ, α^ψ; total optimization steps I; trajectory length T consisting of J timesteps of duration Δt each; number of trajectories N
3: **initialize** choose initial weights θ and ψ, define iteration variables i and j, force and value function gradients δ_P, δ_V, temporal difference δ (can be $R[\mathbf{X}_{t^-,T}]$ or $\delta_\psi[\mathbf{X}_{t^-,T}, t]$ or $\delta'_\psi[\mathbf{X}_{t^-,t+\tau}, t]$ for MCR/MCVB/AC)
4: $i \leftarrow 0$
5: **repeat**
6:     Using chosen method to generate trajectories $\mathbf{X}_{0,T}$ with configurations, times and temporal differences denoted by $\mathbf{x}_j, t_j$ and δ_j respectively.
7:     $j \leftarrow 0$
8:     $\delta_P \leftarrow 0$
9:     $\delta_V \leftarrow 0$
10:     **repeat**
11:         $\delta_P \leftarrow \delta_P + \delta_j \dot{y}_\theta(t_j) \Delta t$
12:         $\delta_V \leftarrow \delta_V + \delta_j \nabla_\psi V_\psi(\mathbf{x}_j, t_j) \Delta t$
13:         $j \leftarrow j + 1$
14:     **until** $j = J$
15:     average δ_P, δ_V over N trajectories to get $\bar{\delta}_P, \bar{\delta}_V$
16:     $\theta \leftarrow \theta + \alpha^\theta \bar{\delta}_P$
17:     $\psi \leftarrow \psi + \alpha^\psi \bar{\delta}_V$
18:     $i \leftarrow i + 1$
19: **until** $i = I$

4.3 Discrete timestep implementations of finite time algorithms

We now describe how the time-continuous equations of the reinforcement learning algorithm are efficiently implemented in simulations with a fixed discrete timestep Δt, though variable timesteps may be easily used. We use an Euler propagator to integrate the SDE in Equation (4.10) as

$$\mathbf{x}_{t+\Delta t} = \mathbf{x}_t + \Delta t \mathbf{F}_\theta(\mathbf{x}_t, t) + \mathbb{G}\Delta \mathbf{W}_t \tag{4.31}$$

where $\Delta\mathbf{W}$ is a Gaussian random variable with mean 0 and variance Δt. The trajectory probability from Eq. (4.11) is now given by products of stepwise probabilities

$$
\begin{aligned}
& p_\theta \left[\mathbf{X}_{t,t+\Delta t} | \mathbf{x}_t \right] \\
& = \frac{\exp\left\{ -\frac{1}{2\Delta t} \left| \mathbb{G}^{-1} \left(\mathbf{x}_{t+\Delta t} - \mathbf{x}_t - \Delta t \mathbf{F}_\theta(\mathbf{x}_t, t) \right) \right|^2 \right\}}{2\pi\Delta t \, \det(\mathbb{G})}
\end{aligned}
\tag{4.32}
$$

Next we discretize the gradient of the logarithm of trajectory probabilities using the Ito convention. We propagate the Malliavin weights from Eq. (4.18) as

$$
\begin{aligned}
y_\theta(t + \Delta t) = y_\theta(t) + & \left[\mathbb{G}^{-1} \left(\mathbf{x}_{t+\Delta t} - \mathbf{x}_t - \Delta t \mathbf{F}_\theta(\mathbf{x}_t, t) \right) \right] \\
& \cdot \left[\mathbb{G}^{-1} \nabla_\theta \mathbf{F}_\theta(t) \right]
\end{aligned}
\tag{4.33}
$$

We also write the full return (4.14) through a sum of stepwise rewards as

$$
R\left[\mathbf{x}_{t^-, t+\tau} \right] = \sum_{j : j\Delta t < \tau} r\left(\mathbf{x}_{j+1}, \mathbf{x}_j, t + j\Delta t \right)
\tag{4.34}
$$

where the timestep index j starts from -1 in this sum, with the notation t^- accounting for the timestep before the current one, and the subscript j refers to the time $t + j\Delta t$. The reward at each step is defined as

$$
\begin{aligned}
& r\left(\mathbf{x}_{j+1}, \mathbf{x}_j, t + j\Delta t \right) \\
& = -s \left(A_j \Delta t + \mathbf{B}_j \cdot (\mathbf{x}_{j+1} - \mathbf{x}_j) + A(\mathbf{x}_{j+1}) \delta_{jn} \right) \\
& \quad + \frac{\left[\mathbb{G}^{-1}(\mathbf{x}_{j+1} - \mathbf{x}_j - \Delta t \mathbf{F}_\theta(\mathbf{x}_j, t_j)) \right]^2}{2} \\
& \quad - \frac{\left[\mathbb{G}^{-1}(\mathbf{x}_{j+1} - \mathbf{x}_j - \Delta t \mathbf{F}(\mathbf{x}_j, t_j)) \right]^2}{2},
\end{aligned}
\tag{4.35}
$$

using the definition of the observable from Eq. (4.2) and accounting for an additional singular reward at the end of the trajectory after the last timestep n. Here the first three terms come from the observable and the last two terms represent the KL divergence between the original and optimized dynamics.

Now we combine the rewards, Malliavin weights and value functions in multiple ways to produce the gradients in the different algorithms. The pseudocodes of efficient implementations of these are presented below.

Monte-Carlo returns

The gradient in the Monte Carlo returns algorithm can be rewritten from Equation (4.19) as

$$
\begin{aligned}
\chi_{\mathrm{MCR}}(\theta, T) &= -\left\langle \int_0^T dt\, R\left[\mathbf{X}_{t^-,T}\right] \dot{y}_\theta(t) \right\rangle_{p_\theta} \\
&= -\left\langle \int_0^T dt\, \dot{y}_\theta(t) \int_{t^-}^T dt'\, \dot{R}(t') \right\rangle_{p_\theta} \\
&= -\left\langle \int_0^T dt\, \dot{R}(t) \int_0^{t^+} dt'\, \dot{y}_\theta(t') \right\rangle_{p_\theta} \\
&= -\left\langle \int_0^T dt\, \dot{R}(t) y_\theta(t^+) \right\rangle_{p_\theta}
\end{aligned}
\tag{4.36}
$$

where the return has been written as a time integral of its differential changes, and t^+ is shorthand for $t+\epsilon$ for some small positive ϵ. This has converted the double time integral into a single time integral, which is then evaluated on-the-fly while propagating the trajectory. An implementation of this algorithm with a fixed timestep Δt is described in the pseudocode in Alg. 3.

Algorithm 3 Finite time MCR

1: **inputs** dynamical approximation $\mathbf{F}_\theta(\mathbf{x}, t)$
2: **parameters** learning rate α^θ; total optimization steps I; trajectory length T consisting of J timesteps of duration Δt each; number of trajectories N
3: **initialize** choose initial weights θ, define iteration variables i and j, force gradient δ_P, stepwise rewards r representing the increments in return
4: $i \leftarrow 0$
5: **repeat**
6:     Using chosen method to generate trajectories $\mathbf{X}_{0,T}$ with configurations, times, noises, Malliavin weights and rewards denoted by $\mathbf{x}_j, t_j, \Delta\mathbf{W}_j, y_\theta(t_j)$ and $r(\mathbf{x}_{j+1}, \mathbf{x}_j, t_j) = r_j$ respectively
7:     $j \leftarrow 0$
8:     $\delta_P \leftarrow 0$
9:     $y_\theta(t_0) \leftarrow 0$
10:     **repeat**
11:         $y_\theta(t_{j+1}) \leftarrow y_\theta(t_j) + \Delta\mathbf{W}_j \cdot [\mathbb{G}^{-1}\nabla_\theta\mathbf{F}_\theta(\mathbf{x}_j, t_j)]$
12:         $\delta_P \leftarrow \delta_P + r_j y_\theta(t_{j+1})$
13:         $j \leftarrow j + 1$
14:     **until** $j = J$
15:     average δ_P over N trajectories to get $\bar{\delta}_P$
16:     $\theta \leftarrow \theta + \alpha^\theta\bar{\delta}_P$
17:     $i \leftarrow i + 1$
18: **until** $i = I$

Monte-Carlo returns with a value baseline

We use a similar technique to rewrite the double time integral for the gradient in the Monte Carlo value baseline algorithm, Equation (4.22), using a single time integral as

$$
\begin{aligned}
&\chi_{\text{MCVB}}(\theta, T) \\
&= -\left\langle \int_0^T dt \; \{R\left[\mathbf{X}_{t^-,T}\right] - V_\psi(\mathbf{x}_t, t)\} \, \dot{y}_\theta(t) \right\rangle_{p_\theta, \psi=\psi_i} \\
&= -\left\langle \int_0^T dt \; \left\{\dot{R}(t)y_\theta(t^+) - V_\psi(\mathbf{x}_t, t)\dot{y}_\theta(t)\right\} \right\rangle_{p_\theta, \psi=\psi_i}.
\end{aligned} \tag{4.37}
$$

We rewrite the gradient of the value error in Eq. (4.30) similarly as

$$
\nabla_\psi L(\psi, \psi_i)\Big|_{\psi=\psi_i}
$$
$$
= -\left\langle \int_0^T dt \left\{ \dot{R}(t) \left(\int_0^{t^+} dt' \nabla_\psi V_\psi(t') \right) \right.\right.
$$
$$
\left.\left. - V_\psi(t) \nabla_\psi V_\psi(t) \right\} \right\rangle_{p_\theta, \psi=\psi_i}
$$
$$
= -\left\langle \int_0^T dt \left\{ \dot{R}(t) z_\psi(t^+) - V_\psi(t) \dot{z}_\psi(t) \right\} \right\rangle_{p_\theta, \psi=\psi_i}, \tag{4.38}
$$

where the arguments of the value function $V_\psi(\mathbf{x}_t, t)$ have been suppressed as $V_\psi(t)$ and the integral of the gradient of the value function upto and including current time has been denoted as $z_\psi(t^+)$. We explicitly set the $V(\mathbf{x}_t, t)$ to 0 for any $t \geq T$, i.e., after the last timestep, in these expressions. The single time integral is then evaluated on-the-fly as the trajectory is propagated. If the force and the value function approximations use the same set of basis functions as we do with a fixed grid of Gaussians, the MCVB algorithm incurs no additional computational cost over the MCR algorithm. An implementation of this algorithm with a fixed timestep Δt is described in the pseudocode in Alg. 4.

Actor-critic

We rewrite the gradient in the Actor-critic algorithm from Equation (4.25) using a shift in time origin as

$$
\chi_{\text{AC}}(\theta, T)
$$
$$
= -\left\langle \int_0^T dt\, \delta'\left[\mathbf{X}_{t^-, t+\tau}, t\right] \dot{y}_\theta(t) \right\rangle_{p_\theta, \psi=\psi_i}
$$
$$
= -\left\langle \int_\tau^{T+\tau} dt\, \delta'\left[\mathbf{X}_{t^--\tau, t}, t-\tau\right] \dot{y}_\theta(t-\tau) \right\rangle_{p_\theta, \psi=\psi_i} \tag{4.39}
$$

where the change in return and the value function for $t \geq T$ is explicitly set to 0. We similarly write the gradient of the value error from Eq. (4.30) as

$$
\nabla_\psi L(\psi, \psi_i)\Big|_{\psi=\psi_i} = -\left\langle \int_\tau^{T+\tau} dt\, \delta'\left[\mathbf{X}_{t^--\tau, t}, t-\tau\right] \right.
$$
$$
\left. \nabla_\psi V_\psi(\mathbf{x}_{t-\tau}, t-\tau) \right\rangle_{p_\theta, \psi=\psi_i} \tag{4.40}
$$

Algorithm 4 Finite time MCVB
___

1: **inputs** dynamical approximation $\mathbf{F}_\theta(\mathbf{x}, t)$, value approximation $V_\psi(\mathbf{x}, t)$
2: **parameters** learning rates α^θ, α^ψ; total optimization steps I; trajectory length T consisting of J timesteps of duration Δt each; number of trajectories N
3: **initialize** choose initial weights θ and ψ, define iteration variables i and j, force and value function gradients δ_P, δ_V, stepwise rewards r representing the increments in return
4: $i \leftarrow 0$
5: **repeat**
6:     Using chosen method to generate trajectories $\mathbf{X}_{0,T}$ with configurations, times, noises, Malliavin weights, integral of value function gradients, and rewards denoted by $\mathbf{x}_j, t_j, \Delta\mathbf{W}_j, y_\theta(t_j), z_\psi(t_j)$ and $r(\mathbf{x}_{j+1}, \mathbf{x}_j, t_j) = r_j$ respectively
7:     $j \leftarrow 0$
8:     $\delta_P \leftarrow 0$
9:     $\delta_V \leftarrow 0$
10:     $y_\theta(t_0) \leftarrow 0$
11:     $z_\psi(t_0 \leftarrow 0)$
12:     **repeat**
13:         $\dot{y}_\theta(t_j) \leftarrow \Delta\mathbf{W}_j \cdot [\mathbb{G}^{-1}\nabla_\theta\mathbf{F}_\theta(\mathbf{x}_j, t_j)]/\Delta t$
14:         $y_\theta(t_{j+1}) \leftarrow y_\theta(t_j) + \Delta t\dot{y}_\theta(t_j)$
15:         $\dot{z}_\psi(t_j) \leftarrow \nabla_\psi V_\psi(\mathbf{x}_j, t_j)$
16:         $z_\psi(t_{j+1}) \leftarrow z_\psi(t_j) + \Delta t\dot{z}_\psi(t_j)$
17:         $\delta_P \leftarrow \delta_P + r_j y_\theta(t_{j+1}) - V_\psi(\mathbf{x}_j, t_j)\dot{y}_\theta(t_j))$
18:         $\delta_V \leftarrow \delta_V + r_j z_\psi(t_{j+1}) - V_\psi(\mathbf{x}_j, t_j)\dot{z}_\psi(t_j)$
19:         $j \leftarrow j + 1$
20:     **until** $j = J$
21:     average δ_P, δ_V over N trajectories to get $\bar{\delta}_P, \bar{\delta}_V$
22:     $\theta \leftarrow \theta + \alpha^\theta\bar{\delta}_P$
23:     $\psi \leftarrow \psi + \alpha^\psi\bar{\delta}_V$
24:     $i \leftarrow i + 1$
25: **until** $i = I$

These integrals are then evaluated on-the-fly along with trajectory propagation. Since the gradients involve correlations of the differential return r with the differential Malliavin weight $\dot{y}_\theta$ and the value function gradient $\dot{z}_\psi = \nabla_\psi V_\psi$ from τ time in the past, this makes it necessary to store and use this history, along with the reward and the value function, for the past $\tau/\Delta t$ timesteps. Aside from this additional memory requirement, given a delay time τ which is much smaller than the trajectory duration, the Actor-critic algorithm has similar computational cost comparable to the MCR and MCVB algorithms. This implementation of the algorithm is described in the pseudocode in Alg. 5.

Algorithm 5 Finite time AC

1: **inputs** dynamical approximation $\mathbf{F}_\theta(\mathbf{x}, t)$, value approximation $V_\psi(\mathbf{x}, t)$
2: **parameters** learning rates α^θ, α^ψ; total optimization steps I; trajectory length T consisting of J timesteps of duration Δt each; temporal delay $M = \tau/\Delta t$; number of trajectories N
3: **initialize** choose initial weights θ and ψ, define iteration variables i and j, force and value function gradients δ_P, δ_V, stepwise rewards r representing the increments in return
4: $i \leftarrow 0$
5: **repeat**
6:     Using chosen method to generate trajectories $\mathbf{X}_{0,T}$ with configurations, times, noises, changes in Malliavin weights, value function gradients, temporal difference, rewards and cumulative rewards denoted by $\mathbf{x}_j, t_j, \Delta\mathbf{W}_j, \Delta y_\theta(t_j), \dot{z}_\psi(t_j), \delta'_j, r(\mathbf{x}_{j+1}, \mathbf{x}_j, t_j) = r_j$ and $R\left[\mathbf{X}_{t_j-\tau, t_j}\right] = R_{j-M,j}$ respectively, and $r_j = V(\mathbf{x}, t_j) = 0$ whenever $j < 0$ or $j \geq J$
7:     $j \leftarrow 0$
8:     $\delta_P \leftarrow 0$
9:     $\delta_V \leftarrow 0$
10:     $R_{-M,0} \leftarrow 0$
11:     **repeat**
12:         $R_{j-M,j} \leftarrow R_{j-M-1,j-1} + r_j - r_{j-M}$
13:         **if** $j < J$ **then**
14:             $\Delta y_\theta(t_j) \leftarrow \Delta\mathbf{W}_j \cdot [\mathbb{G}^{-1}\nabla_\theta\mathbf{F}_\theta(\mathbf{x}_j, t_j)]$
15:             $\dot{z}_\psi(t_j) \leftarrow \nabla_\psi V_\psi(\mathbf{x}_j, t_j)$
16:         **end if**
17:         **if** $j \geq M$ **then**
18:             $\delta'_j \leftarrow V(\mathbf{x}_j, t_j) + R_{j-M,j} - V(\mathbf{x}_{j-M}, t_{j-M})$
19:             $\delta_P \leftarrow \delta_P + \delta'_j \Delta y_\theta(t_{j-M})$
20:             $\delta_V \leftarrow \delta_V + \delta'_j \dot{z}_\psi(t_{j-M})$
21:         **end if**
22:         $j \leftarrow j + 1$
23:     **until** $j = J + M$
24:     average δ_P, δ_V over N trajectories to get $\bar{\delta}_P, \bar{\delta}_V$
25:     $\theta \leftarrow \theta + \alpha^\theta \bar{\delta}_P$
26:     $\psi \leftarrow \psi + \alpha^\psi \bar{\delta}_V$
27:     $i \leftarrow i + 1$
28: **until** $i = I$

4.4 Rare fluctuations in finite time

We have used the algorithms discussed above to examine rare fluctuations of trajectories of fixed duration, starting from a fixed point in configuration space. The specific observable we

have investigated is an indicator function for reaching a desired region, Γ, in configuration space, $O[\mathbf{X}_{0,T}] = h_\Gamma[\mathbf{x}_T]$, where

$$h_\Gamma[\mathbf{x}_T] = \begin{cases} 1 & \mathbf{x}_T \in \Gamma \\ 0 & \text{otherwise} \end{cases},$$

at the final time T. Rare trajectories reaching a target basin in configuration space are often of interest as transition paths for reactive events, and significant development has been undertaken to efficiently generate them [70, 287, 288, 289, 105]. Computing optimal drift forces for generating these rare trajectories enables the study of reactive dynamics in a direct manner. We expect these algorithms to find use in the study of diffusive dynamics where Monte Carlo approaches have difficulty sampling [82, 290, 291, 292]. Further, as the modified force is used with the original noise from the SDE, we have access to the full reactive trajectory ensemble allowing the interrogation of the statistics of the reactive events in a way that other direct path methods like nudged elastic band and zero temperature string methods do not, as they represent only the dominant path [293, 294, 295, 296]. As a consequence, we expect out method will find use when there is a large path space entropy.

The CGF for an indicator variable is given by

$$\lambda(s, T) = \ln \left\langle e^{-s h_\Gamma[\mathbf{x}_T]} \right\rangle_p \tag{4.41}$$

as an average in the original reference dynamics. From Eq. (4.13), the KL divergence being nonnegative implies the average return is bounded above by the value of the CGF $\lambda(s, T)$. The bound can be saturated only by the unique optimal drift force. We compare the value of the optimized return to numerically exact estimates of the CGF given as

$$\lambda(s, T) = \ln \left\{ 1 + (e^{-s} - 1) \int_\Gamma d\mathbf{x}\, \rho(\mathbf{x}, T) \right\}, \tag{4.42}$$

where the definition of the indicator function and the final time distribution $\rho(\mathbf{x}, T)$ evolved from a specific initial condition has been used. This form demonstrates the statistics of a single-time indicator observable is described solely by its mean,

$$\langle h_\Gamma \rangle_p = \int_\Gamma d\mathbf{x}\, \rho(\mathbf{x}, T). \tag{4.43}$$

For a rare fluctuation such that $\langle h_\Gamma \rangle_p < 0.5$, this form indicates that there are two distinct regimes in the biased ensemble with $s < 0$. For a small magnitude of the bias, the indicator function stays close to the unbiased value. Below a critical value of $s^* = -\ln[\langle h_\Gamma \rangle_p / (1 - \langle h_\Gamma \rangle_p)]$ the indicator crosses over to being close to 1. For all of our calculations, we choose a fixed value of s estimated to be smaller then the threshold. With this value of s, we compute the right side of Eq. (4.42) using an eigen-expansion of the propagator of the Fokker-Planck equation of the original dynamics, and compare with the value of the average return from the gradient descent algorithms having the same value of s. Details of this calculation and comparison to an approximate Kramers escape rate are in Section 4.5.
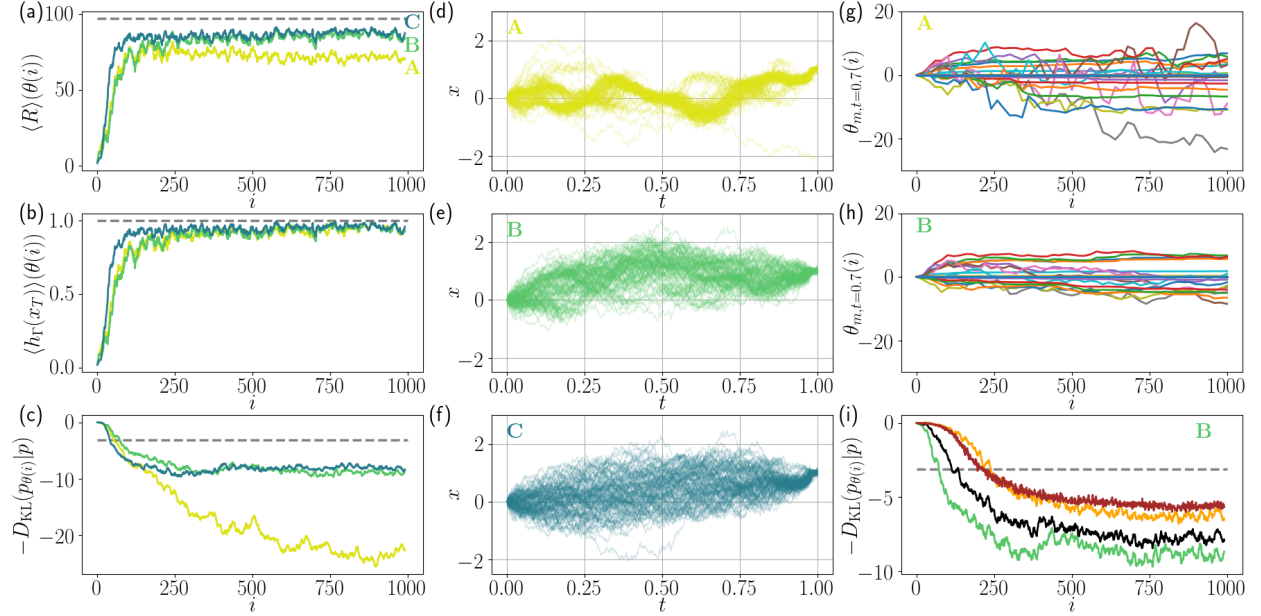
Figure 4.1: **Softened Brownian bridges:** (Left column) Smoothened learning curves showing running estimates of the CGF (a), average value of the indicator observable with the optimized dynamics (b), and the average cost function (c), as functions of optimization steps i, with the MCR('A', yellow), MCVB('B', green) and AC('C',blue) algorithms. The horizontal dashed grey lines denote the numerically exact values. (Middle column) 100 trajectories obtained with the final converged dynamics from the three different algorithms but with the same noise history.(Right column) (g) and (h) show the smoothened convergence of a time-slice of the force parameters, as a function of optimization steps i, in the absence (MCR) and presence (MCVB) of a value function. (i) shows the convergence of the KL divergence cost with finer basis sets optimized with the MCVB algorithm. Green $(31x \times 21t)$, black $(31x \times 41t)$, orange $(31x \times 81t)$ and brown $(41x \times 201t)$ curves show that in the increasing basis limit, the cost-function estimate approaches the value expected from the numerically exact CGF.

Softened Brownian bridges

The first example we consider is a softened version of a so-called Brownian bridge [297, 109], in which a one-dimensional Brownian motion starting from the origin is biased to end near a particular point. The reference dynamics is simply given by free diffusion,

$$dx = \sqrt{2}dW \tag{4.44}$$

where comparing to Eq. 4.1 we have $G = \sqrt{2}$. We consider the target well, $\Gamma(x)$, to be defined as $\{1 - \epsilon \leq x \leq 1 + \epsilon\}$ with $\epsilon = 0.1$. The dynamics is simulated with a discrete

timestep of 0.001. We use a tilting parameter $s = -100$ to bias the original ensemble towards higher occurrence of the rare event.

We optimize a force and value function parametrized by linear combinations of Gaussian distributions with fixed variance and mean. Given a set of means $\{(x_m, t_m)\}_{m=0}^{M}$ and variances $\{\sigma_m\}_{m=0}^{M}$, the force and value function of a position x at time t are given by the coefficients $\{\theta_m\}_{m=0}^{M}$ and $\{\psi_m\}_{m=0}^{M}$ as

$$F_\theta(x, t) = F(x) + \sum_{m=0}^{M} \theta_m e^{-\frac{(x-x_m)^2 + (t-t_m)^2}{2\sigma_m}}$$

$$V_\psi(x, t) = \sum_{m=0}^{M} \psi_m e^{-\frac{(x-x_m)^2 + (t-t_m)^2}{2\sigma_m}}, \tag{4.45}$$

where initially the basis sets are a grid of 31×21 Gaussians in the x-t space. The Gaussians in time are spaced uniformly between $t \in [0, T)$, with standard deviations equal to half the grid-spacing. A third of the Gaussians in space are placed between $x \in [-4, -0.5]$, a third in $x \in (-0.5, 1.5)$ and a third in $x \in [1.5, 5]$. These three families of Gaussians each have standard deviations half of the corresponding grid spacings. We initialize all $\theta_m = \psi_m = 0$.

We consider the performance of the three algorithms differing in the gradient used to optimize them. These include an algorithm that uses no value function (MCR), one that uses a value baseline (MCVB), and one that uses a value function for future returns with $\tau = 0.1$ (AC). We evaluate the efficiency of the algorithms by comparing learning curves, convergence with respect to basis, and properties of the learnt dynamics, shown in Fig 4.1. All figures comparing different algorithms use the same noise history and the same amount of statistics, such that the differences are solely ascribed to the learned dynamics. The MCR algorithm uses a learning rate of $\alpha^\theta = 0.4$. The MCVB algorithm learning rates $\alpha^\theta = 0.4, \alpha^\psi = 50$, and the AC algorithm learning rates $\alpha^\theta = 1, \alpha^\psi = 0.05$.

In Figs. 4.1(a-c), we show learning curves for the total return, the average of the indicator observable, and the KL divergence, generated with 12 trajectories at each optimization step for each of the three algorithms. We have compared the results obtained with this finite basis to the numerically exact value of the optimal return and the corresponding observable average and KL divergence, obtained from Eq. 4.42 where for free diffusion the distribution is known. We find that while all three algorithms quickly achieve a dynamics which mostly fulfills the indicator function conditioning, the MCR algorithm struggles to optimize the KL divergence cost, while the MCVB and AC algorithm achieve converged values efficienctly. As expected, each algorithm provides a variational estimate to the CGF with the MCVB and AC outperforming MCR. Trajectories with the final learned dynamics for the three algorithms are plotted in Fig. 4.1(d-f). The MCR algorithm finds forces that constrain the bridge trajectories too excessively, which results in the suboptimal estimate of the KL divergence. The AC trajectories are closest to the optimal bridge trajectories [109] while the MCVB trajectories lie in between. The main reason for the difference in performance in the three algorithms is the resultant suppression in the statistical errors in the gradient

estimate. This is illustrated in Figs. 4.1(g-h) where the convergence of the gradients of the 31 Gaussian coefficients at a time slice of $t = 0.7$ is shown for both MCR and MCVB. Since the α^θ learning rate is same in both algorithms, the large suppression of fluctuations in the MCVB learning curves results from a more statistically converged gradient estimate using a value function. This suppression of gradient errors at limited statistics in the MCVB and AC algorithms is directly illustrated in Section 4.6.

We have studied the convergence of the KL divergence estimate towards the optimal value extracted from the numerically exact CGF, using the MCVB algorithm with an increasing position and time basis. We increased the number of time Gaussians, from 21 to 41 to 81, to observe the KL divergence cost shrinking as the finer grained force can better support the singular indicator function condition at the end of the trajectory. We also ran the optimization with a much bigger basis of $41x \times 201t$ Gaussians, and used 248 trajectories at every optimization step and learning rates $\alpha^\theta = 5, \alpha^\psi = 1000$. The Gaussians in x have standard deviations equal to half the grid spacing, while the Gaussians in t have standard deviations equal to a third of the grid spacing. While the estimate increased, in this particular problem, obtaining the numerically exact KL divergence would require use of still finer-grained Gaussians in space and time in order to represent the singularities of the edges of the target region and of the last timestep.

Barrier crossing with multiple reaction pathways

We now investigate the ability of the three algorithms to find the optimal dynamics in two-dimensional barrier-crossing problems, the first involving a potential allowing for multiple reaction pathways. The two-dimensional potential $U(\mathbf{x})$ we consider[2] has two minima and two degenerate reaction pathways involving the upper and lower halves of the $\mathbf{x} = (x, y)$ plane as illustrated in Fig. 4.2. Barrier-crossing from one well to another is a rare event occurring with one randomly chosen pathway [72]. Without prior knowledge of the possibility of multiple reaction paths, path sampling algorithms typically need special techniques to discover them [298]. We use our reinforcement learning algorithms to compute an optimal force $F_\theta(\mathbf{x}, t)$ that reproduces unbiased and uncorrelated reaction paths.

The reference equation of motion we consider is

$$d\mathbf{x} = -\nabla U(\mathbf{x}) + \sqrt{2}d\mathbf{W} \tag{4.46}$$

where the matrix $\mathbb{G}$ is proportional to the identity. We use a discretization timestep of 0.001. The trajectories start from the minimum of the left well, at $(x, y) = (-1.11, 0)$, and are allowed to run for a duration of $T = 1.5$ and checked for reaching the right target well defined as $x > 0, U(x, y) < 0$. This small region centered around $(1.11, 0)$ is used as Γ for defining the indicator function observable. The value of T has been chosen to be slightly greater than the typical transition path timescale, such that the optimized force should reproduce trajectories that follow the natural steady-state fluctuations of the system. As long as the choice of

---

[2]The potential we use is $U(x, y) = 4/3[4(1 - x^2 - y^2)^2 + 2(x^2 - 2)^2 + ((x + y)^2 - 1)^2 - ((x - y)^2 - 1)^2 - 2]$

Figure 4.2: **Multiple reaction pathways:** (Left column) Smoothened learning curves showing running estimates of the CGF (a), average value of the indicator observable with the optimized dynamics (b), and the average cost function (c), as functions of optimization steps i, with the MCR(yellow), MCVB(green) and AC(blue) algorithms. The vertical grey lines denote the end of initialization and beginning of optimization run. The horizontal dashed grey lines denote the numerically exact values. The parameter values from the end of the initialization with MCVB and AC have been called B_0 and C_0 respectively. The forces at the end of optimization with AC is called C. (Middle column) 6 representative trajectories obtained with the forces B_0 (panel d), C_0 (panel e), and C (panel f). (Right column) Two-dimensional vectorial representation of the spatially dependent forces as a function of time, at $t = 1$ (g), $t = 1.3$ (h) and $t = 1.5$ (i), obtained from the converged parameters at C.

T is arbitrarily larger than the typical transition path timescale, the optimally generated trajectories will represent unbiased reactive transitions, with additional times being spent in the initial or final metastable states [299]. In the absence of an approximate transition path time estimate, the optimization can be performed over a range of T increasing by orders of magnitude till one enters the regime where side-side correlation functions for the dynamics of barrier crossing behave linearly [72]. We use a value of $s = -500$ to obtain the CGF. The force and the value function are approximated again as a grid of Gaussians with optimizable coefficients, a simple generalization of the one-dimensional Brownian bridge.

The duration of the trajectories we consider, T, is much smaller than the typical first passage time for the rare fluctuation we are interested in studying. As such, a general

complication arises in initializing our algorithms in that in the absence of a modified force, few trajectories satisfy the indicator function condition. Consequently, the gradients for updating the modified forces are generally very small and noisy. In order to initialize our learning process, we start with a softened version of the indicator function of the form

$$\tilde{h}[\mathbf{x}_T] = -[(x_T - x_f)^2 + (y_T - y_f)^2] \tag{4.47}$$

which is quadratic, and non-vanishing across the full domain. After optimizing the return with this observable, we obtain a force that can surpass the barrier, and the optimization with the sharp indicator function observable can begin. This technique of breaking down the optimization of the return into two segments prioritizing each of the two terms of the return is analogous to curriculum learning in reinforcement learning [300]. In many-body systems, the quadratic metric can be defined only in the space of the order parameter that distinguishes the initial and product states. For our multi-channel problem, we initialize learning with $(x_f, y_f) = (1.11, 0)$ in the softened indicator, which is the minimum of the target well. Our approach consists of comparing the performance of the three algorithms MCR, MCVB and AC in the initialization with the quadratic observable, and then using the AC algorithm to optimize the return with the indicator function observable.

Figures 4.2(a-c) demonstrate the learning curves for the full return, the average of the indicator function and the KL divergence cost. The three initializations each use 60 trajectories at every optimization step. The basis functions for the force and value function used are a grid of $21 \times 21 \times 41$ Gaussians in the $\mathbf{x} - t$ space for each component independently. The Gaussians are placed uniformly on the time axis $t \in [0, T)$, while the position Gaussians are distributed uniformly between $x \in [-1.5, 1.5]$ and $y \in [-1.5, 1.5]$. The learning rates used in the initialization are $\alpha^\theta = 1$ for MCR, $\alpha^\theta = 1, \alpha^\psi = 0.5$ for MCVB and $\alpha^\theta = 1, \alpha^\psi = 0.5, \tau = 0.001$ for AC, and the learning rate for the final optimization is $\alpha^\theta = 0.2, \alpha^\psi = 0.08, \tau = 0.1$ in the AC algorithm. In the learning curves, we compare the convergence of the return with numerically exact values obtained by computing the RHS in Eq. 4.42 with a spectral expansion using a Discrete Variable Representation basis [301]. We see that all three algorithms quickly find forces that satisfy the conditioning, but the KL divergence cost is optimized best by the AC algorithm. While each affords a similar variational estimate after the initial optimization, we find qualitative differences in the family of barrier-crossing trajectories obtained from the MCR/MCVB and from the AC algorithm.

Typical trajectories obtained with forces from the end of initialization with MCVB and AC, and at the end of optimization with AC, are shown in Figs. 4.2(d-f). The force obtained from MCVB spontaneously breaks the symmetry in the potential and chooses one reaction path out of the two. This force solution is a local optimum in the MCR and MCVB algorithms, and it does not naturally relax to a symmetric force that would be representative of the degeneracy of the reaction paths. Trajectories from the AC algorithm spend significant amount of time exploring the initial well, such that the discovered forces recognize the presence of multiple pathways approximately. These forces are further refined during the second optimization, such that the reactive trajectories obtained at the end are restored to be almost
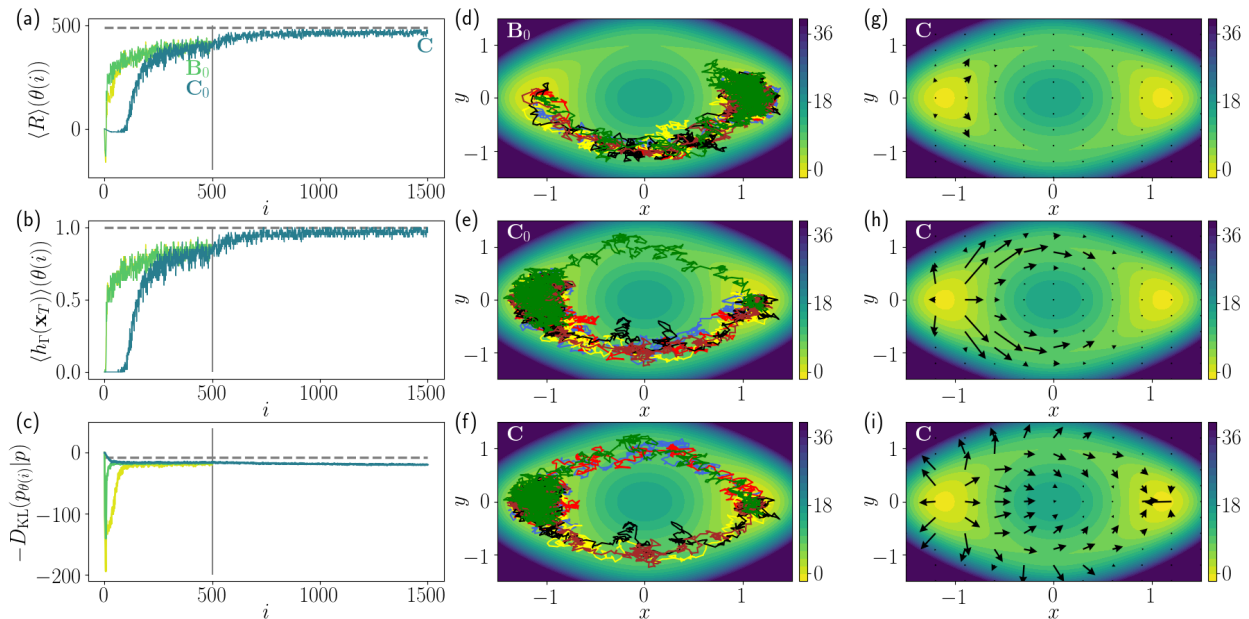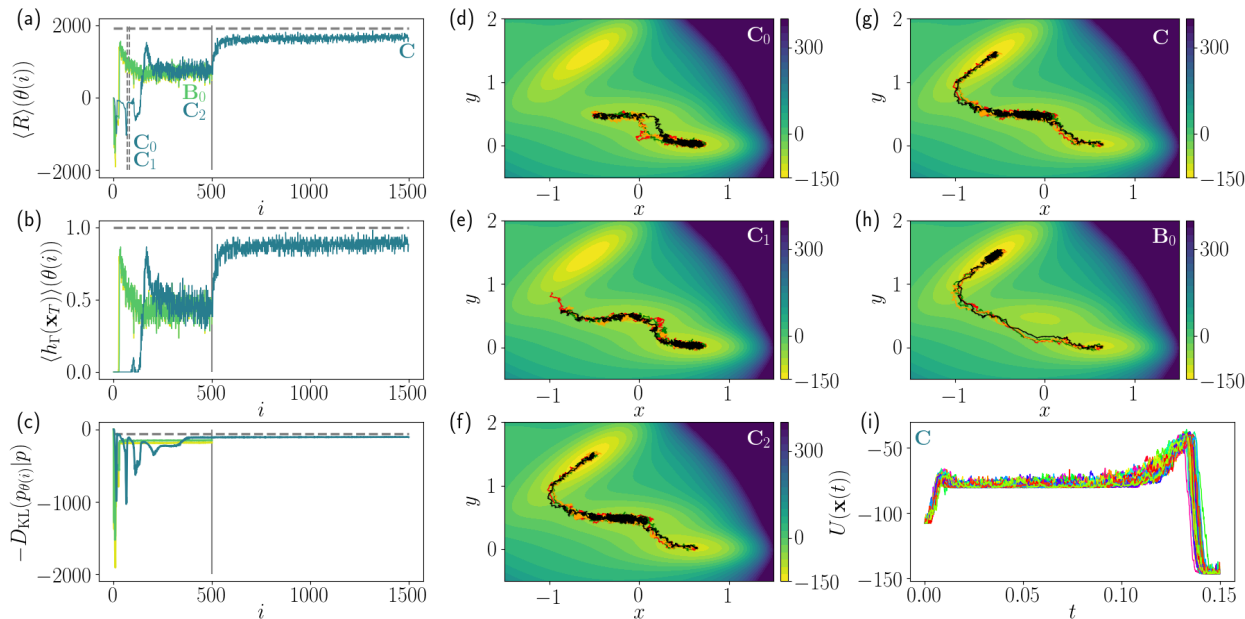
Figure 4.3: **Müller-Brown potential:** (Left column) Smoothened learning curves showing running estimates of the CGF (a), average value of the indicator observable with the optimized dynamics (b), and the average cost function (c), as functions of optimization steps i, with the MCR(yellow), MCVB(green) and AC(blue) algorithms. The vertical grey lines denote the end of initialization and beginning of optimization run. The horizontal dashed grey lines denote the approximate values from a Kramer's escape rate approximation. On the AC learning curve in (a), the parameter values at $i = 70$ and $i = 80$ (the vertical dashed lines) have been called C_0 and C_1 respectively. The values at the end of initialization with MCVB and AC are called B_0 and C_2, and at the end of AC optimization are called C. (Middle column) 4 representative trajectories obtained with the forces C_0(d), C_1(e), and C_2(f). (Right column) 4 representative trajectories obtained with the forces C(g) and B_0(h). (i) Potential energy as a function of time for 100 representative trajectories driven with the force parameters C.

fully symmetric like the natural barrier-crossing fluctuations of the system are expected to be. These symmetric two-dimensional forces obtained at the end of the AC optimization are plotted at three slices of time, in Figs. 4.2(g-i). The forces grow in magnitude as a function of time and generally follow the contours of the underlying potential, and towards the end they gather support in unlikely parts of the potential. The ability of the AC algorithm to discover time-dependent forces that lead to exploration of multiple reaction pathways can prove valuable in uncovering reactive trajectories in systems where such degeneracies are not known *a priori*.

Barrier crossing with a long lived intermediate

Another difficult problem in the generation of transition paths and reactive trajectories typically comes from the presence of long-lived intermediates. In order to study the usefulness of our learning algorithms in this context, we consider as an example the dynamics on the so-called Müller-Brown potential [302]. This two-dimensional potential surface has been used extensively as a testing case for methods relying on the instantonic approximation for barrier-crossing trajectories [299, 303]. The potential is a sum of four Gaussians[3], where three local minima are separated by two barriers as illustrated in Fig. 4.3. We employed our algorithms to find forces that generate uncorrelated trajectories that cross both barriers, starting from a local minimum and ending in the global minimum, that are positioned on either side of the third metastable minimum.

The system evolves with diffusive Langevin dynamics of the same form as Eq. 4.46 using a timestep of 0.0001. We are interested in trajectories starting from $\mathbf{x} = (0.63, 0.03)$ in the rightmost local minumum, and ending near the global minimum, centered around $\mathbf{x} = (-0.5, 1.5)$, with the indicator function region Γ being defined by $U(\mathbf{x}) < 145$). The trajectories are chosen to be of a fixed duration of $T = 0.15$, which is on the order of the expected total transition path timescale from Kramers' theory added to the expected relaxation time in the intermediate well [304, 299]. For initializing the forces we use a softened quadratic modification of the indicator, in Eq. 4.47, with $s = -10000$, while we use a bias value of $s = -2000$ with the indicator observable to compute the CGF. To represent the x and y components independently of the time-dependent optimal force and to represent the value function, we use a basis of Gaussians with optimizable coefficients placed on a $21 \times 21 \times 21$ grid in $\mathbf{x} - t$. The time Gaussians are placed uniformly between $t \in [0, T)$, while the space Gaussians placed uniformly between $x \in [-1.5, 1.5]$ and $y \in [-0.5, 2]$.

In Figs. 4.3(a-c), we have compared the learning curves with MCR, MCVB and AC algorithms during initialization with the smooth indicator function in Eq. 4.47 and the AC algorithm for the final optimization of the full return with the sharp indicator function. Each algorithm uses 60 trajectories at every optimization step to estimate the gradient. The learning rates for the initialization are $\alpha^\theta = 1$ for MCR, $\alpha^\theta = 1, \alpha^\psi = 1$ for MCVB, and $\alpha^\theta = 0.5, \alpha^\psi = 0.2, \tau = 0.0001$ for AC, and the learning rates for the final optimization are $\alpha^\theta = 0.1, \alpha^\psi = 0.01, \tau = 0.01$ for AC. The learning curves have been compared with approximately calculated values of the CGF and the KL div obtained with a Kramer's escape rate estimate along the Minimum Energy Path [293].

We find that all three algorithms optimize the quadratic observable relatively quickly, but the AC algorithm performs the best at optimizing the KL divergence cost. In Figs. 4.3(d-h), we illustrate a few uncorrelated trajectories generated with the modified forces at various stages of the initialization and optimization with the AC method and the end of the initialization with the MCVB method. We find that the forces with the AC algorithm are such

---

[3]The potential takes the form, $U(x, y) = \sum_i A_i \exp[a_i(x - \bar{x}_i)^2 + b_i(x - \bar{x}_i)(y - \bar{y}_i) + c_i(y - \bar{y}_i)^2]$ where $A = (-200, -100, -170, 15)$, $a = (-1, -1, -6.5, 0.7)$, $b = (0, 0, 11, 0.6)$, $c = (-10, -10, -6.5, 0.7)$, $\bar{x} = (1, 0, -0.5, -1)$, and $\bar{y} = (0, 0.5, 1.5, 1)$.

that the trajectories discover and cross the two barriers and the metastable well between them one after another. At the end of the AC initialization, the trajectories have discovered the metastable well and have crossed both barriers to end in the target well. The AC algorithm by this stage of optimization has also moved the major part of the short trajectory from staying in the initial well to the metastable well. This feature is constant throughout the AC optimization, with only minor changes in the force being carried out inside the target end well. The force from the MCVB initialization, on the other hand, only generates trajectories that connect the initial and target well without relaxing significantly in the metastable well. This would be contrary to the instantonic relaxation mechanism in the system, as the stochastic action is minimized by the local relaxation in the metastable well. In Fig. 4.3(i) we have plotted the potential energy as a function of time, for 100 uncorrelated barrier-crossing trajectories, which are driven by the final force from the AC algorithm. The trajectories cross the two barriers at roughly fixed times, and spend majority of the time in the metastable well.

The comparison of the three algorithms illustrates the significant improvement of convergence performance of the MCVB and AC algorithm over the naive MCR approach afforded by value functions. For rare reactive events, we have found that the AC algorithm is suited best to find trajectories that explore configuration space the most in search for the easier barriers to cross, and thus is closest in resembling the natural fluctuations of the system. The errors in the converged values of the CGF depend on the truncation of the force basis and statistical uncertainties. The MCVB and AC algorithms preserve the computational scaling of the MCR with the trajectory duration, and only change the prefactors of the scaling by a small fraction making them viable methods for applications to complex systems. The AC algorithm with a small τ will incur a systematic error in the gradients if the value approximation is not accurate, which goes away at an intermediate τ but at the expense of a larger memory cost that may slow down the algorithm without any change in the scaling. Nevertheless, it is possible to use these algorithms with useful combinations of hyperparameters to achieve efficient convergence with a small amount of averaging. The value functions obtained during the optimizations serve as dynamical equivalents of the committor function, in that they encode the expected value of the probability to reach the target well and the associated KL divergence cost, while starting from any point in configuration space at any point in time. Understanding these connections to reaction coordinate design is likely a fruitful future direction of research.

4.5 Alternative CGF estimates

Numerically exact CGF

We have compared the CGF from the reinforcement learning algorithms in Section 4.4 with numerically exact values obtained from explicitly calculating $\langle h_\Gamma \rangle_p$ in equation (4.43) by solving the corresponding Fokker-Planck operator. The Fokker-Planck operator for the original

dynamics in Eq. (4.46) is given by

$$L = -\nabla.\mathbf{F}(\mathbf{x}) + \nabla^2 \tag{4.48}$$

where $\mathbf{F}(\mathbf{x}) = -\nabla U(x)$ is the underlying conservative force.

We want to use this operator in order to find the probability $\langle h_\Gamma \rangle_p$ as

$$\langle h_\Gamma \rangle_p = \int_\Gamma d\mathbf{x}\, \rho(\mathbf{x}, T) = \int_\Gamma d\mathbf{x}\, e^{LT} \delta(\mathbf{x} - \mathbf{x}_0) \tag{4.49}$$

We exponentiate the operator in its spectral eigenbasis. Since the forces in the original dynamics are conservative, diagonalizing L becomes easier through a similarity transform into a Hermitian operator $\mathcal{L}$ [43, 109],

$$\mathcal{L} = e^{U(\mathbf{x})/2} L e^{-U(\mathbf{x})/2}$$
$$= \nabla^2 - \frac{1}{4}(\nabla U(\mathbf{x}))^2 + \frac{1}{2}\nabla^2 U(\mathbf{x}). \tag{4.50}$$

We diagonalize $\mathcal{L}$ to obtain eigenvalues $-\lambda_n$ and eigenfunctions $\phi_n(\mathbf{x})$,

$$\mathcal{L}\phi_n(\mathbf{x}) = -\lambda_n \phi_n(\mathbf{x}). \tag{4.51}$$

Since $\mathcal{L}$ is Hermitian, the eigenfunctions $\{\phi_n(\mathbf{x})\}$ are mutually orthonormal and can be used to introduce a resolution of identity

$$\delta(\mathbf{x} - \mathbf{x}_0) = \sum_n \phi_n(\mathbf{x}_0)\phi_n(\mathbf{x}) \tag{4.52}$$

The original operator L related by the similarity transform has eigenvalues $-\lambda_n$ and eigenfunctions $e^{-U(\mathbf{x})/2}\phi_n(\mathbf{x})$. This spectral expansion of L can be used to estimate the probability $\langle h_\Gamma \rangle_p$ as

$$\langle h_\Gamma \rangle_p = \int_\Gamma d\mathbf{x}\, e^{LT} \delta(\mathbf{x} - \mathbf{x}_0)$$
$$= e^{U(\mathbf{x}_0)/2} \sum_n e^{-\lambda_n T} \int_\Gamma d\mathbf{x}\, e^{-U(\mathbf{x})/2}\phi_n(\mathbf{x}) \tag{4.53}$$

The final time T that we use in our barrier-crossing simulations is chosen such that $\tau_{\text{rlx}} < T < \tau_{\text{rxn}}$ where τ_{rlx} and τ_{rxn} are respectively the timescale of relaxation in the starting or the ending well, and the timescale of the barrier-crossing reaction, which is expected to be the slowest dynamical mode in the system. Hence when the set $\{\lambda_n\}$ is ordered, the factor $e^{-\lambda_n T}$ should be negligible for all but the few smallest values of n. The sum over n in Equation (4.53) is thus expected to converge within a few terms.

We diagonalize the operator $\mathcal{L}$ using a Discrete Variable Representation basis constructed from Hermite polynomials [301] in two dimensions, $\chi_{M,N}(\alpha x, \alpha y)$, where $\alpha = 5$ is a scaling factor. We obtain identically converged estimates of $\langle h_\Gamma \rangle_p$ with basis sizes ranging from 50×50 to 100×100 using 10 terms in the spectral expansion. The CGF value is then calculated using $\langle h_\Gamma \rangle_p$ in Equations (4.42) and (4.43).

CGF from Kramers escape rate

In one-dimension, corresponding to a dynamics of

$$dq = -U^{'}(q) + \sqrt{2}dW, \tag{4.54}$$

an approximate expression for the barrier-crossing probability in time T is given by the Kramers escape rate in the overdamped limit [11], as

$$\langle h_\Gamma \rangle_p \approx \frac{T}{2\pi}(U^{''}(q_A)|U^{''}(q^\dagger)|)^{1/2}e^{-(U(q^\dagger)-U(q_A))} \tag{4.55}$$

where q is the reaction coordinate and q_A and $q^\dagger$ are the locations of the initial well and the barrier respectively.

In the case of the Müller-Brown potential, we assume the ideal reaction coordinate to be along the Minimum-Energy Path obtained using a Nudged Elastic Band method [293, 294, 295]. With the potential energy $U(q)$ computed along this path q, we use quadratic fits around the initial well (q_A) and around the largest barrier $(q^\dagger)$ to find the double-derivative terms. Finally we use this approximate value of $\langle h_\Gamma \rangle_p$ in Equation (4.42) and (4.43) to obtain the CGF.

4.6 Errors in gradient estimates

In Figure 4.4 we have directly compared the three algorithms for their ability to reduce the variance of the gradient estimates during optimization in the softened Brownian bridge problem. We have chosen the force and value function coefficients θ and ψ from the $i = 100$ step of the MCVB optimization run in Fig. 4.1(b) in the Brownian bridge problem. This value function is thus not exact for the corresponding force but is representative of typical inaccuracies encountered during learning. Keeping these coefficients fixed, we have estimated the gradients of the KL divergence using the three algorithms, while varying the number of uncorrelated trajectories N_w over which the estimates are averaged. Plotted in Fig. 4.4 are the total variance in the gradient estimate summed over all components, $\sum_m \text{Var}[\nabla_{\theta_m} D_{\text{KL}}(p_\theta|p_s)]$, from the different algorithms. The variances are computed from fluctuations over 10 uncorrelated sets of N_w trajectories. The dependence on N_w in log-log scale corresponds to a linear trend with a slope of -1 as expected from the variance of sample means of uncorrelated samples. We find that use of the MCVB and AC algorithms greatly reduces the variance compared to the MCR approach, equivalent to a 5 to 100 times increase in the amount of input trajectory data. We find that the smallest variance corresponds to the AC algorithm with the smallest possible τ, set to the timestep 0.001. However, this choice incurs a systematic error in the expectation of the gradient due to the inaccuracy in the value function, while neither MCVB nor AC with a large τ are susceptible to it. This is manifested in the scaled L^1 norm of the error in the expected gradient from the algorithms. The expectation is calculated over 10^5 trajectories and the error in MCR is zero by definition.
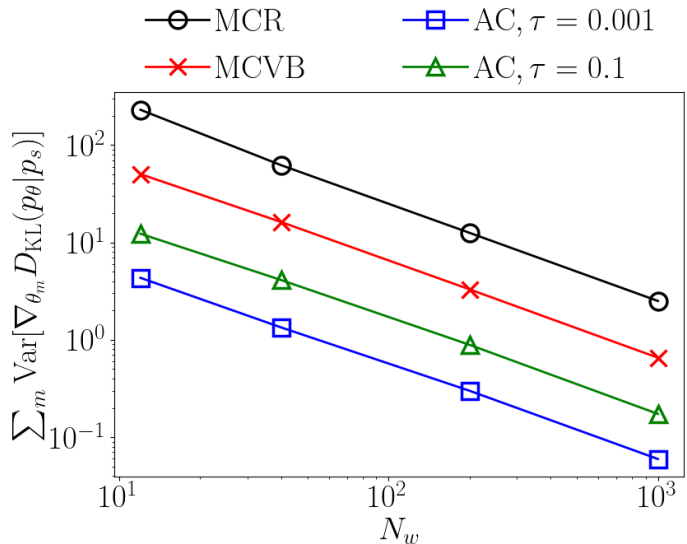
Figure 4.4: **Statistical convergence of gradient estimates** Total variance of the gradient summed over all components, using MCR(black), MCVB(red), AC with $\tau = 0.001$(blue) and AC with $\tau = 0.1$(green), as a function of the number of uncorrelated trajectories N_w for averaging.

The L^1 norms of the errors, divided by that of the true gradient, are 0.22, 7.49 and 1.16 from MCVB, AC($\tau = 0.001$) and AC($\tau = 0.1$) respectively. This shows that the systematic error incurred by AC at small τ can be reduced by having a larger τ, while still having significantly less variance than MCVB and MCR. The crossover between the systematic and statistical error in the AC algorithm depending on τ is also the reason starting the optimization with a small τ and later annealing with a large τ is an efficient strategy, given that the memory requirement scales linearly with τ. We note that the systematic error is formally zero by definition in the expectation of the MCVB gradient estimate as well: the small non-zero value stems from a finite number of samples being used to estimate the expectation.

4.7   Gradient optimization for infinite time dynamics

We now generalize the approach of the previous section to focus on the statistics of time-integrated quantities in the long time limit. While for finite time, the generalized Doob transform is time dependent, under mild assumptions in the long time limit the optimal dynamics is time-homogeneous [63]. As a consequence, the parametrization of the modified force and value function is simplified, and explicitly dependent only on the instantaneous configuration of the system. The generalization of the algorithms to this case consists of two main changes. First, we employ online learning, since there is no end to each trajectory.

Second, a modified definition of return and value are required to avoid divergences in the infinite time limit.

We formulate the infinite time problem by adapting an approach in reinforcement learning based on time-averaged returns [260, 305, 306, 307]. Specifically, we consider the long-time average of the KL divergence of the trajectory ensemble. Under assumptions of time-independence and ergodicity,

$$
\begin{aligned}
d_{\mathrm{KL}}(p_\theta|p_s) &= \lim_{T\to\infty} \frac{1}{T} D_{\mathrm{KL}}(p_\theta|p_s) \\
&= -\langle r(\mathbf{x}, \dot{\mathbf{x}})\rangle_{p_\theta} + \lambda(s),
\end{aligned}
\tag{4.56}
$$

the time average KL divergence reduces to an average over the steady state distribution of the instantaneous change of the return $r(\mathbf{x}, \dot{\mathbf{x}})$. Above, we have defined a scaled CGF,

$$
\lambda(s) = \lim_{T\to\infty} \frac{1}{T} \ln Z(s, T),
\tag{4.57}
$$

that is finite as long as the cumulants of the time-integrated observable are time extensive. The reward, $r(\mathbf{x}, \dot{\mathbf{x}})$, is defined as

$$
\begin{aligned}
r(\dot{\mathbf{x}}, \mathbf{x}) = {}& -sA[\mathbf{x}] - s\mathbf{B}[\mathbf{x}] \cdot \dot{\mathbf{x}} \\
& + \frac{1}{2} \left\{ |\mathbb{G}^{-1} \cdot (\dot{\mathbf{x}} - \mathbf{F}_\theta)|^2 - |\mathbb{G}^{-1} \cdot (\dot{\mathbf{x}} - \mathbf{F})|^2 \right\}
\end{aligned}
\tag{4.58}
$$

and is time-independent and evaluatable within the steady state. A gradient expression analogous to MCR can be derived straightforwardly [117].

The previous definition of the value will diverge in the infinite time limit. A simple modification to address this issue is to remove the average reward scaled by the length of the trajectory segment, defining a differential return

$$
\Delta R\left[\mathbf{X}_{t,t'}\right] = R\left[\mathbf{X}_{t,t'}\right] - (t' - t)\langle r(\dot{\mathbf{x}}, \mathbf{x})\rangle_{p_\theta}
\tag{4.59}
$$

and corresponding differential value function

$$
V(\mathbf{x}) = \lim_{T\to\infty} \langle \Delta R\left[\mathbf{X}_{0,T}\right]\rangle_{p_\theta,\mathbf{x}}.
\tag{4.60}
$$

which satisfies a modified Bellman equation

$$
V(\mathbf{x}) = \langle V(\mathbf{x}_\tau) + \Delta R\left[\mathbf{X}_{0,\tau}\right]\rangle_{p_\theta,\mathbf{x}},
\tag{4.61}
$$

containing the differential return between states, rather than the standard return, and relating the value of states separated by a period of time τ.

This modified Bellman equation can be simply rearranged to give an alternative equation for our time-averaged KL divergence

$$
\begin{aligned}
d_{\mathrm{KL}}(p_\theta|p_s) = {}& - \frac{1}{\tau} \langle V(\mathbf{x}_\tau) + R\left[\mathbf{X}_{0,\tau}\right] - V(\mathbf{x})\rangle_{p_\theta,\mathbf{x}} \\
& + \lambda(s),
\end{aligned}
\tag{4.62}
$$

which we note holds for all $\mathbf{x}$. Differentiating the right side of this equation with respect to θ does not involve the gradient of the stationary state. Therefore, taking the derivative and then averaging over the stationary state under F_θ,[4] we can write an estimate of the dynamical gradient as

$$\nabla_\theta d_{\mathrm{KL}}(p_\theta|p_s) = -\frac{1}{\tau} \left\langle \delta \left[\mathbf{X}_{0,\tau'}\right] y_\theta(\tau) \right\rangle_{p_\theta}, \qquad (4.63)$$

where we have defined the differential temporal difference error

$$\delta \left[\mathbf{X}_{0,\tau'}\right] = V\left(\mathbf{x}_{\tau'}\right) + \Delta R \left[\mathbf{X}_{0,\tau'}\right] - V\left(\mathbf{x}_0\right), \qquad (4.64)$$

reached after introducing an additional baseline in the form of $\tau' \langle r(\dot{\mathbf{x}}, \mathbf{x})\rangle_{p_\theta}$. In this equation we have arrived at a gradient estimate which depends only on the gradient of the transition probabilities, contained in the Malliavin weights $y_\theta(\tau)$, and not the gradient of the stationary state itself. This can thus be easily calculated during a simulation using the parametrized dynamics.

Note the period of time τ' over which the temporal difference is calculated is independent of the period of time τ over which the Malliavin weight is calculated, provided the former is longer. The specific algorithm we consider involves taking the time τ small enough so that the Malliavin weight can be approximated by $\tau \dot{y}_\theta[\mathbf{x}_0]$ which is possible due to the time homogeneous steady state we average within. We thus calculate the estimate as

$$\begin{aligned} \nabla_\theta d_{\mathrm{KL}}(p_\theta|p_s) &= -\left\langle \delta \left[\mathbf{X}_{0,\tau'}\right] \dot{y}_\theta(0) \right\rangle_{p_\theta} \\ &= \chi_{\mathrm{AC}}(\theta) \end{aligned} \qquad (4.65)$$

which we denote as the actor-critic gradient in the long time limit. In practice, we will take $\tau' = \Delta t$, a single time-step in a numerical simulation. A long time limit generalization of the MCVB gradient could be constructed similarly, but this is not considered here.

As in the finite time case, to construct this estimate we also need an approximation to the value function, $V_\psi(\mathbf{x})$. Following a similar construction for the loss function as before, averaging the error over the stationary state, we estimate the gradient by which to update the value function parameters as

$$\nabla_\psi L(\psi) = -\left\langle \delta_\psi \left[\mathbf{X}_{0,\tau'}\right] \nabla_\psi V_\psi\left(\mathbf{x}_0\right) \right\rangle_{p_\theta}, \qquad (4.66)$$

with the approximate temporal difference

$$\delta_\psi \left[\mathbf{X}_{0,\tau'}\right] = V_\psi\left(\mathbf{x}_{\tau'}\right) + \Delta R \left[\mathbf{X}_{0,\tau'}\right] - V_\psi\left(\mathbf{x}_0\right), \qquad (4.67)$$

which also replaces the exact temporal difference in gradient estimates for the dynamics. Finally, we also have flexibility with our estimate of the scaled CGF. This can be done using a running average of the reward,

$$\langle r\rangle_{p_{\theta_i}} = \langle r\rangle_{p_{\theta_{i-1}}} + \alpha_r(\langle r\rangle_{p_{\theta_i}} - \langle r\rangle_{p_{\theta_{i-1}}}) \qquad (4.68)$$

---

[4]Taking this derivative results in gradients of the value function at x and x_T with respect to θ, however, these cancel out when averaging over the stationary state.

where α_r is the learning rate and the subscript p_{θ_i} denotes the parameters from the ith iteration. Alternatively, a lower variance, higher bias estimate may be constructed by noting that we can rearrange Eq. 4.61 to find

$$\langle r \rangle_{p_{\theta_i}} = \langle r \rangle_{p_{\theta_{i-1}}} + \alpha_r \langle \delta_\psi [\mathbf{X}_{0,\tau'}] \rangle_{p_{\theta_i}}, \tag{4.69}$$

an alternative equation for the average. After discretization, an algorithm based on utilising single-transition estimates of these gradients is outlined in pseudocode below in Alg. 6.

Algorithm 6 KL regularized differential actor-critic

1: **inputs** force approximation $\mathbf{F}_\theta(\mathbf{x})$, value approximation $V_\psi(\mathbf{x})$
2: **parameters** learning rates α_i^θ, α_i^ψ, α_i^R; total updates N
3: **initialize** choose initial weights θ and ψ, initial average $\bar{r}$, define iteration variable i, individual error δ
4: $i \leftarrow 0$
5: **repeat**
6:     Generate a transition from $\mathbf{x}$ to $\mathbf{x}'$ according to the dynamics given by $\mathbf{F}_\theta(\mathbf{x})$ and noise vector $\mathbf{w} \sim \mathcal{N}(0,1)$
7:     $\dot{y}_\theta = \frac{\mathbf{w}^T \mathbb{G}^{-1} \nabla_\theta \mathbf{F}_\theta}{\sqrt{\Delta t}}$
8:     $\delta \leftarrow V_\psi(\mathbf{x}') + r(\mathbf{x},\mathbf{x}') - \bar{r} - V_\psi(\mathbf{x})$
9:     $\theta \leftarrow \theta + \alpha_i^\theta \delta \dot{y}_\theta$
10:    $\psi \leftarrow \psi + \alpha_i^\psi \delta \nabla_\psi V_\psi(\mathbf{x})$
11:    $\bar{r} \leftarrow \bar{r} + \alpha_i^R \delta$
12:    $i \leftarrow i + 1$
13: **until** $i = N$

4.8 Rare fluctuations in the long time limit

Here we apply our approach to study the statistics of time-integrated currents in the long time limit. Persistent currents are the hallmark of a nonequilibrium system, and their fluctuations have been studied intensively [27, 308, 309, 199]. Foundational results have been derived that constrain the symmetries of current fluctuations and relate their cumulants. For example, the fluctuation theorems dictate that the CGF satisfies a reflection symmetry about the driving force for the current, due to the microscopic reversibility of the underlying stochastic dynamics [310, 37]. A number of numerical approaches have been developed to evaluate the scaled cumulant generating function, an example of a large deviation function [34, 70, 311, 312, 77, 313, 314]. These functions provide information of the long time behavior of stochastic systems, and encode response relationships and stability. Within this context, our approach is similar to other controlled dynamics [89, 86, 315, 88, 273, 84, 93, 117] based means of evaluating large deviation functions in the continuum and can be used directly as we show below or in concert with Monte Carlo algorithms.

Figure 4.5: **Overdamped current fluctuations:** (a) Learning curves showing running estimates of the SCGF, (b) time-averaged KL divergence to the original dynamics $d_{\mathrm{KL}}(p_{\theta(s,i)}|p)$ during training for bias s at step i, and (c) the time-averaged velocity. The color of each curve indicates the value of of the bias s, corresponding with the colors of the data points in the lower plots. Estimates of (d) the SCGF, (e) time-averages KL divergence with the original dynamics and (f) time-averaged velocity, for the final dynamics found at each value of the bias s indicated on the x axis. The inset of (d) shows the absolute error with numerical diagonalization results, represented by grey circles in (d).

To study the accuracy and efficiency of the algorithm, we consider statistics of the velocity of a particle on a ring of length $L = 2\pi$ with position x moving in a periodic potential. The periodic potential has the form $U(x) = U_0 \cos(x)$ with magnitude U_0, and is driven by a constant force f, such that

$$F(x) = -\frac{dU(x)}{dx} + f \tag{4.70}$$

is the total force for the particle on the ring. The observable we consider is the integrated current, $O[\mathbf{X}_{0,T}] = J[\mathbf{X}_{0,T}]$ given by

$$J[\mathbf{X}_{0,T}] = \int_0^T dt\, \dot{x}(t)\,. \tag{4.71}$$

This observable has a different interpretation depending on whether the dynamics are under- or overdamped, both of which we consider below. In the underdamped case, the current is simply a function of the state with $A(\mathbf{x}) = v$ and $B = 0$, while in the overdamped case it depends on the stochastic increment, $A(\mathbf{x}) = 0$, $B(\mathbf{x}) = 1$.

The corresponding scaled CGF we aim to compute is

$$\lambda(s) = \lim_{T \to \infty} \frac{1}{T} \ln \left\langle e^{-sJ[\mathbf{X}_{0,T}]} \right\rangle_p\,. \tag{4.72}$$

The first derivative of $\lambda(s)$

$$v(s) = -\frac{d\lambda(s)}{ds} \tag{4.73}$$

reports on the average velocity in the tilted ensemble and is a useful indicator of the tails of the reference distribution. The scaled CGF exhibits a Lebowitz-Spohn symmetry [310] such that

$$\lambda(s) = \lambda(-f - s) \tag{4.74}$$

where f is the affinity for the current. The scaled CGF can be computed by the numerical solution of a generalized eigenvalue problem [117] which we use for this low dimensional system to compare the accuracy of our results.

Despite its simplicity this system has been shown to present non-trivial non-equilibrium phenomena due to the competition between ballistic and diffusive motion [314, 129, 130]. Here, the overdamped regime acts as a simple benchmark which can be easily solved by diagonalizating a projection of the Fokker-Planck equation [130]. The underdamped regime is a much more difficult problem to solve, due to a higher dimensional state space and long relaxation time. Indeed, despite access to the SCGF via diagonalization [130], accurate results for the force in the underdamped case have been elusive. However, the actor-critic approach can solve this problem easily.

Current fluctuations of an overdamped particle

In the overdamped case, the evolution equation for the particle on a ring is given by

$$dx = F(x)dt + \sqrt{2}dW \tag{4.75}$$

which is a dimensionless one-dimensional SDE. We integrate this equation with a timestep of 0.001. Since the position is periodic, an ideal representation of both the force and value function is given by a Fourier series

$$F_\theta(x) = F(x) + a^\theta + \sum_{i=1}^{M} b_i^\theta \sin(ix) + c_i^\theta \cos(ix), \tag{4.76}$$

and

$$V_\psi(x) = a^\psi + \sum_{i=1}^{M} b_i^\psi \sin(ix) + c_i^\psi \cos(ix), \tag{4.77}$$

with coefficients a^θ, a^ψ, $\{b_i^\theta, c_i^\theta\}_{i=1}^{M}$ and $\{b_i^\psi, c_i^\psi\}_{i=1}^{M}$ truncated to dimension M.

The results of the differential AC algorithm are shown in Fig. 4.5. We have truncated the basis with $M = 5$ and used learning rates of $\alpha_\theta = 0.1$ and $\alpha_\psi = 0.01$. We annealed across the range s considered, first learning the dynamics at $s = -0.5$, before sweeping across to $s = 1.5$ in steps of $\Delta s = 0.1$. The reward learning rate began at $\alpha_R = 10^{-5}$ and decreased
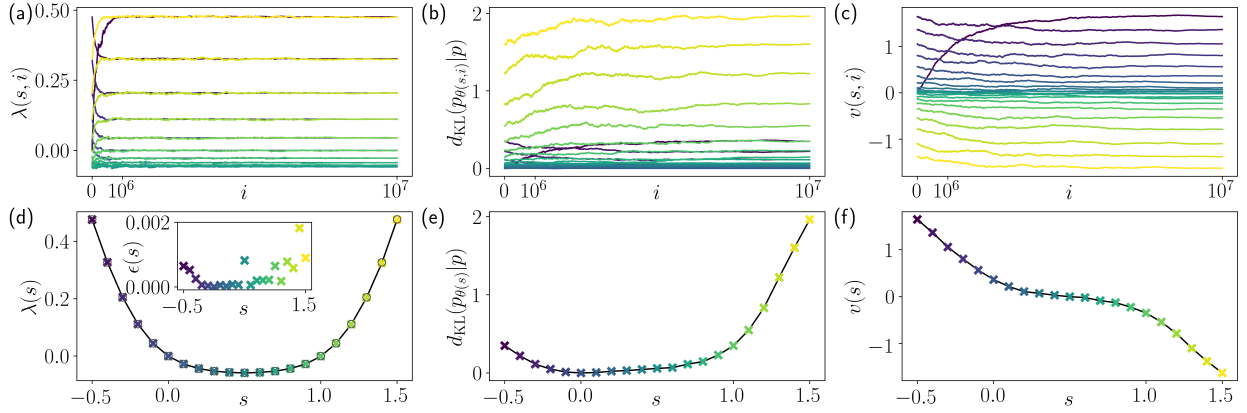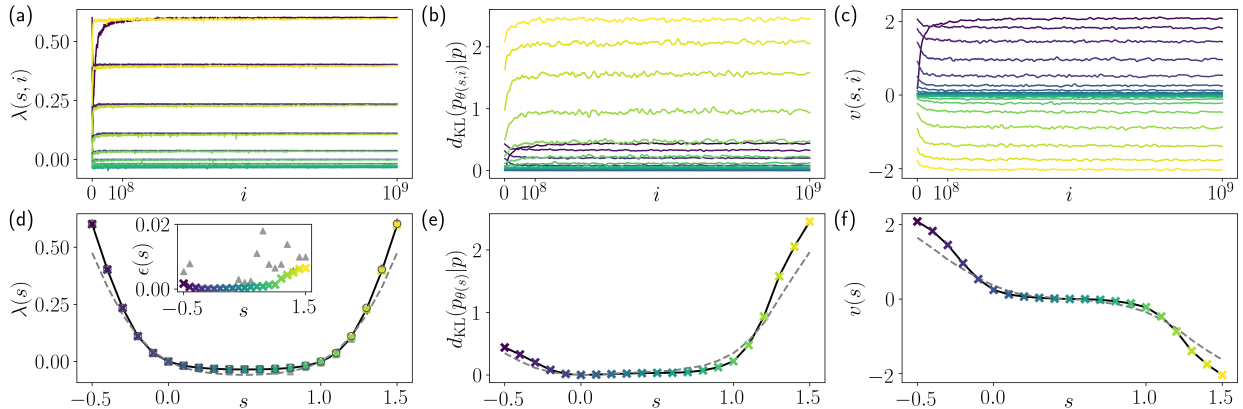
Figure 4.6: **Underdamped current fluctuations:** (a) Learning curves showing running estimates of the SCGF, (b) time-averaged KL divergence to the original dynamics $d_{\mathrm{KL}}(p_{\theta(s,i)}|p)$ during training for bias s at step i, and (c) the time-averaged velocity, calculated as the dynamics is trained. The color of each curve indicates the value of of the bias s, corresponding with the colors of the data points in the lower plots. Estimates of (d) the SCGF, (e) time-averages KL divergence with the original dynamics and (f) time-averaged velocity of the final dynamics for each value of the bias s indicated on the x axis. The inset of (d) shows the absolute error with numerical diagonalization results, represented by grey circles in (d). Results with estimated corrections using the algorithm in Ref.[117] are shown as triangles in (d) and its inset. Dashed curves in (d-f) show the results for the overdamped case for comparison.

linearly to $\alpha_R = 10^{-6}$ throughout training at each value of s, to enable rapid convergence to an accurate result.

We detail estimates of three quantities calculated during the learning process. In Fig. 4.5(a) we show the estimate of $\lambda(s)$, the quantity the algorithm is attempting to maximize. In Fig. 4.5(b) we show an estimate of the time-averaged KL divergence. In Fig. 4.5(c) we show an estimate of the time-averaged velocity. These estimates are running averages calculated using the samples taken from each transition, with learning rates of $0.1\alpha_R$. Learning curves are plotted for training at each individual bias s during the annealing process. For small changes of s, we see that convergence to an accurate estimate of the scale CGF is achieved in approximately 10^6 training steps, each utilizing data from a single transition. This results in a speed of up to two orders of magnitude over the MCR algorithm [117].

In Figs. 4.5(d-f) we plot the end points of each of these learning curves for the three observables plotted in Figs. 4.5(a-c). In Fig. 4.5(d) we see the expected Lebowitz-Spohn symmetry with reflection about $s = 1/2$ for the scaled CGF. The inset shows the absolute error compared to the diagonalization of the Fokker-Planck equation, $\epsilon(s)$, which illustrates

quantitative accuracy across the s values considered. The maximal error is on the order of 1%. Likewise, we see the expected anti-symmetry in the time-averaged KL divergence and velocity in Figs. 4.5(e) and (f). Both of these are also quantitatively accurate. This antisymmetry implies that the optimal force differs from the reference force more for $s > 1$ than $s < 0$. This demonstrates that the regular production of trajectories with significant negative time-integrated velocities requires a substantial change in the systems dynamics, in contrast to those with a significant positive velocity. Nevertheless the learning algorithm employed here is capable of parametrizing the modified force sufficiently well to work across these regimes.

Current fluctuations of an underdamped particle

In the underdamped case, the position and velocity evolve according to two coupled SDEs given by

$$
\begin{aligned}
dx &= v dt, \\
dv &= F(x)dt - v dt + \sqrt{2}dW
\end{aligned}
\tag{4.78}
$$

where the noise acts only on the velocity, v, and the friction, inverse temperature, and mass are taken as unity. As before we discretize our equations with a timestep of 0.001. For the underdamped case, the modified force and value function depends on both the position and velocity of the particle. The approximation need only provide a single output for a force applied to the velocity, as the optimal dynamics can not change the evolution of the position since the position is not directly influenced by noise. To do accomplish this, a simple approach we have taken is to discretize the force and value function approximation along the velocity dimension. More precisely, we can adapt the Fourier series from the overdamped case,

$$
F_\theta(x, v) = a^\theta(v) + \sum_{i=1}^{M_1} b_i^\theta(v)\sin(ix) + c_i^\theta(v)\cos(ix),
\tag{4.79}
$$

with velocity dependent coefficients given by

$$
a^\theta(v) = a_0 I_0(v) + a_{M_2+1} I_{M_2+1}(v) + \sum_{j=1}^{M_2} a_j I_{j,j+1}(v)
\tag{4.80}
$$

where

$$
I_{j,j+1}(v) = \begin{cases} 1 & v_0 + j\Delta v < v < v_0 + (j+1)\Delta v \\ 0 & \text{else} \end{cases}
\tag{4.81}
$$

and the boundary cases $I_0(v)$ and $I_{M_2+1}(v)$ return 1 for v less than v_0 or greater than $v_0 + (M_2 + 1)\Delta v$, respectively. We employ analogous equations for $b_i^\theta(v)$ and $c_i^\theta(v)$. To

achieve accurate results, we find a spacing of $\Delta v = 0.02$ is sufficient, with $v_0 = -8$, $M_2 = 700$ providing a broad enough range to encompass all relevant velocities at the biases considered. We use a Fourier basis with $M_1 = 5$. As before, we use the same functional for the value function as for this modified force.

Figure 4.6 shows estimates of same three quantities as the overdamped case throughout the same annealed learning process. Here we increased the value learning rate to $\alpha_\psi = 0.1$, retain a dynamics learning rate of $\alpha_\theta = 0.1$, and keep the scaled CGF learning rate fixed to $\alpha_R = 10^{-6}$ throughout training. Curves in Figs. 4.6(b,c) are produced from data calculated using the same learning rate as the scaled CGF, before using a windowed average over 100 steps to smooth the curve. We generally see fast convergence to an accurate result in approximately 10^8 transitions worth of updates. The large learning time compared to the overdamped results reflect the significantly finer basis employed for the underdamped model.

The ends of these curves are plotted below in Figs. 4.6(d-f). In the inset of Figs. 4.6(d) we see that we find accurate results compared to the numerically exact answers across the range of s considered. We see analogous results to the overdamped case, reproduced in dashed lines in Figs. 4.6(e,f), the underdamped system obeys the expected Lebowitz-Spohn symmetry. Compared to the overdamped system, the features of the KL divergence and average velocity in underdamped system are sharper.

There are three distinct behaviors for the system as a function of s. For large negative s, the velocity increases significantly. For very large positive s the velocity decreases analogously. For small and intermediate positive s, there is a broad plateau where the velocity is close to zero. These distinct regions are clearly demonstrated in Fig 4.7 where we plot the final optimized forces for a set of s, along with sample trajectories generated by these forces. We see different behavior for biases of $s < 0$, $0 < s < 1$ and $1 < s$. For $s < 0$ the trajectories regularly loop round the ring in the positive direction. For $0 < s < 1$ the trajectories generally do not transition round the ring and instead remain in a small region of space. For $s > 1$ the trajectories loop around the ring in the negative direction.

For comparison, we have optimized the same functional form using the MCR algorithm, as analogous to Ref. [117]. The AC algorithm provides more accurate results than MCR, when optimized using the same amount of statistics [117]. The MCR results are produced by annealing across from $s = 1.5$ down to $s = -0.5$ in steps of 0.1. Training for each value of s involves 20 updates constructed using 50 trajectories with 10^6 time steps each, for a total of 10^9 transitions worth of data. After optimizing the hyperparameters, we see in Fig. 4.8 the convergence in the MCR algorithm is still much slower than the AC algorithm. As a consequence, the best results we can achieve using the same amount of transitions fail to converge to the correct values of the scaled CGF for biases close to $s \gtrsim 1$. This demonstrates one key advantage of utilizing value functions. Due to the reduction in variance of gradient estimates using a small amount of data, we can perform many more updates using the same amount of transitions, improving convergence.
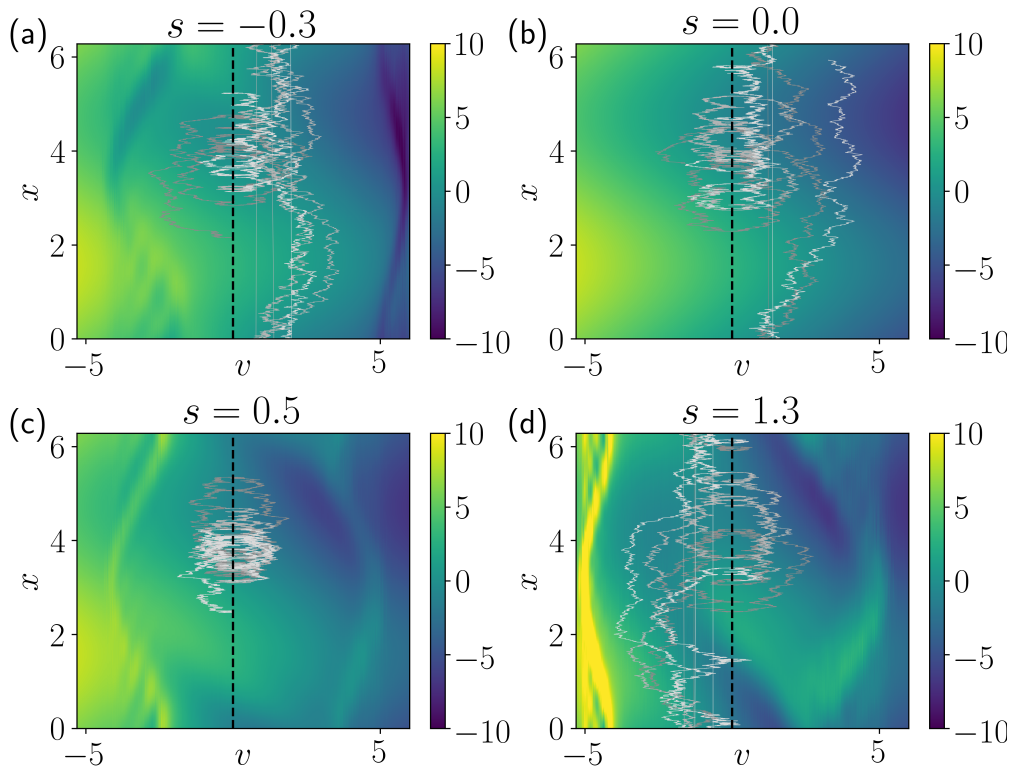
Figure 4.7: **Modified forces and their dynamics:** The final forces learnt during the optimization process for bias $s = -0.3$(a), 0.0(b), 0.5(c) and 1.3(d) with 3 sample trajectories of length $T = 10$ for each force.

4.9 Conclusion

In this paper we have demonstrated how regularized reinforcement learning algorithms can be used to optimize a diffusive dynamics to effectively sample rare trajectories. A key ingredient of our approach is a value function that estimates how relevant each state is to the rare dynamics, a function learnt while simultaneously guiding optimization of the dynamics, allowing for reduced data generation and more detailed function approximations. Across a range of systems and observables, we found that the lower variance estimate of the gradient employing value functions enabled accurate and efficient characterization of rare dynamical fluctuations. In finite time problems, the AC algorithm in particular was able to solve particularly challenging problems associated with multiple reactive channels and long lived intermediates. In the long time limit, the AC algorithm reproduces exact results for the cumulant generating function by directly optimizing to an accurate representation of the Doob dynamics, removing the need to calculate additional corrections or do additional importance sampling.
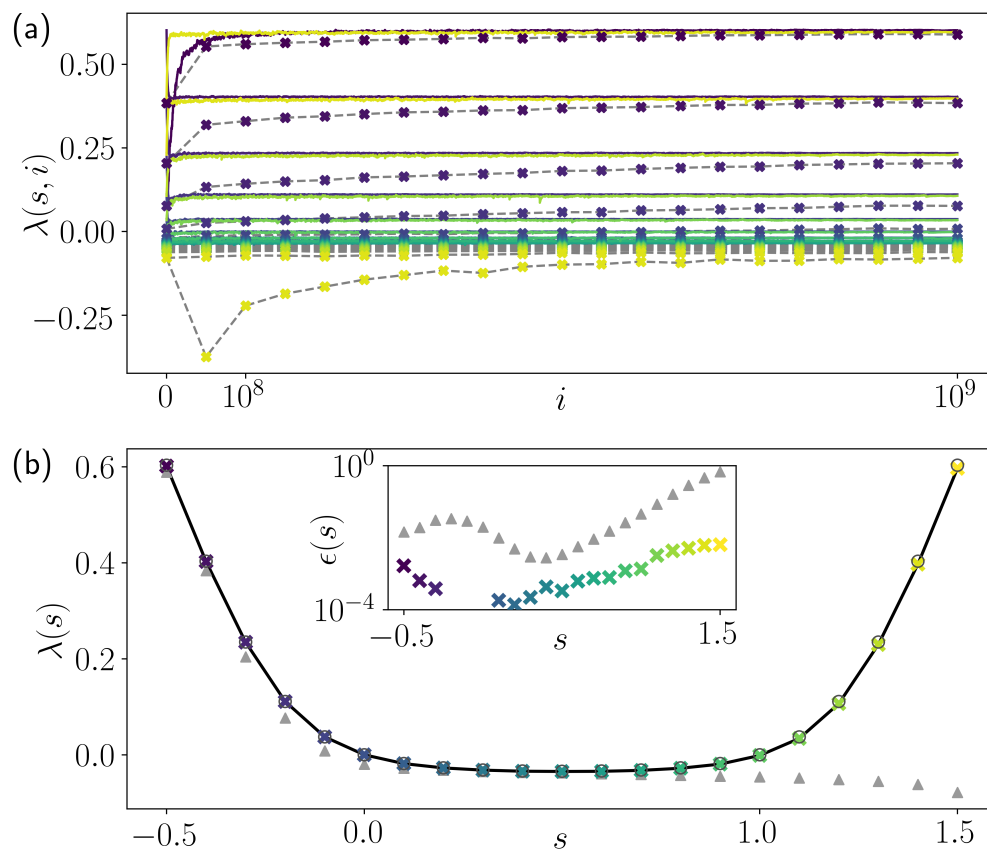
Figure 4.8: **Comparison between AC and MCR algorithms:** (a) learning curves plotted verses the amount of data used during training, for the AC algorithm (solid, colored lines) and the MCR algorithm (colored crosses and dashed gray lines). Curves and crosses are color coded by the value of the bias s being trained for. (b) Final results for the AC algorithm (colored crosses) and the MCR algorithm (gray triangles), with absolute errors to the value from numerical diagonalization shown in the inset.

While we have focused here on the simulation of rare event dynamics and the direct evaluation of their likelihoods, the methods of finding optimized forces developed here can be straightforwardly combined with trajectory importance sampling methods such as transition path sampling [70] or cloning [77] to correct for inaccuracies associated with an incomplete basis. Indeed, previous work has demonstrated that auxiliary dynamics can significantly improve the statistical efficiency of trajectory sampling methods [117, 81, 92, 316]. Further, Monte Carlo approaches can be used to generate data to train the optimal dynamics in a feedback routine as previously demonstrated [86, 315]. This could emphasize the parts of the state space relevant to the rare events earlier than by simply generating data with the current dynamics, thus speeding up optimization. Application to more complex models, such as many-body systems, will be an important development of this line of research. Accurate approximation of the force in many-body problems may require the use of more sophisticated function approximations, such as neural networks, however, a difficult balance will need to be struck between the representative power of the approximation and the computational cost to calculate it. More powerful function approximations will also necessitate the use of more sophisticated algorithms, as training such approximations can become unstable when using correlated data, as we do here.

Chapter 5

Direct evaluation of rare events from variational path sampling

The constituent agents of active matter– biomolecules, colloids, or cells– autonomously consume energy to fuel their motion [317, 318]. Their resultant nonequilibrium states have non-Boltzmann phase-space densities and exhibit exotic structural and dynamical collective fluctuations, including motility-induced phase separation and swarming [20, 319, 320, 48, 321]. Within these nonequilbrium steady-states, fleeting fluctuations can free particles from external potentials [322, 323, 324], nucleate stable phases from metastable ones [325, 65], and assemble passive objects [326, 327]. The study of such rare dynamical events within active matter and the calculation of their associated rates is difficult. Traditional equilibrium rate theories like transition state theory and Kramer's theory require knowledge of the form of the steady-state distribution that is not in general available [328]. Further, only a few numerical methods exist that can be used to tame the exponential computational cost associated with sampling the unlikely fluctuations that lead to transitions between metastable states. Existing methods improve sampling by stratifying or branching stochastic trajectories [102, 103, 311] but do not typically employ driving forces to specifically enhance the sampling of these rare events.

Here we present a perspective and an associated numerical algorithm, termed Variational Path Sampling (VPS), for estimating transition rates in active systems using optimized time-dependent driving forces. Our approach relies on a equality between the rate of a rare event in a reference system and a ratio of path partition functions in the reference system and with a driving force that makes the rare event occur with high probability. The VPS algorithm solves a variational problem to approximate the functional form of an optimal time-dependent driving force for this estimate and is applicable to any stochastic dynamics. With VPS we investigate how driven fluids can direct motion into useful function. We apply this technique to study the rate of conformational changes of a passive dimer in a dense bath of active Brownian particles [329, 330, 331]. This model exemplifies how collective active fluctuations around passive solutes can drive self-assembly and speed up transitions between distinct metastable states [332, 333]. We find the rate to switch between the dimer's two metastable

states increases dramatically with increasing activity in the bath, which we rationalize with a recent dissipation bound from stochastic thermodynamics [74]. We study the computational efficiency of rate estimation with VPS and demonstrate its advantage over existing trajectory stratification based methods like Forward Flux Sampling [102].[1]

5.1   Variational principle for rate computation

We consider a system described by overdamped Brownian dynamics of the form,

$$\gamma_i \dot{\mathbf{r}}_i(t) = \mathbf{F}_i[\mathbf{r}^N(t)] + \boldsymbol{\eta}_i(t) \tag{5.1}$$

where $\dot{\mathbf{r}}_i$ is the rate of change of the i-th particle's position, γ_i is the corresponding friction coefficient, and $\mathbf{F}_i[\mathbf{r}^N(t)]$ is the sum of all conservative, nonconservative and active forces exerted on the i-th particle that depends on the full configuration of the N-particle system, $\mathbf{r}^N$. The final term, $\boldsymbol{\eta}_i(t)$, is a Gaussian white-noise with $\langle \eta_{i\alpha}(t) \rangle = 0$ and

$$\langle \eta_{i\alpha}(t)\eta_{j\beta}(t') \rangle = 2\gamma_i k_{\mathrm{B}} T \delta_{ij} \delta_{\alpha\beta} \delta(t - t') \tag{5.2}$$

for component (α, β) and $k_{\mathrm{B}}T$ is Boltzmann's constant times the temperature. In order to study the transition rate between two long-lived metastable states, denoted A and B, we define each from a given configuration using the indicator functions,

$$h_X[\mathbf{r}^N(t)] = \begin{cases} 1 & \text{if } \mathbf{r}^N(t) \in X \\ 0 & \text{else} \end{cases}, \tag{5.3}$$

for either $X = A, B$. In practice this designation requires an order parameter capable of distinguishing configurations and grouping them into these distinct metastable states like that illustrated in Fig. 5.1(a) in one dimension. Assuming there exists a separation between the time $\tau^\ddagger$ required to traverse the transition region between the two metastable states, and the typical waiting time for the transition, the rate k can be evaluated from the probability to observe a transition, per unit time [335]

$$k = \frac{\langle h_B(t_f) h_A(0) \rangle}{t_f \langle h_A \rangle} = t_f^{-1} \langle h_{B|A}(t_f) \rangle, \tag{5.4}$$

where the angular brackets denote an average over trajectories of duration $\tau^\ddagger < t_f \ll 1/k$ started from a steady-state distribution in A and $\langle h_{B|A}(t_f) \rangle$ denotes the conditional probability for transitioning between A and B in time t_f. When t_f is chosen to satisfy the timescale separation described above, k is independent of time.

   If the transition is rare, most short trajectories are nonreactive leading to difficulties in estimating the rate directly. Instead of trying to evaluate the small transition probability

---

[1]Most of the content of this chapter was originally part of the publication [334].

through stratification as other existing methods do [102, 103], we instead optimize a time-dependent driving force $\boldsymbol{\lambda}(\mathbf{r}^N, t)$ that constrains the transition to occur, and evaluate the probability cost associated with adding that force to the original dynamics. For a general time-dependent force $\boldsymbol{\lambda}$, using the Onsager-Machlup form for the probabilities of stochastic trajectories [336], the rate expression in Eq. 5.4 can be rewritten as [74]

$$k = t_f^{-1} \left\langle e^{-\Delta U_{\boldsymbol{\lambda}}} \right\rangle_{B|A,\boldsymbol{\lambda}} , \tag{5.5}$$

where $\langle\rangle_{B|A,\boldsymbol{\lambda}}$ denotes a conditioned average computed in presence of the additional force. This relation holds for forces $\boldsymbol{\lambda}$ that affect the transition to occur with probability 1, such that the rate in the driven ensemble is $1/t_f$. The average is of the exponential of the change in the path action, $\Delta U_{\boldsymbol{\lambda}}$,

$$\Delta U_{\boldsymbol{\lambda}}[\mathbf{X}] = -\int_0^{t_f} dt \sum_i \frac{[\boldsymbol{\lambda}_i^2 - 2\boldsymbol{\lambda}_i \cdot (\gamma_i \dot{\mathbf{r}}_i - \mathbf{F}_i)]}{4\gamma_i k_{\mathrm{B}} T} , \tag{5.6}$$

between trajectories generated with the added force and in its absence. The path action and all other stochastic integrals are evaluated in the Ito convention.

Equation 5.5 is a direct estimator for a rate employing an auxiliary control system, but it only becomes useful when the protocol $\boldsymbol{\lambda}(\mathbf{r}^N, t)$ generates trajectories in a manner equivalent to the unbiased reactive trajectory distribution. This is because the expectation can be viewed as an overlap between the two reactive path distributions, and without significant overlap the exponential average is difficult to estimate. We express the optimality of $\boldsymbol{\lambda}$ using Jensen's inequality after taking the logarithm of Eq. 5.5 to obtain a variational principle,

$$\ln k \geq -\ln t_f - \langle \Delta U_{\boldsymbol{\lambda}} \rangle_{B|A,\boldsymbol{\lambda}} . \tag{5.7}$$

If the average change in conditioned path action $\langle \Delta U_{\boldsymbol{\lambda}} \rangle_{B|A,\boldsymbol{\lambda}}$ is minimized over all possible functional forms of $\boldsymbol{\lambda}$, the rate can be obtained directly as a simple ensemble average of $\Delta U_{\boldsymbol{\lambda}^*}$ at the minimizer $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$.

The optimal control force $\boldsymbol{\lambda}^*$ that saturates Eq. 5.7 is unique and given by the solution of the backward Kolmogorov equation [337, 338, 339] as detailed in the following subsection. Specifically, the optimal force is $2k_{\mathrm{B}}T$ times the gradient of the logarithm of the commitor probability [340] of ending in state B at t_f. A schematic illustration of the optimal effective time-dependent potential $V_t(R)$ added to a double well potential is illustrated in Fig. 5.1(a). The resultant force gradually destabilizes the reactant well to ensure the transition almost surely within the short duration t_f. Viewed in the backwards direction of time, the potential follows the negative logarithm of the relaxation of an initially localized distribution in B to its steady-state. The force is thus optimal in the sense that reactive trajectories, like those in Fig. 5.1(b), generated with it are drawn from the reference path ensemble with the correct statistical weights. Generically, $\boldsymbol{\lambda}^*(\mathbf{r}^N, t)$ is a function of all particle coordinates, so it is not typically tractable to compute. We demonstrate here that one- and two-body representations of $\boldsymbol{\lambda}$ can be sufficiently close to optimal as to estimate the rate accurately even in cases where the rare event is collective, similar to related observations in large-deviation sampling [84, 86, 87, 117].
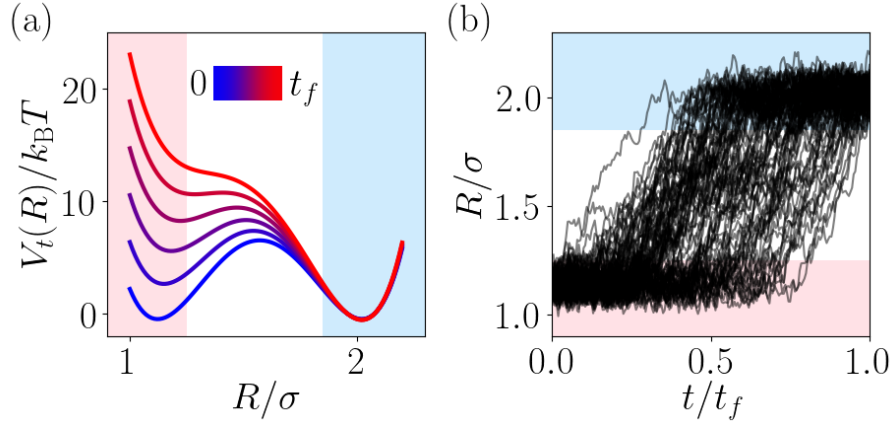
Figure 5.1: Reactive trajectories with VPS. (a)Schematic representation of the total optimal time-dependent potential in an isolated passive dimer as t goes from 0 to t_f. Shaded regions are the compact (A, pink) and extended (B, light blue) states. (b) Unbiased reactive trajectories generated with $\boldsymbol{\lambda}(R, t)$.

Demonstration of optimal control force

For ease of notation, we consider states A and B being specific phase space points, $\mathbf{r}_A^N$ and $\mathbf{r}_B^N$, respectively. Let $\boldsymbol{\lambda}^*(\mathbf{r}^N, t) = 2k_{\mathrm{B}}T\boldsymbol{\nabla}\Phi$, where $\Phi(\mathbf{r}_B^N, t_f|\mathbf{r}^N, t)$ is the log of the probability to end at a single target configuration, $\mathbf{r}_B^N$, conditioned on being at $\mathbf{r}^N$ at time t. This conditioned probability satisfies the logarithmic transform of the backward Kolmogorov equation [109, 337, 338, 339, 268],

$$\partial_t\Phi + \sum_i \left[D_i(\boldsymbol{\nabla}_i\Phi)^2 + \boldsymbol{\nabla}_i\Phi \cdot \mathbf{F}_i/\gamma_i + D_i\boldsymbol{\nabla}_i^2\Phi\right] = 0 \tag{5.8}$$

where $D_i = k_{\mathrm{B}}T/\gamma_i$ is the diffusion constant and the gradients act on $\mathbf{r}_i$. The optimal force achieves the reactive transition by construction, rendering $\langle h_{B|A}\rangle_{\boldsymbol{\lambda}^*} = 1$. The change in path action with the optimal force is

$$\Delta U_{\boldsymbol{\lambda}^*}[\mathbf{X}] = -\int_0^{t_f} dt \sum_i D_i(\boldsymbol{\nabla}_i\Phi)^2 + \boldsymbol{\nabla}_i\Phi \cdot \mathbf{F}_i/\gamma_i - \boldsymbol{\nabla}_i\Phi \cdot \dot{\mathbf{r}}_i \tag{5.9}$$

Using Ito's Lemma for the total time derivative of Φ to eliminate the final term,

$$\Delta U_{\boldsymbol{\lambda}^*}[\mathbf{X}] = -\int_0^{t_f} dt \left\{\sum_i \left[D_i(\boldsymbol{\nabla}_i\Phi)^2 + \boldsymbol{\nabla}_i\Phi \cdot \mathbf{F}_i/\gamma_i + D_i\boldsymbol{\nabla}_i^2\Phi\right] - \dot{\Phi} + \partial_t\Phi\right\} \tag{5.10}$$

and substituting the backward Kolmogorov equation

$$\Delta U_{\boldsymbol{\lambda}^*}[\mathbf{X}] = \int_0^{t_f} dt\, \dot{\Phi} = \Phi(\mathbf{r}_B^N, t_f|\mathbf{r}_B^N, t_f) - \Phi(\mathbf{r}_B^N, t_f|\mathbf{r}_A^N, 0) \tag{5.11}$$

the change of action can be evaluated exactly. The boundary terms from the exact integration are 0 and the log of the transition probability between A and B in the reference system. The latter can be identified with $\ln\langle h_{B|A}\rangle$, resulting in the equality in Eq. 7 using the definition of the rate k in Eq. 4. This reasoning extends linearly to cases where A and B are collections of configurations.

5.2 Optimized control force for passive dimer

We study the accuracy and utility of this formalism in a system comprised of an active bath and a passive dimer that can undergo conformational changes between two metastable states. All particles interact pairwise via a Weeks-Chandler-Andersen (WCA) repulsive potential [341]

$$V_{\text{WCA}}(\mathbf{r}) = \left\{ 4\epsilon \left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6} \right] + \epsilon \right\} \Theta(r_{\text{WCA}} - r) \tag{5.12}$$

with energy scale ϵ, and particle diameter σ, truncated at $r_{\text{WCA}} \equiv 2^{1/6}\sigma$ with the Heaviside function Θ. Active particles experience an additional self-propulsion force of magnitude v_0, $\mathbf{F}_i^a(t) = v_0\mathbf{e}[\theta_i(t)]$ where the director is $\mathbf{e}(\theta_i) = (\cos\theta_i, \sin\theta_i)$ and θ_i obeys $\dot{\theta}_i(t) = \xi_i(t)$ with,

$$\langle\xi_i(t)\rangle = 0, \quad \langle\xi_i(t)\xi_j(t')\rangle = 2D_\theta\delta_{ij}\delta(t - t') \tag{5.13}$$

for angular diffusion constant D_θ. Passive solutes separated by distance R are bound by a double-well potential

$$V_{\text{dw}}(R) = \Delta V \left[1 - (R - r_{\text{WCA}} - w)^2/w^2 \right]^2 \tag{5.14}$$

with an energy barrier of height ΔV between the compact and extended states at $R = r_{\text{WCA}}$ and $R = r_{\text{WCA}} + 2w$ respectively [342]. We study the transition rates between these states, employing indicator functions $h_A(t) = \Theta(R_A - R)$ and $h_B(t) = \Theta(R - R_B)$ for $R_A = 1.25\sigma$ and $R_B = 1.85\sigma$. Conformation transitions like these in dense fluids are collective in origin [342] and serve as a sensitive probe of the bath.

The VPS algorithm estimates an optimal force using a low-rank ansatz by iteratively solving the variational problem in Eq. 5.7, and uses this force to directly obtain a rate estimate. For computing the rate of isomerization of the passive dimer, we approximate $\boldsymbol{\lambda}^*$ with a time-dependent interaction along the dimer bond vector $\mathbf{R}$, expressed as a sum of Gaussians

$$\boldsymbol{\lambda}(\mathbf{R}, t) = \hat{\mathbf{R}} \sum_{p,q=1}^{M_R,M_t} c_{pq}^{(i)} e^{-\frac{(R-\mu_{R,p})^2}{2\nu_R^2} - \frac{(t-\mu_{t,q})^2}{2\nu_t^2}} \tag{5.15}$$

where $c_{pq}^{(1)} = -c_{pq}^{(2)}$ are variational parameters to be tuned, and the locations and widths $\mu_{R,p}$, $\mu_{t,q}$, ν_R and ν_t are held fixed. To impose the conditioning while minimizing $\langle\Delta U_{\boldsymbol{\lambda}}\rangle_{B|A,\boldsymbol{\lambda}}$, we use a Lagrange multiplier s to construct a loss function $\Omega_{\boldsymbol{\lambda}} = \langle\Delta U_{\boldsymbol{\lambda}}\rangle_{\boldsymbol{\lambda}} + s(\langle h_{B|A}\rangle_{\boldsymbol{\lambda}} - 1)$. For a general force that does not ensure the transition with unit probability, there is a

multiplicative contribution to the estimate of the rate in Eq. 5.5 from $\langle h_{B|A}\rangle_{\boldsymbol{\lambda}}$, which for most optimized forces is negligible.

The optimization problem maps onto the computation of a cumulant generating function for the statistics of the indicator $h_B(t_f)$ studied previously [339, 262], with the short trajectories starting from a steady-state distribution in the initial state. As such we can employ generalizations of recent reinforcement learning procedures to efficiently estimate the gradients of the loss function with respect to the variational parameters [125]. Specifically, we modify the Monte-Carlo Value Baseline (MCVB) algorithm [262] which performs a stochastic gradient descent to optimize $c_{pq}^{(i)}$. We add two preconditioning steps over the MCVB algorithm. First, we generate an initial reactive trajectory using a routine reminiscent of well-tempered metadynamics [343]. Then we symmetrize the learned force to ensure time translational invariance of the transition paths. We denote this preconditioning algorithm MCVB-T, which is described in Section 5.4

We first illustrate the systematic convergence of VPS by estimating the isomerization rate of an isolated passive dimer. Such a simplified system allows us to compare to numerically exact results, and study convergence of the force ansatz in the complete basis limit, where $M_R, M_t \to \infty$ and the Gaussians cover the thermally sampled region in R and t. For this simple system, we take $k_{\mathrm{B}}T = \gamma = \sigma = \epsilon = 1$, $w = 0.25\sigma$, with diffusive timescale $\tau = \sigma^2\gamma/k_{\mathrm{B}}T$. We simulate the one-dimensional version of Eq. 5.1 along R, with $V_{\mathrm{dw}}(R)$ only. For simplicity we define state A by the initial condition $R(0) = r_{\mathrm{WCA}}$, and state B via $R_B = 1.45\sigma$. To provide a steady-state value in Eq. 5.4 [262, 299], we use an Euler method and take in this example $t_f = \gamma w \sigma/\sqrt{8k_{\mathrm{B}}T\Delta V}$. We choose $\mu_{R,p}$ and $\mu_{t,q}$ evenly distributed in $R/\sigma \in [0.9, 1.77]$ and $t \in [0, t_f]$, respectively, and ν_R, ν_t to be half the distance between Gaussian centers. We consider basis sizes $M_R = M_t = 2 - 40$, each optimized independently and used to sample $\sim 10^5$ transition paths.

Figure 5.2(a) illustrates a typical learning curve for the control force, showing convergence of the variational rate bound towards the numerically exact rate. The variational estimate requires a basis of $M_R, M_t > 40$ to approach the rate to within the statistical uncertainty of the estimate, however alternative estimates with small basis sets can be refined using a cumulant expansion approximation to Eq. 5.5. Specifically, truncating the exact exponential relation at the ℓth cumulant as

$$\ln k \approx -\ln t_f + \sum_{n=1}^{\ell} \frac{1}{n!} \frac{d^n \ln \langle e^{-\Delta U_{\boldsymbol{\lambda}}}\rangle_{B|A,\boldsymbol{\lambda}}}{d\Delta U_{\boldsymbol{\lambda}}^n} \tag{5.16}$$

provides an approximation to the rate that converges in the limit that ℓ is large. Figure 5.2(b) illustrates this convergence, where we find that even coarse-representations of the control force can yield close estimates of the rate with only the first few cumulants, illustrating a tradeoff between basis set completeness and statistical efficiency. Sweeping across a wide range of barrier heights in Fig. 5.2(c), we find excellent agreement between the log-rate from brute force simulations and a truncation of the cumulant expansion to $\ell = 2$ using $M_R = 80$ and $M_t = 30$.
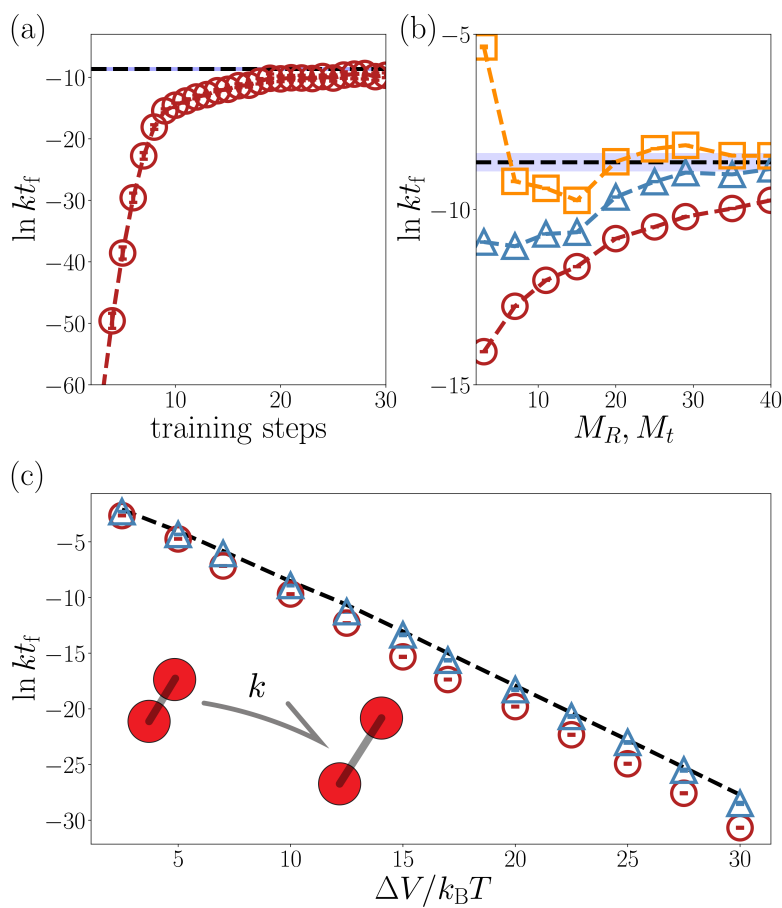
Figure 5.2: Convergence of isomerization rates for an isolated passive dimer. (a) Learning curve for $\Delta V = 10k_{\mathrm{B}}T$ and $M_R, M_t = 20$. (b) Convergence of the variational rate estimate (circles) and cumulant corrections for $\ell = 2$ (triangles) and $\ell = 4$ (squares) with basis size as compared to the numerically exact answer (dashed line). (c) Variational (circles) and $\ell = 2$ (triangles) estimate of the rate compared to the exact value (dashed line) with increasing barrier height.

5.3 Active driving in dimer isomerization

We next compute the isomerization rate with VPS when the dimer is immersed in an explicit solvent of active Brownian particles with $N = 80$ and a total density of $0.6/\sigma^2$. The dimer particles have a friction $\gamma_{\rm d} = 2\gamma$ and the solvent particles have $\gamma_{\rm s} = 4\gamma$. We take $\gamma = \sigma = \epsilon = 1$, $k_{\rm B}T = 0.5$, $\Delta V = 7k_{\rm B}T$, $\tau = \sigma^2\gamma/2k_{\rm B}T = 1$, $D_\theta = 1/\tau$ and timestep $10^{-5}\tau$. We also change $w = 0.45\sigma$ such that the collisional cross-section of the dimer is large. Collisions with active particles transduce energy along the dimer bond and we study the change in the isomerization rate as the bath activity $v_0\sigma/k_{\rm B}T$ is varied from 0 to 18. We use a basis size of $M_R = M_t = 50$ distributed between $R/\sigma \in [0.9, 2.3]$ and $t \in [0, t_f]$ where $t_f = 0.2\tau$. The optimization starts by learning forces $\boldsymbol{\lambda}(\mathbf{R}, t)$ for the isolated dimer with WCA interactions between monomers, followed by the MCVB-T algorithm. Then, $\boldsymbol{\lambda}(\mathbf{R}, t)$ is optimized in the presence of the bath for $v_0 = 0$ and higher values of v_0 are initialized from converged forces at the previous v_0.

The rate is a strong function of activity, increasing twenty-fold over the range of v_0's considered. While the variational rate estimate from Eq. 5.7 is closest for the passive bath, it weakens with increasing v_0, indicating a growing importance of solvent degrees of freedom in the optimal control force. With converged forces at each v_0, we run 10^6 trajectories of length t_f to compute k from Eq. 5.5. This estimate correctly predicts the suppression of k due to passive solvation and can be converged statistically for $v_0\sigma/k_{\rm B}T < 9$, which is supported by direct rate estimates from unbiased simulations in Fig. 5.3(a). Above $v_0\sigma/k_{\rm B}T = 9$, the optimized force is not close enough to $\boldsymbol{\lambda}^*$ to estimate k directly through the exponential average or a low order cumulant expansion.

Provided we have access to the transition path ensemble from direct unbiased simulations or methods like Transition Path Sampling [70, 81, 344] we can supplement the estimate of k using histogram reweighting [42]. k satisfies a reweighting relation of the form,

$$k = \frac{e^{-\Delta U_{\boldsymbol{\lambda}}} P_{B|A,\boldsymbol{\lambda}}(\Delta U_{\boldsymbol{\lambda}})}{t_f P_{B|A,0}(\Delta U_{\boldsymbol{\lambda}})} \tag{5.17}$$

where we have defined $P_{B|A,\boldsymbol{\lambda}}(\Delta U_{\boldsymbol{\lambda}}) = \langle \delta(\Delta U_{\boldsymbol{\lambda}}[\mathbf{X}] - \Delta U_{\boldsymbol{\lambda}}) \rangle_{B|A,\boldsymbol{\lambda}}$ and similarly for its undriven counterpart $\boldsymbol{\lambda} = 0$. We evaluate k with this estimator by sampling 10^4 driven and only 6-100 unbiased reactive paths, using the Bennett Acceptance Ratio [345] to evaluate the ratio of probabilities. Compared with the brute-force estimate in Fig. 5.3(a), we find this reweighting predicts k accurately across all values of v_0 with significantly higher statistical efficiency then a brute force calculation, which validates the accuracy and utility of the control forces. We have compared the VPS rate estimates in Section 5.6, using either Eqs. 5.5 and 5.17, to the Rosenbluth variant of Forward Flux Sampling [102], and find that VPS is statistically more efficient and converges more quickly with the number of reactive trajectories.

Access to an ensemble of transition paths in this active system gains us mechanistic insight into the process. The rate enhancement observed for the compact to extended state transition of the passive dimer with bath activity can be understood using recent results from stochastic thermodynamics. Specifically the rate enhancement achievable by coupling a
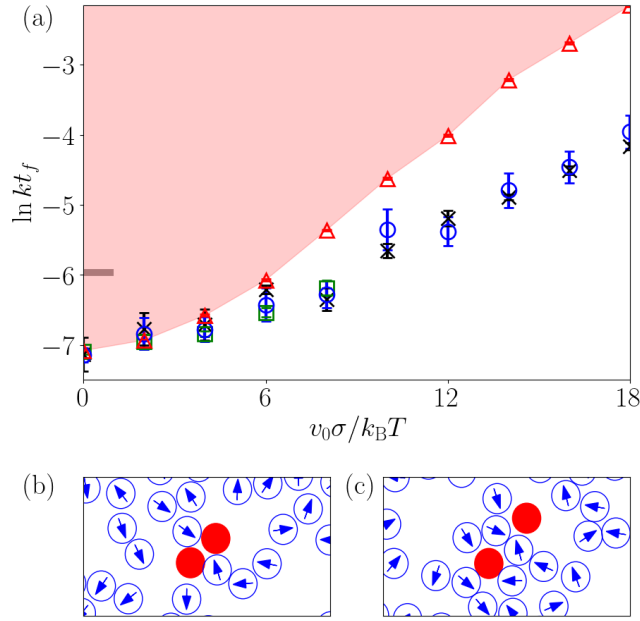
Figure 5.3: Rate enhancement of isomerization in an active fluid. (a) Change in the rate as estimated from direct unbiased simulations (crosses), from exponential estimate (squares), and from histogram reweighting (circles). The excess dissipated heat (triangles) bounds the rate enhancement achievable demarked by the red shaded region. The thick tick mark on the left denotes the rate for the isolated dimer. (b) and (c) Typical snapshots of reactive trajectories of the active bath (blue) and passive dimer (red), at $t = 0$ and $t = t_f$.

reactive mode to a nonequilibrium driving force is bounded from above by the heat dissipated over the course of the transition [74]. In this case the nonequilibrium driving is afforded by the interactions between the dimer and the active bath, so the bound takes the form

$$\ln k \leq \ln k_0 + \frac{1}{2k_B T} \langle Q - Q_0 \rangle_{B|A} \tag{5.18}$$

where k_0 is the rate at $v_0 = 0$ and $\langle Q - Q_0 \rangle_{B|A}$ is the dissipative heat less its average at $v_0 = 0$ given by

$$Q = \int_0^{t_f} dt \sum_{i \in d} \sum_{j \in s} (\dot{\mathbf{r}}_i - \dot{\mathbf{r}}_j) \cdot \mathbf{F}_{\text{WCA}}(\mathbf{r}_{ij}) \tag{5.19}$$

which is a sum of the total force from the WCA potential of the solvent particles (s) on the dimer (d) times the difference in their velocities in an ensemble at fixed v_0 (SM). This bound is verified in Fig. 5.3(a) for all v_0, and saturated at small v_0. The specific mechanism of energy transfer from bath to dimer that promotes transitions is clarified by examining reactive trajectories driven by the biasing force and are typical, after removal of the bias

from the incomplete basis set. Figures 5.3 (b) and (c) show typical snapshots of the solvated
dimer at the start and end of the reaction. Energy transfer results from active particles
accumulating around the dimer, and preferentially in its cross-section, pushing it apart into
an extended state. This mechanism of action is reminiscent of how nonequilibrium agents
collect in the corners of mesoscopic gears to power their directed rotation [332, 333]. At low
v_0, we find the driven isomerization process is efficient, while deviation from the bound at
large v_0 demonstrates that energy is additionally funneled into non-reactive modes. Further
studies showing the unbiased nature of the VPS-sampled transition path ensemble in terms
of duration and distribution of transition paths, and quantification of the changing solvation
environment with v_0 are provided in Sections 5.5 and 5.8.

5.4   Protocol for learning optimal force

We optimize the time-dependent control force $\boldsymbol{\lambda}(\mathbf{R}, t)$ by minimizing $\langle \Delta U_{\boldsymbol{\lambda}} \rangle_{B|A,\boldsymbol{\lambda}}$ over vari-
ational parameters $c_{pq}^{(i)}$ using Lagrange multiplier s to impose the $B|A$ conditioning. If
s is chosen to be more negative than an approximately estimated threshold value $s^* =$
$\ln[\langle h_{B|A} \rangle / (1 - \langle h_{B|A} \rangle)]$, the optimized forces drive the reaction with unit probability and s
need not be individually optimized. For a rare transition, any choice of s with a magni-
tude an order or more larger than the energy barrier height will robustly provide forces that
always satisfy the $B|A$ conditioning [262].

   For optimization we use an extension of a reinforcement learning algorithm called Monte
Carlo Value Baseline (MCVB) [262]. This algorithm computes the correlation of the gra-
dient of the log of the trajectory probability, called Malliavin weights [117, 122], with the
instantaneous change in $\Delta U_{\boldsymbol{\lambda}}$ and h_B over the course of the trajectory. These yield the
exact gradients of the loss-function $\Omega_{\boldsymbol{\lambda}}$ with respect to the tunable parameters, with which
a stochastic gradient descent is performed. The MCVB algorithm simultaneously learns
the driving force and a corresponding value function, $\mathcal{V}(\mathbf{R}, t) = \langle \Delta U_{\boldsymbol{\lambda}, t} + s h_{B|A} \rangle_{\boldsymbol{\lambda}|R(t)=R}$,
where $\Delta U_{\boldsymbol{\lambda}, t}$ contains the integrated action difference only within $[t, t_f]$ and the expectation
is conditioned on starting from R at time t. The value function greatly reduces the the
variance of the gradients at zero cost, allowing better convergence. Our modification to this
algorithm, refered to as MCVB-T, is a preconditioning step that enforces time translational
symmetry for the log of the bridge probability, $\Phi(\mathbf{r}_f^N, t_f | \mathbf{r}^N, t) = \Phi(\mathbf{r}_f^N, t_f - t | \mathbf{r}^N, 0)$, as the
reference forces are not explicit functions of time. This is achieved by randomly choosing
a $t_{\mathrm{mid}} \in [0, t_f]$ for every trajectory used for averaging the force gradient, and applying the
force $\boldsymbol{\lambda}(\mathbf{R}, t \in [t_{\mathrm{mid}}, t_f])$ on it only for a duration $[t_{\mathrm{mid}}, t_f]$. This ensures that trajectories
undergoing the transition at late times are accounted for while training the force.

   Details of the MCVB-T algorithm are available in the pseudocode in Algorithm 7. The
set of Gaussian coefficients parametrizing the force and the value function are denoted in
short by χ and ψ respectively. The MCVB algorithm is a special case of MCVB-T with fixed
$t_{\mathrm{mid}} = 0$.

Algorithm 7 Monte-Carlo Value Baseline with Time-translation invariance (MCVB-T)

1: **inputs** Gaussian coefficients for a general force $\boldsymbol{\lambda}_\chi(\mathbf{r}^N, t)$ and value function $\mathcal{V}_\psi(\mathbf{r}^N, t)$
2: **parameters** learning rates α^χ, α^ψ; total optimization steps I; trajectory length t_f consisting of J timesteps of duration Δt each; number of trajectories N
3: **initialize** choose initial weights χ and ψ, define iteration variables i and j, force and value function gradients δ_P, δ_V, define functional form for stepwise increments (rewards) ξ to the loss-function $\Delta U_{\boldsymbol{\lambda}} + s h_{B|A}$
4: $i \leftarrow 0$
5: **repeat**
6:     Generate trajectories $[\mathbf{X}(t)]$ with first-order Euler propagation starting from uncorrelated steady-state configurations in state A. Every trajectory starts experiencing the force $\boldsymbol{\lambda}$ from a random time t_{mid} which is sampled uniformly from $[0, t_f]$. Configurations, times, noises (with variance $2\boldsymbol{\gamma} k_{\mathrm{B}} T \Delta t$), Malliavin weights, integral of value function gradients, and rewards are denoted by $\mathbf{r}_j^N, t_j, \boldsymbol{\eta}_j, y_\chi(t_j), z_\psi(t_j)$ and $\xi(t_j) = \xi_j$ respectively.
7:     $j \leftarrow 0$
8:     $\delta_P \leftarrow 0$
9:     $\delta_V \leftarrow 0$
10:     $y_\chi(t_0) \leftarrow 0$
11:     $z_\psi(t_0) \leftarrow 0$
12:     **repeat**
13:         $\dot{y}_\chi(t_j) \leftarrow \boldsymbol{\eta}_j \cdot \nabla_\chi \boldsymbol{\lambda}_\chi(\mathbf{r}_j^N, t_j)/2k_{\mathrm{B}} T \Delta t$
14:         $y_\chi(t_{j+1}) \leftarrow y_\chi(t_j) + \Delta t \dot{y}_\chi(t_j)$
15:         $\dot{z}_\psi(t_j) \leftarrow \nabla_\psi \mathcal{V}_\psi(\mathbf{r}_j^N, t_j)$
16:         $z_\psi(t_{j+1}) \leftarrow z_\psi(t_j) + \Delta t \dot{z}_\psi(t_j)$
17:         $\delta_P \leftarrow \delta_P + \xi_j y_\chi(t_{j+1}) - \mathcal{V}_\psi(\mathbf{r}_j^N, t_j)\dot{y}_\chi(t_j))$
18:         $\delta_V \leftarrow \delta_V + \xi_j z_\psi(t_{j+1}) - \mathcal{V}_\psi(\mathbf{r}_j^N, t_j)\dot{z}_\psi(t_j))$
19:         $j \leftarrow j + 1$
20:     **until** $j = J$
21:     average δ_P, δ_V over N trajectories to get $\overline{\delta}_P, \overline{\delta}_V$
22:     $\chi \leftarrow \chi - \alpha^\chi \overline{\delta}_P$
23:     $\psi \leftarrow \psi + \alpha^\psi \overline{\delta}_V$
24:     $i \leftarrow i + 1$
25: **until** $i = I$

Figure 5.4 illustrates all learning curves that led to the results in Fig. 3, plotting $\ln \overline{h_{B|A}} - \overline{\Delta U_{\lambda}}\big|_{B|A}$ as a function of training steps. The averages of this estimator and the gradients are computed over 40 trajectories simulated at each training step. The trajectories are initialized with coordinates randomly chosen from a collection of 10000 steady-state dimer and bath configurations in state A, collected once in every 0.1τ time units from a long trajectory without any driving forces. We first learn optimal forces in the absence of the explicit bath in Fig. 5.4(a), and then optimize these forces in the presence of the bath in Fig. 5.4(b). We start our optimization by first finding an arbitrary force that ensures the transition with a nonzero probability. We learn initial parameters $c_{pq}^{(i)}$ with a routine similar to well-tempered metadynamics [343]. Starting with $c_{pq}^{(i)} = 0$, at fixed frequency we add $c_{pq}^{(i)} \mapsto c_{pq}^{(i)} + \tau_m \omega T_m / [T_m + \omega \mathcal{N}(R, t)]$, where $\mathcal{N}(R, t)$ is a running histogram of order parameter R up to the current time t, and hyperparameters T_m, ω and τ_m determine how quickly the force landscape is filled. The blue learning curve in Fig. 5.4(a) refers to 100 steps of metadynamics run with $\tau_m = 10t/M_t$, $\omega = 4000$ and $T_m = 9000$. We find that the force solely from metadynamics is highly suboptimal compared to the rate bound, indicated by the black dashed line. Starting with a $\boldsymbol{\lambda}$ averaged over all metadynamics steps and with $\mathcal{V} = 0$, we next optimize both sets of parameters with MCVB-T and then MCVB each over 1000 steps with learning rates $\alpha^{\chi} = 40, \alpha^{\psi} = 200$ and $s = -100$. We find that the variational estimate converges tightly to the exact rate bound.
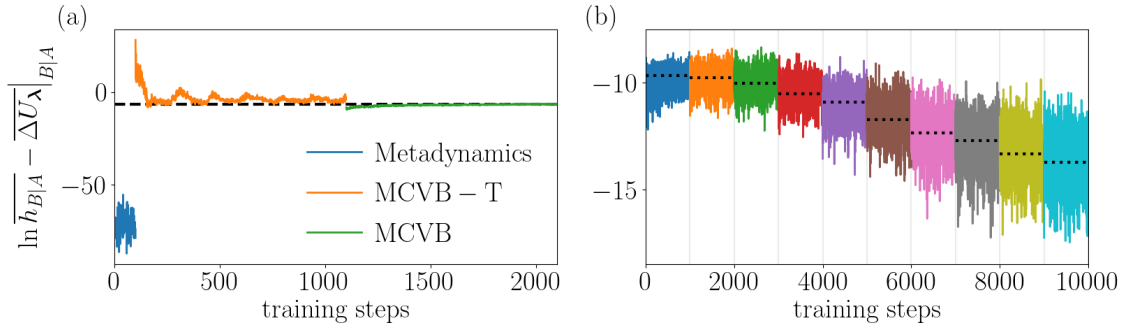
Figure 5.4: Learning curves for variational bound. (a) Optimization for the isolated dimer with 100 steps of well-tempered metadynamics(blue), 1000 steps MCVB-T(orange) and 1000 steps MCVB(green). Black dashed line is $\ln kt_f$ for the isolated dimer. (b) Learning curves for 1000 steps each, in presence of the explicit bath with $v_0\sigma/k_{\mathrm{B}}T = 0, 2, 4, ..., 18$. Black dotted lines denote the corresponding converged values.

Next we use the converged $\boldsymbol{\lambda}$ and $\mathcal{V}$ to start the optimization in presence of the bath, as illustrated in Figure 5.4(b). We successively optimize for each $v_0\sigma/k_{\mathrm{B}}T \in \{0, 2, 4, ..., 18\}$ starting from the previous converged result, each over 1000 steps. Each time we choose $(\alpha^{\chi}, \alpha^{\psi}) = (0, 200)$ for the first 200 steps and $(40, 200)$ for the remaining 800 steps. Learning the value function before starting to change the force in this way avoids a brief period of

divergence at the beginning of each optimization run [262]. The results are robust towards changing the learning rates as long as α^ψ is kept about 5-10 times of α^χ, such that the value function is always approximately accurate whenever the force is being changed.

Results in Fig. 2 were also obtained similar to this protocol, but with no value function. For Fig. 2(c), the initialization parameters τ_m, ω and T_m are chosen at each barrier height so that at least half of the transitions are reactive.

5.5 Unbiased reactive events from VPS

We use Eq. 5 and 13 in the main text to obtain rate estimates from direct simulations using the low-rank optimized force $\boldsymbol{\lambda}$. For the passive dimer in an active bath with $v_0\sigma/k_BT = 8$, we have illustrated in Figure 5.5(a) overlap of the driven distribution $P_{B|A,\boldsymbol{\lambda}}(\Delta U_{\boldsymbol{\lambda}})$ with the unbiased distribution $P_{B|A}(\Delta U_{\boldsymbol{\lambda}})$ after tilting to correct the systematic error. The scaling constant k_{rwt}, which is our estimate for the rate k, has been evaluated from Bennett Acceptance Ratio method [345] by using the tilting exponent as the reduced potential. This overlap is observable only when the driving force $\boldsymbol{\lambda}$ is near-optimal. If the tilted distribution does not contain enough statistics to represent the unbiased distribution, the estimate k_{exp} from Eq. 5 given by the area under the tilted distribution will underestimate the rate. If the basis set is complete and the exact optimal force $\boldsymbol{\lambda}^*$ can be obtained, $\Delta U_{\boldsymbol{\lambda}^*}$ will follow a Dirac delta distribution $P_{B|A,\boldsymbol{\lambda}^*}(\Delta U_{\boldsymbol{\lambda}^*}) = \delta(\Delta U_{\boldsymbol{\lambda}^*} + \ln kt_f)$, and the first cumulant will be sufficient to describe the log of the average of the exponential. This is also evident in Eq. 5.11 where $\boldsymbol{\lambda} = \boldsymbol{\lambda}^*$ makes $\Delta U_{\boldsymbol{\lambda}}[\mathbf{X}]$ trajectory independent. In that case, all three estimates of k from Eqs. 5, 7 and 13 will be equal and the unbiased reactive events will be entirely force-assisted rather than being driven by thermal fluctuations.

Figure 5.5(b) shows the systematic and statistical errors in $\ln kt_f$ calculated as VPS estimates k_{exp} and k_{rwt} from Equations 5 and 13 respectively in the main text. We have computed the errors by comparing to direct unbiased simulation, as the number N_w of uncorrelated trajectories of duration t_f is varied, expressed through the total number of simulation timesteps $N_F = N_w t_f/\delta t$ where δt is a single timestep. Given the optimized driving force, k_{exp} is computed by averaging over N_w trajectories and k_{rwt} is obtained by reweighting $P_{B|A,\boldsymbol{\lambda}}(\Delta U_{\boldsymbol{\lambda}})$ and $P_{B|A}(\Delta U_{\boldsymbol{\lambda}})$ each computed with $N_w/2$ total trajectories, of which only a fraction are reactive without the driving force. At small N_F, k_{exp} systematically underestimates the rate due to the full area under $e^{-\Delta U_{\boldsymbol{\lambda}}}P_{B|A}(\Delta U_{\boldsymbol{\lambda}})$ not being accessible because of incomplete overlap, making k_{exp} formally unbiased but statistically biased. This error disappears with large N_F. However, the full rate can still be successfully obtained by comparing segments of incomplete distributions. Thus even when the undriven trajectory ensemble has ≤ 10 reactive trajectories at smaller values of N_F, k_{rwt} incurs much less error and provides a rate estimate that is both formally and statistically unbiased.

Figure 5.5(c) demonstrates convergence of the transition path ensemble obtained from direct simulations with the optimized forces even before the tilting correction. $P_{B|A,\boldsymbol{\lambda}}(\tau^\ddagger)$ and $P_{B|A}(\tau^\ddagger)$ are distributions of the transition path time $\tau^\ddagger$ measured as the time after leaving
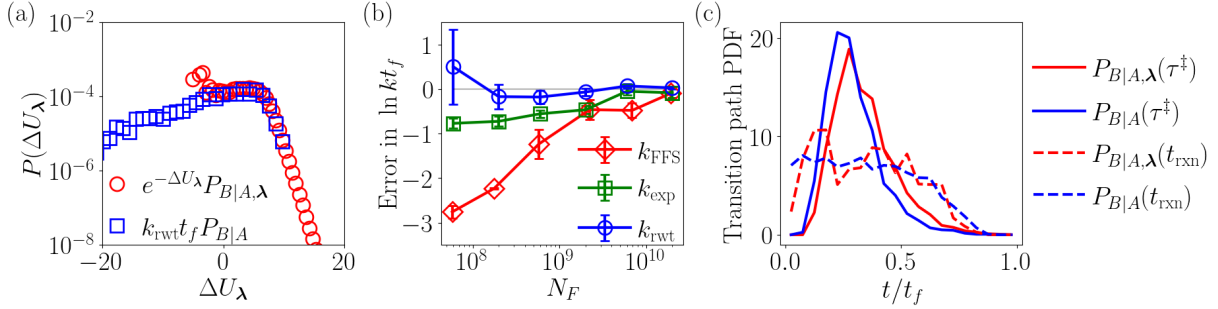
Figure 5.5: Unbiased rates, statistical convergence and transition path ensemble with VPS for $v_0\sigma/k_BT = 8$

. (a) Overlap of the incomplete tilted biased and the unbiased distributions, with the scaling coefficient computed from Bennett Acceptance Ratio. (b) Errors in the rate estimates from Equations 5 (k_{exp}) and 13 (k_{rwt}) and from Forward Flux Sampling (k_{FFS})as the amount of total simulation timesteps N_F is varied. (c) Probability Density Functions (PDF) of transition path times and reactive escape times in the transition path ensemble, computed from the driven trajectories and unbiased reactive trajectories.

state A and before reaching state B without returning to A. $P_{B|A}(\tau^\ddagger)$ is from 2000 reactive trajectories obtained from 10^6 total unbiased simulated trajectories, while $P_{B|A,\boldsymbol{\lambda}}(\tau^\ddagger)$ is from a total of 2000 driven trajectories all of which were reactive. We find convergence in the distribution of transition path times signifying direct access to the nearly unbiased transition path ensemble by using the optimal force. Similarly we compare the distribution of the start time of the reaction $t_{rxn} \in [0, t_f]$ measured as the time the trajectory last leaves A before arriving in B. We again find convergence in the driven ensemble compared to the unbiased reactive ensemble indicating the forces $\boldsymbol{\lambda}(\mathbf{R}, t)$ are near-optimal at all values of t.

The directly evaluated rates without additional forces used to compare VPS estimates have in most cases been computed from 5 trajectories, each of duration $10^4\tau$ with τ being the diffusive timescale. We compute k using Eq. 4 by evaluating the expectation with a rolling window over the trajectories after relaxing to a steady-state. We deviate from this protocol only for Figure 2(c), where the barrier heights are too large to estimate the rate from direct simulations. Here we use a numerically exact escape rate obtained from Kramer's theory [11].

5.6 Comparison with Forward Flux Sampling

In Figure 5.5(b) we have compared the numerical cost of VPS at an active self-propulsion $v_0\sigma/k_BT = 8$ with that of a Rosenbluth-like variant of Forward Flux Sampling (RB-FFS)

[102]. Starting from an ensemble of steady-state configurations in A, RB-FFS uses multiple interfaces between A and B to sequentially generate the transition path ensemble and compute the nonequilibrium reaction rate without an additional driving force [346, 67]. The transition paths generated from RB-FFS are unbranched and each has an associated weight as part of the transition path ensemble, analogous to VPS, from which the rate is estimated.

In order to apply RB-FFS, we define the interfaces along R as $R/\sigma \in \{1.25, 1.29, 1.33, 1.38, 1.43, 1.50, 1.57, 1.65, 1.72, 1.77, 1.81, 1.85\}$ with the first and the last interfaces corresponding to R_A and R_B respectively. We start RB-FFS trajectories from the same ensemble of steady-state configurations in A that we have used for VPS, and record the configurations whenever the trajectories cross R_A from the A side. Every time a trajectory reaches B, we replace it in A at a random steady-state configuration. From each of $\mathcal{M}_0$ recorded configurations located at R_A, we generate reactive paths by shooting $\mathcal{M} = 100$ trajectories from each interface sequentially and randomly choosing one out of those that reach the next interface instead of coming back to A. The rate estimate is given by product of the forward flux of crossing R_A and the conditional probabilities of reaching subsequent interfaces, computed from an average over weighted reactive trajectories from the RB-FFS simulation [102]. We have varied $\mathcal{M}_0$ between 20 and 7000 to study the convergence of the RB-FFS rate estimate k_{FFS} as a function of the total number of simulation timesteps N_F, as shown in Figure 5.5(b). Statistical errors have been estimated over 3 independent parallel runs of the entire RB-FFS procedure.

We find that the VPS rate estimates k_{exp} and k_{rwt} incur much smaller errors than k_{FFS} at small N_F, though at large N_F all estimates converge to the same rate. Specifically, k_{rwt} converges to the true rate fastest among the three estimates, and k_{exp} incurs much smaller systematic errors than k_{FFS} even before convergence. This demonstrates that the use of the optimized force in a simple low-dimensional basis in VPS reduces the computational cost of estimating the exact rate by an order of magnitude or more compared to a trajectory stratification based method like RB-FFS. Further, we find that in RB-FFS, obtaining sufficient statistics given by a large $\mathcal{M}_0$ required the use of a long serial simulation to converge the flux of crossing the first interface at R_A. Parallelization of the $\mathcal{M}$ trajectory segments starting from each interface scaled poorly due to a broad distribution of durations over the trajectory segments, each of which must continue till they reach either the next interface, or A. Since shooting from the next interface can only start after the slowest of the previous trajectory segment has concluded, parallel implementations of RB-FFS scaled poorly and required overall a very long serial simulation. Our attempts to parallelize RB-FFS in an alternate fashion by reducing the serial configurations $\mathcal{M}_0$ worsened the systematic error in k_{FFS} even when averaged over fully independent RB-FFS implementations over many parallel threads. In contrast, every step of VPS is trivially parallelizable because of all trajectories being of the same duration t_f, which corresponds to $N_F = 2 \times 10^4$. As a result, a parallel implementation of VPS reduced its cost linearly and the overall computation required much shorter serial simulations.

5.7 Dissipative rate bounds

Stochastic thermodynamics provides a fundamental speed-limit on the enhancement achievable of the rate k of a rare nonequilibrium transition over a reference equilibrium rate k_0 in terms of the excess heat dissipation in the reactive path ensemble [74],

$$\ln k \leq \ln k_0 + \frac{1}{2k_{\mathrm{B}}T}\langle Q\rangle_{B|A} \tag{5.20}$$

where $\langle Q\rangle_{B|A}$ is defined as the time-reversal asymmetric contribution from the change in path action between the equilibrium reference and the nonequilibrium system in which it is measured. This bound holds under mild assumptions of instantonic or diffusive transitions and follows from a similar change of measure as leads to Eq. 7, with the additional observation of the time-reversal symmetric contribution of the change in path action being negligible near equilibrium. Here we show how to arrive at Eq. 14 employing this bound. As in the main text, we assume a separation of timescales between local relaxation and typical transitions so that the rate problem is well-posed.

We consider the rate enhancement afforded by coupling the dimer to an active solvent over the equilibrium passive bath isomerization rate. However, if we simply compute the excess heat dissipated as resulting from the time reversal asymmetric change in path action in turning v_0 from 0 to some finite value, the heat will be extensive in the number of solvent degrees of freedom and thus not have a well-defined thermodynamic limit. To mitigate this, we note that the isomerization rate of the dimer would be independent of v_0 if the dimer and solvent did not interact. Denoting k^{ni} and k_0^{ni} the rates of isomerization when the dimer is uncoupled to the solvent at finite or zero v_0, respectively, then $k^{\mathrm{ni}} = k_0^{\mathrm{ni}}$ and

$$\ln \frac{k}{k_0} = \ln \frac{k}{k^{\mathrm{ni}}}\frac{k_0^{\mathrm{ni}}}{k_0} \leq \frac{1}{2k_{\mathrm{B}}T}\left(\langle Q\rangle_{B|A} - \langle Q\rangle_{B|A,0}\right) \tag{5.21}$$

where $\langle Q\rangle_{B|A}$ is the excess dissipation resulting from turning on interactions between the dimer and solvent at finite v_0, and $\langle Q\rangle_{B|A,0} = Q_0$ results from turning on interactions between the dimer and solvent at $v_0 = 0$. The inequality is preserved even though a difference of heats is taken since the second ratio of rates k_0^{ni}/k_0 are both evaluated at equilibrium and thus the symmetric part of the action is zero. This second heat subtracts out the dissipation that is uncorrelated with the isomerization, and the remaining excess dissipation is left finite even when the number of solvent particles is large, so long as the dimer is correlated with the solvent over a finite distance.

The full path action for a system at finite v_0 in the presence of dimer-solvent interactions

is

$$U_{v_0} = -\frac{1}{4k_BT}\int_0^{t_f} dt\, \gamma_d^{-1}\sum_{i\in d}\left[\gamma_d\dot{r}_i + \nabla_i V_{dw} - \sum_{j\in d}\mathbf{F}_{WCA}(\mathbf{r}_{ij}) - \sum_{j\in s}\mathbf{F}_{WCA}(\mathbf{r}_{ij})\right]^2$$

$$+ \gamma_s^{-1}\sum_{i\in s}\left[\gamma_s\dot{r}_i - v_0\mathbf{e}[\theta_i] - \sum_{j\in d}\mathbf{F}_{WCA}(\mathbf{r}_{ij}) - \sum_{j\in s}\mathbf{F}_{WCA}(\mathbf{r}_{ij})\right]^2 - \frac{1}{4D_\theta}\int_0^{t_f} dt\sum_{i\in s}\dot{\theta}_i^2$$

$$(5.22)$$

and using the convention that v_0 is invariant under time-reversal [48], the dissipated heat associated with turning on interactions between the solvent and dimer is

$$Q(t_f) = \int_0^{t_f} dt\left[\sum_{i\in d}\sum_{j\in d}\dot{\mathbf{r}}_i\mathbf{F}_{WCA}(\mathbf{r}_{ij}) + \sum_{j\in s}\sum_{i\in d}\dot{\mathbf{r}}_j\mathbf{F}_{WCA}(\mathbf{r}_{ji})\right] \qquad (5.23)$$

and is the same if $v_0 = 0$ or is nonzero. Since $\mathbf{F}_{WCA}(\mathbf{r}_{ji}) = -\mathbf{F}_{WCA}(\mathbf{r}_{ij})$ we find the definition of the Q in Eq. 15.

We measure $\langle Q\rangle_{B|A}(t_f)$ by averaging Eq. 5.23 over reactive trajectories of length $t_f = .2\tau$ sampled from long, 2×10^8 time-step, simulations in the nonequilibrium steady state at fixed v_0, with all other parameters as in the main text. Assuming transitions are uncorrelated, we compile Q samples from $24 - 96$ independent simulations, and use this data to calculate a mean and standard error, as depicted in Fig. 3 (red triangles).

5.8 Nonequilibrium solvation structure

We have studied how the solvation structure around the dimer evolves with activity. We compute the two-dimensional pair distribution functions for the position of the solvent around the dimer bond within a conditioned steady state ensemble average

$$g_X(r,\phi_1) = \frac{1}{\rho_s\rho_d}\frac{\langle\sum_{i\in s}\delta(\mathbf{r}_{cm}^d)\delta(r - |\mathbf{r}_i - \mathbf{r}_{cm}^d|)\delta(\phi_1 - \arccos(\mathbf{r}_i\cdot\mathbf{R}))h_X(R)\rangle}{\langle h_X(R)\rangle}, \qquad (5.24)$$

with the center of the dimer bond $\mathbf{r}_{cm}^d = (\mathbf{r}_1 + \mathbf{r}_2)/2$ as a reference. Here, r is the radial distance between $\mathbf{r}_{cm}^d$ and surrounding bath particles, which make an angle ϕ_1 with the bond vector $\mathbf{R} = \mathbf{r}_1 - \mathbf{r}_2$. The indicator function $h_X(R)$ restricts configurations where the bond length R falls into state X. Similarly, we probe the orientation of active solute particles around the dimer bond vector with the pair distribution

$$g_X(r,\phi_2) = \frac{1}{\rho_s\rho_d}\frac{\langle\sum_{i\in s}\delta(\mathbf{r}_{cm}^d)\delta(r - |\mathbf{r}_i - \mathbf{r}_{cm}^d|)\delta(\phi_1 - \arccos(\mathbf{e}[\theta_i]\cdot\mathbf{R}))h_X(R)\rangle}{\langle h_X(R)\rangle}. \qquad (5.25)$$

where ϕ_2 is the angle between a bath director and the dimer bond, and $\rho_s\sigma^2 = 0.6$ is the density of the solvent, and $\rho_d\sigma^2 = 0.008$ is the density of the dimers. To compute $g_X(r,\phi_i)$
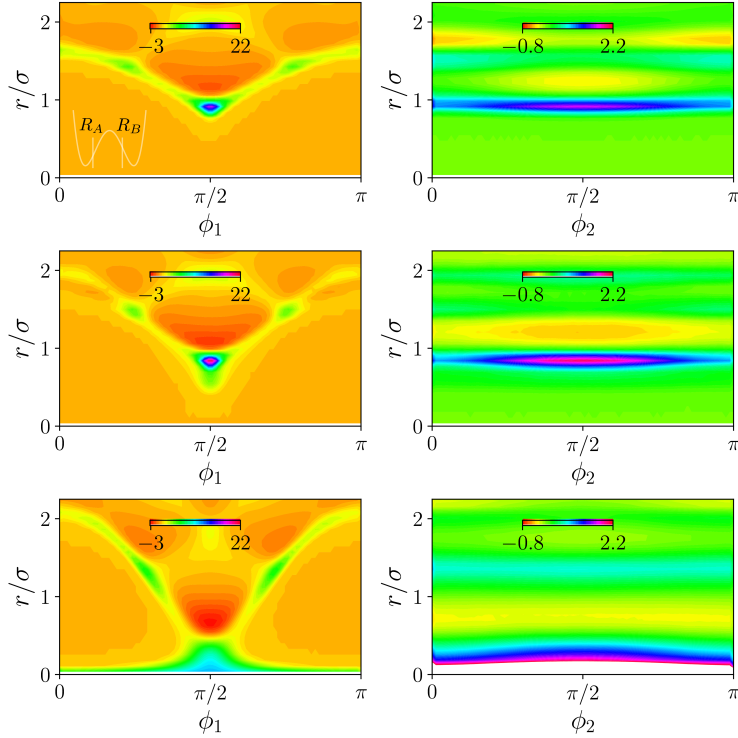
Figure 5.6: Solvation structure of the dimer by the active bath. Difference in pair distributions $\Delta g_X(r, \phi_1)$ (Left) and $\Delta g_X(r, \phi_2)$ (Right). Configurations are conditioned such that the bath is sampled with the dimer in the collapsed state $\Delta g_A(r, \phi_{1,2})$ (Top), the transition region $\Delta g_{AB}(r, \phi_{1,2})$ (Middle) or the extended state $\Delta g_B(r, \phi_{1,2})$ (Bottom).

we average over configurations sampled from 24 simulations each with a length of 2×10^8 time-steps.

In Fig. 5.6, we consider the change in the pair distributions $\Delta g_X(r, \phi_{i=1,2}) = g_X(r, \phi_{i=1,2}, v_0 = 9) - g_X(r, \phi_{i=1,2}, v_0 = 0)$ in an active bath with $v_0 = 9$ and its equilibrium counterpart at $v_0 = 0$. The different rows impose different conditions for the dimer bond distance $R = |\mathbf{R}|$ to be either primarily in state $X = A$ (top row), with $R < 1.55\sigma$, in the transition region $X = AB$ between states $1.55\sigma < R < 1.65\sigma$ (center row), or mostly in state $X = B$ (bottom row), with $R > 1.65\sigma$.

These results demonstrate that the rate enhancement is correlated with active particles dynamically wedging within the cross section of the dimer, pushing it apart into an extended state. The left column of Fig. 5.6 demonstrates that activity greatly enhances the packing of bath particles between the two bonded dimer particles, while the right column illustrates that bath particles preferentially orient perpendicular to the bond vector once far enough within the cross section. In state A, the active nature of the bath causes particles to push

the dimer apart along R, as evidenced by the depletion for $\phi_2 = \pi/2$ and $r/\sigma > 1$ in the top-right panel of Fig. 5.6. The transition region, center-left, shows a significantly higher peak in the radial distribution function around $\phi_2 = \pi/2$, marking a decrease in the height of the effective free energy barrier along R. This analysis also illustrates the mechanism of increased stability in the active dimer extended state. Namely, Fig. 5.6 bottom-left shows that the driven bath particles act to inhibit the extended state from closing.

5.9 Conclusion

In conclusion, we developed a novel formalism and corresponding algorithm termed Variational Path Sampling to compute rate constants in nonequilibrium systems by optimally driving the systems to transition between metastable states. The method consists of variationally obtaining an optimal time-dependent force within a chosen basis set and using it to estimate the rate and elucidate the reaction mechanism from direct simulations. We have applied VPS to predict and explain the enhancement of isomerization rate of a passive dimer in an explicit active bath, and demonstrated the role of dissipation in the rate amplification. VPS can be used to compute rates in arbitrary stochastic systems and extends the use of optimal control forces in large deviation sampling to transient rare events [117, 339, 84, 89, 86]. VPS complements trajectory-level importance sampling methods by generating the rare reactive event through a time-series of driving forces instead of a sequence of rare noise histories. We expect this approach to find broad use in rate computations for rare events in dissipative systems throughout the physical sciences and across scales.

Data availability

The source code and data that reproduce the findings of this study are openly available on Zenodo at https://doi.org/10.5281/zenodo.5763101 [347].

Bibliography

[1]   Phillip L Geissler, Christoph Dellago, and David Chandler. "Kinetic pathways of ion pair dissociation in water". In: *The Journal of Physical Chemistry B* 103.18 (1999), pp. 3706–3710.

[2]   Robert A Kuharski et al. "Stochastic molecular dynamics study of cyclohexane isomerization". In: *The Journal of Physical Chemistry* 92.11 (1988), pp. 3261–3267.

[3]   Maria E Gracheva et al. "Simulation of the electric response of DNA translocation through a semiconductor nanopore–capacitor". In: *Nanotechnology* 17.3 (2006), p. 622.

[4]   Joseph E Basconi and Michael R Shirts. "Effects of temperature control algorithms on transport properties and kinetics in molecular dynamics simulations". In: *Journal of chemical theory and computation* 9.7 (2013), pp. 2887–2899.

[5]   Gengping Jiang et al. "Molecular dynamics simulations of the electric double layer capacitance of graphene electrodes in mono-valent aqueous electrolytes". In: *Nano Research* 9.1 (2016), pp. 174–186.

[6]   Luiz Felipe C Pereira and Davide Donadio. "Divergence of the thermal conductivity in uniaxially strained graphene". In: *Physical Review B* 87.12 (2013), p. 125424.

[7]   Filippos D Sofos, Theodoros E Karakasidis, and Antonios Liakopoulos. "Effects of wall roughness on flow in nanochannels". In: *Physical Review E* 79.2 (2009), p. 026305.

[8]   Pablo F Damasceno, Michael Engel, and Sharon C Glotzer. "Predictive self-assembly of polyhedra into complex structures". In: *Science* 337.6093 (2012), pp. 453–457.

[9]   Goundla Srinivas, Dennis E Discher, and Michael L Klein. "Self-assembly and properties of diblock copolymers by coarse-grain molecular dynamics". In: *Nature materials* 3.9 (2004), pp. 638–644.

[10]   Douglas J Tobias, Kechuan Tu, and Michael L Klein. "Atomic-scale molecular dynamics simulations of lipid membranes". In: *Current opinion in colloid & interface science* 2.1 (1997), pp. 15–26.

[11]   Robert Zwanzig. *Nonequilibrium statistical mechanics*. Oxford university press, 2001.

[12] Frank Noé and Cecilia Clementi. "Kinetic distance and kinetic maps from molecular dynamics simulation". In: *Journal of chemical theory and computation* 11.10 (2015), pp. 5002–5011.

[13] David Chandler. "Barrier crossings: classical theory of rare but important events". In: *Classical and Quantum Dynamics in Condensed Phase Simulations* 523 (1998).

[14] John D Chodera et al. "Long-time protein folding dynamics from short-time molecular dynamics simulations". In: *Multiscale Modeling & Simulation* 5.4 (2006), pp. 1214–1226.

[15] Gabriele C Sosso et al. "Crystal nucleation in liquids: Open questions and future challenges in molecular dynamics simulations". In: *Chemical reviews* 116.12 (2016), pp. 7078–7116.

[16] Lisa M Nash et al. "Topological mechanics of gyroscopic metamaterials". In: *Proc. Natl. Acad. Sci. USA* 112.47 (2015), pp. 14495–14500.

[17] Shaon Chakrabarti et al. "Molecular chaperones maximize the native state yield on biological times by driving substrates out of equilibrium". In: *Proc. Natl. Acad. Sci. USA* 114.51 (2017), E10919–E10927.

[18] David Zwicker et al. "Growth and division of active droplets provides a model for protocells". In: *Nat. Phys.* 13.4 (2017), p. 408.

[19] Udo Seifert. "Stochastic thermodynamics, fluctuation theorems and molecular machines". In: *Rep. Prog. Phys.* 75.12 (2012), p. 126001.

[20] Michael E Cates and Julien Tailleur. "Motility-induced phase separation". In: *Annual Reviews of Condensed Matter Physics* 6.1 (2015), pp. 219–244.

[21] Stephen Whitelam. "Strong bonds and far-from-equilibrium conditions minimize errors in lattice-gas growth". In: *The Journal of chemical physics* 149.10 (2018), p. 104902.

[22] Michael Nguyen and Suriyanarayanan Vaikuntanathan. "Design principles for nonequilibrium self-assembly". In: *Proc. Natl. Acad. Sci. USA* 113.50 (2016), pp. 14231–14236.

[23] C Suryanarayana. *Non-equilibrium processing of materials*. Elsevier, 1999.

[24] Jerry P Gollub and James S Langer. "Pattern formation in nonequilibrium physics". In: *Reviews of Modern Physics* 71.2 (1999), S396.

[25] Matthew J Abplanalp et al. "A study of interstellar aldehydes and enols as tracers of a cosmic ray-driven nonequilibrium synthesis of complex organic molecules". In: *Proceedings of the National Academy of Sciences* 113.28 (2016), pp. 7727–7732.

[26] Aharon Katchalsky and Peter F Curran. "Nonequilibrium thermodynamics in biophysics". In: *Nonequilibrium Thermodynamics in Biophysics*. Harvard University Press, 2013.

[27] Bernard Derrida. "Non-equilibrium steady states: fluctuations and large deviations of the density and of the current". In: *Journal of Statistical Mechanics: Theory and Experiment* 2007.07 (2007), P07023.

[28] Tomaž Prosen and Marko Žnidarič. "Matrix product simulations of non-equilibrium steady states of quantum spin chains". In: *Journal of Statistical Mechanics: Theory and Experiment* 2009.02 (2009), P02035.

[29] Igor Rychkov. "Block copolymers under shear flow". In: *Macromol. Theor. Simul.* 14.4 (2005), pp. 207–242.

[30] Nicolas Bachelard et al. "Emergence of an enslaved phononic bandgap in a non-equilibrium pseudo-crystal". In: *Nat. Mater.* 16.8 (2017), p. 808.

[31] Anabella A Abate et al. "Shear-aligned block copolymer monolayers as seeds to control the orientational order in cylinder-forming block copolymer thin films". In: *Macromolecules* 49.19 (2016), pp. 7588–7596.

[32] Grant M Rotskoff and Phillip L Geissler. "Robust nonequilibrium pathways to microcompartment assembly". In: *Proc. Natl. Acad. Sci. USA* 115.25 (2018), pp. 6341–6346.

[33] Andre?s Arango-Restrepo, J Miguel Rubi, and Daniel Barraga?n. "Understanding gelation as a nonequilibrium self-assembly process". In: *The Journal of Physical Chemistry B* 122.18 (2018), pp. 4937–4945.

[34] Hugo Touchette. "The large deviation approach to statistical mechanics". In: *Phys. Rep.* 478.1-3 (2009), pp. 1–69.

[35] Marco Baiesi, Christian Maes, and Bram Wynants. "Fluctuations and response of nonequilibrium states". In: *Physical review letters* 103.1 (2009), p. 010602.

[36] Chloe Ya Gao and David T Limmer. "Nonlinear transport coefficients from large deviation functions". In: *J. Chem. Phys.* 151.1 (2019), p. 014101.

[37] Gavin E Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". In: *Physical Review E* 60.3 (1999), p. 2721.

[38] Christopher Jarzynski. "Nonequilibrium equality for free energy differences". In: *Physical Review Letters* 78.14 (1997), p. 2690.

[39] Giovanni Gallavotti and Ezechiel Godert David Cohen. "Dynamical ensembles in stationary states". In: *Journal of Statistical Physics* 80.5 (1995), pp. 931–970.

[40] Andre C Barato and Udo Seifert. "Thermodynamic uncertainty relation for biomolecular processes". In: *Physical review letters* 114.15 (2015), p. 158101.

[41] Todd R Gingrich et al. "Dissipation bounds all steady-state current fluctuations". In: *Physical review letters* 116.12 (2016), p. 120601.

[42] Daan Frenkel and Berend Smit. *Understanding molecular simulation: from algorithms to applications.* Vol. 1. Elsevier, 2001.

[43] Hannes Risken. "Fokker-planck equation". In: *The Fokker-Planck Equation.* Springer, 1996, pp. 63–95.

[44]   Jean-Pierre Hansen and Ian Ranald McDonald. *Theory of simple liquids: with applications to soft matter*. Academic press, 2013.

[45]   Lars Onsager and S Machlup. "Fluctuations and irreversible processes". In: *Phys. Rev.* 91.6 (1953), p. 1505.

[46]   Hugo Touchette. "A basic introduction to large deviations: Theory, applications, simulations". In: *arXiv preprint arXiv:1106.4146* (2011).

[47]   L Chris G Rogers and David Williams. *Diffusions, Markov processes and martingales: Volume 2, Itô calculus*. Vol. 2. Cambridge university press, 2000.

[48]   Trevor GrandPre et al. "Entropy production fluctuations encode collective behavior in active matter". In: *Physical Review E* 103.1 (2021), p. 012613.

[49]   Stephen Whitelam, Katherine Klymko, and Dibyendu Mandal. "Phase separation and large deviations of lattice active matter". In: *The Journal of chemical physics* 148.15 (2018), p. 154902.

[50]   Yann-Edwin Keta et al. "Collective motion in large deviations of active particles". In: *Physical Review E* 103.2 (2021), p. 022603.

[51]   R. L. Jack and P. Sollich. "Large Deviations and Ensembles of Trajectories in Stochastic Models". In: *Prog. Theor. Phys. Supp.* 184 (2010), pp. 304–317.

[52]   Robert L. Jack. "Ergodicity and large deviations in physical systems with stochastic dynamics". In: *The European Physical Journal B* (2020).

[53]   Mauro Merolle, Juan P Garrahan, and David Chandler. "Space–time thermodynamics of the glass transition". In: *Proc. Natl. Acad. Sci. USA* 102.31 (2005), pp. 10837–10840.

[54]   David Chandler and Juan P Garrahan. "Dynamics on the way to forming glass: Bubbles in space-time". In: *Annu. Rev. Phys. Chem.* 61 (2010), pp. 191–217.

[55]   Tanguy Laffargue et al. "Large deviations of Lyapunov exponents". In: *J. Phys. A* 46.25 (2013), p. 254002.

[56]   Pavel V Kuptsov and Antonio Politi. "Large-deviation approach to space-time chaos". In: *Phys. Rev. Lett.* 107.11 (2011), p. 114101.

[57]   Carlos Pérez-Espigares and Pablo I. Hurtado. "Sampling rare events across dynamical phase transitions". In: *Chaos* 29.8 (2019), p. 083106.

[58]   Chloe Ya Gao and David T Limmer. "Transport coefficients from large deviation functions". In: *Entropy* 19.11 (2017), p. 571.

[59]   David T Limmer, Chloe Y Gao, and Anthony R Poggioli. "A large deviation theory perspective on nanoscale transport phenomena". In: *The European Physical Journal B* 94.7 (2021), pp. 1–16.

[60]   David Andrieux and Pierre Gaspard. "Fluctuation theorem and Onsager reciprocity relations". In: *J. Chem. Phys.* 121.13 (2004), pp. 6167–6174.

[61] Pierre Gaspard. "Multivariate fluctuation relations for currents". In: *New J. Phys.* 15.11 (2013), p. 115014.

[62] Hugo Touchette. "Introduction to dynamical large deviations of Markov processes". In: *arXiv:1705.06492* (2017).

[63] Raphaël Chetrite and Hugo Touchette. "Nonequilibrium Markov processes conditioned on large deviations". In: *Ann. Inst. Henri Poincaré.* Vol. 16. 9. Springer. 2015, pp. 2005–2057.

[64] Baron Peters. *Reaction rate theory and rare events.* Elsevier, 2017.

[65] D Richard, H Lowen, and T Speck. "Nucleation pathway and kinetics of phase-separating active Brownian particles". In: *Soft Matter* 12.4 (2016), pp. 5257–5264. DOI: https://doi.org/10.1039/C6SM00485G.

[66] Jochen P Müller et al. "Force sensing by the vascular protein von Willebrand factor is tuned by a strong intermonomer interaction". In: *Proceedings of the National Academy of Sciences* 113.5 (2016), pp. 1208–1213.

[67] Rosalind J Allen, Daan Frenkel, and Pieter Rein ten Wolde. "Simulating rare events in equilibrium or nonequilibrium stochastic systems". In: *The Journal of chemical physics* 124.2 (2006), p. 024102.

[68] Mark Iosifovich Freidlin and Alexander D Wentzell. "Random perturbations". In: *Random perturbations of dynamical systems.* Springer, 1998, pp. 15–43.

[69] Tobias Grafke and Eric Vanden-Eijnden. "Numerical computation of rare events via large deviation theory". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 29.6 (2019), p. 063118.

[70] Peter G Bolhuis et al. "Transition path sampling: Throwing ropes over rough mountain passes, in the dark". In: *Annu. Rev. Phys. Chem.* 53.1 (2002), pp. 291–318.

[71] Omar Valsson and Michele Parrinello. "Variational approach to enhanced sampling and free energy calculations". In: *Phys. Rev. Lett.* 113.9 (2014), p. 090601.

[72] Christoph Dellago et al. "Transition path sampling and the calculation of rate constants". In: *J. Chem. Phys.* 108.5 (1998), pp. 1964–1977.

[73] Addison J Schile and David T Limmer. "Rate constants in spatially inhomogeneous systems". In: *The Journal of chemical physics* 150.19 (2019), p. 191102.

[74] Benjamin Kuznets-Speck and David T. Limmer. "Dissipation bounds the amplification of transition rates far from equilibrium". In: *Proceedings of the National Academy of Sciences* 118.8 (2021).

[75] Gianmaria Falasco and Massimiliano Esposito. "Dissipation-time uncertainty relation". In: *Physical Review Letters* 125.12 (2020), p. 120604.

[76] Robert M Turner, Thomas Speck, and Juan P Garrahan. "Meta-work and the analogous Jarzynski relation in ensembles of dynamical trajectories". In: *Journal of Statistical Mechanics: Theory and Experiment* 2014.9 (2014), P09017.

[77] Cristian Giardina, Jorge Kurchan, and Luca Peliti. "Direct evaluation of large-deviation functions". In: *Phys. Rev. Lett.* 96.12 (2006), p. 120603.

[78] Cristian Giardina et al. "Simulating rare events in dynamical processes". In: *Journal of statistical physics* 145.4 (2011), pp. 787–811. DOI: `10.1007/s10955-011-0350-4`.

[79] Martin Tchernookov and Aaron R Dinner. "A list-based algorithm for evaluation of large deviation functions". In: *J. Stat. Mech. Theory Exp.* 2010.02 (2010), P02006.

[80] Aaron R Dinner et al. "Trajectory stratification of stochastic dynamics". In: *SIAM Rev.* 60.4 (2018), pp. 909–938.

[81] Ushnish Ray, Garnet Kin-Lic Chan, and David T Limmer. "Importance sampling large deviations in nonequilibrium steady states. I". In: *J. Chem. Phys.* 148.12 (2018), p. 124120.

[82] Todd R Gingrich and Phillip L Geissler. "Preserving correlations between trajectories for efficient path sampling". In: *J. Chem. Phys.* 142.23 (2015), 06B614_1.

[83] Esteban Guevara Hidalgo, Takahiro Nemoto, and Vivien Lecomte. "Finite-time and finite-size scalings in the evaluation of large-deviation functions: Numerical approach in continuous time". In: *Phys. Rev. E* 95.6 (2017), p. 062134.

[84] Ushnish Ray, Garnet Kin-Lic Chan, and David T Limmer. "Exact fluctuations of nonequilibrium steady states from approximate auxiliary dynamics". In: *Phys. Rev. Lett.* 120.21 (2018), p. 210602.

[85] Takahiro Nemoto and Shin-ichi Sasa. "Computation of large deviation statistics via iterative measurement-and-feedback procedure". In: *Phys. Rev. Lett.* 112.9 (2014), p. 090602.

[86] Takahiro Nemoto et al. "Population-dynamics method with a multicanonical feedback control". In: *Physical Review E* 93.6 (2016), p. 062123. DOI: `10.1103/PhysRevE.93.062123`. URL: `https://arxiv.org/abs/1601.06648`.

[87] Daniel Jacobson and Stephen Whitelam. "Direct evaluation of dynamical large-deviation rate functions using a variational ansatz". In: *Physical Review E* 100.5 (2019), p. 052139.

[88] Stephen Whitelam, Daniel Jacobson, and Isaac Tamblyn. "Evolutionary reinforcement learning of dynamical large deviations". In: *J. Chem. Phys.* 153.4 (2020), p. 044113. DOI: `10.1063/5.0015301`. eprint: `https://doi.org/10.1063/5.0015301`. URL: `https://doi.org/10.1063/5.0015301`.

[89] Jakub Dolezal and Robert L Jack. "Large deviations and optimal control forces for hard particles in one dimension". In: *Journal of Statistical Mechanics: Theory and Experiment* 2019.12 (2019), p. 123208.

[90] Carsten Hartmann and Christof Schütte. "Efficient rare event simulation by optimal nonequilibrium forcing". In: *J. Stat. Mech. Theory Exp.* 2012.11 (2012), P11004.

[91] Jannes Quer et al. "An automatic adaptive importance sampling algorithm for molecular dynamics in reaction coordinates". In: *SIAM J. Sci. Comput.* 40.2 (2018), A653–A670.

[92] Takahiro Nemoto, Robert L Jack, and Vivien Lecomte. "Finite-size scaling of a first-order dynamical phase transition: Adaptive population dynamics and an effective model". In: *Phys. Rev. Lett.* 118.11 (2017), p. 115702.

[93] Grégoire Ferré and Hugo Touchette. "Adaptive sampling of large deviations". In: *J. Stat. Phys.* 172.6 (2018), pp. 1525–1544.

[94] Andreas Härtel, David Richard, and Thomas Speck. "Three-body correlations and conditional forces in suspensions of active hard disks". In: *Phys. Rev. E* 97.1 (2018), p. 012606.

[95] D Simon, V Popkov, and G Schütz. "Asymmetric simple exclusion process on a ring conditioned on enhanced flux". In: *J. Stat. Mech. Theory Exp.* (2010), P10007.

[96] Jeffrey K Weber et al. "Dynamical phase transitions reveal amyloid-like states on protein folding landscapes". In: *Biophys. J.* 107.4 (2014), pp. 974–982.

[97] Gary A Huber and Sangtae Kim. "Weighted-ensemble Brownian dynamics simulations for protein association reactions". In: *Biophysical journal* 70.1 (1996), pp. 97–110.

[98] Daniel M Zuckerman and Lillian T Chong. "Weighted ensemble simulation: review of methodology, applications, and software". In: *Annual review of biophysics* 46 (2017), pp. 43–57.

[99] Christoph Dellago, Peter G Bolhuis, and David Chandler. "Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements". In: *J. Chem. Phys.* 108.22 (1998), pp. 9236–9245.

[100] Gavin E Crooks and David Chandler. "Efficient transition path sampling for nonequilibrium stochastic dynamics". In: *Physical Review E* 64.2 (2001), p. 026109.

[101] Chantal Valeriani et al. "Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling". In: *J. Chem. Phys.* 127.11 (2007), p. 114109.

[102] Rosalind J Allen, Chantal Valeriani, and Pieter Rein Ten Wolde. "Forward flux sampling for rare event simulations". In: *Journal of physics: Condensed matter* 21.46 (2009), p. 463102.

[103] Aryeh Warmflash, Prabhakar Bhimalapuram, and Aaron R Dinner. "Umbrella sampling for nonequilibrium processes". In: *The Journal of chemical physics* 127.15 (2007), p. 114109.

[104] Alex Dickson, Aryeh Warmflash, and Aaron R Dinner. "Nonequilibrium umbrella sampling in spaces of many order parameters". In: *J. Chem. Phys.* 130.7 (2009), 02B605.

[105] Grant M Rotskoff, Andrew R Mitchell, and Eric Vanden-Eijnden. "Active Importance Sampling for Variational Objectives Dominated by Rare Events: Consequences for Optimization and Generalization". In: *arXiv preprint arXiv:2008.06334* (2020).

[106] Grant M. Rotskoff and Eric Vanden-Eijnden. *Learning with rare data: Using active importance sampling to optimize objectives dominated by rare events*. 2020. arXiv: `2008.06334 [physics.data-an]`.

[107] Hao Wang and Ron Elber. "Milestoning with wind: Exploring the impact of a biasing potential in exact calculation of kinetics". In: *The Journal of Chemical Physics* 152.22 (2020), p. 224105.

[108] Surl-Hee Ahn et al. "Gaussian-Accelerated Molecular Dynamics with the Weighted Ensemble Method: A Hybrid Method Improves Thermodynamic and Kinetic Sampling". In: *Journal of Chemical Theory and Computation* 17.12 (2021), pp. 7938–7951.

[109] Satya N Majumdar and Henri Orland. "Effective Langevin equations for constrained stochastic processes". In: *Journal of Statistical Mechanics: Theory and Experiment* 2015.6 (2015), P06039.

[110] Robert L Jack and Peter Sollich. "Effective interactions and large deviations in stochastic processes". In: *Eur. Phys. J. Spec. Top.* 224.12 (Sept. 2015), pp. 2351–2367. ISSN: 1951-6401. DOI: `10.1140/epjst/e2015-02416-9`.

[111] Raphaël Chetrite and Hugo Touchette. "Variational and optimal control representations of conditioned and driven processes". In: *J. Stat. Mech. Theory Exp.* 2015.12 (2015), P12001.

[112] Francis J Alexander and Gregory L Eyink. "Rayleigh-Ritz calculation of effective potential far from equilibrium". In: *Phys. Rev. Lett.* 78.1 (1997), p. 1.

[113] Takahiro Nemoto and Shin-ichi Sasa. "Thermodynamic formula for the cumulant generating function of time-averaged current". In: *Phys. Rev. E* 84.6 (2011), p. 061113.

[114] Takahiro Nemoto and Shin-ichi Sasa. "Variational formula for experimental determination of high-order correlations of current fluctuations in driven systems". In: *Phys. Rev. E* 83.3 (2011), p. 030105.

[115] Monroe D Donsker and SR Srinivasa Varadhan. "Asymptotic evaluation of certain Markov process expectations for large time, I". In: *Commun. Pure Appl. Math.* 28.1 (1975), pp. 1–47.

[116] David Ceperley and Berni Alder. "Quantum monte carlo". In: *Science* 231.4738 (1986), pp. 555–560.

[117] A. Das and D. T. Limmer. "Variational control forces for enhanced sampling of nonequilibrium molecular dynamics simulations." In: *Journal of Chemical Physics* 151.24 (2019), p. 244123. DOI: `https://doi.org/10.1063/1.5128956`.

[118] Raphaël Chetrite and Hugo Touchette. "Nonequilibrium microcanonical and canonical ensembles and their equivalence". In: *Physical review letters* 111.12 (2013), p. 120601.

[119] Joseph L Doob. *Classical potential theory and its probabilistic counterpart: Advanced problems*. Vol. 262. Springer Science & Business Media, 2012.

[120] Mari Carmen Bañuls and Juan P Garrahan. "Using matrix product states to study the dynamical large deviations of kinetically constrained models". In: *arXiv preprint arXiv:1903.01570* (2019).

[121] Phillip Helms, Ushnish Ray, and Garnet Kin-Lic Chan. "Dynamical phase behavior of the single- and multi-lane asymmetric simple exclusion process via matrix product states". In: *Phys. Rev. E* 100 (2 Aug. 2019), p. 022101.

[122] Patrick Warren and Rosalind Allen. "Malliavin Weight Sampling: A Practical Guide". In: *Entropy* 16.1 (Dec. 2013), pp. 221–232. ISSN: 1099-4300. DOI: `10.3390/e16010221`. URL: `http://dx.doi.org/10.3390/e16010221`.

[123] Patrick B Warren and Rosalind J Allen. "Malliavin weight sampling for computing sensitivity coefficients in brownian dynamics simulations". In: *Phys. Rev. Lett.* 109.25 (2012), p. 250601.

[124] Ryogo Kubo, Morikazu Toda, and Natsuki Hashitsume. *Statistical physics II: nonequilibrium statistical mechanics*. Vol. 31. Springer Science & Business Media, 2012.

[125] Dominic C Rose, Jamie F Mair, and Juan P Garrahan. "A reinforcement learning approach to rare trajectory sampling". In: *New Journal of Physics* 23.1 (2021), p. 013013.

[126] Yurii E Nesterov. "A method for solving the convex programming problem with convergence rate O $(1/k\hat{\ } 2)$". In: *Sov. Math. Dokl.* Vol. 27. 1983, pp. 372–377.

[127] Ilya Sutskever et al. "On the importance of initialization and momentum in deep learning". In: *ICML*. 2013, pp. 1139–1147.

[128] Ryota Shinagawa and Kazuo Sasaki. "Enhanced diffusion of molecular motors in the presence of adenosine triphosphate and external force". In: *J. Phys. Soc. Jpn.* 85.6 (2016), p. 064004.

[129] Xiao-guang Ma et al. "Colloidal dynamics over a tilted periodic potential: Forward and reverse transition probabilities and entropy production in a nonequilibrium steady state". In: *Phys. Rev. E* 96.1 (1 July 2017), p. 012601. DOI: `10.1103/PhysRevE.96.012601`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.96.012601`.

[130] Lukas P Fischer, Patrick Pietzonka, and Udo Seifert. "Large deviation function for a driven underdamped particle in a periodic potential". In: *Phys. Rev. E* 97.2 (2 Feb. 2018), p. 022143. DOI: `10.1103/PhysRevE.97.022143`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.97.022143`.

[131] Jorge Kurchan. "Fluctuation theorem for stochastic dynamics". In: *Journal of Physics A: Mathematical and General* 31.16 (1998), p. 3719.

[132] Pelerine Tsobgni Nyawo and Hugo Touchette. "Large deviations of the current for driven periodic diffusions". In: *Phys. Rev. E* 94.3 (2016), p. 032101.

[133] Robert L Jack, Ian R Thompson, and Peter Sollich. "Hyperuniformity and phase separation in biased ensembles of trajectories for diffusive systems". In: *Phys. Rev. Lett.* 114.6 (2015), p. 060601.

[134] E Autieri et al. "Dominant reaction pathways in high-dimensional systems". In: *J. Chem. Phys.* 130.6 (2009), p. 064106.

[135] Christian Maes. "Frenesy". In: *arXiv preprint arXiv:1904.10485* (2019).

[136] Estelle Pitard, Vivien Lecomte, and Frédéric Van Wijland. "Dynamic transition in an atomic glass former: A molecular-dynamics evidence". In: *Europhys. Lett.* 96.5 (2011), p. 56002.

[137] Christopher J Fullerton and Robert L Jack. "Dynamical phase transitions in super-cooled liquids: Interpreting measurements of dynamical activity". In: *J. Chem. Phys.* 138.22 (2013), p. 224506.

[138] Eric Neuscamman. "The Jastrow antisymmetric geminal power in Hilbert space: Theory, benchmarking, and application to a novel transition state". In: *J. Chem. Phys.* 139.19 (2013), p. 194105.

[139] Bernard Derrida. "An exactly soluble non-equilibrium system: the asymmetric simple exclusion process". In: *Physics Reports* 301.1-3 (1998), pp. 65–83.

[140] Joachim Krug. "Boundary-induced phase transitions in driven diffusive systems". In: *Physical review letters* 67.14 (1991), p. 1882.

[141] Andrea Parmeggiani, Thomas Franosch, and Erwin Frey. "Phase coexistence in driven one-dimensional transport". In: *Physical review letters* 90.8 (2003), p. 086601.

[142] V Popkov and GM Schütz. "Transition probabilities and dynamic structure function in the ASEP conditioned on strong flux". In: *J. Stat. Phys.* 142.3 (2011), pp. 627–639.

[143] Phillip Helms and Garnet Kin-Lic Chan. "Dynamical phase transitions in a 2D classical nonequilibrium model via 2D tensor networks". In: *Physical Review Letters* 125.14 (2020), p. 140601.

[144] Andreas Schadschneider, Debashish Chowdhury, and Katsuhiro Nishinari. *Stochastic transport in complex systems: from molecules to vehicles.* Elsevier, 2010.

[145] Carolyn T MacDonald, Julian H Gibbs, and Allen C Pipkin. "Kinetics of biopolymer-ization on nucleic acid templates". In: *Biopolymers: Original Research on Biomolecules* 6.1 (1968), pp. 1–25.

[146] Anatoly B Kolomeisky. "Motor proteins and molecular motors: how to operate machines at the nanoscale". In: *Journal of Physics: Condensed Matter* 25.46 (2013), p. 463101.

[147] Bertil Hille. "Ionic channels in excitable membranes. Current problems and biophysical approaches". In: *Biophysical Journal* 22.2 (1978), pp. 283–294.

[148] Debasish Chaudhuri and Abhishek Dhar. "Stochastic pump of interacting particles". In: *EPL (Europhysics Letters)* 94.3 (2011), p. 30006.

[149] Debasish Chaudhuri, Archishman Raju, and Abhishek Dhar. "Pumping single-file colloids: Absence of current reversal". In: *Physical Review E* 91.5 (2015), p. 050103.

[150] Dominik Lips, Artem Ryabov, and Philipp Maass. "Brownian asymmetric simple exclusion process". In: *Phys. Rev. Lett.* 121.16 (2018), p. 160601.

[151] Dominik Lips, Artem Ryabov, and Philipp Maass. "Single-file transport in periodic potentials: The Brownian asymmetric simple exclusion process". In: *Physical Review E* 100.5 (2019), p. 052121.

[152] Jiawei Yan, Hugo Touchette, Grant M Rotskoff, et al. "Learning nonequilibrium control forces to characterize dynamical phase transitions". In: *Physical Review E* 105.2 (2022), p. 024115.

[153] Dvira Segal, Andrew J Millis, and David R Reichman. "Numerically exact path-integral simulation of nonequilibrium quantum transport and dissipation". In: *Phys. Rev. B* 82.20 (2010), p. 205323.

[154] Addison J Schile and David T Limmer. "Studying rare nonadiabatic dynamics with transition path sampling quantum jump trajectories". In: *J. Chem. Phys.* 149.21 (2018), p. 214109.

[155] Atilim Gunes Baydin et al. "Automatic differentiation in machine learning: a survey". In: *J. Mach. Learn. Res.* 18.153 (2018).

[156] Luigi Bonati, Yue-Yu Zhang, and Michele Parrinello. "Neural networks-based variationally enhanced sampling". In: *Proceedings of the National Academy of Sciences* 116.36 (2019), pp. 17641–17647.

[157] Carsten Hartmann, Christof Schütte, and Wei Zhang. "Model reduction algorithms for optimal control and importance sampling of diffusions". In: *Nonlinearity* 29.8 (2016), p. 2298.

[158] James F Dama et al. "The theory of ultra-coarse-graining. 1. General principles". In: *J. Chem. Theory Comput.* 9.5 (2013), pp. 2466–2480.

[159] Aram Davtyan et al. "The theory of ultra-coarse-graining. 2. Numerical implementation". In: *J. Chem. Theory Comput.* 10.12 (2014), pp. 5265–5275.

[160] James F Dama, Jaehyeok Jin, and Gregory A Voth. "The theory of ultra-coarse-graining. 3. Coarse-grained sites with rapid local equilibrium of internal states". In: *J. Chem. Theory Comput.* 13.3 (2017), pp. 1010–1022.

[161] Salvatore Torquato. "Inverse optimization techniques for targeted self-assembly". In: *Soft Matter* 5.6 (2009), pp. 1157–1173.

[162] Marc Z Miskin et al. "Turning statistical physics models into materials design engines". In: *Proc. Natl. Acad. Sci. USA* 113.1 (2016), pp. 34–39.

[163] Eran Rabani et al. "Drying-mediated self-assembly of nanoparticles". In: *Nature* 426.6964 (2003), pp. 271–274.

[164] Joakim Stenhammar et al. "Continuum theory of phase separation kinetics for active Brownian particles". In: *Physical review letters* 111.14 (2013), p. 145702.

[165] Alexandros Chremos, Kelly Margaritis, and Athanassios Z Panagiotopoulos. "Ultra thin films of diblock copolymers under shear". In: *Soft Matter* 6.15 (2010), pp. 3588–3595.

[166] John Toner, Yuhai Tu, and Sriram Ramaswamy. "Hydrodynamics and phases of flocks". In: *Annals of Physics* 318.1 (2005), pp. 170–244.

[167] Yuting I Li and Michael E Cates. "Non-equilibrium phase separation with reactions: a canonical model and its behaviour". In: *Journal of Statistical Mechanics: Theory and Experiment* 2020.5 (2020), p. 053206.

[168] Qingbao Guan et al. "Fabrication of in situ nanofiber-reinforced molecular composites by nonequilibrium self-assembly". In: *ACS applied materials & interfaces* 10.45 (2018), pp. 39293–39306.

[169] Aleksey S Tsipotan et al. "Laser-induced wavelength-controlled self-assembly of colloidal quasi-resonant quantum dots". In: *Optics express* 24.10 (2016), pp. 11145–11150.

[170] Maelenn Chevreuil et al. "Nonequilibrium self-assembly dynamics of icosahedral viral capsids packaging genome or polyelectrolyte". In: *Nature communications* 9.1 (2018), pp. 1–9.

[171] Michael F Hagan and David Chandler. "Dynamic pathways for viral capsid assembly". In: *Biophysical journal* 91.1 (2006), pp. 42–54.

[172] Stephen Whitelam and Robert L Jack. "The statistical mechanics of dynamic pathways to self-assembly". In: *Annual review of physical chemistry* 66 (2015), pp. 143–163.

[173] Stephen Whitelam et al. "The role of collective motion in examples of coarsening and self-assembly". In: *Soft Matter* 5.6 (2009), pp. 1251–1262.

[174] Arthur C Newton et al. "Rotational diffusion affects the dynamical self-assembly pathways of patchy particles". In: *Proceedings of the National Academy of Sciences* 112.50 (2015), pp. 15308–15313.

[175] Sahand Hormoz and Michael P Brenner. "Design principles for self-assembly with short-range interactions". In: *Proceedings of the National Academy of Sciences* 108.13 (2011), pp. 5193–5198.

[176] Sylvain Vauthey et al. "Molecular self-assembly of surfactant-like peptides to form nanotubes and nanovesicles". In: *Proceedings of the National Academy of Sciences* 99.8 (2002), pp. 5355–5360.

[177] Shuguang Zhang et al. "Design of nanostructured biological materials through self-assembly of peptides and proteins". In: *Current opinion in chemical biology* 6.6 (2002), pp. 865–871.

[178] Michael A Boles, Michael Engel, and Dmitri V Talapin. "Self-assembly of colloidal nanocrystals: from intricate structures to functional materials". In: *Chemical reviews* 116.18 (2016), pp. 11220–11289.

[179] Chad A Mirkin et al. "A DNA-based method for rationally assembling nanoparticles into macroscopic materials". In: *Nature* 382.6592 (1996), pp. 607–609.

[180] Dmitri V Talapin et al. "Quasicrystalline order in self-assembled binary nanoparticle superlattices". In: *Nature* 461.7266 (2009), pp. 964–967.

[181] Daphne Klotsa and Robert L Jack. "Controlling crystal self-assembly using a real-time feedback scheme". In: *The Journal of Chemical Physics* 138.9 (2013), p. 094502.

[182] Christopher J Fullerton and Robert L Jack. "Optimising self-assembly through time-dependent interactions". In: *The Journal of Chemical Physics* 145.24 (2016), p. 244505.

[183] Bartosz A Grzybowski et al. "From dynamic self-assembly to networked chemical systems". In: *Chemical Society Reviews* 46.18 (2017), pp. 5647–5678.

[184] Giulio Ragazzon and Leonard J Prins. "Energy consumption in chemical fuel-driven self-assembly". In: *Nature nanotechnology* 13.10 (2018), pp. 882–889.

[185] Laura Heinen and Andreas Walther. "Programmable dynamic steady states in ATP-driven nonequilibrium DNA systems". In: *Science advances* 5.7 (2019), eaaw0590.

[186] Michael Grünwald et al. "Exploiting non-equilibrium phase separation for self-assembly". In: *Soft Matter* 12.5 (2016), pp. 1517–1524.

[187] Avishek Das and David T Limmer. "Variational design principles for nonequilibrium colloidal assembly". In: *The Journal of chemical physics* 154.1 (2021), p. 014107.

[188] AW Lees and SF Edwards. "The computer study of transport processes under extreme conditions". In: *Journal of Physics C: Solid State Physics* 5.15 (1972), p. 1921.

[189] W Benjamin Rogers and John C Crocker. "Direct measurements of DNA-mediated colloidal interactions and their quantitative modeling". In: *Proceedings of the National Academy of Sciences* 108.38 (2011), pp. 15687–15692.

[190] W Benjamin Rogers and Vinothan N Manoharan. "Programming colloidal phase transitions with DNA strand displacement". In: *Science* 347.6222 (2015), pp. 639–642.

[191] Vinothan N Manoharan. "Colloidal matter: Packing, geometry, and entropy". In: *Science* 349.6251 (2015).

[192] Guangnan Meng et al. "The free-energy landscape of clusters of attractive hard spheres". In: *Science* 327.5965 (2010), pp. 560–563.

[193] Zorana Zeravcic, Vinothan N Manoharan, and Michael P Brenner. "Size limits of self-assembled colloidal structures made using specific interactions". In: *Proceedings of the National Academy of Sciences* 111.45 (2014), pp. 15918–15923.

[194] Weige Xue and Gary S Grest. "Shear-induced alignment of colloidal particles in the presence of a shear flow". In: *Physical review letters* 64.4 (1990), p. 419.

[195] Xiang Cheng et al. "Assembly of vorticity-aligned hard-sphere colloidal strings in a simple shear flow". In: *Proceedings of the National Academy of Sciences* 109.1 (2012), pp. 63–67.

[196] Yingying Dou et al. "A review on self-assembly in microfluidic devices". In: *Journal of Micromechanics and Microengineering* 27.11 (2017), p. 113002.

[197] Arash Nikoubashman. "Self-assembly of colloidal micelles in microfluidic channels". In: *Soft Matter* 13.1 (2017), pp. 222–229.

[198] Søren Toxvaerd and Jeppe C Dyre. *Communication: Shifted forces in molecular dynamics.* 2011.

[199] Trevor GrandPre and David T Limmer. "Current fluctuations of interacting active Brownian particles". In: *Phys. Rev. E* 98.6 (2018), p. 060601.

[200] J Dana Honeycutt and Hans C Andersen. "Molecular dynamics study of melting and freezing of small Lennard-Jones clusters". In: *Journal of Physical Chemistry* 91.19 (1987), pp. 4950–4963.

[201] Grant M Rotskoff, Gavin E Crooks, and Eric Vanden-Eijnden. "Geometric approach to optimal nonequilibrium control: Minimizing dissipation in nanomagnetic spin systems". In: *Physical Review E* 95.1 (2017), p. 012148.

[202] James C Spall. *Introduction to stochastic search and optimization: estimation, simulation, and control.* Vol. 65. John Wiley & Sons, 2005.

[203] Natalie Arkus. *Theoretical approaches to self-assembly and biology.* Harvard University, 2009.

[204] Natalie Arkus, Vinothan N Manoharan, and Michael P Brenner. "Minimal energy clusters of hard spheres with short range attractions". In: *Physical review letters* 103.11 (2009), p. 118303.

[205] Fabian M Hecht and Andreas R Bausch. "Kinetically guided colloidal structure formation". In: *Proceedings of the National Academy of Sciences* 113.31 (2016), pp. 8577–8582.

[206] M Hermes et al. "Nucleation of colloidal crystals on configurable seed structures". In: *Soft Matter* 7.10 (2011), pp. 4623–4628.

[207] Hai Pham-Van et al. "Two-dimensional clusters of colloidal particles induced by emulsion droplet evaporation". In: *Nanomaterials* 10.1 (2020), p. 156.

[208] Ran Niu, Thomas Palberg, Thomas Speck, et al. "Self-assembly of colloidal molecules due to self-generated flow". In: *Physical Review Letters* 119.2 (2017), p. 028001.

[209] Rebecca W Perry et al. "Two-dimensional clusters of colloidal spheres: Ground states, excited states, and structural rearrangements". In: *Physical review letters* 114.22 (2015), p. 228301.

[210] Marco Baiesi, Christian Maes, and Bram Wynants. "Nonequilibrium linear response for Markov dynamics, I: jump processes and overdamped diffusions". In: *Journal of statistical physics* 137.5-6 (2009), p. 1094.

[211] Thomas Speck, Jakob Mehl, and Udo Seifert. "Role of external flow and frame invariance in stochastic thermodynamics". In: *Physical review letters* 100.17 (2008), p. 178302.

[212] Kiryl Asheichyk et al. "Response of active Brownian particles to shear flow". In: *The Journal of chemical physics* 150.14 (2019), p. 144111.

[213] A Imperio and L Reatto. "Microphase separation in two-dimensional systems with competing interactions". In: *The Journal of chemical physics* 124.16 (2006), p. 164712.

[214] Donald LD Caspar and Aaron Klug. "Physical principles in the construction of regular viruses". In: *Cold Spring Harbor symposia on quantitative biology*. Vol. 27. Cold Spring Harbor Laboratory Press. 1962, pp. 1–24.

[215] Shiho Tanaka et al. "Atomic-level models of the bacterial carboxysome shell". In: *science* 319.5866 (2008), pp. 1083–1086.

[216] Erica M Hildebrand and Job Dekker. "Mechanisms and functions of chromosome compartmentalization". In: *Trends in Biochemical Sciences* 45.5 (2020), pp. 385–396.

[217] Peter Fratzl. "Cellulose and collagen: from fibres to tissues". In: *Current opinion in colloid & interface science* 8.1 (2003), pp. 32–39.

[218] Michael F Hagan and Gregory M Grason. "Equilibrium mechanisms of self-limiting assembly". In: *Reviews of Modern Physics* 93.2 (2021), p. 025008.

[219] Michael Grünwald and Phillip L Geissler. "Patterns without patches: Hierarchical self-assembly of complex structures from simple building blocks". In: *ACS nano* 8.6 (2014), pp. 5891–5897.

[220] Debapriya Banerjee et al. "Assembly of particle strings via isotropic potentials". In: *The Journal of Chemical Physics* 150.12 (2019), p. 124903.

[221] Joakim Stenhammar et al. "Phase behaviour of active Brownian particles: the role of dimensionality". In: *Soft matter* 10.10 (2014), pp. 1489–1499.

[222] Carl P Goodrich et al. "Designing self-assembling kinetics with differentiable statistical physics models". In: *Proceedings of the National Academy of Sciences* 118.10 (2021).

[223] Alex Albaugh and Todd R Gingrich. "Simulating a chemically-fueled molecular motor with nonequilibrium molecular dynamics". In: *arXiv preprint arXiv:2102.06298* (2021).

[224] Joseph NE Lucero, Aliakbar Mehdizadeh, and David A Sivak. "Optimal control of rotary motors". In: *Physical Review E* 99.1 (2019), p. 012119.

[225] Todd R Gingrich et al. "Efficiency and large deviations in time-asymmetric stochastic heat engines". In: *New Journal of Physics* 16.10 (2014), p. 102003.

[226] Aidan I Brown and David A Sivak. "Theory of nonequilibrium free energy transduction by molecular machines". In: *Chemical reviews* 120.1 (2019), pp. 434–459.

[227] Emma Lathouwers, Joseph NE Lucero, and David A Sivak. "Nonequilibrium energy transduction in stochastic strongly coupled rotary motors". In: *The Journal of Physical Chemistry Letters* 11.13 (2020), pp. 5273–5278.

[228] Zachary M Sherman et al. "Inverse methods for design of soft materials". In: *The Journal of Chemical Physics* 152.14 (2020), p. 140902.

[229] Stephen Whitelam, Lester O Hedges, and Jeremy D Schmit. "Self-assembly at a nonequilibrium critical point". In: *Physical review letters* 112.15 (2014), p. 155504.

[230] Andrew W Long and Andrew L Ferguson. "Nonlinear machine learning of patchy colloid self-assembly pathways and mechanisms". In: *The Journal of Physical Chemistry B* 118.15 (2014), pp. 4228–4244.

[231] Andrew W Long et al. "Machine learning assembly landscapes from particle tracking data". In: *Soft Matter* 11.41 (2015), pp. 8141–8153.

[232] Xun Tang et al. "Optimal feedback controlled assembly of perfect crystals". In: *ACS nano* 10.7 (2016), pp. 6791–6798.

[233] Stephen Whitelam and Isaac Tamblyn. "Learning to grow: Control of material self-assembly using evolutionary reinforcement learning". In: *Physical Review E* 101.5 (2020), p. 052604.

[234] Stephen Whitelam et al. "Correspondence between neuroevolution and gradient descent". In: *Nature communications* 12.1 (2021), pp. 1–10.

[235] Avishek Das and David T. Limmer. *Variational Design Principles For Nonequilibrium Colloidal Assembly.* Version v1.0. 1. Zenodo, Nov. 2020. DOI: 10.5281/zenodo.4289235.

[236] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction.* 2nd. MIT Press, 2018.

[237] Rémi Munos and Paul Bourgine. "Reinforcement Learning for Continuous Stochastic Control Problems". In: *Adv. Neural. Inf. Process. Syst.* Ed. by M. Jordan, M. Kearns, and S. Solla. Vol. 10. MIT Press, 1998. URL: https://proceedings.neurips.cc/paper/1997/file/186a157b2992e7daed3677ce8e9fe40f-Paper.pdf.

[238] Rémi Munos. "Policy gradient in continuous time". In: *J. Mach. Learn. Res.* 7 (2005), pp. 771–791. URL: `http://jmlr.org/papers/v7/munos06b.html`.

[239] Kenji Doya. "Reinforcement Learning in Continuous Time and Space". In: *Neural Comput.* 12.1 (2000), pp. 219–245. DOI: `10.1162/089976600300015961`. eprint: `https://doi.org/10.1162/089976600300015961`.

[240] Steven J. Bradtke and Michael O. Duff. "Reinforcement Learning Methods for Continuous-Time Markov Decision Problems". In: (1994). URL: `https://papers.nips.cc/paper/889-reinforcement-learning-methods-for-continuous-time-markov-decision-problems.pdf`.

[241] Kyriakos G. Vamvoudakis and Frank L. Lewis. "Online actor–critic algorithm to solve the continuous-time infinite horizon optimal control problem". In: *Automatica* (2010). DOI: `10.1016/j.automatica.2010.02.018`.

[242] Nicolas Frémaux, Henning Sprekeler, and Wulfram Gerstner. "Reinforcement Learning Using a Continuous Time Actor-Critic Framework with Spiking Neurons". In: *PLOS Comput. Biol.* (2013). DOI: `10.1371/journal.pcbi.1003024`.

[243] Randal W. Beard, George N. Saridis, and John T. Wen. "Galerkin approximations of the generalized Hamilton-Jacobi-Bellman equation". In: *Automatica* 33.12 (1997), pp. 2159–2177. ISSN: 0005-1098. DOI: `https://doi.org/10.1016/S0005-1098(97)00128-3`. URL: `https://www.sciencedirect.com/science/article/pii/S0005109897001283`.

[244] Murad Abu-Khalaf and Frank L. Lewis. "Nearly optimal control laws for nonlinear systems with saturating actuators using a neural network HJB approach". In: *Automatica* 41.5 (2005), pp. 779–791. ISSN: 0005-1098. DOI: `https://doi.org/10.1016/j.automatica.2004.11.034`. URL: `https://www.sciencedirect.com/science/article/pii/S0005109805000105`.

[245] Volodymyr Mnih et al. "Human-level control through deep reinforcement learning". In: *Nature* 518 (2015), pp. 529–533. DOI: `10.1038/nature14236`.

[246] Oriol Vinyals et al. "Grandmaster level in StarCraft II using multi-agent reinforcement learning". In: *Nature* 575 (2019), p. 350.

[247] David Silver et al. "A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play". In: *Science* 362.6419 (2018), pp. 1140–1144. ISSN: 0036-8075. DOI: `10.1126/science.aar6404`. eprint: `https://science.sciencemag.org/content/362/6419/1140.full.pdf`. URL: `https://science.sciencemag.org/content/362/6419/1140`.

[248] Tuomas Haarnoja et al. "Soft Actor-Critic: Off-Policy Maximum Entropy Deep Reinforcement Learning with a Stochastic Actor". In: *arxiv.org:1801.01290* (2018).

[249] Tuomas Haarnoja et al. "Soft Actor-Critic Algorithms and Applications". In: *arXiv:1812.05905* (2018).

[250] OpenAI et al. "Solving Rubik's Cube with a Robot Hand". In: *arXiv:1910.07113* (2019).

[251] Marin Bukov et al. "Reinforcement Learning in Different Phases of Quantum Control". In: *Phys. Rev. X* 8 (3 Sept. 2018), p. 031086. DOI: `10.1103/PhysRevX.8.031086`.

[252] Marin Bukov. "Reinforcement learning for autonomous preparation of Floquet-engineered states: Inverting the quantum Kapitza oscillator". In: *Phys. Rev. B* 98.22 (Dec. 2018). ISSN: 2469-9969. DOI: `10.1103/physrevb.98.224305`.

[253] Jiahao Yao, Marin Bukov, and Lin Lin. "Policy gradient based quantum approximate optimization algorithm". In: *Mathematical and Scientific Machine Learning*. PMLR. 2020, pp. 605–634.

[254] Thomas Fösel et al. "Reinforcement Learning with Neural Networks for Quantum Feedback". In: *Phys. Rev. X* 8 (3 Sept. 2018), p. 031084. DOI: `10.1103/PhysRevX.8.031084`.

[255] Feng Chen et al. "Extreme spin squeezing from deep reinforcement learning". In: *Physical Review A* 100.4 (2019), p. 041801.

[256] Mogens Dalgaard et al. "Global optimization of quantum dynamics with AlphaZero deep exploration". In: *npj Quantum Information* (2020). DOI: `10.1038/s41534-019-0241-0`.

[257] Ariel Barr, Willem Gispen, and Austen Lamacraft. "Quantum Ground States from Reinforcement Learning". In: *Proceedings of The First Mathematical and Scientific Machine Learning Conference*. Ed. by Jianfeng Lu and Rachel Ward. Vol. 107. Proceedings of Machine Learning Research. Princeton University, Princeton, NJ, USA: PMLR, 20–24 Jul 2020, pp. 635–653. URL: `http://proceedings.mlr.press/v107/barr20a.html`.

[258] Willem Gispen and Austen Lamacraft. *Ground States of Quantum Many Body Lattice Models via Reinforcement Learning*. 2020. arXiv: `2012.07063 [quant-ph]`.

[259] Richard S. Sutton et al. "Policy gradient methods for reinforcement learning with function approximation". In: *NIPS*. 2000. URL: `http://incompleteideas.net/papers/SMSM-NIPS99.pdf`.

[260] Peter Marbach and John N. Tsitsiklis. "Approximate Gradient Methods in Policy-Space Optimization of Markov Reward Processes". In: *Discrete Event Dyn. S.* 13.1 (Jan. 2003), pp. 111–148. ISSN: 1573-7594. DOI: `10.1023/A:1022145020786`.

[261] Richard S. Sutton. "Learning to predict by the methods of temporal differences". In: *Machine Learning* 3.1 (Aug. 1988), pp. 9–44. ISSN: 1573-0565. DOI: `10.1007/BF00115009`.

[262] Avishek Das et al. "Reinforcement learning of rare diffusive dynamics". In: *The Journal of Chemical Physics* 155.13 (2021), p. 134105.

[263] Adrián A Budini, Robert M Turner, and Juan P Garrahan. "Fluctuating observation time ensembles in the thermodynamics of trajectories". In: *J. Stat. Mech. Theory Exp.* 2014.3 (2014), P03012.

[264] Tooru Taniguchi and EGD Cohen. "Onsager-Machlup theory for nonequilibrium steady states and fluctuation theorems". In: *J. Stat. Phys.* 126.1 (2007), pp. 1–41.

[265] Jacek Grela, Satya N Majumdar, and Grégory Schehr. "Non-intersecting Brownian bridges in the flat-to-flat geometry". In: *Journal of Statistical Physics* 183.3 (2021), pp. 1–35.

[266] Benjamin De Bruyne, Satya N Majumdar, and Grégory Schehr. "Generating discrete-time constrained random walks and Lévy flights". In: *Physical Review E* 104.2 (2021), p. 024117.

[267] Vivek S. Borkar, Sonia Juneja, and Arzad Alam Kherani. "Peformance Analysis Conditioned on Rare Events: An Adaptive Simulation Scheme". In: *Comm. Inf. S.* 3.4 (2003), pp. 256–278.

[268] Federico Carollo et al. "Making rare events typical in Markovian open quantum systems". In: *Physical Review A* 98.1 (1 July 2018), p. 010103. DOI: `10.1103/PhysRevA.98.010103`. URL: `https://link.aps.org/doi/10.1103/PhysRevA.98.010103`.

[269] Luke Causer, Mari Carmen Bañuls, and Juan P Garrahan. "Optimal sampling of dynamical large deviations via matrix product states". In: *Physical Review E* 103.6 (2021), p. 062144.

[270] Hilbert J. Kappen, Vicenç Gómez, and Manfred Opper. "Optimal control as a graphical model inference problem". In: *Machine Learning* 87.2 (2012), pp. 159–182. DOI: `10.1007/s10994-012-5278-7`.

[271] Vladimir Y Chernyak et al. "Stochastic optimal control as non-equilibrium statistical mechanics: calculus of variations over density and current". In: *J. Phys. A* 47.2 (Dec. 2013), p. 022001. DOI: `10.1088/1751-8113/47/2/022001`. URL: `https://doi.org/10.1088%2F1751-8113%2F47%2F2%2F022001`.

[272] Sep Thijssen and H. J. Kappen. "Path integral control and state-dependent feedback". In: *Phys. Rev. E* 91 (3 Mar. 2015), p. 032104. DOI: `10.1103/PhysRevE.91.032104`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.91.032104`.

[273] H. J. Kappen and H. C. Ruiz. "Adaptive Importance Sampling for Control and Inference". In: *J. Stat. Phys.* 162.5 (Mar. 2016), pp. 1244–1266. ISSN: 1572-9613. DOI: `10.1007/s10955-016-1446-7`.

[274] Emanuel Todorov. "Efficient computation of optimal actions". In: *Proc. Natl. Acad. Sci.* 106.28 (2009), pp. 11478–11483. ISSN: 0027-8424. DOI: `10.1073/pnas.0710743106`. eprint: `https://www.pnas.org/content/106/28/11478.full.pdf`. URL: `https://www.pnas.org/content/106/28/11478`.

[275] Gergely Neu, Anders Jonsson, and Vicenç Gómez. "A unified view of entropy-regularized Markov decision processes". In: *ArXiv* abs/1705.07798 (2017). URL: `https://arxiv.org/abs/1805.00909`.

[276] Sergey Levine. "Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review". In: *ArXiv* abs/1705.07798 (2018). URL: `https://arxiv.org/abs/1705.07798`.

[277] Matthieu Geist, Bruno Scherrer, and Olivier Pietquin. "A Theory of Regularized Markov Decision Processes". In: *ICML*. 2019. URL: `https://arxiv.org/abs/1901.11275`.

[278] Doina Precup, Richard S. Sutton, and Satinder P. Singh. "Eligibility Traces for Off-Policy Policy Evaluation". In: *ICML*. ICML '00. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 2000, pp. 759–766. ISBN: 1-55860-707-2. URL: `http://dl.acm.org/citation.cfm?id=645529.658134`.

[279] Thomas Degris, Martha White, and Richard S. Sutton. "Off-Policy Actor-Critic". In: *ICML*. 2012. URL: `https://arxiv.org/abs/1205.4839`.

[280] Chris J. C. H. Watkins. "Learning from Delayed Rewards". PhD thesis. Cambridge University, 1989. URL: `https://www.cs.rhul.ac.uk/home/chrisw/thesis.html`.

[281] Ronald J. Williams. *Reinforcement-Learning Connectionist Systems*. Tech. rep. Northeastern University, 1987.

[282] Ronald J. Williams. "Simple statistical gradient-following algorithms for connectionist reinforcement learning". In: *Machine Learning* 8.3 (May 1992), pp. 229–256. ISSN: 1573-0565. DOI: `10.1007/BF00992696`.

[283] Leemon Baird and Andrew W Moore. "Gradient descent for general reinforcement learning". In: *Adv. Neural. Inf. Process. Syst.* (1999), pp. 968–974.

[284] Richard S. Sutton et al. "Fast Gradient-descent Methods for Temporal-difference Learning with Linear Function Approximation". In: *ICML*. ICML '09. Montreal, Quebec, Canada: ACM, 2009, pp. 993–1000. ISBN: 978-1-60558-516-1. DOI: `10.1145/1553374.1553501`.

[285] Hamid R. Maei et al. "Convergent Temporal-difference Learning with Arbitrary Smooth Function Approximation". In: *NIPS*. NIPS'09. Vancouver, British Columbia, Canada: Curran Associates Inc., 2009, pp. 1204–1212. ISBN: 978-1-61567-911-9. URL: `http://dl.acm.org/citation.cfm?id=2984093.2984229`.

[286] H. R. Maei. "Gradient Temporal-Difference Learning Algorithms". PhD thesis. University of Alberta, 2011.

[287] Michele Invernizzi, Pablo M Piaggi, and Michele Parrinello. "Unified approach to enhanced sampling". In: *Phys. Rev. X* 10.4 (2020), p. 041034.

[288] Yuehaw Khoo, Jianfeng Lu, and Lexing Ying. "Solving for high-dimensional committor functions using artificial neural networks". In: *Research in the Mathematical Sciences* 6.1 (2019), pp. 1–13.

[289] Qianxiao Li, Bo Lin, and Weiqing Ren. "Computing committor functions for the study of rare events using deep learning". In: *The Journal of Chemical Physics* 151.5 (Aug. 2019), p. 054112. ISSN: 1089-7690. DOI: 10.1063/1.5110439. URL: http://dx.doi.org/10.1063/1.5110439.

[290] Michael Grünwald, Christoph Dellago, and Phillip L Geissler. "Precision shooting: Sampling long transition pathways". In: *J. Chem. Phys.* 129.19 (2008), p. 194101.

[291] Nicholas Guttenberg, Aaron R Dinner, and Jonathan Weare. "Steered transition path sampling". In: *J. Chem. Phys.* 136.23 (2012), 06B609.

[292] Gabriel Stoltz. "Path sampling with stochastic dynamics: Some new algorithms". In: *J. Comp. Phys.* 225.1 (2007), pp. 491–508.

[293] Graeme Henkelman, Blas P Uberuaga, and Hannes Jónsson. "A climbing image nudged elastic band method for finding saddle points and minimum energy paths". In: *J. Chem. Phys.* 113.22 (2000), pp. 9901–9904.

[294] Graeme Henkelman and Hannes Jónsson. "Improved tangent estimate in the nudged elastic band method for finding minimum energy paths and saddle points". In: *J. Chem. Phys.* 113.22 (2000), pp. 9978–9985.

[295] Graeme Henkelman. "Methods for calculating rates of transitions with application to catalysis and crystal growth". PhD thesis. 2001.

[296] E Weinan, Weiqing Ren, and Eric Vanden-Eijnden. "String method for the study of rare events". In: *Physical Review B* 66.5 (2002), p. 052301.

[297] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*. Vol. 293. Springer Science & Business Media, 2013.

[298] Hiroshi Fujisaki, Motoyuki Shiga, and Akinori Kidera. "Onsager–Machlup action-based path sampling and its combination with replica exchange for diffusive and multiple pathways". In: *J. Chem. Phys.* 132.13 (2010), p. 134101.

[299] Marc Delarue, Patrice Koehl, and Henri Orland. "Ab initio sampling of transition paths by conditioned Langevin dynamics". In: *The Journal of chemical physics* 147.15 (2017), p. 152703.

[300] Yoshua Bengio et al. "Curriculum learning". In: *Proceedings of the 26th annual international conference on machine learning*. 2009, pp. 41–48.

[301] Viktor Szalay. "Discrete variable representations of differential operators". In: *J. Chem. Phys.* 99.3 (1993), pp. 1978–1984.

[302] Klaus Müller and Leo D Brown. "Location of saddle points and minimum energy paths by a constrained simplex optimization procedure". In: *Theor. Chim. Acta* 53.1 (1979), pp. 75–93.

[303] Silvia Bonfanti and Walter Kob. "Methods to locate saddle points in complex landscapes". In: *J. Chem. Phys.* 147.20 (2017), p. 204104.

[304] Michiel Laleman, Enrico Carlon, and Henri Orland. "Transition path time distributions". In: *J. Chem. Phys.* 147.21 (2017), p. 214103.

[305] Anton Schwartz. "A Reinforcement Learning Method for Maximizing Undiscounted Rewards". In: *ICML.* 1993. DOI: `10.1016/B978-1-55860-307-3.50045-9`. URL: `http://www.sciencedirect.com/science/article/pii/B9781558603073500459`.

[306] Dimitri P. Bertsekas and John N. Tsitsiklis. *Neuro-dynamic programming.* Springer, 1996.

[307] John N. Tsitsiklis and Benjamin Van Roy. "Average cost temporal-difference learning". In: *Automatica* 35.11 (1999), pp. 1799–1808. ISSN: 0005-1098. DOI: `10.1016/S0005-1098(99)00099-0`. URL: `http://www.sciencedirect.com/science/article/pii/S0005109899000990`.

[308] Patrick Pietzonka, Andre C Barato, and Udo Seifert. "Universal bounds on current fluctuations". In: *Phys. Rev. E* 93.5 (2016), p. 052145.

[309] Thierry Bodineau and Bernard Derrida. "Current fluctuations in nonequilibrium diffusive systems: an additivity principle". In: *Phys. Rev. Lett.* 92.18 (2004), p. 180601.

[310] Joel L Lebowitz and Herbert Spohn. "A Gallavotti–Cohen-type symmetry in the large deviation functional for stochastic dynamics". In: *J. Stat. Phys.* 95.1 (1999), pp. 333–365.

[311] Frédéric Cérou and Arnaud Guyader. "Adaptive multilevel splitting for rare event analysis". In: *Stochastic Analysis and Applications* 25.2 (2007), pp. 417–443.

[312] Vivien Lecomte and Julien Tailleur. "A numerical approach to large deviations in continuous time". In: *J. Stat. Mech.* 2007.03 (2007), P03004. URL: `http://stacks.iop.org/1742-5468/2007/i=03/a=P03004`.

[313] Thibault Lestang et al. "Computing return times or return periods with rare event algorithms". In: *J. Stat. Mech. Theory Exp.* 2018.4 (2018), p. 043213.

[314] Pelerine Tsobgni Nyawo and Hugo Touchette. "Large deviations of the current for driven periodic diffusions". In: *Phys. Rev. E* 94 (3 Sept. 2016), p. 032101. DOI: `10.1103/PhysRevE.94.032101`. URL: `https://link.aps.org/doi/10.1103/PhysRevE.94.032101`.

[315] Tom H E Oakes, Adam Moss, and Juan P Garrahan. "A deep learning functional estimator of optimal dynamics for sampling large deviations". In: *Mach. Learn.: Sci. Technol.* 1.3 (July 2020), p. 035004. DOI: `10.1088/2632-2153/ab95a1`. URL: `https://doi.org/10.1088/2632-2153/ab95a1`.

[316] G Bartolucci, S Orioli, and Faccioli P. "Transition path theory from biased simulations". In: *J. Chem. Phys.* 149.7 (2018), p. 072336. DOI: `https://doi.org/10.1063/1.5027253`.

[317] M Cristina Marchetti et al. "Hydrodynamics of soft active matter". In: *Reviews of Modern Physics* 85.3 (2013), p. 1143.

[318] M Reza Shaebani et al. "Computational models for active matter". In: *Nature Reviews Physics* 2.4 (2020), pp. 181–199.

[319] Thomas Speck. "Collective behavior of active Brownian particles: From microscopic clustering to macroscopic phase separation". In: *The European Physical Journal Special Topics* 225.11 (2016), pp. 2287–2299.

[320] Takahiro Nemoto et al. "Optimizing active work: Dynamical phase transitions, collective motion, and jamming". In: *Physical Review E* 99.2 (2 Feb. 2019), p. 022605. DOI: 10.1103/PhysRevE.99.022605. URL: https://link.aps.org/doi/10.1103/PhysRevE.99.022605.

[321] Gerhard Gompper et al. "The 2020 motile active matter roadmap". In: *Journal of Physics: Condensed Matter* 32.19 (2020), p. 193001.

[322] Eric Woillez et al. "Activated escape of a self-propelled particle from a metastable state". In: *Physical review letters* 122.25 (2019), p. 258001.

[323] Andrei Militaru et al. "Escape dynamics of active particles in multistable potentials". In: *Nature Communications* 12.1 (2021), pp. 1–6.

[324] Eric Woillez, Yariv Kafri, and Vivien Lecomte. "Nonlocal stationary probability distributions and escape rates for an active Ornstein–Uhlenbeck particle". In: *Journal of Statistical Mechanics: Theory and Experiment* 2020.6 (2020), p. 063204.

[325] Ahmad K Omar et al. "Phase Diagram of Active Brownian Spheres: Crystallization and the Metastability of Motility-Induced Phase Separation". In: *Physical Review Letters* 126.18 (2021), p. 188002.

[326] Joakim Stenhammar et al. "Activity-induced phase separation and self-assembly in mixtures of active and passive particles". In: *Physical review letters* 114.1 (2015), p. 018301.

[327] Stewart A Mallory, Chantal Valeriani, and Angelo Cacciuto. "An active approach to colloidal self-assembly". In: *Annual review of physical chemistry* 69 (2018), pp. 59–79.

[328] Abraham Nitzan. *Chemical dynamics in condensed phases: relaxation, transfer and reactions in condensed molecular systems*. Oxford university press, 2006.

[329] Yaouen Fily and M Cristina Marchetti. "Athermal phase separation of self-propelled particles with no alignment". In: *Physical review letters* 108.23 (2012), p. 235702.

[330] Gabriel S Redner, Michael F Hagan, and Aparna Baskaran. "Structure and dynamics of a phase-separating active colloidal fluid". In: *Physical review letters* 110.5 (5 Jan. 2013), p. 055701. DOI: 10.1103/PhysRevLett.110.055701. URL: https://link.aps.org/doi/10.1103/PhysRevLett.110.055701.

[331] Julian Bialké, Hartmut Löwen, and Thomas Speck. "Microscopic theory for the phase separation of self-propelled repulsive disks". In: *Europhysics Letters* 103.3 (2013), p. 30008.

[332] SA Mallory, ML Bowers, and A Cacciuto. "Universal reshaping of arrested colloidal gels via active doping". In: *The Journal of Chemical Physics* 153.8 (2020), p. 084901.

[333] Andrey Sokolov et al. "Swimming bacteria power microscopic gears". In: *Proceedings of the National Academy of Sciences* 107.3 (2010), pp. 969–974.

[334] Avishek Das, Benjamin Kuznets-Speck, and David T Limmer. "Direct evaluation of rare events in active matter from variational path sampling". In: *Physical Review Letters* 128.2 (2022), p. 028005.

[335] David Chandler. "Statistical mechanics of isomerization dynamics in liquids and the transition state approximation". In: *The Journal of Chemical Physics* 68.6 (1978), pp. 2959–2970.

[336] L. Onsager and S. Machlup. "Fluctuations and Irreversible Processes." In: *Phys. Rev.* 91 (1953), p. 1505. DOI: `https://doi.org/10.1103/PhysRev.91.1505`.

[337] J Szavits-Nossan and M R Evans. "Inequivalence of nonequilibrium path ensembles: the example of stochastic bridges." In: *Journal of Statistical Mechanics* 2015 (2015), P12008. DOI: `https://doi.org/10.1088/1742-5468/2015/12/P12008`.

[338] R. Chetrite and H. Touchette. "Nonequilibrium Markov Processes Conditioned on Large Deviations". In: *Annales Henri Poincaré* 16 (2015), p. 2005. DOI: `https://doi.org/10.1007/s00023-014-0375-8`.

[339] R. Chetrite and H. Touchette. "Variational and optimal control representations of conditioned and driven processes." In: *Journal of Statistical Mechanics* 2015 (2015), P12001. DOI: `https://doi.org/10.1088/1742-5468/2015/12/P12001`.

[340] Eric Vanden-Eijnden et al. "Transition-path theory and path-finding algorithms for the study of rare events." In: *Annual review of physical chemistry* 61 (2010), pp. 391–420.

[341] John D Weeks, David Chandler, and Hans C Andersen. "Role of repulsive forces in determining the equilibrium structure of simple liquids". In: *The Journal of chemical physics* 54.12 (1971), pp. 5237–5247.

[342] Christoph Dellago, Peter G Bolhuis, and David Chandler. "On the calculation of reaction rate constants in the transition path ensemble". In: *The Journal of chemical physics* 110.14 (1999), pp. 6617–6625.

[343] Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. "Well-tempered metadynamics a smoothly converging and tunable free-energy method". In: *Physical Review Letters* 100.2 (2008), p. 020603.

[344] P Buijsman and PG Bolhuis. "Transition path sampling for non-equilibrium dynamics without predefined reaction coordinates". In: *The Journal of chemical physics* 152.4 (2020), p. 044108.

[345] Michael R Shirts and John D Chodera. "Statistically optimal analysis of samples from multiple equilibrium states". In: *Journal of chemical physics* 129.12 (2008), p. 124105.

[346] Rosalind J Allen, Patrick B Warren, and Pieter Rein Ten Wolde. "Sampling rare switching events in biochemical networks". In: *Physical review letters* 94.1 (2005), p. 018104.

[347] Avishek Das, Ben Kuznets-Speck, and David T. Limmer. *Direct evaluation of rare events in active matter from variational path sampling*. Version v3.0. Zenodo, Aug. 2021. DOI: 10.5281/zenodo.5763101. URL: %7Bhttps://doi.org/10.5281/zenodo.5763101%7D.