

# Chapter 1

## Introduction

Statistical Physics uses probability theory and statistical methods to describe physical systems with many degrees of freedom. It provides a relation of macroscopic quantities in equilibrium like temperature or pressure to probabilities of detailed microscopic configurations of a material. It is intuitively understood from an experimental point of view: Controlling temperature and pressure of a material allows us to control system properties like the state of matter or the heat capacity and measure such quantity of interest. Repeating the experiment will always produce the same result within the error of measurement uncertainty. The exact microscopic configuration of all atoms in the systems on the other hand are unknown and will differ between measurements but do not influence the outcome. Still, we understand that the microscopic behaviour of the system is greatly influenced by a few macroscopic variables. For instance, a liquid phase does not consist of well-ordered atoms. From the microscopic perspective, having full knowledge of all microscopic configuration of a system provides full thermodynamic information of it. Each microscopic state can be assigned a probability to show up depending on macroscopic parameters. The key for the experimentalist and the statistical physicist is to identify these parameters and find their mathematical link to the microscopic probabilistic description of the system. Such relations are well known for systems that do not gain heat from or dissipative heat to its environment, i.e. they are in equilibrium. For the broad number of systems that do not meet this requirement a full theory is not formulated until today [1].

The scope of statistical physics has greatly increased since the emergence of compu-

tational physics in the 1940s [2, 3]. The impact on physics lies in the possibility to support and compare analytical work with numerical and expand it beyond its analytical accessibility. The advantage is found in the possibility to sample microscopic states of a system and connect these to macroscopic thermodynamics via ensemble averages [4]. Simulations provide a method to study off-equilibrium statistical mechanics because one can define non-equilibrium conditions without understanding their macroscopic influence on the system, e.g. one might define a force or a heat source and measure the macroscopic effect on the system to gain off-equilibrium information without having a access to a general theory.

With the rapid development of computational hardware, a number of simulations were performed in fields like mechanics, electrodynamics, particle physics, astrophysics, fluid dynamics and chemistry [5]. The field of soft matter physics with the feature of large fluctuations aims to probe polymers, liquid crystals and the whole area of colloidal system [6]. However, most simulations are performed in thermodynamic equilibrium until today because they are more efficient based on our wide understanding of equilibrium statistical mechanics. In particular Boltzmann and Gibbs formulation of ensembles teaches us that microstates weighted by the Boltzmann factor hold all thermodynamic information on equilibrium systems [7].

Off-equilibrium cannot draw from such a theory and we have to sample microtrajectories without additional information on the ensemble of a system. Macroscopic motion or heat transport from a source are examples of off-equilibrium processes. The dynamics are time-dependent and cannot be understood by static microstates but through a chain of microscopic states. Stochastic thermodynamics provides a conceptual framework to create such weighted trajectories for a large class of soft matter systems under fairly general non-equilibrium conditions [8]. It models aqueous solution of the system implicitly by a coupling to a heat bath while allowing off-equilibrium disturbances based on external forces or unbalanced chemical potentials. This framework allows us to model off-equilibrium systems and estimate macroscopic information of interest.

Increasing computational power gave rise to a number of simulation off-equilibrium like crystal growth morphology [9, 10], phase separation [11] or glassy dynamics [12], just to name a few. We distinguish between different types of non-equilibrium processes: (i) A system can be time-dependent out of equilibrium by external forces. For instance an external alternating electric field disturbs a molecular system and the system response is studied. (ii) A system is driven out of equilibrium by a time-constant external force, for instance a hydrodynamic forces, chemical reactions,

mechanical forces or temperature gradients. An external reservoir provides heat flow to maintain the disturbance. The system will settle in a non-equilibrium steady state (NESS) that shows non-equilibrium behaviour like path-dependence while its macroscopic state does not change in time. The probability of taking a specific microtrajectory will be time-independent, irrespective on assumptions of the history of the system.

This property makes the NESS the ideal case to expand the well-know equilibrium *reweighting* technique from equilibrium to off-equilibrium. Reweighting is a powerful tool in computational physics because it connects data collected from simulation under different thermodynamic conditions. In equilibrium, one draws mathematical relations from statistical mechanics that allow to reassign probabilities of recorded data under changing thermodynamic condition correctly. This technique is a pivot point for efficient sampling methods of equilibrium systems, like multicanonical simulation [13], replica exchange [14] or metadynamics [15]. Such methods rely on reweighting the configuration space between equilibrium systems [16–18]. More recent attempts focus on sampling the dynamics and reweight the data on trajectories between equilibrium ensembles [19–21]. These methods expanded the scope of simulations to complex dynamic processes like folding of proteins [22], protein kinase [23] or ligand binding [24]. Having a comparable tool for NESS would expand the scope of soft-matter simulation further. Possible systems of interest are for instance molecular motors [25] or biological switches [26]. The key for such tool is the identification of controlling parameters defining the state of a NESS and a theory that connects simulation data to such parameters. First results were recently developed, relying on reweighting microscopic trajectories individually [27, 28]. In this thesis, we will coarse-grain these trajectories and shift to an ensembles based reweighting method similar to the ensemble reweighting of microstates in equilibrium.

The dynamics are coarsened by applying Markov State Models (MSM). The idea is to discretise the space and record the trajectories transitioning between discrete states. The formed bundle of short trajectories can be reassembled to form longer trajectories. This approach of collecting dynamic information is more efficient because shorter trajectories are faster to sample than long ones and dynamically similar microstates are clustered while preserving long timescale information. An MSM model can be constructed in the high-dimensional configuration space or — if this space grows too large for complex systems — in system specific collective variables [29]. The MSM approach was applied to analyse simulation sampled from equilibrium [30] and non-equilibrium systems [31, 32].

The reweighting method needs a foundation in off-equilibrium to reweight the data according to the thermodynamic information available to the user. The equilibrium maximum entropy principle of statistical mechanics says that a statistical system tends to the state where most microscopic realisations are possible subject to constraints. The constraints act as the thermodynamic environment like the temperature. Jaynes proposed with the Maximum Caliber an extension of this principle to non-equilibrium processes [33]. This method was shown to recover physical relations in off-equilibrium [34], model dynamical complex systems from limited information [35, 36], correct dynamic data by inferring physical information [37, 38] and more application on statistical systems in physics, chemistry and biology [39]. Here we will use it as the basis for our reweighting method on NESS. Jaynes evolved the equilibrium maximum entropy theory to an information theoretic point of view: Based on information I have on a system, what is the most likely state it is in? Jaynes answers this question assuming the most uncertain (or highest entropy) probability distribution is as noncommittal as possible regarding unknown information. This change in view allowed him to propose an expansion to the equilibrium theory on microstates to any kind of system. It opens the theory to collect microtrajectories and maximise the entropy of all paths.

The thermodynamic environment for NESS is defined by constraining the local dissipation of heat in the system. This constraint describes how the state of the system changes depending on where and how much heat is inserted or withdrawn from the system. The non-dissipative contribution to dynamics are drawn from the reference data. It is shown how this contribution originates from a dynamical invariant for NESS. The quantity is comparable to the density of states in equilibrium. We will discuss how the Maximum Caliber determines a thermodynamic state and how to choose the constraints to model a physical system.

The structure of the thesis is as follows: The first part will give a short introduction to equilibrium statistical mechanics and off-equilibrium stochastic thermodynamic with its consequences. A toy model for a NESS is introduced and theoretical and practical introduction to Markov state modelling will be given on the example of this model. The third part focuses on the introduction to the framework of Jaynes Maximum Caliber and an extensive derivation of the novel reweighting scheme. The following chapters will test the procedure (i) by reweighting systems by conformational variables and (ii) reweighting systems by collective variables. The last chapter will present (i) a discussion on the physical basis of the reweighting scheme based on the underlying invariant in NESS and (ii) a discussion on trajectory-based stochastic

thermodynamics of the chosen NESS ensembles.

# Bibliography

- [1] J. P. Dougherty, “Foundations of non-equilibrium statistical mechanics,” *Philosophical Transactions of the Royal Society of London. Series A: Physical and Engineering Sciences*, vol. 346, no. 1680, pp. 259–305, 1994.
- [2] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, “Equation of state calculations by fast computing machines,” *The journal of chemical physics*, vol. 21, no. 6, pp. 1087–1092, 1953.
- [3] B. J. Alder and T. E. Wainwright, “Studies in molecular dynamics. i. general method,” *The Journal of Chemical Physics*, vol. 31, no. 2, pp. 459–466, 1959.
- [4] D. Frenkel and B. Smit, *Understanding molecular simulation: from algorithms to applications*, vol. 1. Elsevier, 2001.
- [5] D. W. Heermann, “Computer-simulation methods,” in *Computer Simulation Methods in Theoretical Physics*, pp. 8–12, Springer, 1990.
- [6] C. Holm and K. Kremer, *Advanced computer simulation approaches for soft matter sciences III*, vol. 221. Springer, 2008.
- [7] J. W. Gibbs, *Elementary principles in statistical mechanics: developed with especial reference to the rational foundations of thermodynamics*. C. Scribner’s sons, 1902.
- [8] U. Seifert, “Stochastic thermodynamics, fluctuation theorems and molecular machines,” *Reports on progress in physics*, vol. 75, no. 12, p. 126001, 2012.
- [9] M. Radu and K. Kremer, “Enhanced crystal growth in binary Lennard-Jones mixtures,” *Physical Review Letters*, vol. 118, no. 5, p. 055702, 2017.

- [10] B. Lander, U. Seifert, and T. Speck, “Crystallization in a sheared colloidal suspension,” *The Journal of chemical physics*, vol. 138, no. 22, p. 224907, 2013.
- [11] J. Smrek and K. Kremer, “Small activity differences drive phase separation in active-passive polymer mixtures,” *Physical review letters*, vol. 118, no. 9, p. 098002, 2017.
- [12] H. Kawamura, “Dynamical simulation of spin-glass and chiral-glass orderings in three-dimensional Heisenberg spin glasses,” *Physical review letters*, vol. 80, no. 24, p. 5421, 1998.
- [13] W. Janke, “Multicanonical monte carlo simulations,” *Physica A: Statistical Mechanics and its Applications*, vol. 254, no. 1-2, pp. 164–178, 1998.
- [14] Y. Sugita and Y. Okamoto, “Replica-exchange molecular dynamics method for protein folding,” *Chemical physics letters*, vol. 314, no. 1-2, pp. 141–151, 1999.
- [15] A. Laio and M. Parrinello, “Escaping free-energy minima,” *Proceedings of the National Academy of Sciences*, vol. 99, no. 20, pp. 12562–12566, 2002.
- [16] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, “Multidimensional free-energy calculations using the weighted histogram analysis method,” *Journal of Computational Chemistry*, vol. 16, no. 11, pp. 1339–1350, 1995.
- [17] T. M. Schaäfer and G. Settanni, “Data reweighting in metadynamics simulations,” *Journal of Chemical Theory and Computation*, vol. 16, no. 4, pp. 2042–2052, 2020.
- [18] M. R. Shirts and J. D. Chodera, “Statistically optimal analysis of samples from multiple equilibrium states,” *The Journal of chemical physics*, vol. 129, no. 12, p. 124105, 2008.
- [19] H. Wu, A. S. Mey, E. Rosta, and F. Noé, “Statistically optimal analysis of state-discretized trajectory data from multiple thermodynamic states,” *The Journal of Chemical Physics*, vol. 141, no. 21, p. 12B629\_1, 2014.
- [20] H. Wan, G. Zhou, and V. A. Voelz, “A maximum-caliber approach to predicting perturbed folding kinetics due to mutations,” *Journal of chemical theory and computation*, vol. 12, no. 12, pp. 5768–5776, 2016.

- [21] H. Wu, F. Paul, C. Wehmeyer, and F. Noé, “Multiensemble Markov models of molecular thermodynamics and kinetics,” *Proceedings of the National Academy of Sciences*, vol. 113, no. 23, pp. E3221–E3230, 2016.
- [22] V. A. Voelz, G. R. Bowman, K. Beauchamp, and V. S. Pande, “Molecular simulation of ab initio protein folding for a millisecond folder NTL9 (1- 39),” *Journal of the American Chemical Society*, vol. 132, no. 5, pp. 1526–1528, 2010.
- [23] S. Yang and B. Roux, “Src kinase conformational activation: thermodynamics, pathways, and mechanisms,” *PLoS computational biology*, vol. 4, no. 3, 2008.
- [24] N. Plattner and F. Noé, “Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and markov models,” *Nature communications*, vol. 6, p. 7653, 2015.
- [25] M. Schliwa and G. Woehlke, “Molecular motors,” *Nature*, vol. 422, no. 6933, pp. 759–765, 2003.
- [26] A. Goldbeter, *Biochemical oscillations and cellular rhythms: the molecular bases of periodic and chaotic behaviour*. Cambridge university press, 1997.
- [27] L. Donati, C. Hartmann, and B. G. Keller, “Girsanov reweighting for path ensembles and Markov state models,” *The Journal of chemical physics*, vol. 146, no. 24, p. 244112, 2017.
- [28] J. D. Russo, J. Copperman, and D. M. Zuckerman, “Iterative trajectory reweighting for estimation of equilibrium and non-equilibrium observables,” *arXiv preprint arXiv:2006.09451*, 2020.
- [29] F. Noé and C. Clementi, “Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods,” *Current Opinion in Structural Biology*, vol. 43, pp. 141–147, 2017.
- [30] C. R. Schwantes and V. S. Pande, “Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9,” *Journal of chemical theory and computation*, vol. 9, no. 4, pp. 2000–2009, 2013.
- [31] F. Knoch and T. Speck, “Nonequilibrium Markov state modeling of the globule-stretch transition,” *Physical Review E*, vol. 95, no. 1, p. 012503, 2017.
- [32] F. Knoch and T. Speck, “Non-equilibrium Markov state modeling of periodically driven biomolecules,” *The Journal of chemical physics*, vol. 150, no. 5, p. 054103, 2019.



- [33] E. T. Jaynes, “Macroscopic prediction,” in *Complex Systems—Operational Approaches in Neurobiology, Physics, and Computers*, pp. 254–269, Springer, 1985.
- [34] P. D. Dixit, J. Wagoner, C. Weistuch, S. Pressé, K. Ghosh, and K. A. Dill, “Perspective: Maximum caliber is a general variational principle for dynamical systems,” *The Journal of chemical physics*, vol. 148, no. 1, p. 010901, 2018.
- [35] P. D. Dixit, A. Jain, G. Stock, and K. A. Dill, “Inferring transition rates of networks from populations in continuous-time Markov processes,” *Journal of chemical theory and computation*, vol. 11, no. 11, pp. 5464–5472, 2015.
- [36] P. D. Dixit, E. Lyashenko, M. Niepel, and D. Vitkup, “Maximum entropy framework for inference of cell population heterogeneity in signaling networks,” *bioRxiv*, p. 137513, 2018.
- [37] Z. F. Brotzakis, M. Vendruscolo, P. Bolhuis, *et al.*, “A method of incorporating rate constants as kinetic constraints in molecular dynamics simulations,” *arXiv preprint arXiv:2006.00868*, 2020.
- [38] D. Meral, D. Provasi, and M. Filizola, “An efficient strategy to estimate thermodynamics and kinetics of G protein-coupled receptor activation using metadynamics and maximum caliber,” *The Journal of chemical physics*, vol. 149, no. 22, p. 224101, 2018.
- [39] K. Ghosh, P. D. Dixit, L. Agozzino, and K. A. Dill, “The Maximum Caliber variational principle for nonequilibria,” *Annual Review of Physical Chemistry*, vol. 71, pp. 213–238, 2020.