Doctoral thesis ETH No. 16863

Aspects of thermostated and nonautonomous molecular dynamics

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY
ZURICH

for the degree of
Doctor of Natural Sciences

presented by
MICHEL ALAIN CUENDET
M. Sc., Swiss Federal Institute of Technology EPFL,
Lausanne, Switzerland
born August 11, 1974
citizen of Switzerland

accepted on the recommendation of
Prof. Dr. Wilfred F. van Gunsteren, examiner
Prof. Dr. Olivier Michielin, Prof. Dr. Ursula Röthlisberger,
Prof. Dr. Matthias Troyer, co-examiners

2006

To my Grandfather.

Acknowledgments

First and foremost, I would like to thank my three supervisors, without whom this thesis could never have been realized. I started with Prof. Dr. Ursula Röthlisberger as she was still at ETH, before she moved to the EPFL in Lausanne. I would like to acknowledge her giving me the opportunity to start a Ph.D in her group although my background was not in chemistry or molecular dynamics. I am grateful to her for allowing me to go to numerous conferences and workshops, and giving me a lot of freedom in my research. Next, I would like to thank Prof. Dr. Olivier Michielin, who I got to know during a period of civil service in his laboratory. With him I later started a collaboration which gave a new breath to my Ph.D. and led to Chapter 6 of this thesis. I liked the long passionate discussions with him and his communicative motivation for the project. I am grateful for the working place, the generous support and the appreciable computational resources of many kinds that Olivier provided, although I was not an official member of his group. Finally, I would like to express my gratitude to Prof. Dr. Wilfred van Gunsteren, who became my adoptive Doktorvater when I decided to stay at ETH. I gradually felt more and more integrated in the group, and I am thankful to Wilfred as well as his whole group for making me feel in the end like a true member of IGC. Towards the very end of the thesis, a very appreciated and fruitful scientific exchange built up with Wilfred, which finally resulted in Chapter 2 of this thesis. I was honored to work with one of the founders of the field and was impressed by his enthusiasm and thoroughness. I am especially indebted to him for the great attention that he devoted to me in the last couple of months before completion of the thesis. I would also like to thank Prof. Matthias Troyer for being a co-examiner for my thesis defense. In addition, I acknowledge the support of the special grant of the ETH council "Doktorieren an der anderen ETH", which provided funding for most of my thesis. I am grateful to Dr. Christoph Grolimund of the ETH council for allowing an extension of the grant.

Very important to me are the people who were my companions during the four and a half years of this thesis work. Chronologically, I'll start with Ursula's group. My early office mates Marialore Sulpizi and Leonardo Guidoni, who guided my first steps in using CPMD, remained the last survivors of the group in Zürich. Stefano Piana donated me his jungle-like plant, which prospered on my desk until this day. Patrick Maurer taught me the rudiments of administrating a Linux cluster and shaking old SCSI cables with bad contacts until the home disk works again. Maria Carola Colombo and Ute Röhrig were also there since the beginning, and we had nice discussions sharing our impressions about the difficulties of completing a Ph.D thesis. I later pleasantly shared offices in Lausanne with Denis Bucher and Sam Arey, and finally with Christian Gossens with whom I had very friendly discussions. Enrico Tapavicza was my funky flatmate in real life, with whom

I would sometimes practice my german more in Lausanne than in Zürich. I liked the collaboration on the alchemical DFT project with Anatole von Lilienfeld-Toal, who showed me his impressively fast, and also potentially dangerous way of writing papers. I thank Michele Cascella for rescuing the azurin project which caused me so much trouble, taking it over and bringing it to the publication stage. Very valuable to me were the numerous discussions with Pascal Baillod, either relaxed, or very technical about the hiccups of Gromacs or scripting in Matlab and VMD. (You know more than I do now!) I'll also keep a nice memory of all other past or present members of the group, Roberto Lins, Håkan Hugosson, Mauricio Coutinho-Neto, I-Chun Lin, Fanny Masson, Samuele Giani, Marc-Etienne Moret and the legendary italian opera character Matteo Gugliemi. Last but not least, I would like to mention the great work of Ivano Tavernelli, our daddy of all, who is involved in literally every project of the group. I am grateful to Ivano for proofreading some of my early work and for answering a lot of my questions.

Olivier's group also contributed greatly to the cheerful environment which I enjoyed during these last years. I remember the time when we were 8 people working in a 9 m^2 cubicle, the "sheep den", in the former Swiss Institute of Bioinformatics building. I enjoyed the little walks outside in the sun with Theres Fagerberg, when we were getting too hypnotized and stressed out by the computer monitor. Antoine Leimgruber was also episodically there, that old pal from my physics studies at EPFL. In the old building were also Pierre Chodanowski, Mathias Ferber the Warcraft addict, Hamid Hussain-Khan who seems to have read something about almost everything, as well as the Linux guru, the fastest typist in the west, Aurélien Grosdidier. Later came to the group the very knowledgeable Vincent Zoete, and Simon Bernèche, the umbrella sampling specialist, who helped me double checking my weighted histogram algorithms. It's a pleasure to see as new additions to the group Ute Röhrig (back from Rome, Lausanne is cooler!) and Justyna Lagiewka.

As already mentioned, I really appreciated the atmosphere in Wilfred's group as well. The mustache man Riccardo Baron helped me in my first steps with GROMOS. Clara Christ was extremely helpful, in particular when it came to hacking GROMOSXX, which I had known for only a week. She knew how to charm us to an early lunch at 11h55. I also enjoyed interacting with Philippe Hünenberger, who provided a very thorough and useful proofreading of the Jarzynski derivation manuscript. Merijn Schenk put me in a good mood every time I heard his laughter from three offices away, and he invested a lot of effort in a last minute attempt to save the "spaghetti and balls" project. I enjoyed a lot the talks with Vincent Kräutler, the other loud laugher, about the great mysteries of symplecticness. Mika Kastenholtz shared his Poisson-Boltzmann code and sometimes his cynical views. I'll remember "the Dude" Moritz Winger, Daniel "it's a long story" Trzesniak, Jozica Dolenc for her legendary smile, Dirk Bakowies for his critiques about mensa food (but the physics mensa is much better), my chocolate providers Lovorka Perić and Zrinka Gattin, who for sure knows how to cure a cold, the impressive Maria Reif (who else has implemented a reversible pressure scaling integrator after two weeks of Ph.D?), the tidy programmer Markus Christen, the late night working Sankha Ghosh, the sunny accent of Cristina Pereira, the omnipresent jokes of Daan Geerke, the unbeatable french accent of David Kony, the very friendly chats with Bettina Keller and Tim Heinz, and Alice Glättli who had set up my IGC email address, the first mark of my belonging to

the group. I joined the group too late to get to know better, Haibo Yu, Peter Gee, Maria Grazia Giuffreda, and Chris Oostenbrink, who left before I finished. Last but not least, comes my boombastic office mate Bojan Žagrović, who would deserve a whole paragraph of acknowledgments on his own. He was a constant source of intense discussions on topics ranging from the nature of entropy to the latest gossip, or from the meaning of the universe to that of the strange zero in the GROMOS input file. He was on many days my best reason to come to the office.

I should also acknowledge all the people who contributed to providing an excellent computing infrastructure for my thesis. Thanks to the system administrators at different times and locations, Patrick Maurer, Christian Gossens and Matteo Gugliemi in Ursula's group and Aurlien Grosdidier in Olivier's group. The Vital-IT team, Volker Flegel and Roberto Fabbretti also provided great support. With the courtesy of Paul Vickers and Dominique Guillot I was able to use an itanium2 cluster at the Bristol HP lab, with the brilliant assistance of John Henriksson. Pierre Cuendet and Francis Trolliet gave me the opportunity to build a small computing grid with the macs in the EPFL chemistry library and teaching room. Reinhard Kissner administrated the network and my mailserver at ETH. I should also thank Axel Kohlmeyer who remotely helped me to hack CPMD, as well as the people who answered all my questions and challenges on the Gromacs mailing list. Special thanks to Prof. Groove and D. J. Static for providing the vibes. I would also like to thank the administrative assistants for their support, Frau Pfister in the early days at ETH, the excellent Daniela Kalbermatter after my transfer to Wilfred's group, and the friendly and efficient Karin Pasche in Lausanne.

Special thanks to my peeps from Morges who made me stay funky during those years: AN, GR, Webb, Bertrand, Martino, Bibu, Élise, Ader, Amjad, Gael, Aurèle, Raph, and the rest of the crew. Thanks also to those I kept as friends from my physics studies, the "stkritches": Seb, the two Antoine's, Greg, Nico, Phil, Pierre-Ed, Martin, and Joao. The greatest tribute goes of course to my mum, the mother of the world from my point of view, and to my dad, who has been and will be, even if unintendedly, influential. This tribute also goes to my little sister Loïse, who's always there in a little corner of my mind. I would like to thank them from the bottom of my heart for having supported me through my education and during all these years.

Beyond all what words can say, the last dedication goes to the one who has been my marvelous secret garden for almost as long as my Ph.D studies, my beloved Aleksandra.

Contents

Kurzfassung

In den letzten dreissig Jahren hat sich die Molekulardynamik (MD) als ein nützliches Hilfsmittel zum Verständnis biomolekularer Prozesse auf atomarer Ebene und zur Vorhersage bestimmter thermodynamischer Eigenschaften erwiesen. Allerdings beinhalten MD Simulationen die Möglichkeit in einem System auf vielfältige Weise Störungen hervorzurufen, die im Laborexperiment unmöglich wären. Zum einen kann das System gezwungen werden verschiedene thermodynamische Ensemble anzunehmen, in dem Thermostaten für Simulationen bei konstanter Temperatur und Barostate für Simulationen bei konstantem Druck verwendet werden. Zum anderen kann das System beliebigen externen Feldern ausgestezt werden, zum Beispiel um eine Reaktion in die gewünschte Richtung zu treiben. Externe Felder bringen das System in einen Nichtgleichgewichtszustand, in welchem die gewöhnliche statistische Mechanik nicht mehr gültig ist. Durch diese Möglichkeit, das System unter Stress setzen zu können, bieten MD Simulationen eine Basis für Experimente im Bereich der Nichtgleichgewichts-Statistischen Mechanik. Ausser ein Mittel für die Untersuchung molekularer Systeme zu sein hängt MD mit den mehr theoretischen Aspekten der Physik und der Mathematik dynamische Systeme zusammen.

Kapitel 1 dieser Arbeit liefert einen Überblick über MD von einem statistisch-mechanischen Standpunkt. Besonders betont werden die Ergodenhypothese und die Phasenraumverteilungsfunktion für Hamiltonian sche und nicht-Hamiltonian sche Systeme. Die potentielle Energiefunktion, mit der Moleküle in klassischer MD modelliert werden, wird kurz besprochen. Die Kontrolle der Tempertatur eines MD Systems benötigt eine akkurate Berechnung der instantanen kinetischen Energie, welche aus dem Quadrat der Teilchengeschwindigkeiten berechnet wird. Kapitel 2 verdeutlicht, wie die quadrierten Geschwindigkeiten in Zusammenhang mit dem 'leap-frog' Integrator - einem weit verbreiteten diskreten Propagator für die Bewegungsgleichungen - am besten berechnet werden.

Ein wichtiges Ergebnis der Nichtgleichgewichts-Statistischen Mechankik ist die Jarzynski Identität, die es erlaubt Gleichgewichts-Freie-Energieunterschiede aus einer Reihe von Nichgleichgewichtstrajektorien zu berechnen. In Kapitel 3 wird eine neue statistisch-mechanische Herleitung der Jarzynski Identität für die historisch wichtige und häufig verwendete Nosé-Hoover nicht-Hamiltonian sche Dynamik präsentiert. Die Herleitung basiert nur auf den in der MD verwendeten Bewegungsgleichungen und benötigt keine zusätzlichen Hypothesen thermodynamischer Art. Kapitel 4 erweitert diese Herleitung auf eine Vielzahl von Thermostaten und Barostaten basierend sowohl auf Hamiltonianscher als auch nicht-Hamiltonianscher Dynamik. Manche dieser dynamischen Systeme haben praktische Relevanz während andere von theoretischem Interesse für die Verallgemeinerung der Herleitung sind. In Kapitel 5 wird eine Methode um Gleichgewichtseigenschaften aus Nichtgleichgewichtstrajektorien zu extrahieren präsentiert, und eine alternative Fromel wird

gegeben und getestet. Kapitel 6 präsentiert eine Untersuchung des Dissoziationsmechanismus von zwei für Immunologie sehr wichtige Proteinen, das T-Zell Rezeptor und der Haupthistokompatibilitätskomplex. Insbesondere werden freie Energieprofile der Dissoziation für den Wildtyp und einen Mutanten mit der Jarzynski Identität berechnet. Zum Abschluss wird in Kapitel 7 ein Ausblick gegeben.

Summary

In the last thirty years, molecular dynamics (MD) simulation has proved a useful tool for the understanding of biomolecular processes at the atomic level and the prediction of certain of their thermodynamic properties. However, inherent to MD simulation is the additional possibility to perturb the system in a variety of ways impossible in laboratory experiments. First, the system can be forced to adopt desired thermodynamic ensembles, using thermostats for constant temperature, and barostats for constant pressure simulations. Second, arbitrary external fields can be applied to the system, for instance in order to drive a reaction of interest. The external fields bring the system in a nonequilibrium state, where standard statistical mechanics does not apply anymore. Due to this freedom of putting the system under stress, MD simulation provides experimentation grounds for the field of nonequilibrium statistical mechanics. In addition to being an investigative tool for molecular systems, MD relates to more theoretical aspects of the physics and mathematics of dynamical systems.

Chapter 1 of this thesis provides an overview of MD seen from the statistical mechanics standpoint. Emphasis is placed on the ergodic hypothesis and the phase space distribution function for both Hamiltonian and non-Hamiltonian systems. The potential energy function used to model molecules in classical MD is briefly reviewed. Controlling the temperature of an MD system requires an accurate estimation of the instantaneous kinetic energy, calculated from the squared velocity of the particles. Chapter 2, clarifies how the squared velocities are best estimated in conjunction with a discrete propagator widely used to solve the equations of motion, the leap-frog integrator.

An important result of nonequilibrium statistical mechanics is the Jarzynski identity, which allows to estimate equilibrium free energy differences from sets of nonequilibrium trajectories. In Chapter 3, a new statistical-mechanical derivation of the Jarzynski identity is provided for the Nosé-Hoover non-Hamiltonian thermostated dynamics, which is both historically important and widely used in MD. The derivation is only based on the specific equations of motions used in MD, and requires no additional hypothesis of thermodynamic nature, like former derivations do. Chapter 4 extends this derivation to a variety of thermostats and barostats based on both Hamiltonian and non-Hamiltonian dynamics. Some of these dynamical systems are of practical relevance, some others are of theoretical interest for the generalization of the derivation. In Chapter 5, a method to extract equilibrium properties from a weighted averaging along nonequilibrium trajectories is presented, and an alternative formula is provided and tested. Chapter 6 presents the molecular dissociation mechanism of two proteins of great relevance in immunology, the T-cell receptor and the major histocompatibility complex. In particular, dissociation free energy profiles are computed with the Jarzynski identity for the wild-type complex, and a mutant. Finally, an outlook is given in Chapter 7.

Résumé

Dans ces trente dernières années, la simulation de dynamique moléculaire (MD) a fait ses preuves comme instrument utile pour la compréhension de processus biomoléculaires au niveau atomique, et la prédiction de certaines de leurs propriétés thermodynamiques. Cependant, la simulation moléculaire, par son essence même, offre la possibilité de perturber le système de plusieurs façons, impossibles dans une expérience de laboratoire. Premièrement, le système peut être forcé à se conformer à des ensembles thermodynamiques désirés, en utilisant des thermostats pour des simulations à température constantes, et des barostats pour des simulations à pression constante. Deuxièmement, on peut appliquer au système des forces externes arbitraires, par exemple afin de piloter une réaction à laquelle on s'intéresse. Ces forces externes amènent le système dans un état hors-équilibre, où la mécanique statistique standard ne s'applique plus. Grâce à cette liberté d'agir sur le système, la MD fournit un terrain d'expérimentation pour l'étude de la mécanique statistique hors-équilibre. En plus d'être un outil de recherche sur les systèmes moléculaire eux-mêmes, la MD est liée à des aspects plus théoriques des mathématiques et de la physique des systèmes dynamiques.

Le Chapitre 1 de cette thèse fournit une vue d'ensemble de la MD, sous l'angle de la mécanique statistique. L'accent est placé sur l'hypothèse ergodique et la distribution d'espace de phase, pour des systèmes aussi bien Hamiltoniens que non-Hamiltoniens. La fonction d'énergie potentielle utilisée pour modéliser des molécules en MD classique est brièvement récapitulée. Contrôler la température d'un système dynamique demande une estimation précise de son énergie cinétique instantanée, calculée à partir des vitesses des particules au carré. Le Chapitre 2 clarifie la meilleure façon d'estimer ces vitesses au carré en conjonction avec un propagateur discret couramment utilisé pour résoudre les équations du mouvement, l'intégrateur leap-frog.

Un résultat important en mécanique statistique hors-équilibre, l'identité de Jarzynski, permet d'estimer, à partir d'ensembles de trajectoires hors-équilibre, des différences d'énergie libre correspondant à l'état d'équilibre. Au Chapitre 3, une nouvelle dérivation de l'identité de Jarzynski, basée sur la mécanique statistique, est fournie pour la dynamique de Nosé-Hoover. Cette dynamique à température constante a une importance historique, et est de plus couramment utilisée en MD. La dérivation est basée uniquement sur les équations du mouvement spécifiques à la MD, et ne fait appel à aucune hypothèse supplémentaire du type thermodynamique, contrairement aux dérivations existantes. Le Chapitre 4 étend cette dérivation à plusieurs thermostats et barostats basés sur des dynamiques tant Hamiltoniennes que non-Hamiltoniennes. Certains ce ces systèmes dynamiques sont utilisés en pratique, et d'autres revêtent un intérêt théorique pour la généralisation de la dérivation. Le Chapitre 5 présente une méthode pour extraire des pro-

priétés d'équilibre à partir d'une moyenne pondérée le long de trajectoires hors-équilibre. Une nouvelle formule servant le même but est proposée et testée. Le Chapitre 6 présente l'étude du mécanisme moléculaire de dissociation de deux protéines jouant un rôle important en immunologie, le récepteur de cellule T et le complexe majeur d'histocompatibilité. En particulier, des profiles d'énergie libre d'association sont calculés à l'aide le l'identité de Jarzynski pour le complexe naturel ainsi qu'un mutant. Finalement, quelques perspectives sont évoquées au Chapitre 7.

Publications

Chapter 2

Michel A. Cuendet and Wilfred F. van Gunsteren,
"On the calculation of velocity-dependent properties in molecular dynamics simulations using the leap-frog integration algorithm",
Submitted to *J. Chem. Phys.* (2006).

Chapter 3

Michel A. Cuendet,
"Statistical Mechanical Derivation of Jarzynski's Identity for Thermostated Non-Hamiltonian Dynamics",
Phys. Rev. Lett. 96 : 120602 (2006).

Chapter 4

Michel A. Cuendet,
"The Jarzynski identity derived from general Hamiltonian or non-Hamiltonian dynamics reproducing NVT or NPT ensembles",
In press, *J. Chem. Phys.* (2006).

Related publication

Michele Cascella, Michel A. Cuendet, Ivano Tavernelli and Ursula Röthlisberger
"Optical spectra characterization in metalloproteins: Cu(II) photoexcitation in azurin by hybrid TDDFT-molecular dynamics simulations",
Submitted to *Phys. Rev. Lett.* (2006).

Chapter 1

Introduction

It has been said that 'the human mind has never invented a labor-saving machine equal to algebra'. If this be true, it is but natural and proper that an age like our own, characterized by the multiplication of labor-saving machinery, should be distinguished by the unexampled development of this most refined and most beautiful of machines.

Josiah Willard Gibbs, 1887.

1.1 General considerations

Most approaches to material systems in modern science start from the common paradigm that macroscopic properties of matter can be understood by looking at the behavior of its constitutive elements at the microscopic scale. Thermodynamics describes the behavior of large systems (of the order of 10^{23} particles). Statistical mechanics, founded by Boltzmann and Gibbs more than a century ago, bridges the laws of thermodynamics with the microscopic world, populated by atoms and molecules whose motion is governed by classical mechanics. Various models differ in the way specific interactions between atoms are described. But even in very accurate models where forces between atoms are mediated by electrons described on the quantum mechanical level, classical mechanics can still be used with a high level of accuracy for the motions of the nuclei themselves (Born-Oppenheimer approximation). Seeking a deeper and more specific understanding of the properties of matter, we are therefore brought to study the classical dynamics of ensembles of particles.

The computer has become a useful tool to investigate the dynamics of systems with many degrees of freedom, which is not possible analytically. In a molecular dynamics

(MD) computer simulation, many particle-particle interactions are routinely computed, and Newton's equation of motion is integrated at consecutive time steps. The first MD simulation of a fluid of hard disks was published in 1957 by Alder and Wainwright [1]. Since then, MD has spread in various areas of science, including theoretical physics, astrophysics, solid state physics, crystallography, chemistry, and biology. The first successful MD simulation of a protein is due to McCammon, Gelin and Karplus [2] in 1977. An alternative to MD is Monte-Carlo (MC) simulation, in which the dynamics of the system is not integrated, but random conformations are generated, which are compatible with a given thermodynamic state. See Ref. [3] for a comparative introduction to both methods

Despite the ever increasing computer power at hand, both the size of the systems and the time scales accessible to computer simulations are still much below the thermodynamic requirements. Most experiments happen on the macroscopic scales, in which the average collective behavior of a very large number of identical systems is observed. On the other hand, only a small number of particles, or even one single copy of a protein is considered in an MD simulation. Despite this apparent discrepancy, MD simulation results can be related to quantities of thermodynamic nature. The key element bridging the two worlds is the *ergodic hypothesis*. An ergodic system of (few) particles is such that after monitoring the evolution generated by its dynamics for a sufficiently long time, the total time spent by one particle in a given microstate is proportional to the statistical mechanical probability of this microstate. Equivalently, ergodicity means that averaging along one single trajectory of a small system for a long enough time is the same as generating a big number of conformations of the system compatible with a given thermodynamic state, or monitoring a macroscopic system made of many copies of the small system in all its possible microstates. In the following, a more mathematical formulation of this concept is given. Note that it is easier to show that a system is nonergodic than to provide the positive proof of ergodicity, which is available only for very simple systems. It is however commonly believed that all realistic nonlinear many-body systems are ergodic.

A notion related to ergodicity is *mixing*. A dynamical system is called mixing if after a sufficiently long time, trajectories initiated in any small region will spread over the entire available space. Formally, every ergodic system is mixing, but the reverse is not true. The concept of mixing is tightly linked with the *chaotic* character or *Lyapunov instability* of complex dynamical systems. This notion expresses the fact that for many dynamical systems, the solutions to the equations of motion are sensitive to small perturbations in the initial conditions, so sensitive that these perturbations grow exponentially with time. The amount of precision on the initial condition necessary to describe a particular exact trajectory increases linearly with time, and is easily destroyed by the smallest perturbation. Thus, the determinism of mechanics is an illusion. It is the irrelevance of the microscopic initial conditions which makes the systematic study of macroscopic physical properties possible.

Another property of dynamical systems related to ergodicity but formally demonstrated is *Poincaré recurrence*. The Poincaré recurrence theorem states that a system having a finite energy and being confined to a finite volume will, after a sufficiently long time, return to an arbitrarily small neighborhood of its initial microstate. The theorem is based on two premises. First, trajectories of closed dynamical systems do not intersect. Second, the phase space volume of a finite element is conserved under the dynamics (see

Section 1.3). The *Poincaré recurrence time* is the amount of time elapsed before the recurrence happens. In mixing systems, it also gives a sense of how much time is needed for an exhaustive sampling of the available configuration space. Unfortunately, Poincaré recurrence times get huge for any system with interesting complexity. Order of magnitude estimates suggest that Poincaré recurrence times are of the order $\sqrt{\Omega}$ in a discretized space with Ω discrete states. Such times are effectively infinite (exceeding the age of the universe) once the number of particles is of the order ten to a hundred. Given this, it seems that the only way to circumvent the quasi-infinite time necessary for exhaustive sampling and accessing thermodynamic properties is to have macroscopic amounts of molecules and time to average on. This apparently leaves little chance to MD.

The driving potential for all processes in physical systems at constant volume V and temperature T is the free energy $F = U - TS$, with U the average energy of the system, and S the entropy of the system. Entropy can be thought of as the amount of uncertainty (or "mixedupness" in the phrase of Gibbs) which remains about a system, after its observable macroscopic properties have been taken into account. More precisely, according to Shannon's definition, entropy is the amount of further information needed to define the detailed microscopic state of the system that remains uncommunicated by a description solely in terms of the macroscopic variables of classical thermodynamics. Entropy is thus a state function corresponding to the total number of microstates compatible with a given macrostate, (N,V, T for the canonical ensemble), and the Hamiltonian function describing the system. Unlike energy, entropy cannot be determined from a single configuration (positions and energies) of a microscopic system. Thus, computing the entropy of a system is a difficult task, since it involves counting all available microstates, which is impossible without extensive sampling. In regard to the preceding paragraph, this seems impossible for complex systems. There are however a number of reasons explaining how MD can possibly be the useful tool it is.

- In practice we are not interested in absolute free energies. Only the free energy difference ΔF_{AB} between two states A and B matters to determine their relative stability. Therefore, an MD simulation only needs to sample the regions where the density of microstates varies between condition A and condition B.

- We are often interested in only a part of the phase space, not its full extent. Such a local macrostate can be for example demarcated by a high energy barrier that the system cannot overcome. A typical example of this is the case of proteins. Often only the folded conformation is of interest, and the much more extended space of all unfolded conformations is ignored. Ergodicity limited to such a local macrostate is called quasi-ergodicity.

- The fact that the dynamics is mixing ensures that after a sufficient time, the set of sampled microstates becomes dense in the set of all accessible microstates. Formally, a set Y is dense in X if for any point x in X, any neighborhood of x contains at least a point from Y. Despite the ruggedness the free energy surface of a complex system can display, it is locally smooth, with features not much smaller than the size of a particle. This, together with dense sampling favors at least approximate sampling of all relevant free energy features.

With these three considerations, it becomes more plausible that MD simulation is able to give useful free energy and entropy results [3, 4, 5, 6], as it has done over about fifty years for a variety of interesting problems.

Besides the ergodicity problem central to MD, there is another question that needs to be addressed before one can accept MD as a faithful description of processes obeying the general laws of thermodynamics. According to the second law of thermodynamics, all macroscopic phenomena follow an irreversible evolution, which leads inexorably to a future where the order of the past has been replaced by relative disorder, and where the entropy of the system is bigger. This irreversible growth of disorder can be described as a loss of the information required to describe or recover the past. This loss makes it impossible to reverse the evolution. However, Newton's equations of motion describing the movement of the constitutive particles of matter are intrinsically time reversible, as is Maxwell's equations. This means that, considering a given trajectory, Newton's equations admit a second solution with the sign of time and velocities reversed, which traces out exactly the same coordinate values in reverse order and obeys the same relation between force and acceleration. Loschmidt's paradox is simply stated: "How can the irreversible second law of thermodynamics be compatible with, and result from, an underlying time-reversible mechanics?". Zermélo expressed a similar argument: The Poincaré recurrence implies that isolated systems cannot display irreversible behavior, because they are bound to return to exactly the same microscopic configuration at some time point.

Recent advances have shown however that coupling mechanical systems to correctly modeled boundary conditions is enough to explain the symmetry breaking associated with irreversible processes and resolve the conceptual problems linked with conservative mechanics. Hoover [7] calls the augmented mechanics "thermomechanics" to emphasize its links to thermodynamics through the explicit incorporation of thermal effects. Tightly linked to this symmetry breaking is the theory of chaos, which says that the amount of information necessary to find back a given initial condition with a given accuracy by reversing Newton's equations increases exponentially with time. In relation to thermomechanics and Loschmidt's paradox, recent finding obtained by computer simulation showed [8, 9, 10] that when the system is allowed to dissipate heat, its tendency to destroy information is larger in the forward time evolution. Typically, the forward dissipative trajectory is less sensitive to perturbations, and thus more stable, than is the time reversed backward one. The entropy produced by a process is linked to the Lyapunov exponents of the system, which represent the average rate of divergence of two neighboring trajectories (see Section 1.5). The spectrum of these Lyapunov exponents in turn depends on the coupling of the system to its environment (or thermostat). Therefore, this coupling should be carefully included in the model. Conversely to Hoover's view, Evans argues that the answer to Loschmidt's paradox lies in the fluctuation theorem [11, 12] (see Table 1.1), which quantifies the relative probabilities of observing a trajectory producing or consuming entropy.

We have seen that in macroscopic experiments or in equilibrium MD simulations, standard theories to describe chemical and conformational equilibria as well as reaction kinetics apply. However, some recent experiments of a different kind were able to look at very small scale or single-molecule events, in a time resolved fashion. For example, Searles, Evans, and others [13], by following the trajectory of a colloidal particle captured in an optical trap, were able to show that the second law of thermodynamics can be violated

over short time scales. Another experiment by Liphardt *et al.* [14], in which the work necessary to unfold a single RNA strand is recorded, shows that there are big fluctuations among different realizations of a nonequilibrium process. These two experiments show that for short time scales and small system sizes the standard laws of thermodynamics do not apply. Interestingly, while this kind of phenomena is difficult to observe experimentally, MD simulation is even better suited to study them than thermodynamic equilibria.

1.2 Phase distribution function

The evolution of a N-particle system in 3 dimensions is described by $3N$ coordinates $q = (q_1, ..., .q_N)$ and $3N$ momenta $p = (p_1, ...p_N)$. The system at a given time t is uniquely described by a point $\Gamma \equiv (q, p)$ in the $6N$ dimensional space of coordinates and momenta called phase space. The evolution of γ is governed by the dynamics, a set of differential equations of the form $\dot{\Gamma} = \dot{\Gamma}(\Gamma)$. As time progresses, the phase point Γ traces out a path which we call the phase space trajectory of the system. A trajectory is uniquely defined by its initial condition, namely the microstate $\Gamma(0)$ of the system at time $t = 0$.

A thermodynamic state describes an ensemble of individual N-particle systems, that is an essentially infinite number of systems characterized by identical dynamics and identical state variables (e.g. N,V,T) but different initial conditions $\Gamma(0)$. We wish to consider the average behavior of a collection of macroscopically identical systems distributed over a range of initial microstates. In generating the ensemble we make the usual assumptions of classical mechanics. We assume that it is possible to know all the positions and momenta of an N-particle system to arbitrary precision at some initial time, and that the motion can be calculated exactly from the equations of motion. In continuous dynamics, the ensemble contains an infinite number of individual systems, so that the number of systems in a particular microstate may be considered to change continuously as we pass to neighboring microstates. With this assumption, we consider an infinitesimal element of phase space located at Γ. The fraction δN of systems, which at time t have coordinates and momenta within dq of q and dp of p is used to define the phase space distribution function $\rho(q, p, t)$, by

$$\frac{\delta N}{N} = \rho(q, p, t)dV = \rho(q, p, t)dqdp \tag{1.1}$$

The second equality is valid only in Cartesian coordinate systems, where the volume element $dV = dqdp$, i.e. the metric factor is unity. Although trivial in the case of Hamiltonian systems, we will see that the choice of such a coordinate system is not always most convenient or even possible in the case of non-Hamiltonian systems. The total number of particles in the system being fixed, integrating over the whole phase space provides a normalization of the distribution function, $\int \rho(\Gamma, t))d\Gamma = 1$. The conservation law for ρ states that there must be an equivalent number of trajectories entering and leaving a volume element [15],

$$\left.\frac{\partial \rho}{\partial t}\right|_{\Gamma} + \nabla_{\Gamma} \cdot \left(\dot{\Gamma}\rho\right) = 0. \tag{1.2}$$

This is the *Liouville equation*. It can equivalently be expressed using the streaming or

total time derivative of the distribution function,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = -\rho\kappa(\Gamma). \tag{1.3}$$

The quantity

$$\kappa(\Gamma) \equiv \frac{\partial}{\partial\Gamma} \cdot \dot{\Gamma}(\Gamma) \tag{1.4}$$

is called the *phase space compression factor*, because it expresses the rate of change of the phase space density at point Γ. $\kappa(\Gamma)$ is also sometimes called the divergence of the the flow generated by the dynamics. It is readily obtained from any given equations of motion. It should be noted that the Liouville equation is a very general result, since it has been obtained without reference to the actual equations of motion. Its correctness doesn't require the existence of a Hamiltonian to generate the equations of motion. The equation relies on only two conditions: that ensemble members cannot be created or destroyed and that the distribution function is sufficiently smooth for the appropriate derivatives to exist.

Based on the phase distribution function, we can give a definition for the entropy mentioned in the previous paragraph. The Gibbs entropy is given by [16]

$$S_G = -k_B \int d\Gamma \rho(\Gamma, t) \ln \rho(\Gamma, t). \tag{1.5}$$

For many systems it is apparent that after possible initial transients, the N particle distribution function $\rho(\Gamma, t)$, becomes essentially time independent, $\frac{\partial\rho}{\partial t} = 0$. This is evidenced by the fact that the macroscopic properties of the system relax to fixed average values. This obviously happens for equilibrium systems. It also occurs in some nonequilibrium systems, so-called nonequilibrium steady states [15]. We will call all such systems stationary.

For a stationary system, we may define the ensemble average of a phase variable or observable $\mathcal{A}(\Gamma)$, using the stationary distribution function $\rho(\Gamma)$, so that,

$$\langle\mathcal{A}\rangle = \int d\Gamma \mathcal{A}(\Gamma)\rho(\Gamma). \tag{1.6}$$

On the other hand, a time average of the same phase variable is defined as

$$\bar{\mathcal{A}} = \lim_{\tau\to\infty} \frac{1}{\tau} \int_0^\tau dt\ \mathcal{A}(\Gamma(t))$$

An ergodic system is a stationary system for which $\bar{\mathcal{A}} = \langle\mathcal{A}\rangle$. Liouville's theorem ensures that the notion of time average makes sense, but ergodicity does not follow from Liouville's theorem. We call *ergodic transformation* the replacement of a phase average derived from statistical mechanics by a time average to be used in an MD simulation. When studying transient nonequilibrium processes for which the distribution function $\rho(\Gamma, t)$ is not stationary, the ergodic transformation is not possible. Therefore, the only route to phase space sampling is to effectively start different simulations with a set of different initial conditions Γ_0. This is the spirit of the Jarzynski identity [17] (JI), which allows to estimate the equilibrium free energy difference $F_{AB} = F_B - F_A$ between an initial state A and a final state B by averaging the exponential of the work W_{AB} performed to drive the

system out of equilibrium from A to B. This average is precisely taken over the initial conditions of a collection of independent trajectories,

$$e^{-\beta F_{AB}} = \langle e^{-\beta W_{AB}(\Gamma_0)} \rangle_{\Gamma_0} \tag{1.7}$$

This result belongs to a family of four recent nonequilibrium relations (see Table 1.1), which are briefly introduced at the end of Section 1.5. In Chapters 3 and 4, an original derivation of the JI is provided, which follows directly from the specific dynamics of the system. The choice of the initial conditions is critical as it should represent the best possible sampling of the initial state. In most reported applications of the JI, the set of initial conditions is composed of decorrelated configurations with momenta taken from a long equilibrium trajectory in the initial state A. In Chapter 6 the JI is applied with a different protocol, more in a Monte-Carlo spirit, in which each initial condition comes from a separate short equilibration started with independent random initial velocities.

We have seen that the ergodic transformation is not directly possible when following a single nonequilibrium trajectory. In Chapter 5 we show non-stationary phase averages can be reweighted in such a way that the ergodic transformation can again be applied. This means that it is possible to obtain the equilibrium free energy difference ΔF_{AB} by following a single nonequilibrium trajectory switching between states A and B.

1.3 Hamiltonian systems

If the system can be described with an autonomous Hamiltonian $H(q,p)$, the associated canonical equations of motion are

$$\dot{q}_i = \frac{\partial H}{\partial p_i}(p,q), \tag{1.8}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}(p,q). \tag{1.9}$$

It is a simple matter to show that $\kappa(\Gamma) = 0$ for a Hamiltonian system. The phase space is then called incompressible, and the Liouville equation (1.3) takes on a simpler form,

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = 0. \tag{1.10}$$

The existence of a Hamiltonian is a sufficient, but not necessary condition for the phase space compression factor to vanish. The Liouville equation (1.10) implies that the volume element following a point (p,q) cannot change with time. This is related to a series of mathematical conditions describing the evolution of an isolated Hamiltonian system. The Hamiltonian evolution is said to generate *canonical transformations* in time, which means that if the initial coordinates $(p(0), q(0))$ satisfy relations (1.8) and (1.9), then the coordinates at time t, $(p(t), q(t))$ will also satisfy relations (1.8) and (1.9). Hamilton's equations can be rewritten using the so called *symplectic* notation [18],

$$\dot{\Gamma} = \mathbb{M}\frac{\partial H}{\partial \Gamma}, \qquad \mathbb{M} = \begin{pmatrix} 0 & I \\ -I & 0 \end{pmatrix},$$

where I is the identity matrix of dimension $3N$. If $\mathbf{J}_{ij} = \partial\Gamma_i/\partial\Gamma_j(0)$ denotes the Jacobian matrix of the transformation $\Gamma(0) \to \Gamma(t)$ along a Hamiltonian trajectory, then the *symplectic condition* holds,

$$\mathbf{J}\mathbb{M}\mathbf{J}^{\mathbf{T}} = \mathbb{M}. \tag{1.11}$$

$\mathbb{M}$ can also be seen as the metric tensor for the Hamiltonian system. It is easy to see that (1.11) implies phase space volume conservation, since the infinitesimal volume elements at times 0 and t are related by $d\Gamma(t) = Jd\Gamma(0)$, and (1.11) imposes that $J = |\det \mathbf{J}| = 1$. The symplectic condition is however more general than just phase space volume conservation. For example, in the language of differential geometry on manifolds, the canonical transformation generated by Hamilton's equations is said to be a symplectomorphism, meaning that it conserves the symplectic two-form $dp \wedge dq$. [19]

If we introduce the Liouville operator $i\mathcal{L} \equiv \dot{\Gamma} \cdot \nabla_\Gamma$, the evolution of the Hamiltonian system can be written as [15]

$$\Gamma(t) = e^{i\mathcal{L}}\Gamma(0). \tag{1.12}$$

This is a formal solution of (1.8) and (1.9). The exponential operator $e^{i\mathcal{L}}$ defines the flow of the Hamiltonian system, which was shown to satisfy the symplectic condition (1.11). We also note that the adjoint of the Liouville operator corresponds to its inverse, that is $e^{i\mathcal{L}}$ is unitary. This implies that the trajectory is exactly time reversible.

Based on the Trotter decomposition of the Liouville operator on small time steps Δt, discrete time propagators can be obtained [18, 20], for which each time step is a transformation of the type (1.12). This class of propagators, called *symplectic integrators*, formally satisfies condition (1.11), conserve the phase space volume element, and are time-reversible. A key property of symplectic integrators is that they integrate exactly an approximate Hamiltonian, which is a small perturbation of the exact physical Hamiltonian. This means for example that symplectic integrators are not inclined to numerical energy drift. A variety of integrators of different orders can be derived [21] using the above method, including the velocity Verlet and position Verlet [22] algorithms.

In MD simulation, practical integrators must only involve a single force calculation, since this is by far the most computationally intensive part of the calculation. An efficient integrator commonly implemented in MD packages is the leap-frog [23] scheme,

$$x(t + \Delta t) = x(t) + v(t + \tfrac{\Delta t}{2})\Delta t + \mathcal{O}\left((\Delta t)^3\right), \tag{1.13}$$

$$v(t + \tfrac{\Delta t}{2}) = v(t - \tfrac{\Delta t}{2}) + \frac{1}{m} f(t)\Delta t + \mathcal{O}\left((\Delta t)^3\right). \tag{1.14}$$

Interestingly, it seems [21] that the leap-frog algorithm cannot be derived directly with the Trotter decomposition method. It is rather derived by subtracting truncated Taylor expansions at consecutive time steps. However, the leap-frog produces the same trajectories as the velocity Verlet. It should therefore share its good properties.

The leap-frog algorithm provides accurate velocities only at half time steps. In Chapter 2, the most appropriate way to approximate v_n^2 at full time steps is discussed. This question is of importance for practical application, since using the square of the standard leap-frog velocity $v_n^2 = \frac{1}{4}\left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}}\right)^2$ instead of $v_n^2 = \frac{1}{2}\left(v_{n-\frac{1}{2}}^2 + v_{n+\frac{1}{2}}^2\right)$ is shown to result in a systematic bias in the system kinetic energy and pressure.

1.4 Simulating NVT and NPT ensembles: extended dynamics.

The Hamiltonian dynamics described in the previous paragraph generates system conformations at a fixed energy, determined by the initial conditions. We say that the Hamiltonian dynamics generates the NVE ensemble, which is described by the microcanonical partition function [16]

$$\Omega(N,V,E) = \frac{1}{h^{3N}N!}\int d\Gamma \delta\left[H(\Gamma) - E\right].$$

In the prefactor, the term containing Planck's constant h ensures that the partition function is dimensionless and consistent with counting quantum states. The $N!$ term accounts for indistinguishable particles. The corresponding microcanonical distribution function is $\rho(\Gamma) = \frac{1}{\Omega}\delta\left[H(\Gamma) - E\right]$. However, many laboratory experiments are conducted in solution, where the system is thermally coupled to its environment. In this case, the temperature of the system is fixed, not its energy. This corresponds to the NVT or canonical ensemble. A fundamental reason to study systems coupled to heat baths is that isolated systems cannot display any dissipative behavior. For the system to be able to relax to a state of maximum entropy, as the second law of thermodynamics describes for all macroscopic systems, there has to be a way to exchange energy with the environment [7]. Indeed, the canonical distribution function can be defined as the distribution maximizing the Gibbs entropy S_G (1.5),

$$\rho(\Gamma) = \frac{1}{Z}e^{-\beta H(\Gamma)}. \tag{1.15}$$

The corresponding canonical partition function is

$$Z(N,V,T) = \frac{1}{h^{3N}N!}\int d\Gamma\ e^{-\beta H(\Gamma)}.$$

The Helmholtz free energy F corresponding to the canonical ensemble is

$$F = U - TS, \tag{1.16}$$

where U represents the average energy of the system. The free energy is related to the partition function through $F = -\frac{1}{\beta}\ln Z$. Many methods have been proposed to calculate free energies and entropies, because these quantities are among both the most sought-after and most elusive properties of a system. In molecular simulations, it is of interest to study the evolution of the system along a given reaction coordinate ξ. The propensity of the system to move from A to B along the reaction coordinate depends on the free energy difference between the two sub-states in which ξ is found close to A or B. The free energy profile relating sub-states along the reaction coordinate is called a potential of mean force (PMF). It is expressed by a projection of the partition function on the hyperplane $(\xi(\Gamma) = \xi)$,

$$\Delta F(\xi) = -\frac{1}{\beta}\ln\int d\Gamma \delta\left[\xi(\Gamma) - \xi\right] e^{-\beta H(\Gamma)}.$$

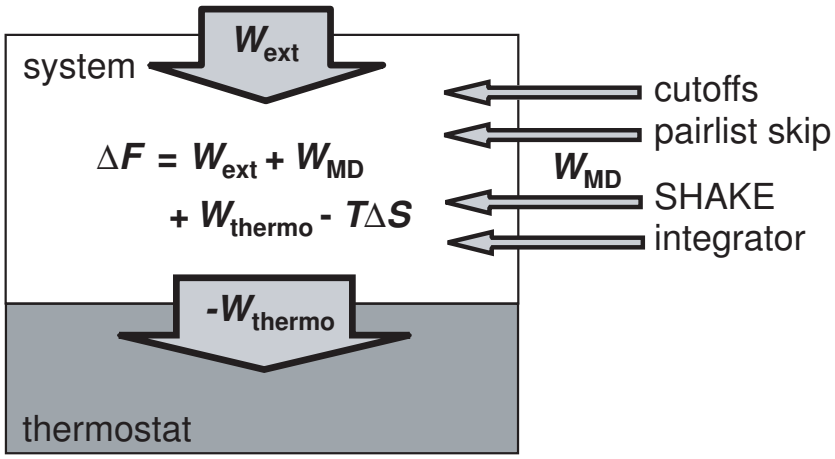

Figure 1.1: *Schematic representation of a nonautonomous thermostated MD system. Red arrows represent heat or work flows. A steady state corresponds to the situation where* $W_{\mathrm{ext}} + W_{\mathrm{MD}} = -W_{\mathrm{thermo}}$.

Note that the PMF is defined up to an additive constant. Chapter 6 uses the Jarzynski identity (1.7) and the umbrella sampling method [24] to calculate the dissociation PMF of a protein-protein complex. Chapter 5 focuses on a nonequilibrium free energy method which is also suited for the calculation of PMFs.

Often in laboratory experiments, the volume of the experimental system is not fixed, but the external pressure is. This situation corresponds to the isobaric-isothermic or NPT ensemble. The corresponding distribution function is

$$\rho(\Gamma) = \frac{1}{Z} e^{-\beta\left[H(\Gamma)+PV\right]}.$$

The Gibbs free energy defined for the NPT ensemble is

$$G = U - TS + PV.$$

A number of factors impede a precise computer simulation of the standard statistical mechanical ensembles. First, one can obviously not simulate a realistic heat bath with a large number of degrees of freedom. Second, the nature of the coupling between the system and the bath, which is deliberately disregarded in statistical mechanical and thermodynamical approaches, needs to enter explicitly the MD equations of motion, which are then no longer Newtonian. The modified dynamics that serve this purpose are called thermostats. A thermodynamic picture of a MD system coupled to a thermostat is shown on Fig. 1.1. There, the energy variation ΔU of the system expected from (1.16) is mediated by its different couplings to the outside. These couplings are expressed as different kinds of work: the work W_{ext} performed by an external agent (nonautonomous dynamics), W_{thermo} provided by the thermostat, or W_{MD} resulting from the undesired side effects of the algorithms used (see Section 1.6).

In the early eighties, Nosé introduced [25] a temperature dependent Hamiltonian, which was formally proven to produce a canonical distribution for the physical degrees of freedom (q, p). This was an improvement over former *ad hoc* temperature control methods [26]. The Nosé thermostat works by coupling the physical system to a couple of additional degrees of freedom s and p_s,

$$H_N(q, \tilde{p}, s, p_s) = H(q, \frac{\tilde{p}}{s}) + \frac{p_s^2}{2Q_s} + N_{df} k_B T \ln s.$$

Here Q_s is a pseudo-mass determining the time scale of the thermostat. The Hamiltonian equations of motion for $\tilde{p}$, q, p_s and s are derived from the Hamiltonian in the standard fashion. These variables follow a microcanonical distribution, and are integrated with respect to the Nosé time $\tilde{t}$. The variable $s = \frac{d\tilde{t}}{dt}$ can be seen as a time scaling factor. The true physical momentum $p = \frac{\tilde{p}}{s}$ is canonically distributed, but corresponds to sampling at uneven intervals in the real time t. This uneven time sampling is not very convenient for the calculation of dynamical properties.

Hoover [27] rewrote the Nosé equations of motion with the physical momentum p instead of the momentum $\tilde{p}$ in Nosé time, and the variable $\eta = \ln s$. This gives the Nosé-Hoover (NH) thermostat

$$\begin{aligned} \dot{r}_i &= \frac{p_i}{m_i}, \qquad i = 1, ..., N, \\ \dot{p}_i &= -\frac{\partial H}{\partial r_i} - \frac{p_\eta}{Q} p_i, \\ \dot{\eta} &= \frac{p_\eta}{Q}, \\ \dot{p}_\eta &= \sum_i \frac{p_i^2}{m_i} - N_{df} k_B T \ , \end{aligned} \tag{1.17}$$

with Q a pseudo-mass. These dynamical equations are no longer Hamiltonian. But the physical momentum p is now canonically sampled at regular time intervals. The variable η has become auxiliary, since none of the other variables is coupled to it. In addition, the NH dynamics makes it possible to easily couple different regions of the system to different thermostats, possibly at different temperatures. The NH thus became a method of choice in thermomechanics, for the theoretical study of nonequilibrium MD and chaos phenomena [7, 28, 29, 30, 31, 32, 8, 9, 33], as well as in standard MD simulations. A number of improved thermostats based on the NH principle have been proposed. These include chains of thermostats, control of higher moments of the velocity distribution, generalized coupling terms, or control of the pressure via coupled box volume scaling. Some of these methods are reviewed in Chapter 3. Note that approximately at the same time as the NH, the weak coupling or Berendsen thermostat was proposed [34], based on a first-order (exponential) relaxation of the system temperature to the reference temperature, as opposed to the second-order (oscillatory) relaxation of NH. Although the Berendsen thermostat was never shown to reproduce any known thermodynamical ensemble [35], it is still widely used in MD applications, because of its fast convergence and the absence of spurious oscillations.

Twelve years after the discovery of the NH equations, Dettmann and Morriss [36] found that there exists a Hamiltonian which generates equations equivalent to the scaled Nosé equations, while retaining the Hamiltonian character. Bond *et al.* [37] independently found that this Hamiltonian corresponds to a Poincaré transformation of time, and called it the Nosé-Poincaré (NP) Hamiltonian,

$$H_{NP} \equiv sH_N.$$

In Chapters 3 and 5, it is emphasized that the NH thermostat also "works", i.e. yields canonical sampling of the physical variables, in the case where the potential energy term has an explicit time dependence (nonautonomous system). In a recent contribution, Struckmeier [38] showed that the Poincaré transformation allows to map any time dependent Hamiltonian system onto a time independent one. At the end of Chapter 3, it is shown that a nonautonomous version of the NP dynamics can be clearly formulated using the formalism of Struckmeier. Based on these premises, the main object of Chapter 3 is to show that various general forms of the NH and NP thermostated dynamics directly imply the JI (1.7).

1.5 Non-Hamiltonian and nonautonomous dynamics

The NH scheme modifies the dynamics, such that it becomes non-Hamiltonian. The Jacobian of a coordinate transformation generated by the evolution is not anymore unity, and the symplectic condition (1.11) doesn't hold anymore. Therefore, Tuckerman *et al.* [39] proposed a formalism in which the system phase Γ evolves in a Riemannian manifold with metric tensor $\mathcal{G}$, instead of the Euclidean space. Accordingly, the invariant volume element will be

$$dV = \sqrt{g(\Gamma, t)} d\Gamma,$$

where, the metric factor $g(\Gamma, t)$ is the determinant of the metric tensor $\mathcal{G}$, and is determined from the dynamics by

$$\sqrt{g(\Gamma, t)} = \exp\left\{ - \int dt \; \kappa(\Gamma(t), t) \right\},$$

with $\kappa(\Gamma, t)$ the phase space compression factor defined in (1.4). The phase space metric factor effectively follows the compression effects generated by the dynamics. In this coordinate system, considering the non-Hamiltonian system isolated, a microcanonical phase distribution function $f(\Gamma, t)$ is determined from the motion invariants, such that instead of (1.1), the number of microstates per volume element is expressed as

$$\delta N = f(\Gamma, t) \sqrt{g(\Gamma, t)} d\Gamma.$$

Here, $\frac{\mathrm{d}f}{\mathrm{d}t} = 0$ and all phase space compression effects are included in the metric factor. The conservation of the number of microstates is expressed by a generalized Liouville equation which is similar to (1.2), with $\rho(\Gamma, t) = f(\Gamma, t)\sqrt{g(\Gamma, t)}$. Formally however, both formalisms are not completely equivalent, and there is a controversy [40], as to which of

the ρ or $\sqrt{g}$ representations should be used (see Chapter 5). The formalism of Tuckerman *et al.* [39, 41] however allows a straightforward construction of the density function. In the case where the dynamics represents a physical system coupled to an environment, it provides a convenient analysis tool to find the distribution spanned by the physical variables [41], which are are central to the developments in Chapter 3 and 5.

Note that recent developments have shown that the metric tensor can be determined [42] explicitly for any dynamical system. This formalism can be expressed [43] in a very general geometrical framework [44] using differential forms on manifolds. In particular, a condition similar to the symplectic condition (1.11) holds along the evolution [45],

$$\mathcal{G}_{ij}(\Gamma(0)) = \frac{\partial \Gamma_k}{\partial \Gamma_i(0)} \frac{\partial \Gamma_l}{\partial \Gamma_j(0)} \mathcal{G}_{kl}(\Gamma). \tag{1.18}$$

Similarly to the symplectic integrators mentioned above, it is possible for arbitrary dynamics to construct [45, 46] integrators based on the Trotter decomposition of the Liouville operator (1.12), that are consistent with condition (1.18). As symplectic integrators provide energy and phase space volume conservation in the Hamiltonian case, correct integrators for non-Hamiltonian and nonautonomous systems should provide energy and phase space volume variations consistent with what is expected from the exact thermostated equations of motion, i.e. consistent with (1.18).

Irreversible work performed on the system will inevitably transform into heat, which is in turn extracted by the thermostat (see Fig. 1.1. This flow of energy through the system has as a consequence that the average phase space compression factor $\langle \kappa(\Gamma, t) \rangle$ takes a negative value. This can be seen for the Nosé-Hoover dynamics (1.17), where $\kappa(\Gamma) = -p_\eta/Q$. If the instantaneous temperature is higher than the target temperature T, p_η is positive, and thus κ negative. The Liouville equation (1.3),

$$\frac{\mathrm{d} \ln \rho}{\mathrm{d}t} = -\kappa,$$

shows that, on average, the phase space density increases. Alternatively, one can say that the infinitesimal volume element around Γ shrinks, which corresponds to the so-called *phase space compression*. The phase-space distribution effectively collapses onto an ergodic multifractal attractor [47]. This observation that any stable stationary state must correspond to average positive friction and to shrinkage of the phase volume was called by Hoover [7] the mechanical analog of the second law of thermodynamics. This tendency is linked to the Lyapunov instability phenomenon. The Lyapunov spectrum of the system is the set of Lyapunov exponents $\{\lambda_i\}_{i=1}^{3N}$ which express the exponential rate of divergence of a small initial perturbation along each dimension. Though both negative and positive Lyapunov exponents coexist in chaotic problems, the negative exponents expressing the loss of information have the upper hand. The following relation holds for the entropy change of the external environment surrounding the system,

$$\frac{\mathrm{d}S^{\mathrm{external}}}{\mathrm{d}t} = -k_B \sum_{i=1}^{3N} \lambda_i > 0.$$

This corresponds to the net rate of entropy extraction from the system, which is on average positive. The resulting overall *entropy production* is characteristic of nonequilibrium flows [33, 7].

The definition of the entropy of the system itself poses a conceptual problem. As ρ tends to infinity, the Gibbs entropy as defined in (1.5) seems to diverge to $-\infty$. Since a nonequilibrium steady state (for example, constant flow of heat through the system of Fig. 1.1) has a defined energy, it should have a defined entropy as well. This could indicate that the Gibbs formulation of the entropy is not appropriate for nonequilibrium systems. Interestingly, it was shown [48] that introducing an infinitesimal white noise term in the dynamical equations prevents the Gibbs entropy to effectively diverge. Evans [49] also noted that the projection of the phase space density does not diverge in subspaces of dimension lower than the dimension of the steady state attractor. For macroscopic systems where the dimensional loss is tiny and the weak field approximation holds, the contribution of these lower order densities sum up to the local equilibrium Gibbs entropy of linear irreversible thermodynamics.

The situation is different in the formalism of Tuckerman *et al.* [50, 39]. The separation of the density f from the metric $\sqrt{g}$ allows a formulation of the Gibbs entropy other than (1.5),

$$\tilde{S} = -k_B \int d\Gamma \sqrt{g} f \ln f. \tag{1.19}$$

Since $\sqrt{g}d\Gamma$ is invariant under the evolution and $\dot{f} = 0$, the entropy $\tilde{S}$ (1.19) takes a finite value in a nonequilibrium steady state, where

$$\frac{\mathrm{d}\tilde{S}}{\mathrm{d}t} = 0.$$

Thus, the entropy calculated according to (1.19) satisfies the following condition: The entropy of a *closed* system – even if it is out of equilibrium – is conserved. The form $\tilde{S}$ is in addition covariant under coordinate changes. Some authors have however argued [51, 40] that all compression effects end up in $\sqrt{g}$, which ultimately becomes ill-defined as a metric factor. It is not clear which one of the the two qualitatively very different entropy formulations (1.5) and (1.19) is the most appropriate, both reducing to the same expression for Hamiltonian systems. The question of the entropy of a nonequilibrium state remains open.

Desipte this lack of proper definitions for the nonequilibrium entropy and free energy, important new relationships describing nonequilibrium processes have recently been discovered. First Evans and coworkers devised the fluctuation theorem [11] in 1993 and the Kawazaki identity [52] in 1995. This seminal work was followed by the discovery of the Jarzynski identity [17] in 1997 and the Crooks theorem [53] in 1999. These equations are valid no matter how far from equilibrium the process occurs, and therefore form a family of results which might become the fundament of a brand new branch of statistical physics in the coming years. In the following, we shortly introduce the four nonequilibrium relations, which are summarized in Table 1.1.

The nonequilibirum relations display a striking degree of symmetry among them, as well as with prior equilibrium results shown in the first column of Table 1.1. The Zwanzig formula [54] allows to calculate the free energy difference ΔF_{AB} between two states A and B described by Hamiltonians H_A and H_B with $\Delta H_{AB} = H_B - H_A$. It is effectively a perturbative approach, in which the density of state B is averaged over conformations

Table 1.1: *Synthetic view of four nonequilibrium statistical mechanical relations which remain valid far from equilibrium (second and third columns). The first column displays the standard Zwanzig and Bennett equilibrium relations for comparison.*

Equilibrium energy	*Work from state A to B*	*Dissipation*
Zwanzig [54] $\left\langle e^{-\beta\Delta H_{AB}}\right\rangle_A = e^{-\beta\Delta F_{AB}}$	Jarzynski [17] $\left\langle e^{-\beta W_{AB}}\right\rangle_A = e^{-\beta\Delta F_{AB}}$	Kawazaki identity [52] $\left\langle e^{-\Sigma_t}\right\rangle = 1$
Bennett [55] $\frac{\left\langle \min(1,e^{-\beta\Delta H_{AB}})\right\rangle_A}{\left\langle \min(1,e^{-\beta\Delta H_{BA}})\right\rangle_B} = e^{-\beta\Delta F_{AB}}$	Crooks [53] $e^{-\beta W}\frac{P_{AB}(W)}{P_{BA}(-W)} = e^{-\beta\Delta F_{AB}}$	fluctuation theorem [11] $e^{-\Sigma_t}\frac{P(\Sigma_t)}{P(-\Sigma_t)} = 1$

of state A. An accurate calculation requires that the two states A and B overlap well in phase space. The Bennett ratio [55] expresses the optimal way of combining a perturbation from A to B obtained from a trajectory in A and the reverse perturbation from B to A obtained from a trajectory in B.

The first nonequilibirum relation in Table 1.1 is the JI, which we have already introduced above (1.7). A more complete discussion of this result can be found in the introduction of Chapters 3, which provides a derivation of the JI in the case of thermostated dynamics. Note that the average in the JI is over equilibrium conformations in state A, like in the Zwanzig equation. W_{AB} is the dissipative work performed in the process from A to B. If the reverse process is also available, the work W_{BA} can also be measured, and the Crooks relation [53] can be applied (see Table 1.1). Here, $P_{AB}(W)$ represents the probability of observing a given work value W in the forward process from A to B, and $P_{BA}(W)$ is the probability of observing this value in the reverse process from B to A. One application of the Crooks identity is to measure the forward and reverse work distributions $P_{AB}(W)$ and $P_{BA}(W)$ over many realizations of a given process, and then find the value W at which the distributions intersect, which gives the free energy difference ΔF_{AB}.

The argument of the fluctuation theorem [11] and the Kawazaki identity [52, 56], the dissipation function Σ_t, is a measure of the system's reversibility under an external field. The exact form of Σ_t depends on the ensemble which one considers [12, 57]. In the canonical ensemble,

$$\Sigma_t = \beta\big[H\left(\Gamma(\tau)\right) - H\left(\Gamma(0)\right)\big] - \int_0^\tau dt\ \kappa(\Gamma(t)).$$

The ensemble average of the dissipation function $\langle\Sigma_t\rangle$ is the time integral of the irreversible entropy production rate. A trajectory with $\Sigma_t < 0$ is entropy-consuming. The transient fluctuation theorem shown in table 1.1 provides an analytic expression for the probability that the dissipative heat flux flows through the system in the direction opposite to that required by the second law of thermodynamics which imposes $\langle\Sigma_t\rangle \geq 0$. Note that in both the numerator an the denominator, $P(\Sigma_t)$ refers to the probability of observing a given value of Σ_t along a forward trajectory, unlike in the Crooks theorem. The fluctuation theorem requires only extremely general hypotheses to hold: (1) the initial distribution

must be known; (2) the evolution must be time-reversible; (3) apart from a set of zero measure, the phases at time t must have a nonzero probability according to the initial distribution function. The fluctuation theorem applies to situations where the free energy change is zero, and therefore does not in general hold simultaneously to the JI. The fluctuation theorem takes various forms, among which an integrated version, and a form applying to nonequilibrium steady states [12].

The Kawazaki identity follows from the fluctuation theorem, while the reverse is not true. The Kawazaki identity expresses a normalization condition on the values Σ_t can take over the ensemble. The unavoidable but infrequent negative-Σ_t trajectories contribute rarely to the average, but each contribution is exponentially significant. The exponential rarity of observing negative-Σ_t trajectories is exactly compensated by the negative exponential in the Kawasaki function.

All four of the nonequilibrium statistical mechanical relations of Table 1.1 have recently been demonstrated experimentally. In 2002, Liphardt *et al.* [14] applied the Jarzynski identity to the unfolding of a single RNA molecule whose ends were attached to optical tweezers. Collin *et al.* [58] used the same system in 2005 to verify the Crooks theorem. The fluctuation theorem was first demonstrated in 2002 by Wang *et al.* [13] with computer simulations and experiment. The experiment consisted in dragging 6.3 μm latex particles in water with an optical trap, and measuring Σ_t on a collection of trajectories. In a subsequent 2004 experiment where the strength of an optical trap holding a colloidal particle is disontinuously varied, Carberry *et al.* [56] were able to simultaneously demonstrate the Kawazaki identity and the fluctuation theorem.

1.6 Molecular dynamics potential energy function

In the previous sections, we have reviewed the rationale behind MD, the statistical mechanical notions necessary to relate MD to experiments, and how the equations of motion are integrated in a MD simulation. Up to this point, no mention was made of the potential energy function which determines the forces between the particles in the system. This is however where all the physics of the model simulated resides, and it is also by far the most time-consuming part of a simulation. All the considerations in the above sections are valid for a large spectrum of models ranging from hard spheres to high accuracy quantum methods. Given the limited (although even increasing) computational power at hand, there is a tradeoff between the accuracy of the model and the scale of the system we can consider. This scale includes both the physical size of the system (number of particles), and the extent in time of the simulation. For example, a branch of MD is concerned with Born-Oppenheimer or Carr-Parrinello [59] dynamics associated with different flavors of density functional theory. These models are limited to the study of systems of a few hundred particles, on time scales of a few tens of picoseconds.

In this thesis we focus on another branch of MD, namely classical molecular mechanics. In classical MD all electronic effects are replaced by empirically parametrized classical potential energy functions acting on the atoms. This allows to roughly multiply by a thousand both the size and time extent of the models mentioned above, reaching several 100'000 of particles for durations of tens of nanoseconds. The main drawbacks are that no

bond-breaking reaction can be simulated in classical MD, metals with complicated valence behaviors are poorly modeled, and, of course, all effects relying on redistribution of electrons, such as photochemistry or charge transfer are ruled out. There are different levels of refinement within classical MD methods. In most cases, a fixed charge is attributed to each atom. On the one hand more refined models, such as for example polarizable force fields, have an additional mobile charge center per atom. On the other hand, coarse grained models merge several atoms in bigger particles in order to reduce the computational effort and explore extended length and time scales. Some biomolecular models replace the solvent around the molecule of interest with a dielectric continuum and possibly empirical entropic terms depending on the solvent-exposed surface of the molecule (implicit solvation [60]). We focus here on explicit solvent models, where all solvent molecules are taken into account.

A typical atomistic potential MD energy function is of the form:

$$\begin{aligned}\Phi(r) &= \sum_{\text{bonds}} K_r(r-r_{eq})^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_{eq})^2 + \sum_{\text{dihedrals}} K_\phi[1+\cos(n\phi-\gamma)] \\ &+ \sum_i \sum_{\substack{j<i \\ j\notin \text{excl}(i)}} \left[\frac{C_{ij}^{(12)}}{r_{ij}^{12}} - \frac{C_{ij}^{(6)}}{r_{ij}^{6}} + \frac{1}{4\pi\epsilon_r\epsilon_0}\frac{q_i q_j}{r_{ij}}\right] \end{aligned} \tag{1.20}$$

Bond lengths r and angles θ are treated harmonically with reference points r_{eq} and θ_{eq} and harmonic constants K_r and K_θ, respectively. Dihedral angles are restrained with a periodic term bearing n minima, and scaled by the constant K_ϕ. Van der Waals interactions are modelled by a Lennard-Jones (LJ) potential energy function with atom-pair specific parameters $C^{(6)}$ and $C^{(12)}$. The last term of (1.20) represents the electrostatic interaction between two atoms of charge q_i and q_j, separated by a distance r_{ij}. The relative dielectric permittivity, ϵ_r is typically set to 1 in explicit solvent simulations. The terms contributing to the energy in (1.20) are common to the majority of currently used biomolecular force fields. A review of available force fields can be found in Ref. [61].

There are a number of technical aspects related to the evaluation of the potential energy function, which have to be briefly discussed. First, we mention some points related to neighboring atoms in a molecule:

- Inside a molecule, first, second, and third covalently bound neighboring atoms are excluded from the LJ and Coulomb pairs. Instead, their interactions are accounted for in the corresponding bond, angle and dihedral terms. The rightmost sum in (1.20) therefore runs over all non-excluded atom pairs.

- Some force fields, including the GROMOS 43A1 parameter set [62] used in Chapter 6 , use the united atom model. In this case, apolar CH_n groups are considered as one single atom with a larger LJ radius. This has proved a valid approximation, and saves a significant amount of computational effort.

- Inside a molecule, neighboring atoms are grouped in small charge groups of zero total charge. Using charge-group based cutoffs facilitates the calculation of long range electrostatic interactions. It prevents for example taking only into account only one strongly charged atom of a polar group, without the other.

- In order to increase the allowed integration time step, the bonds with the highest oscillation frequencies, or even all bonds, can be constrained to their ideal lengths. Using the SHAKE [63] procedure, these constraints are incorporated in the integration algorithm in such a way that the energy of the system is conserved.

- Note that the phenomenon of hydrogen bonding is not explicitly modeled in most force fields. The effect is mimicked by assigning stronger electrostatic charges to donor and acceptor groups.

The number of long-range or nonbonded pairwise interactions in (1.20) increases as N^2, where N is the number of particles in the system. These interactions represent a major part of the computational burden, and their treatment requires special care. To begin with, in order to avoid artifacts due to finite-size effects, the boundary is usually completely removed by treating the system with periodic boundary conditions. This places the system in an environment similar to a infinite solution. Nevertheless, it imposes a crystalline order across copies of the system, which can lead to periodicity-induced artifacts (see the end of Chapter 6). Non-bonded interactions can be divided into the rapidly decaying LJ interactions, and the long-range electrostatic interactions, which have, even at a distance of the order of the box length, usually not decayed to a negligible value. The LJ interactions are generally truncated after a convenient cutoff distance, which has to be taken shorter than half the box length in order to avoid interactions of atoms with more than one periodic image of other atoms (minimum image convention). To further reduce the computational effort, a twin-range cutoff scheme is usually applied. In this case, LJ interactions within a short-range cutoff are computed at every time step, while the remaining interaction up to a long-range cutoff are only updated every few steps. This allows to perform the search of neighbors, which is a time consuming task, also only every few time steps. For electrostatic interactions, there are two approaches to the problem, which both come with severe approximations. The first approach again truncates the Coulombic interactions beyond a cutoff distance, either with an abrupt truncation, or smoothed by a switching or shifting function, or using the reaction field correction which accounts for a mean polarization effect outside of the cutoff. The second approach makes use of the exact periodicity of the system and includes all electrostatic interactions (also with periodic copies) by lattice sum techniques based on Ewald [64] summation, among which the particle mesh Ewald (PME) method [65], a derivative of Hockney and Eastwood's [66] particle-particle particle-mesh method, is widely used.

The approximations mentioned above have two kinds of impacts on the reliability of the simulation. First, the model effectively implemented might not reflect the physical behavior of the system. For example, choosing a cutoff on the Coulomb interaction neglects long range interactions, which can lead to an underestimation of the solvation free energy of a charged molecule. Second, the implementation of force calculation can alter intrinsic properties of the dynamics. For example, the use of cutoffs for the LJ interactions leads to a positive drift of the system energy. Conversely,omitting the recalculation of certain forces at intermediate time steps in a twin-range scheme results in a negative drift of the energy. These effects can be seen as an additional source of work on the system, as shown on Figure 1.1. The consequence is that even if no external work is applied to the system in a intent to perform a nonequilibrium simulation, there will be heat flowing

through the system. According to the above discussion on phase space compression, this will have uncontrolled consequences on the entropy of the system. In fact, force field approximations and the related impact on the dynamics can often outweigh all the inaccuracies the integrator itself might cause by e.g. not being measure preserving.

1.7 References

[1] B. J. Alder and T. E. Wainwright. "Phase transition for a hard sphere system". *J. Chem. Phys.*, **27**, (1957) 1208.

[2] J. A. McCammon, B. R. Gelin, and M. Karplus. "Dynamics of folded proteins". *Nature*, **267**, (1977) 585.

[3] D. Frenkel and B. Smit. *Understanding molecular simulation : from algorithms to applications* (Academic Press, San Diego, 2002).

[4] C. Chipot and D. A. Pearlman. "Free energy calculations. the long and winding gilded road". *Mol. Sim.*, **28**, (2002) 1.

[5] W. F. van Gunsteren, T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith. "Computation of free energy in practice: choice of approximations and accuracy limiting factors". In: "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications", eds. W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, vol. 2 (Escom Science Publishers, Leiden, The Netherlands, 1993) 315.

[6] W. F. van Gunsteren, X. Daura, and A. E. Mark. "Computation of free energy". *Helv. Chim. Acta*, **85**, (2002) 3113.

[7] W. G. Hoover. *Time Reversibility, Computer Simulation, and Chaos* (World Scientific Publishing, Singapore, 1999).

[8] W. G. Hoover, H. A. Posch, and C. G. Hoover. "Fluctuations and asymmetry via local lyapunov instability in the time-reversible doubly thermostated harmonic oscillator". *J. Chem. Phys.*, **115**, (2001) 5744.

[9] W. G. Hoover, C. G. Hoover, and H. A. Posch. "Dynamical instabilities, manifolds, and local Lyapunov spectra far from equilibrium". *Comput. Meth. Sci. Tech.*, **7**, (2001) 55.

[10] W. G. Hoover, H. A. Posch, K. Aoki, and D. Kusnezov. "Remarks on non-Hamiltonian statistical mechanics: Lyapunov exponents and phase-space dimensionality loss". *Europhys. Lett.*, **60**, (2002) 337.

[11] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. "Probability of second law violations in shearing steady states". *Phys. Rev. Lett.*, **71**, (1993) 2401.

[12] D. J. Evans and D. J. Searles. "The fluctuation theorem". *Adv. Phys.*, **51**, (2002) 1529.

[13] G. M. Wang, E. M. Sevick, E. Mittag, D. J. Searles, and D. J. Evans. "Experimental demonstration of violations of the second law of thermodynamics for small systems and short time scales". *Phys. Rev. Lett.*, **89**, (2002) 050 601.

[14] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamente. "Equilibrium information from nonequilibrium measurements in an experimental test of jarzynski's identity". *Science*, **296**, (2002) 1832.

[15] D. J. Evans and G. P. Morriss. *Statistical Mechanics of NonEquilibrium Liquids*. Theoretical Chemistry Monograph Series (Academic Press, London, London, 1990).

[16] D. Chandler. *Introduction to Modern Statistical Mechanics* (Oxford University Press, 1987).

[17] C. Jarzynski. "A nonequilibrium equality for free energy differences". *Phys. Rev. Lett.*, **78**, (1997) 2690.

[18] P. Procacci and M. Marchi. "Multiple time steps algorithms for the atomistic simulations of complex molecular systems". In: "Proceedings of the NATO-ASI School, Erice 11-21 June 1998 : Advances in the Computer Simulations of Liquid Crystals", eds. C. Zannoni and P. Pasini, Proceedings of the NATO-ASI School (Kluwer Academic publishers, Dodrecht, the Netherlands, 1999) 333–387.

[19] B. Leimkuhler. *Simulating Hamiltonian Dynamics* (Cambridge University Press, 2005).

[20] S. K. Gray, D. W. Noid, and B. G. Sumpter. "Symplectic integrators for large scale molecular dynamics simulations: A comparison of several explicit methods". *J. Chem. Phys.*, **101**, (1994) 4062.

[21] M. Tuckerman, B. J. Berne, and G. J. Martyna. "Reversible multiple time scale molecular dynamics". *J. Chem. Phys.*, **97**, (1992) 1990.

[22] L. Verlet. "Computer experiments on classical fluids : I. thermodynamical properties of Lennard-Jones molecules". *Phys. Rev.*, **159**, (1967) 98.

[23] R. W. Hockney. "The potential calculation and some applications". *Methods Comput. Phys.*, **9**, (1970) 136.

[24] G. M. Torrie and J. P. Valleau. "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling". *J. Comp. Phys.*, **23**, (1977) 187.

[25] S. Nosé. "A unified formulation of the constant temperature molecular dynamics methods". *J. Chem. Phys.*, **81**, (1984) 511.

[26] P. H. Hünenberger. "Thermostat algorithms for molecular dynamics simulations". *Adv. Polym. Sci.*, **173**, (2005) 105.

[27] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". *Phys. Rev. A*, **31**, (1985) 1695.

[28] A. Baranyai. "Phase space compression and entropy of nonequilibrium steady states". *J. Chem. Phys.*, **105**, (1996) 7723.

[29] E. Cohen and L. Rondoni. "Note on phase space contraction and entropy production in thermostated Hamiltonian systems". *arXiv:cond-mat*, **9712213**.

[30] M. Dolowschiak and Z. Kovacs. "Fluctuation formula in the Nosé–Hoover thermostated lorentz gas". *arXiv:nlin.CD*, **0401036**.

[31] B. L. Holian, G. Ciccotti, W. G. Hoover, B. Moran, and H. A. Posch. "Nonlinear-response theory of time-independent fields: Consequences of the fractal nonequilibrium distribution function". *Phys. Rev. A*, **39**, (1989) 5414.

[32] B. L. Holian, H. A. Posch, and W. G. Hoover. "Nonequilibrium free energy, coarse-graining, and the liouville equation". *Phys. Rev. A*, **42**, (1990) 3196.

[33] D. Ruelle. "Positivity of entropy production in nonequilibrium statistical mechanics." *J. Stat. Phys.*, **85**, (1996) 1.

[34] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular dynamics with coupling to an external bath". *J. Chem. Phys.*, **81**, (1984) 3684.

[35] T. Morishita. "Fluctuation formulas in molecular-dynamics simulations with the weak coupling heat bath". *J. Chem. Phys.*, **113**, (2000) 2976.

[36] C. P. Dettmann and G. P. Morriss. "Hamiltonian reformulation and pairing of Lyapunov exponents for Nosé-Hoover dynamics". *Phys. Rev. E*, **55**, (1997) 3693.

[37] S. D. Bond, B. J. Leimkuhler, and B. B. Laird. "The Nosé–poincaré method for constant temperature moleculardynamics". *J. Comp. Phys.*, **151**, (1999) 114.

[38] J. Struckmeier. "Hamiltonian dynamics on the symplectic extended phase space for autonomous and non-autonomous systems". *J. Phys. A*, **38**, (2005) 1257.

[39] M. E. Tuckerman, C. J. Mundy, and G. J. Martyna. "On the classical statistical mechanics of non-Hamiltonian systems". *Europhys. Lett.*, **45**, (1999) 149–155.

[40] J. D. Ramshaw. "Remarks on non-Hamiltonian statistical mechanics". *Europhys. Lett.*, **59**, (2002) 319.

[41] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna. "Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems". *J. Chem. Phys.*, **115**, (2001) 1678.

[42] V. E. Tarasov. "Phase-space metric for non-Hamiltonian systems". *J. Phys. A*, **38**, (2005) 2145.

[43] G. S. Ezra. "On the statistical mechanics of non-Hamiltonian systems: the generalized Liouville equation, entropy, and time-dependent metrics". *J. Math. Chem.*, **35**, (2004) 29.

[44] L. Casetti, M. Pettini, and E. G. D. Cohen. "Geometric approach to Hamiltonian dynamics and statistical mechanics". *Phys. Reports*, **337**, (2000) 237.

[45] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna. "A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble". *J. Phys. A*, **39**, (2006) 5629.

[46] G. S. Ezra. "Reversible measure-preserving integrators for non-Hamiltonian systems". *J. Chem. Phys.*, **125**, (2006) 034 104.

[47] B. L. Holian, W. G. Hoover, and H. A. Posch. "Resolution of Loschmidt's paradox: The origin of irreversible behavior in reversible atomistic dynamics". *Phys. Rev. Lett.*, **59**, (1987) 10.

[48] D. Daems and G. Nicolis. "Entropy production and phase space volume contraction". *Phys. Rev. E*, **59**, (1999) 4000.

[49] D. J. Evans and L. Rondoni. "Comments on the entropy of nonequilibrium steady states". *J. Stat. Phys.*, **109**, (2002) 895.

[50] M. E. Tuckerman, C. J. Mundy, and M. L. Klein. "Toward a statistical thermodynamics of steady states". *Phys. Rev. Lett.*, **78**, (1997) 2042.

[51] W. G. Hoover, D. J. Evans, H. A. Posch, B. L. Holian, and G. P. Morriss. "Comment on "toward a statistical thermodynamics of steady states"". *Phys. Rev. Lett.*, **80**, (1998) 4103.

[52] D. J. Evans and D. J. Searles. "Steady states, invariant measures, and response theory". *Phys. Rev. E*, **52**, (1995) 5839.

[53] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". *Phys. Rev. E*, **60**, (1999) 2721.

[54] R. W. Zwanzig. "High-temperature equation of state by a perturbation method. I. Nonpolar gases". *J. Chem. Phys.*, **22**, (1954) 1420.

[55] C. H. Bennett. "Efficient estimation of free energy differences from Monte Carlo data". *J. Comp. Phys.*, **22**, (1976) 245.

[56] D. M. Carberry, S. R. Williams, G. M. Wang, E. M. Sevick, and D. J. Evans. "The Kawasaki identity and the fluctuation theorem". *J. Chem. Phys.*, **121**, (2004) 8179.

[57] J. C. Reid, E. M. Sevick, and D. J. Evans. "A unified description of two theorems in non-equilibrium statistical mechanics: The fluctuation theorem and the work relation". *Europhys. Lett.*, **72**, (2005) 726.

[58] D. Collin, F. Ritort, C. Jarzynski, S. B. Smith, I. Tinoco Jr., and C. Bustamante. "Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies". *Nature*, **437**, (2005) 231.

[59] R. Car and M. Parrinello. "Unified approach for molecular dynamics and density-functional theory". *Phys. Rev. Lett.*, **55**, (1985) 2471–2474.

[60] B. Roux and T. Simonson. "Implicit solvent models". *Biophys. Chem.*, **78**, (1999) 1.

[61] A. D. MacKerell. "Empirical force fields for biological macromolecules: Overview and issues". *J. Comp. Chem.*, **25**, (2004) 1584.

[62] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).

[63] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. "Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". *J. Comp. Phys.*, **23**, (1977) 327.

[64] P. P. Ewald. "Die Berechnung optischer und electrostatistischer Gitterpotentiale". *Ann. Phys.*, **64**, (1921) 253.

[65] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. "A smooth particle mesh ewald method". *J. Chem. Phys.*, **103**, (1995) 8577–8593.

[66] R. W. Hockney and J. W. Eastwood. *Computer simulation using particles* (Int. Physics Publishing, Taylor & Francis, Inc., Bristol, PA, USA, 1988).

Chapter 2

On the calculation of velocity-dependent properties in molecular dynamics simulations using the leap-frog integration algorithm

Summary

Widely used programs for molecular dynamics simulation of (bio)molecular systems are the Verlet and leap-frog algorithms. In these algorithms, the particle velocities are less accurately propagated than the positions. Important quantities for the simulation such as the temperature and the pressure involve the squared velocities at full time steps. Here, we derive an expression for the squared particle velocity at full time step in the leap-frog scheme, which is more accurate than the standardly used one. In particular, this allows to show that the full time step kinetic energy of a particle is more accurately computed as the average of the kinetic energies at previous and following half steps than as the square of the average velocity as implemented in various molecular dynamics codes. Use of the square of the average velocity introduces a systematic bias in the calculation of the instantaneous temperature and pressure of a molecular dynamics system. We show the consequences when the system is coupled to a thermostat and a barostat.

2.1 Introduction

Classical dynamics for a set of N particles or atoms in 3 dimensions is most generally formulated [1] as Hamilton's equations of motion for the generalized positions $q = (q_1, ..., q_{3N})$ and momenta $p = (p_1, ..., p_{3N})$,

$$\dot{q}_i = \frac{\partial H}{\partial p_i}(p, q), \tag{2.1}$$

$$\dot{p}_i = -\frac{\partial H}{\partial q_i}(p, q). \tag{2.2}$$

The Hamiltonian function describing conservative systems of particles with mass m_i and Cartesian coordinates $r = (r_1, ..., r_{3N})$ is

$$H(p, r) = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + V(r), \tag{2.3}$$

i.e. the kinetic energy term only depends on the momenta or velocities $v_i = \frac{p_i}{m_i}$, and the potential energy term depends only on the positions r. Inserting (2.3) into (2.1) and (2.2), Newton's equations of motion are obtained,

$$\dot{r}_i(t) = v_i(t) \tag{2.4}$$

$$\dot{v}_i(t) = \frac{f_i(t)}{m_i} \tag{2.5}$$

$$f_i(t) = -\frac{\partial}{\partial r_i} V(r(t)). \tag{2.6}$$

Due to the separation of the p- and q-dependence in (2.3), equations (2.4) and (2.5) are equivalent to a second-order differential equation

$$\ddot{r}_i(t) = \frac{1}{m_i} f_i(r(t)), \tag{2.7}$$

in which no terms in $\dot{r}(t)$ are present. This implies that the computation of velocities is not required to solve Eq. (2.7). This equation also shows that Newton's equations of motion are invariant under time reversal. Both properties should be taken into account when selecting an algorithm to integrate Eqs. (2.4) and (2.5).

Numerical methods to solve sets of ordinary differential equations can be found in the literature [2, 3, 4]. They are based on finite differences and solve the equations by stepping forward in time using time steps Δt. Since the computation of the forces, Eq. (2.6), is much more expensive than manipulation of positions and velocities, only integration schemes that involve only one force evaluation per time step are of use in molecular dynamics (MD) simulation. This rules out the generally very efficient Runge-Kutta algorithms. A second consideration when selecting an integration algorithm derives from the nature of the energy hypersurface $V(r)$ for molecular systems in the condensed phase. The function $V(r)$ is very rugged, i.e. it has a mountainous character with many valleys. Since a system often visits the bottoms of these valleys characterized by positive second derivatives of the

potential energy $V(r)$ or negative first derivatives of the forces with respect to r, an efficient MD algorithm should at least include a correct treatment of the first derivative of the forces in order to avoid systematic errors [5]. This means that the order of the algorithm, defined as the highest order of the time step Δt included in the equations for the propagation of the positions, should at least be three. In complex condensed phase molecular systems, it is not efficient to use an integration algorithm beyond third or fourth order [6], because the higher derivatives of the forces display a non-systematic behavior.

These considerations (time-reversibility, one force calculation per step, and up to third-order accuracy in Δt) have led to the wide-spread use of the Verlet [7, 8, 9, 10] and leap-frog [11] algorithms in MD simulation. We note that these algorithms satisfy the condition of symplecticity [12, 13, 14, 15], i.e. they preserve the Poincaré invariants characteristic of continuous Hamiltonian systems. In addition, the resulting discrete trajectories represent the exact evolution of a "shadow" Hamiltonian system with a Hamiltonian function [16, 17, 18] arbitrarily close to the physical Hamiltonian, which hinders systematic energy drifts. The symplectic property of the Verlet and leap-frog integrators is most directly evidenced when deriving [19, 15] them using a Trotter expansion of the Liouville evolution operator. This formalism also allows to show [20, 21] that both algorithms generate trajectories with equivalent positions but slightly different velocities. Indeed, as shown in the following, Verlet and leap-frog integrators both propagate the positions with an accuracy of third order in Δt, whereas the velocities are propagated to first and to second order respectively.

Since most molecular properties of interest depend only on the particle positions, the relatively low accuracy of the velocity propagation is often not problematic. Yet, a higher-accuracy expression for the velocities would be welcome, e.g. when calculating kinetic energies, where inaccurate expressions may introduce a systematic bias. This is of concern when using the kinetic energies to couple the system to a heat bath in constant temperature simulations, or when the internal pressure of the system as calculated from the virial is regulated. Here we derive an expression for the particle velocities at full time steps, which is more accurate than the standard leap-frog expression.

In section 2, the relevant formulae of the Verlet and leap-frog integration algorithms are derived using the Taylor expansion approach. The more accurate expression for the squared velocities in the leap-frog scheme is derived in section 3 and illustrated with simple numerical examples in section 4. Section 5 contains a short discussion of the applicability of the proposed expressions.

2.2 The Verlet and leap-frog algorithms for MD simulation

In order to simplify the notation, we consider in the following only Cartesian components x, v and f of the position, velocity and force of a single particle and use the subindex n to indicate the time point t_n, i.e. $t_{n+1} \equiv t_n + \Delta t$. The Verlet integration scheme is most easily derived [5] by expanding the positions at time $t_n + \Delta t$ in a Taylor series in Δt at time t_n,

$$x_{n+1} = x_n + \dot{x}_n \Delta t + \frac{1}{2!}\ddot{x}_n (\Delta t)^2 + \frac{1}{3!}\dddot{x}_n (\Delta t)^3 + \mathcal{O}\left((\Delta t)^4\right).$$

Adding the corresponding expansion for $t_n - \Delta t$, rearranging terms and using Eq. (2.7), one obtains the Verlet algorithm,

$$x_{n+1} = 2x_n - x_{n-1} + \frac{1}{m} f_n (\Delta t)^2 + \mathcal{O}\left((\Delta t)^4\right), \tag{2.8}$$

which indeed fulfills the three criteria for an efficient MD algorithm mentioned above. An expression for the velocity can be obtained by subtracting the Taylor expansion for $t_n - \Delta t$ from the one for $t_n + \Delta t$, rearranging and using Eq. (2.7),

$$v_n = \frac{1}{2\Delta t}\left(x_{n+1} - x_{n-1}\right) - \frac{1}{6m}\dot{f}_n (\Delta t)^2 + \mathcal{O}\left((\Delta t)^4\right).$$

In the standard Verlet algorithm, the term containing $\dot{f}_n$ is omitted, leading to a first-order accuracy of the Verlet velocity propagation,

$$v_n = \frac{1}{2\Delta t}\left(x_{n+1} - x_{n-1}\right). \tag{2.9}$$

The accuracy can be enhanced by using a time-reversal invariant expression for $\dot{f}_n$,

$$\dot{f}_n = \frac{1}{2\Delta t}\left(f_{n+1} - f_{n-1}\right),$$

leading to

$$v_n = \frac{1}{2\Delta t}\left(x_{n+1} - x_{n-1}\right) - \frac{1}{12m}\left(f_{n+1} - f_{n-1}\right)\Delta t + \mathcal{O}\left((\Delta t)^4\right). \tag{2.10}$$

Since the Verlet algorithm (2.8) to propagate the positions does not depend explicitly on velocities, a coupling of the molecular system to a heat bath or thermostat, which will influence the velocities, is not straightforward. Such a coupling is more easily implemented using the leap-frog integration algorithm.

The leap-frog integration scheme is most easily derived by expanding the positions or velocities at time $t_n + \frac{\Delta t}{2}$ in a Taylor series in $\frac{\Delta t}{2}$ at time point t. Subtracting the expansions of the positions for $\frac{\Delta t}{2}$ and $-\frac{\Delta t}{2}$ from each other, rearranging terms, using Eq. (2.7) and then shifting all time points by $\frac{\Delta t}{2}$ one obtains

$$x_{n+1} = x_n + v_{n+\frac{1}{2}}\Delta t + \frac{1}{24m}\dot{f}_{n+\frac{1}{2}}(\Delta t)^3 + \mathcal{O}\left((\Delta t)^5\right). \tag{2.11}$$

Using a time reversal invariant expression for $f_{n+\frac{1}{2}}$,

$$\dot{f}_{n+\frac{1}{2}} = \frac{1}{\Delta t}\left(f_{n+1} - f_n\right),$$

one obtains

$$x_{n+1} = x_n + v_{n+\frac{1}{2}}\Delta t + \frac{1}{24m}\left(f_{n+1} - f_n\right)(\Delta t)^2 + \mathcal{O}\left((\Delta t)^5\right). \tag{2.12}$$

The expression for the velocity propagation is obtained by subtracting the expansions of the velocities for $\frac{\Delta t}{2}$ and $-\frac{\Delta t}{2}$ from each other, rearranging terms and using Eq. (2.7),

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + \frac{1}{m} f_n \Delta t + \frac{1}{24m} \dddot{f}_n (\Delta t)^3 + \mathcal{O}\left((\Delta t)^5\right). \tag{2.13}$$

Using a time-reversal invariant expression for $\dddot{f}_n$,

$$\dddot{f}_n = \frac{1}{(\Delta t)^2} \left(f_{n+1} - 2f_n + f_{n-1} \right),$$

one obtains

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + \frac{1}{m} f_n \Delta t + \frac{1}{24m} \left(f_{n+1} - 2f_n + f_{n-1} \right) \Delta t + \mathcal{O}\left((\Delta t)^5\right).$$

The standard leap-frog algorithm,

$$v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}} + \frac{1}{m} f_n \Delta t + \mathcal{O}\left((\Delta t)^3\right) \tag{2.14}$$

$$x_{n+1} = x_n + v_{n+\frac{1}{2}} \Delta t + \mathcal{O}\left((\Delta t)^3\right) \tag{2.15}$$

is obtained by omitting the terms containing $\dot{f}$ and $\ddot{f}$ in Eqs. (2.11) and (2.13) respectively.

The leap-frog algorithm, Eqs. (2.15) and (2.14), is exactly equivalent [5] to the Verlet algorithm, Eqs. (2.8) and (2.9). This can be shown by writing the velocity $v_{n-\frac{1}{2}}$ in terms of positions : shift the time in Eq. (2.15) by $-\Delta t$ and solve for $v_{n-\frac{1}{2}}$; then substitute the resulting expression in Eq. (2.14) and the latter in Eq. (2.15). This leads to Eq. (2.8), the Verlet position formula. The Verlet velocity formula (2.9) is obtained by using Eq. (2.15) to express $v_{n\frac{1}{2}}$ and $v_{n+\frac{1}{2}}$ in terms of positions and then inserting the result into

$$v_n = \frac{1}{2} \left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}} \right),$$

which is the formula for the leap-frog velocity at full time steps.

A higher-order expression for v_n in the leap-frog scheme can be obtained by subtracting the Taylor expansion for the position at $t_n - \Delta t$ from the one at $t_n + \Delta t$, and substituting the result in Eq. (2.10),

$$v_n = \frac{1}{2} \left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}} \right) - \frac{1}{16m} \left(f_{n+1} - f_{n-1} \right) \Delta t + \mathcal{O}\left((\Delta t)^4\right). \tag{2.16}$$

Comparing this expression for v_n to the one in Eq. (3.45) in [5], it is observed that the latter contains (wrongly) a factor $\frac{1}{12}$ instead of $\frac{1}{16}$. The factor $\frac{1}{12}$ is obtained when using Eq. (2.9) instead of the more accurate Eq. (2.10) in the derivation of Eq. (2.16). Due to the occurrence of f_{n+1} in Eq. (2.16), which is unknown at t_n, Eq. (2.16) is not of much practical use.

An alternative higher-order expression for v_n in the leap-frog scheme can be obtained by subtracting from x_{n+1} as given in Eq. (2.12) the analogous expression for x_{n-1},

$$x_{n+1} - x_{n-1} = \left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}} \right) \Delta t + \frac{1}{24m} \left(f_{n+1} - f_{n-1} \right) (\Delta t)^2 + \mathcal{O}\left((\Delta t)^5\right).$$

Solving for $(f_{n+1} - f_{n-1})$ and substituting the result into Eq. (2.16) yields

$$v_n = \left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}}\right) - \frac{3}{2\Delta t}\left(x_{n+1} - x_{n-1}\right) + \mathcal{O}\left((\Delta t)^4\right),$$

which can be used to obtain a posteriori higher-accuracy velocities from a standard leap-frog trajectory. Due to the presence of x_{n+1}, its use in generating a trajectory is not practical.

2.3 A simple higher-order formula for squared velocities in the leap-frog scheme

A first expression for the squared velocity at time t_n can be obtained by taking the square of Eq. (2.16) and omitting terms of order $(\Delta t)^2$ and higher,

$$\begin{aligned} v_n^2 &= \frac{1}{4}\left(v_{n-\frac{1}{2}}^2 + 2v_{n-\frac{1}{2}}v_{n+\frac{1}{2}} + v_{n+\frac{1}{2}}^2\right) \\ &\quad -\frac{1}{16m}\left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}}\right)\left(f_{n+1} - f_{n-1}\right)\Delta t + \mathcal{O}\left((\Delta t)^2\right). \end{aligned} \tag{2.17}$$

In order to obtain an expression for $v_{n-\frac{1}{2}}v_{n+\frac{1}{2}}$ in terms of squared velocities and forces, we multiply Eq. (2.13) by $v_{n-\frac{1}{2}}$, multiply the time-inverted equivalent of Eq. (2.13) by $v_{n+\frac{1}{2}}$, and add the resulting equations, thereby omitting terms of order $(\Delta t)^3$ and higher,

$$2v_{n-\frac{1}{2}}v_{n+\frac{1}{2}} = v_{n-\frac{1}{2}}^2 + v_{n+\frac{1}{2}}^2 - \frac{1}{m}\left(v_{n+\frac{1}{2}} - v_{n-\frac{1}{2}}\right)f_n\Delta t + \mathcal{O}\left((\Delta t)^3\right). \tag{2.18}$$

Inserting this result into Eq. (2.17) gives an alternative higher-order expression for v_n^2,

$$v_n^2 = v_n^2(\mathrm{I}) + v_n^2(\mathrm{II}) + v_n^2(\mathrm{III}) + \mathcal{O}\left((\Delta t)^2\right), \tag{2.19}$$

with the three terms I, II and III defined respectively as

$$\begin{aligned} v_n^2(\mathrm{I}) &= \frac{1}{2}\left(v_{n-\frac{1}{2}}^2 + v_{n+\frac{1}{2}}^2\right), \\ v_n^2(\mathrm{II}) &= -\frac{1}{4m}\left(v_{n+\frac{1}{2}} - v_{n-\frac{1}{2}}\right)f_n\Delta t, \\ v_n^2(\mathrm{III}) &= -\frac{1}{16m}\left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}}\right)\left(f_{n+1} - f_{n-1}\right)\Delta t. \end{aligned}$$

Term I is always positive. In order to estimate the signs of terms II and III, we replace finite difference expressions for v and f by their equivalent averages or derivatives. Term II represents the squared forces,

$$v_n^2(\mathrm{II}) \approx -\frac{1}{4m^2}f_n^2(\Delta t)^2 \leq 0, \tag{2.20}$$

and term III expresses the correlation between the velocity and the force variation,

$$v_n^2(\mathrm{III}) \approx -\frac{1}{4m}v_n\dot{f}_n(\Delta t)^2.$$

Using the relations $\dot{f} = \frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t}$ and $f = -\frac{\partial V}{\partial x}$, we find for the last term

$$v_n^2(\mathrm{III}) \approx +\frac{1}{4m} v_n^2 \frac{\partial^2 V}{\partial x_n^2} (\Delta t)^2,$$

which will be positive for a convex ($\frac{\partial^2 V}{\partial x^2} > 0$) potential energy surface. As noted in the introduction, an MD system in the condensed phase spends most time in the bottoms of the energy hypersurface valleys, which are characterized by positive second derivatives. This implies that $v_n^2(\mathrm{III})$ is generally positive.

In the standard leap-frog scheme, the kinetic energy at full time step is evaluated using the average velocity at t_n, which using Eq. (2.18) is equal to

$$\begin{aligned} v_n^2(\mathrm{LF}) &= \frac{1}{4}\left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}}\right)^2 \\ &= \frac{1}{2}\left(v_{n-\frac{1}{2}}^2 + v_{n+\frac{1}{2}}^2\right) - \frac{1}{4m}\left(v_{n+\frac{1}{2}} - v_{n-\frac{1}{2}}\right) f_n \Delta t \\ &= v_n^2(\mathrm{I}) + v_n^2(\mathrm{II}). \end{aligned} \tag{2.21}$$

The terms I and II are of order $(\Delta t)^0$ and $(\Delta t)^1$ respectively. Compared to the correct higher-order expression (2.19), term III, which is also of order $(\Delta t)^1$, is missing. Because terms II and III generally have opposite signs, a systematic negative bias is associated with $v_n^2(\mathrm{LF})$. In order to estimate v_n^2, it is consequently more accurate to only use term I,

$$v_n^2(\mathrm{AVG}) = v_n^2(\mathrm{I}) = \frac{1}{2}\left(v_{n-\frac{1}{2}}^2 + v_{n+\frac{1}{2}}^2\right), \tag{2.22}$$

than the usual $v_n^2(\mathrm{LF})$, which yields a slightly too low kinetic energy.

Computing the squared velocities with the full higher-order expression

$$v_n^2(\mathrm{HO}) = v_n^2(\mathrm{I}) + v_n^2(\mathrm{II}) + v_n^2(\mathrm{III}) \tag{2.23}$$

is impractical, because term III contains f_{n+1} which is not available at step n. An alternative is to keep both terms of order $(\Delta t)^1$, at the cost of abandoning the time reversibility in the third term of Eq. (2.19) by approximating $(f_{n+1} - f_{n-1})$ with $2\,(f_n - f_{n-1})$. This gives the higher-order non-reversible estimate

$$\begin{aligned} v_n^2(\mathrm{HOnr}) &= \frac{1}{2}\left(v_{n-\frac{1}{2}}^2 + v_{n+\frac{1}{2}}^2\right) - \frac{1}{4m}\left(v_{n+\frac{1}{2}} - v_{n-\frac{1}{2}}\right) f_n \Delta t \\ &\quad - \frac{1}{8m}\left(v_{n-\frac{1}{2}} + v_{n+\frac{1}{2}}\right)\left(f_n - f_{n-1}\right)\Delta t + \mathcal{O}\left((\Delta t)^2\right), \end{aligned} \tag{2.24}$$

in which the third term will be denoted as $v_n^2(\mathrm{III}')$.

2.4 Applications

2.4.1 One-dimensional systems : harmonic oscillator, and Lennard-Jones particle pair

The effects of the different approximations discussed in the previous section can be illustrated using a 1-dimensional harmonic oscillator with force constant $m\omega^2$. An exact trajectory is given by $x(t) = \cos(\omega t)$. The amplitude, which depends on the initial conditions, is set to unity. The exact squared velocity at time t_n is then

$$v_n^2 = \omega^2 \sin^2(\omega t_n), \tag{2.25}$$

and the three terms of Eq. (2.19) are

$$\begin{aligned} v_n^2(\mathrm{I}) &= \tfrac{1}{2}\omega^2 \left[\sin^2(\omega t_{n-\frac{1}{2}}) + \sin^2(\omega t_{n+\frac{1}{2}})\right] \\ &= \tfrac{1}{2}\omega^2 \left[1 - \cos(2\omega t_n)\cos(\omega\Delta t)\right], \qquad (2.26) \\ v_n^2(\mathrm{II}) &= -\tfrac{1}{4}\omega^3\Delta t \left[\sin(\omega t_{n+\frac{1}{2}}) - \sin(\omega t_{n-\frac{1}{2}})\right]\cos(\omega t_n) \\ &= -\tfrac{1}{4}\omega^3\Delta t \left[1 + \cos(2\omega t_n)\right]\sin\omega\tfrac{\Delta t}{2}, \qquad (2.27) \\ v_n^2(\mathrm{III}) &= -\tfrac{1}{16}\omega^3\Delta t \left[\sin(\omega t_{n-\frac{1}{2}}) + \sin(\omega t_{n+\frac{1}{2}})\right]\left[\cos(\omega t_{n+1}) - \cos(\omega t_{n-1})\right] \\ &= +\tfrac{1}{8}\omega^3\Delta t \left[1 - \cos(2\omega t_n)\right]\cos(\omega\tfrac{\Delta t}{2})\sin(\omega\Delta t). \qquad (2.28) \end{aligned}$$

Using Taylor series expansions around zero for the cosine and sine functions with arguments containing Δt, we find

$$\begin{aligned} v_n^2(\mathrm{I}) &= \omega^2\sin^2(\omega t_n) + \tfrac{1}{4}\omega^2\cos(2\omega t_n)\left[(\omega\Delta t)^2 - \tfrac{1}{12}(\omega\Delta t)^4\right] + \mathcal{O}\left((\Delta t)^6\right), \\ v_n^2(\mathrm{II}) &= -\tfrac{1}{4}\omega^2\cos^2(\omega t_n)\left[(\omega\Delta t)^2 - \tfrac{1}{24}(\omega\Delta t)^4\right] + \mathcal{O}\left((\Delta t)^6\right), \\ v_n^2(\mathrm{III}) &= \tfrac{1}{4}\omega^2\sin^2(\omega t_n)\left[(\omega\Delta t)^2 - \tfrac{7}{24}(\omega\Delta t)^4\right] + \mathcal{O}\left((\Delta t)^6\right). \end{aligned}$$

Summing up these terms yields

$$v_n^2(\mathrm{HO}) = \omega^2\sin^2(\omega t_n) - \omega^2\left[\tfrac{1}{96} + \tfrac{1}{24}\sin^2(\omega t_n)\right](\omega\Delta t)^4 + \mathcal{O}\left((\Delta t)^6\right),$$

which shows that $v_n^2(\mathrm{HO})$ matches the exact squared velocity up to order $(\Delta t)^4$.

For the approximation $v_n^2(\mathrm{III}')$ we find

$$v_n^2(\mathrm{III}') = -\tfrac{1}{4}\omega^3\Delta t\sin(\omega t_n)\cos(\omega\tfrac{\Delta t}{2})\Big[\cos(\omega t_n)\big(1 - \cos(\omega\Delta t)\big) - \sin(\omega t_n)\sin(\omega\Delta t)\Big],$$

or using Taylor series expansions around zero for the cosine and sine functions with arguments containing Δt,

$$\begin{aligned} v_n^2(\mathrm{III}') &= +\tfrac{1}{4}\omega^2\sin^2(\omega t_n)\Big[(\omega\Delta t)^2 - \tfrac{1}{2}\cot(\omega t_n)(\omega\Delta t)^3 \\ &\quad -\tfrac{7}{24}(\omega\Delta t)^4 + \tfrac{5}{48}\cot(\omega t_n)(\omega\Delta t)^5\Big] + \mathcal{O}\left((\Delta t)^6\right). \qquad (2.29) \end{aligned}$$

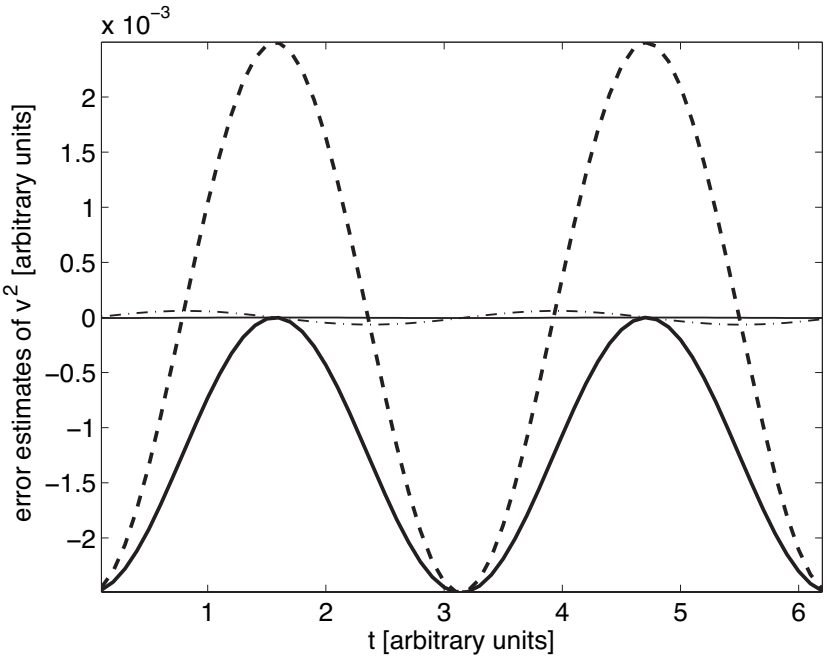

Figure 2.1: *Error of v^2 estimators for an harmonic oscillator with force constant $m\omega^2$, $\omega = 1$, $m = 1$, and $\Delta t = 0.1$. Thick line: $\Delta v_n^2(\mathrm{LF})$, Eq. (2.21); dashed line: $\Delta v_n^2(\mathrm{AVG})$, Eq. (2.22); dash-dotted line: $\Delta v_n^2(\mathrm{HOnr})$, Eq. (2.24); thin line: $\Delta v_n^2(\mathrm{HO})$, Eq. (2.23).*

Summing up the terms $v_n^2(\mathrm{I})$, $v_n^2(\mathrm{II})$, and $v_n^2(\mathrm{III}')$ yields

$$\begin{aligned} v_n^2(\mathrm{HOnr}) &= \omega^2 \sin^2(\omega t_n) - \tfrac{1}{16}\omega^2 \sin(2\omega t_n)(\omega\Delta t)^3 \\ &\quad -\tfrac{1}{32}\omega^2\Big[1 - \tfrac{2}{3}\cos(2\omega t_n)\Big](\omega\Delta t)^4 \\ &\quad +\tfrac{5}{384}\omega^2 \sin(2\omega t_n)(\omega\Delta t)^5 + \mathcal{O}\left((\Delta t)^6\right), \end{aligned}$$

which shows that $v_n^2(\mathrm{HOnr})$ matches the exact squared velocity up to order $(\Delta t)^3$.

We define the absolute error of estimator XX as $\Delta v_n^2(\mathrm{XX}) = v_n^2(\mathrm{XX}) - v_n^2$, where v_n^2 stands for the exact squared velocity at step n. The time averaged absolute error or bias of an estimator is then denoted $\langle \Delta v_n^2(\mathrm{XX}) \rangle$. Using Eqs. (2.26) to (2.28) and (2.29), we can compute the bias of the various v_n^2 estimators defined in Eqs. (2.22), (2.21), (2.23), and (2.24) respectively,

$$\begin{aligned} \left\langle \Delta v_n^2(\mathrm{AVG}) \right\rangle &= 0\,, \\ \left\langle \Delta v_n^2(\mathrm{LF}) \right\rangle &= -\tfrac{1}{4}\omega^3 \Delta t \sin\left(\omega\tfrac{\Delta t}{2}\right) \qquad (2.30) \\ &= -\tfrac{1}{8}\omega^4(\Delta t)^2 + \tfrac{1}{192}\omega^6(\Delta t)^4 + \mathcal{O}\left((\Delta t)^6\right), \\ \left\langle \Delta v_n^2(\mathrm{HO}) \right\rangle &= -\tfrac{1}{4}\omega^3 \Delta t \sin^3\left(\omega\tfrac{\Delta t}{2}\right) \qquad (2.31) \\ &= -\tfrac{1}{32}\omega^6(\Delta t)^4 + \mathcal{O}\left((\Delta t)^6\right) \\ &= \left\langle \Delta v_n^2(\mathrm{HOnr}) \right\rangle. \end{aligned}$$

Bond	k_{bond} 10^5 $\left[\frac{\text{kJ}}{\text{mol nm}^2}\right]$	$m_\star$ [amu]	Period [fs]	Relative bias [%] $\Delta t = 1$ fs	$\Delta t = 0.5$ fs	$\Delta t = 0.1$ fs
C - C	3.347	6.005	26.60	1.39	0.35	0.01
C - O	5.021	6.861	23.21	1.83	0.46	0.02
H - N	3.745	0.940	9.93	9.83	2.49	0.10

Table 2.1: *Relative kinetic energy bias due to the use of v_n^2(LF), Eq. (2.21), for typical bonds of the GROMOS 43A1 force field [22], with energy k_BT, $T = 300$ K. Here, k_{bond} represents the harmonic constant and $m_\star$ the reduced mass of the two atoms.*

This shows that the leap-frog estimator has a systematic bias of order $(\Delta t)^2$, whereas v_n^2(AVG) is unbiased. The higher-order terms have a bias of order $(\Delta t)^4$. Figure 2.1 shows the absolute errors of the various v_n^2 estimators defined in Eq. (2.21) to (2.24). The time step chosen is $\Delta t = 0.1$, $m = 1$, and $\omega = 1$. These parameters are in arbitrary units, and correspond to approximately 20π integration points per period. We observe that v_n^2(LF) oscillates around a negative value consistent with the prediction of Eq. (2.30), namely $-1.25 \cdot 10^{-3}$.

It is informative to calculate the effect of estimating the kinetic energy with v_n^2(LF) for some typical harmonic oscillators in an MD simulation. Table 2.1 shows the results expected for the C-C, C-O, H-N bonds as represented in the GROMOS 43A1 force field [22], together with typical time steps. The relative kinetic energy bias is calculated as $\langle \Delta v_n^2(\text{LF}) \rangle \,/\, \langle v^2 \rangle$, with the numerator evaluated using Eq. (2.30), and $\langle v^2 \rangle$ the exact average squared velocity corresponding to an energy of $\frac{1}{2}k_BT$ at $T = 300K$. In case of a realistic motion, an amplitude prefactor has to be added to all equations from (2.25) to (2.31), which are based on a trivial harmonic oscillator with position $x(t) = \cos \omega t$. For a total oscillator energy of k_BT, the amplitude is $x_0 = \sqrt{\frac{2k_BT}{m\omega^2}}$.

As a further example, we look at the interaction of two Lennard-Jones (LJ) particles. The parameters are chosen such that this example represents a typical interaction found in an MD simulation. The C6 and C12 parameters are those of the oxygen-oxygen interaction in a pair of water molecules, $2.617 \cdot 10^{-3}$ kJ mol^{-1}nm^6 and $2.634 \cdot 10^{-6}$ kJ mol^{-1}nm^{12}, respectively. Particles have a mass of 16 amu. The energy is set to k_BT at $T = 300K$, and the time step is $\Delta t = 1$ fs, as often used in MD simulations. The initial positions and velocities are such that a centered collision is induced. No exact solution is available for the motion of Lennard-Jones particles. We use instead the leap-frog algorithm to integrate the equations of motion and generate a trajectory. Expression v_n^2(HO), Eq. (2.23), is evaluated a posteriori, and serves as a reference for the other less accurate kinetic energy estimators. Figure 2.2 shows the behavior of the resulting errors for the kinetic energies calculated with the different v_n^2 estimators. The error of an estimator XX is calculated as $\Delta v_n^2(\text{XX}) = v_n^2(\text{XX}) - v_n^2(\text{HO})$. It is again apparent that v_n^2(HOnr) is the most accurate estimator. In addition, v_n^2(LF) displays a negative error during the whole collision, indicating that this estimator is also biased on average for Lennard-Jones interactions.

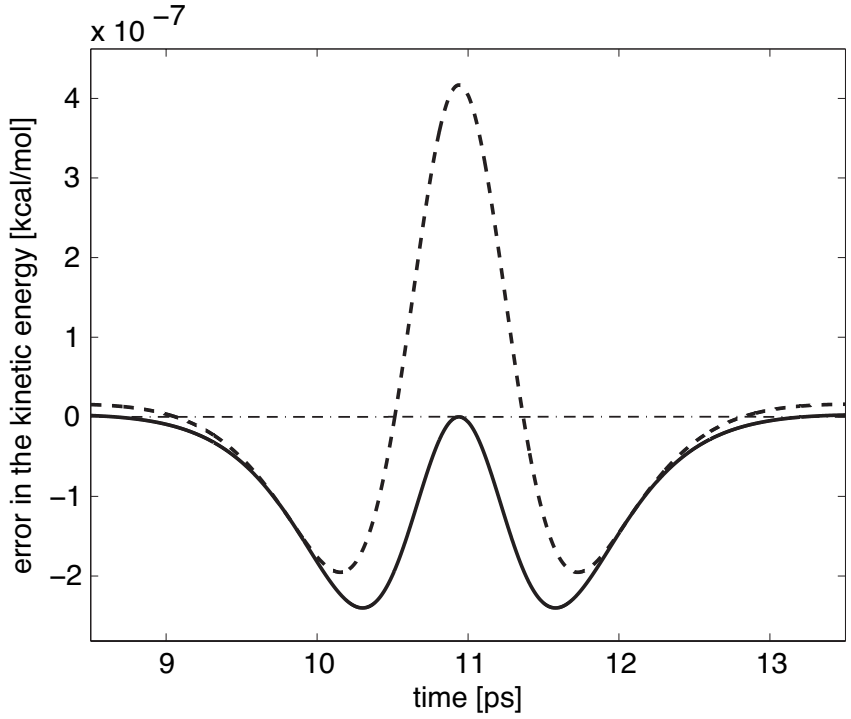

Figure 2.2: *Error of kinetic energy estimators for a collision of two Lennard-Jones particles, water oxygens with energy k_BT for $T = 300$ K, with $\Delta t = 1$ fs. Thick line: $\frac{1}{2}m\Delta v_n^2(\mathrm{LF})$, Eq. (2.21); dashed line: $\frac{1}{2}m\Delta v_n^2(\mathrm{AVG})$, Eq. (2.22); dash-dotted line: $\frac{1}{2}m\Delta v_n^2(\mathrm{HOnr})$, Eq. (2.24). Here, $\frac{1}{2}mv_n^2(\mathrm{HO})$, Eq. (2.23), is the reference.*

2.4.2 Realistic MD simulations

Based on the results of Table 2.1, we can expect that the true kinetic energy of a molecular system modeled only by harmonic bonded terms could be approximately 3% above the value estimated with $v_n^2(\mathrm{LF})$ of Eq. (2.21). A realistic explicit solvent MD system however includes many other types of interactions, including softer angle and dihedral-angle potential energy terms, electrostatic as well as Lennard-Jones interactions. Next, we assess the behavior of the different kinetic energy estimators for realistic MD systems. The instantaneous temperature of a dynamical system of N particles with N_{df} degrees of freedom is related to the average kinetic energy through

$$T(t) = \frac{1}{k_B N_{df}} \sum_{i=1}^{3N} m_i v_i^2(t). \tag{2.32}$$

In systems with many degrees of freedom, the instantaneous errors of lower-order estimators for the squared velocities, such as $v_n^2(\mathrm{AVG})$ (2.22), tend to average out. Notwithstanding, we have established that the estimator $v_n^2(\mathrm{LF})$ (2.21) has a systematic bias, which remains apparent even when averaged over the system's degrees of freedom. In

System	Number of atoms		Feedback temperature	LF temperature
	Solute	Solvent	[K]	[K]
Na^+ in water	1	3408	300.00	299.54
TCR-pMHC	6286	77943	300.00	299.35

Table 2.2: *Average temperature calculated with the biased expression v_n^2(LF), Eq. (2.21), from MD simulations at 300 K, with $\Delta t = 1$ fs, and using an unbiased thermostat feedback.*

the following we look at the impact on the temperature $T(t)$ resulting from the choice of the biased v_n^2(LF) versus non-biased v_n^2(AVG) for estimating the full time step squared velocities.

The first system is a box of 1136 water molecules with one sodium ion. The bonds and the bond angle in the water molecules are constrained to their ideal values. There are no harmonic terms in this system, only electrostatic and LJ interactions. The second system is a large protein assembly of 605 residues, the A6 T-cell receptor bound to the HLA-A2 major histocompatibility complex[23]. The protein is immersed in a rectangular periodic box with 25981 water molecules, and the force field used is GROMOS 43A1 [22]. All bonds involving hydrogens are constrained. In both cases the simulation is done with the GROMACS 3.3 software [24], which uses v_n^2(AVG) to couple the system temperature to a Nosé-Hoover thermostat [25] at 300K and time constant 0.15 ps. The electrostatic interactions are treated with the particle-mesh Ewald scheme [26] with a direct-space cutoff of 0.8 nm. The LJ interactions are updated every step up to a cutoff distance of 0.8 nm, and every 5 steps up to a cutoff distance of 1.4 nm. The integration time step is 1 fs, and data is averaged over 100 ps. Table (2.2) shows the average temperature calculated using v_n^2(LF) (2.21). As expected from the sign of expression (2.20) it lies below the true temperature. This effect is more pronounced for the protein system, which contains harmonic bonds.

2.5 Discussion

We have suggested the use of the more accurate expression v_n^2(HOnr) (2.24) or of the unbiased one v_n^2(AVG) (2.22) instead of the standard leap-frog one v_n^2(LF) (2.21) in MD simulation. This will improve the accuracy of molecular or system properties that depend on the squared velocities at full time steps. Often, the accuracy of an integration scheme is evaluated by considering the degree of conservation of the total energy, e.g. $x^2(t)+v^2(t)$ for the one-dimensional harmonic oscillator. Using a low-order or biased formula for estimating the squared velocities might lead to erroneous conclusions [27] regarding the accuracy of the formula for propagating the positions [28].

The leap-frog integration scheme can also be formulated for stochastic dynamics (SD) simulation [29], in which instead of Newton's equations (2.5) the Langevin equation

$$\dot{v}_i(t) = \frac{1}{m_i} f_i(t) + \frac{1}{m_i} f_i^{\text{stoch}}(t) - \gamma_i v_i(t) \tag{2.33}$$

System	Δt [fs]	True temperature [K]	Feedback pressure [bar]	True pressure [bar]
Na^+ in water	1	300.46	1.01	5.14
Na^+ in water	2	301.83	1.24	17.83
TCR-pMHC	1	300.65	1.05	6.97
protein		302.23		
solvent		300.48		

Table 2.3: *Unbiased temperature and pressure from MD simulations with reference temperature 300 K and reference pressure 1 atm, but using the biased v_n^2(LF), Eq. (2.21), for the thermostat and barostat feedback.*

is integrated forward in time. The particle i is subjected to a force f_i^{stoch} and a frictional force, the size of which is determined by the particle friction coefficient γ_i. Because of the velocity term occurring in Eq. (2.33), the invariance of the equation of motion under time reversal is lost and the SD leap-frog algorithm becomes more complex than the MD one [22, 30]. Yet, use of v_n^2(HOnr) (2.24) or v_n^2(AVG) (2.22) instead of v_n^2(LF) (2.21) is also recommended in this case.

In order to speed up MD or SD simulations in which motions of different, well separated time scales occur, the use of multiple-time-step algorithms has been proposed [31, 19, 32]. The use of higher-accuracy expressions for the squared velocities should also be considered in those cases.

As shown in Table 2.2, using the biased expression v_n^2(LF) (2.21) leads to a systematic undervaluation of the temperature of the system (at full time steps). This effect impacts the simulation if the biased full time step temperature is used to couple the system to a heat bath. In such a constant temperature simulation, a thermostat is used to keep the average biased temperature close to the reference temperature (300 K). Table 2.3 illustrates that in this case the true average temperature of the system systematically lies above the reference temperature. The difference is more important in the protein itself, where harmonic bonds are present.

The consequences are more severe if the system is in addition subjected to pressure coupling. In this case, the volume of the simulation box is scaled in such a way that the average internal pressure of the system matches a reference pressure. As a consequence of the virial theorem, the internal scalar pressure at time t, which can be used for pressure coupling in isotropic systems, is calculated as

$$p(t) = \frac{1}{3V(t)} \sum_{i=1}^{3N} \Big[m_i v_i^2(t) + x_i(t) \cdot f_i(t) \Big].$$

Here, the squared velocities should be taken at full time steps, since the positions and forces come at full time steps. The second term usually being negative for condensed phase systems, a small absolute error on $v_i^2(t)$ may result in a relatively higher error on $p(t)$. Table 2.3 shows the true average pressure in molecular simulations in which the biased v_n^2(LF) is fed to the barostat. The same Nosé-Hoover thermostat as before was used, together with a Parrinello-Rahman [33, 34] barostat with time constant 0.5 ps.

2.6 Conclusion

We have derived a higher-order expression for full time step squared velocities in the leap-frog algorithm, where only velocities at half time steps are accurately known. This expression has one term of zeroth order in Δt, and two first-order terms. This decomposition allows to show that the usual leap-frog expression for the squared velocities is a biased estimator. Indeed it is equivalent to the first and second term only, and the latter is shown to be always negative. Appropriate practical expressions for the full time step squared velocity are either the first, zero-order, term only, or the expression including both first-order terms. The zero-order term is readily available as average of the preceding and following squared half step velocities. The full first-order expression can only be evaluated at the cost of an approximation which destroys its formal time-reversal invariance.

We have demonstrated with one-dimensional harmonic and Lennard-Jones systems, that the higher-order expression is more accurate. While the zero-order term shows larger instantaneous errors, it is unbiased. In practical molecular systems with many degrees of freedom, the instantaneous errors cancel out, but the bias of the standard leap-frog estimator remains. We assessed the impact of using this biased estimator (as it is done in some state of the art MD software packages) on temperature and pressure regulation of molecular systems. While the induced temperature error is within one percent, the resulting pressure is manyfold higher than the reference pressure, by a factor of between 5 to 7 with a time step of 1 fs, and a factor of 14 with a time step of 2 fs. For systems with numerous degrees of freedom, which allow for error compensation, use of the unbiased formula (2.22) for the full time step squared velocities seems appropriate. On the other hand, for low-dimensional systems, with less error compensation, use of the higher-order formula (2.24) is recommended.

2.A Appendix A : Leap-frog integrators for coupled systems

In this appendix, we show that the issues related to the estimation of v_n^2 can be bypassed for the coupling of the system to a heat bath. If appropriate time-reversible integration schemes are chosen, only the temperature at half time steps is needed for the thermostat feedback. As a first example, we consider the Berendsen [35] weak coupling scheme,

$$\begin{aligned} v_{n+\frac{1}{2}} &= v_{n-\frac{1}{2}} + \frac{1}{m} f_n \Delta t, \\ v_{n+\frac{1}{2}}^{\text{scaled}} &= v_{n+\frac{1}{2}} \sqrt{1 + \frac{\Delta t}{\tau_T} \left[\frac{T_0}{T_{n+\frac{1}{2}}} - 1 \right]}, \\ x_{n+1} &= x_n + v_{n+\frac{1}{2}}^{\text{scaled}} \Delta t. \end{aligned}$$

The parameter τ_T sets the time scale of the first-order velocity relaxation induced by the scaling term, and T_0 is the reference temperature. We see that the temperature $T_{n+\frac{1}{2}}$, calculated using Eq. (2.32) with the velocities $v_{n+\frac{1}{2}}$, is used.

As a further example, we consider the Nosé-Hoover [25] (NH) thermostated dynamics,

$$\begin{aligned} \dot{r}_i(t) &= v_i(t), \\ \dot{v}_i(t) &= \frac{f_i(t)}{m_i} - \xi v_i, \\ \dot{\xi}(t) &= \frac{1}{Q}(T(t) - T_0). \end{aligned}$$

Here, Q is a pseudo-mass regulating the time scale of the second-order relaxation induced by the coupling variable ξ. A consistent leap-frog scheme can be obtained [36, 37, 38] for the NH dynamics using the same approach as for the standard leap-frog, Eqs. (2.14) and (2.15). This approach involves expanding each of the above variables in Taylor series in $\frac{\Delta t}{2}$ and $-\frac{\Delta t}{2}$ at time point t, subtracting the resulting expansions and rearranging. One finds

$$\xi_n = \xi_{n-1} + \frac{1}{Q}\left[T_{n-\frac{1}{2}} - T_0\right]\Delta t, \tag{2.34}$$

$$v_{n+\frac{1}{2}} = \frac{v_{n-\frac{1}{2}}\left[1 - \frac{1}{2}\xi_n \Delta t\right] + \frac{1}{m} f_n \Delta t}{1 + \frac{1}{2}\xi_n \Delta t}, \tag{2.35}$$

$$x_{n+1} = x_n + v_{n+\frac{1}{2}}\Delta t. \tag{2.36}$$

It is clear that this integrator is fully time-reversible. Again, only $T_{n-\frac{1}{2}}$ is needed, which is accurately calculated from the velocities at half time steps. Only fully consistent integrators such as Eqs. (2.34) to (2.36) should be used in MD software. Alternatively, one could choose the integrators for extended dynamics proposed by Martyna *et al.* [39], based on the Trotter decomposition of the Liouville operator. Note that these integrators are similar to the velocity Verlet [40], and thus provide accurate velocities at full time steps.

2.B Appendix B : Conservation of the shadow Hamiltonian

In Section 2.4.1, the accuracy of discrete kinetic energy estimators with respect to the analytical solution was studied. Instead, one can look at the conservation of the physical Hamiltonian function (2.3) of the system, evaluated at full time steps. The kinetic energy term can be evaluated with estimators v^2(AVG) (2.22), v^2(LF) (2.21), and v^2(HO) (2.23). The corresponding results are shown on the left panel of Figure 2.3 for the same harmonic oscillator as in Section 2.4.1. The degree of conservation of the physical Hamiltonian follows similar trends as the accuracy of the v^2 estimators shown on Figure 2.1. The higher order v(HO) yields a two-order of magnitude better conserved Hamiltonian.

For symplectic integrators of general systems, the strictly conserved shadow Hamiltonian is not known analytically. It can only be approximated via a Lie expansion of the discrete exponential propagation operator [15, 17]. The higher order terms of this expansion are obtained by multiple application of the Baker-Campbell-Hausdorf formula, which gives rise to a large number of cumbersome Poisson brackets. This renders convergence

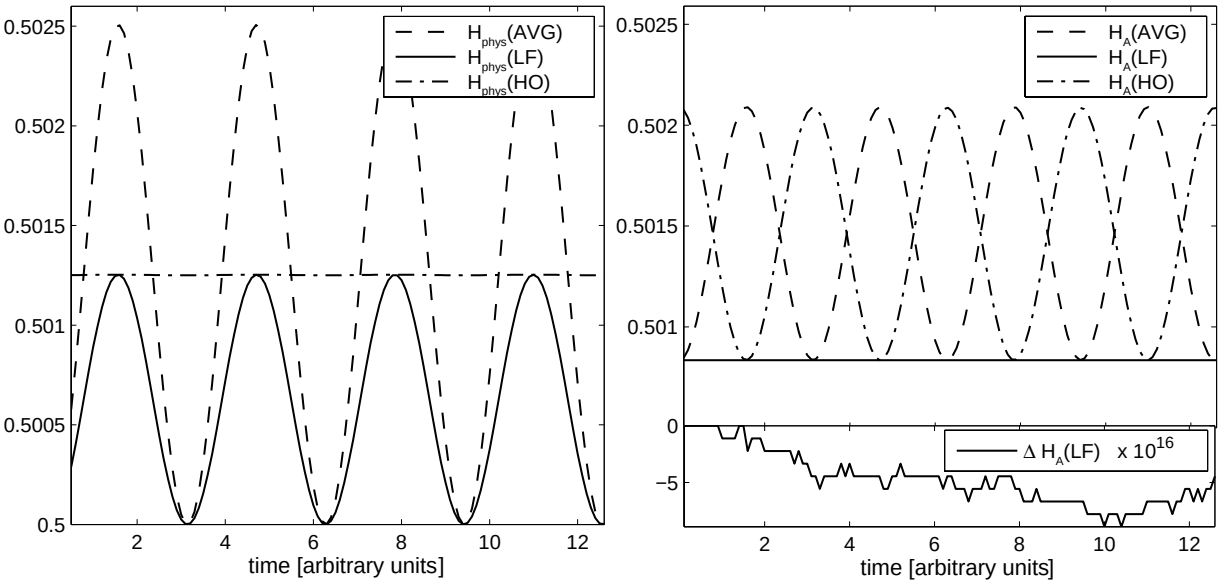

Figure 2.3: *Left panel : For a harmonic oscillator with same parameters as in Figure 2.1, evolution of the physical Hamiltonian calculated with different v_n^2 estimators corresponding to Eqs. (2.22), (2.21), and (2.23), respectively. Right panel : Evolution of the shadow Hamiltonian $H_A(v, r, \Delta t)$ (2.37) with the same v_n^2 estimators. The inset shows the variation of H_A(LF) at machine precision.*

studies and applications difficult. In the special case of the harmonic oscillator however, it is possible to find the exact expression for the shadow Hamiltonian [18, 41],

$$H_A(v, r, \Delta t) = \frac{m}{2|\omega \Delta t|} \arccos\left[1 - \tfrac{\omega^2 \Delta t^2}{2}\right] \left(\frac{v^2}{C} + C\omega^2 r^2\right), \tag{2.37}$$

with

$$C = \sqrt{1 - \left(\frac{\omega \Delta t}{2}\right)^2}.$$

The right panel of Figure 2.3 shows the behavior of $H_A(v, r, \Delta t)$ evaluated at full time steps with the v^2 estimators v^2(AVG) (2.22), v^2(LF) (2.21), and v^2(HO) (2.23). The behavior is strikingly different from the left panel of the same figure. Using v^2(LF) yields a perfect conservation of $H_A(p, r, \Delta t)$ up to machine precision, as is apparent in the lower inset. The higher order estimator v^2(HO) does not bring any improvement in the conservation of $H_A(v, r, \Delta t)$ over the first order v^2(AVG).

Based on this, we can say that the leap-frog is consistent with the Verlet integrator, and thus symplectic, only in the sense of v^2(LF). On the other hand, v^2(HO) gives squared velocities which are closer to the analytical solution. This is true although the velocity $\sqrt{v^2(\mathrm{HO})}$ is not on the discrete trajectory of the symplectic integrator. This suggests that

in spite of their very useful long time stability, symplectic integrators do not always give the best instantaneous accuracy.

Acknowledgments

The work presented in this chapter was directed by W. F. van Gunsteren. We thank Daniel Trzesniak and Philippe H. Hünenberger for stimulating discussions regarding the accuracy of MD algorithms.

2.3 References

[1] H. Goldstein. *Classical Mechanics* (Addison Wesley, Reading, MA, USA, 1950).

[2] C. W. Gear. *Numerical Initial Value Problems in Ordinary Differential Equations* (Prentice-Hall, Engelwood Cliffs, NJ, USA, 1971).

[3] L. Lapidus and J. H. Seinfeld. *Numerical Solutions of Ordinary Differential Equations* (Academic Press, New York, USA, 1971).

[4] G. Dahlquist and A. Björk. *Numerical Methods* (Prentice-Hall, Engelwood Cliffs, NJ, USA, 1974).

[5] H. J. C. Berendsen and W. F. van Gunsteren. "Practical algorithms for dynamic simulations". In: "Proceedings of the International School of Physics "Enrico Fermi", course 97", eds. G. Ciccotti and W. Hoover. North-Holland, Amsterdam, (1986) 43.

[6] W. F. van Gunsteren and H. J. C. Berendsen. "Algorithms for macromolecular dynamics and constraint dynamics". *Mol. Phys.*, **34**, (1977) 1311.

[7] J. Delambre. *Mem. Acad. Turin*, **5**, (1790-93) 143.

[8] C. Störmer. "Sur les trajectoires des corpuscules électrisés dans l'espace sous l'action du magnétisme terrestre, avec application aux aurores boréales". *Arch. Sci. Phys. Nat.*, **24**, (1907) 5,113,221.

[9] L. Verlet. "Computer experiments on classical fluids : I. thermodynamical properties of Lennard-Jones molecules". *Phys. Rev.*, **159**, (1967) 98.

[10] D. Levesque and L. Verlet. "Molecular dynamics and time reversibility". *J. Stat. Phys.*, **72**, (1993) 519.

[11] R. W. Hockney. "The potential calculation and some applications". *Methods Comput. Phys.*, **9**, (1970) 136.

[12] P. J. Channell and C. Scovel. "Symplectic integration of Hamiltonian systems". *Nonlinearity*, **3**, (1990) 231.

[13] J. M. Sanz-Serna. "Symplectic integrators for Hamiltonian problems: an overview". *Acta Numerica*, 243.

[14] H. Yoshida. "Recent progress in the theory and application of symplectic integrators". *Celest. Mech. Dyn. Astro.*, **56**, (1993) 27.

[15] B. Leimkuhler. *Simulating Hamiltonian Dynamics* (Cambridge University Press, 2005).

[16] S. Toxvaerd. "Hamiltonians for discrete dynamics". *Phys. Rev. E*, **50**, (1994) 2271.

[17] R. D. Skeel, G. Zhang, and T. Schlick. "A family of symplectic integrators: Stability, accuracy, and molecular dynamics applications". *SIAM Journal on Scientific Computing*, **18**, (1997) 203.

[18] S. A. Chin and S. R. Scuro. "Exact evolution of time-reversible symplectic integrators and their phase errors for the harmonic oscillator". *Phys. Lett. A*, **342**, (2005) 397.

[19] M. Tuckerman, B. J. Berne, and G. J. Martyna. "Reversible multiple time scale molecular dynamics". *J. Chem. Phys.*, **97**, (1992) 1990.

[20] S. Toxvaerd. "Comment on: Reversible multiple time scale molecular dynamics". *J. Chem. Phys.*, **99**, (1993) 2277.

[21] M. Tuckerman, B. J. Berne, and G. J. Martyna. "Reply to comment on: Reversible multiple time scale molecular dynamics". *J. Chem. Phys.*, **99**, (1993) 2278.

[22] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).

[23] D. N. Garboczi, P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. "Structure of the complex between human t-cell receptor, viral peptide and HLA-A2". *Nature*, **384**, (1996) 134.

[24] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. "Gromacs: Fast, flexible, and free". *J. Comp. Chem.*, **26**, (2005) 1701.

[25] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". *Phys. Rev. A*, **31**, (1985) 1695.

[26] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. "A smooth particle mesh ewald method". *J. Chem. Phys.*, **103**, (1995) 8577–8593.

[27] M. Levitt and H. Meirovitch. "Integrating the equations of motion". *J. Mol. Biol.*, **168**, (1983) 617.

[28] J. Åqvist, W. F. van Gunsteren, M. Leijonmarck, and O. Tapia. "A molecular dynamics study of the C-terminal fragment of the L7/12 ribosomal protein secondary structure motion in a 150 picosecond trajectory". *J. Mol. Biol.*, **183**, (1985) 461.

[29] W. F. van Gunsteren and H. J. C. Berendsen. "A leap-frog algorithm for stochastic dynamics". *Mol. Simulat.*, **1**, (1988) 173.

[30] W. R. P. Scott, P. H. Hünenberger, I. G. Tironi, A. E. Mark, S. R. Billeter, J. Fennen, A. E. Torda, T. Huber, P. Krüger, and W. F. van Gunsteren. "The GROMOS biomolecular simulation program package". *J. Phys. Chem. A*, **103**, (1999) 3596.

[31] W. F. van Gunsteren and H. J. C. Berendsen. "Computer simulation of molecular dynamics: Methodology, applications and perspectives in chemistry". *Angew. Chem. Int. Ed. Engl.*, **29**, (1990) 992.

[32] X. Qian and T. Schlick. "Efficient multiple-time-step integrators with distance-based force splitting for particle-mesh-Ewald molecular dynamics simulations". *J. Chem. Phys.*, **116**, (2002) 5971.

[33] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals : A new molecular dynamics method". *J. Appl. Phys.*, **52**, (1981) 7182.

[34] S. Nosé and M. L. Klein. "Constant pressure molecular dynamics for molecular systems". *Mol. Phys.*, **50**, (1983) 1055.

[35] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak. "Molecular dynamics with coupling to an external bath". *J. Chem. Phys.*, **81**, (1984) 3684.

[36] S. Toxvaerd. "Algorithms for canonical molecular dynamics simulations". *Mol. Phys.*, **72**, (1991) 159.

[37] S. Toxvaerd. "Molecular dynamics at constant temperature and pressure". *Phys. Rev. E*, **47**, (1993) 343.

[38] B. L. Holian, A. F. Voter, and R. Ravelo. "Thermostatted molecular dynamics: How to avoid the Toda demon hidden in Nosé–Hoover dynamics". *Phys. Rev. E*, **52**, (1995) 2338.

[39] G. J. Martyna, M. E. Tuckerman, D. J. Tobias, and M. L. Klein. "Explicit reversible integrators for extended systems dynamics". *Mol. Phys.*, **87**, (1996) 1117.

[40] W. C. Swope, H. C. Andersen, P. H. Berens, and K. R. Wilson. "A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters". *J. Chem. Phys.*, **76**, (1982) 637.

[41] T. F. Miller, G. J. Martyna, M. Eleftheriou, P. Pattnaik, A. Ndirango, and D. Newns. "Symplectic quaternion scheme for biophysical molecular dynamics". *J. Chem. Phys.*, **116**, (2002) 8649.

Chapter 3

Statistical mechanical derivation of the Jarzynski identity for thermostated non-Hamiltonian dynamics

Summary

The Jarzynski identity (JI) relates thermodynamic free energy differences to nonequilibrium work averages. Several proofs of the JI have been provided on the thermodynamic level. They rely on assumptions such as equivalence of ensembles in the thermodynamic limit, or weakly coupled infinite heat baths. However, the JI is widely applied to NVT computer simulations involving finite numbers of particles, whose equations of motion are strongly coupled to few extra degrees of freedom modeling a thermostat. In this case, the above assumptions are no longer valid. We propose a statistical mechanical approach to the JI solely based on the specific equations of motion, without any further assumption. We provide a detailed derivation for the non-Hamiltonian Nosé-Hoover dynamics, which is routinely used in computer simulations to produce canonical sampling.

3.1 Introduction

The second law of thermodynamics implies that, at constant temperature, the average work $\langle W_{AB} \rangle$ to drive a system from state A to state B cannot be smaller than the Helmholtz free energy difference $\Delta F_{AB} = F_B - F_A$ between the two states, $\Delta F_{AB} \leq \langle W_{AB} \rangle$. Equality holds only if the process is quasi-static or reversible, in which case the work is independent of the transition path between A and B. Note that the quasi-static regime is beyond reach of computer simulation for most interesting systems. In contrast, the recent Jarzynski identity [1] (JI) states that

$$e^{-\beta \Delta F_{AB}} = \langle e^{-\beta W_{AB}(\Gamma_0)} \rangle_{\Gamma_0}. \tag{3.1}$$

The phase Γ describes a system in the generalized coordinates and momenta space. A trajectory starts at time 0 from the initial condition Γ_0 in state A, and is brought to state B at time t. Here the work is path-dependent, and the average $\langle \cdot \rangle_{\Gamma_0}$ is taken over different trajectories with independent canonically distributed Γ_0. Here, $\beta = (k_B T)^{-1}$, with k_B the Boltzmann constant and T the temperature of the initial equilibrium ensemble. The JI opens the possibility of calculating equilibrium free energy differences from nonequilibrium processes. In theory, the JI holds regardless of how far from equilibrium the process occurs. However, practical application is limited by the requirement to have a large enough collection of trajectories for accurate estimation of the ensemble average involved in (3.1).

In most applications, the system is thermally coupled to an environment. Sometimes this is unavoidable, when performing or modeling experiments in solution. Sometimes this is desirable, for example to control the kinetic energy in a computer simulation (note that temperature itself is ill defined in a nonequilibrium process). In both cases, the dissipative part of the work applied to the physical system is converted into heat, which is absorbed by the environment. This heat bath, or thermal reservoir, has to be included in the model. On the one hand, in thermodynamical descriptions, heat baths are considered very large, and the nature of the coupling to the system is deliberately disregarded. On the other hand, in computer simulation the physical equations of motion are explicitly modified to include coupling to a few additional dynamic variables, which model heat baths and external pressure. Hoover coined the term "thermomechanics" [2] to emphasize the theoretical importance of such thermostated Newtonian systems. Indeed, integrating the resulting coupled equations of motion allows to solve Loschmidt's paradox, i.e. explain the emergence of irreversible processes from time-reversible dynamics. In this sense, thermostated dynamical systems open up new horizons in statistical mechanics.

Most former derivations approach the JI from the thermodynamical point of view. The original 1997 proof [1] and its later refinement [3] by Jarzynski both show that the JI holds for Hamiltonian systems coupled to infinite heat baths. The first proof uses the strong hypothesis of weak coupling of the system to the temperature bath. The later proof lifts this hypothesis, but explicitly uses the equivalence of microcanonical and canonical ensembles in the limit of large number of particles. Further proofs were provided under similar thermodynamic-like conditions using path integrals and the Feynman-Kac theorem [4], as well as for Markov-chain dynamics [5] or Langevin evolution [6]. Park and Schulten [7] demonstrated that the relation holds for the Gibbs free energy in an isothermal-isobaric ensemble. An alternate derivation proposed by Jarzynski [8] is based on the master equation with only two general hypotheses: Markovian evolution and detailed

balance. The NH dynamics is shown to satisfy these hypotheses, but the resulting route to the JI is different and less direct than the one proposed here[1]. The link of the JI with the transient fluctuation theorem [11] was elucidated for stochastic systems by Crooks [5] and for deterministic systems by Jarzynski [12] and Evans [13], who later proposed a unified reformulation [14] of the two results. Liphardt *et al.* [15] showed the relevance of the JI to experimental data as well.

A large spectrum of temperature or pressure coupling schemes, each altering the dynamics in a different way, are routinely used in computer simulation of systems of interacting particles [16]. In deterministic thermostats, which do not introduce random external forces, the physical equations of motion are coupled to a few additional dynamic variables modeling heat bath and external pressure. The resulting schemes are shown to reproduce the desired ensembles in the infinite time limit, but the alterations to the Hamiltonian dynamics cannot be *a priori* considered a weak coupling. Furthermore, the number of degrees of freedom considered is always finite, and sometimes very small. In this case, the above thermodynamic assumptions are not met. There is, to our knowledge, no direct derivation of the JI from the equations of motion themselves. Starting from the idea that they contain all the information about the system, we propose to investigate whether they directly imply the JI, from the statistical mechanical (thermomechanical) standpoint.

For the present derivation, we use the well studied and widely implemented Nosé-Hoover [17, 18] (NH) dynamics. Improved thermostats exist, but NH is a prototypical example providing deterministic canonical sampling. The NH thermostat perturbs the dynamics in such a way that it becomes non-Hamiltonian. Tuckerman*et al.*[10] provided an efficient framework to study such systems and to derive the partition function of the physical variables from the extended system equations of motion. Although sometimes controversial [19], the framework of Tuckerman*et al.* [20, 10] underpins our developments, and we summarize it in Section 3.1.2 of this introduction.

3.1.1 Nonequilibrium molecular dynamics

Let us focus now on the case of a thermostated system under nonequilibrium conditions. We consider a physical system modeled by a set of positions r and momenta p in contact with a heat bath. The condensed notation $x = (r, p)$ represents the physical variables, which are complemented by a set of thermostating variables Γ^{thermo}, such that the phase vector is $\Gamma = (x, \Gamma^{\text{thermo}})$. Consider the general dynamical system

$$\dot{\Gamma} = \dot{\Gamma}(\Gamma, t). \tag{3.2}$$

The ensemble of all possible values for the vector Γ constitutes the phase space. A solution of (3.2) is a vector $\Gamma = \Gamma(t, \Gamma_0)$, function of time t and an initial condition Γ_0. A trajectory or path starting from Γ_0 at $t = 0$ and ending at time τ is defined as the set $\{\Gamma(t, \Gamma_0)\}_{t\in[0,\tau]}$. The evolution can also be denoted with the flow ϕ,

$$\Gamma = \phi_t(\Gamma_0). \tag{3.3}$$

[1]The assertion made in Ref. [9], that all prior proofs of the JI relied on hypotheses not verified in molecular simulation, was a bit overstated. Jarzynski had addressed the case of the NH dynamics assuming a canonical distribution for (r, p, p_η). The framework of Ref. [10] used here seems more complete, since it allows to show that this assumption is not true in the presence of a second motion invariant.

Suppose that the physical system is described by a standard Hamiltonian of the form $H(r,p,t) = \sum_i \frac{p_i^2}{2m_i} + \Phi(r,t)$. The nonequilibrium condition is reached by varying the potential energy term $\Phi(r,t)$ via its explicit time dependence. This is analogous to an external agent performing mechanical work on the system in order to drive it away from equilibrium. Under this stress, the system inevitably tends to heat up due to the irreversible conversion of work into heat [21]. This is compensated by the thermostat, which absorbs the excess energy. In this perspective, the thermostat coupling mediates the reaction of the physical system to the action of the time varying potential. This can be seen in detail by looking at the time variation of the energy $H(x,t)$ of the physical system between time 0 and τ,

$$\begin{aligned} H(x(\tau),\tau) - H(x_0,0) &= \int_0^\tau dt \frac{\partial H}{\partial x}(x(t),t)\,\dot{x}(t) + \int_0^\tau dt \frac{\partial H}{\partial t}(x(t),t) \\ &= \mathcal{Q} \quad + \quad W\ . \end{aligned} \tag{3.4}$$

The first integral is commonly interpreted as the heat $\mathcal{Q}$ transferred from the heat bath to the system, while the second integral is the work W performed on the system. Relation (3.4) can then be seen as a statement of the first law of thermodynamics. As pointed out by Jarzynski [3], $\mathcal{Q}$ and W are defined here in terms of the microscopic history of the physical system alone, $\{x(t)\}_{t\in[0,\tau]}$. Even though the system is part of an extended phase space including all degrees of freedom of the environment, an observer needs to monitor only the internal degrees of freedom $\{x\}$, in order to deduce how much work was performed on the system and how much heat was absorbed by it during a given realization of the process. But this implies knowing the whole history of the physical system up to time τ. Alternatively, if the dynamics of the bath is known, the work can be specified by one point in the extended phase space at a given time (typically the initial condition Γ_0), and a restriction of the flow ϕ (3.3) of the form $x = \varphi_t(\Gamma_0)$. In this case, the work accumulated at time τ along a nonequilibrium trajectory is defined as

$$W_\tau(\Gamma_0) = \int_0^\tau dt \frac{\partial H}{\partial t}(\varphi_t(\Gamma_0),t)\ . \tag{3.5}$$

Here we express the evolution of $x(t)$ with a restriction of the flow ϕ of the form $x(t) = \varphi_t(\Gamma_0)$. The extended initial condition Γ_0 has to be given as argument, since x is coupled to the bath.

As a side note, the meaning of $\mathcal{Q}$ in (3.4) becomes explicit if we admit standard equations of motion for a system coupled to a heat bath of the form $\dot{r} = \frac{p}{m}, \dot{p} = -\frac{\partial\Phi}{\partial r} - \alpha(\Gamma^{\text{thermo}})p$, where α is an arbitrary function of the heat bath variables. According to (3.4), the heat term takes the form $\mathcal{Q} = -\int dt\ \alpha(\Gamma^{\text{thermo}})\frac{p^2}{m}$. This is equivalent to considering the coupling term as a friction force $f^{\text{thermo}} = -\alpha(\Gamma^{\text{thermo}})p$, and calculating its work on the system via the usual formula $\int dt\ f^{\text{thermo}} \cdot v$. We see that $\mathcal{Q}$ really represents the work of the thermostat, i.e. the energy pumped by it into the system.

3.1.2 Theory of nonautonomous non-Hamiltonian dynamics

Thermostats of the NH family as well as the KFC thermo-barostat perturb the dynamics in such a way that it becomes non-Hamiltonian. Tuckerman *et al.* [10] provided an efficient

framework to study such non-Hamiltonian extended dynamical systems. In the following, this approach is recapitulated, in a way that emphasizes its validity in the nonautonomous case. Under the dynamics (3.2), the initial phase space volume element $d\Gamma_0$ around Γ_0 is transformed into $d\Gamma$ at time t by the Jacobian of the coordinate transformation $d\Gamma = J(\Gamma, \Gamma_0)d\Gamma_0$. This relation can be rearranged [20] in such a way that both sides of the equation take an invariant form,

$$\sqrt{g(\Gamma, t)}d\Gamma = \sqrt{g(\Gamma_0, 0)}d\Gamma_0. \tag{3.6}$$

The term $\sqrt{g(\Gamma, t)}$ is called the metric factor, and $\sqrt{g(\Gamma, t)}d\Gamma$ constitutes a time-invariant measure of the phase space. This notation comes from the Riemannian formalism, where $g(\Gamma, t)$ is the determinant of the metric tensor $\mathcal{G}$ which describes the geometry of space. $\mathcal{G}$ can be determined [22] explicitly from the dynamics. It was also shown [23] using the general formalism of differential forms, that an invariant phase space measure can be determined even without defining a metric. These aspects however reach beyond the scope of the present work, but we keep the notation $\sqrt{g(\Gamma, t)}$ for consistency with Refs. [20] and [10]. As far as we are concerned here, the metric factor can be determined [20, 10, 23] directly from the dynamics (3.2) through[2]

$$\sqrt{g(\Gamma, t)} = \exp\left\{-\int dt\ \kappa\left(\Gamma(t), t\right)\right\}, \tag{3.7}$$

where $\kappa(\Gamma, t)$ is the phase space compression factor, defined as

$$\kappa(\Gamma, t) = \nabla_\Gamma \cdot \dot{\Gamma}(\Gamma, t). \tag{3.8}$$

This formalism shows that in the description of non-Hamiltonian or nonequilibrium systems, the phase space need not be Euclidean, and the metric need not be constant. Note that for Hamiltonian systems on the other hand, the compression factor vanishes and the metric factor remains constant, $\sqrt{g(\Gamma, t)} \equiv 1$, meaning that the phase space is incompressible.

Suppose that the dynamical system (3.2) possess a set of n_c conservation laws or conserved quantities $\{c_j(\Gamma, t) \mid j = 1, ..., n_c\}$, which satisfy $\frac{d}{dt}c_j(\Gamma, t) = 0, \forall j$. In this case, a trajectory will not sample the entire phase space, but a subspace determined at each time t by the intersection of the hypersurfaces $\{c_j(\Gamma, t) = C_j\}$, where $\{C_j\}$ represents a set of constants determined by the initial conditions. Therefore, the dynamics (3.2) generate a microcanonical phase density per unit of Cartesian phase space volume, constructed as a product of δ-functions expressing the conservation laws,

$$f(\Gamma, t) = \prod_{j=1}^{n_c} \delta\left[c_j(\Gamma, t) - C_j\right]. \tag{3.9}$$

It is important to identify all conservation laws for the system of interest, otherwise $f(\Gamma, t)$ will not properly describe the phase space distribution of the system [10]. According to

[2]Note that the indefinite integral in equation (3.7) raises questions about the unicity of $\sqrt{g(\Gamma, t)}$. See the discussion in Ref. [23].

our definition (3.6) of the phase space measure, the time-dependent partition function is given by

$$\Omega_t \propto \int d\Gamma \; f(\Gamma, t)\sqrt{g(\Gamma, t)}. \tag{3.10}$$

The complete definition of the partition function includes a constant prefactor containing the number of degrees of freedom and a combinatorial term accounting for indistinguishable particles. For simplicity, the prefactor will be omitted throughout the rest of this paper.

Tuckerman *et al.* [10] propose a general procedure for constructing the partition function generated by a non-Hamiltonian system satisfying the conditions mentioned above. Here we rewrite this procedure in a form that emphasizes its validity in the explicitly time-dependent case :

1. Determine all conservation laws satisfied by the equations of motion, and construct the microcanonical phase space density $f(\Gamma, t)$ according to equation (3.9).

2. Using the conservation laws and the equation of motion, identify and eliminate linearly dependent variables, as well as driven or uncoupled variables.

3. Calculate the phase space compression factor $\kappa(\Gamma, t)$ using (3.8) and the metric factor $\sqrt{g(\Gamma, t)}$ using (3.7), and generate the invariant volume element $\sqrt{g(\Gamma, t)}d\Gamma$.

4. Construct the partition function at time t using the definition (3.6) of the phase space measure,

$$\Omega_t(C_1, ..., C_{n_c}) \propto \int d\Gamma \sqrt{g(\Gamma, t)} \prod_{j=1}^{n_c} \delta\left[c_j(\Gamma, t) - C_j\right]. \tag{3.11}$$

5. If equation (3.2) corresponds to a physical system coupled to an extended phase space, integrate expression (3.11) over the extra degrees of freedom, in order to determine the distribution function sampled by the physical variables.

In this framework, the average of any instantaneous observable $A(\Gamma)$ over the ensemble at time t is given by

$$\langle A \rangle_t = \frac{1}{\Omega_t} \int d\Gamma \; A(\Gamma) f(\Gamma, t)\sqrt{g(\Gamma, t)}. \tag{3.12}$$

3.2 The Jarzynski identity as a property of the Nosé-Hoover dynamics

Consider now a system of N physical particles in 3 dimensions with Cartesian positions $r = \{r_1, ..., r_N\}$, momenta $p = \{p_1, ..., p_N\}$, and masses $\{m_1, ..., m_N\}$. The canonical ensemble is described by the partition function

$$Z(N, V, T) = \int dr dp \; e^{-\beta H(r,p)}, \tag{3.13}$$

where $H(r,p)$ is the Hamiltonian of the physical system, and the constant prefactor is omitted. The NH thermostating method involves an extended phase space $\{\Gamma = (r, p, \eta, p_\eta)\}$, where physical positions r and momenta p ($3N$-vectors) are coupled to the real-valued thermostating variable η and its associated momentum p_η. Evolution is governed by the non-Hamiltonian dynamical system

$$\begin{aligned} \dot{r}_i &= \frac{p_i}{m_i}, \qquad i = 1, ..., N \\ \dot{p}_i &= -\frac{\partial \Phi}{\partial r_i}(r,t) - \frac{p_\eta}{Q} p_i \\ \dot{\eta} &= \frac{p_\eta}{Q} \\ \dot{p}_\eta &= \sum_i \frac{p_i^2}{m_i} - N_{df} k_B T \ . \end{aligned} \tag{3.14}$$

N_{df} is the number of physical degrees of freedom. It is shown [24] that an optimal choice for the thermostat mass is $Q = N_{df} k_B T \tau_{\mathrm{NH}}^2$, with τ_{NH} a typical timescale for the system under study. The equation for η is often omitted in the literature (it was not present in the original formulations [17, 18]) and it is not integrated in simple implementations. This is due to the fact that none of $\{r, p, p_\eta\}$ are coupled to η. We will refer to this property by saying that η is *noncoupling*. This particularity allows to write $x_t = \varphi(x_0, p_{\eta_0})$, independently of η, which is a key prerequisite for the following derivation. Nonetheless, η has an important physical meaning. Indeed, according to (3.7) the metric factor becomes

$$\sqrt{g(\Gamma_t, t)} = e^{N_{df}\eta_t}. \tag{3.15}$$

According to this formalism, the phase space compression does not directly depend on any explicit time variation of the potential, but on the state of the thermostat itself, as described by η. Note that instantaneous compression exists even if the system is unperturbed.

At this point, it is useful to recall a couple of properties of the autonomous NH dynamics, i.e. with the explicit time-dependence of the potential removed, $\Phi(r) \equiv \Phi(r,0)$ (equilibrium conditions). First, the pseudo-Hamiltonian

$$H'(\Gamma) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \Phi(r) + \frac{p_\eta^2}{2Q} + N_{df} k_B T \eta \tag{3.16}$$

is an invariant of the motion [18]. This can be readily verified by calculating the direct time derivative of $H'(\Gamma(t))$. The second and key property of the autonomous NH thermostat is the following: The microcanonical distribution in the extended phase space projects to canonically distributed variables (r,p) in the physical phase space [17], with the usual partition function (3.13). We briefly demonstrate this property by applying the method of Tuckerman *et al.* [10] described above. The distribution function in the extended phase space is constructed according to (3.9) from the constant of the motion (3.16). Then, using the metric factor (3.15), the extended partition function becomes

$$\Omega = \int dx dp_\eta d\eta \ e^\eta \delta \left[H'(\Gamma) - C \right].$$

A property of the Dirac δ-function is going to be useful several times in the following: for any smooth function $\ell(\eta)$ with one simple root $\eta_\star$,

$$\delta[\ell(\eta)] = \delta[\eta - \eta_\star]/|\ell'(\eta_\star)| \ . \tag{3.17}$$

Using this and the conserved quantity (3.16) yields

$$\delta\left[H'(\Gamma) - C\right] = \beta\delta\left[-\frac{\beta}{N_{df}}\left(H(x) + \sum_{k=1}^{M}\frac{p_{\eta_k}^2}{2Q_k} - C\right) - \eta\right] .$$

Integrating Ω over η allows to eliminate the Dirac delta function,

$$\Omega = \beta e^{\beta C}\int dp_\eta e^{-\beta\sum\frac{p_{\eta_k}^2}{2Q_k}}\int dx e^{-\beta H(x)} \propto Z(N,V,T) \ . \tag{3.18}$$

The integral over the thermostat momenta p_η is a constant. We see that the autonomous NH thermostat reproduces the canonical partition function for the physical variables. As explained in [10], if there are conserved quantities other than H', such as the center of mass momentum ($\sum_i f_i = 0$), the simple NH thermostat no longer generates canonical sampling (but Nosé-Hoover chains [24] do).

In the following, we are going to calculate the free energy difference between two thermodynamic states A and B described by the two Hamiltonians $H_A(x)$ and $H_B(x)$. The equilibrium free energy difference is given by

$$e^{-\beta\Delta F_{AB}} = \frac{\int e^{-\beta H_B(x)}dx}{\int e^{-\beta H_A(x)}dx} = \frac{Z_B}{Z_A} \ . \tag{3.19}$$

We assume that $H_A(x)$ can be transformed into $H_B(x)$ by a continuous variation of the potential $\Phi(r,t)$. The nonequilibrium Hamiltonian $H(x,t)$ is initially equivalent to $H_A(x)$, and varies such that $H(x,\tau) = H_B(x)$ at final time τ. The system is correctly described at $t = 0$ by the equilibrium partition function corresponding to state A. We generalize the expression of the NH pseudo-Hamiltonian to the case of a nonautonomous potential $\Phi(r,t)$ driving the system out of equilibrium, by replacing $\Phi(r)$ with $\Phi(r,t)$ in (3.16) and noting $H'(\Gamma_t,t)$. Of course, $H'(\Gamma_t,t)$ is not invariant anymore, but it can be shown by computing explicitly its time derivative that we have the following invariant of the motion,

$$c(\Gamma_t,t) = H'(\Gamma_t,t) - W_t(\Gamma_0) = C \ ,$$

where we have used the definition (3.5) of the work. In a process starting at $t = 0$ and ending at $t = \tau$, we can express this as a conservation law with only quantities at τ on the left hand side. Using the inverse phase space flow ϕ^{-1} and noting that $W_0(\Gamma_0) = 0$, we get

$$H'(\Gamma_\tau,\tau) - W_\tau(\phi_\tau^{-1}(\Gamma_\tau)) = H'(\Gamma_0,0) \ . \tag{3.20}$$

3.2.1 Derivation of the Jarzynski identity

We perform the following experiment: We prepare the system in an equilibrium state with the initial Hamiltonian $H_A(\Gamma)$. Starting from different initial conditions taken from this equilibrium ensemble, we change the Hamiltonian from $H(\Gamma, 0) = H_A(\Gamma)$ to $H(\Gamma, \tau) = H_B(\Gamma)$.

We start the derivation by writing out the mathematical object $\langle e^{-\beta W_\tau(\Gamma_0)}\rangle_{\Gamma_0}$ according to (3.12):

$$\frac{1}{Z_A}\int d\Gamma_0 \sqrt{g(\Gamma_0,0)} e^{-\beta W_\tau(\Gamma_0)} \delta\left[H'(\Gamma_0,0) - C\right].$$

This is a correct ensemble average for Γ_0. We now perform a change of variables $\Gamma_0 \to \Gamma_\tau = \phi_\tau(\Gamma_0)$, with $\Gamma_0 = \phi_\tau^{-1}(\Gamma_\tau)$. We know from (3.6) that the phase space measure is invariant. This yields

$$\frac{1}{Z_A}\int d\Gamma_\tau \sqrt{g(\Gamma_\tau,\tau)} e^{-\beta W_\tau(\phi_\tau^{-1}(\Gamma_\tau))} \delta\big[H'(\phi_\tau^{-1}(\Gamma_\tau),0) - C\big].$$

Since x is independent of η in NH, $W_\tau(\phi_\tau^{-1}(\Gamma_\tau)) = W_\tau(\varphi_\tau^{-1}(x_\tau, p_{\eta_\tau}))$. To lighten the notation, we use the shorthand $W_\tau(x, p_\eta)$, and drop the τ index on all phase space variables. Using conservation law (3.20), as well as the explicit form (3.15) of the metric factor, we write out the integral over the different components of Γ_τ,

$$\frac{1}{Z_A}\int dx\ dp_\eta\ e^{-\beta W_\tau(x,p_\eta)} \int d\eta\ e^{N_{df}\eta}\ \delta\left[H'(\Gamma,\tau) - W_\tau(x,p_\eta) - C\right].$$

Using (3.17) and remembering that $H'(\Gamma,\tau) = H(x,\tau) + \frac{p_\eta^2}{2Q} + \frac{N_{df}}{\beta}\eta$, we integrate over η,

$$\begin{aligned}\frac{1}{Z_A}\frac{\beta}{N_{df}}\int dx\ dp_\eta\ e^{-\beta W_\tau(x,p_\eta)} e^{-\beta\left[H(x,\tau)+\frac{p_\eta^2}{2Q} - W_\tau(x,p_\eta) - C\right]} \\ = \frac{1}{Z_A}\frac{\beta D e^{\beta C}}{N_{df}}\int dx\ e^{-\beta H(x,\tau)}\ .\end{aligned}$$

The contribution of p_η is a Gaussian, which integrates out in the constant D. The last integral corresponds to the equilibrium partition function Z_B at state B, because the same prefactor arises when deriving the equilibrium partition function alone [10]. We finally have the result

$$\langle e^{-\beta W_\tau(\Gamma_0)}\rangle_{\Gamma_0} = \frac{Z^B}{Z^A} = e^{-\beta \Delta F_{AB}}\ .$$

3.3 Conclusion

We have presented a new derivation scheme for the JI, and applied it to the Nosé-Hoover dynamics, which is implemented in a number of widely used molecular dynamics simulation packages. This work provides the simulator who plans to use the JI with a formal proof that it strictly holds for the equations of motion used. Of course, this does not alleviate the main caveat for the practical application of the JI, namely the requirement to produce enough trajectories to ensure convergence of the work average in (3.1).

With respect to the thermodynamic proofs of the JI, the present derivation suffers a generality loss, because it is specific to the particular dynamics considered. On the other hand, our derivation is free of any assumptions on the coupling to and the size of a theoretical heat bath, the thermodynamic limit or the equivalence of microcanonical and canonical ensembles. On the theoretical level, the present work also illustrates the usefulness of the controverted formalism of Tuckerman et al. [20, 10], which recovers for the NH dynamics a generally admitted result.

Acknowledgments

Thanks to G. Ciccotti for an insightful discussion, as well as P. H. Hünenberger, B. Žagrović and I. Tavernelli for proofreading the manuscript.

3.4 References

[1] C. Jarzynski. "A nonequilibrium equality for free energy differences". *Phys. Rev. Lett.*, **78**, (1997) 2690.

[2] W. G. Hoover. *Time Reversibility, Computer Simulation, and Chaos* (World Scientific Publishing, Singapore, 1999).

[3] C. Jarzynski. "Response to cohen and mauzerall". *arXiv:cond-mat*, 0407340.

[4] G. Hummer and A. Szabo. "Free energy reconstruction from nonequilibrium single-molecule pulling experiments". *PNAS*, **98**, (2001) 3658–3661.

[5] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". *Phys. Rev. E*, **60**, (1999) 2721.

[6] J. M. Schurr and B. S. Fujimoto. "Equalities for the nonequilibrium work transferred from an external potential to a molecular system. analysis of single-molecule extension experiments". *J. Phys. Chem. B*, **107**, (2003) 14 007.

[7] S. Park and K. Schulten. "Calculating potentials of mean force from steered molecular dynamics information". *J. Chem. Phys.*, **120**, (2004) 5946.

[8] C. Jarzynski. "Equilibrium free energy differences from nonequilibrium measurements: a master equation approach". *Phys. Rev. E*, **56**, (1997) 5018.

[9] M. A. Cuendet. "Statistical mechanical derivation of Jarzynski's identity for thermostated non-Hamiltonian dynamics". *Phys. Rev. Lett.*, **96**, (2006) 120 602.

[10] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna. "Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems". *J. Chem. Phys.*, **115**, (2001) 1678.

[11] D. J. Evans and D. J. Searles. "The fluctuation theorem". *Adv. Phys.*, **51**, (2002) 1529.

[12] C. Jarzynski. "Hamiltonian derivation of a detailed fluctuation theorem". *J. Stat. Phys.*, **98**, (2000) 77.

[13] D. J. Evans. "A nonequilibrium free energy theorem for deterministic systems". *Mol. Phys.*, **101**, (2003) 1551.

[14] J. C. Reid, E. M. Sevick, and D. J. Evans. "A unified description of two theorems in non-equilibrium statistical mechanics: The fluctuation theorem and the work relation". *Europhys. Lett.*, **72**, (2005) 726.

[15] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamente. "Equilibrium information from nonequilibrium measurements in an experimental test of jarzynski's identity". *Science*, **296**, (2002) 1832.

[16] P. H. Hünenberger. "Thermostat algorithms for molecular dynamics simulations". *Adv. Polym. Sci.*, **173**, (2005) 105.

[17] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". *Phys. Rev. A*, **31**, (1985) 1695.

[18] S. Nosé. "An extension of the canonical ensemble molecular dynamics method". *Mol. Phys.*, **57**, (1986) 187.

[19] J. D. Ramshaw. "Remarks on non-Hamiltonian statistical mechanics". *Europhys. Lett.*, **59**, (2002) 319.

[20] M. E. Tuckerman, C. J. Mundy, and G. J. Martyna. "On the classical statistical mechanics of non-Hamiltonian systems". *Europhys. Lett.*, **45**, (1999) 149–155.

[21] W. G. Hoover. *Computational Statistical Mechanics*. Studies in Modern Thermodynamics (Elsevier Science, Amsterdam, 1991).

[22] V. E. Tarasov. "Phase-space metric for non-Hamiltonian systems". *J. Phys. A*, **38**, (2005) 2145.

[23] G. S. Ezra. "On the statistical mechanics of non-Hamiltonian systems: the generalized Liouville equation, entropy, and time-dependent metrics". *J. Math. Chem.*, **35**, (2004) 29.

[24] G. J. Martyna, M. L. Klein, and M. Tuckerman. "Nosé–Hoover chains: The canonical ensemble via continuous dynamics". *J. Chem. Phys.*, **97**, (1992) 2635.

Chapter 4

The Jarzynski identity derived from general Hamiltonian or non-Hamiltonian dynamics reproducing NVT or NPT ensembles

Summary

The Jarzynski identity (JI) relates nonequilibrium work averages to thermodynamic free energy differences. It was shown in a recent contribution [Cuendet, Phys. Rev. Lett. 96, 120602 (2006)] that the JI can, in particular, be derived directly from the Nosé-Hoover thermostated dynamics. This statistical mechanical derivation is particularly relevant in the framework of molecular dynamics simulation, because it is based solely on the equations of motion considered and is free of any additional assumption on system size or bath coupling. Here, this result is generalized to a variety of dynamics, along two directions. On the one hand, specific improved thermostating schemes used in practical applications are treated. These include Nosé-Hoover chains, higher moment thermostats, as well as an isothermic-isobaric scheme yielding the JI in the NPT ensemble. On the other hand, the theoretical generality of the new derivation is explored. Generic dynamics with arbitrary coupling terms and an arbitrary number of thermostating variables, both non-Hamiltonian and Hamiltonian, are shown to imply the JI. In particular, a nonautonomous formulation of the generalized Nosé-Poincaré thermostat is proposed. Finally, general conditions required for the JI derivation are briefly discussed.

4.1 Introduction

In this chapter, we explore the validity domain of the statistical mechanical derivation of the JI, along the lines proposed in Chapter 3. This derivation provides direct evidence that the JI can be used to compute free energy differences for systems of finite numbers of particles, evolved by integration of the specific thermostated dynamical equations considered. We look at two groups of dynamics. First we consider non-Hamiltonian thermostating, with different efficient extensions of the NH scheme of practical use, as well as a generic theoretical formulation. As an important addition to this first group, the case of barostated dynamics generating the NPT ensemble is studied. The second group is represented by a general class of Hamiltonian thermostats, also relevant to molecular dynamics (MD) applications. A brief review of each dynamics considered follows, and a summary can be found in table 4.1.

In the first place, a detailed derivation is provided for the widely used Nosé-Hoover chain thermostat [1] (NHC). In the NHC dynamics, the original NH variable is coupled to a second thermostat, which is the first one of a chain of additional thermostats, each regulating the temperature of the previous one. The NHC is an improvement of the original NH thermostat because it is formally proven to provide canonical sampling even in difficult systems with few degrees of freedom [2]. The NHC also improves thermalization, and eliminates spurious temperature modulations which can appear with NH (Toda demon [3]). The JI derivation obtained for the NHC turns out to hold almost unchanged for two further types of extensions of the NH scheme. The first type of extension aims at controlling higher moments of the physical momenta distribution, which proved to yield more effective thermostating in various applications. The kinetic moments method (KMM) of Hoover and Holian [4] and the generalized Gaussian moments thermostat (GGMT) of Liu and Tuckerman [5] are considered. The second type of extension involves arbitrary generic phase functions as coupling terms. Among others [6, 7], the generalized formulation of Bulgac and Kusnezov [8] (BK) models an infinity of possible couplings of both positions or momenta to the temperature bath, all providing canonical sampling. The BK case is interesting, because it approaches the most general representation of the NH family of thermostats (without chains) retaining all characteristics necessary for the JI derivation to hold.

Second, we address the case of the isothermal-isobaric (NPT) ensemble, which represents the conditions under which most experiments are carried out, and is therefore often used in computer simulations. The NPT ensemble is generated by allowing the system volume to fluctuate, and introducing it as an additional dynamical variable in the equations of motion [9]. A number of time-reversible methods based on the NH approach have been proposed to simulate the NPT ensemble, such as for example the Parrinello-Rahman [10], or Martyna-Tobias-Klein [11] dynamics. Although similar in essence, they vary in several aspects: the method to scale the coordinates and to calculate the internal pressure (by atomic versus molecular virial [12]), the precise way volume is coupled to pressure, or the addition of thermostat or barostat chains. The present derivation focuses on the thermo-barostat proposed by Kalibaeva, Ferrario and Ciccotti [13] (KFC). It is based on the molecular virial, and bears a relatively compact notation. Chains of thermostats or barostats can be incorporated into the KFC system, but these are not necessary to repro-

Table 4.1: *Summary of dynamical systems for which the Jarzynski identity is derived. Abbreviations are explained in the text, and references to the original publications are given as superscripts.*

Non-Hamiltonian	*Hamiltonian*
NHC [1] Nosé-Hoover chains	
KMM [4] / **GGMT** [5] higher moment thermostating	
BK [8, 14] generalized Nosé-Hoover	**GNP** [15] generalized Nosé-Poincaré
KFC [13] NPT ensemble (with $\sum f = 0$)	

duce the NPT ensemble exactly even when an extra conservation law is imposed. This property is exploited here: the additional constraint $\sum_i f_i = 0$ is imposed, in order to show that the JI derivation generalizes to cases with more than one conserved quantity.

The third type of dynamics considered is the so called Nosé-Poincaré (NP) scheme. It differs in essence from the previous cases, because the extended dynamics remains Hamiltonian. The Hamiltonian character has the advantage that symplectic numerical integrators can be designed, which are shown to provide stability, in addition to time-reversibility and phase space volume conservation. The original Nosé [16] thermostat is actually already Hamiltonian, but it samples the physical variables at uneven time intervals, which is a serious drawback for the study of dynamical properties. Dettmann and Morriss [17] were the first to devise a Hamiltonian which reproduces NH trajectories with even time sampling. Independently, Bond *et al.* [18] found that a Poincaré transformation allows to rescale the time in the Nosé dynamics while preserving the Hamiltonian structure. The dynamics resulting from both approaches turn out to be equivalent, and form the NP thermostat. This name was coined by Bond *et al.* [18] who also provided the first extensive study including numerical validation.

The case of nonautonomous NP dynamics has not yet been covered in the literature. Here, an interesting link is made to a result of Zare and Szebehely [19] recently rediscovered by Struckmeier [20], which shows that the Poincaré transformation allows to map any nonautonomous Hamiltonian to an autonomous one. This property of the Poincaré transformation imbricates well with the results of Dettmann and Morriss [17] and Bond *et al.* [18]. Applied to the nonautonomous Nosé Hamiltonian, the general formalism of extended Hamiltonian mappings provides a useful constant of the motion, and places the NP thermostat in a larger theoretical framework. The NP thermostat is in fact a particular case of the more generic Generalized Nosé-Poincaré [15] (GNP) family of thermostats. The thermostat momenta term in the GNP Hamiltonian takes an arbitrary form and includes an arbitrary number of additional thermostating variables, without altering the properties of NP. These generalizations account for a whole class of Hamiltonian temperature couplings, including chains of Nosé thermostats [21], similar to the NHC. Although ther-

mostats of the NP family are not used as commonly as those of the NH family, they have been thoroughly studied [22], and might become a method of choice for MD simulations.

Brief review of the theory of non-Hamiltonian dynamics

We give here a summary of formalism of Tuckerman *et al.* [23, 2], which was described more in detail in Chapter 3. We consider a physical system modeled by a set of positions r and momenta p in contact with a heat bath. The condensed notation $x = (r, p)$ represents the physical variables, which are complemented by a set of thermostating variables Γ^{thermo}, such that the phase vector is $\Gamma = (x, \Gamma^{\text{thermo}})$. Consider the general dynamical system

$$\dot{\Gamma} = \dot{\Gamma}(\Gamma, t). \tag{4.1}$$

If we assume that the physical system has an energy function $H(x)$ which is conserved in the autonomous case, the work accumulated at time τ along a nonequilibrium trajectory is defined as

$$W_\tau(\Gamma_0) = \int_0^\tau dt \frac{\partial H}{\partial t}(\varphi_t(\Gamma_0), t) \ . \tag{4.2}$$

Here we express the evolution of $x(t)$ with a restriction of the flow ϕ of the form $x(t) = \varphi_t(\Gamma_0)$, with Γ_0 the initial condition of the extended system. The extended initial condition Γ_0 has to be given as argument, since x is coupled to the bath.

Tuckerman *et al.* [2] propose a general procedure for constructing the partition function generated by a non-Hamiltonian system satisfying the conditions mentioned above:

1. Determine all conservation laws $\{c_j(\Gamma, t) = C_j\}$ satisfied by the equations of motion (4.1).

2. Using the conservation laws and the equation of motion, identify and eliminate linearly dependent variables, as well as driven or uncoupled variables. Construct the microcanonical phase space density $f(\Gamma, t)$ according to

$$f(\Gamma, t) = \prod_{j=1}^{n_c} \delta\left[c_j(\Gamma, t) - C_j\right]. \tag{4.3}$$

3. Calculate the phase space compression factor

$$\kappa(\Gamma, t) = \nabla_\Gamma \cdot \dot{\Gamma}(\Gamma, t), \tag{4.4}$$

and the metric factor

$$\sqrt{g(\Gamma, t)} = \exp\left\{-\int dt \ \kappa\left(\Gamma(t), t\right)\right\}. \tag{4.5}$$

Generate the invariant volume element

$$\sqrt{g(\Gamma, t)} d\Gamma = \sqrt{g(\Gamma_0, 0)} d\Gamma_0. \tag{4.6}$$

4. Construct the partition function at time t using the definition (4.6) of the phase space measure and the phase space density (4.3),

$$\Omega_t(C_1, ..., C_{n_c}) \propto \int d\Gamma \sqrt{g(\Gamma, t)} \prod_{j=1}^{n_c} \delta\left[c_j(\Gamma, t) - C_j\right]. \tag{4.7}$$

In this framework, the average of any instantaneous observable $A(\Gamma)$ over the ensemble at time t is given by

$$\langle A \rangle_t = \frac{1}{\Omega_t} \int d\Gamma \; A(\Gamma) f(\Gamma, t) \sqrt{g(\Gamma, t)}. \tag{4.8}$$

4.2 The NVT ensemble : the Nosé-Hoover chain and generalized thermostats

Consider now a system of N particles in 3 dimensions with Cartesian positions $r = (r_1, ..., r_N)$, momenta $p = (p_1, ..., p_N)$, and masses $\{m_1, ..., m_N\}$. The canonical ensemble is described by the partition function

$$Z(N, V, T) = \int dr dp \; e^{-\beta H(r,p)} \; , \tag{4.9}$$

where $H(r, p)$ is the Hamiltonian of the physical system, and the constant prefactor has been omitted. The NHC scheme couples the physical positions and momenta (r, p) to the thermostating momenta $p_\eta = (p_{\eta_1}, ..., p_{\eta_M})$ and the associated variables $\eta = (\eta_1, ..., \eta_M)$, for a chain of M thermostats. While r and p are $(3N)$-vectors, η_k and p_{η_k} are real numbers. The extended phase can be expressed as $\Gamma = (r, p, \eta, p_\eta)$. The system evolution is represented by the non-Hamiltonian dynamical system

$$\begin{aligned}
\dot{r}_i &= \frac{p_i}{m_i}, \qquad i = 1, ..., N \\
\dot{p}_i &= -\frac{\partial \Phi}{\partial r_i}(r, t) - \frac{p_{\eta_1}}{Q_1} p_i \\
\dot{\eta}_k &= \frac{p_{\eta_k}}{Q_k}, \qquad k = 1, ..., M \\
\dot{p}_{\eta_1} &= \sum_i \frac{p_i^2}{m_i} - N_{df} k_B T - \frac{p_{\eta_2}}{Q_2} p_{\eta_1} \\
\dot{p}_{\eta_k} &= \frac{p_{\eta_{k-1}}^2}{Q_{k-1}} - k_B T - \frac{p_{\eta_{k+1}}}{Q_{k+1}} p_{\eta_k} \; , \qquad k = 2, ..., M-1 \\
\dot{p}_{\eta_M} &= \frac{p_{\eta_{M-1}}^2}{Q_{M-1}} - k_B T \; .
\end{aligned} \tag{4.10}$$

N_{df} is the number of degrees of freedom of the system. An optimal choice [1] for the thermostat masses is $Q_1 = N_{df} k_B T \tau_{NH}^2$ and $Q_k = k_B T \tau_{\rm NH}^2$ for $k = 2, ..., M$. The parameter $\tau_{\rm NH}$ represents a typical timescale for the system under study. The NH thermostat is recovered for $M = 1$. It is apparent that the evolution of (r, p, p_η) is not coupled to any of

the η_k, which are consequently *noncoupling* variables. This is an important prerequisite for the following derivation. In the following, the η_k will appear in a linear combination,

$$\bar{\eta} = N_{df}\eta_1 + \sum_{k=2}^{M} \eta_k \ .$$

The compressibility of equations (4.10), calculated according to (4.4), is

$$\kappa(\Gamma) = -N_{df}\frac{p_{\eta_1}}{Q_1} - \sum_{k=2}^{M} \frac{p_{\eta_k}}{Q_k} = -\ \dot{\bar{\eta}} \ .$$

Thus, according to (4.5) the metric factor becomes

$$\sqrt{g(\Gamma,t)} = e^{\bar{\eta}}. \tag{4.11}$$

It turns out that the phase space compression does not directly depend on any explicit time variation of the potential, but rather on the state of the thermostat itself, as described by $\bar{\eta}$. Note that instantaneous compression exists even in unperturbed systems, while in this case the time average of κ vanishes.

At this point, it is useful to recall a couple of properties of the autonomous NHC dynamics, i.e. with the explicit time-dependence of the potential removed, $\Phi(r) = \Phi(r,0)$ (equilibrium conditions). The second and key property of the autonomous NHC thermostat is that the microcanonical distribution in the extended phase space projects to canonically distributed variables (r,p) in the physical phase space [1], with the usual partition function (4.9). As explained in Ref. [2], one of the advantages of the NHC thermostat is that this property is retained if there are conserved quantities other than H', such as the center of mass momentum ($\sum_i f_i = 0$). This is not the case for the simple NH thermostat ($M = 1$). The second property is that the pseudo-Hamiltonian

$$H'(\Gamma) = \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \Phi(r) + \sum_{k=1}^{M} \frac{p_{\eta_k}^2}{2Q_k} + k_B T\bar{\eta} \tag{4.12}$$

is an invariant of the motion [24, 2]. Assuming that $H'(\Gamma)$ is the only conserved quantity, it is apparent in (4.10), (4.11) and (4.12), that the η_k interact[1] with the dynamics only through their linear combination $\bar{\eta}$.

In the following, we are going to calculate the Helmholtz free energy difference $\Delta F_{AB} = F_B - F_A$ between two states A and B of the physical system, described by the two Hamiltonians $H_A(x)$ and $H_B(x)$, and the corresponding canonical partition functions Z_A and Z_B of the form 4.9. The equilibrium free energy difference is given by

$$e^{-\beta\Delta F_{AB}} = \frac{\int e^{-\beta H_B(x)}dx}{\int e^{-\beta H_A(x)}dx} = \frac{Z_B}{Z_A} \ . \tag{4.13}$$

[1]Following Ref. [2], this means that only one of the η_k is independent, while the others are driven. If there were a second conserved quantity (as in Section 4.3), all but two of the η_k would be driven. Note that this property is different from the *noncoupling* defined above.

It is assumed that $H_A(x)$ can be transformed into $H_B(x)$ by a continuous variation of the potential $\Phi(r,t)$. The nonequilibrium Hamiltonian $H(x,t)$ is initially equivalent to $H_A(x)$, and varies such that $H(x,\tau) = H_B(x)$ at final time τ. The system is considered at equilibrium before $t = 0$, so the partition function (4.9) describes correctly the system at $t = 0$, corresponding to an equilibrium situation in state A, with a fixed potential energy function $\Phi(r,0)$.

In the case of a nonautonomous potential $\Phi(r,t)$ driving the system out of equilibrium, expression (4.12) of the NHC pseudo-Hamiltonian can be generalized by replacing $\Phi(r)$ with $\Phi(r,t)$, and denoting it $H'(\Gamma,t)$. Of course, $H'(\Gamma,t)$ is not invariant anymore. However, by computing explicitly its time derivative, it can be shown that the following quantity is an invariant of the motion,

$$\begin{aligned} H''(\Gamma,t) &= \sum_{i=1}^{N} \frac{p_i^2}{2m_i} + \Phi(r,t) + \sum_{k=1}^{M} \frac{p_{\eta_k}^2}{2Q_k} + k_B T \bar{\eta} \\ &\quad - \int_0^t \frac{\partial \Phi}{\partial t'}(r(t'),t')dt' \\ &= H'(\Gamma,t) - W_t(\Gamma_0) = C \ . \end{aligned} \tag{4.14}$$

The definition (4.2) of the work was used. As we have seen, W_τ depends on the path $\{x(t)\}_{t\in[0,\tau]}$ only. Since the η_k are all *noncoupling* variables in (4.10) and the dynamics are deterministic, giving a point (x, p_η) at any time t is sufficient to specify the path along which W_τ is accumulated. Therefore, W_τ can be expressed as a function of x and p_η only, and is independent of $\bar{\eta}$. Formally, $W_\tau(\Gamma_0)$ can take as argument the inverse $\varphi_\tau^{-1}(x,p_\eta)$ of the reduced flow φ and the shorthand notation $W_\tau(x,p_\eta) \equiv W_\tau(\varphi_\tau^{-1}(x,p_\eta))$ will be used. Noting that $W_0(\Gamma_0) = 0$, the conservation law (4.14) can be expressed with only quantities at τ on the right hand side,

$$H'(\Gamma_0,0) = H'(\Gamma,\tau) - W_\tau(x,p_\eta) \ . \tag{4.15}$$

4.2.1 Derivation of the Jarzynski identity for NHC

Let us perform the following experiment. The system is prepared in an equilibrium state with the initial Hamiltonian $H_A(\Gamma)$. Starting from different initial conditions Γ_0 taken from this equilibrium ensemble, we change the Hamiltonian from $H(\Gamma,0) = H_A(\Gamma)$ to $H(\Gamma,\tau) = H_B(\Gamma)$. The JI derivation closely follows Ref. [25] and starts by writing the mathematical object $\langle e^{-\beta W_\tau(\Gamma_0)}\rangle_{\Gamma_0}$ according to (4.8),

$$\frac{1}{\Omega_A}\int d\Gamma_0 \sqrt{g(\Gamma_0,0)} e^{-\beta W_\tau(\Gamma_0)} \delta\left[H'(\Gamma_0,0) - C\right].$$

This is a correct ensemble average for Γ_0. We now perform a change of variables $\Gamma_0 \to \Gamma_\tau = \phi_\tau(\Gamma_0)$, with $\Gamma_0 = \phi_\tau^{-1}(\Gamma_\tau)$. We know from (4.6) that the phase space measure is invariant, which yields

$$\frac{1}{\Omega_A}\int d\Gamma_\tau \sqrt{g(\Gamma_\tau,\tau)} e^{-\beta W_\tau(\phi_\tau^{-1}(\Gamma_\tau))} \delta\left[H'(\phi_\tau^{-1}(\Gamma_\tau),0) - C\right].$$

Since from now on all phase variables are expressed at time τ, the τ index is dropped to lighten the notation. As previously discussed, $W_\tau(\phi_\tau^{-1}(\Gamma_\tau))$ can be expressed as $W_\tau(x, p_\eta)$. Using the conservation law (4.15) of the pseudo-Hamiltonian, as well as the explicit form (4.11) of the metric factor, we write out the integral over the different components of Γ,

$$\frac{1}{\Omega_A}\int dx dp_\eta e^{-\beta W_\tau(x,p_\eta)}\int d\bar{\eta} e^{\bar{\eta}}\delta\left[H'(\Gamma,\tau)-W_\tau(x,p_\eta)-C\right].$$

A property of the Dirac δ-function is going to be useful several times in the following: for any smooth function $\ell(\eta)$ with one simple root $\eta_\star$,

$$\delta[\ell(\eta)] = \delta[\eta-\eta_\star]/|\ell'(\eta_\star)|\ . \tag{4.16}$$

Using this and remembering that $H'(\Gamma,\tau) = H(x,\tau) + \sum_k \frac{p_{\eta_k}^2}{2Q} + \frac{1}{\beta}\bar{\eta}$, we integrate over $\bar{\eta}$,

$$\begin{aligned}\frac{\beta}{\Omega_A}\int dx dp_\eta e^{-\beta W_\tau(x,p_\eta)} e^{-\beta\left[H(x,\tau)+\sum\limits_k \frac{p_{\eta_k}^2}{2Q}-W_\tau(x,p_\eta)-C\right]}\\ = \frac{\beta D e^{\beta C}}{\Omega_A}\int dx e^{-\beta H(x,\tau)}.\end{aligned}$$

The work terms cancel out in the upper equation. The contributions of the p_{η_k} are Gaussians, which integrate out in the constant D. The last integral corresponds to the equilibrium partition function Z_B at state B. Because the same prefactor arises in the denominator when deriving the physical equilibrium partition function Z_A from the extended partition function Ω_A [2], we finally have the result

$$\langle e^{-\beta W_\tau(\Gamma_0)}\rangle_{\Gamma_0} = \frac{Z_B}{Z_A} = e^{-\beta\Delta F_{AB}}\ .$$

4.2.2 Higher moments thermostating

The idea of regulating higher moments of the kinetic energy was proposed by Hoover [26, 27] already few years after the publication of the original Nosé thermostat. However, no application was reported until Hoover and Holian's kinetic moments method [4] (KMM) in 1996, which was shown to allow improved mixing for stiff low-dimensional systems and better control of kinetic temperature fluctuations in nonequilibrium simulations. Whereas the original publications give condensed equations in reduced units[2], the KMM dynamics are considered here with explicit parameters such as particle number, masses, temperature and k_B. The equations are also adapted to accommodate any combination of arbitrarily

[2]Note that in Hoover's 1989 comment [27], the denominators in the equations for the 4th and 6th moments contain wrong powers of (mkT). A dimensional analysis shows that the exponents should be 3 and 5 instead of 2 and 3, respectively. Furthermore, for the conservation of the pseudo-Hamiltonian, the pseudo-masses Q_k have to appear in the friction term of the $\dot{p}$ equation. The same applies to N_{df} and the $(2k-1)$ term which have to appear in the $\dot{p}_\eta$ equation.

high moments $\{1...M\}$, and the notation $S = \sum_{i=1}^{N} \frac{p_i^2}{m_i}$ is used,

$$\begin{aligned}
\dot{r}_i &= \frac{p_i}{m_i}, \qquad i = 1, ..., N \\
\dot{p}_i &= -\frac{\partial \Phi}{\partial r_i}(r,t) - \sum_{k=1}^{M} \frac{p_{\eta_k}}{Q_k} S^{k-1} p_i \\
\dot{p}_{\eta_k} &= S^k - (2k-1) k_B T N_{df} S^{k-1}, \qquad k = 1, ..., M \\
\dot{\eta}_k &= (2k-1) \frac{p_{\eta_k}}{Q_k} S^{k-1}
\end{aligned}$$

The dynamics of the $2M$ extended variables is designed to control the fluctuations in the even moments of the Gaussian distribution in momentum. Note that the $\dot{\eta}_k$ equation is not present in the KMM paper. Its form is chosen such that, according to (4.4), the phase space compression factor is

$$\kappa(\Gamma) = -\sum_{k=1}^{M} (2k-1) \frac{p_{\eta_k}}{Q_k} S^{k-1} = -\sum_{k=1}^{M} \dot{\eta}_k \ .$$

If the thermostat collective variable is redefined as $\bar{\eta} = \sum_{k=1}^{M} \eta_k$, it turns out that the metric factor takes the same form (4.11) as in the NHC case, $\sqrt{g(\Gamma,t)} = e^{\bar{\eta}}$. This definition of η_k implies in addition that the NHC pseudo-Hamiltonian H' as defined in (4.12) is an invariant of the motion of the autonomous KMM dynamics. Similarly, H'' from (4.14) is also an invariant for the nonautonomous version of KMM. This is easily seen by calculating the time derivative of H'' explicitly. So, both the KMM metric factor and invariant of the motion have the same form as in NHC. In addition, the η_k are also *noncoupling*, such that $W_\tau(\Gamma_0) = W_\tau(x, p_\eta)$. Based on these common starting points, the derivation developed above for the NHC thermostat applies here as well, and the JI follows from the KMM dynamics.

As already pointed out by Hoover [4], the KMM scheme becomes unstable in practice when control of higher than the lowest two moments is attempted. This is due to the fact that high powers of the kinetic energy fluctuate about the *instantaneous* values of lower powers. Liu and Tuckerman [5] proposed the generalized Gaussian moments method (GGMT), which allows stable control of higher kinetic energy moments. The GGMT was shown to lead to improved convergence of path integrals, better energy equipartitioning and spatial sampling, as well as superior performance in nonequilibrium simulations with respect to the NHC. The GGMT also bears reversible integrators based on a Trotter-type decomposition. This makes it a good candidate for demanding MD applications. The equations are as follows:

$$\begin{aligned}
\dot{r}_i &= \frac{p_i}{m_i}, \qquad i = 1, ..., N \\
\dot{p}_i &= -\frac{\partial \Phi}{\partial r_i}(r,t) - \sum_{k=1}^{M} \sum_{j=1}^{k} \frac{p_{\eta_k}}{Q_k} \frac{S^{j-1}}{\beta^{k-j} C_{j-1}} p_i,
\end{aligned}$$

$$\begin{aligned} \dot{p}_{\eta_k} &= \frac{S^k}{C_{k-1}} - \frac{N_{df}}{\beta^k}, \qquad k = 1, ..., M, \\ \dot{\eta}_k &= \left[\frac{1}{\beta^{k-1}} + \sum_{j=2}^{k} \frac{S^{j-1}}{\beta^{k-j} dN C_{j-2}} \right] \frac{p_{\eta_k}}{Q_k}. \end{aligned} \tag{4.17}$$

The constants C_k are defined as $C_k = \prod_{j=1}^{k}(dN + 2j)$, with $C_0 = 1$. An optimal choice for the thermostat masses Q_k is suggested in Ref. [5]. If we set $\bar{\eta} = N_{df} \sum_{k=1}^{M} \eta_k$, the metric factor takes again the same form (4.11) as in the NHC, and the invariant of the motion identifies with (4.14). Furthermore, W_τ is independent of the η_k, and from this point the NHC derivation of the JI applies to the GGMT as well.

4.2.3 Generic form of the Nosé-Hoover scheme

From the theoretical point of view, it is interesting to investigate the degree of generality of the present JI derivation. In other words, investigate what minimal set of necessary characteristics the thermostating dynamics must have for the derivation to hold. Bulgac and Kusnezov [8, 14] (BK) proposed very generic dynamical equations which provide a canonical distribution based on the NH principle. BK use two thermostating variables, one coupled to positions r and the other coupled to physical momenta p. Here, a slightly generalized version of the formalism is proposed, in which a set of M extended momenta p_{η_k}, $k = 1...M$, is considered. The first M_r extended momenta act on r, while the remaining $(M - M_r)$ are coupled to p. This generalization has no impact on the construction of the BK scheme, which relies on four hypotheses. First, a canonical phase density is imposed for (x, p_η),

$$\rho(x, p_\eta, t) = e^{-\beta \left[H(x,t) + \sum_{k=1}^{M} \frac{h_k(p_{\eta_k})}{C_k} \right]}.$$

Here, the $h_k(p_{\eta_k})$ are arbitrary functions[3] such that $e^{-\beta h_k}$ is in L_1, i.e. its integral is finite. The C_k are arbitrary constants. The second hypothesis requires that the density ρ be a stationary point of the Liouville equation. The third hypothesis is that $\frac{\partial}{\partial p_{\eta_k}} \dot{p}_{\eta_k} = 0, \forall k$. The fourth hypothesis relies on the form of the coupled equations for $\dot{r}$ and $\dot{p}$,

$$\begin{aligned} \dot{r}_i &= \frac{\partial H}{\partial p_i} - \sum_{k=1}^{M_r} h'_k(p_{\eta_k})\, G_{ik}(x)\ , \quad i = 1, ..., N \\ \dot{p}_i &= -\frac{\partial H}{\partial r_i} - \sum_{k=M_r+1}^{M} h'_k(p_{\eta_k})\, G_{(N+i)k}(x)\ . \end{aligned}$$

The two first hypotheses above imply that $h'_k \equiv \frac{\mathrm{d}h_k}{\mathrm{d}p_{\eta_k}}$ appears in the coupling terms, whereas the G_{ik} remain arbitrary functions of x. BK showed that starting from these two

[3] In order to keep the notations consistent throughout the present paper, symbols different from BK [8, 14] had to be chosen. The equations of BK are recovered with the following mapping: $h_k \to g_k$; $\frac{\mathrm{d}h_k}{\mathrm{d}p_{\eta_k}} \to h_k$; $p_\eta \to (\xi, \zeta)$; $G_i \to (F_i, G_i)$; $C_k \to (\alpha, \beta)$.

equations, the $\dot{p}_{\eta_k}$ equation follows,

$$\begin{aligned}\dot{p}_{\eta_k} &= C_k \sum_{i=1}^{2N} \left[\frac{\partial H}{\partial x_i} G_{ik}(x) - k_B T \frac{\partial G_{ik}}{\partial x_i}(x) \right], \\ \dot{\eta}_k &= \sum_{i=1}^{2N} h'_k(p_{\eta_k}) \, \frac{\partial G_{ik}}{\partial x_i}(x) \ , \quad k = 1, ..., M\end{aligned}$$

Here, the original BK dynamics is completed with the equation for $\dot{\eta}_k$, inferred from the form of the motion invariant and of the phase space compression factor. The phase vector is again $\Gamma = (x, p_\eta, \eta)$. The explicit time dependence of the potential has no impact on the form of the motion equations or the thermostating property. If we define $\bar{\eta} = \sum_{k=1}^{M} \eta_k$, the nonautonomous BK conserved quantity takes the form

$$H''(\Gamma, t) = H(x, t) + \sum_{k=1}^{M} \frac{h_k(p_{\eta_k})}{C_k} + k_B T \bar{\eta} - W_t(x, p_\eta).$$

According to (4.4), the phase space compression factor is

$$\kappa(\Gamma) = -\sum_{i=1}^{2N} \sum_{k=1}^{M} h'_k(p_{\eta_k}) \, \frac{\partial G_{ik}}{\partial x_i}(x) = -\dot{\bar{\eta}} \ ,$$

and we find again the metric factor $e^{\bar{\eta}}$. Furthermore, the η_k are *noncoupling* variables, and the work can again be written $W_t(x, p_\eta)$. Despite the generic character of the BK equations, the ingredients for the JI derivation, $\sqrt{g(\Gamma, t)}$, $H''(\Gamma, t)$ and $W_t(x, p_\eta)$ retain the same form as in the NHC case. The only difference is the p_{η_k} term in $H''(\Gamma, t)$, which is now expressed in terms of h_k. But h_k is still independent of η_k, and by construction integrates to a constant when placed in a negative exponentiation. Finally, the NHC JI derivation can be reused once more for the BK generic dynamics.

The BK equations do not say anything about how one should thermostate most efficiently a given system for a given purpose. However, the result is interesting because the BK dynamics cover the simple NH and both higher moments methods, KMM and GGMT, as special cases. For example, the original NH is obtained by keeping only one coupling term in the $\dot{p}$ equation and setting $h = \frac{p_\eta^2}{2}$ and $G_i = p_i$. Other special cases include for instance the thermostat of Brańka *et al.* [28]. Their method resembles NH, except for the coupling term, which takes the form $-p_i \, p_{\eta_k}^{2n-1}/Q_k$, with n an arbitrary integer. With respect to NH, this amounts to changing h to $\frac{p_\eta^{2n}}{2n}$. A further interesting example is the configurational temperature thermostat proposed by Braga and Travis [29], aimed at controlling better the thermodynamic temperature, rather than the kinetic temperature (they differ far from equilibrium). The temperature is estimated from derivatives of the potential energy, and the dynamics uses one r-dependent coupling term in the $\dot{r}$ equation, recovered by setting $h = \frac{p_\eta^2}{2}$ and $G_i = \frac{\partial \Phi}{\partial r_i}$ in the BK scheme.

Note however that all chained methods such as NHC are no special cases of BK, because the requirement that $\frac{\partial}{\partial p_{\eta_k}} \dot{p}_{\eta_k} = 0 \ \forall k$ is not met. The KFC thermo-barostat presented in the

next section has also a feature similar to chaining in the $\dot{p}_V$ equation, and consequently does not belong to the BK family. It would however be possible [1] and interesting to construct a more generic representation of the method which would allow chained schemes as special cases.

4.3 The NPT ensemble : the KFC thermo-barostat

In the KFC formalism [13], the absolute position and momentum of atom i with mass m_i belonging to molecule μ with total mass M_μ are given by $r_{\mu,i}$ and $p_{\mu,i}$. The position and momentum of the center of mass of each molecule are denoted by R_μ and P_μ. V is the volume of the simulation cell, and Π is the isotropic internal molecule-based pressure,

$$\Pi = \frac{1}{3V}\sum_\mu \left[\frac{P_\mu^2}{M_\mu} + f_\mu \cdot R_\mu\right] + \frac{\partial \Phi}{\partial V}(r,V,t) \ .$$

The force f_μ on molecule μ is derived from the potential $\Phi(r,V,t)$. Note that $\Phi(r,V,t)$ has an explicit dependence on the volume, necessary to account for long range interactions in periodic systems [12]. To study the case $\sum_i f_i = 0$, we define the new phase space variables $\mathbf{R} = \frac{1}{M}\sum_{\mu,i} m_{\mu,i} r_{\mu,i}$ as the center of mass position, and $\mathbf{P} = \sum_{\mu,i} p_{\mu,i}$ as the total momentum of the system, with $M = \sum_{\mu,i} m_{\mu,i}$. We take the relative positions $r'_{\mu,i} = r_{\mu,i} - \mathbf{R}$ and the relative momenta $p'_{\mu,i} = p_{\mu,i} - \mathbf{P}$ as a new coordinate system. Then $\mathbf{R}$ is a driven variable, and two components of $\mathbf{P}$ are linearly dependent, such that only its modulus P is kept as an independent variable. The corresponding partition function for an external or reference pressure Π_{ext} is, up to a constant,

$$Z(N,\Pi_{\rm ext},T) = \int dP dV dr' dp' P^2 V \ e^{-\beta[H(r',p',V,t)+\Pi_{\rm ext}V]} \ . \tag{4.18}$$

To describe the dynamics of this system, we need $6(N-1)$ equations for p' and r' and one for the modulus of P. The system of KFC employs 6 additional dynamical variables : the volume V and the associated momentum p_V, the temperature coupling variables η and p_η, and the volume coupling variables ξ and p_ξ. To each of these additional momenta is associated a pseudo mass, Q_V, Q_η, Q_ξ. The resulting phase vector is $\Gamma = (r', p', P, V, p_V, \eta, p_\eta, \xi, p_\xi)$, and the equations of motion are:

$$\begin{aligned}
\dot{r'}_{\mu,i} &= \frac{p'_{\mu,i}}{m_{\mu,i}} + R'_\mu \frac{p_V}{3VQ_V}, \qquad i = 1,...,N \\
\dot{p'}_{\mu,i} &= -\frac{\partial \Phi}{\partial r'_{\mu,i}}(r',V,t) - p'_{\mu,i}\frac{p_\eta}{Q_\eta} - \frac{m_{\mu,i}}{M_\mu} P'_\mu \frac{p_V}{3VQ_V} \\
\dot{P} &= -P\frac{p_\eta}{Q_\eta} - P\frac{p_V}{3VQ_V} \\
\dot{V} &= \frac{p_V}{Q_V} \\
\dot{p}_V &= [\Pi - \Pi_{\rm ext}] - p_V \frac{p_\xi}{Q_\xi}
\end{aligned} \tag{4.19}$$

$$\begin{aligned}
\dot{\eta} &= \frac{p_\eta}{Q_\eta} \\
\dot{p}_\eta &= \sum_{\mu,i} \frac{{p'}^2_{\mu,i}}{m_{\mu,i}} - Lk_BT \\
\dot{\xi} &= \frac{p_\xi}{Q_\xi} \\
\dot{p}_\xi &= \frac{p_V^2}{k_BT} \ .
\end{aligned}$$

The constant L is determined at a later point. This system has two invariants of the motion [13]. The first invariant is the pseudo Hamiltonian, which, in the autonomous case, is similar to (4.12),

$$\begin{aligned}
\tilde{H} &= \sum_{\mu,i} \frac{{p'}^2_{\mu,i}}{m_{\mu,i}} + \frac{P^2}{2M} + \Phi(r', V) + \Pi_{\text{ext}} V + \\
&\quad + \frac{p_V^2}{2Q_V} + \frac{p_\eta^2}{2Q_\eta} + \frac{p_\xi^2}{2Q_\xi} + Lk_BT\eta + k_BT\xi \ .
\end{aligned}$$

In the case of an explicit time-dependence of the potential $\Phi(r', V, t)$, the direct time differentiation of $\tilde{H}$ gives $\frac{d\tilde{H}}{dt} = \frac{\partial \Phi}{\partial t}(r', V, t)$. With the work $W_t(\Gamma_0)$ defined as in (4.2), integrating this expression yields the conservation law

$$\tilde{H}(\Gamma, t) - W_t(\Gamma_0) = C_1 \ . \tag{4.20}$$

The second invariant is derived from the condition $\sum_i f_i = 0$ and is similar to the one pointed out by Nosé [24] in the NVT case. By integrating the $\dot{P}$ equation from the dynamics (4.19), we find

$$\begin{aligned}
P(t) &= C_2 \ \exp\left\{ - \int_0^t dt \left[\frac{p_\eta}{Q_\eta} + \frac{p_V}{3VQ_V} \right] \right\} && (4.21) \\
&= C_2 \ \exp\left\{ - \int_0^t dt \left[\dot{\eta} + \frac{\dot{V}}{3V} \right] \right\} && (4.22) \\
&= C_2 \ e^{-\eta - \frac{1}{3}\ln V} . && (4.23)
\end{aligned}$$

Inverting the last expression yields

$$PV^{1/3}e^{\eta} = C_2 \ . \tag{4.24}$$

The compressibility of equations (4.19) is

$$\kappa(\Gamma) = -(3N-2)\dot{\eta} - \frac{\dot{V}}{3V} - \dot{\xi} \ ,$$

and the associated metric factor is

$$\sqrt{g(\Gamma, t)} = V^{1/3} e^{(3N-2)\eta + \xi} \ . \tag{4.25}$$

4.3.1 Derivation of the Jarzynski identity for KFC

The derivation follows closely the one presented for NHC. We start by writing the equilibrium phase space average $\langle e^{-\beta W_\tau(\Gamma_0)}\rangle_{\Gamma_0}$ at time $t = 0$, using the two invariants of the motion (4.20) and (4.24). For compactness, $\tilde{H}(\Gamma, t)$ is written $\tilde{H}_t$.

$$\frac{1}{\Omega_A}\int d\Gamma_0 \sqrt{g(\Gamma_0,0)} e^{-\beta W_\tau(\Gamma_0)} \delta\left[\tilde{H}_0 - C_1\right] \delta\left[P_0 V_0^{1/3} e^{\eta_0} - C_2\right]$$

$$= \frac{1}{\Omega_A}\int d\Gamma_\tau V_\tau^{1/3} e^{(3N-2)\eta_\tau + \xi_\tau} e^{-\beta W_\tau} \delta\left[\tilde{H}_\tau - W_\tau(\phi_\tau^{-1}(\Gamma_\tau)) - C_1\right] \delta\left[P_\tau V_\tau^{1/3} e^{\eta_\tau} - C_2\right].$$

on the second line, the change of variables $\Gamma_0 \to \Gamma_\tau = \phi_\tau(\Gamma_0)$ has been performed, with $\Gamma_0 = \phi_\tau^{-1}(\Gamma_\tau)$. We have again used the fact that the phase space measure $\sqrt{g(\Gamma,t)}d\Gamma$ is invariant under the flow ϕ, and expressed the compression factor according to (4.25). We have finally used (4.20) to substitute $\tilde{H}(\phi_\tau^{-1}(\Gamma_\tau), 0)$ with $\tilde{H}_\tau - W_\tau(\phi_\tau^{-1}(\Gamma_\tau))$.

Seeing that ξ and η are *noncoupling* variables in (4.19) and following the argumentation leading to (4.15) establishes that the work is independent of ξ and η, and can be expressed as a function of $\hat{\Gamma} = (r', p', P, V, p_V, p_\eta, p_\xi)$ at any time t. This allows to note $W_\tau \equiv W_\tau(\phi_\tau^{-1}(\Gamma_\tau)) \equiv W_\tau(\hat{\Gamma}_\tau)$. In the following, the τ index on phase variables is left out. The next step is the integration over η, applying property (4.16) to the rightmost δ-function,

$$\frac{1}{C_2\Omega_A}\int d\hat{\Gamma}\ V^{1/3}\left[\frac{C_2}{PV^{1/3}}\right]^{3N-2} e^{-\beta W_\tau}\int d\xi\ e^\xi\ \delta\left[\tilde{H}_\tau - W_\tau - C_1\right]$$

$$= \frac{\beta e^{\beta C_1}}{C_2\Omega_A}\int d\hat{\Gamma}\ V^{1/3}\left[\frac{C_2}{PV^{1/3}}\right]^{3N-2-L} e^{-\beta W_\tau}\ e^{-\beta\left[H(r',p',V,\tau) + \frac{p_V^2}{2Q_V} + \frac{p_\eta^2}{2Q_\eta} + \frac{p_\xi^2}{2Q_\xi} + \Pi_{\mathrm{ext}} V - W_\tau\right]}.$$

On the second line, the integration over ξ has been performed, based again on the key fact that W_τ is independent of ξ. The integrals over the Gaussian auxiliary momenta terms are constants. We see again that the two W_τ terms cancel out. Finally, the choice $L = 3N$ gives

$$\frac{C}{\Omega_A}\int dV dr' dp' dP\ V P^2\ e^{-\beta[H(r',p',V,\tau) + \Pi_{\mathrm{ext}} V]}.$$

The V term is equivalent to the integral over the center of mass position, $\int_{D(V)} dR$. All constants gathered in C, which resulted from the above integrations, are the same as in the calculation of the equilibrium partition function alone [13]. We have again obtained the JI.

4.4 NVT Hamiltonian dynamics : the generalized Nosé-Poincaré thermostat

In this section, it is shown that the JI can as well be derived from Hamiltonian thermostated equations of motion. The GNP dynamics is based on the generalized Nosé (GN) Hamiltonian

$$H_{GN}(\Gamma, \tilde{t}) = H(q, \frac{\tilde{p}}{s}, \tilde{t}) + Lk_BT \ln s + h(p_s, \{\sigma_k, \frac{\tilde{p}_{\sigma_k}}{s}\}).$$

The Hamiltonian of the physical system is $H(r, \frac{\tilde{p}}{s}, \tilde{t}) = \frac{\tilde{p}^2}{2ms^2} + \Phi(r, t(\tilde{t}))$, and the extended phase is now $\Gamma = (q, \tilde{p}, s, p_s, \{\sigma_k, \tilde{p}_{\sigma_k}\})$. The variable $s = \frac{d\tilde{t}}{dt}$ is a time scaling factor, with conjugate momentum p_s. The equations of motion for Γ derived from H_{GN} are integrated with respect to the Nosé time $\tilde{t}$. The particle positions q and true momenta $p = \frac{\tilde{p}}{s}$ are canonically distributed, but sampled at irregular intervals of the real time t.

The function h is arbitrary, but must be continuous, bounded below, and such that $e^{-\beta h}$ is in L_1. An arbitrary number of additional thermostating variables $\{\sigma_k, p_{\sigma_k}\},\ k = 1, ..., M$, can be included in h. Note that in the present study, h is defined in a slightly different way than in Laird and Leimkuhler's [15] original paper. Here, $h\left(p_s, \{\sigma_k, p_{\sigma_k}\}\right)$ is a function of the auxiliary momenta $p_{\sigma_k} = \frac{\tilde{p}_{\sigma_k}}{s}$ in real time, whereas in Ref. [15] h is a function of $\tilde{p}_{\sigma_k}$. While this modification makes no difference in the proof that the GNP dynamics generates canonically distributed (q, p), it is necessary for the JI derivation, as will become apparent later. The original Nosé Hamiltonian [16] is recovered for $h = \frac{p_s^2}{2Q_s}$, where Q_s is the pseudo-mass associated with s. Similarly, setting $h = \frac{p_s^{2n}}{2nQ_s}$ with n integer, gives the scheme of Brańka *et al.* [28]. For conciseness all arguments of h are from now on grouped in the collective variable $\Lambda = (p_s, \{\sigma_k, p_{\sigma_k}\})$.

Zare and Szebehely [19] and later Struckmeier [20] showed that any nonautonomous (with respect to time $\tilde{t}$) Hamiltonian $H(Q, P, \tilde{t})$ can be mapped to an extended Hamiltonian $H_1(Q, P, \tilde{t}, \mathcal{H})$ which is autonomous with respect to time t. The original canonical variables (Q, P) are augmented by a couple of conjugate variables $\tilde{t}$ and $-\mathcal{H}$, while the new time t becomes the free parameter. Here $\mathcal{H}$ is the value of H, regarded as a pure function of time t. The mapping is obtained by a Poincaré transformation of the original Hamiltonian, $H_1 = s\left[H - \mathcal{H}\right]$, with $s = \frac{d\tilde{t}}{dt}$. The extended Hamiltonian $H_1(Q, P, \tilde{t}, \mathcal{H}) \equiv 0$ is a conserved quantity along the evolution. Applying this mapping to the GN Hamiltonian with $Q = (r, s, \{\sigma_k\})$ and $P = (\tilde{p}, p_s, \{\tilde{p}_{\sigma_k}\})$ yields

$$H_{GNP}(\Gamma, \tilde{t}, \mathcal{H}) = s\left[H_{GN}(\Gamma, \tilde{t}) - \mathcal{H}\right] \ .$$

H_{GNP} is equivalent to the GNP Hamiltonian proposed by Laird and Leimkuhler [15]. Interestingly, the mapping to an extended Hamiltonian provides two benefits. First, multiplying a zero-valued Hamiltonian results in a scaling of time which relieves the problem of uneven time sampling, while leaving the dynamics intact. This is the core of the NP formalism as found by Dettmann and Morriss [17] and Bond *et al.* [18]. Second, the transformation provides a conserved quantity and two additional equations of motion. The dynamical system for $(\Gamma, \tilde{t}, -\mathcal{H})$ is derived from the GNP Hamiltonian in the usual way. The equations for Γ are almost[4] identical to those of the GNP [15], and are not repeated here. The equations for the additional conjugate variables $(\tilde{t}, -\mathcal{H})$ are

$$\begin{aligned} \frac{\mathrm{d}\tilde{t}}{\mathrm{d}t} &= -\frac{\partial H_{GNP}}{\partial \mathcal{H}} = s \\ \frac{\mathrm{d}\mathcal{H}}{\mathrm{d}t} &= \frac{\partial H_{GNP}}{\partial \tilde{t}} = s\frac{\partial \Phi}{\partial t}\frac{\mathrm{d}t}{\mathrm{d}\tilde{t}} = \frac{\partial \Phi}{\partial t}(r, t) \ . \end{aligned}$$

[4]The difference with respect to Ref. [15] lies in the $\dot{p}_s$ equation, which here has the additional term $-\sum_k \frac{\tilde{p}_{\sigma_k}}{s}\partial_{p_{\sigma_k}} h$. This is due to the slightly different definition of h used.

The first equation is an identity, and the second equation describes the variation of the energy H_{GN}. It can be integrated from time 0 to τ, which gives

$$\mathcal{H}(\tau) = \int_0^\tau \frac{\partial \Phi}{\partial t}(r(t),t)dt + H^0_{GN} = W_\tau(\Gamma_0) + H^0_{GN} \ , \tag{4.26}$$

where H^0_{GN} is the initial value of the GN Hamiltonian. Since $H_{GNP} \equiv 0$ and $s > 0$ strictly, we can write the conservation law

$$H_{GN}(\Gamma,\tau) = H_{GN}(\Gamma_0,0) + W_\tau(\Gamma_0) \ .$$

We see that both $\tilde{t}$ and $\mathcal{H}$ are driven variables which can be left aside from the analysis, and the remaining Hamiltonian is noted $H_{GNP}(\Gamma,t)$. It is informative to write the Γ dynamics in terms of momenta in real time, p and p_{σ_k}, using $\dot{p} = \frac{\dot{\tilde{p}}}{s} - p\frac{\dot{s}}{s}$ and $\dot{p}_{\sigma_k} = \frac{\dot{\tilde{p}}_{\sigma_k}}{s} - p_{\sigma_k}\frac{\dot{s}}{s}$,

$$\begin{aligned}
\dot{q}_i &= \frac{p_i}{m_i} \\
\dot{p}_i &= -\frac{\partial \Phi}{\partial q_i} - p_i \ \partial_{p_s} h \\
\dot{s} &= s \ \partial_{p_s} h \\
\dot{p}_s &= \sum_{i=1}^{N} \frac{p_i^2}{m_i} - g k_B T - \sum_{k=1}^{M} \partial_{p_{\sigma_k}} h - \Delta H_{GN} \\
\dot{\sigma}_k &= \partial_{p_{\sigma_k}} h \\
\dot{p}_{\sigma_k} &= -\partial_{\sigma_k} h - p_{\sigma_k} \partial_{p_s} h .
\end{aligned} \tag{4.27}$$

Here, the compact notation $\partial_x h = \frac{\partial h}{\partial x}(\Lambda)$ is used. The term $\Delta H_{GN} = H_{GN}(q,p,s,\Lambda,t) - \mathcal{H}(t)$ formally vanishes at all times[5], since $H_{GNP} \equiv 0$ and $s > 0$. This implies that p_s as well as $h(\Lambda)$ are independent of s. It is finally apparent that s is a *noncoupling* variable in this representation, which is a key element for the following JI derivation. Note that the original definition of h in Ref. [15] does not allow to evidence this property.

4.4.1 Derivation of the Jarzynski identity for GNP

As the GNP dynamics is Hamiltonian, the phase space is Euclidean, $\sqrt{g} \equiv 1$, and the machinery of Tuckerman *et al.* [2] is not required anymore. The partition function over the extended phase space at time t simply takes the microcanonical form

$$\Omega_t = \int d\Gamma \ \delta\left[H_{GNP}(\Gamma,t)\right] \ .$$

We consider again a system starting at equilibrium with Hamiltonian H_A, and brought slowly to H_B within time τ, via the explicit time-dependence of the potential. Like in the previous sections, we start by writing at $t = 0$ the equilibrium phase space average $\langle e^{-\beta W_\tau(\Gamma_0)} \rangle_{\Gamma_0}$,

$$\frac{1}{\Omega_A} \int d\Gamma_0 \ e^{-\beta W_\tau(\Gamma_0)} \delta\left[H_{GNP}(\Gamma_0,0)\right] \ .$$

[5] In a numerical integrator of the GNP system, the term ΔH_{GN} will not be exactly zero, and taking it into account might improve stability.

Under the change of variable $\Gamma_0 \to \Gamma_\tau = \phi_\tau(\Gamma_0)$, this expression becomes

$$\frac{1}{\Omega_A}\int d\Gamma_\tau \; e^{-\beta W_\tau(\phi^{-1}(\Gamma_\tau))}\delta\left[s_\tau\left(H_{GN}(\Gamma_\tau,\tau) - W_\tau(\phi^{-1}(\Gamma_\tau)) - H^0_{GN}\right)\right],$$

where the conserved $H_{GNP}(\Gamma_\tau,\tau)$ in the δ-function has been substituted according to (4.26). The rest of the derivation follows closely Ref. [15]. The next step is a second change of variables, $p \leftarrow \frac{\tilde{p}}{s}$ and $p_{\sigma_k} \leftarrow \frac{\tilde{p}_{\sigma_k}}{s}$. Equations (4.27) show that in this representation the evolution of (q,p,Λ) is independent of s. Accordingly, the restricted flow φ is sufficient to determine the work, $W_\tau(\phi^{-1}(\Gamma_\tau)) \equiv W_\tau(\varphi_\tau^{-1}(q_\tau,p_\tau,\Lambda_\tau))$, which can be noted $W_\tau(q,p,\Lambda)$. From now on, all integrals are over Γ_τ, and the τ index on phase variables is left out. Using $d\tilde{p} = s^{N_{df}}dp$, and $d\tilde{p}_{\sigma_k} = s\; dp_{\sigma_k}$ for each of the $k = 1,...,M$ auxiliary momenta , we obtain

$$\frac{1}{\Omega_A}\int drdpd\Lambda \; e^{-\beta W_\tau(q,p,\Lambda)} \times \int ds \; s^{N_{df}+M}\delta\left[s\left(H_B(q,p) + \frac{L}{\beta}\ln s + h(\Lambda) - W_\tau(q,p,\Lambda) - H^0_{GN}\right)\right].$$

The δ-function can be transformed using (4.16) and the integral over s becomes

$$\int ds \; s^{N_{df}+M}\frac{\beta}{L}\delta\left[s - e^{-\frac{\beta}{L}\left[H_B(q,p)+h(\Lambda)-W_\tau(q,p,\Lambda)-H^0_{GN}\right]}\right].$$

By choosing $L = N_{df} + M$ and performing the integration, we get

$$\frac{\beta e^{\beta H^0_{GN}}}{\Omega_A(N_{df}+M)}\int d\Lambda \; e^{-\beta h(\Lambda)}\int drdp \; e^{-\beta W_\tau(q,p,\Lambda)} \; e^{-\beta[H_B(r,p)-W_\tau(q,p,\Lambda)]} \; .$$

Again, W_τ cancels out. The integral over Λ yields a constant, given the hypothesis made on h. The rightmost integral identifies with the canonical partition function Z_B. The same constants appear [15] when integrating the extended variables out of Ω_A to get Z_A. The result follows again,

$$\langle e^{-\beta W_\tau}\rangle_{\Gamma_0} = \frac{Z^B}{Z^A} = e^{-\beta\Delta F_{AB}} \; .$$

4.5 Conclusion

In this paper, the validity domain of the statistical mechanical derivation of the Jarzynski identity presented in Ref. [25] is extended along two axes. The first axis addresses thermostating schemes of practical relevance in MD. As such, the NHC thermostat and the KFC thermo-barostat are implemented in state-of-the-art simulation codes. Similarly, both the Hamiltonian NP scheme and the GGMT should become popular, since they have demonstrated practical advantages. Treating these particular cases, it is shown that different useful features can be added to thermostats without altering the derivation. These include thermostat chains, pressure coupling, and addition of conserved quantities such as the total momentum. One aspect which has not been addressed here is the addition of

holonomic constraints. But as shown in Ref. [2], the reweighting of the phase space due to constraints arises naturally in the metric term $\sqrt{g}$ from the non-Hamiltonian equations of motion. Although it makes the mathematical expressions significantly more complex, the presence of holonomic constraints changes neither the construction of the partition function [13] nor the essence of the derivation. In addition, all possible combinations of the above characteristics are expected to leave the derivation intact, although this was not demonstrated explicitly.

The second extension axis points to probing the theoretical generality of the present JI derivation. Generic classes of thermostats including arbitrary coupling terms are considered. The BK scheme represents a whole family of non-Hamiltonian thermostats based on the NH principle, with a minimal set of hypotheses assuring canonical sampling. In parallel, the derivation is shown to hold for a family of Hamiltonian thermostats represented by the GNP scheme. It turns out that whether the equations of motion are Hamiltonian or non-Hamiltonian is not determinant for the validity of the proof. Based on the cases covered here, one can conjecture that the following conditions are required. First, the autonomous version of the thermostat must admit a proof that the desired phase density is produced for physical variables. In this proof, thermostat variables must be integrated out of the extended microcanonical partition function, which implies several sub-conditions : (i) In non-Hamiltonian systems, the motion invariant contains a term linear in $\bar{\eta}$, which is associated with a metric factor of the form $e^{\bar{\eta}}$. The equivalent requirement for dynamics of the Nosé family is a term of the form $\ln(s)$ in the Hamiltonian (and $\sqrt{g} \equiv 1$). (ii) The thermostat momentum terms in the conserved quantity must be additive and have a well behaved integral $\int e^{-y} dy$. (iii) The phase density is a stationary solution of the Liouville equation. Provided that the above requirements for the thermostat are fulfilled, one additional condition is required for the JI derivation. A coordinate system must be found, in which the thermostating variable ($\bar{\eta}$ or s) is *noncoupling*. In this representation, for all cases considered, the coupling term takes the form $h(p_\eta)G(x)$, and is additive to the dynamics of the physical system. The soundness and the meaning of these hypotheses still need to be investigated, in particular their relation to the detailed balance hypothesis invoked in Ref. [30].

With respect to the thermodynamic proofs of the JI, the present derivation relieves the need for hypotheses on the system size, or on the nature of the heat bath and the coupling to it. Starting directly from the dynamics, no further assumption is needed. The specific conservation laws considered, which can impact the validity of the thermostat and JI, are explicitly taken into account. On the other hand, the present derivation still depends on the particular form of the dynamics considered. However, the lack of generality outlined in Ref. [25] appears now considerably reduced, due to the generic character of the dynamics covered. Further theoretical work is needed to characterize precisely the most general conditions necessary for the proof to hold. These conditions could then be related to their counterparts in the thermodynamic representation.

Acknowledgments

I am grateful to G. Ciccotti for the insightful discussions we had during the reviewing process of the two Jarzynski derivation manuscripts. I thank B. Leimkuhler for useful hints on the NP formalism, as well as W. G. Hoover, C. Jarzynski, P. H. Hünenberger, and B. Žagrović for constructive comments on the manuscript.

4.6 References

[1] G. J. Martyna, M. L. Klein, and M. Tuckerman. "Nosé–Hoover chains: The canonical ensemble via continuous dynamics". *J. Chem. Phys.*, **97**, (1992) 2635.

[2] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna. "Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems". *J. Chem. Phys.*, **115**, (2001) 1678.

[3] B. L. Holian, A. F. Voter, and R. Ravelo. "Thermostatted molecular dynamics: How to avoid the Toda demon hidden in Nosé–Hoover dynamics". *Phys. Rev. E*, **52**, (1995) 2338.

[4] W. G. Hoover and B. L. Holian. "Kinetic moments method for the canonical ensemble distribution". *Phys. Lett. A*, **211**, (1996) 253.

[5] Y. Liu and M. E. Tuckerman. "Generalized Gaussian moment thermostatting: A new continuous dynamical approach to the canonical ensemble". *J. Chem. Phys.*, **112**, (2000) 1685.

[6] J. Jellinek and R. S. Berry. "Generalization of Nosé's isothermal molecular dynamics". *Phys. Rev. A*, **38**, (1988) 3069.

[7] A. C. Branka and K. W. Wojciechowski. "Generalization of Nosé and Nosé–Hoover isothermal dynamics". *Phys. Rev. E*, **62**, (2000) 3281.

[8] A. Bulgac and D. Kusnezov. "Canonical ensemble averages from pseudocanonical dynamics". *Phys. Rev. A*, **42**, (1990) 5045.

[9] H. C. Andersen. "Molecular dynamics simulations at constant pressure and/or temperature". *J. Chem. Phys.*, **72**, (1980) 2384.

[10] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals : A new molecular dynamics method". *J. Appl. Phys.*, **52**, (1981) 7182.

[11] G. J. Martyna, D. J. Tobias, and M. L. Klein. "Constant pressure molecular dynamics algorithms". *J. Chem. Phys.*, **101**, (1994) 4177.

[12] B. Oliva and P. H. Hünenberger. "Calculation of the group-based pressure in molecular simulations. II. numerical tests and application to liquid water". *J. Chem. Phys.*, **116**, (2002) 6898.

[13] G. Kalibaeva, M. Ferrario, and G. Ciccotti. "Constant pressure–constant temperature molecular dynamics: a correct constrained NPT ensemble using the molecular virial". *Mol. Phys.*, **101**, (2003) 765.

[14] D. Kusnezov, A. Bulgac, and W. Bauer. "Canonical ensembles from chaos". *Ann. Phys.*, **204**, (1990) 155.

[15] B. B. Laird and B. J. Leimkuhler. "Generalized dynamical thermostating technique". *Phys. Rev. E*, **68**, (2003) 016 704.

[16] S. Nosé and M. L. Klein. "Constant pressure molecular dynamics for molecular systems". *Mol. Phys.*, **50**, (1983) 1055.

[17] C. P. Dettmann and G. P. Morriss. "Hamiltonian reformulation and pairing of Lyapunov exponents for Nosé-Hoover dynamics". *Phys. Rev. E*, **55**, (1997) 3693.

[18] S. D. Bond, B. J. Leimkuhler, and B. B. Laird. "The Nosé–poincaré method for constant temperature moleculardynamics". *J. Comp. Phys.*, **151**, (1999) 114.

[19] K. Zare and V. Szebehely. "Time transformations in the extended phase-space". *Celestial Mechanics*, **11**, (1975) 469.

[20] J. Struckmeier. "Hamiltonian dynamics on the symplectic extended phase space for autonomous and non-autonomous systems". *J. Phys. A*, **38**, (2005) 1257.

[21] B. J. Leimkuhler and C. R. Sweet. "The canonical ensemble via symplectic integrators using Nosé and Nosé–poincaré chains". *J. Chem. Phys.*, **121**, (2004) 108.

[22] B. J. Leimkuhler and C. R. Sweet. "A Hamiltonian formulation for recursive multiple thermostats in a common timescale". *SIAM J. Appl. Dyn. Sys.*, **4**, (2005) 187.

[23] M. E. Tuckerman, C. J. Mundy, and G. J. Martyna. "On the classical statistical mechanics of non-Hamiltonian systems". *Europhys. Lett.*, **45**, (1999) 149–155.

[24] S. Nosé. "An extension of the canonical ensemble molecular dynamics method". *Mol. Phys.*, **57**, (1986) 187.

[25] M. A. Cuendet. "Statistical mechanical derivation of Jarzynski's identity for thermostated non-Hamiltonian dynamics". *Phys. Rev. Lett.*, **96**, (2006) 120 602.

[26] W. G. Hoover. *Molecular Dynamics*. No. 258 in Lecture Notes in Physics (Springer, Berlin, 1986).

[27] W. G. Hoover. "Generalization of Nosé's isothermal dynamics: Non-Hamiltonian dynamics for the canonical ensemble". *Phys. Rev. A*, **40**, (1989) 2814.

[28] A. C. Branka, M. Kowalik, and K. W. Wojciechowski. "Generalization of the Nosé–Hoover approach". *J. Chem. Phys.*, **119**, (2003) 1929.

[29] C. Braga and K. P. Travis. "A configurational temperature Nosé-Hoover thermostat". *J. Chem. Phys.*, **123**, (2005) 134 101.

[30] C. Jarzynski. "Equilibrium free energy differences from nonequilibrium measurements: a master equation approach". *Phys. Rev. E*, **56**, (1997) 5018.

Chapter 5

Compensating Phase Space Compression

Summary

In a recent paper, J. VandeVondele and U. Röthlisberger [1] (VR) introduced a new statistical-mechanical method allowing to retrieve equilibrium properties from path integrals on nonequilibrium trajectories. The idea is to include a term in the phase space average, which compensates for the phase space compression inherent to nonequilibrium processes. In VR, this term depends on the divergence of the dynamics, which is expressed by the thermostat variable. In the present study, we give an alternative formulation for the phase space weight, which depends directly on the work performed to drive the system. Both formulations are tested by calculating free energy differences of a toy system. The approach of VR, which works with an Euclidean phase space element $d\Gamma$ is then linked to the general framework of Tuckerman *et al.* [2, 3], which works with a phase space volume element evolving with the system.

5.1 Densities on the phase space

Consider a dynamic system which can be described at each time t by a phase vector Γ. For a system of N particles, $\Gamma = (q_1, ..., q_{3N}, p_1, ..., p_{3N})$, where q_i and p_i are the position and the momentum of particle i. A very general dynamic system is governed by the differential equations

$$\dot{\Gamma} = \xi(\Gamma, t). \tag{5.1}$$

A solution of (5.1) is a vector $\Gamma = \Gamma(t, \Gamma_0)$ which is a function of time t and of an initial condition Γ_0. A trajectory or path starting from Γ_0 at $t = 0$ and ending at time τ is defined as the set of points $\{\Gamma(t, \Gamma_0)\}_{t\in[0,\tau]}$. The ensemble of all possible values for Γ, i.e. the ensemble of all possible microstates for the system, constitutes the phase space. In the description of non-Hamiltonian or nonequilibrium systems, the phase space need not be Euclidean. It is generally described by the measure

$$dV(\Gamma, t) = \sqrt{g(\Gamma, t)}d\Gamma.$$

The metric factor that we use here is denoted $g(\Gamma, t)$ because in Riemannian terms it is the determinant of the metric tensor $\mathcal{G}$, which describes the geometry of the space. As shown by Tarasov [4], an invariant $\mathcal{G}$ can be determined explicitly for any given dynamics (5.1). Tuckerman showed [5] that metric conserving integrators can be designed for molecular dynamics (MD). Ezra [6], on another hand, pointed out that defining a metric is not strictly necessary for the following developments, which can be described in the general formalism of differential forms on manifolds. However at this stage, the full machinery of tensor analysis or differential geometry is not needed for our purposes, and $\sqrt{g}$ just remains a notation, Under the constraint that it is smooth and positive for all Γ, $\sqrt{g}$ can be chosen at will. Different authors have made different choices for $\sqrt{g}$. Tuckerman *et al.* [2, 3] (TEA) let $\sqrt{g}$ follow the system coordinate evolution according to the associated Jacobian, such that $\sqrt{g}d\Gamma$ is a motion invariant. On the other hand, Ramshaw [7], Evans and Morriss [8], or Holian, Posh, and Hoover [9] implicitly consider $\sqrt{g} \equiv 1$.

Once a particular functional form for $\sqrt{g}$ has been chosen, it becomes possible to define a probability density on the phase space. We follow here the approach presented by Ramshaw [10], which conciliates the different formulations of the problem. Let $\tilde{f}(\Gamma, t)$ be the time dependent phase space distribution function per unit volume dV. This means that, if Ω determines the total number of microscopic states available to the system, the fraction $\delta\Omega$ of microstates per volume element is

$$\frac{\delta\Omega}{\Omega} = \tilde{f}(\Gamma, t)dV = \tilde{f}(\Gamma, t)\sqrt{g(\Gamma, t)}d\Gamma.$$

Here, $\tilde{f}(\Gamma, t)$ is everywhere positive and normed, $\int \tilde{f}(\Gamma, t)dV = 1$. TEA use a slightly different definition for f where $f(\Gamma, t)$ is not normed,

$$\delta\Omega = f(\Gamma, t)dV = f(\Gamma, t)\sqrt{g(\Gamma, t)}d\Gamma. \tag{5.2}$$

In this sense, f is technically speaking not a density function. The relation between the two definitions is simply $\tilde{f} = f/\Omega$. The TEA choice for f is rather unusual, but being

aware of this peculiarity, we will use their definition, since the present study is partly based on the TEA formalism. In the TEA view, a general partition function can be written as

$$\Omega = \int f(\Gamma, t)\sqrt{g(\Gamma, t)}d\Gamma. \tag{5.3}$$

The phase space average of an observable $\mathcal{A}(\Gamma)$ at time t takes the form

$$\langle A\rangle_t = \frac{1}{\Omega}\int A(\Gamma)f(\Gamma, t)\sqrt{g(\Gamma, t)}d\Gamma.$$

The alternative and more usual formalism adopted in [7, 8, 9] and also used[1] by VR [1] always relates the phase space distribution function to the Cartesian volume element $d\Gamma$ instead of the general volume element dV,

$$\frac{\delta\Omega}{\Omega} = \tilde{\rho}(\Gamma, t)d\Gamma$$

Here, $\tilde{\rho}$ is normed, $\int d\Gamma\tilde{\rho}(\Gamma, t) = 1$. Alternatively, one can consider a form of distribution function, which is not normalized, such that

$$\delta\Omega = \rho(\Gamma, t)d\Gamma \tag{5.4}$$

The advantage of ρ is that the partition function can be expressed

$$\Omega = \int \rho(\Gamma, t)d\Gamma.$$

If we consider this definition of the partition function to be equivalent to the TEA definition (5.3) on an arbitrary volume of the phase space, we have that

$$\rho(\Gamma, t) = \sqrt{g(\Gamma, t)}\ f(\Gamma, t). \tag{5.5}$$

With respect to the distribution $\rho(\Gamma, t)$, the phase space average of any phase space observable $\mathcal{A}(\Gamma)$ is expressed by

$$\langle\mathcal{A}\rangle_t = \frac{1}{\Omega}\int \mathcal{A}(\Gamma)\rho(\Gamma, t)d\Gamma. \tag{5.6}$$

As shown in [7, 8], the state density obeys a continuity condition, which can be expressed by the generalized Liouville equation

$$\frac{\partial\rho}{\partial t} + \nabla_\Gamma \cdot (\rho\dot{\Gamma}) = 0. \tag{5.7}$$

Here ∇_Γ is the $6N$-dimensional gradient operator. Very early numerical studies by Holian and Hoover [11] have shown that simulations of thermostated nonequilibrium systems are

[1]Note that in [1] , ρ is denoted f.

consistent with the Liouville equation (5.7). According to the covariant derivative rule $\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \dot{\Gamma}\nabla\rho$, equation (5.7) can be rewritten

$$\frac{d\rho}{dt} = -\rho\kappa(\Gamma), \tag{5.8}$$

where

$$\kappa(\Gamma) = \nabla_\Gamma \cdot \dot{\Gamma} \tag{5.9}$$

is the phase space compression factor, or compressibility of the phase space flow. As a formal solution of (5.8), the phase space density function takes the form

$$\rho(\Gamma, t) = e^{-K_0(\Gamma,t)}\rho(\Gamma_0, 0), \tag{5.10}$$

where we have defined the phase space compression

$$K_0(\Gamma, t) = \int_0^t \kappa(\Gamma(t'))dt'. \tag{5.11}$$

Note that for Hamiltonian systems, the Liouville theorem guarantees that there is an invariant phase space measure $d\Gamma$. The compressibility vanishes, $\nabla_\Gamma \cdot \dot{\Gamma} = 0$, and the Liouville equation is the one of an incompressible flow,

$$\frac{d\rho}{dt} = \frac{\partial\rho}{\partial t} + \dot{\Gamma}\nabla\rho = 0.$$

5.2 The framework of Tuckerman *et al.* for non-Hamiltonian dynamics

Let's now focus on the approach in which the density f is kept separated from the metric factor $\sqrt{g}$. Along the evolution, the dynamics (5.1) generates a coordinate transformation from the coordinates at time 0 to time t, $\Gamma = \Gamma(t, \Gamma_0)$. The initial phase space volume element $d\Gamma_0$ evolves according to the Jacobian of the coordinate transformation:

$$d\Gamma = J(\Gamma, \Gamma_0)d\Gamma_0. \tag{5.12}$$

It can be shown [2, 3], that $J(\Gamma, \Gamma_0)$ satisfies an equation of the form

$$\frac{d}{dt}J(\Gamma, \Gamma_0) = J(\Gamma, \Gamma_0)\kappa(\Gamma, t). \tag{5.13}$$

For a non-Hamiltonian system, the general solution of (5.13) is

$$J(\Gamma, \Gamma_0) = \exp\left\{\int_0^t \kappa(\Gamma(t'))dt'\right\}. \tag{5.14}$$

If we denote $K(\Gamma, t)$ the indefinite time integral of $\kappa(\Gamma, t)$ defined in (5.9),

$$K(\Gamma, t) = \int \kappa(\Gamma(t'))dt', \tag{5.15}$$

the Jacobian (5.14) can be written in the form $e^{K(\Gamma,t)-K(\Gamma_0,0)}$. Substituting this expression in (5.12) leads to

$$e^{-K(\Gamma,t)}d\Gamma = e^{-K(\Gamma_0,0)}d\Gamma_0.$$

This shows that the measure $e^{-K(\Gamma,t)}d\Gamma$ is invariant under the evolution, and not simply $d\Gamma$. The factor $e^{-K(\Gamma,t)}$ can be associated with the metric factor $\sqrt{g(\Gamma,t)}$, and we have

$$\sqrt{g(\Gamma,t)} = e^{-K(\Gamma,t)}. \tag{5.16}$$

Note that, for Hamiltonian systems, $\kappa = 0$ and the metric factor remains constant, $\sqrt{g(\Gamma,t)} \equiv 1$. This implies that the two definitions (5.2) and (5.4) of the density are equivalent, $f \equiv \rho$ for Hamiltonian systems. Note also that the Jacobian governs the transformation

$$\sqrt{g(\Gamma,t)}d\Gamma = \sqrt{g(\Gamma,t)}J(\Gamma,\Gamma_0)d\Gamma_0 = \sqrt{g(\Gamma_0,0)}d\Gamma_0,$$

and thus the Jacobian is a ratio of metric factors, which is by definition unity at $t = 0$. It is however not possible in general to find a set of phase space coordinates such that $\sqrt{g(\Gamma,t)}$ is constant for non-Hamiltonian systems.

A system evolving according to the dynamics (5.1) and described by a phase space distribution function $f(\Gamma,t)$ satisfies [2] the generalized Liouville equation

$$\frac{\partial(f\sqrt{g})}{\partial t} + \nabla_\Gamma \cdot (f\sqrt{g}\ \dot{\Gamma}) = 0. \tag{5.17}$$

According to (5.5), we see that this generalized Liouville equation is formally equivalent to the standard formulation (5.7) in terms of ρ. In fact, because (5.13) is essentially the same equation as (5.7), $\sqrt{g}$ is also itself a phase space distribution of the same type as ρ. This choice for $\sqrt{g}$ effectively transfers all compression effects from the probability density f into the metric. Because f is a ratio of two such metric factors, all compression effects cancel out. Equations (5.17) and (5.13) together imply that even for a non-Hamiltonian system, f is conserved, i.e. $df/dt = 0$. The separation of f from the metric $\sqrt{g}$ and the invariant measure $\sqrt{g}d\Gamma$ allow to define a coordinate-independent entropy, which is takes a finite value for nonequilibrium steady states (see Chapter 1).

Ramshaw [10] and other authors [12] raised a point against this framework. The fact that $\sqrt{g}$ is itself an phase space density has severe consequences for systems with attractors, for which the trajectories condense into a fractal structure at $t \to \infty$. In this case, the metric factor loses its smoothness and becomes singular. Thus, $\sqrt{g}d\Gamma$ will no longer define an acceptable volume element dV for systems with an attractor in steady state. This issue is still in debate, and it reaches beyond the scope of this study.

Suppose that the set of dynamical equations (5.1) possesses a set of n_c conservation laws or conserved quantities $c_j(\Gamma,t), j = 1,...,n_c$, which satisfy

$$\frac{\mathrm{d}}{\mathrm{d}t}c_j(\Gamma,t) = 0.$$

Thus a trajectory will not sample the entire phase space, but a subspace determined at each time t by the intersection of the hypersurfaces $\{c_j(\Gamma,t) = C_j\}$, where C_j is a set

of constants determined by the initial conditions. Therefore, the distribution function generated by the dynamics can be constructed as a product of δ-functions expressing the conservation laws,

$$f(\Gamma, t) = \prod_j \delta\left[c_j(\Gamma, t) = C_j\right]. \tag{5.18}$$

Equation (5.18) satisfies the generalized Liouville equation (5.17). Note that a distribution constructed from an arbitrary subset of conservation laws also satisfies the generalized Liouville equation. But this solution will not properly describe the phase space distribution of the system. Therefore, satisfying equation (5.17) is a necessary, but not sufficient condition to guarantee that a distribution function is generated by a given dynamical system.

5.3 Weighted phase space averaging

In this section, we recall the ideas of VR [1], on which this work is based. We consider a system which is initially at equilibrium and is forced out of it, starting at $t = 0$. The initial phase space average (5.6) of the observable $\mathcal{A}$ is given by

$$\langle \mathcal{A} \rangle_0 = \frac{\int \mathcal{A}(\Gamma_0)\rho(\Gamma_0, 0)d\Gamma_0}{\int \rho(\Gamma_0, 0)d\Gamma_0}$$

Because the system is out of equilibrium, $\langle \mathcal{A} \rangle_t$ will be different from $\langle \mathcal{A} \rangle_0$. Therefore no ergodicity assumption can be made on such a system, and $\langle \mathcal{A} \rangle$ can not be estimated by following a single trajectory, like it is usually done in equilibrium molecular dynamics. To be able to perform consistent time averages on trajectories, we need a *time-independent* averaging functional on the *time-dependent* phase space. Suppose that the system (together with the thermostat) is ergodic when left at equilibrium. The idea of VR [1] is to devise a weight factor $w(\Gamma, t)$ which allows to define a weighted phase space averaging functional $\langle \cdot \rangle^w$, in such a way that the value of $\langle \mathcal{A} \rangle^w$ is constant over time,

$$\langle \mathcal{A} \rangle_t^w = \frac{\int \mathcal{A}(\Gamma) w(\Gamma, t)\rho(\Gamma, t)d\Gamma}{\int w(\Gamma, t)\rho(\Gamma, t)d\Gamma} = \langle \mathcal{A} \rangle_0. \tag{5.19}$$

The weight factor is defined by

$$w(\Gamma, t) = \frac{\rho(\Gamma, 0)}{\rho(\Gamma, t)}. \tag{5.20}$$

When substituted into formula (5.19), $w(\Gamma, t)$ brings the integrals back to the initial distribution $\rho(\Gamma, 0)$, and we have $\langle \mathcal{A} \rangle_t^w \equiv \langle \mathcal{A} \rangle_0, \forall t$. Using the solution (5.10) of the Liouville equation for $\rho(\Gamma, t)$, we get

$$w(\Gamma, t) = e^{K_0(\Gamma, t)} \frac{\rho(\Gamma, 0)}{\rho(\Gamma_0, 0)}. \tag{5.21}$$

After replacement, the functional takes the form

$$\langle \mathcal{A} \rangle_t^w = \frac{\int \mathcal{A}(\Gamma)\ e^{K_0(\Gamma, t)}\rho(\Gamma, 0)\ \rho^{-1}(\Gamma_0, 0)\ \rho(\Gamma, t)d\Gamma}{\int e^{K_0(\Gamma, t)}\ \rho(\Gamma, 0)\ \rho^{-1}(\Gamma_0, 0)\ \rho(\Gamma, t)d\Gamma}. \tag{5.22}$$

It is now possible to make the ergodic transformation from this time independent phase space average to a time average over a single trajectory sampled according to $\rho(\Gamma, t)$. This leads to estimating the average of A with

$$\langle \mathcal{A} \rangle^w = \lim_{\tau \to \infty} \frac{\int_0^\tau \mathcal{A}(\Gamma) e^{K_0(\Gamma,t)}\ \rho(\Gamma,0)\ \rho^{-1}(\Gamma_0,0)\ dt}{\int_0^\tau e^{K_0(\Gamma,t)}\ \rho(\Gamma,0)\ \rho^{-1}(\Gamma_0,0)\ dt}. \tag{5.23}$$

In practice, the evaluation of $\langle \mathcal{A} \rangle$ from this expression is difficult. Indeed, it can be shown that for a system kept in a nonequilibrium steady state, the phase space compression factor $\kappa(\Gamma, t)$ (5.9) is on average less than zero [13, 14, 15]. Consequently, for a typical member of the ensemble, $K_0(\Gamma, t)$ oscillates around a linearly decreasing function of time, and the weighting term $e^{K_0(\Gamma,t)}$ decreases exponentially during the evolution. In a time average, consequent points on the trajectory are supposed to progressively refine the estimation of $\langle \mathcal{A} \rangle$, and eventually make it converge to its expected value. But in our case, the weight $e^{K_0(\Gamma,t)}$ of a later point on the trajectory is much lower than the weight of the previous ones and the contributions of the last points cannot compensate for the early points. This makes the average dependent on the initial conditions (for finite times).

The second idea of VR [1] is to eliminate this exponential weight decrease, by introducing a phase space compression compensation function. This trick relies on the fact that $\langle \mathcal{A} \rangle_t^w$ in (5.22) is a ratio of two phase space averages. Thus both the numerator and the denominator can be multiplied by an identical factor. If this factor is independent of Γ, it can be introduced into the integrals. Choosing the factor $e^{\gamma(t)}$, where $\gamma(t)$ is a pure function of time, we introduce it in (5.22) and make again the ergodic transformation. From (5.23) we see that after the ergodic transformation, both integrals in the numerator and denominator follow the same trajectory with the the same initial condition, such that the term $\rho^{-1}(\Gamma_0, 0)$ cancels out. This leads to the final result,

$$\langle A \rangle^{\gamma,K} = \lim_{\tau \to \infty} \frac{\int_0^\tau dt\ A(\Gamma)\ e^{K_0(\Gamma,t)+\gamma(t)}\ \rho(\Gamma,0)}{\int_0^\tau dt\ e^{K_0(\Gamma,t)+\gamma(t)}\ \rho(\Gamma,0)}. \tag{5.24}$$

The idea is to choose $\gamma(t)$ such that it keeps the exponent of the weight function small. It has to track the evolution of $-K_0(\Gamma, t)$, while remaining independent of Γ. The ideal choice would be $\gamma(t) = -\langle \kappa \rangle t + c$. The weight of a trajectory point would thus be related to the integral of the fluctuations of the phase space compression around its average value.

At this point, some differences between the present derivation and the one of VR need to be clarified. Instead of using the natural initial distribution of the system for $\rho(\Gamma, 0)$, which we consider canonical, VR use an uniform distribution, $\bar{\rho}(\Gamma, 0) \equiv C$. In this case, the phase space weight (5.20) becomes simply $\bar{w}(\Gamma, t) = e^{K_0(\Gamma,t)}$. In order to compute a phase space average in a realistic ensemble, VR introduce a perturbation term $D(\Gamma)$, such that

$$\langle \mathcal{A} \rangle_t^{\bar{w}} = \frac{\int \mathcal{A}(\Gamma)\ D(\Gamma) e^{K_0(\Gamma,t)}\ \bar{\rho}(\Gamma,t) d\Gamma}{\int D(\Gamma) e^{K_0(\Gamma,t)}\ \bar{\rho}(\Gamma,t) d\Gamma} \tag{5.25}$$

In the end, by choosing the perturbation term as the canonical distribution, introducing $\gamma(t)$, and doing the ergodic transformation, VR arrive after this detour to expression (5.24).

The choice of the phase space compression compensation function $\gamma(t)$ is arbitrary. VR choose a function $\gamma(t)$ which is a second-order damped oscillator, constructed to follow the value of $\ln\left(D(\Gamma)\right) - K_0(\Gamma, t)'$. The oscillator equation has two parameters, which need to be adjusted. However, it offers poor tracking capabilities for drifting functions such as $K_0(\Gamma, t)$. It requires a large equilibration time, and large oscillations can be triggered by abrupt fluctuations of $K_0(\Gamma, t)$. In the present study, we determine the tracking function $\gamma(t)$ as a low pass filter, which takes into account $K_0(\Gamma(t), t)$ over a time interval $[t-\tau_\gamma, t]$, complemented by a drift term. Let Δt be the integration time step of the MD simulation, and τ_γ a multiple of Δt such that $\tau_\gamma \gg \Delta t$. At each iteration, $\gamma(t)$ is calculated in two steps,

$$\begin{aligned} \gamma_1(t) &= \frac{1}{\tau_\gamma}\Big[K_0\big(\Gamma(t-\Delta t), t-\Delta t\big) - \gamma_1(t-\Delta t)\Big] \\ \gamma(t) &= -\gamma_1(t) - \frac{1}{\tau_\gamma}\sum_{t'=t-\tau_\gamma}^{t}\Big[K_0(\Gamma(t'), t') - \gamma_1(t')\Big]\Delta t \end{aligned} \tag{5.26}$$

The idea is that if K_0 drifts linearly, $\gamma_1(t)$ is going to follow with the same slope, but with a constant lag. The sum term is an average estimation of this lag over the last $\frac{\tau_\gamma}{\Delta t}$ steps, which is used as a correction on $\gamma_1(t)$ to get $\gamma(t)$. The negative signs come from the fact that in the present case, the target is $-K_0(\Gamma, t)$. This scheme provides a faithful tracking of a drifting function, with only one parameter, τ_γ.

5.4 Systems with a stationary canonical distribution

Suppose that a function $H'(\Gamma, t))$ represents the energy of the system of interest described by the dynamics (5.1), without the system being necessarily Hamiltonian. The variation of H' can be expressed as

$$\frac{\mathrm{d}H'}{\mathrm{d}t} = \nabla_\Gamma H' \cdot \dot{\Gamma} + \frac{\partial H'}{\partial t}.$$

By integrating each side from $t = 0$ to $t = \tau$, we get

$$\begin{aligned} H'(\Gamma_\tau, \tau) - H'(\Gamma_0, 0) &= \int_0^\tau dt\ \nabla_\Gamma H' \cdot \dot{\Gamma}_t + \int_0^\tau dt \frac{\partial H'}{\partial t} \\ &= \quad \mathcal{Q}_\Gamma \quad + \quad W_a. \end{aligned} \tag{5.27}$$

Relation (5.27) can be seen as a statement of the first law of thermodynamics for the system Γ. Here, Γ represents all degrees of freedom in presence, possibly including some thermostating variables (see Section 5.5). The first integral is commonly interpreted as the heat $\mathcal{Q}_\Gamma$ absorbed by the system and the second integral is the work W_a accumulated by the system.

Consider that Γ is canonically distributed, $\rho \propto e^{-\beta H'(\Gamma, t)}$, and that in addition this distribution function is stationary, $\frac{\partial \rho}{\partial t} = 0$. General classes of dynamics providing these conditions have been reported [16, 17, 18, 19]. Assuming this, the Liouville equation (5.7) implies

$$\nabla_\Gamma \cdot (\rho\dot{\Gamma}) = \rho\left[-\beta\nabla_\Gamma H' \cdot \dot{\Gamma} + \nabla_\Gamma \cdot \dot{\Gamma}\right] = 0.$$

Integrating from $t = 0$ to $t = \tau$ yields

$$\beta \int_0^\tau dt \nabla_\Gamma H' \cdot \dot{\Gamma} = \int_0^\tau dt \nabla_\Gamma \dot{\Gamma}.$$

Relating this equation to the definitions (5.27) of $\mathcal{Q}_\Gamma$ and (5.11) of K_0, we find that the heat absorbed by the system is proportional to its phase space compression,

$$\beta \mathcal{Q}_\Gamma = K_0 \tag{5.28}$$

Note that with the elements developed above, we are in position to write a conserved quantity $H''(\Gamma, t)$ for any non-Hamiltonian system with an energy function $H'(\Gamma, t)$ and held at constant temperature, i.e. thermostated,

$$H''(\Gamma, t) = H'(\Gamma, t) - k_B T K_0(\Gamma, t) - W_a(\Gamma, t) = C \tag{5.29}$$

Although the derivation is different, this result is similar to the invariant identified by Terada and Kidera [20]. It applies to various thermostats including the Gaussian isokinetic thermostat [21], the Nosé-Hoover [22], and the Nosé-Hoover chains [23]. The quantity (5.29) should in particular be used to assess the quality of numerical integrators for thermostated MD. This conserved quantity will become handy in Section 5.5.

We now go back to the weighted phase space average, with result (5.28) in mind. We write expression (5.21) of the weighting function for a canonical initial distribution as

$$\begin{aligned} w(\Gamma, t) &= e^{K_0(\Gamma,t)} e^{-\beta[H'(\Gamma,0) - H'(\Gamma_0,0)]} \\ &= e^{K_0(\Gamma,t)} e^{-\beta[H'(\Gamma,0) - H'(\Gamma,t) + H'(\Gamma,t) - H'(\Gamma_0,0)]} \\ &= e^{K_0(\Gamma,t)} e^{-\beta[H'(\Gamma,0) - H'(\Gamma,t) + \mathcal{Q}_\Gamma + W_a]} \\ &= e^{-\beta[\Phi(r,0) - \Phi(r,t) + W_a]}. \end{aligned} \tag{5.30}$$

The last line corresponds to the case where $H'(\Gamma, t)$ is the energy of system, whose only explicitly time-dependent part is a potential energy term $\Phi(r, t)$ depending only on positions r (a subspace of Γ), such that $\frac{\partial H'}{\partial t} = \frac{\partial \Phi}{\partial t}$. We can now write a new expression for the weighted average (5.24),

$$\langle \mathcal{A} \rangle_t^w = \frac{\int \mathcal{A}(\Gamma) e^{-\beta[\Phi(r,0) - \Phi(r,t) + W_a]} \rho(\Gamma, t) d\Gamma}{\int e^{-\beta[\Phi(r,0) - \Phi(r,t) + W_a]} \rho(\Gamma, t) d\Gamma}.$$

We introduce again a phase space compression compensation function $\gamma_W(t)$, like in (5.24). The difference is that $\gamma_W(t)$ tracks the work $-\beta W_a$ performed on the system, instead of $-K_0$. A scheme similar to (5.26) is used to calculate $\gamma_W(t)$ in the simulation.

$$\langle \mathcal{A} \rangle^{\gamma,W} = \lim_{\tau \to \infty} \frac{\int_0^\tau dt \; A(\Gamma) \; e^{-\beta[\Phi(r,0) - \Phi(r,t) + W_a] + \gamma_W(t)}}{\int_0^\tau dt \; e^{-\beta[\Phi(r,0) - \Phi(r,t) + W_a] + \gamma_W(t)}}. \tag{5.31}$$

This expression will be compared to $\langle \mathcal{A} \rangle^{\gamma,K}$ (5.24) by computing free energy differences in thermostated systems.

5.5 The Nosé-Hoover chain thermostat

Consider again our N-particle system with cartesian positions $r = \{r_1, ..., r_{3N}\}$ and momenta $p = \{p_1, ..., p_{3N}\}$ in 3 dimensions. In MD simulations, the Nosé-Hoover chain (NHC) thermostat [23] is used to regulate the temperature of the system and generate canonically distributed physical variables. This method involves an extended phase space. The physical positions and momenta, which we denote $x = (r, p)$ are coupled to the thermostating variables η_k and their associated momenta p_{η_k}, where $k = 1...M$ for a chain of M thermostats. The extended phase space state can be expressed as

$$\hat{\Gamma} = (r, p, \eta, p_\eta).$$

The System evolution is then represented by the non-Hamiltonian dynamical system

$$\begin{aligned}
\dot{r} &= \frac{p}{m}, && (5.32)\\
\dot{p} &= -\frac{\partial \Phi}{\partial r}(r,t) - \frac{p_{\eta 1}}{Q_1} p, \\
\dot{\eta}_k &= \frac{p_{\eta k}}{Q_k} \qquad k = 2, ..., M, && (5.33)\\
\dot{p}_{\eta_1} &= \frac{p^2}{m} - N_{df} k_B T - \frac{p_{\eta_2}}{Q_2} p_{\eta_1}, \\
\dot{p}_{\eta_k} &= \frac{p^2_{\eta_{k-1}}}{Q_{k-1}} - k_B T - \frac{p_{\eta_{k+1}}}{Q_{k+1}} p_{\eta_k}, \\
\dot{p}_{\eta_M} &= \frac{p^2_{\eta_{M-1}}}{Q_{M-1}} - k_B T.
\end{aligned}$$

It is shown in [23] that an optimal choice for the thermostat masses is $Q_1 = N_{df} k_B T \tau^2_{\mathrm{NHC}}$, and $Q_k = k_B T \tau^2_{\mathrm{NHC}}$ for $k = 2, ..., M$, where N_{df} is the number of degrees of freedom of the system. The parameter τ_{NHC} represents a typical timescale for the system under study. The compression factor of equations (5.33) is, according to (5.9),

$$\begin{aligned}
\kappa(\Gamma) &= -N_{df} \frac{p_{\eta_1}}{Q_1} - \sum_{k=2}^{M} \frac{p_{\eta_k}}{Q_k} \\
&= -N_{df} \dot{\eta}_1 - \sum_{k=2}^{M} \dot{\eta}_k \\
&= -\dot{\bar{\eta}},
\end{aligned}$$

where we have defined the collective variable

$$\bar{\eta} = N_{df} \eta_1 + \sum_{k=2}^{M} \eta_k.$$

In this case, the phase space compression (5.15) becomes

$$K = -\bar{\eta}.$$

Thus, according to (5.16) we have that

$$\sqrt{g(\Gamma,t)} = e^{\bar{\eta}}. \tag{5.34}$$

The energy terms of interest include the Hamiltonian of the physical system,

$$H(r,p,t) = \sum_{i=1}^{3N} \frac{p_i^2}{2m_i} + \Phi(r,t),$$

and the energy of the system $\Gamma = (r, p, p_\eta)$,

$$H'(r,p,p_\eta,t) = H(r,p,t) + \sum_{k=1}^{M} \frac{p_{\eta_k}^2}{2Q_k}$$

The most interesting property of the Nosé-Hoover chain thermostat is the following: A microcanonical ensemble for $\hat{\Gamma}$ (isolated extended system) projects to a canonical distribution on the physical phase space $\{r, p\}$, as well as on the phase space of $\Gamma = (r, p, p_\eta)$. We consider that the thermostat variables model a heat bath in contact with the physical system. Under modification of the potential $\Phi(r,t)$, the physical system is pushed to change its energy. Mechanical work is performed on the system $x = (r, p)$, which is driven away from equilibrium. Then the system inevitably tends to heat up due to the irreversible conversion of work into heat [24]. But this is compensated by the thermostat, which absorbs the excess heat. To clarify how the thermostat extracts heat from the physical system, we look at the variation of the energy $H(x,t)$:

$$\begin{aligned} H(x,\tau) - H(x_0,0) &= \int_0^\tau \nabla_x H(x,t)\cdot \dot{x}_t + \int_0^\tau dt \frac{\partial H}{\partial t}(x,t) \\ &= \quad \mathcal{Q}_x \quad + \quad W_a \end{aligned} \tag{5.35}$$

This is again an analog of the second law of thermodynamics, but this time for the physical system only. We first note that the expression for the work is the same as in (5.27), because the explicit time dependence affects only the physical potential $\Phi(r,t)$. To get a sense of the physical meaning of $\mathcal{Q}_x$, we go back to the equations of motion of the physical particles (5.32), and identify the term $-\frac{p_{\eta_1}}{Q_1}p$ as the (non conservative) force f^{thermo} applied by the thermostat on the physical system. We can then compute the work of this force in the usual way

$$W^{\text{thermo}} = \sum_{i=1}^{3N} \int_0^\tau dt f_i^{\text{thermo}} \cdot \dot{r}_i = -\sum_{i=1}^{3N} \int_0^\tau dt \frac{p_{\eta_1} p_i^2}{Q_1 m_i}. \tag{5.36}$$

If we work out the expression of the exchanged heat $\mathcal{Q}_x$ defined in (5.35), it turns out, not surprisingly, that $W^{\text{thermo}} = \mathcal{Q}_x$.

If we now focus on the $\Gamma = (x, p_\eta)$ system, we can apply the results found in Section 5.4. The variation of $H'(r,p,p_\eta,t)$ is given by (5.27). Subtracting expressions (5.27) and (5.35), we find

$$\mathcal{Q}_x + \sum_{k=1}^{M} \frac{1}{2Q_k}\left[p_{\eta_k}^2 - p_{\eta_k}^2(0)\right] = \mathcal{Q}_\Gamma \tag{5.37}$$

This means that the heat $\mathcal{Q}_x$ extracted from the physical system is equal to the heat $\mathcal{Q}_\Gamma$ extracted from the $\Gamma = (x, p_\eta)$ system, up to the fluctuations of kinetic energy of the p_η variables. On average, this kinetic energy is constant, and the heat flux is conserved through the thermostat momenta. According to (5.28), the heat pumped by the thermostat into the system $\Gamma = (r, p, p_\eta)$ is $\mathcal{Q}_\Gamma = \frac{1}{\beta}K_0 = -\frac{1}{\beta}\left[\bar{\eta} - \bar{\eta}(0)\right]$. Using this and (5.36) to calculate $\mathcal{Q}_x$, relation (5.37) can be used in practice to check an implementation of the NHC thermostat.

For the nonautonomous NHC dynamics, the general conserved quantity (5.29) takes the following form (which appears already in Ref. [25]),

$$H''(\hat{\Gamma}) = H'(\Gamma, t) + k_B T\bar{\eta} - W_a(\Gamma, t). \tag{5.38}$$

All terms in this conservation law are available in practical implementations, and $H''(\hat{\Gamma})$ can readily be calculated. Monitoring the evolution of $H''(\hat{\Gamma})$ constitutes a good quality check for a simulation. Note that $H''(\hat{\Gamma})$ is usually not constant in MD simulations but exhibits steady drifts, because there are hidden source terms in (5.38) due to numerical drift and cutoff truncation of forces, among others (see Chapter 6).

5.6 Application of the formalism of Tuckerman *et al.*

We now consider the full extended phase space $\hat{\Gamma} = \{r, p, p_\eta, \eta\}$, and apply the method of TEA [2, 3] described above, in order to express the distribution function of the physical system. According to (5.18), the distribution function of our extended system has to be constructed from the constant of the motion (5.38)

$$f(\hat{\Gamma}, t) = \delta\left[H''(\hat{\Gamma}, t) - C\right].$$

Using the metric factor (5.34), we can write the partition function over the extended space

$$\hat{\Omega}_t(N, V, T) = \int dpdrdp_\eta d\bar{\eta}\; e^{\bar{\eta}}\delta\left[H''(\hat{\Gamma}, t) - C\right].$$

We then reformulate the Dirac function using (5.38),

$$\delta\big[-\beta\big(H'(\Gamma, t) - W_a(\Gamma, t) - C\big) - \bar{\eta}\big].$$

A property of the Dirac δ-function is that for any smooth function $\ell(\eta)$ with one simple root $\eta_\star$, $\delta[\ell(\eta)] = \delta[\eta - \eta_\star]/|\ell'(\eta_\star)|$. Using this, we integrate over $\bar{\eta}$ to get rid of the Dirac function,

$$\hat{Z}_t(N, V, T) = \beta e^{\beta C}\int dpdrdp_\eta e^{-\beta\left[H'(r,p,p_\eta,t) - W_a(r,p,p_\eta,t)\right]}.$$

At this point, we cannot further follow the derivation provided [3] by TEA for the equilibrium NHC. In this derivation, the kinetic energy terms for p_η integrate out in a constant, and the distribution function for the physical variables is found. Here, however, W_a depends on p_η. This is due to the fact that, although W_a can be defined in terms of a path $\{r\}_{t=0}^{\tau}$ only, the determination of this path via the dynamics (5.32) requires knowledge of

at least one point on the trajectory of $p_\eta(0)$, since r is coupled to p_η. The integration over $\bar{\eta}$ could be performed, since none of r, p, p_η is coupled to it. So we cannot further separate the expression of $\hat{Z}_t(N,V,T)$ to find the distribution function of (r,p). We have to satisfy ourselves with the time-dependent density of the Γ system,

$$\rho_{\mathrm{TEA}}(\Gamma,t) = e^{-\beta\left[H'(\Gamma,t)-W_a(\Gamma,t)\right]}.$$

We can now define a weighted phase space average for an observable $\mathcal{A}(x)$, according to the approach of VR,

$$\langle\mathcal{A}\rangle_t^{\mathrm{TEA}} = \frac{\int \mathcal{A}(\Gamma) w_{\mathrm{TEA}}(\Gamma,t)\rho(\Gamma,t)d\Gamma}{\int w_{\mathrm{TEA}}(\Gamma,t)\rho(\Gamma,t)d\Gamma} = \langle\mathcal{A}\rangle_0.$$

The weight factor is defined by

$$\begin{aligned} w_{\mathrm{TEA}}(\Gamma,t) &= \frac{\rho_{\mathrm{TEA}}(\Gamma,0)}{\rho_{\mathrm{TEA}}(\Gamma,t)} \\ &= \frac{e^{-\beta H'(\Gamma,0)}}{e^{-\beta[H'(\Gamma,t)-W_a(\Gamma,t)]}} \\ &= e^{-\beta[\Phi(r,0)-\Phi(r,t)+W_a(\Gamma,t)]}. \end{aligned} \tag{5.39}$$

This form is equivalent to (5.30), which was found following a different route. This means that the VR approach cast into the TEA formalism, points to the same expression (5.31) for a time average of an observable in a nonequilibrium NHC system. The TEA approach seems less general than the one presented in Section 5.4, because (5.39) has been derived using the NHC dynamics explicitly. However, the derivation can be applied unchanged for any dynamical system satisfying the hypothesis of Section 5.4.

5.7 Calculating free energy differences

The goal is to calculate the free energy difference between two (equilibrium) Hamiltonians $H_A(r,p) = \sum \frac{p_i^2}{2m_i} + \Phi_A(r)$, and $H_B(r,p) = \sum \frac{p_i^2}{2m_i} + \Phi_B(r)$,

$$\exp\{-\beta\Delta F_{AB}\} = \frac{\int e^{-\beta H_B(x)}dx}{\int e^{-\beta H_A(x)}dx}.$$

The nonequilibrium dynamics is constructed with a potential energy function $\Phi(r,t)$ that is switched between Φ_A and Φ_B in an oscillatory fashion. VR use a free energy perturbation approach to calculate the Boltzmann averages of H_A and H_B from the nonequilibrium ensemble. In fact, since the nonequilibrium phase space averages (5.24) and (5.31) have already been formulated in terms of the initial distribution function $\rho(\Gamma,0)$, the perturbation is also formally with respect to $\rho(\Gamma,0)$ instead of $\rho(\Gamma,t)$. This is equivalent to taking an observable

$$\mathcal{A}(\Gamma) = e^{-\beta[H_A(x)-H(x,0)]} = e^{-\beta[\Phi_A(r)-\Phi(r,0)]},$$

and a similar observable $\mathcal{B}$ corresponding to H_B and calculating

$$e^{-\beta\Delta F_{AB}} = \frac{\langle\mathcal{B}\rangle_t^w}{\langle\mathcal{A}\rangle_t^w}$$

with $\langle\cdot\rangle_t^w$ the weighted phase space average functional (5.19). Note that the partition functions are the same for $\mathcal{A}$ and $\mathcal{B}$, and cancel out. For an NHC thermostated system, $K_0 = -\bar{\eta} + \bar{\eta}(0)$, and the term in $\bar{\eta}(0)$ cancels out. Performing the ergodic transformation, and adopting the reweighting method of VR, we replace the denominator with $\langle\mathcal{A}\rangle^{\gamma,K}$ and the numerator with $\langle\mathcal{B}\rangle^{\gamma,K}$, both defined with (5.24). With $\rho(\Gamma,0) = e^{-\beta H'(\Gamma,0)}$ this gives

$$e^{-\beta\Delta F_{AB}^K} \simeq \frac{\int_0^\tau dt\ e^{H'_B(\Gamma)}\ e^{-\bar{\eta}+\gamma_\eta(t)}}{\int_0^\tau dt\ e^{H'_A(\Gamma)}\ e^{-\bar{\eta}+\gamma_\eta(t)}} \tag{5.40}$$

Here we note the tracking function $\gamma_\eta(t)$ to emphasize the fact that in the NHC case, it tracks $\bar{\eta}$. The superscript K reminds that ΔF_{AB}^K uses a reweighting function based on the phase space compression K_0 (5.11). On the other hand, using the new formulation for the weighted phase space average $\langle\mathcal{A}\rangle^{\gamma,W}$ (5.31), we get

$$e^{-\beta\Delta F_{AB}^W} \simeq \frac{\int_0^\tau dt\ e^{-\beta\left[\Phi_B(r)-\Phi(r,t)+W_a(\Gamma,t)\right]+\gamma_W(t)}}{\int_0^\tau dt\ e^{-\beta\left[\Phi_A(r)-\Phi(r,t)+W_a(\Gamma,t)\right]+\gamma_W(t)}}. \tag{5.41}$$

This provides a way to estimate the equilibrium free energy difference ΔF from a path integral on a nonequilibrium trajectory $\{\Gamma(t)\}_{t=0}^\tau$, in which the Hamiltonian is switched between states A and B.

In a discrete time frame, (5.41) can be rewritten

$$e^{-\beta\Delta F_{AB}^W} \simeq \frac{\sum_{j=1}^n\ e^{-\beta\left[\Phi_B(r)-\Phi(r,t_j)\right]}e^{-\beta W_a(\Gamma,t_j)+\gamma_W(t_j)}}{\sum_{j=1}^n e^{-\beta\left[\Phi_A(r)-\Phi(r,t_j)\right]}e^{-\beta W_a(\Gamma,t_j)+\gamma_W(t_j)}}. \tag{5.42}$$

This formulation displays similarities with the Jarzynski identity [25] (JI), which can be written as an arithmetic average over n trajectories,

$$e^{-\beta\Delta F_{AB}^{\mathrm{Jar}}} = \frac{\frac{1}{n}\sum_{j=1}^n\ e^{-\beta W_{AB}(\Gamma_j,t)}}{Z_A}. \tag{5.43}$$

Here, the Γ_j represent n initial conditions sampled in state A. Formula (5.42) uses the same process and same initial condition with different starting times t_j. The JI on the other hand uses the same starting time for all averages, but different starting conditions Γ_j. Further theoretical investigations are necessary to establish the exact nature of the relationship between the two expressions.

5.7.1 Validation : the harmonic oscillator

To test the free energy difference estimators (5.40) and (5.41), we calculate the free energy difference of a material point between to states A and B defined by harmonic potential

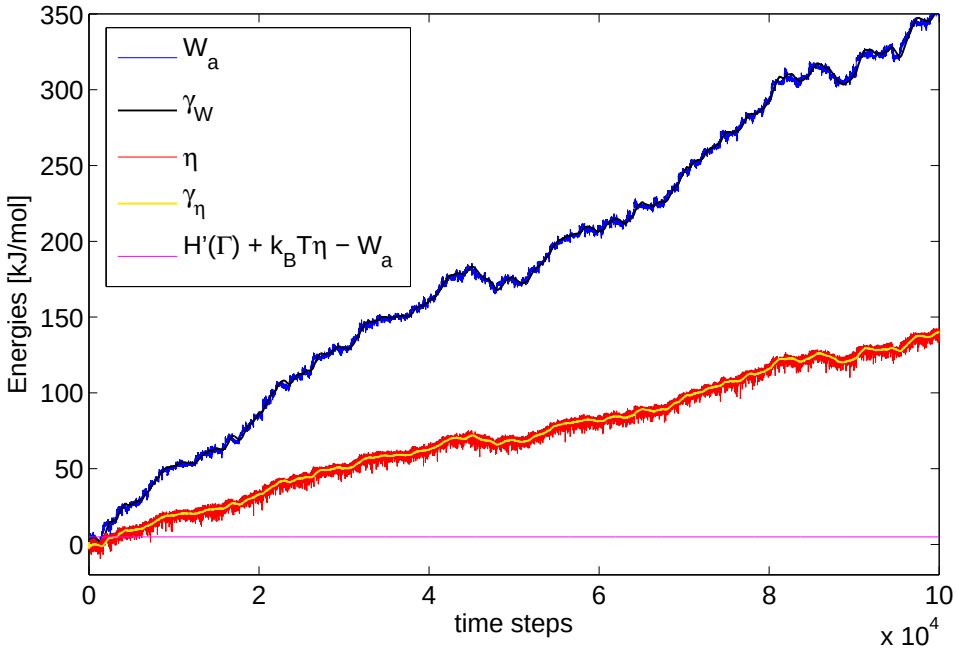

Figure 5.1: *Evolution of the external work W_a performed on an harmonic oscillator by periodically changing its harmonic constant. The resulting phase space compression $\bar{\eta}$ is also shown, as well as the tracking functions $\gamma_W(t)$ and $\gamma_\eta(t)$. The purple line shows the conservation of the NHC quantity $H''(\hat{\Gamma}, t)$ (5.38).*

energy functions with constants k^A and k^B. The theoretical free energy difference is

$$\Delta F^{\text{theo}} = -\frac{1}{\beta} \ln \left(\sqrt{\frac{k^A}{k^B}} \right). \tag{5.44}$$

The nonequilibrium dynamics use a time-dependent harmonic constant, which sweeps the interval between k^A and k^B with a period $\omega_k/2\pi$,

$$k(t) = \tfrac{1}{2}\Big[(k^A + k^B) + (k^A - k^B)\cos(\omega_k t)\Big]. \tag{5.45}$$

Fig. 5.1 shows the evolution of $\bar{\eta}$ and $W_a(\Gamma, t)$ as well as their respective tracking functions $\gamma_\eta(t)$ and $\gamma_W(t)$ calculated using (5.26). The approximate ratio between the curves is β, which is consistent with equations (5.40) and (5.41). As an additional control, the NHC conserved quantity $H''(\hat{\Gamma})$ (5.38) is plotted. We see that it can be considered constant with respect to the quantities of energy flowing through the system. The parameters chosen here are $k^A = 2000$ kJmol^{-1}nm^{-2}, $k^B = 5000$ kJmol^{-1}nm^{-2}, temperature $T = 300$ K, mass $m = 23$ a.u., $\tau_{\text{NHC}} = 0.5$ ps, $\omega_k = 2\pi$, $\tau_\gamma = 2$ ps, $\Delta t = 0.001$ ps.

In order to assess the convergence properties of the free energy difference estimators F^K_{AB} (5.40) and F^W_{AB} (5.41), 20 trajectories are generated with random initial conditions

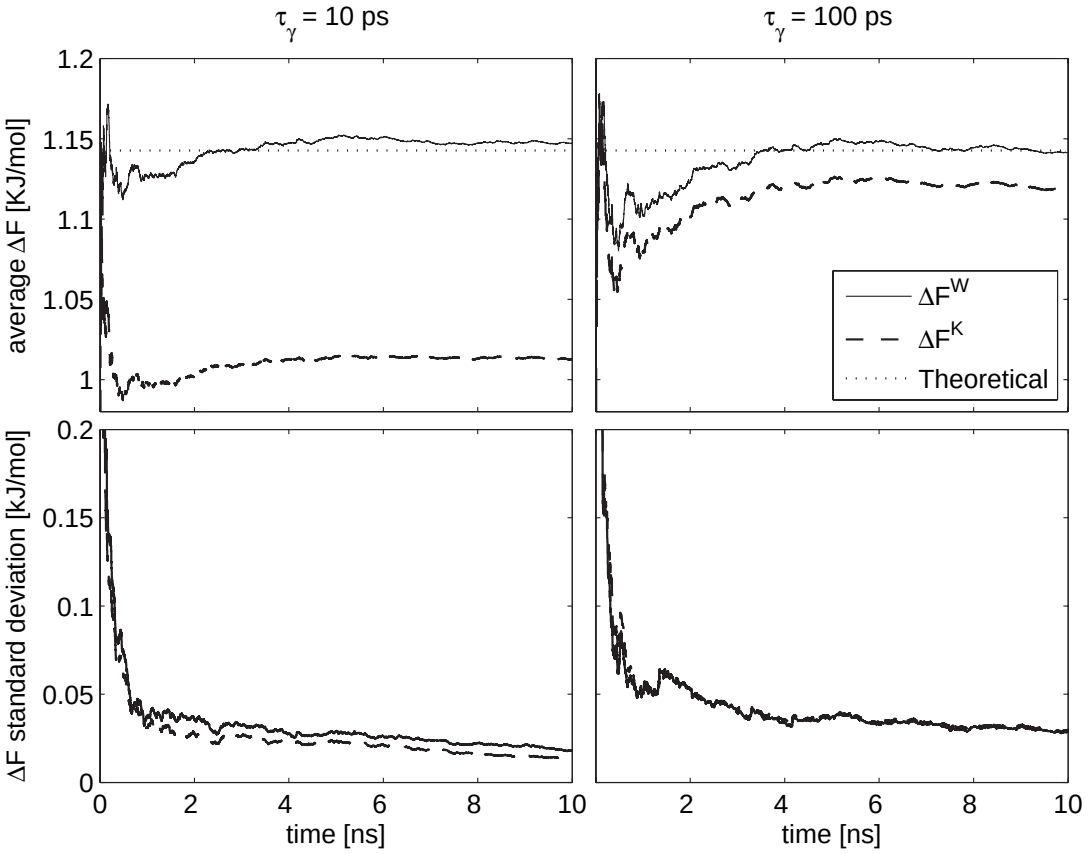

Figure 5.2: *Free energy difference of a material point between two harmonic potential energy functions with different harmonic constants: convergence study on 20 trajectories. Here ΔF^W refers to Eq. (5.41), and ΔF^K to Eq. (5.40) (see text for details).*

for both $\tau_\gamma = 10$ ps and $\tau_\gamma = 100$ ps. The two upper panels of Fig. 5.2 show the average behavior, compared to the theoretical value (5.44). The two lower panels on Fig. 5.2 show the standard deviation over the 20 trajectories. While F^W_{AB} reliably converges to the theoretical value, F^K_{AB} displays a systematic bias. This bias seems to decrease when τ_γ is increased. This could be related to the fact that $\bar{\eta}$ has a much more rugged behavior than W_a (see Fig. 5.1). This feature is apparently consistent with VR [1], who find that the free energy estimator based on $\bar{\eta}$ converges better with a slow oscillating compensating function $\gamma_\eta(t)$. According to the present test, F^W_{AB} is a better estimator than F^K_{AB}.

A good consistency check for the simulation is to plot the free energy estimators without the phase space compression compensating functions $\gamma_\eta(t)$ and $\gamma_W(t)$ (not shown). These uncompensated time averages very quickly converge to a meaningless value, without the further points on the trajectory being able to influence it (see Fig. 5.3). An exact overlap of the uncompensated F^W_{AB} and F^K_{AB} is expected, as predicted by the conserved quantity (5.38).

5.7.2 The multiple nonequilibrium path integral method

In the weighted phase space averages (5.24) and (5.31), using the phase space compression compensation functions $\gamma_\eta(t)$ and $\gamma_W(t)$ to track either $\bar{\eta}$ or $w_a(\Gamma, t)$, constitutes an approximation. Indeed, this violates the hypothesis that $\gamma_\eta(t)$ and $\gamma_W(t)$ are pure functions of time. The impact of this violation depends on the tracking scheme chosen, for example a damped oscillator like in VR, or a low pass filter like in the present study. If the tracking is too tight, $\gamma_\eta(t)$ and $\gamma_W(t)$ cannot be considered independent of Γ anymore. It is impossible to assess *a priori* the impact of this approximation as a function of the various parameters of the tracking scheme. This observation remains valid for the free energy estimators (5.40) and (5.41).

At the cost of violating another hypothesis, one can avoid the use of tracking functions. To calculate the free energy difference between states A and B we run a nonequilibrium simulation in which the potential energy function $\Phi(r, t)$ oscillates between Φ_A and Φ_B. In the above calculations, the path integrals in equations (5.40) and (5.41) start at $t = 0$, time at which the perturbation is turned on and $\Phi(r, t)$ leaves state A. This is consistent with the hypothesis that Γ is initially canonically distributed (in state A). In practice however, one can start the sampling of a path integral at any time $t > 0$. We have seen on Fig. 5.2 that without $\gamma_\eta(t)$ and $\gamma_W(t)$, the time averages are rapidly trapped at an inaccurate free energy value. But as the simulation goes on, several averages can be started, separated by a given time interval. At the end of the process, we have a collection of averages, each trapped at different values.

Fig. 5.3 shows an example of such a multiple path integral free energy estimation. This example was computed in arbitrary units, with a particle of mass $m = 1$, at a temperature $T = 1$ in an harmonic potential with an harmonic constant oscillating between $k^A = 6$ and $k^B = 14$ with a pulsation $\omega_k = 0.3$, and a time step $\Delta t = 0.01$. The corresponding theoretical equilibrium free energy difference between A and B is 0.4227. The arithmetic mean of the inaccurate free energy averages shown on Fig. 5.3 yields $\Delta F_{AB}^{K,\text{multi}} = 0.4242$, with only 18 path integrals estimated. The example showed is representative of the kind of convergence observed with a variety of parameters.

The present multiple nonequilibrium path integral method, in its version based on ΔF_{AB}^{W} (5.41) can be formally expressed as

$$e^{-\beta \Delta F_{AB}^{W,\text{multi}}} = \frac{1}{n} \sum_{j=1}^{n} \frac{\int_{t_j}^{\tau} dt \; e^{-\beta \left[\Phi_B(r) - \Phi(r,t) + W_a(\Gamma, t_j, t) \right]}}{\int_{t_j}^{\tau} dt \; e^{-\beta \left[\Phi_A(r) - \Phi(r,t) + W_a(\Gamma, t_j, t) \right]}}, \tag{5.46}$$

with t_j, the time at which the jth path integral is started. In each of these n path integrals, the work is accumulated from t_j to t,

$$W_a(\Gamma, t_j, t) = \int_{t_j}^{t} dt \; \frac{\partial \Phi}{\partial t}.$$

Note that there is no theoretical proof that (5.46) converges to the equilibrium free energy difference. The reliability and the performance of the method (5.46) need to be further assessed, and its links to the Jarzynski identity (5.43) should be studied.

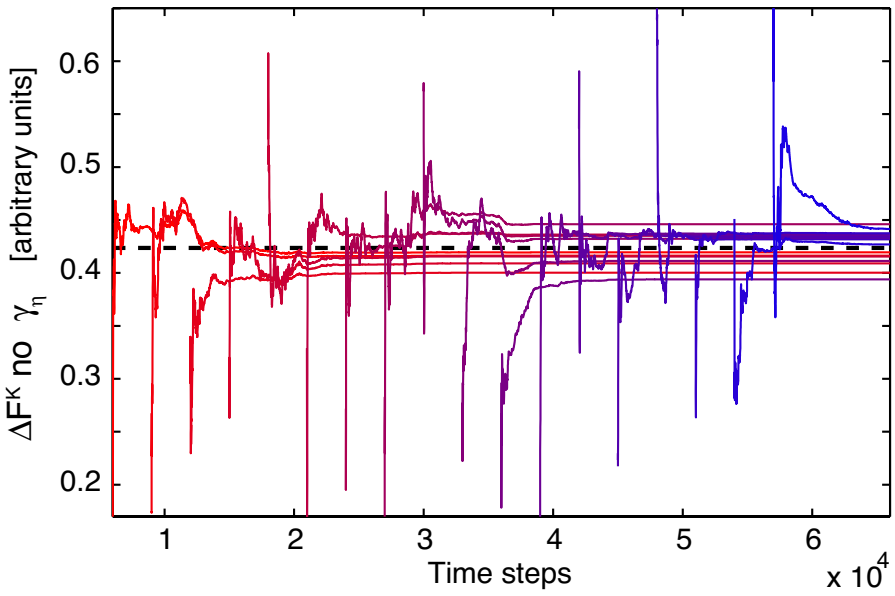

Figure 5.3: *Free energy difference of a material point between two harmonic potential energy functions with different harmonic constants. ΔF^K is estimated with (5.40) without the phase space compression tracking function $\gamma_\eta(t)$. The procedure is repeated at intervals of 3000 steps. The dashed line shows the theoretical free energy difference. See text for parameters of the simulation.*

5.8 Calculating potentials of mean force

For a physical system described by an equilibrium Hamiltonian $H_0(\Gamma)$, the potential of mean force (PMF) at a given position Ξ along a reaction coordinate $\xi(r)$, which is a function of the spacial coordinates r, is defined as

$$e^{-\beta F(\Xi)} = \frac{\int d\Gamma \; e^{-\beta H_0(\Gamma)} \delta\big[\xi(r) - \Xi\big]}{\int d\Gamma \; e^{-\beta H_0(\Gamma)}} \tag{5.47}$$

In the presence of free energy barriers, the system will not spontaneously explore the Ξ range of interest. We therefore add an external restraint of the form $u(\xi(r), \xi_0(t))$, which keeps the physical or intrinsic system close to $\xi_0(t)$. The parameter $\xi_0(t)$ can then be switched in such a way that the system visits the relevant region of Ξ. The resulting Hamiltonian describes the biased or composite system,

$$H(\Gamma, t) = H_0(\Gamma) + u(\xi(r), \xi_0(t)).$$

Using a collection of equilibrium simulations with fixed restraints centered at neighboring positions ξ_0 along the reaction coordinate Ξ amounts to performing an umbrella sam-

pling [26] calculation. Here, we are interested in doing a single simulation with a time-dependent restraint which sweeps the reaction coordinate, and brings the system out of equilibrium. Using the phase space weighted expression (5.19), we can calculate the phase space average of the particular observable $\mathcal{A}(\Gamma) = \delta\big[\xi(r) - \Xi\big]$ along a nonequilibrium trajectory. This yields the PMF along Ξ for the intrinsic system H_0. To evidence this, we reproduce the unbiasing procedure of Ref. [27] in the nonequilibrium case. Consider the nonequilibrium distribution function $\rho(\Gamma, t)$ corresponding to the extended system. The idea of Ref. [27] is to multiply both the numerator and the denominator of (5.47) with the two weighted integrals

$$\int d\Gamma\ w(\Gamma,t)\rho(\Gamma,t),$$
$$\int d\Gamma\ \delta\big[\xi(r) - \Xi\big] w(\Gamma,t)\rho(\Gamma,t).$$

The result can be arranged in three terms

$$\begin{aligned} e^{-\beta F(\Xi)} &= \frac{\int d\Gamma\ \delta\big[\xi(r) - \Xi\big] w(\Gamma,t)\rho(\Gamma,t)}{\int d\Gamma\ w(\Gamma,t)\rho(\Gamma,t)} \\ &\quad \times \frac{\int d\Gamma\ w(\Gamma,t)\rho(\Gamma,t)}{\int d\Gamma\ e^{-\beta H_0}} \\ &\quad \times \frac{\int d\Gamma\ e^{-\beta H_0(\Gamma)}\delta\big[\xi(r) - \Xi\big]}{\int d\Gamma\ \delta\big[\xi(r) - \Xi\big] w(\Gamma,t)\rho(\Gamma,t)}. \end{aligned}$$

The first term represents the weighted biased PMF for the extended system. If we choose

$$w(\Gamma,t)) = \frac{e^{-\beta H_0(\Gamma)}}{\rho(\Gamma,t)},$$

we see that both the second and third terms reduce to unity. The weight $w(\Gamma, t))$ effectively converts a nonequilibrium phase average for the composite system to an equilibrium phase average for the intrinsic system. Using the two different expressions we found for the denominator $\rho(\Gamma, t)$ of the weight function, either (5.21) or (5.30), we get two expressions for the PMF. First, using $\langle A\rangle^{\gamma,K}$ (5.24) obtained with weighting function (5.21) based on the phase space compression $K_0(\Gamma, t) = -\bar{\eta} + \bar{\eta}(0)$ for a system coupled to a NHC thermostat, we get

$$e^{-\beta F^K(\Xi)} \propto \int_0^\tau dt\ \delta\big[\xi(r) - \Xi\big] e^{-\beta H_0(\Gamma)} e^{-\bar{\eta}+\gamma_\eta(t)} \tag{5.48}$$

Second, using $\langle A\rangle^{\gamma,W}$ (5.31) with the weighting function (5.21) based on the work $W_a(\Gamma, t)$, we get

$$e^{-\beta F^W(\Xi)} \propto \int_0^\tau dt\ \delta\big[\xi(r) - \Xi\big] e^{+\beta u(\Gamma,t)} e^{-\beta W_a(\Gamma,t)+\gamma_W(t)} \tag{5.49}$$

In both cases, the denominator does not depend on Ξ. Since we are only interested in differences between different positions along Ξ, we have left it out.

5.9 Outlook

The present work is elaborated on the method proposed by VR [1] to calculate equilibrium phase space averages from nonequilibrium path integrals. The method uses a reweighting function based on the phase space compression factor. First, the derivation of the original method was reformulated on firmer grounds. Second, based on general properties of dynamical systems conserving a canonical distribution, an alternative result was provided, which uses a reweighting function based directly on the work performed on the system. The two formulations were compared on a toy system. Third, a theoretical connection with the formalism of TEA [3, 2] was made and we showed that the derivations in both formalisms arrive at the same result. The multiple nonequilibrium path integral method was proposed, as an alternative way to calculate free energies. Finally formulas are proposed to calculate potentials of mean force, based on both formulations of phase space weighted averaging.

The efficiency of both phase space reweighting free energy formulas (5.40) and (5.41) has been compared on one toy example, which shows that the W_a based method performs better than the K_0 based method. However, this project is still open, and many questions await answers. First, a more extensive study of the two free energy estimators is needed, including systematic variation of the parameters and application to other systems. In addition to free energy differences between two states A and B, potential of mean force calculations should be performed to assess the performance of expressions (5.48) and (5.49). A proposed molecular dynamics system to study potentials of mean force would be the water exchange process close to a sodium ion in aqueous solution. This is a good test system, since reference potentials of mean force can be calculated from the radial distribution function of water around the ion. The performance of (5.48) and (5.49) could then be compared to that of other nonequilibrium methods, such as slow growth and the Jarzynski [25] method. Moreover, theoretical comparisons should be pursued between both phase space reweighting methods (5.40) and (5.41), the multiple nonequilibrium path integral method (5.46) and other free energy methods such as the Jarzynski identity [25], or the Crooks theorem [28].

Acknowledgments

The work presented in this chapter was supervised by U. Röthlisberger. I am also grateful to G. Ciccotti for insightful exchanges, I. Tavernelli for his inputs on early versions of the manuscript, as well as J. VandeVondele for sharing his views on the original method.

5.10 References

[1] J. VandeVondele and U. Rothlisberger. "Estimating equilibrium properties from non-Hamiltonian dynamics". *J. Chem. Phys.*, **115**, (2001) 7859.

[2] M. E. Tuckerman, C. J. Mundy, and G. J. Martyna. "On the classical statistical mechanics of non-Hamiltonian systems". *Europhys. Lett.*, **45**, (1999) 149–155.

[3] M. E. Tuckerman, Y. Liu, G. Ciccotti, and G. J. Martyna. "Non-Hamiltonian molecular dynamics: Generalizing Hamiltonian phase space principles to non-Hamiltonian systems". *J. Chem. Phys.*, **115**, (2001) 1678.

[4] V. E. Tarasov. "Phase-space metric for non-Hamiltonian systems". *J. Phys. A*, **38**, (2005) 2145.

[5] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna. "A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble". *J. Phys. A*, **39**, (2006) 5629.

[6] G. S. Ezra. "On the statistical mechanics of non-Hamiltonian systems: the generalized Liouville equation, entropy, and time-dependent metrics". *J. Math. Chem.*, **35**, (2004) 29.

[7] J. D. Ramshaw. "Remarks on entropy and irreversibility in non-Hamiltonian systems". *Phys. Lett. A*, **116**, (1986) 110.

[8] D. J. Evans and G. P. Morriss. *Statistical Mechanics of NonEquilibrium Liquids*. Theoretical Chemistry Monograph Series (Academic Press, London, London, 1990).

[9] B. L. Holian, H. A. Posch, and W. G. Hoover. "Nonequilibrium free energy, coarse-graining, and the liouville equation". *Phys. Rev. A*, **42**, (1990) 3196.

[10] J. D. Ramshaw. "Remarks on non-Hamiltonian statistical mechanics". *Europhys. Lett.*, **59**, (2002) 319.

[11] B. L. Holian and W. G. Hoover. "Numerical test of the liouville equation". *Phys. Rev. A*, **34**, (1986) 4229.

[12] W. G. Hoover, D. J. Evans, H. A. Posch, B. L. Holian, and G. P. Morriss. "Comment on "toward a statistical thermodynamics of steady states"". *Phys. Rev. Lett.*, **80**, (1998) 4103.

[13] B. L. Holian, W. G. Hoover, and H. A. Posch. "Resolution of Loschmidt's paradox: The origin of irreversible behavior in reversible atomistic dynamics". *Phys. Rev. Lett.*, **59**, (1987) 10.

[14] B. L. Holian, G. Ciccotti, W. G. Hoover, B. Moran, and H. A. Posch. "Nonlinear-response theory of time-independent fields: Consequences of the fractal nonequilibrium distribution function". *Phys. Rev. A*, **39**, (1989) 5414.

[15] D. Ruelle. "Positivity of entropy production in nonequilibrium statistical mechanics." *J. Stat. Phys.*, **85**, (1996) 1.

[16] J. Jellinek and R. S. Berry. "Generalization of Nosé's isothermal molecular dynamics". *Phys. Rev. A*, **38**, (1988) 3069.

[17] A. Bulgac and D. Kusnezov. "Canonical ensemble averages from pseudocanonical dynamics". *Phys. Rev. A*, **42**, (1990) 5045.

[18] D. Kusnezov, A. Bulgac, and W. Bauer. "Canonical ensembles from chaos". *Ann. Phys.*, **204**, (1990) 155.

[19] B. B. Laird and B. J. Leimkuhler. "Generalized dynamical thermostating technique". *Phys. Rev. E*, **68**, (2003) 016 704.

[20] T. Terada and A. Kidera. "Generalized form of the conserved quantity in constant-temperature molecular dynamics". *J. Chem. Phys.*, **116**, (2002) 33.

[21] D. Brown and J. H. R. Clarke. "A comparison of constant energy, constant temperature and constant pressure ensembles in molecular dynamics simulations of atomic liquids". *Mol. Phys.*, **51**, (1984) 1243.

[22] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". *Phys. Rev. A*, **31**, (1985) 1695.

[23] G. J. Martyna, M. L. Klein, and M. Tuckerman. "Nosé–Hoover chains: The canonical ensemble via continuous dynamics". *J. Chem. Phys.*, **97**, (1992) 2635.

[24] W. G. Hoover. *Computational Statistical Mechanics*. Studies in Modern Thermodynamics (Elsevier Science, Amsterdam, 1991).

[25] C. Jarzynski. "A nonequilibrium equality for free energy differences". *Phys. Rev. Lett.*, **78**, (1997) 2690.

[26] G. M. Torrie and J. P. Valleau. "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling". *J. Comp. Phys.*, **23**, (1977) 187.

[27] W. F. van Gunsteren, T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith. "Computation of free energy in practice: choice of approximations and accuracy limiting factors". In: "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications", eds. W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, vol. 2 (Escom Science Publishers, Leiden, The Netherlands, 1993) 315.

[28] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". *Phys. Rev. E*, **60**, (1999) 2721.

Chapter 6

A steered molecular dynamics study of the TCR-pMHC complex

Summary

CD8+ T cells recognize peptides bound to major histocompatibility complex (MHC) molecules on the surface of antigen presenting cells. Under appropriate conditions, binding of a peptide-MHC complex (pMHC) to the T cell receptor (TCR) results in T cell activation. While the molecular mechanisms underlying T cell activation are not yet fully understood, it is clear that both the affinity and the association/dissociation kinetics of TCR-pMHC interactions are key parameters. In the present study, we look at two TCR-pMHC complexes which only differ by the mutation of a single amino acid on the peptide, and yet produce opposite T cell response. We use a novel steered molecular dynamics scheme adapted to the large size of the system to investigate the details of the TCR dissociation from the MHC. The Jarzynski identity is in particular applied to recover equilibrium dissociation free energy profiles from collections of nonequilibrium dedocking trajectories. The free energy results do not stand quantitative comparison with experiment or umbrella sampling simulations. However, the behaviors of the two wild type and mutated peptide are clearly differentiated by the simulations, which in addition provide valuable insights the TCR-pMHC dissociation mechanisms.

6.1 Introduction

6.1.1 The TCR-pMHC system

Recognition by the CD8+ T cell receptor (TCR) of antigenic peptides (p) presented by the major histocompatibility complex (MHC) class I molecules is the key step leading to T cell activation and, ultimately, target cell killing. The molecular events taking place at the TCR-pMHC interface are, therefore, subject of active research (see review in Ref. [1]). This system is of great interest for medical applications and, in particular, for cancer immunotherapy. The activity of the T cell in cytotoxic assays was shown [2, 3] to depend on the thermodynamic affinity of the TCR for the pMHC complex. Furthermore, TCR binding and unbinding kinetic constants also proved to be key factors of cytotoxicity [4, 5, 6, 7]. In the prospect of optimizing TCR sequences that can be genetically incorporated into patient's lymphocytes and confer tumor immunity [8], there is a strong need for techniques to characterize quantitatively the thermodynamic and kinetic behavior of any given TCR in association with any peptide and any MHC. In addition to the computation of global properties, molecular dynamics (MD) simulations allow to dissect details of the interaction at the atomic level, giving valuable insights into the molecular basis of TCR recognition, and guiding rational design.

The TCR-pMHC studied here is the human A6 TCR, specific for the HTLV-1 virus Tax nonapeptide (LLFGYPVYV) bound to HLA-A0201. The complex is shown in Fig. 6.1, in the bound and unbound state. Recognition of pMHC is mediated by the very specific complementarity determining region (CDR) loops [9] of the TCR. The Tax peptide is a strong agonist, i.e. it induces T cell activation at very low concentrations. The recognition mechanism is extremely sensitive, since a single mutation on the Tax peptide can turn it into a weak antagonist, which inhibits the T cell function instead of triggering it.

The A6-Tax-HLA-A0201 system was chosen because it has been extensively studied experimentally, together with four peptide mutants. In addition to X-ray structures [10, 11], experimental data on T cell activation response, kinetic constants from surface plasmon resonance (SPR) experiments [11] and equilibrium constants from sedimentation equilibrium analytical ultracentrifugation [12] are available for the Tax peptide and the mutants. In the present work, we focus on the Tax wild type and the P6A mutant, in which the proline residue in the sixth position is replaced by an alanine. This mutation, shown in Fig. 6.2, induces minute changes in the structure of the complex. No contact between the mutated proline residue and the TCR is apparent in the X-ray structure. The only significant difference between the TCR-pMHC structures with the wild type and the P6A-substituted Tax peptide is a packing defect consisting of an enlarged cavity partially filled with a bound water molecule[11]. However, the mutation is able to change the peptide from a strong agonist to an antagonist. The measured equilibrium constants correspond to free energy differences from bound to unbound of 34.5 kJ/mol for Tax and 22.6 kJ/mol for P6A. Kinetic measurements show that there is a free energy barrier with $\Delta G_{\text{off}} = 79.5$ kJ/mol and $\Delta G_{\text{on}} = 44.5$ kJ/mol for the Tax peptide. The kinetic constants for P6A could not be measured, but it is suspected that the free energy barrier is significantly lower [11].

The typical half life of a TCR-pMHC complex is of the order of seconds [13, 14], a time scale currently out of reach of standard MD simulations. Indirect methods have been used

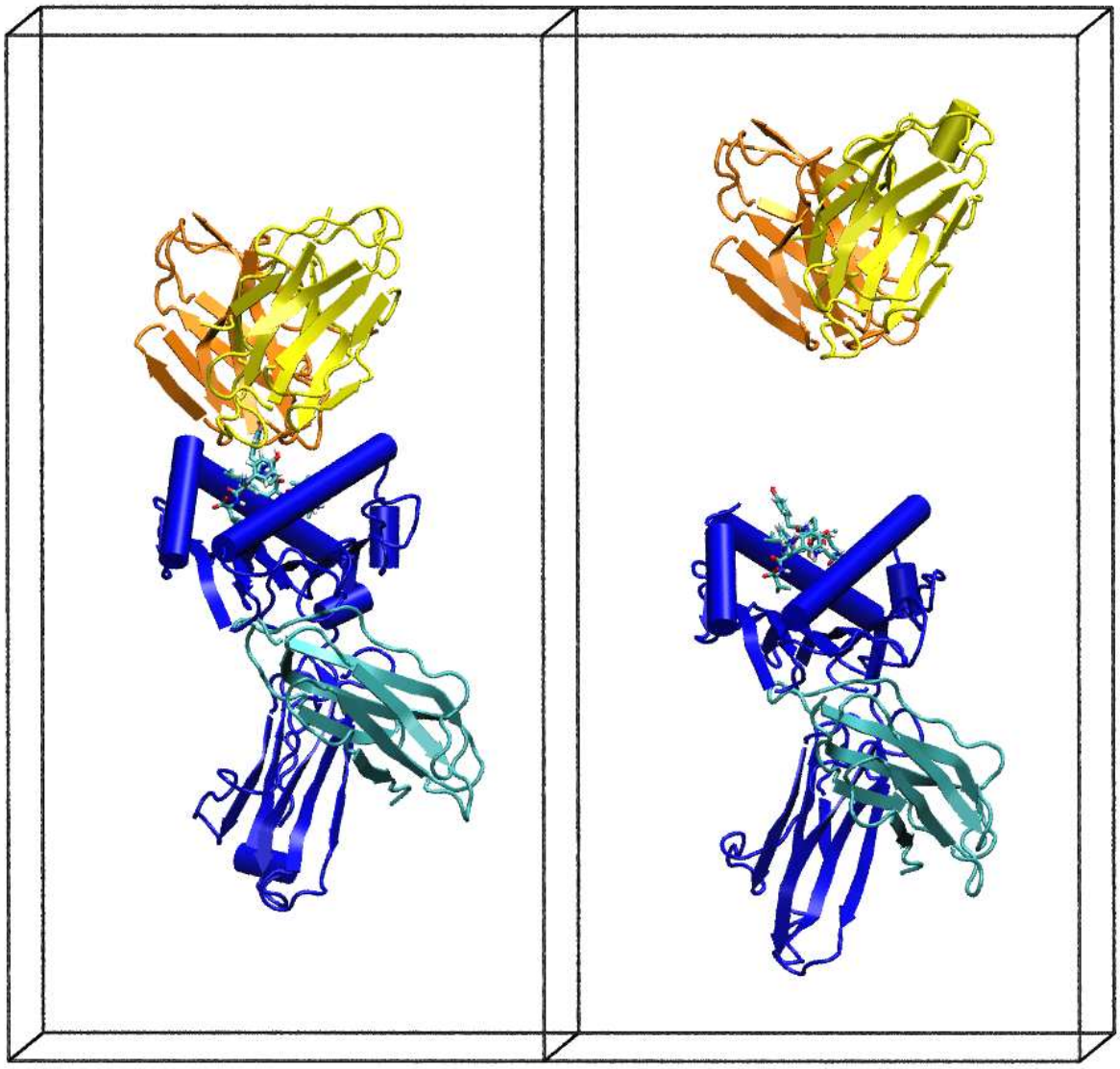

Figure 6.1: *The TCR-p-MHC system. Left panel: bound state at the beginning of the simulation. Right panel: final unbound state. Yellow and orange represent the α and β chains of the TCR. Dark blue is the α chain of the MHC, and light blue is the β2-microglobulin domain of the MHC. The Tax peptide is shown in all atoms licorice representation. The lines represent the simulation box. Water molecules are not shown for clarity. The entire system comprises 84'238 atoms.*

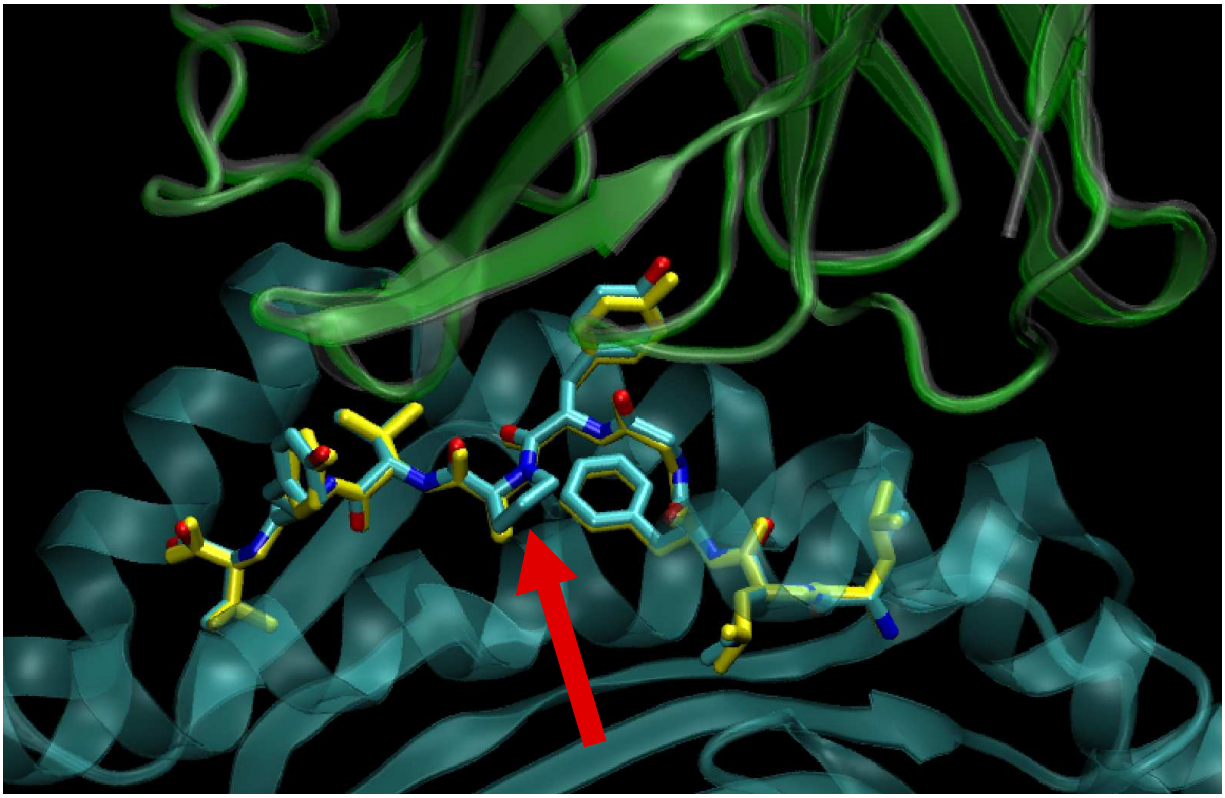

Figure 6.2: *Structural changes at the TCR-pMHC interface upon P6A mutation of the Tax peptide. The wild type peptide is represented with colors coding atom types, and the P6A is drawn in yellow. The red arrow shows the position of the mutation. The MHC groove is represented by the blue ribbon structure at the bottom. The TCR structure corresponding to the Tax and P6A peptides are shown in green and grey ribbon representation.*

to compute equilibrium free energy differences between bound states of mutants [15, 16]. These methods take advantage of Hess' law on the binding/mutation thermodynamic cycle, and use thermodynamic integration to calculate the mutation free energy in the bound and unbound states. On the other hand, when the entire dissociation potential of mean force (PMF) is wanted, various MD methods are available to circumvent the timescale limitation. Standard methods rely on series of equilibrium simulations restrained or constrained at different fixed positions along the reaction coordinate. In the present case the reaction coordinate is the distance between the center of mass (CM) of the pMHC and the CM of the TCR. In this study, one of the methods we use to assess the dissociation PMF is the well established umbrella sampling method [17]. On the other hand, a very recent class of methods rely on nonequilibrium dynamics, where the motion along the reaction coordinate is progressively driven by an external potential (targeted MD [18] or steered MD [19, 20, 21]). Dissipative work is performed on the system, and a thermostat couples the system to a heat bath which maintains the equilibrium temperature.

6.1.2 The Jazynski identity

The second law of thermodynamics states that the average work of a process cannot be smaller than the difference of free energies between the initial and the final states, $\Delta F \leq \langle W \rangle$. Equality holds if the process is quasi-static or reversible. In this case, the work is independent of the path of the transition between A and B. On the contrary, the Jarzynski identity (JI) [22] holds regardless of the speed of the process

$$e^{-\beta \Delta F} = \langle e^{-\beta W} \rangle, \tag{6.1}$$

where $\beta = 1/k_B T$. Now the work is path-dependent, and the average $\langle . \rangle$ is over different trajectories with independent canonically distributed initial conditions. This opens the possibility of calculating equilibrium free energy differences from nonequilibrium processes. A strong requirement is to have a sufficiently large collection of trajectories to allow an accurate estimation of the exponential average in (6.1), which is a major concern for practical applications.

The JI was first first reported [22] in 1997, and proved to apply to a variety of Hamiltonian and non-Hamiltonian dynamics [23], such as the Nosé-Hoover [24] thermostated dynamics, Markov-chain dynamics [25] or Langevin evolution [26]. The relation with the transient fluctuation theorem [27] was elucidated by Crooks [25] for stochastic systems and Jarzynksi [28] and Evans [29] for deterministic systems. An alternative derivation of 6.1 using path integrals and the Feynman-Kac theorem was proposed by Hummer and Szabo [19]. Park and Schulten [21] also demonstrated explicitly that the relation holds for the Gibbs free energy in an isothermal-isobaric ensemble. Recently the JI was derived [30, 31] in a purely statistical mechanical way for thermostated and barostated dynamical systems. Liphardt *et al.* [32] showed the relevance of the JI to experimental data as well.

6.1.3 Steered molecular dynamics

Suppose our system of interest is described by the Hamiltonian $H_0(\mathbf{q}, \mathbf{p})$ and that it is thermostated at temperature T. We would like to study the transition from state A to state B (e.g. the dissociation of two proteins), whose free energy barrier is several times $k_B T$. The probability of observing a spontaneous barrier-crossing event within a reasonable simulation time is extremely low. To overcome this, a perturbation is applied to the system to drive the reaction from A to B. This perturbation takes the form of a time-dependent external potential energy function u acting on a reaction coordinate ξ. The system is then described by the extended Hamiltonian

$$H_{\xi_0}(\mathbf{r}, \mathbf{p}) = H_0(\mathbf{r}, \mathbf{p}) + u(\mathbf{r}, \xi_0(t)). \tag{6.2}$$

The parameter ξ_0 is a pure function of time, and is varied such that the system is restrained in state A at time 0, and in state B at time τ. We denote $u(\mathbf{r}, t) = u(\mathbf{r}, \xi_0(t))$. If the reaction coordinate $\xi(\mathbf{r})$ is a function of atom positions, $u(\xi(\mathbf{r}), t)$ is called a steering function. The physical system described by $H_0(\mathbf{q}, \mathbf{p})$ is called the *intrinsic* system, while the perturbed system described by $H_{\xi_0}(\mathbf{r}, \mathbf{p})$ is called the *extended* or *composite* system.

The steering function is generally chosen harmonic and centered on a given reference reaction coordinate $\xi_0(t)$,

$$u(\mathbf{r},t) = \frac{k}{2}(\xi(\mathbf{r}) - \xi_0(t))^2. \tag{6.3}$$

Other forms of steering function are possible, as will be described later. The parameter $\xi_0(t)$ is changed typically with a constant velocity

$$\xi_0(t) = \xi_A + v_\xi t, \tag{6.4}$$

so that it covers the relevant region of ξ. The harmonic constant k and the steering velocity v_ξ are parameters of the simulation.

Let W_A be the external work done on the system from time 0 to τ,

$$W_A(\tau) = \int_0^\tau dt \frac{\partial H(\mathbf{r},\mathbf{p},\xi_0(t))}{\partial t} \; = \int_0^\tau dt \frac{\partial u(\mathbf{r},t)}{\partial t}. \tag{6.5}$$

It is denoted W_A according to [26] because it represents the work *accumulated* by the composite system. As pointed out in [26], $W_A(\tau)$ differs from the work *transferred* from the perturbation potential to the intrinsic system along the trajectory, which is instead given by the work of the perturbation force,

$$W_T(\tau) = \int_0^\tau f^u(\mathbf{r}_t, \xi_0(t)) \cdot d\mathbf{r}_t = -\int_0^\tau \frac{\partial u(\mathbf{r}_t,t)}{\partial \mathbf{r}_t} \cdot d\mathbf{r}_t. \tag{6.6}$$

The two forms of work are very different quantities. Suppose that the steering function is switched from $u(\mathbf{r},0)$ to its final form $u(\mathbf{r},\tau)$ in a time sufficiently short, so that the system coordinates remain practically unchanged. In this case, $W_T(\tau)$ necessarily vanishes, whereas $W_A(\tau)$ does not. In fact, $W_A(\tau)$ and $W_T(\tau)$ are both transferred works, but for different systems. $W_T(\tau)$ is the work transferred from the external perturbation $u(\mathbf{r},\xi_0)$ on the bare or *intrinsic* system $H_0(\mathbf{r},\mathbf{p})$. $W_A(t)$ is the work transferred from the surroundings to the perturbed or *composite* system defined by $H_0(\mathbf{r},\mathbf{p}) + u(\mathbf{r},t)$. The relation between $W_A(t)$ and $W_T(t)$ follows from $du = \frac{\partial u(\mathbf{r},t)}{\partial t}dt + \frac{\partial u(\mathbf{r},t)}{\partial \mathbf{r}}\dot{d\mathbf{r}}$. Rearrangement and integration from $(\mathbf{r}_0,0)$ to $(\mathbf{r},t)$ yields [19]

$$W_A(t) = W_T(t) + u(\mathbf{r},t) - u(\mathbf{r}_0,0). \tag{6.7}$$

6.1.4 Individual pulling

Special care has to be taken in the choice of the form of the steering function (6.3), and the choice of atoms to which it applies. Pulling of molecules is usually done by applying a force on one single atom [33, 34] to mimic an AFM experiment, in which case ξ is the distance between the pulled atom and a fixed atom. Another common way is to define ξ as the CM distance of a protein and a ligand, which means applying uniformly to each atom in the molecules a force proportional to its mass.

For big proteins with a rather loose structure and bound to a ligand by a strong interaction such as the TCR-pMHC complex, these approaches are not appropriate. Applying a single point force or a mass weighted uniform force on such a system can induce big

distortions in its tertiary structure, and partial unfolding. Additional problems arise if the interaction between the protein and its ligand is spread over a large surface perpendicular to the pulling direction. If during the pulling, one side of the interface unbinds while the other side remains bound, a force moment appears with respect to the CM. This results in a rotation of the protein and its ligand, which roll on each other instead of separating. When probing the interaction between two proteins normally attached to the membranes of two different cells, such big rotation movements seem unlikely.

In order to avoid these issues, we introduce a new scheme to actuate the dissociation, which we call *individual pulling*. It proceeds as follows. In an average structure of the bound complex, the reference position of an atom is determined with respect to the CM of its respective part (TCR or pMHC). A harmonic potential energy term centered on this reference position is then applied to the z coordinate of the atom, while lateral movements remain completely free. To achieve pulling, the reference positions of all restrained atoms are shifted uniformly along the z-axis, each part in opposite direction. During this process, the interatomic reference distances remain unchanged in each part, while the CM distance is increased. This scheme prevents big artifact changes in the tertiary structure of each part by distributing the pulling forces where reaction forces are strongest. In addition, this keeps the system well aligned during the unbinding, which is useful in an elongated simulation box.

The individual pulling is well suited to the TCR-pMHC system since the structures of both MHC and TCR in solution are known to be very similar to those in the bound state [13], except for the peptide and the very flexible [35] CDR loops. These parts are left unrestrained. In the remaining of the system, only backbone and C_α atoms are subjected to a steering function, while sidechains are left unrestrained. Favoring straight unbinding over hinge opening or rolling types of pathways is supported by evidence that the TCR-pMHC interaction actually happens between multimers in vivo [36, 37], and by the fact that both parts are attached to two parallel membranes.

We now formally define the individual pulling scheme. We note $\xi(\mathbf{r}_t)$ the instantaneous reaction coordinate, which is the actual distance along z between the CMs of the two parts $j = 1, 2$ described by r^1_{CM} and r^2_{CM},

$$\xi(\mathbf{r}) = |z^2_{CM} - z^1_{CM}|. \tag{6.8}$$

We will also use the notation shortcut $\xi(t)$. The corresponding driving parameter ξ_0,

$$\xi_0 = |\bar{z}^2_{CM} - \bar{z}^1_{CM}|, \tag{6.9}$$

sets the reference distance between two hypothetical positions of the CMs $\bar{z}^1_{CM}$ and $\bar{z}^2_{CM}$. For each part j, the reference structure determined before the pulling is expressed as a set of internal coordinates $\{\bar{\zeta}^j_i\}^N_{i=1}$, which denote the reference distance (along the z-axis) of each atom i to the reference CM $\bar{z}^j_{CM}$ of part j. The $\{\bar{\zeta}^j_i\}$ are constant during the entire pulling experiment, and allow to define the reference position $\bar{z}_i$ of atom i,

$$\bar{z}_i = \bar{z}^j_{CM} + \bar{\zeta}^j_i. \tag{6.10}$$

The individual restraint of atom i is then defined as

$$u_i(z_i, t) = \frac{k_i}{2}(z_i - \bar{z}_i(t))^2 \tag{6.11}$$

The individual force constant k_i is defined such that two requirements are met. First, the sum of all individual forces $f_i = -k_i(z_i - \bar{z}_i)$ in each part satisfies the action-reaction principle along the z-axis,

$$\sum_{i \in \text{part1}} f_i = - \sum_{i \in \text{part2}} f_i. \tag{6.12}$$

Second, we want that the total center of mass perturbation energy in each part resulting from all individual restraints is equivalent to a standard steering potential energy,

$$u_{\text{CM}}(\xi(\mathbf{r}), \xi_0) = \frac{k_{\text{CM}}}{2}(\xi(\mathbf{r}) - \xi_0)^2, \tag{6.13}$$

where k_{CM} is the equivalent harmonic constant driving the reaction coordinate. In this way, the individual pulling scheme can easily be compared to standard CM steered MD. These two requirements are met by setting

$$k_i^j = \frac{m_i}{M_j} k_{CM}, \tag{6.14}$$

with M_j the mass of part j. With this choice of k_i^j, the sum of the internal forces vanishes in each part.

During the simulation, $\xi_0(t)$ and the system's instantaneous reaction coordinate $\xi(r)$ set the overall steering force f_{CM}. For a given f_{CM}, which depends only on the relative distance $\xi(r)$, we have to determine the individual forces f_i, in terms of the absolute atomic coordinates z_i. In other words, we have to set the absolute reference positions $\bar{z}_{CM}^{1,2}$ at a given time step, given ξ_0 and $\xi(\mathbf{r})$. The $\bar{z}_{CM}^{1,2}$ are determined by attributing a half of $(\xi(\mathbf{r}) - \xi_0)$ to each part. From these $\bar{z}_{CM}^{1,2}$, all the $\bar{z}_i$ are found with (6.10), and the forces on individual atoms can be computed. In the end, we have of course

$$u(\mathbf{r}, \xi_0) = \sum_{i=1}^{N} u_i(z_i, \xi_0), \tag{6.15}$$

$$W_A(t) = \sum_{i=1}^{N} W_i(t), \tag{6.16}$$

with $W_i(t)$ the accumulated work performed by the individual restraint $u_i(z_i, t)$, defined similarly to (6.5).

Upon conformational changes, the internal potential energy $u_{\text{int}}(\mathbf{r}, \xi_0)$ is accumulated in the individual restraint energy terms of each part. The individual restraints have been constructed in such a way that

$$u_{\text{int}}(\mathbf{r}, \xi_0) = u(\mathbf{r}, \xi_0) - u_{\text{CM}}(\xi(\mathbf{r}), \xi_0). \tag{6.17}$$

Considering that important conformational changes happen on relatively long time scales, and supposing the fluctuations of the internal component of the steering function are fast with respect to the CM motion, we can make the hypothesis that for a given trajectory the internal potential energy is only relevant on average as

$$\tilde{u}_{\text{int}}(\Xi, \xi_0) = \int \delta\left[\xi(\mathbf{r}) - \Xi\right] u_{\text{int}}(\mathbf{r}, \xi_0)\ e^{-\beta H(r, \xi_0)} d\mathbf{r}. \tag{6.18}$$

6.1.5 Obtaining a PMF from steered dynamics with the Jarzynski method

Considering a reaction from state A to state B, we are interested not only in the free energy difference between those two states, but also in the free energy profile of the transition. In the canonical ensemble, the potential of mean force (PMF) $\Delta F_0(\Xi)$ is the Helmholtz free energy profile corresponding to the integrated state density $\langle\rho(\Xi)\rangle$ along the reaction coordinate Ξ [38],

$$\langle\rho(\Xi)\rangle = e^{-\beta\Delta F_0(\Xi)} = \int d\mathbf{r}d\mathbf{p}\ \delta[\xi(\mathbf{r}) - \Xi]\exp\{-\beta H_0(\mathbf{r},\mathbf{p})\}. \tag{6.19}$$

Note that the PMF (6.19) of interest is a property of the intrinsic system. On the other hand, applying Jarzynski's equality (6.1) directly to steered MD provides the free energy change $F_\Xi^C - F_0^C$ of the composite system,

$$e^{-\beta(F_\Xi^C - F_0^C)} = \langle\exp\{-\beta W_A(t)\}\rangle_0. \tag{6.20}$$

Here, $W_A(t)$ is the work done on the composite system, as defined in (6.5). The average $\langle\cdot\rangle_0$ is taken over the ensemble of trajectories in initial states $(\mathbf{r}_0,\mathbf{p}_0)$ sampled from the canonical ensemble corresponding to the composite Hamiltonian $H_{\xi_A}(\mathbf{r},\mathbf{p})$. In order to obtain the PMF for the intrinsic system, we have to find efficient ways to perform three operations :

(i) Estimate the exponential average $\langle e^{-\beta W}\rangle$.

(ii) Reduce from the composite to the intrinsic system (unbias).

(iii) Average over all $\xi(\mathbf{r})$ visited during the evolution to find ΔF_0 as a function of Ξ.

Hummer and Szabo [19] showed that

$$e^{-\beta(\Delta F_0(\Xi) - F_0^C)} = \big\langle\delta[\xi(\mathbf{r}) - \Xi]\exp\{-\beta(W_A(t) - u\,(\mathbf{r},\xi_0))\}\big\rangle. \tag{6.21}$$

This corresponds to points (ii) and (iii) above. The Weighted Histogram Averaging Method (WHAM) [39, 40] is an efficient way to estimate the average (6.21) from steered MD time series [41, 19, 42] This approach has been used to estimate free energy profiles in AFM experiments [19] as well as computer simulations [43, 44]. However, the direct use of (6.21) is problematic when the free energy barrier that the system has to overcome is high. Indeed, the exponential average $\langle e^{-\beta W}\rangle$ is dominated by small work values that arise only very seldom. Let $P(W)$ be the probability distribution of the work, which can typically be approximated by a Gaussian with a standard deviation σ. Then $P(W)e^{-\beta W}$ has a normal distribution of same width, but with its peak shifted by $\beta\sigma^2$ towards smaller values of W. Most work values are sampled around the peak of $P(W)$. If σ is large, the region around the peak of $P(W)e^{-\beta W}$ is not properly sampled, which results in a bias in the estimation of the exponential average. In practice, the estimator (6.21) can be used only when the work fluctuation σ is not much larger than k_BT.

In the present study we use a hybrid averaging scheme. We first perform operations (ii) and (iii) on separate trajectories, using a modified WHAM which takes the internal bias potential energy into account. This gives for each trajectory a work profile $\bar{W}_i(\Xi)$ corresponding to the intrinsic system. In a second step, the average $\langle e^{-\beta \bar{W}(\Xi)} \rangle$ is estimated using the cumulant expansion method [22, 45, 21], which provides an unbiased statistical estimate of the exponential average. In the two next paragraphs, we give a detailed description of the modified WHAM and the cumulant expansion methods.

Note that points (ii) and (iii) above equally apply to other free energy methods. Among them, the umbrella sampling method [17] is a standard method to calculate PMFs, and we use it here for comparison with the JI. Umbrella sampling consists in running several equilibrium simulation at a series of restrained positions along the reaction coordinate. Each simulation gives a piece of PMF for the composite system, which is biased by the restraining or umbrella potential. Pieces of overlapping PMFs need to be unbiased and spliced together, corresponding to operations (ii) and (iii) above. The WHAM method is appropriate to do these two operations also in the case of umbrella sampling.

6.1.6 A modified weighted histogram method

The PMF obtained in steered MD around a given position Ξ is biased, since it pertains to the composite system $H_0(\mathbf{r}, \mathbf{p}) + u(\mathbf{r}, \xi_0)$, instead of the intrinsic system. From the corresponding biased distribution function $\rho^{\text{biased}}_{\xi_0}(\Xi)$, the unbiased one can be obtained (operations (*ii*) and (*iii*) above) by [38]

$$\begin{aligned}
\rho_{\xi_0}(\Xi) &= \int \delta\left[\xi(\mathbf{r}) - \Xi\right] e^{+\beta u(\mathbf{r},\xi_0)}\ \rho^{\text{biased}}_{\xi_0}(\Xi)\ e^{+\beta F_{\xi_0}}\ e^{-\beta H(\mathbf{r},\xi_0)} d\mathbf{r} \\
&= e^{+\beta u_{\text{CM}}(\Xi,\xi_0)}\ \rho^{\text{biased}}_{\xi_0}(\Xi)\ e^{+\beta F_{\xi_0}} \\
&\quad \times \int \delta\left[\xi(\mathbf{r}) - \Xi\right]\ e^{+\beta u_{\text{int}}(\mathbf{r},\xi_0)}\ e^{-\beta H(r,\xi_0)} d\mathbf{r} \qquad (6.22) \\
&= e^{+\beta u_{\text{CM}}(\Xi,\xi_0)}\ e^{+\beta \tilde{u}_{\text{int}}(\Xi,\xi_0)} \rho^{\text{biased}}_{\xi_0}(\Xi)\ e^{+\beta F_{\xi_0}}. \qquad (6.23)
\end{aligned}$$

Here, the values of $\tilde{u}_{\text{int}}(\Xi, \xi_0)$ are empirically defined such that $e^{+\beta \tilde{u}_{\text{int}}(\Xi,\xi_0)} = \left\langle e^{+\beta u_{\text{int}}(\mathbf{r},\xi_0)} \right\rangle$ in bins of $\xi(\mathbf{r})$ around values of Ξ. The undetermined constant F_{ξ_0} defined

$$e^{-\beta F_{\xi_0}} = \langle -\beta u(\mathbf{r}, \xi_0) \rangle, \qquad (6.24)$$

represents the free energy associated with the introduction of the biasing potential function. In WHAM [39], the total PMF at Ξ is determined as the sum of all unbiased PMF parts, weighted according to the distance between Ξ and their reference point ξ_0 [46], and the number n_{ξ_0} of sample points in window ξ_0,

$$\langle \rho(\Xi) \rangle = \sum_{\xi_0} \rho_{\xi_0}(\Xi) \frac{n_{\xi_0} e^{-\beta\left(u_w(\Xi,\xi_0) - F_{\xi_0}\right)}}{\sum_{\xi_0'} n_{\xi_0'} e^{-\beta\left(u_w(\Xi,\xi_0') - F_{\xi_0'}\right)}}. \qquad (6.25)$$

Here, the weighting mediated by the function $u_w(\Xi, \xi_0)$, which in regular WHAM is harmonic, with the same harmonic constant as the biasing function. In the present case, at

least two choices arise for u_w. The most immediate is $u_w = u_{\mathrm{CM}} + \tilde{u}_{\mathrm{int}}$, which means that the weighting is done with the same function as the steering. The other choice is to take only the CM distance part for the weighting, $u_w = u_{\mathrm{CM}}$, which omits the noisy internal potential energy. In practice, this second choice turns out not to converge properly, so we focus on the first. Combining equations (6.23) and (6.25) yields

$$\langle \rho(\Xi) \rangle = \frac{\sum_{\xi_0} n_{\xi_0} \rho_{\xi_0}^{\mathrm{biased}}(\Xi)}{\sum_{\xi_0} n_{\xi_0} e^{-\beta\left(u_{\mathrm{CM}}(\Xi,\xi_0)+\tilde{u}_{\mathrm{int}}(\Xi,\xi_0)-F_{\xi_0}\right)}} \tag{6.26}$$

$$e^{-\beta F_{\xi_0}} = \int d\Xi \; e^{-\beta(u_{\mathrm{CM}}(\Xi,\xi_0)+\tilde{u}_{\mathrm{int}}(\Xi,\xi_0))} \langle \rho(\Xi) \rangle. \tag{6.27}$$

This constitutes the specific WHAM system of equations for individual restraints, which must be solved iteratively. In the case of umbrella sampling, an initial guess for F_{ξ_0} is provided by estimating offsets which make the PMF of each ξ_0 overlap approximately.

In the case of steered MD, the picture is slightly different, since ξ_0 sweeps the whole range of the reaction coordinate during one trajectory, instead of being fixed. Accordingly, $\xi(\mathbf{r})$ coming from a single trajectory is also distributed over the whole range. Histograms are collected in bins located at fixed Ξ, and data coming from different trajectories are summed up indifferently. According to the JI, we define

$$\rho_{\xi_0}^{\mathrm{biased}}(\Xi) = \frac{\langle \delta[\xi(\mathbf{r}) - \Xi] \; e^{-\beta W_A(\xi_0)} \rangle_0}{\langle e^{-\beta W_A(\xi_0)} \rangle_0}, \tag{6.28}$$

and the initial guess for the offset of window ξ_0 is $e^{-\beta F_{\xi_0}} = \langle e^{-\beta W_A(\xi_0)} \rangle_0$. In addition, the number of points n_{ξ_0} in a window is considered approximately constant, and is taken out of the weighting term. Inserting these modifications in (6.27) yields an expression similar to the adapted WHAM of Hummer and Szabo [19, 42] with the addition of the $\tilde{u}_{\mathrm{int}}(\Xi, \xi_0)$ term. This expression can be used directly to find $\Delta F_0(\Xi)$ from simulations in which the dissipative part of the steering work is small and the direct estimation of the exponential average is possible.

In the present case, we leave the inter-trajectory averaging (i.e. operation (i)) for the next step, namely the cumulant expansion method. We nonetheless use the modified WHAM on each trajectory i separately to find unbiased and smoothed work profiles $\bar{W}_i(\Xi)$ (i.e. operations (ii) and (iii)). The corresponding WHAM equations are

$$\bar{W}_i(\Xi) = \frac{\sum_{\xi_0} \langle \delta[\xi(\mathbf{r}) - \Xi] \; e^{-\beta W_i(\xi_0)} \rangle_0 \; e^{+\beta F_{\xi_0}}}{\sum_{\xi_0} e^{-\beta(u_{\mathrm{CM}}(\Xi,\xi_0)+\tilde{u}_{\mathrm{int}}(\Xi,\xi_0))} \; e^{+\beta F_{\xi_0}}} \tag{6.29}$$

$$e^{-\beta F_{\xi_0}} = \int d\Xi \; e^{-\beta(u_{\mathrm{CM}}(\Xi,\xi_0)+\tilde{u}_{\mathrm{int}}(\Xi,\xi_0))} \; e^{-\beta \bar{W}_i(\Xi)}. \tag{6.30}$$

The evaluation of the numerator in the above equation consists in building a histogram of $e^{-\beta W_i(\xi_0)}$ with bins of width Δ_Ξ centered on positions Ξ. In the single trajectory case, the one-step estimation using the first guess $e^{-\beta F_{\xi_0}} = e^{-\beta W_A(\xi_0)}$ allows a simplification in the numerator,

$$\bar{W}_i(\Xi) = \frac{\sum_{\xi_0} \Theta_\Xi \left[\xi(\mathbf{r})\right]}{\sum_{\xi_0} e^{-\beta(u_{\mathrm{CM}}(\Xi,\xi_0)+\tilde{u}_{\mathrm{dist}}(\Xi,\xi_0))} \; e^{+\beta W_i(\xi_0)}}. \tag{6.31}$$

The indicator function Θ_Ξ is one if $\xi(\mathbf{r})$ is in the interval $[\xi - \Delta_\Xi, \xi + \Delta_\Xi]$, and zero otherwise.

6.1.7 The cumulant expansion method

To bypass the direct estimation of an inter-trajectory exponential average (operation (i)), the cumulant expansion method is often used [22, 47, 20, 21]. The idea is to expand the logarithm of the exponential average in terms of cumulants,

$$\Delta F_0(\Xi) = -\frac{1}{\beta}\ln\langle \bar{W}_i(\Xi)\rangle = \sum_{i=1}^{\infty} C_i \frac{(-\beta)^i}{i!}. \tag{6.32}$$

The first two cumulants are $C_1 = \langle \bar{W}\rangle$ and $C_2 = \sigma^2_{\bar{W}}$, the average and the variance of the work distribution. The second order expansion reads

$$\Delta F_0(\Xi) = \langle \bar{W}\rangle - \frac{\beta}{2}\sigma^2_{\bar{W}}. \tag{6.33}$$

Marcinkiewicz's theorem[48] states that if and only if the work distribution is Gaussian, all but the first two cumulants in the expansion vanish. In all other cases, there are an infinite number of non-vanishing cumulants. This property is of relevance here, because higher moments are difficult to estimate from limited samples. If the work distribution is close to Gaussian, the second-order expansion provides an accurate estimation. There are a couple of observations pointing in that direction. First, in near-equilibrium regimes the work distribution approaches a Gaussian. In this perspective[45], the JI reduces to predating near-equilibrium results[49] similar to (6.33), which are valid in the linear response regime, and are related to the fluctuation-dissipation theorem. Second, the Gaussian character of the work distribution, even far from equilibrium, is favored by the choice of a stiff steering function [21].

Both estimators $\langle \bar{W}\rangle$ and $\sigma^2_{\bar{W}}$ are statistically unbiased. Assuming that we have n work values drawn from a Gaussian distribution, the statistical errors can be estimated from the expected errors on the mean and variance. For this we need to further assume that the statistical errors on the mean and variance are uncorrelated.

$$\mathrm{var}(\Delta F_{\xi_0(t)}) = \frac{\sigma^2}{n} + \frac{\beta\sigma^4}{n-1}. \tag{6.34}$$

When using a truncated cumulant expansion (6.32), two kinds of error are involved: the error due to neglecting higher order terms, and the error due to the finite sampling of each term. By averaging directly the exact formula 6.21, we have no truncation error, but we have seen that the sampling error can become prohibitive. On the contrary, low-order cumulants are easier to estimate from a finite data set, because they are unbiased.

6.1.8 Simulation setup

The wild type A6-Tax-HLA-A0201 X-ray structure [10] was taken from the 1AO7 entry of the Protein Data Bank, and the P6A mutant structure[11] from the 1QRN entry. To

reduce the system size, only the V_α and V_β domains of the TCR are included in the simulation. This is justified by experimental evidence [50] that the V_α and V_β domains alone keep the same structure and reactivity with the MHC as in the full TCR molecule. The resulting model includes a total of 605 residues. The proteins are modelled with the GROMOS 43A1 force field [51] and SPC water [52] with constrained [53] bonds and angle. Explicit water molecules are necessary for an accurate description of buried surface solvation upon unbinding, including water bridges and entropic effects due to the liberation of trapped water molecules. Indeed, a number of bound water molecules have been resolved at the TCR-pMHC interface [10, 11]. In addition, a detailed molecular dynamics study evidenced [54] the important role of some water molecules at the TCR-pMHC interface. All water molecules resolved in the crystal structures are kept in their original initial position as the protein is inserted in pre-equilibrated water. 9 Na^+ counterions are added to balance the net charge of the system. The simulation box is chosen such that the CMs of the MHC and the TCR are aligned along the z axis. A mild restraint of 100 kJ $mol^{-1}nm^{-2}$ in the (x, y) plane is applied to the CMs of the MHC and the TCR to keep them aligned along the z-axis. This auxiliary restraint is perpendicular to the reaction coordinate and thus never accumulates work and does not interfere with the free energy measurement. The protein is surrounded by at least 1.2 nm of water on each of the x and y sides of the box. An additional 1.6 nm is left in the z direction, to allow space for the dissociation. This results in 25990 water molecules in a $7.6 \times 8.1 \times 14.4$ nm box. We use the single precision version of the gromacs 3.3 MD package [55, 56, 57], modified for individual atom pulling. Long-range electrostatic interactions are treated with the particle mesh Ewald (PME) summation [58]. A twin-range cutoff scheme is used for the Lennard-Jones interactions in which interactions are calculated every step up to a cutoff distance of 0.8 nm and every 5 steps up to 1.4 nm. All covalent bonds involving hydrogens are constrained at their ideal lengths using SETTLE [59], a version of the SHAKE [60] algorithm. The protein and solvent are coupled to two separate Nosé-Hoover [24] thermostats with time constant 0.1 ps. The pressure is controlled by a Parrinello-Rahman [61, 62] barostat, with time constant 0.5 ps. This combination of thermostat and barostat should preserve the correct NPT ensemble.

The system is prepared as follows. Random velocities are first assigned according to a Maxwell distribution at 100 K. Then the system is heated up to 300 K in 100 ps at constant volume, and with atoms of the backbone and side chains up to the C_β carbon restrained to the crystal positions. A stable RMSD of 0.197 nm on backbone atoms is observed with respect to the crystal structure after 800 ps of equilibration.

The protocol of a standard Jarzynski simulation involves a long reference equilibrium simulation with the initial reaction coordinate. Decorrelated configurations with velocities are then chosen from this equilibrium trajectory as starting points for the pulling runs. When a computer grid is used, calculating the reference trajectory is very inefficient. Instead, we use the following protocol for a distributed computing of the steered dedocking simulations. Different replicas are started on independent processors at the heating up phase described above, and different initial random velocities are assigned to each of them. After a temperature of 300K is reached, 700 ps of equilibration are performed on each replica. This protocol has two advantages over the standard one, besides the practical aspect related to grid computing. First, the structure at the end of each short

equilibration cannot have deviated much from the crystal structure, as it could happen after one extremely long trajectory, due to MD inaccuracies. Second, the diversity of structures obtained at the end of the independent equilibrations is much bigger that the diversity of a sample of consecutive frames of one long equilibration [63]. This is likely to provide a better convergence of the average in (6.1).

The last 200 ps of each equilibration run are used to compute the average CM distance ξ_A, as well as the average internal coordinates $\{\bar{\eta}_i^j\}_{i=1}^N$. This average structural information is specific to each trajectory, and is used as initial distance and reference structure for the individual restraints (6.11). Next, the individual restraints are applied on backbone atoms of the regions discussed above with a resulting harmonic constant of $k_{CM} = 2 \cdot 10^4$ kJ mol^{-1}nm^{-2}. This k_{CM} is chosen stiff enough to have a good spatial resolution in the free energy profile and to keep the work distribution Gaussian [21], while not damping the thermal behavior of the system. The corresponding harmonic constant k_i (6.14) on one carbon atom is 69 kJ mol^{-1}nm^{-2}, which allows for a typical displacement of 0.27 nm at k_BT. This shows that in addition to being totally free in the (x, y) plane, local parts of the protein backbone are allowed to relax in the z direction as well, in spite of the structural restraints,. The composite system is then further equilibrated during 300 ps with ξ_0 fixed at ξ_A. Finally, the pulling itself is started. A pulling speed of $v = 5 \cdot 10^{-4}$ nm/ps is chosen as a tradeoff between staying as close to equilibrium as possible, and the computer resources at hand. With these settings, 2 nm are covered in 4 ns of simulation. One steering trajectory including equilibration as described above requires about 6 weeks of calculation on a 1.3 GHz Itanium2 processor. The simulation protocol for the P6A mutant is exactly the same.

In order to compare the Jarzynski with a more standard method to compute a PMF, we have performed umbrella sampling calculations. One first calculation uses the same individual restraints as in the Jarzynski setup, and a second calculation uses standard CM restraints, which apply uniform forces on the atoms. For the umbrella sampling runs, initial frames were taken from a preliminary 40 ns slow dedocking trajectory. For the runs subjected to individual restraints, 112 windows were chosen at ξ_0 intervals of approximately 0.013 nm, covering the reaction coordinate from 0 to 1.45 nm. For the runs restrained by a usual potential energy function on the CMs, only 48 windows were used in the same interval. Sampling was accumulated during 3.2 to 3.6 ns in the individual restraint case, and during 2 ns in the CM case. Finally, the modified WHAM defined in expression (6.27) was used to splice and unbias the individual restraint PMF, and the standard WHAM of [39, 40] was used for the CM case.

6.2 Results and discussion

6.2.1 Steered molecular dynamics

The work profiles obtained for 152 pulling trajectories of the TCR-Tax-MHC complex are shown in Fig. 6.3. For each trajectory i independently, values of $\bar{W}_i$ have been smoothed and unbiased by applying the iterative modified WHAM (6.29). In the first 0.4 nm of displacement, the resulting work curves undergo a steep increase as most hydrogen bonds and van der Waals contacts are broken. At 1 nm and further, when all interactions have

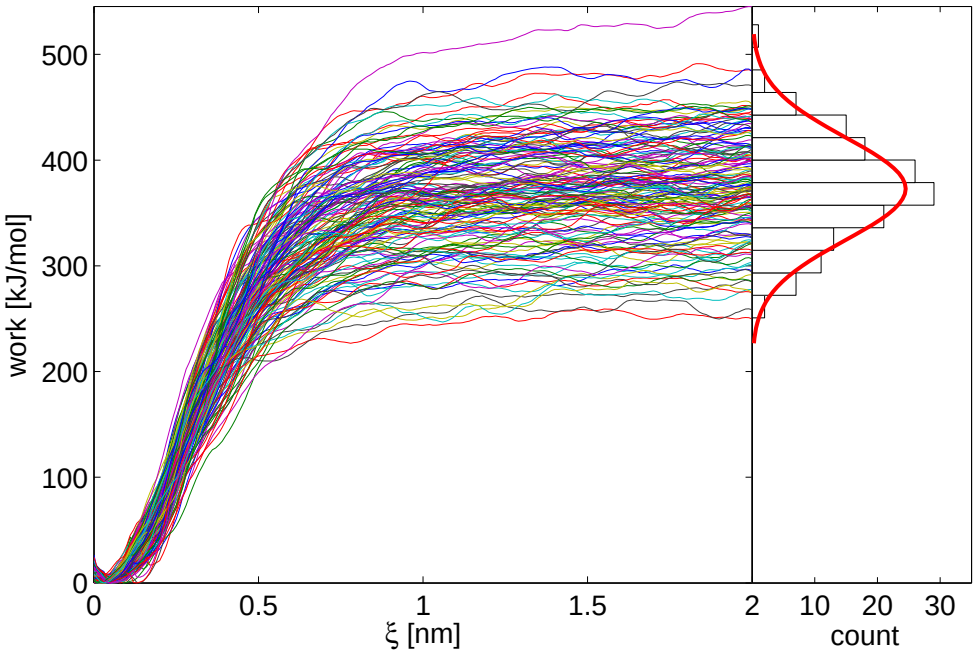

Figure 6.3: *Collection of 152 work profiles along the reaction coordinate, smoothed and unbiased using (6.29). The right panel shows an histogram of this distribution at $\xi = 1.7$ nm, with the fitted Gaussian used in the cumulant expansion method.*

vanished, a plateau value is reached. At all distances, the work distribution is consistent with a Gaussian distribution according to a 5% Kolmogorov-Smirnov test (data not shown). Indeed, the Gaussian character of the work distribution even far from equilibrium is favored by the choice of a stiff steering potential energy function[21].

Fig. 6.4 shows the evolution of the steering potential energy along the reaction coordinate. Two contributions are shown. First, $u_{CM}(\xi(\mathbf{r}), \xi_0)$ directly reflects the resistance of the system to the dissociation force. It increases sharply to a maximum at approximately 0.2 nm distance at which most contacts have to be broken. Then $u_{CM}(\xi(\mathbf{r}), \xi_0)$ decreases again, to reach zero approximately at a distance of 1 nm, at which all interactions have faded (see Figs. 6.7, 6.8, and 6.9). The Tax peptide induces a higher steering potential energy than the P6A mutant, indicating that the TCR dissociation energy barrier is higher (with this steering velocity). The second contribution to the steering potential energy is $u_{\text{int}}(\mathbf{r}, \xi_0)$, which is related to the internal rearrangement of each part. We first observe a sharp increase due to the fact that atoms far from the TCR-pMHC interface are free to follow the steering restraint and atoms close to the interface are pulled back by the attraction. This increase may also be related to a propensity of the TCR and the pMHC to rotate with respect to each other (see Section 6.1.4), which is prevented by the individual restraints. Beyond 1.0 nm, $u_{\text{int}}(\mathbf{r}, \xi_0)$ continues to increase, due to slow internal rearrangement. Because no strong external force acts on the proteins at this point in the

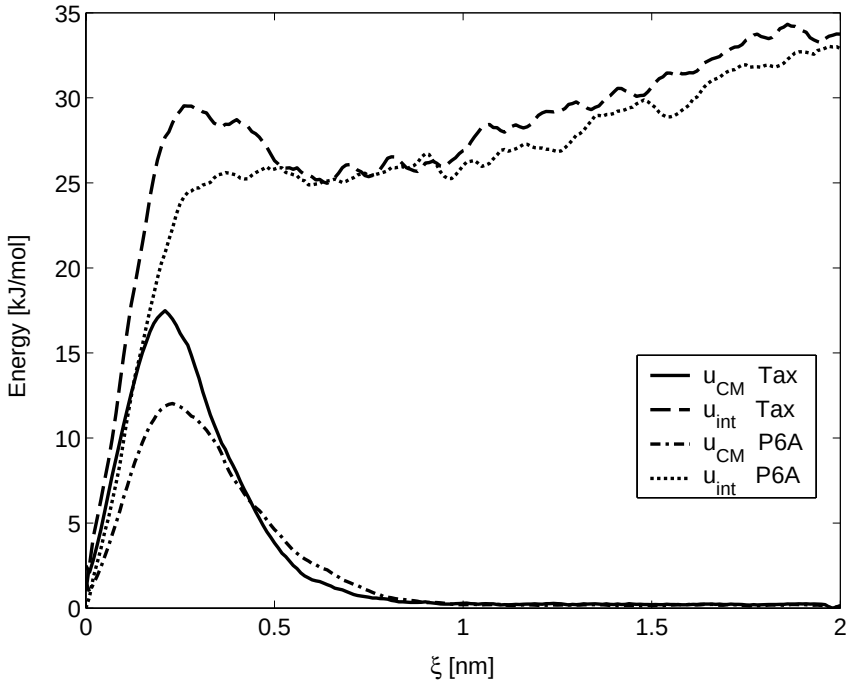

Figure 6.4: *Contributions to the steering potential energy $u(\boldsymbol{r}, \xi)$ along the reaction coordinate ξ, for the Tax and the P6A mutant complexes. Red lines show the part $u_{\mathrm{CM}}(\xi(\boldsymbol{r}), \xi)$ (6.13), which depends on the CM distance. Black lines show $u_{\mathrm{int}}(\boldsymbol{r}, \xi)$ (6.17), which corresponds to internal rearrangements in the TCR and pMHC.*

process, this shows that the system is able to rearrange by itself in spite of the individual restraints.

In order to better understand the dedocking process of the TCR and the pMHC, it is of interest to assess the existence of a funnel in space, along which the complex would preferably dedock. If this was the case, trajectories with the lowest accumulated work would be clustered in some regions. In Fig. 6.5, the final work of each trajectory is plotted as a function of several parameters. These include initial and final x, y and z relative CM positions of each part, as well as initial and final dedocking angle θ. We define θ as the angle between a reference vector linking residues 84 and 170 of the MHC, and the vector connecting the CMs of the two TCR chains. θ characterizes the orientation of the TCR relative to the MHC in the (x, y) plane. No correlation of the final work with any of these initial or final positional parameters is observed.

Fig. 6.6 offers a view of the ensemble of positions obtained after dedocking. Each trajectory is represented by 3 connected segments. The first segment, located in the MHC grove, connects the two reference residues mentioned above. We see that after superposing the MHC α_1 and α_2 chains coming from the last frame of each trajectory, the orientation of this reference segment is well preserved. The long segment perpendicular to the MHC surface links the CM of the the MHC α_1 and α_2 chains to the CM of the TCR. The

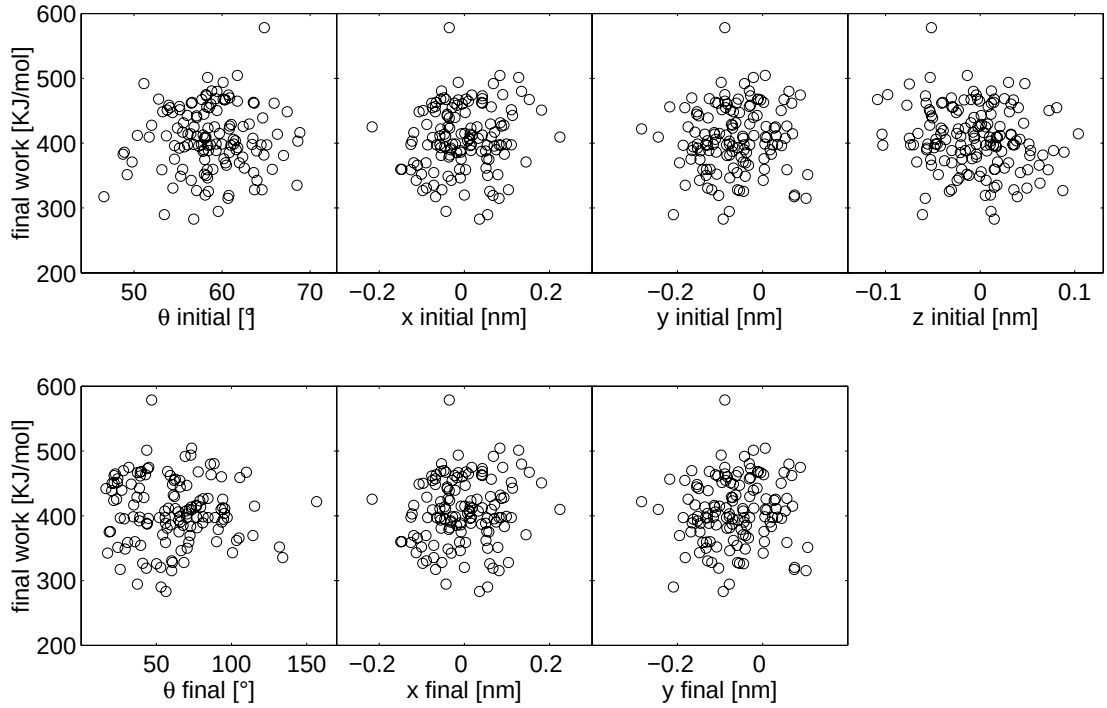

Figure 6.5: *Final work values as a function of initial and final TCR positions with respect to the pMHC.*

segments in the foreground connect the CM of the whole TCR to the CM of the TCR α chain (in orange in Fig. 6.1). The color code red corresponds to a high final work value, and blue to a low final work value. We see, in agreement with Fig. 6.5 that there is no correlation between final position or orientation and the work.

Next, we look at the average electrostatic interaction between the TCR and the pMHC. The protein coordinates were extracted from configurations taken at 200 ps time intervals from each steered trajectory, which corresponds to 0.1 nm intervals of ξ_0. For each of these configurations, the solvent was stripped away and the Coulomb interaction energy was calculated without cutoff in a non-periodic setup. A relative permittivity of 54.0 was reported [64] for SPC water. However, this value is not expected to hold for charge-charge interactions at short distance and in the structured first solvation layers around the protein. Given the fact that an SPC water molecule is itself not polarizable, one expects charge dielectric effects to appear only at distance allowing several water molecules to reorient. Therefore, a relative permittivity of 1 was retained, as a first order measure. For all configurations at a given time point, the electrostatic energy was then averaged over all trajectories. The separate contributions of the Tax peptide and the MHC itself were computed, and are shown in Fig. 6.7. It appears that the contribution of the peptide to the total interaction Coulomb energy is negligible. We see also that the interaction energy reaches a plateau after 1.7 nm. It is slightly repulsive at long distances, probably due to the fact that the MHC has a total charge of -8 e and the TCR of -1 e. This long-distance

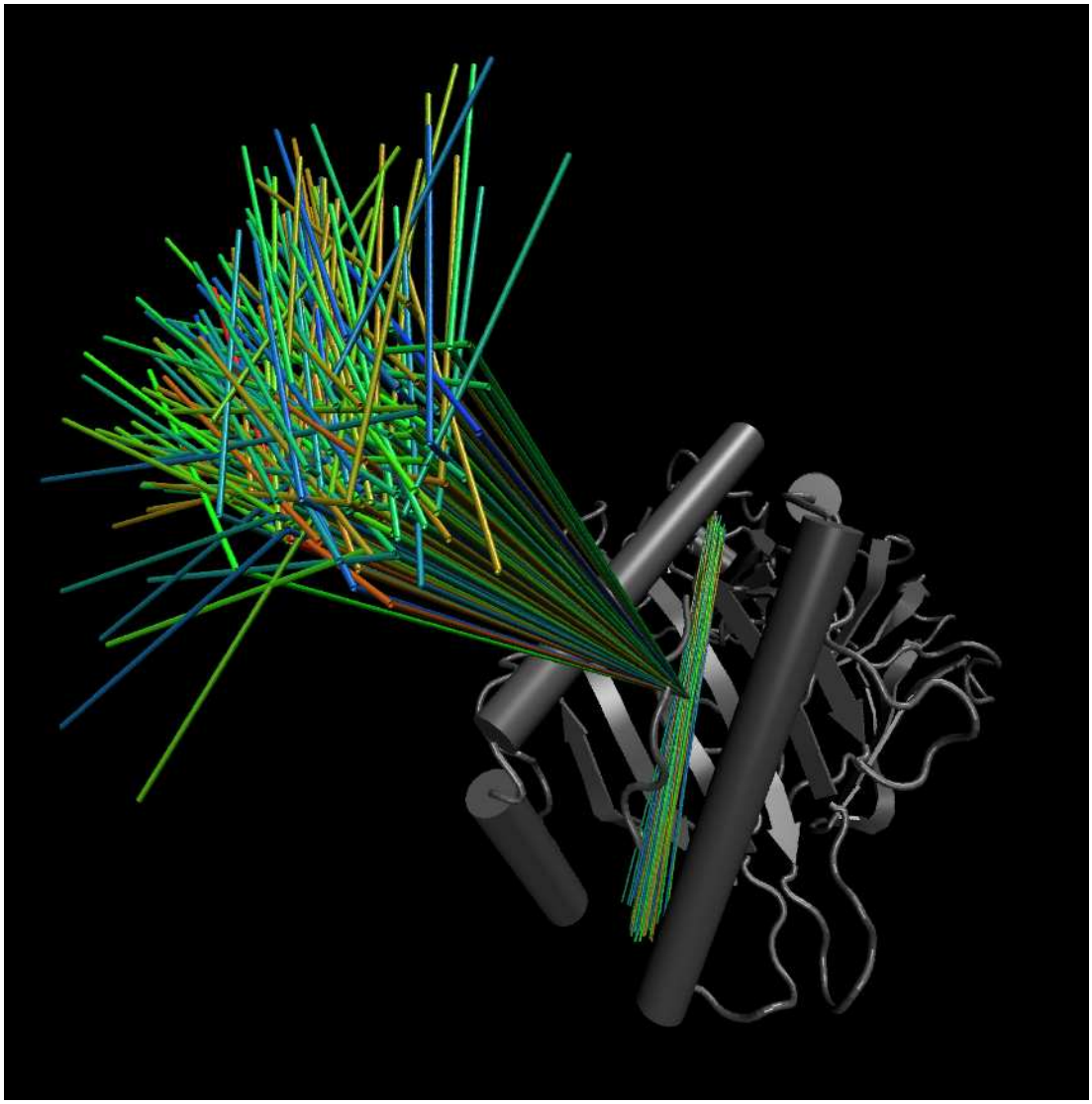

Figure 6.6: *View of the pMHC interface showing the final position and angle of the TCR after separation from the pMHC. The long segments connect the MHC reference point to the TCR center of mass. The short segments in the foreground connect the centers of mass of the whole TCR and the TCR-α chain. The color reflects the magnitude of the final work performed in a particular trajectory, with red corresponding to large work. See text for details.*

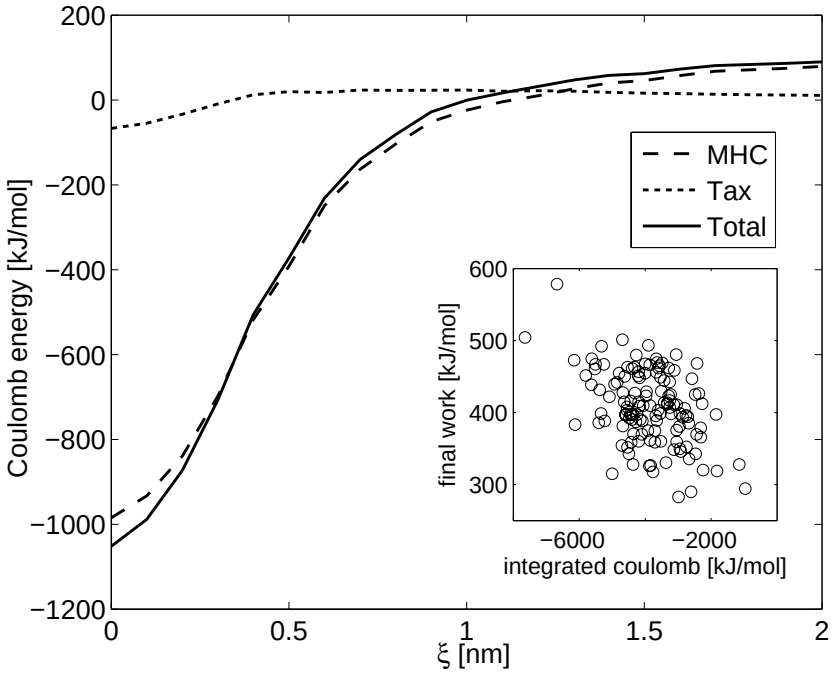

Figure 6.7: *Electrostatic interaction energies (excluding water and with a relative permittivity of* 1*) as a function of the reaction coordinate, averages over 152 unbinding trajectories with the Tax peptide. The inset shows, for each trajectory, the final work as a function of the integrated electrostatic energy.*

contribution can easily be shielded by ions and the solvent. The strong attraction at short distance is explained by complementary surface charges and dipoles. The inset shows that there is little correlation between the electrostatic energy integrated along the trajectory and the final work. The electrostatic interaction energies of the P6A mutant exhibit very similar behaviors (data not shown). We note that the Tax complex has a 75 kJ/mol more favorable bound state electrostatic interaction than the P6A complex. This could be a contribution to the difference in association free energies.

The same analysis was done for the Lennard-Jones interaction energy between the TCR and the pMHC, as a function of the reaction coordinate. Fig. 6.8 shows that the relative contribution of the peptide to the total interaction is much larger than in the electrostatic case. We see that beyond 1.0 nm CM distance all van der Waals contacts are broken. The inset shows the correlation that exists between the Lennard-Jones energy integrated along the trajectories and the final work. This indicates that the van der Waals and steric interactions play a determining role in the dedocking mechanism. The P6A complex has a similar behavior (data not shown), with a LJ interaction energy just 11 kJ/mol above the Tax complex in the bound state. The resolvation LJ energy is expected to be of the same order as the pMHC-TCR LJ interaction energy, and both contributions should be taken

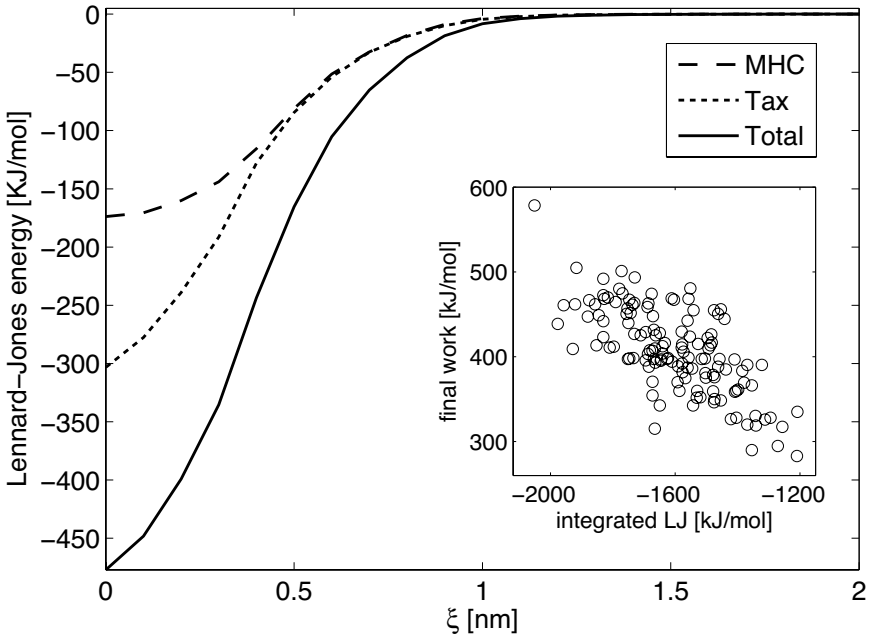

Figure 6.8: *Van der Waals interaction energy (excluding water) as a function of the reaction coordinate, averaged over 152 unbinding trajectories with the Tax peptide. The inset shows, for each trajectory, the final work as a function of the integrated van der Waals energy.*

into account prior to drawing any conclusion on the role of the van der Waals contacts in the dissociation process.

Another indicator of the strength of the TCR-pMHC binding is the number of hydrogen bonds between the two molecules. Fig. 6.9 shows the variation of the average number of hydrogen bonds as a function of ξ. The criteria used are : (i) hydrogen-acceptor distance smaller than 0.25 nm, (ii) angle donor - hydrogen - acceptor larger than 135°. We see that beyond 1 nm, all hydrogen bonds have disappeared. The inset on the right panel shows that, in the bound state, the P6A mutant has on average one interaction hydrogen bond less than the wild type complex. On another hand, the P6A retains an additional hydrogen bond in the second phase of the dissociation process. Fig. 6.9 also shows the resolvation of each part upon dissociation. The average variation of the number of hydrogen bonds between either TCR or pMHC and any water molecule is plotted, with the bound state as a reference. We see that the TCR in solution recovers the same number of hydrogen bonds as when bound. With both Tax and P6A peptides, the pMHC is able to form approximately two hydrogen bonds more with the solvent than with the TCR.

The dotted line on Fig. 6.9 represents the sum of the number of hydrogen bonds between parts of the complex and the number of solvation hydrogen bonds (mean of TCR and pMHC). It shows that, at the first stage of the dissociation, the solvent is able to

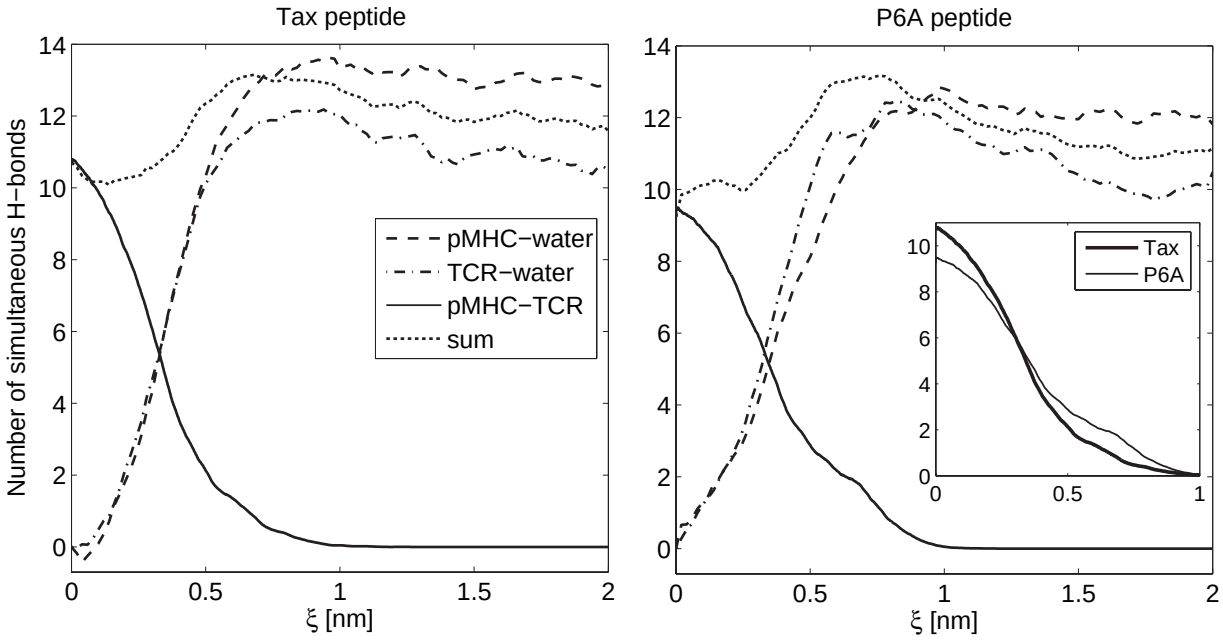

Figure 6.9: *Number of hydrogen bonds upon dissociation for the Tax and P6A complexes. The dotted line represents the sum of the pMHC -TCR hydrogen bonds and the average of the solvation hydrogen bonds for each pMHC and TCR. The inset on the right panel reproduces the pMHC-TCR H-bond curves to facilitate comparison between Tax and P6A.*

progressively compensate for the TCR-pMHC hydrogen bonds lost. There is thus no desolvated intermediate state, which would have indicated that the pulling speed used was excessive compared to the solvent characteristic diffusion time. Between 1 and 1.5 nm separation, a slow decrease of the dotted line indicates that the surface of the proteins reorganizes upon solvation and looses one to two hydrogen bonds with the solvent. The P6A complex displays a peak in the total number of hydrogen bonds (compared to the bound or unbound states), which is higher than for the Tax complex. This could contribute to lowering the observed dissociation free energy barrier of the P6A complex with respect to the Tax complex.

6.2.2 Binding free energy profile estimations

The Jarzynski method, as described in the introduction, was applied to the 152 trajectories of the TCR-Tax-MHC complex. An additional set of 162 trajectories was run in the same conditions for the P6A mutant. The PMF obtained are shown in Fig. 6.10. We see that the calculated dissociation free energy differences, -110 and -210 kJ mol^{-1} for the Tax and P6A respectively, are largely underestimated as compared to the experimental results. The calculated ΔG^{off} for the Tax complex is also underestimated by approximately 30 kJ mol^{-1}. The statistical error related to the use of the second order cumulant expansion (6.33) can be estimated with formula (6.34). This results in an expected standard deviation

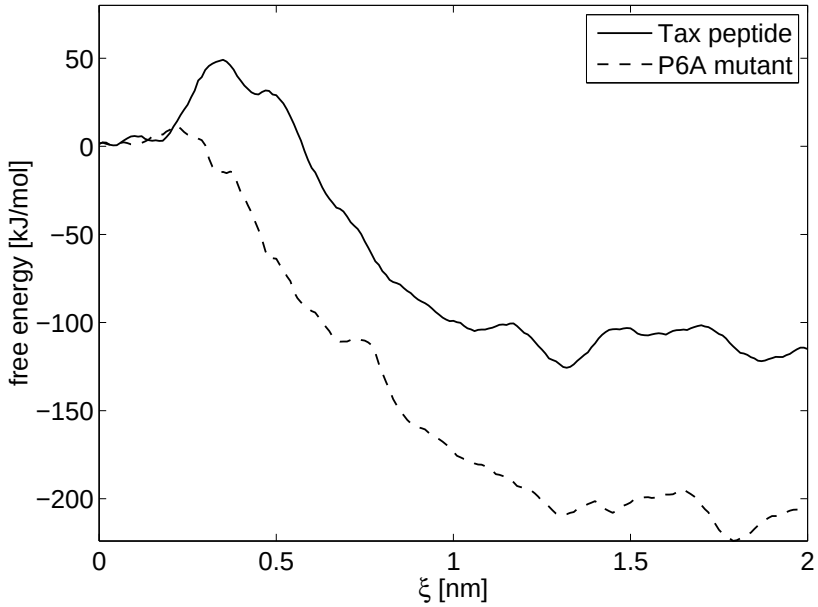

Figure 6.10: *Potentials of mean force obtained from the Jarzynski method with 152 trajectories for the Tax peptide and 162 for the P6A mutant.*

for ΔG of 17 kJ/mol at the barrier, and 55 kJ/mol in the unbound state. Even with such large error bars, the calculated results cannot be reconciled with the experiment.

Our simulations are nonetheless able to unambiguously differentiate the behavior of the P6A mutant, and give free energy differences of the right sign, both at the barrier and in the dissociated state. This is already not trivial given the very minute and local character of the mutation, as well as the great structural similarity of both complexes [11]. The P6A complex seems to have a very low free energy barrier, which might be corroborated by the fact that no binding of this mutant could be observed in the plasmon resonance experiment [11].

The contributions to the second order cumulant expansion (6.33) used in 6.11 to estimate the free energy are shown in the right panel of Fig. 6.11. The first term is the mean unbiased steering work. We see that beyond approximately 1.0 nm, very little work is performed on the system, indicating that the reaction coordinate range chosen is enough to reach a stable unbound state. This is consistent with observations in Figs. 6.7, 6.8, and 6.9. The P6A mutant requires on average less work for the dissociation, which is consistent with the steering potential energy differences observed in Fig. 6.4. The second term in (6.33), which represents the fluctuations in the unbiased work distribution, is also shown in the left panel of Fig. 6.11. The values are appropriately scaled, so that the free energy profile of 6.11 corresponds to the mean work curve minus the fluctuation curve. We see that the lower free energy profile of the P6A mutant is due to both a lower mean work and a higher fluctuation term.

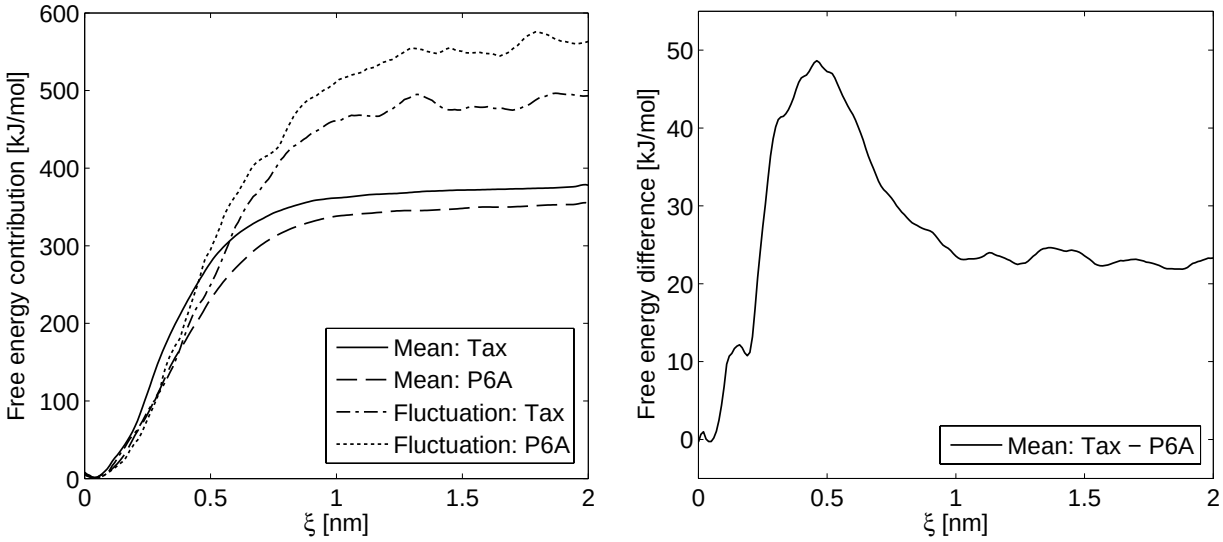

Figure 6.11: *Left panel: Black lines show the mean* $\frac{1}{N}\sum_i \bar{W}_i$ *of the unbiased work necessary to unbind the TCR from the pMHC for the Tax peptide and the P6A mutant, as a function of the reaction coordinate. The red lines show the fluctuation contribution to the second order cumulant expansion free energy estimate, i.e. the right most term in (6.33). Right panel: Difference of the mean work profiles for the Tax and the P6A mutant.*

The right panel of Fig. 6.11 shows the difference between the two mean work curves on the left panel, for the Tax and P6A peptides. Assuming that the dissipative effects are the same for both peptides, this provides a crude approximation of the TCR-pMHC dissociation free energy difference between the Tax and P6A complexes. The advantage of this approximation is that it is simple and avoids the estimation of the variance of the work distributions, which is sensitive to extreme values. Fig. 6.11 shows a strong difference at the free energy barrier, which is consistent with the contrasted SPR experiments [11] reporting a high free energy barrier for the Tax complex, but a low barrier for the P6A complex (no value measured, the complex was too unstable). The equilibrium free energy value is 22 kJ/mol, which compares better to the experimental value of 12 kJ/mol [11, 12] than the approximate 90 kJ/mol of Fig. 6.10

The box used for the present steering MD simulation is of limited size, due to computing power limitations. Thus, periodicity artifacts are one of the factors which could impair the reliability of the results. The periodic boundary conditions used in the simulation place the system in a different context as in the experiment, where it is solvated in a quasi-infinite solution, and doesn't see ordered neighbors. Upon TCR-pMHC dissociation, the lattice shape changes from one big protein per cell to two separated parts. Therefore, the periodicity artifacts - if any - can be different in the bound and unbound states, and may consequently affect the calculated PMF. The artifacts can be of two kinds [65, 66]. First the long-range electrostatic energy can be affected by the presence of a periodic lattice of charges. Second, the rearrangement of the water molecules participating in the protein

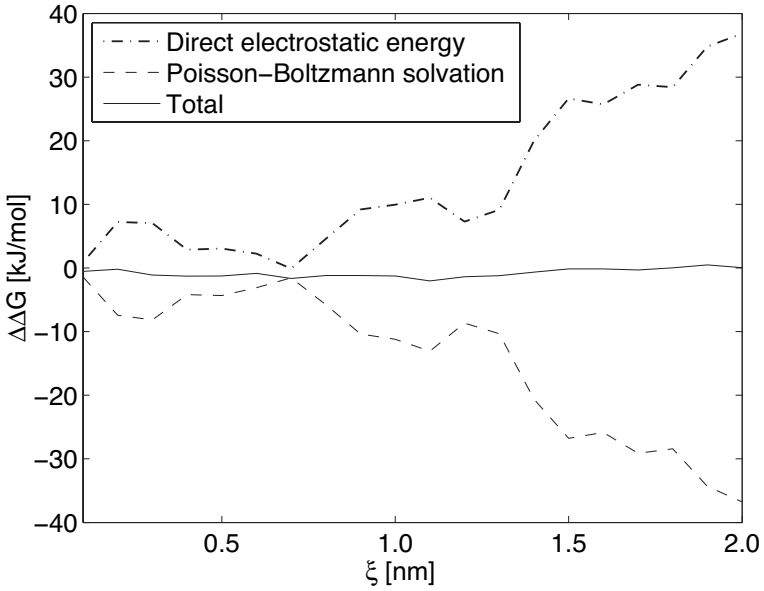

Figure 6.12: *Difference in unbinding free energy differences due to long-range electrostatics or Poisson-Boltzmann polar solvation effects.*

solvation is disturbed by the presence of the periodic images, which influences the polar part of the solvation free energy.

In order to assess these effects, selected configurations are taken from each trajectory, and the solvent is removed. For the electrostatic part, the difference between the full Coulomb energy (without cutoff) of the isolated protein and the PME energy of a periodic lattice of proteins is calculated. This gives the energy cost for going from an ideally isolated molecule to the periodic configuration of the simulation. For the solvation part, the difference between Poisson-Boltzmann terms in the isolated and periodic configurations is calculated. The results are shown in Fig. 6.12. We see that although each individual artifact has a sizable magnitude, the two compensate, and the resulting effect on the PMF is negligible.

Another possible cause for the free energy undervaluation in Fig. 6.10 is the large drift observed in the Nosé-Hoover pseudo-Hamiltonian $H'(\Gamma)$ (see (3.16) in Chapter 3). For the whole system, the value of this drift at equilibrium is of 188 kJ $\text{mol}^{-1}\text{ps}^{-1}$. This is an average computed on the 162 equilibration trajectories for the P6A mutant. This large negative drift means that if the energy of the physical system remains on average constant, it is because the thermostat is continuously pumping energy into the system. This is due to the twin-range cutoff scheme used. Indeed a control simulation where long-range interactions were computed at every step showed a negligible pseudo-Hamiltonian drift. The derivation of the JI for the Nosé-Hoover thermostat proposed in Chapter 3 assumes that $H'(\Gamma)$ is constant at equilibrium. A similar derivation can be done with a sink term added to $H'(\Gamma)$. The result is not surprizing : the energy sink term simply adds

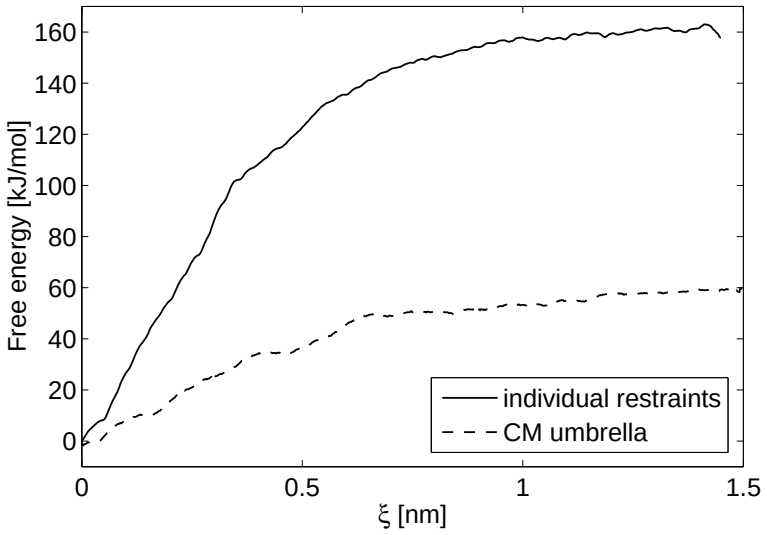

Figure 6.13: *Potentials of mean force obtained from the umbrella sampling method. The upper line is obtained using the individual restraint scheme described in the text. The lower curve is obtained with a usual restraining of the CM distances.*

up to the free energy estimate. In the present case, this would lead to much more negative free energy values at the end of the process. However, the sink term in $H'(\Gamma)$ is collective to all degrees of freedom in the system. In practice, it seems reasonable that the JI is only sensitive to the degrees of freedom that are coupled to the reaction coordinate. It is impossible to estimate the fraction of the total that this represents. Thus we can say that we expect a bias towards lower dissociation free energies due to the $H'(\Gamma)$ drift, but it is not quantifiable. For this reason, it would be preferable to choose an energy conserving MD setup for future applications of the JI. Of course, this will be at a significantly higher computational cost.

The PMFs obtained from umbrella sampling calculations with individual restraints or CM restraints are shown in Fig. 6.13. The data collected were postprocessed with the WHAM (6.27) method. Strikingly, these PMFs are qualitatively very different from the Jarzynski results of Fig. 6.10). Although umbrella sampling is a more established method than steered MD, the absence of a free energy barrier prevents a favorable comparison with experiment, even qualitatively. The two umbrella PMFs also differ from each other, which indicates that there might still remain methodological problems related to the application of individual restraints. However, the PMF obtained from the CM umbrella sampling also does not show the free energy barrier, and cannot be considered more reliable than the umbrella sampling with individual restraints. No further conclusions can be drawn at this point.

6.3 Conclusion

Behind this study were two motivations. The first one is of methodological nature. The recent steered molecular dynamics methods had never been applied to a system of the size of the TCR-pMHC. The height of the free energy barrier to overcome represents a challenge for the Jarzynski method. An original pulling scheme was introduced, which allows to prevent unrealistic molecular distortions and rotations. Related to this, an appropriate averaging method was developed to calculate the exponential average inherent to the Jarzynski method. However, questions regarding the methodology remain open, in particular in view of the large underestimation of the dissociation free energies, and the discrepancies between the Jarzynski and the umbrella sampling results. The steered molecular dynamics approach however is able to provide potentials of mean force, which, even if biased with respect to the experiment, are able to differentiate between large protein complexes differing only by one point mutation.

The second motivation for this work was to foster a better understanding of the TCR-pMHC interaction, which is central to the immunological response mechanism. The particular system considered here, the A6-Tax-HLA-A2 complex, is a challenge due to the total T cell response change induced by such a minute and local mutation as the replacement of a proline by an alanine residue in the Tax peptide. The present study shows the relative contribution of the MHC interface and the peptide to the electrostatic and van der Waals interactions along the dissociation path. The work measured on collections of trajectories for both the Tax peptide and the P6A mutant allows to unambiguously differentiate the two, but the Jarzynski method is not able to reproduce the experimental free energy barriers or equilibrium dissociation free energies. Looking closer at the forced dedocking simulations, it is difficult to determine which positional or energetic factors influence the final work of a given trajectory. About 300 dedocking trajectories of 4 ns each, and 160 umbrella restrained trajectories of 3 ns each on average are reported here. This represents a total of more than 1.6 microsecond of simulation time, let alone the equilibration runs, and the preliminary trials. This huge amount of data still needs to be analyzed more in detail to understand the role played by specific residues or specific hydrogen bonds during the process. Since a system-wide difference in behavior is observed between the Tax and the P6A peptides, it should be possible to evidence the key molecular events responsible for it. Even if the absolute free energies are not reproduced, the mechanisms at work in the simulation might give some insight on the biological process.

6.A Appendix : Additional Jarzynski PMF results

This appendix provides PMF results obtained during preliminary or subsequent control simulations of the TCR-Tax-MHC complex. These are shown on Fig. 6.14. The additional data sets differ by the Gromacs [56] (gmx) version used, the number of trajectories in each set, and the protocol for initial conditions:

- Protocol S : Sequential configurations taken every 18 ps from a long equilibrium reference trajectory.

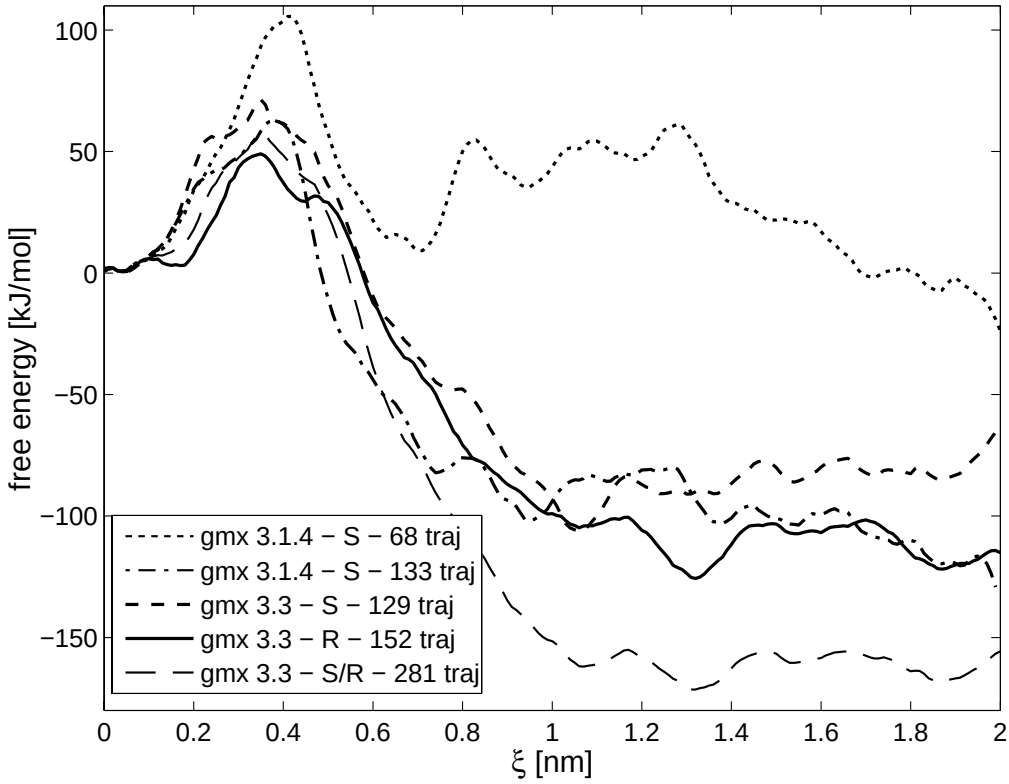

Figure 6.14: *Potentials of mean force obtained from additional data sets. The legend gives the Gromacs [56] version used, the protocol for initial conditions (S = sequential, R = random), and the number of trajectories.*

- Protocol R : Random initial velocities and short independent equilibration.

The data set used on Fig. 6.10 corresponds to conditions (*gmx 3.3 - R - 152 traj*). Note that the 68 trajectories used in the set (*gmx 3.1.4 - S - 68 traj*) are also contained in the set (*gmx 3.1.4 - S - 133 traj*), whose initial conditions were subsequent configurations from the same reference trajectory. The upper curve on Fig. 6.14, (*gmx 3.1.4 - S - 68 traj*), can lead to two different observations. On the one hand, it could show that using a small number of trajectories leads to undersampling of the work distribution and unreliable PMF. On the other hand, the first part of the reference trajectory could have been closer to the crystal structure, with better conserved contacts between the pMHC and the TCR. This hypothesis needs to be tested, but it would imply that subsequent conformations of the reference trajectory are not as reliable as the first 68. In this case, the upper curve on Fig. 6.14 would be a more accurate result (justifying that it is more consistent with the experimental ΔG).

The lowest curve on Fig. 6.14, (*gmx 3.3 - S/R - 281 traj*), combines the trajectories of the (*gmx 3.3 - S - 129 traj*) and (*gmx 3.3 - R - 152 traj*) sets and constitutes the biggest data set. It yields a low ΔG value around -160 kJ mol^{-1}. There are again two possible interpretations. On the one hand, this curve could be the most accurate, since it is based on the largest dataset. On the other hand, it could show that two data sets with different initial condition protocols (S or R) cannot be combined. A slight but significant difference between the mean work observed in these two data sets (not shown) supports this view.

Fig. 6.14 finally shows that using the S or R protocols with gmx 3.3 or using the S protocol with 133 trajectories of gmx 3.1.4 yield similar results. Given the statistical nature of the cumulant expansion method, it is difficult to say whether the difference between protocols or gmx versions is significant. Overall, the three central curves on Fig. 6.14 confirm the robustness of the observations previously made on Fig. 6.10: ΔG^{off} lies between 50 and 60 kJ mol^{-1} and the calculated ΔG is around -100 kJ mol^{-1}. The discrepancy with the experimental values remains to be explained.

Acknowledgements

The work presented in this chapter was supervised by O. Michielin. We are grateful to the Cluster Versus Cancer foundation, the Swiss Institute of Bioinformatics Vital-IT infrastructure, and the HP Labs in Bristol for supplying the computing power necessary for this work. We thank Vincent Zoete and Simon Bernèche for their useful comments.

6.2 References

[1] M. G. Rudolph, R. L. Stanfield, and I. A. Wilson. "How TCRs bind MHCs, peptides, and coreceptors". *Annu. Rev. Immunol.*, **24**, (2006) 419.

[2] D. Valmori, V. Dutoit, V. Schnuriger, A. L. Quiquerez, M. J. Pittet, P. Guillaume, V. RubioGodoy, P. R. Walker, D. Rimoldi, D. Lienard, J. C. Cerottini, P. Romero, and P. Y. Dietrich. "Vaccination with a melanA peptide selects an oligoclonal T cell population with increased functional avidity and tumor reactivity". *J. Immunol.*, **168**, (2002) 4231.

[3] P. Y. Dietrich, F. A. L. Gal, V. Dutoit, M. J. Pittet, L. Trautman, A. Zippelius, I. Cognet, V. Widmer, P. R. Walker, O. Michielin, P. Guillaume, T. Connerotte, F. Jotereau, P. G. Coulie, P. Romero, J. C. Cerottini, M. Bonneville, and D. Valmori. "Prevalent role of TCR α-chain in the selection of the preimmune repertoire specific for a human tumorassociated selfantigen". *J. Immunol.*, **170**, (2003) 5103.

[4] D. S. Lyons, S. A. Lieberman, J. Hampl, J. J. Boniface, Y. Chien, L. J. Berg, and M. M. Davis. "A TCR binds to antagonist ligands with lower affinities and faster dissociation rates than to agonists". *Immunity*, **5**, (1996) 53.

[5] J. D. Rabinowitz, C. Beeson, D. S. Lyons, M. M. Davis, and H. M. McConnell. "Kinetic discrimination in t-cell activation". *PNAS*, **93**, (1996) 1401.

[6] D. Hudrisier, B. Kessler, S. Valitutti, C. Horvath, J. C. Cerottini, and I. F. Luescher. "The efficiency of antigen recognition by CD81 CTL clones is determined by the frequency of serial TCR engagement". *J. Immunol.*, **161**, (1998) 553.

[7] C. Rosette, G. Werlen, M. A. Daniels, P. O. Holman, S. M. Alam, P. J. Travers, N. R. J. Gascoigne, E. Palmer, and S. C. Jameson. "The impact of duration versus extent of TCR occupancy on T cell activation, a revision of the kinetic proofreading model". *Immunity*, **15**, (2001) 59.

[8] T. N. Schumacher. "T-cell-receptor gene therapy". *Nat. Rev. Immunol.*, **2**, (2002) 512.

[9] L. K. Chlewicki, P. D. Holler, B. C. Monti, M. R. Clutter, and D. M. Kranz. "High-affinity, peptide-specific T cell receptors can be generated by mutations in CDR1, CDR2 or CDR3". *J. Mol. Biol.*, **346**, (2005) 223.

[10] D. N. Garboczi, P. Ghosh, U. Utz, Q. R. Fan, W. E. Biddison, and D. C. Wiley. "Structure of the complex between human t-cell receptor, viral peptide and HLA-A2". *Nature*, **384**, (1996) 134.

[11] Y. H. Ding, B. M. Baker, D. N. Garboczi, W. E. Biddison, and D. C. Wiley. "Four A6-TCR/peptide/HLA-A2 structures that generate very different T cell signals are nearly identical". *Immunity*, **11**, (1999) 45.

[12] B. M. Baker, S. J. Gagnon, W. E. Biddison, and D. C. Wiley. "Conversion of a T cell antagonist into an agonist by repairing a defect in the TCR/peptide/MHC interface: implications for TCR signaling." *Immunity*, **13**, (2000) 475.

[13] M. G. Rudolph, J. G. Luz, and I. A. Wilson. "Structural and thermodynamic correlates of T cell signaling". *Annu. Rev. Biophys. Biomol. Struct.*, **31**, (2002) 121.

[14] A. Lanzavecchia, G. Lezzi, and A. Viola. "From TCR engagement to T cell activation: a kinetic view of T cell behavior". *Cell*, **96**, (1999) 1.

[15] O. Michielin and M. Karplus. "Binding free energy differences in a TCR-peptide-MHC complex induced by a peptide mutation: A simulation analysis". *J. Mol. Biol.*, **324**, (2002) 547.

[16] S. Z. Wan, P. V. Coveney, and D. R. Flower. "Molecular basis of peptide recognition by the TCR: Affinity differences calculated using large scale computing". *J. Immunol.*, **175**, (2005) 1715.

[17] G. M. Torrie and J. P. Valleau. "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling". *J. Comp. Phys.*, **23**, (1977) 187.

[18] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer. "Targeted molecular dynamics simulation of conformational change-application to the TR transition in insulin". *Mol. Simulat.*, **10**, (1993) 291.

[19] G. Hummer and A. Szabo. "Free energy reconstruction from nonequilibrium single-molecule pulling experiments". *PNAS*, **98**, (2001) 3658–3661.

[20] S. Park, F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. "Free energy calculation from steered molecular dynamics simulations using jarzynski's equality". *J. Chem. Phys.*, **119**, (2003) 3559.

[21] S. Park and K. Schulten. "Calculating potentials of mean force from steered molecular dynamics information". *J. Chem. Phys.*, **120**, (2004) 5946.

[22] C. Jarzynski. "A nonequilibrium equality for free energy differences". *Phys. Rev. Lett.*, **78**, (1997) 2690.

[23] C. Jarzynski. "Equilibrium free energy differences from nonequilibrium measurements: a master equation approach". *Phys. Rev. E*, **56**, (1997) 5018.

[24] W. G. Hoover. "Canonical dynamics: Equilibrium phase-space distributions". *Phys. Rev. A*, **31**, (1985) 1695.

[25] G. E. Crooks. "Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences". *Phys. Rev. E*, **60**, (1999) 2721.

[26] J. M. Schurr and B. S. Fujimoto. "Equalities for the nonequilibrium work transferred from an external potential to a molecular system. analysis of single-molecule extension experiments". *J. Phys. Chem. B*, **107**, (2003) 14 007.

[27] D. J. Evans and D. J. Searles. "The fluctuation theorem". *Adv. Phys.*, **51**, (2002) 1529.

[28] C. Jarzynski. "Hamiltonian derivation of a detailed fluctuation theorem". *J. Stat. Phys.*, **98**, (2000) 77.

[29] D. J. Evans. "A nonequilibrium free energy theorem for deterministic systems". *Mol. Phys.*, **101**, (2003) 1551.

[30] M. A. Cuendet. "Statistical mechanical derivation of Jarzynski's identity for thermostated non-Hamiltonian dynamics". *Phys. Rev. Lett.*, **96**, (2006) 120 602.

[31] M. A. Cuendet. "The Jarzynski identity derived for general non-Hamiltonian and Hamiltonian dynamics generating the NVT or NPT ensembles". *J. Chem. Phys.*, **submitted**.

[32] J. Liphardt, S. Dumont, S. B. Smith, I. Tinoco Jr., and C. Bustamente. "Equilibrium information from nonequilibrium measurements in an experimental test of jarzynski's identity". *Science*, **296**, (2002) 1832.

[33] M. Rief and H. Grubmüller. "Force spectroscopy of single biomolecules". *Chem. Phys. Chem.*, **3**, (2002) 255.

[34] L. Shen, J. S. and*X. Luo, F. Cheng, Y. Xu, K. Chen, E. Arnold, J. Ding, and H. Jiang. "Steered molecular dynamics simulation on the binding of NNRTI to HIV-1 RT". *Biophys. J.*, **84**, (2003) 3547.

[35] L. Kjer-Nielsen, C. S. Clements, A. W. Purcell, A. G. Brooks, J. C. Whisstock, S. R. Burrows, J. McCluskey, and J. Rossjohn. "A structural basis for the selection of dominant $\alpha\beta$ T cell receptors in antiviral immunity". *Immunity*, **18**, (2003) 53.

[36] M. Cebecauer, P. Guillaume, S. Mark, O. Michielin, N. Boucheron, M. Bezard, B. H. Meye, J. M. Segura, H. Vogel, and I. F. Luescher. "CD8+ cytotoxic T lymphocytes activation by soluble major histocomatibility complex (MHC)-peptide dimers". *J. Biol. Chem.*, **280**, (2005) 23 820.

[37] M. Cebecauer, P. Guillaume, S. Mark, N. Boucheron, O. Michielin, M. Bezard, B. H. Meyer, J. M. Segura, H. Vogel, and I. F. Luescher. "Dimeric T cell receptor antigen recognition mode endows CD8+ CTL with high sensitivity". *Immunity*, (in print).

[38] W. F. van Gunsteren, T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith. "Computation of free energy in practice: choice of approximations and accuracy limiting factors". In: "Computer Simulation of Biomolecular Systems, Theoretical and Experimental Applications", eds. W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, vol. 2 (Escom Science Publishers, Leiden, The Netherlands, 1993) 315.

[39] A. M. Ferrenberg and R. H. Swendsen. "Optimized Monte Carlo data analysis". *Phys. Rev. Lett.*, **63**, (1989) 1195.

[40] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg. "The weighted histogram analysis method for free-energy calculations on biomolecules. i. the methods". *J. Comp. Chem.*, **13**, (1992) 1011.

[41] J. R. Gullingsrud, R. Braun, and K. Schulten. "Reconstructing potentials of mean force through time series analysis of steered molecular dynamics simulations". *J. Comp. Phys.*, **151**, (1999) 190.

[42] G. Hummer and A. Szabo. "Free energy surfaces from single-molecule force spectroscopy". *Acc. Chem. Res.*, **38**, (2005) 504.

[43] M. Balsera, S. Stepaniants, S. Izrailev, Y. Oono, and K. Schulten. "Reconstructing potential energy functions from simulated force-induced unbinding processes". *Biophys. J.*, **73**, (1997) 1281.

[44] M. Cascella, L. Guidoni, A. Maritan, U. Rothlisberger, and P. Carloni. "Multiple steering molecular dynamics applied to chemical reactions: Water exchange at sodium and potassium ions". *J. Phys. Chem. B*, **106**, (2002) 13 027.

[45] D. A. Hendrix and C. Jarzynski. "A "fast growth" method of computing free energy differences". *J. Chem. Phys.*, **114**, (2001) 5974.

[46] B. Roux. "The calculation of the potential of mean force using computer simulations". *Comp. Phys. Comm.*, **91**, (1995) 275.

[47] G. Hummer. "Fast-growth thermodynamic integration: Error and efficiency analysis". *J. Chem. Phys.*, **114**, (2001) 7330.

[48] J. Marcinkiewicz. "Sur une propriété de la loi de Gauss". *Math. Z.*, **44**.

[49] J. Hermans. "Simple analysis of noise and hysteresis in (slow-growth) free energy simulations". *J. Phys. Chem.*, **95**, (1991) 9029.

[50] J. B. Reiser, C. Darnault, C. Grégoire, T. Mosser, G. Mazza, A. Kearney, P. A. van der Merwe, J. C. Fontecilla-Camps, D. Housset, and B. Malissen. "CDR3 loop flexibility contributes to the degeneracy of TCR recognition". *Nature Immunology*, **4**, (2003) 241.

[51] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. *Biomolecular Simulation: The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).

[52] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, and J. Hermans. *Intermolecular forces* (Reidel, Dordrecht, 1981).

[53] S. Miyamoto and P. A. Kollman. "Settle: An analytical version of the shake and rattle algorithms for rigid water models". *J. Comp. Chem.*, **13**, (1992) 952.

[54] P. M. Petrone and A. E. Garcia. "Mhc-peptide binding is assisted by bound water molecules". *J. Mol. Biol.*, **338**, (2004) 419.

[55] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. "Gromacs: A message-passing parallel molecular dynamics implementation". *Comp. Phys. Comm.*, **91**, (1995) 41.

[56] E. Lindahl, B. Hess, and D. van der Spoel. "Gromacs 3.0: A package for molecular simulation and trajectory analysis". *J. Mol. Mod.*, **7**, (2001) 306.

[57] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. "Gromacs: Fast, flexible, and free". *J. Comp. Chem.*, **26**, (2005) 1701.

[58] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. "A smooth particle mesh ewald method". *J. Chem. Phys.*, **103**, (1995) 8577–8593.

[59] B. Hess, H. Bekker, H. J. C. Berendsen, and J. G. E. M. Fraaije. "Lincs: A linear constraint solver for molecular simulations". *J. Comp. Chem.*, **18**, (1997) 1463.

[60] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen. "Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes". *J. Comp. Phys.*, **23**, (1977) 327.

[61] M. Parrinello and A. Rahman. "Polymorphic transitions in single crystals : A new molecular dynamics method". *J. Appl. Phys.*, **52**, (1981) 7182.

[62] S. Nosé and M. L. Klein. "Constant pressure molecular dynamics for molecular systems". *Mol. Phys.*, **50**, (1983) 1055.

[63] L. S. D. Caves, J. D. Evanseck, and M. Karplus. "Locally accessible conformations of proteins: Multiple molecular dynamics simulations of crambin". *Protein Sci.*, **7**, (1998) 649.

[64] P. E. Smith and W. F. van Gunsteren. "Consistent dielectric properties of the simple point charge and extended simple point charge water models at 277 and 300 K". *J. Chem. Phys.*, **100**, (1994) 3169.

[65] P. H. Hünenberger and J. A. McCammon. "Ewald artifacts in computer simulations of ionic solvation and ion–ion interaction: A continuum electrostatics study". *J. Chem. Phys.*, **110**, (1999) 1856.

[66] M. A. Kastenholz and P. H. Hünenberger. "Influence of artificial periodicity and ionic strength in molecular dynamics simulations of charged biomolecules employing lattice-sum methods". *J. Phys. Chem. B*, **108**, (2004) 774.

Chapter 7

Outlook

If you put tomfoolery into a computer, nothing comes out of it but tomfoolery. But this tomfoolery, having passed through a very expensive machine, is somehow ennobled and no-one dares criticize it.

Pierre Gallois

All molecular dynamics (MD) simulations have limitations and are always a tradeoff between three aspects :

- Accuracy of the model
- Size of the system
- Duration of the simulation

These three features are best envisioned when placed on three perpendicular coordinate axes. A given simulation can be anywhere in space, provided that its distance to the origin is smaller than a certain radius corresponding to the maximum computing power available. For example, an MP2 [1] *ab initio* quantum-mechanical calculation may be highly ranked on the accuracy axis, but it will only include a few atoms and its duration will probably be limited to one geometry optimization. On the other hand, a recent classical MD simulation of a full virus [2] included more than a million atoms for about 10 ns. This is a questionable duration for the convergence of any observable on a system of this size. So, despite the tremendous computing power invested in it, this simulation remains rather low on the accuracy and duration axes, but sets a record on the size axis. Another example where size is maximized is the drug-design related problem of docking [3] an appropriate ligand on a protein. In this case, rather than the number of particles in a single system, the number of different ligands scanned for a given protein is key to the success, at the expense of a low accuracy model and poor or non-existent sampling. Finally, as an example where

the duration of the simulation or the sampling is maximized for a small system at modest accuracy, one can mention the *folding at home* project, in which folding rates of peptides have been determined [4] from microseconds of dynamics gathered from independent parts generated on thousands of home computers. These examples represent extreme cases, in which the limits are pushed along a single one of the (accuracy/size/duration) axes. Most simulations are however compromises, but the overall limitation remains, as expressed by the distance from the origin achieved in the (accuracy, size, duration) space.

Two important factors allow to push back the (accuracy, size, duration) limitation. First, the raw computer power, as well as the computing architecture at disposal, obviously determine the total number of particles modeled and of integration steps performed for a given level of accuracy. This means that the area accessible to an MD research group is largely determined by the budget allocated to hardware. But given Moore's law [5], computer power is constantly increasing for a fixed price. This gives a natural boost to the field of MD, where even without methodological improvement, one may find new solutions to the same problem addressed a few years before, due to a tenfold extension of the simulation reach. Hardware is constantly changing, providing new challenges and opportunities. Recent solutions include distributed computing over grids of distant computers such as the *folding at home* project [4], as well as massively parallel machines such as the IBM Blue Gene [6] made of thousands of processors. Hardware specially dedicated to MD is also currently being developed, in which force calculations are hard-coded. These include the MD GRAPE-3 boards or the NEC MD Engine, for both of which tenfold accelerations with respect to conventional processors [7] have been announced.

The second and certainly more interesting factor of progress comes from the methodological side of MD simulation. For a given computing time, the particular algorithm used to perform a given task, as well as the implementation details, greatly influence the simulation efficiency or/and reliability. For example, a grid-based algorithm [8, 9] to build the list of atoms within a cutoff distance of each other can run orders of magnitude faster than an exhaustive search. Alternatively, using an unbiased temperature regulation as shown in Chapter 2 or using a reversible and phase space measure preserving integrator [10, 11] will provide sampling more faithful to the physical ensemble at negligible additional computational cost. A very active front in MD research is the development of ingenious methods to enhance the sampling of relevant regions of the phase space, thus reducing the computing time needed for the convergence of a property of interest. The nonequilibrium free energy methods presented in Chapters 3 and 5 belong to this category. Besides algorithms, model parameters describing the physical system improve constantly. For example in classical MD, better force-field parametrization, together with the appropriate approximations can improve the reliability of a simulation [12], regardless of the computing power at hand. Both algorithms and models need to be continuously validated, improved, and adapted to the ever-increasing scale of simulations. Often the inaccuracies of MD calculations result more of a lack of sampling than of the model itself. Thus, the actual tendency to go to ever bigger systems as computer power develops should be followed only with caution. The extra computing power could better be invested in studying systems of modest size, but with proper sampling.

7.1 Thermostated dynamical systems

The field of thermostated dynamics was born a little more than twenty years ago, in close relationship with MD simulation. These dynamics allowed to study aspects of dynamical systems under external fields, which were unreachable before the computer age. These aspects include nonequilibrium steady states [13, 14], Lyapunov instability [15, 16, 17], and the symmetry breaking allowing time-reversible systems to display irreversible behaviors [18, 19]. In addition, thermostated nonautonomous dynamics relate to more general problems of mathematical interest. For example, Struckmeier recently provided a way to find an invariant for arbitrary nonautonomous Hamiltonian systems [20]. It was shown in Chapter 4 that this property relates to Hamiltonian dynamics such as the Nosé-Poincaré thermostat [21, 10]. As in the historical case of the Nosé [22], Nosé-Hoover and Nosé-Poincaré thermostats, there seems to exist a triangular mapping between the uneven-time-sampling Hamiltonian, the non-Hamiltonian, and the even-time-sampling Hamiltonian equations of motion, which all generate the same trajectories. It would be interesting to establish such a mapping between the respective generalized forms [23, 24, 25] of these dynamics.

The interest for Hamiltonian thermostats comes from the fact that they connect to the mathematical theory of symplectic transformations [26], and therefore bear symplectic integrators [10, 27]. Non-Hamiltonian thermostats are more challenging on the theoretical level, since no such notion as symplecticity exists to describe the behavior expected of an optimal integrator. In recent developments, the Riemannian metric tensor associated with any non-Hamiltonian system has been determined [28], and integrators have been proposed, which preserve the corresponding phase space volume element [11, 29]. On a more general level, the connection has been made [30] between MD dynamics and differential geometry [31]. Beyond their great theoretical interest, these bridges between mathematics and MD can help identifying the best algorithms for demanding applications in which the phase space compression is directly taken into account, such as those described in Chapters 3 and 5. Practical investigations are still needed to assess the impact of the entropy production due to algorithmic inaccuracies on conventional free energy and entropy calculations. This would lead to the specification of standards for high accuracy "calorimetric" MD.

7.2 Methods to calculate free energy differences

It was shown in Chapters 3 and 4 that the Jarzynski identity [32] (JI) emerges as a property of a general class of thermostated dynamical systems. However, the proofs provided still rely on the form of the equations of motion themselves. Following the connection of thermostated systems with the more general formalism of differential forms [30], it should be possible to write a derivation of the JI in geometrical terms. This derivation would perhaps be more independent of the form of the dynamics, and allow to outline the minimal necessary conditions for the JI to hold.

The nonequilibrium method of VandeVondele and Röthlisberger was rederived in two different forms in Chapter 5. The theoretical link between this method and the JI method

needs to be clarified. The parameters (pulling speed, number of sweeps, averaging window for the tracking function) need to be optimized, and the overall performance and the limitations of the method needs to be assessed for a variety of systems. There is currently a controversy on the applicability of the JI for accurate potential of mean force calculations [33, 34, 35, 36, 37]. Studying the convergence behavior of the VR method would help seeing if this controversy applies to the JI only, or more generally to nonequilibrium methods.

7.3 TCR-pMHC complex

The large amount of simulation data provided in Chapter 6 provides valuable insights on the interaction between the TCR and the pMHC. The large amount of data produced can still be used to answer specific questions that might arise in the future. The dissociation free energy profiles obtained for the wild type and the P6A mutant peptide are not consistent with either experiments or umbrella sampling calculations, which raises questions on the methodology and its application. In particular, the individual pulling scheme together with the associated weighted histogram averaging method need to be validated on a smaller system. The nonequilibrium simulations should be rerun with "calorimetric" MD standards (limited energy sinks and wells, and phase space compression), to see if the same trends appear. Various pulling speeds and various actuation methods could be tested as well.

It is however the first time that the JI is applied to such a big system. Therefore, the resulting free energy profiles maybe tell us about the limitations of the method itself. The particularity of such a big system is that there are a lot of degrees of freedom that are involved in the dissociation, but not directly coupled to the reaction coordinate, and possibly of much longer relaxation time. The JI is in principle a thermodynamic result about the system as a whole with all its degrees of freedom, valid for any pulling speed. It is however possible that in such cases the JI breaks down, or gets so difficult to converge that it becomes an impractical method. To tackle this kind of issues, one should separate the method-oriented approach from the system-oriented approach. The individual pulling method should be tested first on a system small enough that many trials can be performed. In a second phase, if the first phase was successful, the system of interest should be addressed.

7.4 References

[1] C. Møller and M. S. Plesset. "Note on an approximation treatment for many-electron systems". *Phys. Rev.*, **46**, (1934) 618.

[2] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten. "Molecular dynamics simulations of the complete satellite tobacco mosaic virus". *Structure*, **14**, (2006) 437.

[3] N. Brooijmans and I. D. Kuntz. "Molecular recognition and docking algorithms". *Annu. Rev. Biophys. Biomol. Struct.*, **32**, (2003) 335.

[4] B. Zagrovic, C. D. Snow, M. R. Shirts, and V. S. Pande. "Simulation of folding of a small alpha-helical protein in atomistic detail using worldwide-distributed computing". *J. Mol. Biol.*, **323**, (2002) 927.

[5] G. E. Moore. "Cramming more components onto integrated circuits". *Electronics*, **38**.

[6] A. Gara, M. A. Blumrich, D. Chen, G. L.-T. Chiu, P. Coteus, M. E. Giampapa, R. A. Haring, P. Heidelberger, D. Hoenicke, G. V. Kopcsay, T. A. Liebsch, M. Ohmacht, B. D. Steinmacher-Burow, T. Takken, and P. Vranas. "Overview of the Blue Gene/L system architecture". *IBM J. Res. & Dev.*, **49**, (2005) 195.

[7] R. Susukita, T. Ebisuzaki, B. G. Elmegreen, H. Furusawa, K. Kato, A. Kawai, Y. Kobayashi, T. Koishi, G. D. McNiven, T. Narumi, and K. Yasuoka. "Hardware accelerator for molecular dynamics: MDGRAPE-2". *Comput. Phys. Comm.*, **155**, (2003) 115.

[8] T. Heinz and P. H. Hünenberger. "A fast pairlist-construction algorithm for molecular simulations under periodic boundary conditions." *J. Comp. Chem.*, **25**, (2004) 1474.

[9] D. van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. "Gromacs: Fast, flexible, and free". *J. Comp. Chem.*, **26**, (2005) 1701.

[10] S. D. Bond, B. J. Leimkuhler, and B. B. Laird. "The Nosé–poincaré method for constant temperature moleculardynamics". *J. Comp. Phys.*, **151**, (1999) 114.

[11] M. E. Tuckerman, J. Alejandre, R. López-Rendón, A. L. Jochim, and G. J. Martyna. "A Liouville-operator derived measure-preserving integrator for molecular dynamics simulations in the isothermal–isobaric ensemble". *J. Phys. A*, **39**, (2006) 5629.

[12] A. D. MacKerell. "Empirical force fields for biological macromolecules: Overview and issues". *J. Comp. Chem.*, **25**, (2004) 1584.

[13] B. L. Holian, G. Ciccotti, W. G. Hoover, B. Moran, and H. A. Posch. "Nonlinear-response theory of time-independent fields: Consequences of the fractal nonequilibrium distribution function". *Phys. Rev. A*, **39**, (1989) 5414.

[14] D. Ruelle. "Positivity of entropy production in nonequilibrium statistical mechanics." *J. Stat. Phys.*, **85**, (1996) 1.

[15] W. G. Hoover, H. A. Posch, and C. G. Hoover. "Fluctuations and asymmetry via local lyapunov instability in the time-reversible doubly thermostated harmonic oscillator". *J. Chem. Phys.*, **115**, (2001) 5744.

[16] W. G. Hoover, C. G. Hoover, and H. A. Posch. "Dynamical instabilities, manifolds, and local Lyapunov spectra far from equilibrium". *Comput. Meth. Sci. Tech.*, **7**, (2001) 55.

[17] W. G. Hoover, H. A. Posch, K. Aoki, and D. Kusnezov. "Remarks on non-Hamiltonian statistical mechanics: Lyapunov exponents and phase-space dimensionality loss". *Europhys. Lett.*, **60**, (2002) 337.

[18] D. J. Evans, E. G. D. Cohen, and G. P. Morriss. "Probability of second law violations in shearing steady states". *Phys. Rev. Lett.*, **71**, (1993) 2401.

[19] W. G. Hoover. *Time Reversibility, Computer Simulation, and Chaos* (World Scientific Publishing, Singapore, 1999).

[20] J. Struckmeier and C. Riedel. "Exact invariants for a class of three-dimensional time-dependent classical Hamiltonians". *Phys. Rev. Lett.*, **85**, (2000) 3830.

[21] C. P. Dettmann and G. P. Morriss. "Hamiltonian reformulation and pairing of Lyapunov exponents for Nosé-Hoover dynamics". *Phys. Rev. E*, **55**, (1997) 3693.

[22] S. Nosé and M. L. Klein. "Constant pressure molecular dynamics for molecular systems". *Mol. Phys.*, **50**, (1983) 1055.

[23] J. Jellinek and R. S. Berry. "Generalization of Nosé's isothermal molecular dynamics". *Phys. Rev. A*, **38**, (1988) 3069.

[24] A. Bulgac and D. Kusnezov. "Canonical ensemble averages from pseudocanonical dynamics". *Phys. Rev. A*, **42**, (1990) 5045.

[25] B. B. Laird and B. J. Leimkuhler. "Generalized dynamical thermostating technique". *Phys. Rev. E*, **68**, (2003) 016 704.

[26] B. Leimkuhler. *Simulating Hamiltonian Dynamics* (Cambridge University Press, 2005).

[27] B. J. Leimkuhler and C. R. Sweet. "The canonical ensemble via symplectic integrators using Nosé and Nosé–poincaré chains". *J. Chem. Phys.*, **121**, (2004) 108.

[28] V. E. Tarasov. "Phase-space metric for non-Hamiltonian systems". *J. Phys. A*, **38**, (2005) 2145.

[29] G. S. Ezra. "Reversible measure-preserving integrators for non-Hamiltonian systems". *J. Chem. Phys.*, **125**, (2006) 034 104.

[30] G. S. Ezra. "On the statistical mechanics of non-Hamiltonian systems: the generalized Liouville equation, entropy, and time-dependent metrics". *J. Math. Chem.*, **35**, (2004) 29.

[31] L. Casetti, M. Pettini, and E. G. D. Cohen. "Geometric approach to Hamiltonian dynamics and statistical mechanics". *Phys. Reports*, **337**, (2000) 237.

[32] C. Jarzynski. "A nonequilibrium equality for free energy differences". *Phys. Rev. Lett.*, **78**, (1997) 2690.

[33] H. Oberhofer, C. Dellago, and P. L. Geissler. "Biased sampling of nonequilibrium trajectories: Can fast switching simulations outperform conventional free energy calculation methods?" *J. Phys. Chem. B*, **109**, (2005) 6902.

[34] D. M. Zuckerman and T. B. Woolf. "Systematic finite-sampling inaccuracy in free energy differences and other nonlinear quantities". *J. Stat. Phys.*, **114**, (2004) 1303.

[35] J. Gore, F. Ritort, , and C. Bustamante. "Bias and error in estimates of equilibrium free-energy differences from nonequilibrium measurements". *PNAS*, **100**, (2003) 12 564.

[36] F. M. Ytreberg and D. M. Zuckerman. "Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems". *J. Comp. Chem.*, **25**, (2004) 1749.

[37] M. R. Shirts and V. S. Pande. "Comparison of efficiency and bias of free energies computed by exponential averaging, the bennett acceptance ratio, and thermodynamic integration". *J. Chem. Phys.*, **122**, (2005) 144 107.

Curriculum Vitae

Name	Michel Alain Cuendet.
Date of birth	11 August 1974.
Place of birth	Morges, Switzerland.
Citizenship	Swiss

Education

April 2002 - present	Ph.D studies supervised by Prof. Dr. W. F. van Gunsteren, Laboratory of Physical Chemistry, Swiss Federal Institute of Technology ETH, Zürich, by Prof. Dr. U. Röthlisberger, Laboratory of Computational Chemistry and Biochemistry, Swiss Federal Institute of Technology EPFL, Lausanne, and by Prof. Dr. O. Michielin, Molecular Modeling Group, Swiss Institute of Bioinformatics, Lausanne.
November 2001	Postgraduate Degree in Mathematical Engineering, Swiss Federal Institute of Technology EPFL, Lausanne, and Ecole Polytechnique, Paris, France.
March 1998	Master of Science Degree in Physics, with thesis in Astrophysics, Swiss Federal Institute of Technology EPFL, Lausanne.
July 1993	Maturité Fédérale, section scientifique, Gymnase de la Cité, Lausanne.

Practical Experience

2001 (7 months)	Postgraduate thesis, Intelligent Business Infrastructure Group, IBM Zürich Research Laboratory.
2000 (5 months)	Research assistant, Genomation Laboratory. University of Washington, Seattle, USA.
1998 (5 months)	Traineeship, Applied Physics Group, Electrotechnologies Division, ABB Corporate Research Ltd, Baden-Dättwil.