Mathematical Issues in Molecular Dynamics

Robert D. Skeel (Purdue University),
Paul Tupper (McGill University)

June 4–9, 2005

1 Introduction

Computations for molecular systems work with a set of values Γ representing the microscopic state of the systems, e.g., atomic positions, atomic velocities, volume of the simulation box. Most such computations can be classified into one of two categories:

1. Sampling. Given the relative probability for different values of Γ, calculate the expectation of some observable $O(\Gamma)$. An example is the probability of two biomolecules being (noncovalently) bound versus unbound.

2. Dynamics. Given the relative probability for initial values $\Gamma(0)$ and equations of motion for $(\mathrm{d}/\mathrm{d}t)\Gamma(t)$, calculate the expectation of $O(\Gamma(t))$. An example is the position autocorrelation function. The equations of motion may be deterministic or stochastic.

The probability for different values of Γ is often specified by the ensemble the system is, see Section 4. It is important to note sampling computations commonly employ dynamics but the dynamics is typically unphysical.

We begin in Section 2 with a consideration of the models we use in molecular dynamics, and whether we can justify their use over other, more physically realistic, models.

A major topic, discussed in Section 3, concerns the effects of discretization errors in numerical trajectories for deterministic molecular dynamics (MD). Due to the highly chaotic nature of the dynamics, the numerical trajectory completely departs from the analytical trajectory very early in the simulation. Yet numerical experiments show that averages calculated from such erroneous trajectories are close the correct ones.

Section 4 considers problem 1, sampling, above. There are roughly two approaches: dynamical sampling, which using uses continuous time deterministic or stochastic dynamics, and Markov Chain Monte Carlo (MCMC) sampling, which uses a Markov Chain to explore state space. In this section we consider primarily the former class, which includes deterministic extended Hamiltonian approaches and stochastic Langevin dynamics. We also consider Hybrid Monte Carlo methods which combines aspects of dynamical sampling with MCMC techniques.

Section 5 considers other topics in sampling. In particular, various Monte Carlo Markov Chain (MCMC) methods are considered along with methods for speeding up their rate of convergence.

Often we do not want to just sample configurations or states of the system, but actual trajectories of the system. An example of this is when we specify an initial configuration and a final configuration and we wish to be able to sample from the set of all trajectories that run between them. Section 6 discusses this kind of situation, which is known as sampling from path space.

For many systems microscopic models are simply to restrictive in terms of the time and length scales that are attainable with current computations. Section 7 discusses coarse graining: replacing microscopic models with mesoscopic ones that preserve the same effective behaviour of the system.

2 Models and Justifications

The main focus of the meeting was classical molecular dynamics, whether determinsitic or stochastic. Since we believe that fundamentally molecular systems are quantum and deterministic, using classical models—particularly stochastic ones— requires some justification.

2.1 Classical vs. Quantum

At the most fundamental level the dynamics of atoms and molecules must follow the rules of *quantum* mechanics and the dynamics prescribed by Schrödinger's or Heisenberg's equations of motion. The presentation of J. Straub described the results of a careful study of the molecular dynamics of vibrational energy transfer within a protein. The predictions of classical models for the molecular dynamics, based on both direct molecular dynamics as well as the theory of "small vibrations" or normal modes of motion, were compared with those of quantum mechanical perturbation theory. It was noted that for the case of vibrational energy transfer in a protein, classical dynamics can lead to poor results for the time scales and pathways of energy transfer. More attention must be paid to this issue if the ultimate objective is to develop accurate numerical methods for realistic simulations of the molecular dynamics of biomolecular systems.

2.2 Stochastic vs. Deterministic

The use of stochastic equations of motion for (real) dynamics also requires an explanation: Let us start from the assumption that the desired molecular system can be described by classical mechanics entirely and correctly. This implies the existence of a microscopic state $\overline{\Gamma}$ and an associated Liouville operator $\overline{\mathcal{L}}$ which completely characterize the molecular system and the time evolution of distributions $\overline{\rho}(\overline{\Gamma}, t)$ in phase space in particular, i.e.,

$$\overline{\rho}_t = \overline{\mathcal{L}}\overline{\rho}.$$

Almost all numerical simulations will work with a reduced representation over a smaller phase space $\Gamma \subset \overline{\Gamma}$. Crucial is the assumption that the time evolution over such a reduced space is still Markovian and the existence of a (perhaps approximative) Liouville operator $\mathcal{L}$ is generally taken as granted. See work by Mori and Zwanzig and, more recently, by Stuart and co-workers for rigorous results. Note that we have to, in general, assume that $\mathcal{L}$ corresponds to some form of stochastic ODE with the noise process representing forces from the missing degrees of freedom. The feeling in that regard is that Langevin/Brownian dynamics is not appropriate and that a more general form of the stochastic and dissipative coupling terms is required (see dissipative particle dynamics (DPD)). Also, in practice the missing degrees of freedom are in contact with the system only on the boundaries. Although simulations are sometimes implemented this way, it is far more common to use periodic boundary conditions and it is an interesting question how to handle this. Constant energy simulations make sense really only for molecules in gas phase.

3 The Accuracy of Long Time Numerical Trajectories —Why MD Works?

The primary focus on the mathematical aspects of molecular dynamics was placed on the long-time accuracy of integrating Newton's or Hamilton's *classical* equations of motion. Many of the paradigms in that field of research were derived from the early work on celestial mechanics, where the assumption of classical dynamics is most certainly a good one.

In the case of classical MD, this question is posed for the difficult case of deterministic dynamics, specifically Hamiltonian dynamics. The equations for Hamiltonian dynamics are

$$\frac{\mathrm{d}}{\mathrm{d}t}\Gamma = J\nabla H(\Gamma), \qquad \Gamma = \begin{bmatrix} q \\ p \end{bmatrix}, \qquad J = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix},$$

for example, $H(q,p) = \frac{1}{2}p^{\mathsf{T}}M^{-1}p + U(q)$ where M is a diagonal matrix of masses and $U(q)$ is potential energy. Numerical evidence indicates that both (i) time averages for steady-state sampling and (ii) ensemble averages for time correlation functions are approximated well from numerical trajectories. The challenge is to explain why this is so.

3.1 One view

In the opening talk R. Skeel sketched an explanation for the success of Hamiltonian molecular dynamics for calculating time correlation functions in terms of the underlying Liouville equations for the probability density. Calculating numerical trajectories is basically the method of characteristics for approximating the probability density at any point in time. For accuracy it is not necessary to have long accurate trajectories, e.g., if two trajectories were exchanged this would not change the density. Because each time step generates a temporal discretization error, long-time accuracy is possible only if these errors are damped. This would seem to require that the Hamiltonian system have the mixing property. However, this is not easy to demonstrate: Suppose that the initial density is not very close to the stationary density but that after a long time t it reaches a density $\rho(q,p,t)$ which is very close. Now consider $\rho(q,-p,t)$. In any traditional norm this is equally close to the stationary density. However, if we started from here and did dynamics, the density would ultimately be equal to the original initial density. So no matter how close we are to the stationary density, it could evolve to be far away. So a convergence proof based on mixing seems difficult because an argument based on showing contractivity in some metric would have to use a metric that distinguishes between $\rho(q,p,t)$ and $\rho(q,-p,t)$. Perhaps, an argument that distinguishes between stable and unstable manifolds would work. Alternatively, the following is plausible: a solution $\rho(q,p,t)$, for smooth initial values $\rho(q,p,0)$, spends almost all of its time very near to $\rho(q,p)$:

$$\lim_{t\to\infty}\frac{1}{t}\int_0^t \|\rho(\cdot,\tau) - \rho(\cdot)\|\mathrm{d}\tau = 0$$

—for some suitably weak norm. This means that nearly all initial conditions produce mixing dynamics with the “expected” rate of convergence.

The preceding discussion is oversimplified and needs to be amended because ergodicity, and mixing, are not generic properties for Hamiltonian systems. It is only on most of phase space that the almost-always mixing property can hold.

3.2 Another view

In the closing discussion another explanation for the success of MD was given by S. Reich. To explain what makes MD work, we need to look at the spectral properties of $\mathcal{L}$. Mixing would correspond to an operator $\mathcal{L}$ with a spectrum $\sigma(\mathcal{L})$ and $\sigma_{\mathrm{ess}}(\mathcal{L}) := \sigma(\mathcal{L}) - 0$ with $\Re(\sigma_{\mathrm{ess}}(\mathcal{L})) \le \gamma < 0$. The constant γ characterizes the decay of correlation. Biomolecular systems would probably not be called mixing in the above sense as γ would be very close to zero (at least relative to simulation times) and $\sigma_{\mathrm{ess}}(\mathcal{L})$ contains both mixing (the essential spectrum) and metastable states (isolated eigenvalues).

In MD the time evolution of a density ρ could be approximated by a sum of individual trajectories and weighted Dirac delta functions, i.e.,

$$\rho_{\mathrm{num}}(\Gamma,t) = \sum_i \alpha_i \delta(\Gamma - \Gamma_i(t)).$$

An interesting question is the convergence of ρ_{num} to ρ under an appropriate norm (e.g., Wasserstein norm). It is optimistic to think that this interpretation of MD would bring “weaker” requirements with respect to trajectory accuracy. Rather it would be expected that one needs to require the convergence of numerical trajectories to shadow trajectories of a modified system for a vast majority of such trajectories. This is also supported by the fact that hyperbolic systems are mixing and almost all trajectories of hyperbolic systems can be shadowed.

One talk that supported this view was W. Hayes'. He presented the state-of-the-art on practical shadowing of real trajectories of Hamiltonian systems. The work originates in the area of celestial mechanics, where there has been great success in showing that computed trajectories are close to nearby analytic trajectories. (In a related talk, W. Newman showed how long celestial mechanics trajectories can be computed to machine precision. Can this be done for MD too?) Some, but not a lot of hope was expressed for the application of Hayes' work to MD, though results are preliminary.

3.3 Yet another view

There is another view, presented by A. Stuart, that does not require that numerical trajectories are shadowed over long time intervals by true trajectories. Begin with the observation that we are typically not interested in the statistical behaviour of the entire system, but only a one or two degree-of-freedom subsystem. The remainder of the system acts as a heat bath for this component. Often the trajectories of the one component are well approximated in distribution by those of a stochastic process, such as the solution to a stochastic differential equation. The remainder of the system acts as a source of random noise and damping. A rigorous results of this sort is given in [25]. Now, for numerically computed trajectories to give the right statistics, it is only necessary that they too are close in distribution to the trajectories of this stochastic process. This has been shown to be true for special systems through numerical experiments in [37] and [38], though as of yet there are no general theorems.

4 Dynamical Sampling

The problem of sampling is to generate states Γ^n, $n = 1, 2, \ldots$ from a prescribed distribution. Assuming ergodicity, one can compute them from the solution of an ODE or stochastic differential equation (SDE). Configurations generated this way are (i) tainted by the initial conditions, (ii) correlated, making it more difficult to estimate variance, and (iii) biased due to the use of a finite step size. Markov chain Monte Carlo methods, which we consider in Section 5, do not suffer the last of these drawbacks.

For a given Hamiltonian system there are a few different distributions on the state space that we might want to sample from. These different distributions are typically referred to as ensembles and correspond to different boundary conditions. The two that we discuss in this report are the NVE ensemble and the NVT ensemble. In each case it is important to note that the probability density can only be expressed in closed form up to a multiplicative constant, known as the partition function. This constant is unknown for most interesting systems.

The NVE or microcanonical ensemble describes an isolated system and corresponds to a fixed number of particles, fixed volume, and fixed energy for the system. If the fixed energy is E, the density of the distribution is given by

$$\frac{\delta(H(q,p)-E)}{\int \delta(H(u,v)-E)\mathrm{d}u\mathrm{d}v}.$$

So zero weight is given to all states where $H(q,p) \neq E$. This density is invariant under Hamiltonian dynamics. Assuming that the Hamiltonian dynamics is ergodic on the energy level set, exact Hamiltonian dynamics sample from the NVE distribution. Hence, when the differential equations of a Hamiltonian system are numerically integrated (that is, we do MD), we are attempting to compute NVE ensemble averages. Of course, many things could go wrong with this computation, and there will typically be a bias. MD is the only widely used procedure for sampling from the NVE ensemble.

Two talks specifically addressed errors in microcanonical simulation. In his talk, P. Tupper described some work attempting to justify the assumption of ergodicity in microcanonical dynamics and MD. The talk by S. Bond surveyed some results from backward error analysis and showed how (under certain assumptions) these results can be applied to compute estimates of the error in averages from molecular dynamics simulations. Results from several test problems were explored including examples from constant temperature molecular dynamics, which corresponds to the next ensemble we consider.

The NVT or canonical ensemble corresponds to a fixed number of particles, fixed volume, and fixed temperature. The density associated with this ensemble is proportional to $\exp(-\beta H(q,p))$ where $\beta = 1/kT$, k is Boltzmann's constant, and T is temperature. This ensemble is believed to be good representation of the distribution of the Hamiltonian system if it is in thermal contact with a much larger system of temperature T. A nice feature of this ensemble is that position q and momentum p are independent for separable Hamiltonians.

There are many different ways to sample from this distribution including

1. deterministic thermostats,
2. Langevin thermostats,
3. MCMC, (see Section 5) and
4. hybrid Monte Carlo (Monte Carlo using MD for proposals).

In the case of dynamical sampling, ergodicity is an important issue, and for both types of methods the rate of convergence is also very important. We consider methods 1, 2, and 4 in subsequent subsections.

An important point to note about all these methods is that unlike MD applied to unadorned Hamilton's equations, the dynamics is not *real*. In each case something has been added to the dynamics that does not correspond directly to any component of the real physical system we are attempting to model. This is done so that states generated by the trajectories sample the canonical distribution. However, it is no longer clear what the trajectories generated can be used for other than this. For example, is there any physical validity to a velocity autocorrelation function computed with Langevin dynamics? The answer may be no.

4.1 Dynamical thermostats

Nosé Dynamics. Leimkuhler, Barth and Sweet are developing extended Hamiltonian formulations for thermostating molecular dynamics. It was the observation of the physicist Nosé [33] that we can augment the energy function $H(q,p)$, by incorporating a single additional phase variable, s, together with its canonical momentum, π, in the following way:

$$H^{\text{Nose}} = H(q, \tilde{p}/s) + \frac{\pi^2}{2\mu} + U(q) + gkT \ln s.$$

A simple integration argument convinces us that,

$$\iint \delta\left[H^{\text{Nose}}(q, sp, s, \pi) - E\right] \mathrm{d}s\mathrm{d}\pi = \text{constant} \times \exp\left(-\frac{N}{gkT} H(q,p)\right),$$

moreover,

$$\begin{aligned} &\iint f(q, \tilde{p}/s)\delta\left[H^{\text{Nose}}(q, \tilde{p}, s, \pi) - E\right] \mathrm{d}s\mathrm{d}\pi\mathrm{d}\tilde{p} \\ &\qquad = \text{constant} \times \int f(q,p) \exp\left(-\frac{N}{gkT} H(q,p)\right) \mathrm{d}p, \end{aligned}$$

so if we choose $g = N$, we are able to reduce the microcanonical density function for H^{Nose} to the canonical one for H.

Time transformation. Accurate sampling of the NVT ensemble for certain types of systems calls for large fluctuations of the thermostating variable s, including potentially very small values. When this happens, the equations of motion from Nosé's Hamiltonian are poorly scaled and standard numerical methods become unstable. For this reason, a time transformation $dt/dt' = s$ is used in the derivation of the Nosé-Hoover equations. However, the way this is traditionally done destroys the symplectic structure. The basis of the Nose-Poincaré method of [6] is rather a *Poincaré transformation*, modifying the original Nosé Hamiltonian to:

$$\begin{aligned} H^{\text{NP}}(q, \tilde{p}, s, \pi) &= s\left[H^{\text{Nose}} - E\right] \\ &= s\left[H(q, \tilde{p}/s) + \frac{\pi^2}{2\mu} + U(q) + gkT \ln s - E\right]. \end{aligned}$$

It is a better alternative starting point for the development of numerical methods than Nosé-Hoover. Now the construction of numerical methods is slightly complicated by the modification of the Hamiltonian–we cannot directly use the Verlet integrator here, for example–but there are several ways to solve $H^{\rm NP}$ which work well.

Nosé-Poincaré Chains. An improvement on the Nosé approach is based on what are termed *Nosé-Poincaré chains*. In Nosé-Hoover chains, one thermostats the thermostating variable, introducing an additional variable and controlling it via an additional temperature equation. This process can be repeated. For example we can use a Hamiltonian like this:

$$H^{\rm NPC;1} = s_0 \left[H(q, \frac{\tilde{p}}{s_0}) + \frac{\pi_0^2}{2\mu s_1^2} + \frac{\pi_1^2}{2\mu_1} + gkT\ln s_0 + kT\ln s_1 + f(s_1) - E \right].$$

In this formula the time-rescaling has been included. If the function f is bounded below with bounded exponential integral, then the integration argument goes through. A canonical sampling argument exists in this case. We can make these chains as complicated as we like by extending them with more variables. The design of the regularizing function and choice of thermal masses is important for good results. For the right choice of parameters, Nosé-Poincare chains have reasonably good ergodicity properties: if we use 3 or 4 augmenting variables and choose carefully the thermal masses $\mu_1, \mu_2, \ldots,$, then we can get an accurate recovery of canonical sampling from these chains. The need for a careful choice of thermostat masses is a flaw, though, and the NPC methods tend to be unstable.

RMT Chains. Better methods are possible based on a more complicated interaction between the bath and the physical variables, termed *Recursive Multiple Thermostating* (RMT). The theory of this model is considered in detail in a recent article of Leimkuhler and Sweet, and practical selection of coefficients is discussed in the work of Barth,Leimkuhler and Sweet presented at the meeting. In nonlinear models the RMT formulation is found to be more sensitive to the details of the underlying system. Heuristics have been presented for selection of thermostat masses. Some preliminary encouraging data were presented on the use of these methods for study of an alanine dipeptide model, based on an implementation of RMT in CHARMM, the popular molecular dynamics code.

In separate work, Leimkuhler and Jia have proposed a general framework of thermostating dynamics for multiscale problems. Using these ideas, we can flexibly introduce canonical sampling over particular components while preserving physically relevant multiscale structural characteristics of the application and maintaining these characteristics in the design of efficient numerical algorithms. Certain classes of problems with fast and slow variable separation were examined in detail. Moreover, a method was proposed for following the slow evolution in an nonequilibrium setting, where only the fast degrees of freedom are assumed to be in equilibrium. The sampling properties of these new formulations were tested in numerical experiments using an enhanced reversible averaging scheme and a Nosé-Poincaré chain. For some choices of model setup, a fully Hamiltonian formulation is impractical.

In more recent work of Gill, Jia, Leimkuhler and Cocks, this multiscale partial thermostating method was applied with modified RMT thermostat for a simplified version of the quasicontinuum molecular dynamics (QCMD) model of materials science. Effectively the combination of this coarse-graining strategy with advanced thermostats provides a powerful adaptive boundary condition for a localized, detailed atomistic calculation. This type of method is useful for studies of nanoindentation, and for defect nucleation (e.g., cracks). This work was also discussed at the meeting.

The summary of the state of the art in dynamical thermostats is that they have interesting properties and can work in practical situations. Their ergodic properties are verifiable in at least some numerical studies, and they provide a flexible adaptive framework for molecular simulation while benefiting from the well known reliability of the classical MD simulation framework. More work is needed to probe their robustness, to understand their ergodic properties, and to evaluate their efficiency vis a vis stochastic dynamics methods. Some of these questions were raised in discussions at the meeting.

4.2 Langevin Thermostating

The Langevin equation in the case of sampling is

$$M\frac{\mathrm{d}^2}{\mathrm{d}t^2}q = -\nabla U(q) - M\gamma\frac{\mathrm{d}}{\mathrm{d}t}q + \sqrt{2k_{\rm B}TM\gamma}\frac{\mathrm{d}}{\mathrm{d}t}W(t) \tag{1}$$

where γ is a damping constant and $W(t)$ is a vector of independent standard Wiener processes.

Like for deterministic thermostating, the Langevin equations have the canonical ensemble as an invariant distribution. Unlike for deterministic dynamics, there are plenty of situations where Langevin dynamics are proven to be ergodic [29]. (One situation where Langevin dynamics has not been proven to be ergodic is that of Dissipative Particle Dynamics (DPD). T. Shardlow explained the problems here and gave an overview of his proof of ergodicity for 1d DPD. More on DPD in section on coarse graining.)

It is, of course, necessary to have an integration scheme for Langevin dynamics. M. Tretyakov described his work with Milstein on quasi-symplectic integrators. These integrators contract volume with the same rate as the original SDEs, and become symplectic methods in the limit of zero noise ($\gamma \to 0$). However, these methods applied to non-globally Lipschitz Langevin equations are not themselves ergodic; they will eventually diverge to infinity with probability one. Tretyakov discussed how this turns out to not be a problem in practice.

An issue that might be worth considering is the relationship between the Langevin based methods, with the coupling to a "physical" heat bath through the Caldeira-Leggett-Zwanzig Hamiltonian form, and the Nosé based methods, with the extended Lagrangian.

One special limiting case of Langevin dynamics is Brownian dynamics, obtained from (1) by letting $M \to 0$ while $M\gamma$ is a constant. This gives:

$$g\frac{\mathrm{d}}{\mathrm{d}t}q = -\nabla U(q) + \sqrt{2k_{\mathrm{B}}TM\gamma}\frac{\mathrm{d}}{\mathrm{d}t}W(t).$$

Brownian dynamics samples the from canonical distribution for the configuration variable q. One issue on which there was some disagreement was under which conditions it is better to just use Brownian rather than Langevin dynamics to sample configurations. A related question is what is the best parameter γ to use in Langevin dynamics for the fastest sampling.

Comparison of Thermostats. Both deterministic thermostats (such as Nosé) and Langevin (stochastic) thermostats can be used successfully to sample configurations from the canonical distribution. Both analytic dynamics have the canonical measure as an invariant. Langevin dynamics are provably ergodic in some cases. This not true of dynamical thermostated dynamics, but it may not be important in practice. In both cases numerical methods applied to them will lead to a bias (that decreases with reduced step length) and potentially instability over long simulations. A natural question is which— if any– is better for practical problems. In particular, which allows more efficient sampling of the canonical ensemble? One issue is that generating a complete set of random numbers per time-step may be costly in some circumstances. Some pointers for measuring ergodicity and convergence are given in [2], in particular, they describe an empirical "ergodic measure," which can be used to compare algorithms and to optimize them. Another possibility is to compare the two methods when used as proposals in Hybrid Monte Carlo (see next subsection). More work is needed to carefully and systematically compare the effectiveness of the best algorithms for both deterministic and stochastic dynamics.

4.3 Hybrid Monte Carlo

Hybrid Monte Carlo (HMC) is an MCMC method that uses MD to generate proposals. We saw many interesting developments in this area. J. Izaguirre presented some work on using the shadow Hamiltonian to reduce the number of rejections. The idea is that in typical HMC you use the amount of energy drift experienced to decide whether or not to accept of MC move. However, it is hard to differentiate between energy drift and a short-term fluctuation just by looking at the energy. However, it has been shown, e.g, [9], how to compute the shadow Hamiltonian value on the fly, and this can be used to determine whether drift is really occurring. Izaguirre and his collaborators have incorporated this into HMC.

HMC in this form, like other Monte Carlo methods, only sample over configuration space. It seems unlikely that there is any way to interpret long trajectories as 'real' dynamics in even some weakened spectral sense (see 3.2). S. Reich and co-worker addressed this issue in taking the approach of Izaguirre even further by reconsidering the means of generating new proposals in HMC. Typically, a proposal is made by discarding all momentum information and re-sampling it from the canonical distribution. Akhamtskaya & Reich suggested (i) to implement HMC over phase space, (ii) to increase the acceptance rate by using a shadow Hamiltonian (see Izaguirre & Hampton), and (iii) to only partially re-sample momenta. It seems feasible that the associated Markov chain has spectral properties that are similar to discrete-time Langevin dynamics and/or DPD in the limit of weak thermal coupling. A main theoretical obstacle is the necessary momentum reversal after a rejected MD step. However, using a sufficiently high-order shadow Hamiltonian, rejections should almost become negligible. This is clearly an area for further research.

As with all MC methods, there is still the issue in HMC of how to optimize parameters to maximize performance. Many authors optimize on the rejection rate. What should be optimized (minimized) is the integrated autocorrelation function, which gives the variance of estimators. Rejection rate affects this, but it is not the whole story.

5 Other Strategies for Efficient Configurational Sampling

An abundance of effective techniques have been recently proposed. There is a need to put these methods in an abstract framework and do a theoretical comparison of efficiency.

Many of these techniques are importance sampling schemes. Importance sampling uses points drawn with biased probability density $\rho_{\mathrm{b}}(q)$ to estimate expectations with respect to the distribution with density $\rho(q)$ using a weighted average, with weights proportional to $\rho(q)/\rho_{\mathrm{b}}(q)$. However, for high-dimensional problems it is difficult to find an easily-sampled $\rho_{\mathrm{b}}(q)$ that is close enough to $\rho(q)$, and for which the density function $\rho_{\mathrm{b}}(q)$ can easily be computed. Commonly the density is known only up to a multiplicative constant, in which case one samples from $\rho_{\mathrm{b}}(q)$ using an MCMC method and uses

$$\langle O(q)\rangle = \frac{\langle O(q)\sigma(q)/\sigma_{\mathrm{b}}(q)\rangle_{\mathrm{b}}}{\langle \sigma(q)/\sigma_{\mathrm{b}}(q)\rangle_{\mathrm{b}}}$$

where σ, σ_{b} denote unnormalized densities. In practice, with finite sampling, this can be approximated by $(\sum_n O(q^n)\sigma(q^n)/\sigma_{\mathrm{b}}(q^n))/(\sum_n \sigma(q^n)/\sigma_{\mathrm{b}}(q^n))$ (which is a biased estimate).

5.1 Fast growth methods

These are importance sampling methods that use the Jarzynski identity [22]. The Hamiltonian is parameterized and changed it as a function of time while doing Markovian dynamics, which must satisfy a "balance condition." Typically the normalizing constant of the initial distribution for the Markovian dynamics is not known and it is sampled using dynamics or an MCMC method.

The talk by R. Neal proposes the use of temperature as a parameter for enabling a sampling not hindered by potential energy barriers. Also, it permits the initialization of the trajectories using configurations drawn exactly from a given distribution (as opposed to an approximate equilibration), so the method becomes one of perfect sampling (no bias due to initial configurations).

Theoretical studies such as [34] indicate that fast growth methods as they are currently formulated offer no advantage over equilibrium sampling, e.g., umbrella sampling.

5.2 Hamiltonian importance sampling

The problem of finding an easily-sampled $\rho_{\mathrm{b}}(q)$ that is close enough to $\rho(q)$, and for which the density $\rho_{\mathrm{b}}(q)$ can easily be computed can be overcome using a $\rho_{\mathrm{b}}(q)$ that is defined in terms of Hamiltonian dynamics. For a molecular system, we sample particle positions uniformly over some (wrapped-around) region, and also sample momenta for these particles from their distribution at a high temperature. We then simulate Hamiltonian dynamics for this system, while periodically multiplying the momenta by some factor slightly less than one, which eventually cools the system to whatever temperature, T, we are interested in. This procedure defines a distribution $\rho_{\mathrm{b}}(q)$, which can be used to estimate expectations with respect to $\rho(q)$, the canonical distribution at temperature T. Crucially, for each sampled q, we can compute $\rho_{\mathrm{b}}(q)$, and hence the appropriate weight to attach to this point in the average: The density of the initial point sampled from the high-temperature distribution is easily calculated; since Hamiltonian dynamics conserves phase space volume, the density of a point found after simulating Hamiltonian dynamics for some time is the same as that of the original point; and finally, multiplying a momentum variable by a factor less than one simply increases the probability density of the resulting point by the same factor.

5.3 Replica exchange

In the discussion on the last day, R. Neal compared parallel tempering with the use of the Jarzynski equality. He said that Jarzynski's method tends to require a finer spacing of distributions than parallel tempering. But on the other hand, information in parallel tempering propagates between distributions via a random walk, which tends to take n^2 steps to move a distance of n. These two effects more-or-less cancel out, so that there is no clear advantage of one method over the other (at least in this respect—their properties differ in other ways that may be relevant in some problems). This points to a research direction of trying to modify one of these methods to obtain the advantageous property possessed by the other method.

5.4 Adaptive biased-force method

Eric Darve presented novel techniques to calculate the potential of mean force along a reaction coordinate, the so-called Adaptive Biasing Force (ABF) and an extension to Monte-Carlo simulations (MC-ABF). The goal of ABF is to improve the sampling of phase space when calculating the free energy along a reaction coordinate $\xi(r)$ or potential of mean force:

$$A = -k_B T \ln\langle\delta(\xi(r) - \xi)\rangle$$

In typical molecular systems, the molecules remain trapped in low energy basins for extensive periods of time escaping only rarely. See Figure 1.

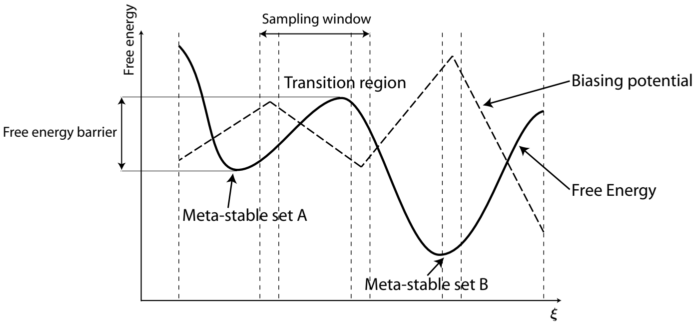


Figure 1: Typical free energy profile showing two basins of attraction and a free energy barrier at transition states. In the method of Umbrella Sampling, a biasing potential is used to improve the sampling of the system. The simulation is often performed inside of windows to improve the efficiency.

ABF improves the efficiency of this type of calculation by applying an external biasing force which leads to a uniform sampling along the reaction coordinate. The biasing force is obtained by applying the following equation for the derivative of the free energy:

$$\frac{dA}{d\xi} = -\left\langle \frac{d}{dt}\left(m_\xi \frac{d\xi}{dt}\right)\right\rangle_\xi, \quad m_\xi^{-1} = \sum_r \frac{1}{m_r}\left(\frac{\partial \xi}{\partial x_r}\right)^2$$

The applied force is taken approximately equal to $(dA/d\xi)\nabla\xi$. It is continuously updated using samples gathered during the simulation.

Darve presented results on dichloro-ethane, fluoro-methane with a water-hexane interface and results on the insertion of a helix inside a membrane: see Figure 2.

Christophe Chipot presented simulation results on the unfolding of deca-alanine using ABF. See Figure 3.

In addition to Molecular Dynamics, Darve presented a new algorithm based on ABF for Metropolis Monte-Carlo simulations, called MC-ABF. This algorithm is based on applying a bias to the transition probabilities. MC-ABF was applied to calculate the density of states for the Ising model. The Ising model is a square system of L^2 spins (up or down) whose energy is given by: $E = -\sum_{i,j} \sigma_i \sigma_j$ where the indices i and j correspond to neighboring spins. In MC-ABF, the density of states $g(E)$ is estimated using a recurrence formula:

$$g(E_i) = g(E_{i-1}) \frac{\Pi^+(E_{i-1})}{\Pi^-(E_i)} \tag{2}$$

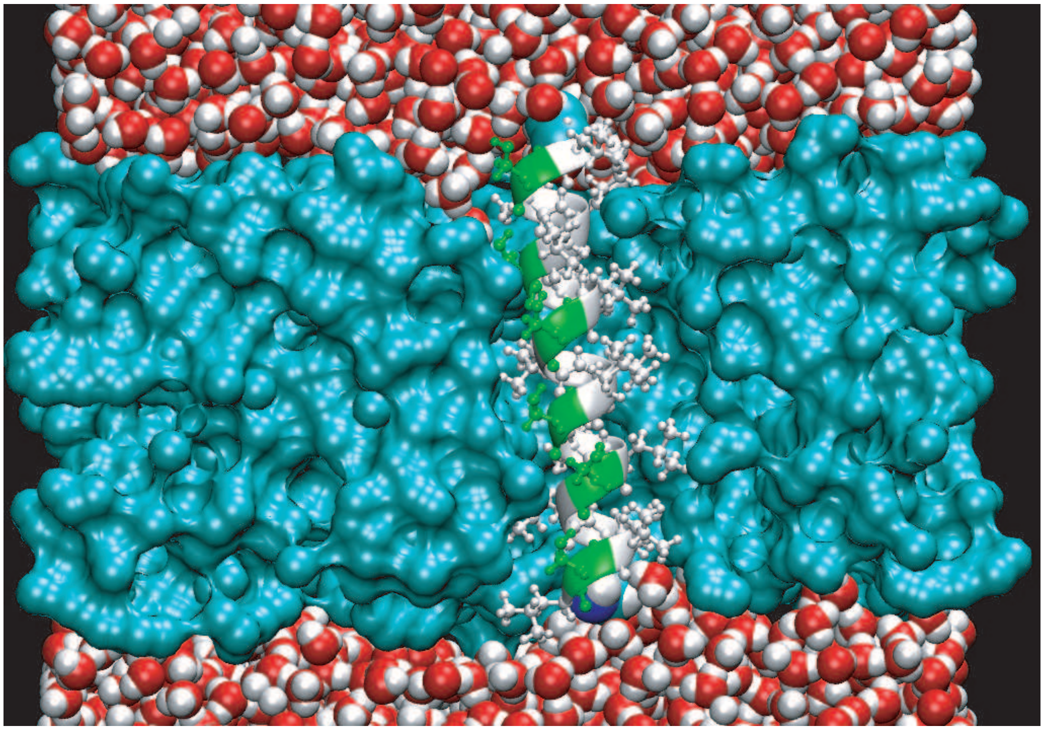
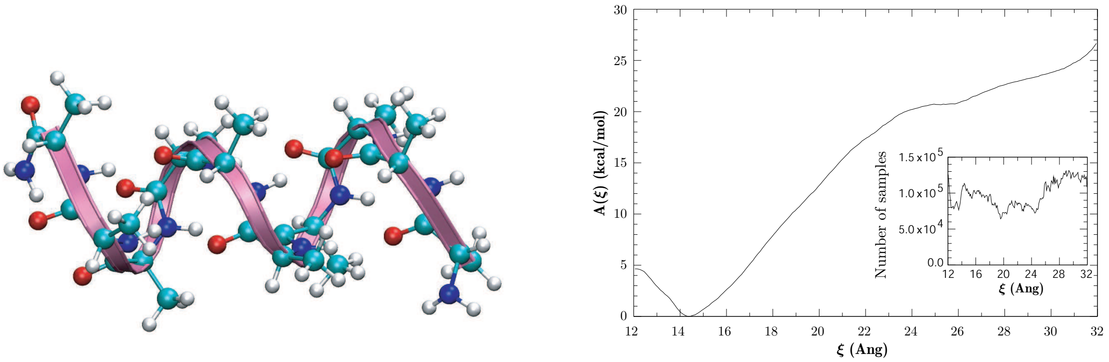


Figure 2: Helix inside a mimetic membrane. This system is used to understand the self-assembly of helices in cell membranes to form ion channels.



(a) Deca–L–alanine in its folded configuration

(b) Free energy profile for Deca–L–alanine calculated using the adaptive biasing force. The inset shows the number of samples as a function of ξ.

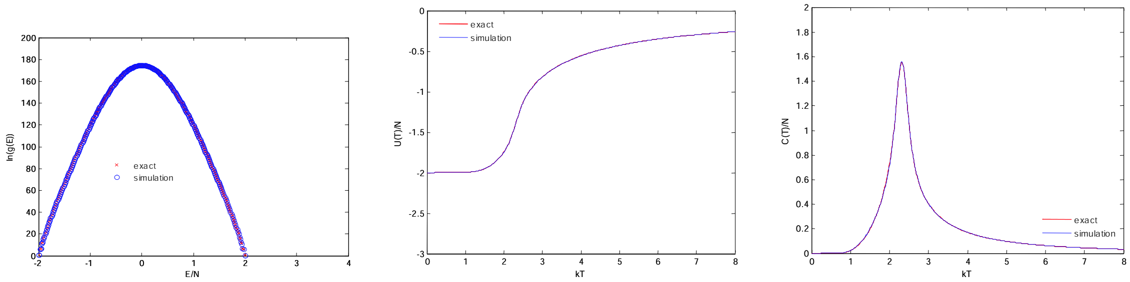
Figure 3: Deca-L-alanine

where $\Pi^+(E_{i-1})$ is the probability to transition from energy E_{i-1} to E_i and $\Pi^-(E_i)$ the probability to transition from E_i to E_{i-1}. The modified acceptance rule is then given by:

$$p_b(E_i \to E_j) = \min\left(\frac{g^{ABF}(E_i)}{g^{ABF}(E_j)}, 1\right)$$

where g^{ABF} is the current estimate of $g(E)$ computed using Equation 2. This biasing leads to a uniform sampling in energy space. In particular states at high and low energies ($2L^2$ and $-2L^2$) are visited as often as intermediate states with energy close to zero. This is despite a difference in population at those energies of about $\exp(180) \approx 10^{80}$.

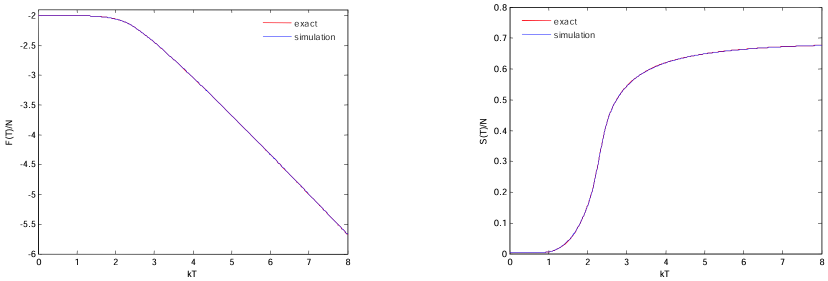
Results were presented for the calculation of the density of states as well as internal energy, specific heat, Helmholtz free energy and Entropy. See Figure 4.



(a) Density of states for Ising model and comparison with analytical solution

(b) Internal energy for Ising model as a function of temperature

(c) Specific heat



(d) Helmholtz free energy

(e) Entropy

Figure 4: Ising model of size 16×16.

5.5 Transformation of Configuration Space Variables

One of the computational grand challenge problems is to develop methodology capable of sampling conformational equilibria in systems with rough energy landscapes. If met, many important problems, most notably protein folding and protein aggregation, could be significantly impacted. In his talk, Tuckerman present a new approach in which molecular dynamics is combined with a novel variable transformation designed to warp configuration space in such a way that barriers are reduced and attractive basins stretched. The essence of the method is as follows: Consider a one-dimensional potential $V(x)$ with a barrier. The canonical partition function is

$$Q = \int dp\, dx\, \exp\left\{-\beta\left[\frac{p^2}{2m} + V(x)\right]\right\}$$

If the barrier in $V(x)$ is high, then if we try to evaluate this partition function using molecular dynamics or Monte Carlo, the presence of the barrier will make crossing events between the two wells on either side of the barrier rare and, hence, hinder the sampling. However, consider a nonlinear variable transformation in the above partition function:

$$u = f(x) = x_0 + \int_{x_0}^{x} dy\; e^{-\beta V_{\rm ref}(y)}$$

where $V_{\rm ref}(x)$ is an arbitrary reference potential. When substituted into the partition function, one obtains

$$Q = \int dp\; du\; \exp\left\{-\beta\left[\frac{p^2}{2m} + (V(x(u)) - V_{\rm ref}(x(u))\right]\right\}$$

One can then use Monte Carlo or thermostated molecular dynamics with a Hamiltonian of the form $p^2/2m + V(x) - V_{\rm ref}(x)$. It is, therefore, very likely that the techniques being developed by Leimkuhler and coworkers will be applicable with this method. Note that this transformation yields exactly the same partition function and therefore, preserves all of the thermodynamic and equilibrium properties of the system. From the form of the transformed partition function, the optimal choice of the reference potential becomes clear: One should choose $V_{\rm ref}(x)$ to be equal to the true potential in the barrier region and zero outside the this region. Note that this is not equivalent to umbrella sampling or guiding potential methods, as the variable u naturally moves on the difference potential $V(x) - V_{\rm ref}(x)$ so that no re-weighting of the phase space is needed. By applying transformations of this type on the full set of backbone dihedral angles in polymers and proteins, Tuckerman was able to show that very large gains in the efficiency of sampling configuration space could be obtained for large polymer chains and small model proteins. The method is further enhanced by including adaptive transformations that remove barriers that arise “on the fly” in a simulation by neighboring solvent atoms or short-range non-bonded type interactions. Currently, Tuckerman and coworkers are implementing the method in their PINY_MD code, a code that contains a full “bio-builder” that will allow all-atom models of proteins to be treated, thereby allowing the method to be tested on realistic problems.

5.6 Multicanonical simulations

Numerical experiments have demonstrated that combining replica exchange with multicanonical Monte Carlo leads to much more effective sampling [30].

6 Sampling from Path Space

Many reactions in (bio)molecular systems occur on time-scales outside the range of current direct molecular dynamics simulations. In simulations of slow molecular reactions, much of the simulation time is in effect ”wasted” waiting for a large enough fluctuation to carry the system from the reactant state to the product state. Following Dellago, Chandler and co-workers [8], one can instead attempt to create only transition paths, i.e., those trajectory segments that connect reactant and product states and exclude the waiting time in the reactant well. However, even if one has “harvested” many such reactive trajectories, it is not always immediately clear where the “bottleneck” of the reaction is (the transition state), and which measure best describes the progress of the reaction (the reaction coordinate). Bayes relation for conditional probabilities can be used to extract transition states and reaction coordinates from an ensemble of transition paths (obtained, e.g., by path sampling) and an equilibrium ensemble (obtained, e.g., by umbrella sampling) [21]:

$$p(\mathrm{TP}|x) = \frac{p(x|\mathrm{TP})p(\mathrm{TP})}{p_{\rm eq}(x)} \tag{3}$$

where $p(x|\mathrm{TP})$ is the probability density of the phase-space variable x, $p_{\rm eq}(x)$ is the equilibrium density, and $p(\mathrm{TP})$ is the fraction of time spent in transition paths, averaged over long equilibrium trajectories. The conditional probability of being on a transition path, $p(\mathrm{TP}|x)$, can be expressed in terms of splitting (or commitment) probabilities, and assumes a maximum at the transition state. The Bayesian relation can be used to locate transition states, optimize reaction coordinates, and calculate rate coefficients directly from path sampling [21, 4]. Moreover, it leads to a simple transition-path sampling algorithm in which trajectories are created by shooting from phase points near a presumed dividing surface between reactant and product states [21, 4].

Another approach to path-sampling was presented by A. Stuart in his talk. He considered the problem in a more general framework: molecular dynamics is just one possible application; another is nonlinear filtering in signal processesing. He described an abstract MCMC method for sampling in such problems, based on generalizing the Metropolis-adjusted Langevin algorithms to infinite dimensions. This leads naturally to the study of stochastic reaction-diffusion equations which, in their invariant measure, sample from the required distribution [16].

7 Coarse graining

Classical MD simulations are typically the method of choice for studying biophysical and soft matter systems at the molecular level. Characteristically, they are limited to system sizes of approximately 10^4 atoms and to times of around 100 ns. Thus, to study systems such as polymer melts, the computationally accessible time and length scales are simply far too short for the system to be able to reach equilibrium as dynamic processes during equilibration occur under hydrodynamic conditions.

A plethora of coarse graining and multiscale modeling methods have been developed to overcome the above difficulties. The approaches range from position space coarse-graining to free energy methods and field theory. Polymer research has been probably the leading field in multiscale modeling in both practical and theoretical aspects. For example, the projection operator formalism has been used by Akkermans and Briels [1] to investigate the fluctuating forces in coarse graining, and and the so-called GENERIC approach [14, 15, 35], which has very different nature but is also based on operator projection formalism, provides an analytically rigorous method for coarse-graining. GENERIC (General Equation for Non-Equilibrium Reversible-Irreversible Coupling) is based on the idea is that there is a general form for the time-evolution of non-equilibrium systems and that it can be given in a general form

$$\frac{dx}{dt} = L(x)\frac{\delta E(x)}{\delta x} + M(x)\frac{\delta S(x)}{\delta x},$$

where x characterizes the state of the system, $L(x)$ is an antisymmetric matrix and $M(x)$ is a symmetric and positive definite matrix. They are connected to the second and the first law of thermodynamics, respectively. $E(x)$ and $S(x)$ are functionals for the total energy and entropy, respectively. It is important to to notice that although GENERIC is rigorous it is not unique. There is no unique way to coarse grain and that constitutes one of the main conceptual difficulties. Other commonly used analytical approaches rely on the Ornstein–Zernike equation [17] and the hypernetted chain closure [5, 26]. A good review of some of the recent developments is given in Ref. [23].

7.1 Dissipative Particle Dynamics (DPD)

As a conceptually simple approach, Dissipative Particle Dynamics (DPD) [10, 20, 24] has recently become popular in soft matter simulations. In DPD the pairwise interaction potentials are “soft” in contrast to the Lennard–Jones -type potentials. “Softness” means that the DPD potential has a finite value at zero particle separation, i.e., the Fermi exclusion principle is not accounted for. It is not obvious that the above approximation is reasonable but Forrest and Suter [13] showed that by explicitly averaging over fluctuations in over long times that approximation becomes justifiable.

Another motivation behind DPD is that it conserves hydrodynamics, i.e., all the interactions are pairwise conservative. In DPD the force exerted on particle i by particle j is

$$\vec{F}_{ij} = \vec{F}^D_{ij} + \vec{F}^R_{ij} + \vec{F}^C_{ij}, \tag{4}$$

where $\vec{F}^D_{ij}$, $\vec{F}^R_{ij}$, and $\vec{F}^C_{ij}$ are the dissipative, random and conservative forces, respectively. The different components are given as

$$\begin{array}{rcl} \vec{F}^D_{ij} & = & -\gamma\,\omega^D(r_{ij})(\vec{v}_{ij}\cdot\vec{e}_{ij})\,\vec{e}_{ij} \quad \text{and} \\ \vec{F}^R_{ij} & = & \sigma\,\omega^R(r_{ij})\,\xi_{ij}\,\vec{e}_{ij}, \end{array} \tag{5}$$

where $\vec{r}_{ij} \equiv \vec{r}_i - \vec{r}_j$, $r_{ij} \equiv |\vec{r}_{ij}|$, $\vec{e}_{ij} \equiv \vec{r}_{ij}/r_{ij}$, and $\vec{v}_{ij} \equiv \vec{v}_i - \vec{v}_j$ for particles with positions $\vec{r}_i$ and velocities $\vec{v}_i$. The ξ_{ij} are symmetric Gaussian random variables with zero mean and unit variance. They are independent for different *pairs* of particles and different times.

The coupling of the the dissipative and random forces, $\vec{F}^D_{ij}$ and $\vec{F}^R_{ij}$, comes from the fact that the thermal heat generated by the random force must be balanced locally by dissipation. The precise relationship between these two forces is determined by the fluctuation-dissipation theorem [10] and is given as

$$\omega^D(r_{ij}) = [\omega^R(r_{ij})]^2 \quad \text{and} \quad \sigma^2 = 2\gamma k_\mathrm{B} T^*, \tag{6}$$

where T^* is the canonical temperature of the system.

The most common choice for the weight functions ω^D and ω^R is the soft-repulsive form

$$\omega^R(r_{ij}) = \begin{cases} 1 - r_{ij}/r_c & \text{for } r_{ij} < r_c; \\ 0 & \text{for } r_{ij} > r_c, \end{cases} \tag{7}$$

where r_c is the cut-off distance and $\omega^D(r_{ij})$ is given by the fluctuation-dissipation relation above. This is also the form that the conservative force takes in the standard DPD formulation.

In addition, DPD can be used as a momentum conserving thermostat. That, and issues related to integration of the DPD equations of motion are discussed in Refs. [3, 32, 39].

7.1.1 Systematic derivation of DPD

The standard DPD is purely phenomenological. Recently, Flekkøy et al. [11, 12] were able to formally link DPD to molecular level properties by using a Voronoi tessellation based technique. The advantage of their method is that it can be used to resolve different length scales simultaneously. The method is formally akin to the well-known renormalization group procedure extensively used in the theory of critical phenomena.

7.2 Effective interactions from the pair correlations

Inverting the radial distribution functions $g(r)$ in order to obtain pair potentials offers another starting point. That approach can be used to obtain the pair potentials for DPD simulations in a more realistic and systematic way. Using pair correlations is based on the so-called Henderson theorem [18] which stipulates that under fixed conditions two pair potentials which give rise to the same $g(r)$ cannot differ by more than a constant. This constant is determined by the condition

$$V(r \to \infty) \to 0, \tag{8}$$

where r is the interparticle distance and V is the pair potential. The Henderson theorem analogous to the Hohenberg–Kohn theorem [19], i.e., all ground state properties are determined by the electron density. What makes the application of the Henderson theorem appealing is that the radial distribution function obtained from a simulation includes effects from the many-body interactions. Furthermore, this way it is possible to define new interaction sites and to compute the radial distribution function between them, and thus to systematically obtain new coarse-grained models at different levels of description.

The Inverse Monte Carlo (IMC) procedure of Lyubartsev and Laaksonen [28] is practical implementation of Henderson's idea. In IMC, one inverts the radial distribution functions – experimental or from microscopic simulations – to obtain effective potentials for a coarse-grained model with a fewer number of degrees of freedom. This approach has been recently used to study salt solutions and lipids [27, 31]

It is important to notice that the effective potential includes corrections from many-body interactions to the well-defined potential of mean force (PMF) [17], which is defined as

$$V^{\mathrm{pmf}}(r) \equiv -k_\mathrm{B} T \ln g(r), \tag{9}$$

where r is the interparticle distance and $g(r)$ the pair correlation function. In other words, the effective potential is *not* the same as the potential of mean force in Eq. (9). Inclusion of the many-body corrections is the reason why an iterative IMC scheme is needed. The iterative IMC procedure guarantees self-consistency, i.e., the effective potentials lead to the same pair correlation behavior as the underlying MD simulations.

7.3 Coarse graining – the future

The above provides just a scratch on the surface. There are a lot of other interesting and promising approaches. In general, multiscale modeling is a very rapidly developing field and progress is partly driven by the fact that despite the increase in CPU power, it is algorithm and method development that is crucial for treating complex problems such as protein folding, polymers, or cellular membranes.

Acknowledgment

The organizers are grateful to Eric Darve, Gerhard Hummer, Mikko Karttunen, Ben Leimkuhler, Radford Neal, Sebastian Reich, John Straub, Andrew Stuart, Mark Tuckerman for contributions of discussion and text to this report.

References

[1] R. L. C. Akkermans and W. J. Briels, J. Chem. Phys. **113**, 6409 (2000).

[2] B. J. Berne and J. E. Straub. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opinion Struct. Biol.*, 7:181–189, 1997.

[3] G. Besold, I. Vattulainen, M. Karttunen, and J. M. Polson, Phys. Rev. E **62**, R7611 (2000).

[4] R. B. Best and G. Hummer, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6732 (2005).

[5] P. G. Bolhuis, A. A. Louis, J.-P. Hansen, and E. J. Meijer, J. Chem. Phys. **114**, 4296 (2001).

[6] S. D. Bond, B. J. Leimkuhler, and B. B. Laird. The Nosé-Poincaré method for constant temperature molecular dynamics. *J. Comput. Phys*, 151(1):114–134, May 1, 1999.

[7] W. J. Briels and R. L. C. Akkermans, Mol. Sim. **28**, 145 (2002).

[8] C. Dellago, P. G. Bolhuis, F. S. Csajka and D. Chandler, *J. Chem. Phys.* **108**, 1964 (1998)

[9] R. D. Engle, R. D. Skeel, and M. Drees. Monitoring energy drift with shadow Hamiltonians. *J. Comput. Phys.*, 206(2):432–452, July 1, 2005.

[10] P. Español and P. Warren, Europhys. Lett. **30**, 191 (1995).

[11] E. G. Flekkoy, P. V. Coveney, and G. De Fabritiis, Phys. Rev. E **62**, 2140 (2000).

[12] E. G. Flekkoy and P. V. Coveney, Phys. Rev. Lett. **83**, 1775 (1999).

[13] B. M. Forrest and U. W. Suter, J. Chem. Phys. **102**, 7256 (1995).

[14] M. Grmela and H. C. Öttinger, Phys. Rev. E **56**, 6620 (1997).

[15] M. Grmela and H. C. Öttinger, Phys. Rev. E **56**, 6633 (1997).

[16] M. Hairer, A. M. Stuart, and J. Voss. A Bayesian Approach to Data Assimilation. *Physica D*, to appear.

[17] J. P. Hansen and I. R. McDonald, *Theory of Simple Liquids*, Academic press, London (1986).

[18] R. L. Henderson, Phys. Lett. **49**A, 197 (1974).

[19] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).

[20] P. J. Hoogerbrugge and J. M. V. A. Koelman, Europhys. Lett. **19**, 155 (1992).

[21] G. Hummer, *J. Chem. Phys.* **120**, 516 (2004).

[22] C. Jarzynski. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, 78(14):2690–2693, Apr. 7, 1997.

[23] M. Karttunen, I. Vattulainen, and A. Lukkarinen (eds.), *Novel Methods in Soft Matter Simulations*, Springer-Verlag, Berlin (2004).

[24] J. M. V. A. Koelman and P. J. Hoogerbrugge, Europhys. Lett. **21**, 363 (1993).

[25] R. Kupferman, A. M. Stuart, J. R. Terry, and P. F. Tupper. Long Term Behaviour of Large Mechanical Systems with Random Initial Data. *Stochastics and Dynamics*, 2(4):533–562, 2002.

[26] A. A. Louis, P. G. Bolhuis, J.-P. Hansen, and E. J. Meijer, Phys. Rev. Lett. **82**, 2522 (2000).

[27] A. P. Lyubartsev, M. Karttunen, I. Vattulainen, and A. Laaksonen, Soft Materials **1**, 121 (2003).

[28] A. P. Lyubartsev and A. Laaksonen, Phys. Rev. E **52**, 3730 (1995).

[29] J. C. Mattingly, A. M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: Locally Lipschitz vector fields and degenerate noise. *Stochastic Process. Appl.*, 101(2):185–232, 2002.

[30] A. Mitsutake, Y. Sugita, and Y. Okamoto. Replica-exchange multicanonical and multicanonical replica-exchange Monte Carlo simulations of peptides. I. Formulation and benchmark test. *J. Chem. Phys.*, 118(14):6664–6675, Apr. 8, 2003.

[31] T.Murtola, E.Falck, M.Patra, M.Karttunen, and I.Vattulainen, J. Chem. Phys **121**, 9156 (2004).

[32] P. Nikunen, M. Karttunen, and I. Vattulainen, Comput. Phys. Commun. **153**, 407 (2003).

[33] S. Nosé. A unified formulation of the constant temperature molecular dynamics methods. *J. Chem. Phys.*, 81:511, 1984.

[34] H. Oberhofer, C. Dellago, and P. L. Geissler. Biased sampling of non-equilibrium trajectories: Can fast switching simulations outperform conventional free energy calculation methods? *J. Phys. Chem. B*, 109:6902–6915, 2005.

[35] H. C. Öttinger, Phys. Rev. E **57**, 1416 (1998).

[36] N. Singhal, C. D. Snow, and V. S. Pande. Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.*, 121(1):415–425, 2004.

[37] A. M. Stuart and J. O. Warren. Analysis and Experiments for a Computational Model of a Heat Bath *Journal of Statistical Physics*, 97, 1999.

[38] P. F. Tupper. A Test Problem for Molecular Dynamics Integrators. *IMA J. Num. Analysis*, 25, 2005.

[39] I. Vattulainen, M. Karttunen, G. Besold, and J. M. Polson, J. Chem. Phys. **116**, 3967 (2002).