An Interacting Particle System Approach for Molecular Dynamics

Mathias Rousset
Laboratoire de Statistique et Probabilités, Université Paul Sabatier
118, Route de Narbonne,
31062 TOULOUSE Cedex 4 FRANCE
rousset@cict.fr

Gabriel Stoltz
CERMICS, Ecole Nationale des Ponts et Chaussées,
6/8 avenue Blaise Pascal,
77455 Marne-la-Vallée, France
and
CEA/DAM Ile-de-France,
BP 12, 91680 Bruyères-le-Châtel, France
stoltz@cermics.enpc.fr

August 9, 2005

Abstract

We present here some applications of an Interacting Particle System (IPS) methodology to the field of Molecular Dynamics. This IPS method allows several simulations of a same out of equilibrium process to keep closer to equilibrium at each time, thanks to a selection mechanism based on the relative virtual work induced on the system. It is therefore an improvement of usual simulated annealing type processes used to compute canonical distributions of the system and free energy differences.

Keywords: Canonical distribution, molecular dynamics, simulated annealing, interacting particle system, genetic algorithms, Sequential Monte-Carlo, Feynman-Kac formula.

AMS: 65C05, 65C35, 80A10.

Phase-space integrals are widely used in Statistical Physics to relate the macroscopic properties of a system to the elementary phenomena at the mi-

croscopic scale [13]. In constant temperature (NVT) molecular simulations, these integrals often take the form

$$\mu(A) = \langle A \rangle = \int_{T^*\mathcal{M}} A(q,p)\, d\mu(q,p). \tag{1}$$

where $\mathcal{M}$ denotes the position space (also called the *configuration space*), and $T^*\mathcal{M}$ denotes its cotangent space. A generic element of the position space $\mathcal{M}$ will be denoted by $q = (q_1, \cdots, q_N)$ and a generic element of the momentum space by $p = (p_1, \cdots, p_N)$. We will consider here that $\mathcal{M} \sim \mathbb{R}^{3N}$ or $\mathbb{T}^{3N}$ (a torus of dimension $3N$, which arises when using periodic boundary conditions), and that $T^*\mathcal{M} \sim \mathbb{R}^{3N} \times \mathbb{R}^{3N}$ or $\mathbb{T}^{3N} \times \mathbb{R}^{3N}$, though in general more complicated situations should be considered, when performing Blue Moon sampling [5, 6] for example.

The measure μ is the canonical probability measure

$$d\mu(q,p) = Z^{-1} \exp(-\beta H(q,p))\, dq\, dp, \tag{2}$$

where $\beta = 1/k_{\mathrm{B}}T$ (T denotes the temperature and k_{B} the Boltzmann constant) and where H denotes the Hamiltonian of the molecular system:

$$H(q,p) = \frac{1}{2} p^T M^{-1} p + V(q). \tag{3}$$

In the above expression, V is the potential experienced by the N particles, and $M = \mathrm{Diag}(m_1, \cdots, m_N)$ where m_i is the mass of the i-th particle. The constant Z in (2) is the normalization constant defined as

$$Z = \int_{T^*\mathcal{M}} \exp(-\beta H(q,p))\ dq\, dp.$$

Some quantities can not be expressed through relations such as (1). One important example is the free energy of a system, defined as

$$F = -\beta^{-1} \ln Z. \tag{4}$$

We refer to [4] for a review of different sampling methods in molecular dynamics for the computation of phase space integrals of type (1). However, it is often the case in practice that a straightforward sampling of μ is difficult. Indeed, high dimensional systems exhibit many local minima in which the system remains trapped, especially when the temperature is low. In those cases, alternative approaches have to be used, such as those building on the *simulated annealing* [18] paradigm. An interesting review of such Monte Carlo methods can be found in [15]. The idea of simulated annealing is to start a simulation at a high temperature, so that the system can evolve freely without remaining stuck in a metastable state, and then to decrease gradually the temperature. Of course, this method can be generalized by interpolating between an initial simple sampling problem, and the

target sampling problem. This allows to attain deeper local minima, but is not efficient as such to sample the whole target measure. Improvements have therefore been proposed to this end. For example, *parallel tempering* is a well-known method in molecular dynamics (see [15] for references) which consists in simulating several replicas of a same system at different temperatures, and exchanging replicas simulated at neighboring temperatures, according to some Metropolis-Hastings rule. Therefore, energetic barriers at the lower temperature can be crossed thanks to the crossings done at higher temperatures. The method requires however some tuning in the distribution of the temperatures, which is usually achieved by running preliminary simulations. Another method is *simulated tempering* [20], where the temperature itself is treated as a dynamical variable. It is to some extent a 'sequential' version of the previous method, particularly interesting when the systems to simulate are large. However, this methods asks for an a priori function controlling the distribution of the temperature, which is also estimated through some preliminary runs [15]. A last method is *multicanonical Monte Carlo* [2] which relies on the importance sampling paradigm. As all importance sampling procedures however, a tuning of the importance function is crucial, and unless a better idea, an iterative procedure of preliminary runs has also to be done first in order to obtain a reasonable importance function.

We present here a complementary approach to the above simulated annealing type strategies, which needs no tuning of parameters. It consists in running M simulations of the system (called 'replicas' or 'walkers') in parallel, resorting typically to a Markovian dynamic, and considering exchanges between the replicas, according to a certain probability depending on the work done on each system. These heuristic explanations are precised in section 2. The set of all replicas (or walkers) is called an 'Interacting Particle System' (IPS) [9], and the methodology is widely used in the fields of Quantum Monte Carlo [1, 24] or Bayesian Statistics, where it is referred to as Sequentiel Monte Carlo [10, 8].

The article is organized as follows. We first precise classical simulated annealing type methods in section 1. We then describe the associated IPS method in section 2, as well as its numerical implementation. Possible applications and some numerical results are then presented in section 3.

1 Simulated annealing type methods

Consider a family of Hamiltonian functions $H_\lambda : T^*\mathcal{M} \to \mathbb{R}$ indexed by a parameter $\lambda \in [0,1]$. The corresponding Hamiltonian dynamics are

$$\begin{cases} \dfrac{dq}{dt} = \dfrac{\partial H_\lambda}{\partial p} \\ \dfrac{dp}{dt} = -\dfrac{\partial H_\lambda}{\partial q} \end{cases} \tag{5}$$

Typically, we can consider $H_\lambda(q,p) = (1-\lambda)H_0(q,p) + \lambda H_1(q,p)$. The family $(H_\lambda)_{\lambda\in[0,1]}$ indexes a path between the original state described by a Hamiltonian H_0 and the final state charaterized by a Hamiltonian H_1. A canonical probability measure μ_λ can be associated to each Hamiltonian H_λ :

$$d\mu_\lambda(q,p) = \frac{1}{Z_\lambda} \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp, \tag{6}$$

where the normalizing constant Z_λ is

$$Z_\lambda = \int_{\mathcal{M}} \mathrm{e}^{-\beta H_\lambda(q,p)}\, dq\, dp.$$

Our aim is the following. We wish to sample according to $d\mu_1$, which may be a difficult task, whereas sampling according to $d\mu_0$ is assumed to be easy. A natural idea is to use a sampling of the measure $d\mu_0$ to obtain a sampling of $d\mu_1$. This philosophy is reminiscent of the simulated annealing method [18]. For this method, the simulation is started in conditions such that the dynamic is ergodic, with a sufficiently fast numerical convergence. A sample of the initial measure $d\mu_0$ can then be computed. For each point of the previous sample, the corresponding configuration of the system is brought slowly to the end state along a path $(\lambda(t))_{t\in[0,T]}$ for a (large) time $T > 0$. Therefore, the final sample of configurations is distributed according to $d\mu_1$.

Such a transition can refer to a change of temperature of the system from β to β', in which case

$$H_0(q,p) = H(q,p), \quad H_1(q,p) = \frac{\beta'}{\beta} H(q,p).$$

It can also represent a modification of the potential, sometimes called 'alchemical transition' in the physics and chemistry litterature. The folding of a protein can be studied this way for example, by setting initially all the long-range interactions to zero, whereas the final state corresponds to a Hamiltonian were all interactions are set on. In this case,

$$H_\lambda(q,p) = \frac{1}{2} p^T M^{-1} p + V_\lambda(q).$$

The usual way to achieve simulated annealing is to perform a time inhomogeneous irreducible Markovian dynamic

$$t \mapsto X_t^{\lambda(t)}, \quad X_0^{\lambda(0)} \sim \mu_0, \tag{7}$$

for $t \in [0,T]$, and a smooth schedule $t \mapsto \lambda(t)$ verifying $\lambda(1) = 0$ and $\lambda(T) = 1$, and such that $d\mu_\lambda$ is for all $\lambda \in [0,1]$ an invariant measure under the time homogeneous dynamic $s \mapsto X_s^\lambda$.

The variable x can represent the whole degrees of freedom (q,p) of the system, or only the configuration part q. Depending on the context, the invariant measure μ will therefore be the canonical measure (2), or its marginal with respect to the momenta, which reads

$$d\tilde{\mu}_\lambda(q) = \frac{1}{\tilde{Z}_\lambda} \mathrm{e}^{-\beta V_\lambda(q)}\, dq, \tag{8}$$

with

$$\tilde{Z}_\lambda = \int_{\mathcal{M}} \mathrm{e}^{-\beta V_\lambda(q)}\, dq.$$

When we do not wish to precise further the dynamics, we simply call $d\mu_\lambda(x)$ the invariant measure, and x the configuration of the system. The actual invariant measure should be clear from the context.

For all $t \in [0,T]$, the dynamic (7) will be usefully characterized by its infinitesimal generator $L_{\lambda(t)}$, defined on a domain of continuous bounded test functions by:

$$L_{\lambda(t)}(f)(x) = \lim_{h \to 0} \frac{1}{h} \big(\mathbb{E}(f(X_{t+h}^{\lambda(t+h)}) | X_t^{\lambda(t)} = x) - f(x) \big)$$

The invariance of $\mu_{\lambda(t)}$ under the instantaneous dynamic can be expressed through the balance condition:

$$\forall f, \quad \mu_{\lambda(t)}(L_{\lambda(t)}(f)) = 0. \tag{9}$$

The dynamics we have in mind are (for a $f$ixed $\lambda \in [0,1]$):

- The hypo-elliptic Langevin dynamic on $T^*\mathcal{M}$

$$\left\{ \begin{array}{ll} dq_t^\lambda & = \dfrac{\partial H_\lambda}{\partial p}(q_t^\lambda, p_t^\lambda)\, dt \\ dp_t^\lambda & = -\dfrac{\partial H_\lambda}{\partial q}(q_t^\lambda, p_t^\lambda)\, dt - \xi M^{-1} p_t^\lambda\, dt + \sigma\, dW_t \end{array} \right. \tag{10}$$

  where W_t denotes a standard $3N$-dimensional Brownian motion. The paradigm of Langevin dynamics is to introduce in the Newton equations of motion (5) some fictitious brownian forces modelling fluctuations, balanced by viscous damping forces modelling dissipation. The parameters $\sigma, \xi > 0$ represent the magnitude of the fluctuations

and of the dissipation respectively, and are linked by the fluctuation-dissipation relation:

$$\sigma = (2\xi/\beta)^{1/2}. \tag{11}$$

Therefore, there remains one adjustable parameter in the model. The infinitesimal generator is given by:

$$L_\lambda f = \frac{\partial H_\lambda}{\partial p} \cdot \nabla_q f - \frac{\partial H_\lambda}{\partial q} \cdot \nabla_p f - \xi M^{-1} p \cdot \nabla_p f + \frac{\xi}{\beta} \Delta_p f.$$

- The elliptic overdamped Langevin dynamic[1] in the configuration space $\mathcal{M}$:

$$dq_t^\lambda = -\nabla V_\lambda(q_t^\lambda)\, dt + \sigma\, dW_t, \tag{12}$$

where the magnitude of the random forcing is given here by

$$\sigma = \sqrt{\frac{2}{\beta}}.$$

The corresponding infinitesimal generator is given by:

$$L_\lambda f = \frac{1}{\beta} \Delta_q f - \nabla V_\lambda(q) \cdot \nabla_q f.$$

Let us remark that the overdamped Langevin dynamic (12) is obtained from the Langevin dynamic (10) by letting the mass matrix M go to zero and by setting $\xi = 1$, which amounts here to rescaling the time.

It is well known that, for a *f*ixed $\lambda \in [0, 1]$, these dynamics are ergodic under mild assumptions on the potential V [4].

When the schedule is sufficiently slow, the dynamic is said quasi-static, and the law of the process $X_t^{\lambda(t)}$ is assumed to stay close to its local steady state throughout the transformation. Unfortunately, this is out of reach at low temperature where local minima of H define metastable states from which the typical escape time is very long. Since a good transition path must then be very slow and long, the associated computational cost becomes prohibitive (more precisely, large deviation results [12] ensure that the typical escape time from metastable states grows exponentially fast with β, and quasi-static transformations should then be exponentially slow with β).

It is therefore interesting to consider approaches building on the simulated annealing formalism, but able to deal with much faster transition schemes. The IPS method is such an approach.

[1] This dynamic is actually known as the 'Langevin dynamic' in the probability and statistics fields. We adopt here the physical names of these stochastic processes, which are more natural when dealing with molecular dynamics.

2 The Interacting Particle System method

Our strategy is inspired by the Sequential Monte Carlo philosophy [8, 11]. The idea is to rewrite the Boltzmann probability path $t \mapsto d\mu_{\lambda(t)}(x)$ as the evolution law of a non-linear jump-diffusion process, and to deduce an associated interacting particle system approximation.

We present in section 2.1 the arguments motivating a selection function based on the virtual work done on the system. The non-linear jump-diffusion process is then described in section 2.2. The IPS approximation is dealt with in section 2.3, as well as convergence results of the discretized measure to the target measure. We end up with some considerations regarding the numerical implementation of the IPS method in section 2.4.

2.1 A virtual work based selection

Let us differentiate the *non-normalized* Boltzmann path $t \mapsto \Pi_{\lambda(t)}(dx) = \mathrm{e}^{-\beta H_{\lambda(t)}(x)}\, dx$ with respect to t:

$$\partial_t \Pi_{\lambda(t)}(f) = -\Pi_{\lambda(t)}\left(\beta \frac{\partial H_{\lambda(t)}}{\partial \lambda}\lambda'(t) f\right).$$

Using the balance condition (9),

$$\partial_t \Pi_{\lambda(t)}(f) = \Pi_{\lambda(t)}\left(L_{\lambda(t)}(f) - \beta \frac{\partial H_{\lambda(t)}}{\partial \lambda}\lambda'(t) f\right). \tag{13}$$

This leads to the following Feynamn-Kac representation [17] of the canonical distribution (which satisfies the same evolution equation):

$$\frac{\Pi_{\lambda(t)}(f)}{\Pi_{\lambda(0)}(1)} = \mathbb{E}\left(f(X_t^{\lambda(t)})\mathrm{e}^{-\beta \int_0^t \frac{\partial H_{\lambda(s)}}{\partial \lambda}(X_s^{\lambda(s)})\lambda'(s)\, ds}\right). \tag{14}$$

This result is essentially the same as the one of [16] and can be expressed through the same thermodynamical relations. Indeed, denoting by

$$W_t = \int_0^t \frac{\partial H_{\lambda(s)}}{\partial \lambda}(X_s^{\lambda(s)})\lambda'(s)\, ds \tag{15}$$

the out of equilibrium virtual work induced on the system on the time schedule $[0, t]$, and taking $f = 1$, it follows

$$\mathbb{E}(\mathrm{e}^{-\beta W_t}) = \mathrm{e}^{-\beta(F(\lambda(t)) - F(0))}. \tag{16}$$

Jensen's inequality then gives

$$\mathbb{E}(W_t) \geq F(\lambda(t)) - F(0). \tag{17}$$

This inequality is an equality if and only if the transformation is quasi-static on $[0,t]$; in this case the random variable W_t is actually constant and equal to ΔF. When the evolution is reversible, this means that equilibrium is maintained at all times.

This highlights the possible use of the virtual work W_t as a quantification of closeness to a quasi-static transformation. We will use this quantity to perform a selection between replicas of the same system evolving according to (7).

2.2 The non-linear jump-diffusion process.

The key idea is to differentiate the *normalized* Boltzmann (or Feynman-Kac) path $t \mapsto \mu_{\lambda(t)}$ and to rewrite it in a non-linear Markovian evolution form. The differentiation of the *normalized* Boltzmann gives

$$\partial_t \mu_{\lambda(t)}(f) = \mu_{\lambda(t)}\left(L_{\lambda(t)}(f) + \beta\left(\mathcal{F}_{\lambda(t)} - \frac{\partial H_{\lambda(t)}}{\partial \lambda}\right)\lambda'(t) f\right), \qquad (18)$$

where $\mathcal{F}_\lambda = \mu_\lambda\left(\frac{\partial H_\lambda}{\partial \lambda}\right)$. This equation can be rewritten as

$$\partial_t \mu_{\lambda(t)}(f) = \mu_{\lambda(t)}\left(L_{\lambda(t)}(f) + J_t(f)\right), \qquad (19)$$

where the jump generator J_t is defined as

$$J_t(f)(x) = \int_{\mathcal{M}} (f(y) - f(x))(\alpha_t^-(x) + \alpha_t^+(y))\mu_{\lambda(t)}(dy),$$

with transition intensities

$$\begin{aligned} \alpha_t^-(x) &= \beta\lambda'(t)\left(\mathcal{F}_{\lambda(t)} - \frac{\partial H_{\lambda(t)}}{\partial \lambda}\right)^-(x), \\ \alpha_t^+(y) &= \beta\lambda'(t)\left(\mathcal{F}_{\lambda(t)} - \frac{\partial H_{\lambda(t)}}{\partial \lambda}\right)^+(y). \end{aligned}$$

The equivalence of (18) and (19) can be checked by a straightforward integration.

The dynamic (19) gives rise to a jump-diffusion process $t \mapsto Y_t$ which evolves according to the following stochastic rules (some facts about pure Markov jump processes are recalled in the Appendix):

Process 1. *Generate Y_0 from $d\mu_0(x)$. Generate idependent clocks $(\tau_n^b, \tau_n^d)_{n\geq 1}$ from an exponential law of mean 1 (the upperscripts b and d refer to 'birth' and 'death' respectively), and initialize the jump times $T^{b/d}$ as $T_0^d = 0, T_0^b = 0$.*

For $0 \leq t \leq T$,

- *Between each jump time, $t \mapsto Y_t$ evolves according to the dynamic* (7)*;*
- *At random times T^d_{n+1} defined by*

$$\int_{T^d_n}^{T^d_{n+1}} \alpha^-_s(Y_s)ds = \tau^d_{n+1},$$

the process jumps to a configuration y, chosen according to the probability measure $d\mu_{\lambda(T^d_{n+1})}(y)$;

- *At random times T^b_{n+1} defined by*

$$\int_{T^b_n}^{T^b_{n+1}} \mu_{\lambda(s)}(\alpha^+_s)ds = \tau^b_{n+1},$$

the process jumps to a configuration y, chosen according to the probability measure $\dfrac{\alpha^+_{T^b_{n+1}}(y)}{\mu_{\lambda(T^b_{n+1})}(\alpha^+_{T^b_{n+1}})} d\mu_{\lambda(T^b_{n+1})}(y)$.

Then, for all $t \geq 0$, the law of Y_t is by construction $\mu_{\lambda(t)}$.

2.3 The Interacting Particle System approximation

We now present a particle interpretation of the process (19) enabling a numerical computation through the use of empirical distributions. Consider M Markovian systems described by variables X^k_t $(0 \leq k \leq M)$. Each system is called a 'walker' or 'particle' in the probability and statistics fields. We use here the name 'replica', which is more apppropriate to the Molecular Dynamics context.

The global evolution of the M replicas can be characterized by the fact that every single replica is subjected to the dynamic (19) of generator $L_{\lambda(t)} + J_t$ in a mean-field interpretation. To this purpose, we approximate the virtual force by

$$\mathcal{F}^M_{\lambda(t)} = \frac{1}{M}\sum_{k=1}^{M} \frac{\partial H_{\lambda(t)}}{\partial \lambda}(X^k_t)$$

and the Boltzmann distribution by

$$d\mu^M_{\lambda(t)}(x) = \frac{1}{M}\sum_{k=1}^{M} \delta_{X^k_t}(dx),$$

which are their empirical versions.

The replicas evolve according to the following stochastic rules (see [24, 25] for further details):

Process 2. *Consider an initial distribution* $(X_0^1, \dots, X_0^M)$ *generated from* $d\mu_0(x)$. *Generate idependent times* $\tau_1^{k,b}, \tau_1^{k,d}$ *from an exponential law of mean* 1 *(the upperscripts* b *and* d *refer to 'birth' and 'death' respectively), and initialize the jump times* $T^{b/d}$ *as* $T_0^{k,d} = 0, T_0^{k,b} = 0$.

For $0 \leq t \leq T$,

- *Between each jump time, evolve independently the replicas* X_t^k *according to the dynamic* (7)*;*

- *At random times* $T_{n+1}^{k,d}$ *defined by*

$$\beta \int_{T_n^{k,d}}^{T_{n+1}^{k,d}} \left(\mathcal{F}_{\lambda(s)}^M - \frac{\partial H_{\lambda(s)}}{\partial \lambda}(X_s^k) \right)^- \lambda'(s)\, ds = \tau_{n+1}^{k,d},$$

  an index $l \in \{1, \dots, M\}$ *is picked at random, and the configuration of the* k*-th replica is replaced by the configuration of the* l*-th replica. A time* $\tau_{n+2}^{k,d}$ *is generated from an exponential law of mean* 1*;*

- *At random times* $T_{n+1}^{k,b}$ *defined by*

$$\beta \int_{T_n^{k,b}}^{T_{n+1}^{k,b}} \left(\mathcal{F}_{\lambda(s)}^M - \frac{\partial H_{\lambda(s)}}{\partial \lambda}(X_s^k) \right)^+ \lambda'(s)\, ds = \tau_{n+1}^{k,b},$$

  an index $l \in \{1, \dots, M\}$ *is picked at random, and the configuration of the* l*-th replica is replaced by the configuration of the* k*-th replica. A time* $\tau_{n+2}^{k,b}$ *is generated from an exponential law of mean* 1.

The selection mechanism therefore favors replicas which are sampling values of the virtual work W_t lower than the empirical average. The system of replicas is 'self-organizing' to keep closer to a quasi-static transformation.

Remark 1. *Process (2) is different and much more symmetric in its presentation than Process (1). This is due to the fact that, for Process (1), we take for the treatment of the 'birth' part (the positive part of* $\mathcal{F}_{\lambda(s)} - \frac{\partial H_{\lambda(s)}}{\partial \lambda}$*), the point of view of the jumping replica which is attracted by another one; whereas in Process (2), we take the point of view of the latter attracting replica which induces a branching.*

In [9, 24], several convergence results and statistical properties of the replicas distribution are proven. They are summarized in the following

Proposition 2.1. *Assume that* $(t, x) \mapsto \frac{\partial H_{\lambda(t)}}{\partial \lambda}(x)$ *is a continuous bounded function on* $[0, T] \times T^*\mathcal{M}$ *(or* $[0, T] \times \mathcal{M}$ *in the case of overdamped Langevin dynamics), and that the dynamic* (7) *is Fellerian and irreducible. Then*

- *For any* $t \in [0,T]$,

$$\mu^M_{\lambda(t)}(f) \exp\left(-\beta \int_0^t \mathcal{F}^M_{\lambda(s)} \lambda'(s)\, ds\right)$$

is an unbiased estimator of (14)*;*

- *For all test function* f *and any* $t \in [0,T]$, $\mu^M_{\lambda(t)}(f)$ *is an asymptotically normal estimator of* $\mu_{\lambda(t)}(f)$*, with bias and variance of order* M^{-1}*.*

The proof follows from Lemma 3.20, Proposition 3.25 and Theorem 3.28 of [9] (see also [24, 25] for further details).

2.4 Numerical implementation

In the previous section, we discretized the measure by considering an empirical approximation. For a numerical implementation to be tractable, it remains to discretize time.

Notice already that the IPS method induces no extra computation of the forces, and is therefore unexpensive to implement. However, although the IPS can be parallelized, the processors have to exchange informations at the end of each time step, which can slow down the simulation.

2.4.1 Discretization of the dynamics

There are several ways to discretize the dynamics (10) or (12). The most common schemes used in molecular dynamics are the Euler-Maruyama discretization for (12), and the BBK scheme [3] for (10). We refer to [4] for alternative approaches in the field of molecular dynamics. In the sequel, we will denote by $x^{i,k}$ a numerical approximation of a realization of $X^k_{i\Delta t}$.

Euler discretization of the overdamped Langevin dynamics. The Euler-Maruyama numerical scheme associated to (12) reads, when taking integration time steps Δt,

$$q^{n+1} = q^n - \Delta t\, \nabla V_\lambda(q^n) + \sqrt{\frac{2\Delta t}{\beta}}\, R^n, \tag{20}$$

where $(R^n)_{n\in\mathbb{N}}$ is a sequence of independent and identically distributed (i.i.d.) $3N$-dimensional standard Gaussian random vectors. The numerical convergence of this scheme can be ensured in some circonstances [4].

Discretization of the Langevin dynamics. When considering an integration time step Δt, the BBK discretization of (10) reads componentwise (recall that the underscripts j refer here to the components of a given $x^k \equiv x = (q,p)$),

$$\left\{ \begin{array}{l} p_j^{n+1/2} = p_j^n + \dfrac{\Delta t}{2}\left(-\nabla_{q_j} V(q^n) - \xi \dfrac{p_j^n}{m_j} + \dfrac{\sigma_j}{\sqrt{\Delta t}} R_j^n\right) \\ q_j^{n+1} = q_j^n + \Delta t\, \dfrac{p_j^{n+1/2}}{m_j} \\ p_j^{n+1} = \dfrac{1}{1+\frac{\xi \Delta t}{2m_j}}\left(p_j^{n+1/2} - \dfrac{\Delta t}{2}\nabla_{q_j} V(q^{n+1}) + \sigma_j \dfrac{\sqrt{\Delta t}}{2} R_j^{n+1}\right) \end{array} \right. \tag{21}$$

where the random forcing terms R_j^n ($j \in \{1,\dots,N\}$ is the label of the particles, n is the iteration index) are standard i.i.d. Gaussian random variables. The fluctuation/dissipation relation (11) must be corrected so that the kinetic temperature is correct in the simulations [4]. To this end, we set

$$\sigma_j^2 = \frac{2\xi}{\beta}\left(1 + \frac{\xi \Delta t}{2m_j}\right). \tag{22}$$

Notice that the relation (11) is recovered in the limit $\Delta t \to 0$.

2.4.2 Numerical algorithm

We consider for example the following discretization of the force exerted on the k-th replica on the time interval $[i\Delta t, (i+1)\Delta t]$:

$$\frac{\partial H^{k,\Delta t}_{\lambda_{i+1/2}}}{\partial \lambda} = \frac{1}{2}\left(\frac{\partial H_{\lambda(i\Delta t)}}{\partial \lambda}(x^{i,k}) + \frac{\partial H_{\lambda((i+1)\Delta t)}}{\partial \lambda}(x^{i+1,k})\right).$$

The mean force is then approximated by

$$\mathcal{F}^{M,\Delta t}_{\lambda_{i+1/2}} = \frac{1}{M}\sum_{k=1}^{M} \frac{\partial H^{k,\Delta t}_{\lambda_{i+1/2}}}{\partial \lambda}.$$

The following algorithm shows how the process (2.1) can be implemented in practice.

Algorithm 1. *Consider an initial distribution $(x_0^1,\dots,x_0^M)$ generated from $d\mu_0(x)$. Generate idependent times $\tau^{k,b}, \tau^{k,d}$ from an exponential law of mean 1. Consider two additional variables $\Sigma^{k,b}, \Sigma^{k,d}$ per replica, initialized at 0.*

For $0 \le i \le \frac{T}{\Delta t}$,

- *Evolve independently the replicas $x^{i,k}$ according to the schemes (20) or (21), using $\lambda = \lambda((i+1/2)\Delta t)$, and obtain $x^{i+1,k}$;*

- *Update the variables* $\Sigma^{k,b}$ *and* $\Sigma^{k,d}$ *as*

$$\Sigma^{k,b} = \Sigma^{k,b} + \beta \left(\mathcal{F}^{M,\Delta t}_{\lambda_{i+1/2}} - \frac{\partial H^{k,\Delta t}_{\lambda_{i+1/2}}}{\partial \lambda} \right)^{-} \lambda'((i+1/2)\Delta t)\, \Delta t,$$

and

$$\Sigma^{k,d} = \Sigma^{k,d} + \beta \left(\mathcal{F}^{M,\Delta t}_{\lambda_{i+1/2}} - \frac{\partial H^{k,\Delta t}_{\lambda_{i+1/2}}}{\partial \lambda} \right)^{+} \lambda'((i+1/2)\Delta t)\, \Delta t.$$

- *If* $\Sigma^{k,d} \geq \tau_n^{k,d}$*, select an index* $l \in \{1, \ldots, M\}$ *at random, and replace the configuration of the* k*-th replica by the configuration of the* l*-th replica. Generate a new time* $\tau^{k,d}$ *from an exponential law of mean* 1*, and set* $\Sigma^{k,d} = 0$*;*

- *If* $\Sigma^{k,b} \geq \tau_n^{k,d}$*, select an index* $l \in \{1, \ldots, M\}$ *at random, and replace the configuration of the* l*-th replica by the configuration of the* k*-th replica. Generate a new time* $\tau^{k,b}$ *from an exponential law of mean* 1*, and set* $\Sigma^{k,b} = 0$*;*

- *Set* $i = i + 1$*.*

3 Applications of the Interacting Particle System method

We present in this section two application of the IPS method in the field of molecular dynamics. The first one is the most obvious. Since the IPS method improves usual simulated annealing through an additional selection process, even quite short transitions between two states are enough to obtain a good approximation of the final density. This is illustrated in section 3.1, where we present numerical results for the cooling process of a pentane molecule. However, as already explained in the introduction, some quantities can not be computed with the only knowledge of the canonical probability measure. Such an example is the free energy of a system. An interesting approach to compute free-energy differences is the Jarzynski non equilibrium estimation [16], which relies on the formula (16). However, it may happen that this procedure is not efficient as such, especially when the target distribution is really far away from the initial distribution, with many local minima appearing during the transition. IPS can be useful in those situations. A toy simulation supports this idea in section 3.2.

3.1 Computation of canonical distributions

We use the so-called united-atom model [23] to simulate linear alkanes. The conformation of the molecule is completely characterized by the positions of the Carbon atoms in this model. The presence of the Hydrogen atom is implicitely taken into account in the definition of the interaction potential the Carbon atoms are subjected to. The Carbon atoms of the linear alkane molecule are indexed from 1 to N. For pentane, $N = 5$. In this case, there are two dihedral angles (ϕ_1, ϕ_2), which determine the conformations of the molecule.

In the model presented here, the interatomic potential involves two-, three-, and four-body interactions: two Carbon atoms connected by a covalent bond interact *via* a harmonic potential; two Carbon atoms that are separated by three more covalent bonds or more interact *via* a Lennard-Jones potential; the three-body interaction is taken into account through a harmonic potential for the bending angle; finally, the four-body interaction is related to a potential for the dihedral angles. We refer to [4] for further precisions on simulation parameters. The normalization is such that $\beta = 1$ corresponds to a temperature of 300 K.

We consider here a cooling process from $\beta = 1$ to $\beta = 2$, in the case when the Lennard-Jones interactions involve only extremal atoms in the chain, so that $\epsilon_{\text{CH}_3\text{-CH}_3} = 0.29$ and $\epsilon_{\text{CH}_3\text{-CH}_2} = 0$ in the reduced units of [4]. Figure 1 presents some reference empirical distributions for the dihedral angles generated with an importance sampling technique.

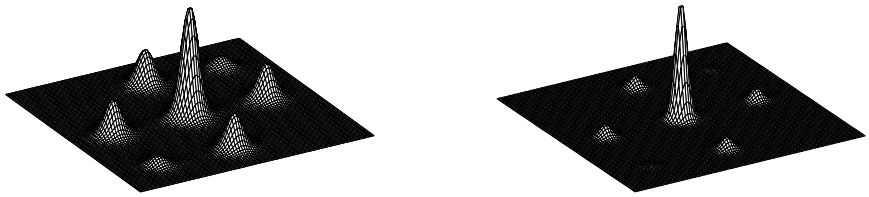

Figure 1: Empirical probability distribution of the dihedral angles (ϕ_1, ϕ_2) of the pentane molecule, for $\epsilon_{\text{CH}_3\text{-CH}_3} = 0.29$ and $\epsilon_{\text{CH}_3\text{-CH}_2} = 0$, generated with Importance sampling, for $\beta = 1$ (Left) and $\beta = 2$ (Right), with sample size $M = 10^9$.

The simulations are done as follows. We first generate an initial distribution of configurations from the canonical measure at inverse temperature $\beta = 1$. This can be done by several techniques, as explained in [4]. Since the aim here is to compare the out of equilibrium cooling processes, we

choose a classical rejection method so that no initial bias is introduced. Let us however emphasize that, contrarily to the case $\beta = 2$, almost all reasonable sampling schemes are convenient. We then first perform a bare simulated annealing from $\beta = 1$ to $\beta = 2$, using the Langevin dynamics (10) with $\xi/m = 1$. We then compare the resulting empirical distribution for the dihedral angles with the one arising from an IPS simulation. Figure 2 presents the results for $M = 10,000$, $\Delta t = 0.01$ and $T = 1$, with a linear scheme $\lambda(t) = t/T$.

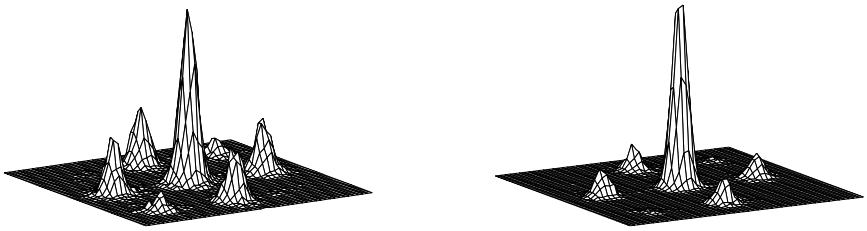

Figure 2: Empirical probability distribution of the dihedral angles (ϕ_1, ϕ_2) at $\beta = 2$ of the pentane molecule generated from a sample at $\beta = 1$, using simulated annealing (Left), and IPS (Right), with sample size $M = 10,000$. The reference distribution is drawn in Figure 1 (Right).

As can be seen in Figure 2, the distribution generated with IPS is much closer to the reference distribution than the distribution generated with simulated annealing. This simple application shows the interest of IPS for computing distributions at low temperature starting from distributions at a higher temperature, even if the driving scheme is quite fast. This is indeed almost always the case in practice when there are several important metastable states.

3.2 Computation of free energy

Estimation of free energy differences. The free energy of a system cannot be computed with a single sample of μ_λ. Only free energy differences can be computed easily. Since the free energy of certain states is known (This is the case for perfect gases, or for solids at low temperature [22]), the free energy of any state can in principle be obtained by an integration between a state whose free energy is known, and the state of interest. Usual methods to this end are Umbrella sampling [27, 26], Thermodynamic integration [19], or Jarzynski's non equilibrium dynamics [16].

In the work of Jarzynski [16], M independent realizations $(X_t^1, ..., X_t^M)$

of a bare out of equilibrium dynamic (7) are used to compute free energy differences through (16), with the estimator

$$\Delta \hat{F}_J = -\frac{1}{\beta} \ln \left(\frac{1}{M} \sum_{k=1}^{M} e^{-\beta W_1^k} \right).$$

Since this point is unclear in [16], we want to lay the emphasis on the possible use of an alternative yet similar estimator relying on a thermodynamical integration:

$$\Delta \hat{F}'_J = \int_0^T \mathcal{F}_{\lambda(t)}^{M_{\mathrm{ind}}} \lambda'(t)\, dt,$$

where

$$\mathcal{F}_{\lambda(t)}^{M_{\mathrm{ind}}} = \mu_{\lambda(t)}^{M_{\mathrm{ind}}} \left(\frac{\partial H_{\lambda(t)}}{\partial \lambda} \right) \quad \text{with} \quad \mu_{\lambda(t)}^{M_{\mathrm{ind}}}(dx) = \frac{\sum_{k=1}^{M} \delta_{X_t^k}(dx)\, \mathrm{e}^{-\beta W_t^i}}{\sum_{k=1}^{M} \mathrm{e}^{-\beta W_t^i}}.$$

However, both estimators $\Delta \hat{F}_J$ and $\Delta \hat{F}'_J$ suffer from the fact that only a few values of W_t^i are really important. Indeed, because of the exponential weighting, only the lower tail of the work distribution is taken into account. The quality of the estimation then relies on those rare values, which may be a problem in practice (see e.g. [21]).

In the case of interacting replicas, we use similarly

$$\Delta \hat{F}_{\mathrm{IPS}} = \int_0^T \mathcal{F}_{\lambda(t)}^{M} \lambda'(t)\, dt,$$

which shares by Proposition 2.1 the same statistical properties as $\Delta \hat{F}_J$: The estimator $\mathrm{e}^{-\beta \Delta \hat{F}_{\mathrm{IPS}}}$ is an unbiased estimator of $\mathrm{e}^{-\beta \Delta F}$, and $\Delta \hat{F}_{\mathrm{IPS}}$ is asymptotically normal with bias and variance of order M^{-1}. Let us however emphasize that the sample is not degenerate for IPS since all points have the same weights.

Toy example. Consider the following family of Hamiltonians $(H_\lambda)_{\lambda \in [0,1]}$:

$$H_\lambda(x) = \frac{x^2}{2} + \lambda Q_1(x) + \frac{\lambda^2}{2} Q_2(x) + \frac{\lambda^3}{6} Q_3(x) + \frac{\lambda^4}{24} Q_4(x) \tag{23}$$

with

$$Q_1(x) = \frac{-1}{8x^2+1}, \quad Q_2(x) = \frac{-4}{8(x-1)^2+1},$$
$$Q_3(x) = \frac{-18}{32(x-3/2)^2+1}, \quad Q_4(x) = \frac{-84}{64(x-7/4)^2+1}.$$

Figure 3 presents some of those Hamiltonians. This toy one-dimensional model is reminiscent of the typical difficulties encountered when μ_0 is very

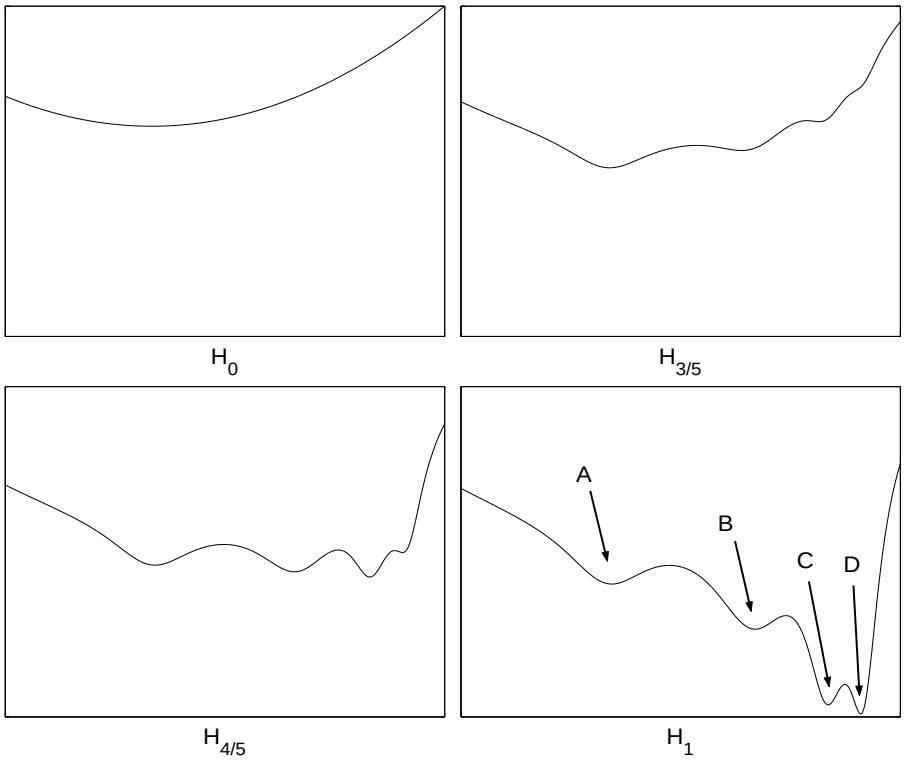

Figure 3: Plot of some Hamiltonian functions, as defined by (23).

different from μ_1. Notice indeed that several transitional metastable states (denoted A and B in Figure 3) occur in the canonical distribution when going from $\lambda = 0$ to $\lambda = 1$. The probability of presence in the basins of attraction of the main stable states of H_1 (C and D in Figure 3) is only effective when λ is close to 1.

Simulations were performed at $\beta = 13$ with the overdamped Langevin dynamic (12), and the above Hamiltonian family (23). The number of replicas was $M = 1000$, the time step $\Delta t = 0.003$, and λ is considered to be linear: $\lambda(t) = t/T$. Figure 4 presents the distribution of replicas during a slow out of equilibrium plain dynamic: $T = 30$. Figure 5 presents the distribution of replicas during a faster dynamic with interaction: $T = 15$.

When performing a plain out of equilibrium dynamic (even 'slow') from $\lambda = 0$ to $\lambda = 1$, almost all replicas are trapped by the energy barrier of these transitional metastable states (see Figure 4). In the end, a very small (almost null) proportion of replicas have performed interesting paths associated with low values of virtual work W. When using (14) to compute thermodynamical quantities, these replicas bear almost all the weight of the degenerate sample, in view of the exponential weighting. The quality of the result therefore depends crucially on these rare values.

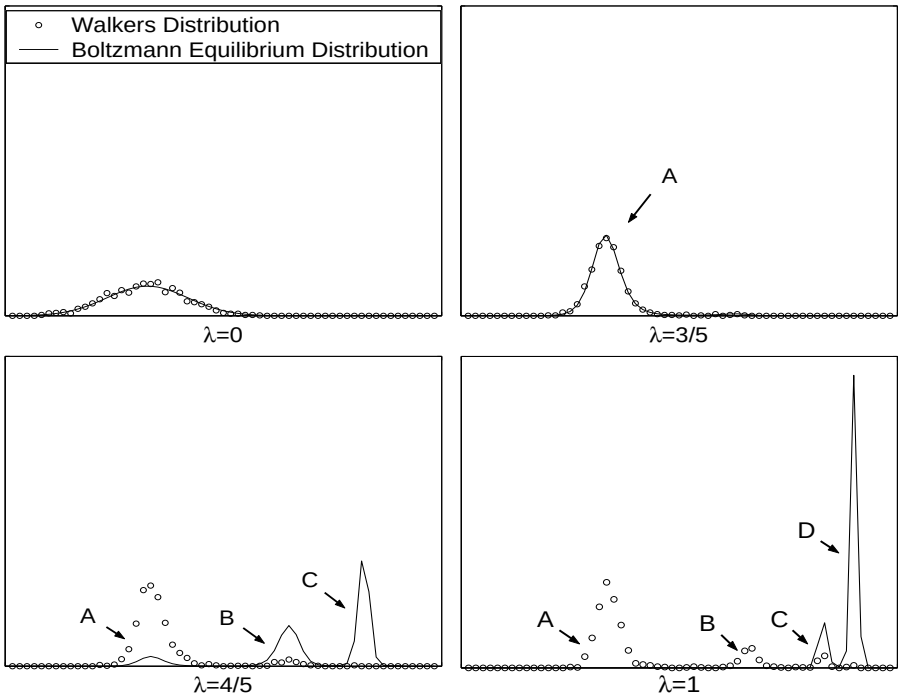

Figure 4: Empirical densities obtained using independant replicas.

Method	Bias	Variance
Plain	+0.25	0.19
Interacting	+0.15	0.10

Table 1: Error in free energy estimation.

On the contrary, in the interacting version, the replicas can perform jumps in the configuration space thanks to the selection mechanism, and go from one metastable basin to another. In our example, as new transition states appear, only few clever replicas are necessary to attract the others in good areas (see Figure 5). In the end, all replicas have the same weight, and the sample is not degenerate.

We have also made a numerical estimation of error in free energy estimation, with 40 realizations of the above simulation. The results are presented in Table 1, and show an important reduction of standard deviation and bias up to a factor 2 when using the IPS method.

References

[1] R. Assaraf, M. Caffarel, A. Khelif, Diffusion Monte Carlo with a fixed number of walkers, *Phys. Rev. E* **61**(4) (2000) 4566–4575.

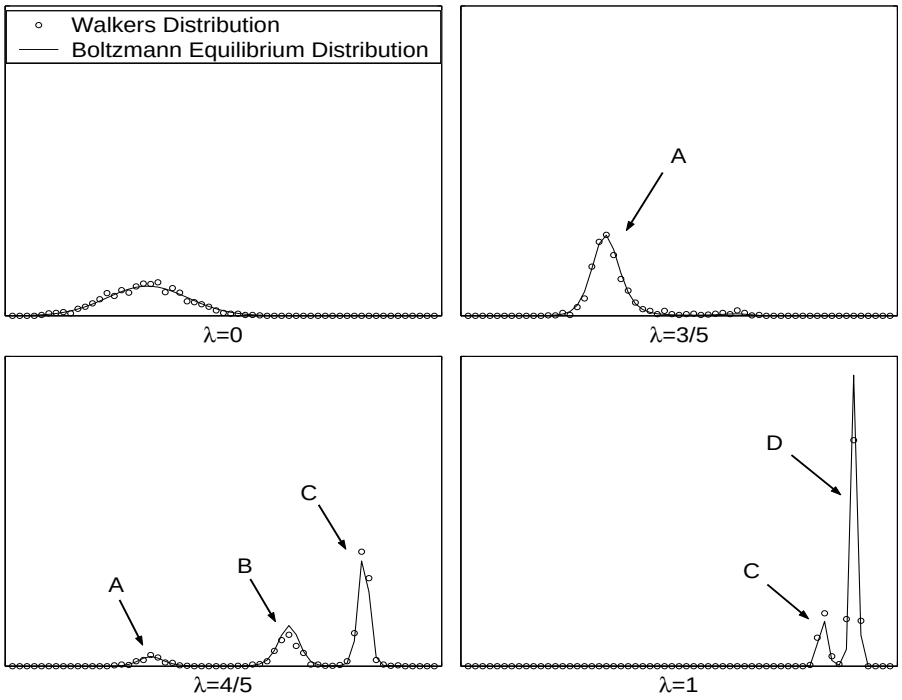

Figure 5: Empirical densities obtained using interacting replicas.

[2] B.A. Berg and T. Neuhaus, Multicanonical algorithms for first order phase transitions, *Phys. Lett. B* **267** (1991) 249–253.

[3] A. Brünger, C.B. Brooks, and M. Karplus, Stochastic boundary conditions for molecular dynamics simulations of ST2 water, *Chem. Phys. Lett.* **105** (1983) 495–500.

[4] E. Cancès, F. Legoll, and G. Stoltz, Theoretical and numerical comparison of sampling methods for molecular dynamics, *IMA Preprint* **2040** (2005).

[5] E.A. Carter, G. Ciccotti, J.T. Hynes, and R. Kapral, Constrained reaction coordinate dynamics for the simulation of rare events, *Chem. Phys. Lett.* **156** (1989) 472–477.

[6] G. Ciccotti, R. Kapral, and E. Vanden-Eijnden, Blue Moon sampling, vectorial reaction coordinates, and unbiased constrained dynamics, *Preprint version*, available at the URL `http://www.cims.nyu.edu/~eve2/bluemoonvect.pdf` (2005).

[7] G.E. Crooks, Nonequilibrium measurements of free energy differences for microscopically reversible markovian systems, *J. Stat. Phys.* **90** (1998) 1481–1487.

[8] P. Del Moral, A. Doucet, and G.W. Peters, Sequential Monte Carlo samplers, *Preprint version*, available on request at the URL `http://www.cs.ubc.ca/~arnaud/` (2004).

[9] P. Del Moral and L. Miclo, Branching and Interacting Particle Systems approximations of Feynman-Kac formulae with applications to nonlinear filtering, Séminaire de Probabilités XXXI, *Lecture notes in Mathematics* **1729** (2000) 1-145.

[10] A. Doucet, N. de Freitas, and N.J. Gordon, *Sequential Monte Carlo Methods in Practice* , Series Statistics for Engineering and Information Science (Springer, 2001).

[11] A. Doucet, M. Rousset, Time continuous particle approximation of probability flows, *in preparation*.

[12] M. I. Freidlin and A.D. Wentzell, *Random perturbations of dynamical systems*, second edition (Springer-Verlag, 1998).

[13] D. Frenkel and B. Smit, *Understanding Molecular Simulation* (Academic Press, 2002).

[14] A. Gelman and X.L. Meng, Simulating normalizing constants: From importance sampling to bridge sampling to path sampling, *Stat. Sci.* **13**(2) (1998) 163–185.

[15] Y. Iba, Extended ensemble Monte Carlo, *Int. J. Mod. Phys. C* **12** (2001) 623–656.

[16] C. Jarzynski, Equilibrium free energy differences from nonequilibrium measurements: a master equation approach, *Phys. Rev. E* **56**(5) (1997) 5018–5035.

[17] I. Karatzas and S. Shreve, *Brownian Motion and Stochastic Calculus* (Springer-Verlag, 1991).

[18] S. Kirkpatrick, C.G. Gelatt, amd M.P. Vecchi, Optimization by simulated annealing, *Science* **220** (1983) 671–680.

[19] J.G. Kirkwood, Statistical mechanics of fluid mixtures, *J. Chem. Phys.* **3** (1935) 300–313.

[20] E. Marinari and G. Parisi, Simulated tempering: a new Monte Carlo scheme, *Europhys. Lett.* **19** (1992) 451–458.

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

[22] J.M. Rickman and R. LeSar, Free-energy calculations in materials research, *Annu. Rev. Matter. Res.* **32** (2002) 195–217.

[23] J.P. Ryckaert and A. Bellemans, Molecular dynamics of liquid alkanes, *Faraday Discuss.* **66** (1978) 95-106.

[24] M. Rousset, On the control of an Interacting Particle estimation of Schrödinger groundstates, *Preprint*, available at the URL `www.lsp.ups-tlse.fr/Fp/Rousset/article3.pdf`

[25] M. Rousset, *Phd Thesis*, in preparation.

[26] G.M. Torrie and J.P. Valleau, Nonphysical sampling distributions in Monte-Carlo free energy estimation: Umbrella sampling, *J. Comp. Phys.* **23** (1977) 187–199.

[27] R. Zwanzig, High-temperature equation of state by a perturbation method: I. Nonpolar gases, *J. Chem. Phys.* **22** (1954) 1420–1426.

Appendix : Pure jump processes

Consider a Markov process X_t of infinitesimal generator

$$J(f)(x) = \int (f(y) - f(x))\alpha(x)d\mu(y),$$

where α is a bounded positive function and μ a probability measure. Denote by $(T_n)_{n\geq 1}$ the jump times (with $T_0 = 0$), and $(\tau_n)_{n\geq 1}$ independant clocks of exponential law of mean 1. Then the system evolves according to the following stochastic rules:

- the jump times are defined by

$$\alpha(X_{T_n})(T_{n+1} - T_n) = \tau_n;$$

- at jump times, the process jumps to a configuration chosen according to the probability measure $d\mu(y)$.