Eur. Phys. J. B **58**, 83–95 (2007)
DOI: 10.1140/epjb/e2007-00201-9

THE EUROPEAN PHYSICAL JOURNAL B

Web ensemble averages for retrieving relevant information from rejected Monte Carlo moves

M. Athènes[a]

Service de Recherches de Métallurgie Physique, CEA/Saclay, 91191 Gif-sur-Yvette Cedex, France

Received 16 February 2007 / Received in final form 22 May 2007
Published online 28 July 2007 –

Abstract. We study the relevance of including information about rejected Monte-Carlo moves in path-sampling computations of free energies. For this purpose, we define webs as sets of paths linked by the path-sampling scheme and introduce an associated statistical ensemble. Within this web ensemble, we derive and test several statistical averages enabling to include information about configurational and path quantities belonging to the unselected trial moves. We numerically observe that retrieving this information does not always result in variance reduction, as theoretically predicted by Delmas and Jourdain. To explain the possible detrimental effect of information-retrieving from web sampling, an action for the webs is introduced. The behaviour of the statistical variance is observed to correlate to an overlapping area of a web action histogram. This area represents the probability that a generated web is such that the difference of its action between the targeted and reference ensembles is lower than the corresponding difference of free energy. Variance reductions are numerically observed for increased areas, as it is the case for the residence weight method proposed previously. More generally, web ensembles yield a rigorous framework for rationalizing existing methods and also for deriving potentially new methods enabling to retrieve relevant information from rejected trial moves.

PACS. 05.10.Ln Monte Carlo methods – 07.05.Tp Computer modeling and simulation – 05.20.Jj Statistical mechanics of classical fluids

1 Introduction

Markov-Chain Monte Carlo (MCMC) methods are used extensively to compute the thermodynamic properties of many-body systems. Any MCMC method is based on an algorithm that generates a Markov chain of configurations in which the probability to visit any particular configuration is made proportional to an adequate statistical weight, here the Boltzmann weight, in order to insure the desired form of importance sampling. To achieve this, the sequential Metropolis algorithm can be implemented: a trial configuration is first generated and then added to the Markov chain using a judicious acceptance rule. As a result a proportion of the trial configurations are rejected and information about these rejected configurations is wasted.

At variance, parallel MCMC algorithms, such as residence time algorithms [1,2], generalized residence time algorithms [3,4], residence weight algorithms [5,6] and the waste-recycling algorithm [7], which were developed to improve the efficiency of the Metropolis algorithm, retrieve information related to the wasted trial moves. These algorithms generate several trial moves in parallel at each Monte Carlo step prior to selecting a favorable one using the Rosenbluth rule. Except for the waste-recycling algorithm, the configuration chain that is constructed using these algorithms differs from the Boltzmann distribution. Hence, a re-weighing factor accommodates the deviation from the targeted statistical weight and enables to recover the desired ensemble average. This re-weighing factor corresponds to a residence time or residence weight and incorporates information related to the unselected moves.

In the waste-recycling algorithm [7], the Markov chain consisting of the selected configurations is made proportional to the Boltzmann weight. Here, information is retrieved by including the unselected configurations in the ensemble average and weighing their contributions using the Rosenbluth probabilities of the selecting procedure. This way of proceeding indeed generalizes a modification of the sequential Metropolis algorithm that was proposed by Ceperley et al. [8] thirty years ago. Information about rejected trial moves is also retrieved within an ensemble average that includes the acceptance rule and whose analytical form was conjectured. It was suggested that the

[a] e-mail: Manuel.Athenes@cea.fr

use of normally rejected information should result in a statistical variance reduction. From a probabilistic point of view, some properties of this information-retrieving scheme have been elucidated recently [9]: the ensemble average is shown to converge normally asymptotically and its asymptotic variance is proved to be not smaller in general than the one of the standard Metropolis procedure, contrarily to what was expected.

Boulougouris and Frenkel [10,11] further generalized the sequential information-retrieving average. They showed that not only one but an infinite number of valid averages could be used in practice in order to retrieve information about rejected trial configurations. The proposed derivation consists in inserting, within the Monte Carlo average, a balance equation linking an accepted configuration to its subsequent trial configuration [10,11]. From a probabilistic point of view, this approach is equivalent to perform the ensemble average over a Markov chain of *linked configurations* rather than of *configurations*.

The aim of this article is to show that the concepts related to averaging *linked configuration* ensembles also apply to the residence algorithms aforementioned. It will result in a rigorous framework enabling to rationalize the residence algorithms and to derive additional information-retrieving averages. We will also examine their variances numerically, and explain in which cases unselected information is relevant.

The article is organized as follows. In Section 2, we present the problem and introduce the statistical ensembles in which residence algorithms operate. Information-retrieving averages are formulated in Sections 3 and 4, and implemented in Section 5 for computing free energy differences between two values of a control parameter within a toy model system. In Section 6, we reformulate the original approach that allowed to derive the residence weight algorithm. In particular, we show that the weighted detailed balance equation that was introduced to justify these algorithms is not valid in general. We give some final remarks in Section 7.

2 Configuration and path ensembles

As a common feature, residence algorithms are based on Boltzmannian path-sampling schemes [12]: this means that, provided a suitable path ensemble is defined, the chain constituted by the selected trial paths is Markovian and Boltzmann distributed. This feature suggests that including information about the unselected trial moves will be made by estimating the averages in an ensemble of *linked paths*, instead of an ensemble of *linked configurations* as for the Metropolis algorithms [10]. To simplify the denominations, linked paths will be called webs and the associated statistical ensemble will be called web ensemble, inspiring from the terminology of reference [10]. Configuration and path ensembles are defined now, while the web ensemble averages are introduced in Section 3.

Let $\mathbf{r}$ designate the particle positions of the many-body system. We assume that the configurational energy, noted $\mathcal{E}_\lambda(\mathbf{r})$, is parametrized by an external control parameter λ.

The ensemble, noted $\mathcal{Z}_\lambda$, with constant particle number N, inverse temperature $\beta = 1/(kT)$ and control parameter λ, has partition function

$$Z_\lambda = \frac{1}{\Lambda^{dN} N!} \int \mathrm{d}\mathbf{r} \exp[-\beta \mathcal{E}_\lambda(\mathbf{r})], \tag{1}$$

where $\Lambda = \sqrt{h^2/[2\pi mkT]}$ is de Broglie's thermal wavelength, (h is Planck's constant and m is the particle mass) and d is dimension. We also have $\mathrm{d}\mathbf{r} = \prod_{k=1}^{d} \mathrm{d}r_k$, where r_k are the components of $\mathbf{r}$. The free energy is $F_\lambda = -kT \ln Z_\lambda$. The unnormalized Boltzmann weight

$$\mathcal{N}_\lambda(\mathbf{r}) = \exp[-\beta \mathcal{E}_\lambda(\mathbf{r})]/[\Lambda^{3N} N!] \tag{2}$$

allows to define the canonical ensemble average as

$$\langle A_\lambda \rangle = \frac{\int A(\mathbf{r}) \mathcal{N}_\lambda(\mathbf{r}) \mathrm{d}\mathbf{r}}{\int \mathcal{N}_\lambda(\mathbf{r}) \mathrm{d}\mathbf{r}}, \tag{3}$$

where A is any configurational quantity. Let us now define paths as pairs of configurations, $\mathbf{z} \equiv (\mathbf{r}_0, \mathbf{r}_1)$ where $\mathbf{r}_0$ and $\mathbf{r}_1$ are any two configurations. We also consider a stochastic process allowing to construct a new configuration starting from an old one. The generating probabilities associated to this process will be considered as conditional probabilities in a path ensemble. Hence, we note $P_{\mathrm{cond}}(\mathbf{z}|\mathbf{r}_0)$ and $P_{\mathrm{cond}}(\mathbf{z}|\mathbf{r}_1)$ the two possible probabilities to construct path $\mathbf{z}$ given that either $\mathbf{r}_0$ or $\mathbf{r}_1$ is known.

To define the path ensemble $\widetilde{\mathcal{Z}}_0$, we associate to the paths $\mathbf{z}$ the equivalent of a Boltzmann weight

$$\mathcal{K}_0(\mathbf{z}) = \mathcal{N}_0(\mathbf{r}_0) P_{\mathrm{cond}}(\mathbf{z}|\mathbf{r}_0). \tag{4}$$

This unnormalized Boltzmann weight is proportional to the probability to observe the corresponding path in a simulation, provided that paths are initiated from configurations $\mathbf{r}_0$ distributed according to ensemble $\mathcal{Z}_0$. Paths of $\widetilde{\mathcal{Z}}_0$ are said to be forward.

Because the conditional probabilities are normalized to one, ensemble averages can also be taken in the path ensemble

$$\langle A_0 \rangle = \frac{\int A(\mathbf{r}_0) \mathcal{N}_0(\mathbf{r}_0) d\mathbf{r}_0 \int_{\Omega(\mathbf{r}_0)} P_{\mathrm{cond}}(\mathbf{z}|\mathbf{r}_0) \mathcal{D}\mathbf{z}}{\int \mathcal{N}_0(\mathbf{r}_0) d\mathbf{r}_0 \int_{\Omega(\mathbf{r}_0)} P_{\mathrm{cond}}(\mathbf{z}|\mathbf{r}_0) \mathcal{D}\mathbf{z}} \tag{5}$$

$$= \frac{\int A(\mathbf{r}_0) \mathcal{K}_0(\mathbf{z}) \mathcal{D}\mathbf{z}}{\int \mathcal{K}_0(\mathbf{z}) \mathcal{D}\mathbf{z}} \tag{6}$$

where $\Omega(\mathbf{r}_0)$ is the subset of all paths starting from $\mathbf{r}_0$. Here, $\int ... \mathcal{D}\mathbf{z}$ denotes an integration over paths. Similarly, if the final configurations $\mathbf{r}_1$ of the paths are distributed according to $\mathcal{Z}_1$ ensemble and if paths are constructed backward ($\mathbf{r}_0$ is generated from $\mathbf{r}_1$), then the Boltzmann weight for the paths becomes

$$\mathcal{K}_1(\mathbf{z}) = \mathcal{N}_1(\mathbf{r}_1) P_{\mathrm{cond}}(\mathbf{z}|\mathbf{r}_1), \tag{7}$$

which defines a path ensemble, noted $\widetilde{\mathcal{Z}}_1$. Paths of $\widetilde{\mathcal{Z}}_1$ are said to be backward. One can also express the canonical

average in the path ensemble $\widetilde{\mathcal{Z}}_1$ instead of $\mathcal{Z}_1$,

$$\langle A_1 \rangle = \frac{\int A(\mathbf{r}_1)\mathcal{K}_1(\mathbf{z})\mathcal{D}\mathbf{z}}{\int \mathcal{K}_1(\mathbf{z})\mathcal{D}\mathbf{z}}. \tag{8}$$

The two Boltzmann weights related to a given path may be expressed as the exponential of an action

$$\mathcal{K}_0(\mathbf{z}) = \exp\left[-\frac{\beta}{2}U(\mathbf{z}) + \frac{\beta}{2}W(\mathbf{z})\right] \tag{9}$$

$$\mathcal{K}_1(\mathbf{z}) = \exp\left[-\frac{\beta}{2}U(\mathbf{z}) - \frac{\beta}{2}W(\mathbf{z})\right], \tag{10}$$

The action of the path has been decomposed into a symmetric contribution, $\frac{\beta}{2}U(\mathbf{z})$, and an antisymmetric contribution, $\frac{\beta}{2}W(\mathbf{z})$. One interprets $W(\mathbf{z})$ as an effective work because it always satisfies a second principle, even when the stochastic process is not micro-reversible [12]. The decomposition allows to define the following parametrized path weight

$$\mathcal{K}_\theta(\mathbf{z}) = \exp\left[-\beta U(\mathbf{z}) + \beta\left(\frac{1}{2} - \theta\right)W(\mathbf{z})\right], \tag{11}$$

together with an associated path partition function [12]

$$\widetilde{Z}_\theta = \int \mathcal{K}_\theta(\mathbf{z})\mathcal{D}\mathbf{z}, \tag{12}$$

where $\theta \in [0;1]$. This defines a associated path ensemble, noted $\widetilde{\mathcal{Z}}_\theta$.

Formally, any path quantity B can be averaged in the path ensemble $\widetilde{\mathcal{Z}}_\theta$

$$\langle B \rangle_\theta = \frac{\int B(\mathbf{z})\mathcal{K}_\theta(\mathbf{z})\mathcal{D}\mathbf{z}}{\int \mathcal{K}_\theta(\mathbf{z})\mathcal{D}\mathbf{z}}. \tag{13}$$

In practice, path averages are useful to compute free energy differences [13] or Landau free energies [14]. In this study, we will later compute the free energy difference $\Delta F = F_1 - F_0$ between system (1) and system (0) using path average (13) and $\theta = 1/2$. To do so, one first notices that the conditional path probabilities are normalized, hence $\widetilde{Z}_0 = Z_0$, $\widetilde{Z}_1 = Z_1$, which gives

$$\exp\left[-\beta\Delta F\right] = \frac{\widetilde{Z}_1}{\widetilde{Z}_0}. \tag{14}$$

One then takes advantage of the bridging property of $\widetilde{\mathcal{Z}}_{\frac{1}{2}}$ with respect to $\widetilde{\mathcal{Z}}_0$ and $\widetilde{\mathcal{Z}}_1$ ensembles, to obtain an efficient way to compute free energy differences [12,15–19]

$$\exp\left[-\beta\Delta F\right] = \frac{\widetilde{Z}_1/\widetilde{Z}_{\frac{1}{2}}}{\widetilde{Z}_0/\widetilde{Z}_{\frac{1}{2}}} \tag{15}$$

$$= \frac{\langle \Upsilon_1(\mathbf{z})\rangle_{\frac{1}{2}}}{\langle \Upsilon_0(\mathbf{z})\rangle_{\frac{1}{2}}} \tag{16}$$

$$= \frac{\langle \exp\left[-\frac{\beta}{2}W(\mathbf{z})\right]\rangle_{\frac{1}{2}}}{\langle \exp\left[+\frac{\beta}{2}W(\mathbf{z})\right]\rangle_{\frac{1}{2}}} \tag{17}$$

where the path averages in the numerator and denominator of equation (16) are obtained by substituting the path function $\Upsilon_\alpha(\mathbf{z}) = \mathcal{K}_\alpha(\mathbf{z})/\mathcal{K}_{\frac{1}{2}}(\mathbf{z})$ using $\alpha = 1$ and $\alpha = 0$ respectively, into equation (13) with $\theta = \frac{1}{2}$.

2.1 Path-sampling scheme

Estimating the path-averages above requires a Monte Carlo scheme that enables to construct a Markov chain of paths with $\widetilde{\mathcal{Z}}_{\frac{1}{2}}$ statistics. Here, we describe the parallel path-sampling scheme that will be implemented.

We note $\mathbf{z}_0$ the current path and $\mathbf{z}_1$,.., $\mathbf{z}_i$,..,$\mathbf{z}_I$ the I trial paths generated in parallel by the path-sampling scheme. We call web any set $\boldsymbol{\pi} = (\mathbf{z}_0, .., \mathbf{z}_i, .., \mathbf{z}_I)$ consisting in $I+1$ paths that can be constructed by the path-sampling scheme starting from the current path $\mathbf{z}_0$. Moreover, the path generating procedure is revertible: $\boldsymbol{\pi}$ can be constructed starting from any path $\mathbf{z}_i$ with an associated conditional probability such that $P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_i) > 0$.

The Metropolis algorithm is used. Among the I trial paths, one path, let say $\mathbf{z}_i$, is pre-selected with probability $P_{\rm s}(\mathbf{z}_i|\boldsymbol{\pi}|\mathbf{z}_0)$. It is then accepted with the Metropolis probability

$$P_{\rm acc}(\mathbf{z}_i) = \min\left(1, \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_i)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_0)P_{\rm s}(\mathbf{z}_i|\boldsymbol{\pi}|\mathbf{z}_0)}{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_0)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_i)P_{\rm s}(\mathbf{z}_0|\boldsymbol{\pi}_0|\mathbf{z}_i)}\right).$$

If path $\mathbf{z}_i$ is rejected, then path $\mathbf{z}_0$ is selected instead. The Metropolis acceptance rule obeys the detailed balance condition

$$\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_0)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_0)P_{\rm s}(\mathbf{z}_i|\boldsymbol{\pi}|\mathbf{z}_0)P_{\rm acc}(\mathbf{z}_i) = \mathcal{K}_{\frac{1}{2}}(\mathbf{z}_i)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_i)P_{\rm s}(\mathbf{z}_0|\boldsymbol{\pi}|\mathbf{z}_i)P_{\rm acc}(\mathbf{z}_0). \tag{18}$$

This condition insures that the Markov chain of paths converges towards the $\widetilde{\mathcal{Z}}_{\frac{1}{2}}$ statistics. Since the selecting probability is chosen to be

$$P_{\rm s}(\mathbf{z}_k|\boldsymbol{\pi}|\mathbf{z}_j) = \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_k)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_k)}{\sum_{0\le l\le I}^{l\neq j}\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_l)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_l)} \tag{19}$$

the acceptance rule simplifies to

$$P_{\rm acc}(\mathbf{z}_i) = \min\left(1, \frac{\sum_{0\le l\le I}^{l\neq 0}\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_l)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_l)}{\sum_{0\le l\le I}^{l\neq i}\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_l)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_l)}\right). \tag{20}$$

This acceptance rule will be used in the present study. It was proposed previously [5] in an equivalent form and has been studied theoretically in reference [9].

Note that a symmetrical rule

$$P'_{\rm s}(\mathbf{z}_k|\boldsymbol{\pi}|\mathbf{z}_j) = \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_k)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_k)}{\sum_{0\le l\le I}\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_l)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_l)} = p_R(\mathbf{z_k}|\boldsymbol{\pi}) \tag{21}$$

can be used as a selecting probability provided that we set

$$P'_{\rm acc}(\mathbf{z}_i) = P'_{\rm acc}(\mathbf{z}_0) = 1.$$

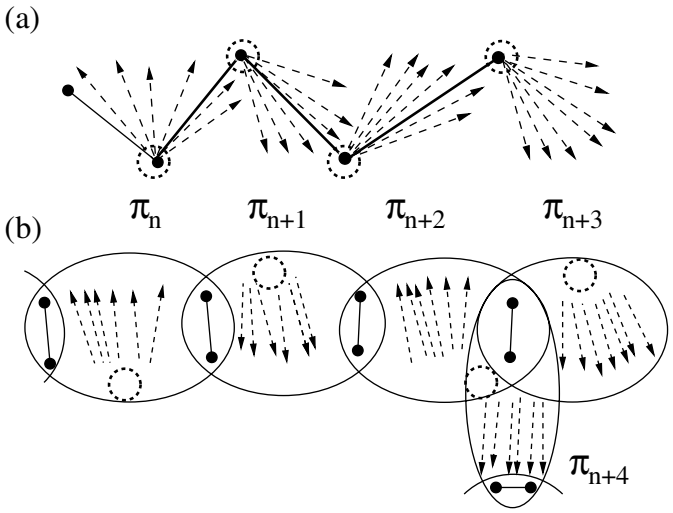

Fig. 1. Alternating procedure [5,6] for generating trial paths in parallel with $I = 7$: (a) selected paths are represented by dumbbells; unselected trial paths are represented by dashed arrows and are generated starting from the terminal configurations (dashed circles) of the selected paths; (b) the corresponding successive webs are represented by ellipses. Inside a web, the dashed circle represents the configuration from which all trial paths are generated. This property guarantees the reversibility of the algorithm. Note also that the selected path in web $\boldsymbol{\pi}_{n+2}$ is identical to the one of web $\boldsymbol{\pi}_{n+3}$.

in equation (18) in order to satisfy the condition of detailed balance. However, for this symmetric rule, the probability to transit to a new path $\mathbf{z}_i$ $(i > 0)$ is always lower than the one given by the Metropolis algorithm since we always have

$$p_R(\mathbf{z_i}|\boldsymbol{\pi}) \leq P_{\rm s}(\mathbf{z}_i|\boldsymbol{\pi}|\mathbf{z}_0)P_{\rm acc}(\mathbf{z_i}).$$

For this reason, the Metropolis rule $P_{\rm acc}(\mathbf{z}_i)$ is traditionally preferred to the symmetric rule $p_R(\mathbf{z_i}|\boldsymbol{\pi})$.

Several procedures for generating the trial paths have been used in the literature. The alternating procedure, proposed and used in references [5–7,12], consists in generating the trial paths starting from the final configuration of the last accepted path, as depicted in Figure 1a. This way of proceeding amounts to implement the waste-recycling algorithm in an ensemble of paths rather than of configurations, as shown in Figure 1b.

The alternating procedure has been generalized by the shooting procedure [16,18,20,21] that consists in generating a trial path both forward and backward in time starting from a randomly selected time slice and from a randomly generated configuration.

In the present study, the parallel shooting procedure schematized in Figure 2 will be used: $2I$ configurational replicas are considered in parallel: half with $\mathcal{Z}_0$ and half with $\mathcal{Z}_1$ statistics. With probability p_0, one draws the I configurational replica that are represented by the black circles using $\mathcal{Z}_0$ statistics; from them one generates the I forward trial paths that are represented by the black dashed arrows using $\widetilde{\mathcal{Z}}_0$ statistics, in which case the constructed web is called bottom. Otherwise, i.e. with probability $p_1 = 1 - p_0$, one draws I replica with $\mathcal{Z}_1$ statistics prior to generating trial paths backward with $\tilde{\mathcal{Z}}_1$ statistics, in which case the constructed web is called top, replica are represented in Figure 2 by grey circles and paths by grey arrows. Hence, knowing path $\mathbf{z}_j \in \boldsymbol{\pi}$, the I remaining paths $\mathbf{z}_l$ $(l \neq j)$ of the web are generated with a conditional probability equal to the Boltzmann weights in $\left[\widetilde{\mathcal{Z}}_\alpha\right]^I$

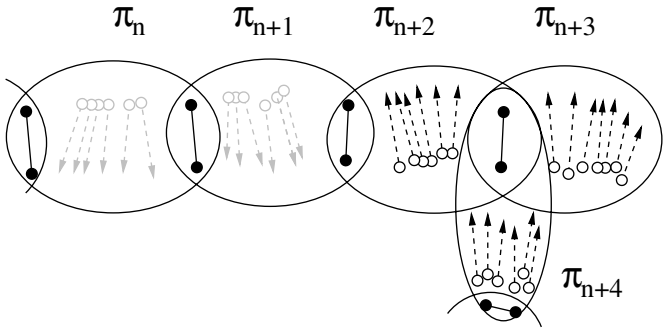

Fig. 2. Schematic representation of the parallel shooting procedure. Trial paths generated forward with $\widetilde{\mathcal{Z}}_0$ statistics are represented by black dashed arrows pointing upward, while those generated backward with $\widetilde{\mathcal{Z}}_1$ statistics are represented by grey dashed arrows pointing downward. $\boldsymbol{\pi}_n$ and $\boldsymbol{\pi}_{n+1}$ are top webs, while $\boldsymbol{\pi}_{n+2}$, $\boldsymbol{\pi}_{n+3}$ and $\boldsymbol{\pi}_{n+4}$ are bottom webs. Note that at variance with the alternating procedure, trial paths are generated from different configurations, represented by the small circles. Reversibility of the algorithm is nevertheless preserved owing to the use of adequate non-symmetric a priori probabilities in the acceptance procedure.

$$P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_j) = \frac{p_\alpha}{[Z_\alpha]^I} \prod_{0\leq l\leq I}^{l\neq j} \mathcal{K}_\alpha(\mathbf{z}_l). \tag{22}$$

with $\alpha = 0$ or $\alpha = 1$ which respectively corresponds to a bottom or top web.

In this study, equation (22) is strictly obeyed because the configurational replicas will be generated with $\mathcal{Z}_\alpha$ statistics directly. If the parallel replicas were obtained from standard Metropolis or Langevin simulations in ensemble $\mathcal{Z}_\alpha$, equation (22) would be only fulfilled up to an unknown normalization constant. Since the same unknown constant would appear in both sides of detailed balance (Eq. (18)), reversibility would still hold and the present procedure would still apply to this more general case.

3 Web ensembles

Provided a probability density is associated to the webs introduced above, a web-ensemble and web-averages can be defined. Web-averages will enable to retrieve information about unselected paths. Since the statistical weights targeted by the path-sampling scheme are $\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_0)$, the webs

constructed from $\mathbf{z}_0$ occur with probability density

$$\Phi(\boldsymbol{\pi}|\mathbf{z}_0) = \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_0)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_0)}{\widetilde{Z}_{\frac{1}{2}}} \tag{23}$$

in the course of the Monte Carlo simulation. The ensemble average of a path quantity B can then be formally transposed into the web-ensemble

$$\langle B\rangle_{\frac{1}{2}} = \frac{\int B(\mathbf{z}_0)\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_0)\mathcal{D}\mathbf{z}_0}{\int \mathcal{K}_{\frac{1}{2}}(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\mathbf{z}_0} \equiv$$
$$\langle B\rangle^\star_{\frac{1}{2}} = \int B(\mathbf{z}_0)\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}. \tag{24}$$

Here, we have introduced $\int ...\mathcal{D}\boldsymbol{\pi} \equiv \int ... \prod_{i=0}^{I}\mathcal{D}\mathbf{z}_i$ to denote a summation over webs. Restricting the integration volume to top and bottom webs respectively yields two additional web-averages

$$\langle B\rangle_{\frac{1}{2}} \equiv \langle B\rangle^1_{\frac{1}{2}} = \frac{1}{p_1}\int_\top B(\mathbf{z}_0)\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi} \tag{25}$$
$$\equiv \langle B\rangle^0_{\frac{1}{2}} = \frac{1}{p_0}\int_\bot B(\mathbf{z}_0)\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi} \tag{26}$$

that will also be considered. The ensemble average of a configurational quantity A defined for $\mathcal{Z}_\theta$ but considered with respect to the present web ensemble $\widetilde{\mathcal{Z}}^\star_{\frac{1}{2}}$ can be transposed from the path average of equation 8 as follows

$$\langle A_\theta\rangle = \frac{\int A(\mathbf{r}_{\theta,0})\Upsilon_\theta(\mathbf{z}_0)\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_0)\mathcal{D}\mathbf{z}_0}{\int \Upsilon_\theta(\mathbf{z}_0)\mathcal{K}_{\frac{1}{2}}(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\mathbf{z}_0}$$
$$= \frac{\int A(\mathbf{r}_{\theta,0})\Upsilon_\theta(\mathbf{z}_0)\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}}{\int \Upsilon_\theta(\mathbf{z}_0)\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}}$$
$$= \frac{\langle A_\theta\Upsilon_\theta\rangle^\star_{\frac{1}{2}}}{\langle\Upsilon_\theta\rangle^\star_{\frac{1}{2}}}, \tag{27}$$

where $\langle A_\theta\Upsilon_\theta\rangle^\star_{\frac{1}{2}}$ is a web average given by equation (24) since it involves the path quantity $\mathbf{z}\to A(\mathbf{r}_\theta)\Upsilon_\theta(\mathbf{z})$.

3.1 Information retrieving

The web-average introduced above will make it possible to retrieve information about rejected trial paths. To account for the unselected paths $\mathbf{z}_{0<i\le I}$ in the web-average, we first consider the adequate web density

$$\Phi(\boldsymbol{\pi}|\mathbf{z}_i) = \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_i)P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_i)}{\widetilde{Z}_{\frac{1}{2}}}, \tag{28}$$

that allows to define $I+1$ equivalent ways of carrying out the web-ensemble average

$$\langle B\rangle^\star_{\frac{1}{2}} = \int B(\mathbf{z}_i)\Phi(\boldsymbol{\pi}|\mathbf{z}_i)\mathcal{D}\boldsymbol{\pi}. \tag{29}$$

Let $p(\mathbf{z}_i|\boldsymbol{\pi})$ ($\mathbf{z}_i \in \boldsymbol{\pi}$ $0 \le i \le I$) denote a real positive function defined on the web ensemble and satisfying the set of detailed balance conditions

$$p(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_j) = p(\mathbf{z}_j|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_i), \tag{30}$$

where $0 \le j \le I$. The web average can be re-expressed as

$$\langle B\rangle^\star_{\frac{1}{2}} = \int B(\mathbf{z}_0)\left[1-\sum_{i=1}^{I}p(\mathbf{z}_i|\boldsymbol{\pi})\right]\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}$$
$$+\sum_{i=1}^{I}\int B(\mathbf{z}_0)p(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}. \tag{31}$$

If we define the permutation

$$\boldsymbol{\pi} = (\mathbf{z}_0, .., \mathbf{z}_i, .., \mathbf{z}_I) \to \bar{\boldsymbol{\pi}}^i = (\bar{\mathbf{z}}^i_0, .., \bar{\mathbf{z}}^i_j, .., \bar{\mathbf{z}}^i_I),$$

where

$$\bar{\mathbf{z}}^i_j = \begin{cases}\mathbf{z}_i & \text{if } j=0\\ \mathbf{z}_0 & \text{if } j=i\\ \mathbf{z}_j & \text{otherwise}\,,\end{cases}$$

then, the integrals inside the summation in equation (31) can be restated as

$$\int B(\mathbf{z}_0)p(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi} = \int B(\mathbf{z}_0)p(\mathbf{z}_0|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_i)\mathcal{D}\boldsymbol{\pi}$$
$$= \int B(\bar{\mathbf{z}}^i_i)p(\bar{\mathbf{z}}^i_i|\bar{\boldsymbol{\pi}}^i)\Phi(\bar{\boldsymbol{\pi}}^i|\bar{\mathbf{z}}^i_0)\mathcal{D}\bar{\boldsymbol{\pi}}^i$$
$$= \int B(\mathbf{z}_i)p(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}.$$

The first transformation consists in inserting the detailed balance condition of equation (30). The second transformation substitutes web $\bar{\boldsymbol{\pi}}^i$ for web $\boldsymbol{\pi}$. This is made possible because the web integration volume is invariant under permutation of paths inside the webs. For any integrated web $\boldsymbol{\pi}$, web $\bar{\boldsymbol{\pi}}^i$ is also to be integrated, and vice versa. This feature leaves the value of the web-integral unchanged. Finally, the third transformation is simply a notation change for a dummy variable. Substituting the obtained integral into equation (31) yields

$$\langle B\rangle^\star_{\frac{1}{2}} = \int\left\{B(\mathbf{z}_0)\left[1-\sum_{i=1}^{I}p(\mathbf{z}_i|\boldsymbol{\pi})\right]\right.$$
$$\left.+\sum_{i=1}^{I}B(\mathbf{z}_i)p(\mathbf{z}_i|\boldsymbol{\pi})\right\}\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}. \tag{32}$$

This average is valid for any web function satisfying the set of detailed balance conditions (Eq. (30)). Hence, if we also impose the additional normalization constraint

$$1-\sum_{i=1}^{I}p(\mathbf{z}_i|\boldsymbol{\pi}) = p(\mathbf{z}_0|\boldsymbol{\pi}), \tag{33}$$

then the web-ensemble average becomes

$$\langle B\rangle^\star_{\frac{1}{2}} = \int\sum_{i=0}^{I}B(\mathbf{z}_i)p_R(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}, \tag{34}$$

where $p_R(\mathbf{z}_i|\boldsymbol{\pi})$ is the unique solution corresponding to the linear system of $I+1$ independent equations formed by equations (30) and (33). We have

$$p_R(\mathbf{z}_i|\boldsymbol{\pi}) = \frac{\Phi(\boldsymbol{\pi}|\mathbf{z}_i)}{\sum_{i=0}^{I}\Phi(\boldsymbol{\pi}|\mathbf{z}_i)}. \tag{35}$$

Finally, transposing the path-ensemble average of a configurational quantity defined in $\mathcal{Z}_\theta$ (Eq. (27)) into a web-ensemble average reads

$$\langle A_\theta\rangle = \frac{\int\sum_{i=0}^{I}A(\mathbf{r}_{\theta,i})\Upsilon_\theta(\mathbf{z}_i)p_R(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}}{\int\sum_{i=0}^{I}\Upsilon_\theta(\mathbf{z}_i)p_R(\mathbf{z}_i|\boldsymbol{\pi})\Phi(\boldsymbol{\pi}|\mathbf{z}_0)\mathcal{D}\boldsymbol{\pi}}. \tag{36}$$

3.2 Monte Carlo estimates

Now let consider that we have implemented the Metropolis algorithm with $P_{\rm cond}(\boldsymbol{\pi}|\mathbf{z}_0)$ as a priori generating probabilities. The Markov chain of paths $\{\mathbf{z}_{n,0}\}$ is Boltzmann distributed in $\widetilde{\mathcal{Z}}_{\frac{1}{2}}$, while the Markov chain of webs $\{\boldsymbol{\pi}_n\}$ is distributed with probability density $\Phi(\boldsymbol{\pi}_n|\mathbf{z}_{n,0})$. Hence, the formal average of a path quantity B (Eq. (34)) can be estimated from the following MCMC average

$$\langle B\rangle^\star_{\frac{1}{2}} = \frac{1}{N}\sum_{n=1}^{N}\sum_{i=0}^{I}\left[B(\mathbf{z}_{i,n})p_R(\mathbf{z}_{i,n}|\boldsymbol{\pi}_n)\right]. \tag{37}$$

Additionally, if one restricts the MCMC average to bottom and top webs, one obtains

$$\langle B\rangle^0_{\frac{1}{2}} = \frac{1}{N^\perp}\sum_{n=1}^{N^\perp}\sum_{i=0}^{I}\left[B(\mathbf{z}^\perp_{i,n})p_R(\mathbf{z}^\perp_{i,n}|\boldsymbol{\pi}^\perp_n)\right] \tag{38}$$

$$\langle B\rangle^1_{\frac{1}{2}} = \frac{1}{N^\top}\sum_{n=1}^{N^\top}\sum_{i=0}^{I}\left[B(\mathbf{z}^\top_{i,n})p_R(\mathbf{z}^\top_{i,n}|\boldsymbol{\pi}^\top_n)\right], \tag{39}$$

where $\boldsymbol{\pi}^\perp_n$ and $\boldsymbol{\pi}^\top_n$ are the bottom and top webs of the chain, $N^\perp$ and $N^\top$ their respective numbers, and, $\mathbf{z}^\perp_{i,n}$ and $\mathbf{z}^\top_{i,n}$ their respective paths. Finally, the web-ensemble average of a configurational quantity A considered in $\mathcal{Z}_\theta$ (Eq. (62)) can be estimated using

$$\langle A_\theta\rangle^\star_{\frac{1}{2}} = \frac{\sum_{n=1}^{N}\sum_{i=0}^{I}A(\mathbf{r}_{\theta,i,n})\Upsilon_\theta(\mathbf{z}_{i,n})p_R(\mathbf{z}_{i,n}|\boldsymbol{\pi}_n)}{\sum_{n=1}^{N}\sum_{i=0}^{I}\Upsilon_\theta(\mathbf{z}_{i,n})p_R(\mathbf{z}_{i,n}|\boldsymbol{\pi}_n)}. \tag{40}$$

4 Web averages for free energy differences

The partition function ratio $\widetilde{Z}_\theta/\widetilde{Z}_{\frac{1}{2}}$ can be computed in the web ensemble using respectively equations (37–39)

$$\frac{\widetilde{Z}_\theta}{\widetilde{Z}_{\frac{1}{2}}} = \langle\Upsilon_\theta\rangle^\star_{\frac{1}{2}} = \langle\Upsilon\rangle^0_{\frac{1}{2}} = \langle\Upsilon\rangle^1_{\frac{1}{2}} \tag{41}$$

recalling that the path function is $\Upsilon_\theta(\mathbf{z}) = \mathcal{K}_\theta(\mathbf{z})/\mathcal{K}_{\frac{1}{2}}(\mathbf{z})$. The free energy difference can then be obtained using three different methods

$$\beta\Delta F = \ln\langle\Upsilon_0\rangle^\star_{\frac{1}{2}} - \ln\langle\Upsilon_1\rangle^\star_{\frac{1}{2}} \tag{42}$$

$$= \ln\langle\Upsilon_0\rangle^1_{\frac{1}{2}} - \ln\langle\Upsilon_1\rangle^0_{\frac{1}{2}} \tag{43}$$

$$= \ln\langle\Upsilon_0\rangle^0_{\frac{1}{2}} - \ln\langle\Upsilon_1\rangle^1_{\frac{1}{2}}. \tag{44}$$

The average of equation (42) corresponds to an all information retrieving method (AIR). At variance, the average of equation (43) corresponds to a partial information retrieving method (PIR): unselected paths generated in $\widetilde{\mathcal{Z}}_\theta$ are discarded when estimating the $\widetilde{Z}_\theta/\widetilde{Z}_{\frac{1}{2}}$ ratio ($\theta = 0$ or 1). Finally, the web-average of equation (44) corresponds to the residence weight method, which was derived previously [5,6]. This method consists in retrieving information only of the trial paths generated in $\widetilde{\mathcal{Z}}_\theta$ (or approximately generated in $\widetilde{\mathcal{Z}}_\theta$ if the alternating shooting procedure is used as in Ref. [6,16]) when estimating the $\widetilde{Z}_\theta/\widetilde{Z}_{\frac{1}{2}}$ ratio ($\theta = 0$ or 1).

As an example, we show how to practically implement a web average. Combining equations (22), (28) and (35) and using the identity $\Upsilon_{1-\alpha} = [\Upsilon_\alpha]^{-1}$ yields

$$\Upsilon_\theta(\mathbf{z}_i)p_R(\mathbf{z}_i|\boldsymbol{\pi}) = \Upsilon_\theta(\mathbf{z}_i)\Upsilon_{1-\alpha}(\mathbf{z}_i)\left[\sum_{j=0}^{I}\Upsilon_{1-\alpha}(\mathbf{z}_j)\right]^{-1} = \frac{\exp\left[\beta(\alpha-\theta)W(\mathbf{z}_i)\right]}{\sum_{j=0}^{I}\exp\left[\beta(\alpha-\frac{1}{2})W(\mathbf{z}_j)\right]} \tag{45}$$

whatever $0 \le i \le I+1$ and where $\alpha = 1$ for a top web and $\alpha = 0$ for a bottom web. Then inserting equation (45) using $(\alpha,\theta) = (0,1)$ and $(1,0)$ into equations (38) and (39), respectively, yields the PIR estimates of the two following partition function ratios

$$\frac{Z_1}{\widetilde{Z}_{\frac{1}{2}}} = \frac{1}{N^\perp}\sum_{n=1}^{N^\perp}\frac{\sum_{i=0}^{I}\exp[-\beta W(\mathbf{z}^\perp_{i,n})]}{\sum_{i=0}^{I}\exp[-\frac{\beta}{2}W(\mathbf{z}^\perp_{i,n})]} \tag{46}$$

$$\frac{Z_0}{\widetilde{Z}_{\frac{1}{2}}} = \frac{1}{N^\top}\sum_{n=1}^{N^\top}\frac{\sum_{i=0}^{I}\exp[\beta W(\mathbf{z}^\top_{i,n})]}{\sum_{i=0}^{I}\exp[\frac{\beta}{2}W(\mathbf{z}^\top_{i,n})]}. \tag{47}$$

Using instead $(\alpha,\theta) = (1,1)$ and $(0,0)$ yields the RW estimates of these two partition function ratios

$$\frac{Z_1}{\widetilde{Z}_{\frac{1}{2}}} = \frac{I+1}{N^\top}\sum_{n=1}^{N^\top}\left[\sum_{i=0}^{I}\exp[\frac{\beta}{2}W(\mathbf{z}^\top_{i,n})]\right]^{-1} \tag{48}$$

$$\frac{Z_0}{\widetilde{Z}_{\frac{1}{2}}} = \frac{I+1}{N^\perp}\sum_{n=1}^{N^\perp}\left[\sum_{i=0}^{I}\exp[-\frac{\beta}{2}W(\mathbf{z}^\perp_{i,n})]\right]^{-1}. \tag{49}$$

They correspond to arithmetic averages of normalized residence weights. Then the RW estimates of the free energy difference are

$$\beta\Delta F = \ln\frac{N^\perp\sum_{n=1}^{N^\top}[\sum_{i=0}^{I}\exp[+\frac{\beta}{2}W(\mathbf{z}^\top_{i,n})]^{-1}}{N^\top\sum_{n=1}^{N^\perp}[\sum_{i=0}^{I}\exp[-\frac{\beta}{2}W(\mathbf{z}^\perp_{i,n})]^{-1}}. \tag{50}$$

The AIR estimates are deduced from the PIR and RW estimates by taking their arithmetic averages if partition function ratios are involved, or their exponential averages if free energy differences are instead concerned.

4.1 Web action identity

In order to cast the various information-retrieving averages (Eqs. (46–49)) into a single identity, let us define the following action

$$A^{\alpha}_{\theta}(\boldsymbol{\pi}) = -\frac{1}{\beta}\ln\left[\frac{1}{1+I}\sum_{i=0}^{I}\mathcal{K}_{\theta}(\mathbf{z}_i)\mathrm{P}^{\alpha}_{\mathrm{cond}}(\boldsymbol{\pi}|\mathbf{z}_i)\right] \quad (51)$$

where $\alpha = 0$ for a bottom web and $\alpha = 1$ for a top web. Then, the action difference written below simplifies to

$$\begin{aligned} A^{\alpha}_{\theta}(\boldsymbol{\pi}) - A^{\alpha}_{\frac{1}{2}}(\boldsymbol{\pi}) &= -\frac{1}{\beta}\ln\frac{\sum_{i=0}^{I}\exp\left[\beta(\alpha-\theta)W(\mathbf{z}_i)\right]}{\sum_{i=0}^{I}\exp\left[\beta(\alpha-\frac{1}{2})W(\mathbf{z}_i)\right]} \\ &= -\frac{1}{\beta}\ln\sum_{i=0}^{I}\Upsilon_{\theta}(\mathbf{z}_i)p_R(\mathbf{z}_i|\boldsymbol{\pi}), \end{aligned} \quad (52)$$

which allows to express the partition function ratio as an exponential average

$$\begin{aligned} \frac{\widetilde{Z}_{\theta}}{\widetilde{Z}_{\frac{1}{2}}} &= \exp\left[-\beta(\widetilde{F}_{\theta} - \widetilde{F}_{\frac{1}{2}})\right] \\ &= \left\langle \exp\left[-\beta(A^{\alpha}_{\theta}(\boldsymbol{\pi}) - A^{\alpha}_{\frac{1}{2}}(\boldsymbol{\pi})\right]\right\rangle^{\alpha}_{\frac{1}{2}} \end{aligned} \quad (53)$$

considered with respect to either top or bottom webs. Histograms of the action-difference will be constructed in order to localize the intersecting values with respect to the free energy differences $(\widetilde{F}_{\theta} - \widetilde{F}_{\frac{1}{2}})$ and to quantify the overlapping properties [22] of the sampled distributions.

5 Simulation results

5.1 Toy model

The three information-retrieving averages are implemented on a toy model system and compared to the standard MCMC average (M). The system consists in 50 harmonic oscillators. A parametrized internal energy is defined as follows

$$\mathcal{E}_{\lambda}(\mathbf{r}) = (1-\lambda)\frac{k_0}{2}\left[\mathbf{r} - \mathbf{c}/k_0\right]^2 + \lambda\frac{k_1}{2}\left[\mathbf{r} + \mathbf{c}/k_1\right]^2 \quad (54)$$

where $0 \leq \lambda \leq 1$ and whose parameters are given by

$$\begin{aligned} \mathbf{c} &= \mathbf{1}\sqrt{k_0}/3 \qquad & \mathbf{1} &= (1, ..., 1) \\ k_1 &= 1.2k_0 \qquad & k_0 &= 1. \end{aligned}$$

Here, the switching procedure between systems (0) and (1) is instantaneous, hence $\mathbf{z} = (\mathbf{r}, \mathbf{r})$, $W(\mathbf{z}) = \mathcal{E}_1(\mathbf{r}) - \mathcal{E}_0(\mathbf{r})$. Moreover, since there is a single path associated to any configuration, this path occurs with probability one. Therefore, $\widetilde{F}_{\alpha} = F_{\alpha}$ whatever $\alpha \in [0, 1]$, where $\widetilde{F}_{\alpha} = -\beta^{-1}\ln\widetilde{Z}_{\alpha}$. The relevant free energy differences of the system are

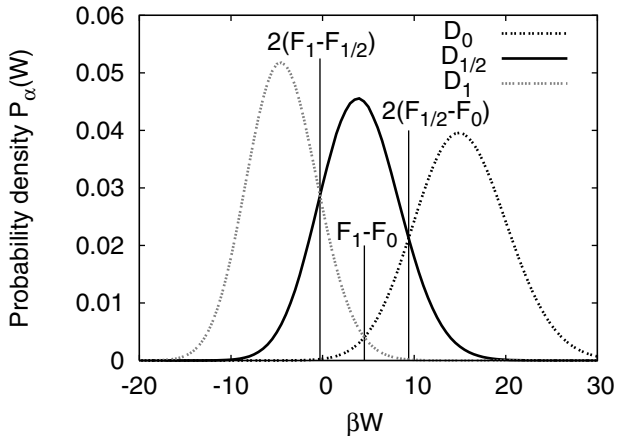

Fig. 3. Work-histograms obtained when instantaneously switching the particle system from $\mathcal{Z}_0$ to $\mathcal{Z}_1$ statistics (distribution D_0) or from $\mathcal{Z}_1$ to $\mathcal{Z}_0$ statistics (distribution D_1). The intermediate work-histogram $D_{1/2}$ corresponds to statistics $\widetilde{\mathcal{Z}}_{\frac{1}{2}}$.

$$F_1 - F_0 = 25\beta^{-1}\ln\frac{k_1}{k_0} \quad (55)$$

$$F_{\frac{1}{2}} - F_0 = 25\beta^{-1}\ln\frac{k_0+k_1}{2k_0} + E_{\frac{1}{2}} \quad (56)$$

$$F_1 - F_{\frac{1}{2}} = 25\beta^{-1}\ln\frac{2k_1}{k_0+k_1} - E_{\frac{1}{2}} \quad (57)$$

where $E_{\frac{1}{2}} = \mathcal{E}_{\frac{1}{2}}(\mathbf{r} = \mathbf{0})$.

Figure 3 displays the three work-histograms, D_0, $D_{\frac{1}{2}}$ and D_1, of the probability $P_{\theta}(W) = \langle\delta\left(W(\mathbf{z}) - W\right)\rangle_{\theta}$ with $\theta = 0$, $\frac{1}{2}$ and 1, respectively. Note that any two histograms D_{θ} and $D_{\theta'}$ are related by a Legendre transform

$$\left[\widetilde{F}_{\theta} - \frac{1}{\beta}\ln\mathrm{P}_{\theta}(W)\right] = \left[\widetilde{F}_{\theta'} - \frac{1}{\beta}\ln\mathrm{P}_{\theta'}(W)\right] - (\theta' - \theta)W$$

and, consequently, they intersect at the value

$$\frac{\widetilde{F}_{\theta} - \widetilde{F}_{\theta'}}{\theta - \theta'} = \frac{F_{\theta} - F_{\theta'}}{\theta - \theta'}. \quad (58)$$

These two properties stem from generalizing the non-equilibrium fluctuation theorem [16].

5.2 Free energy differences

We have estimated the free energy differences using the path-sampling scheme detailed in Section 2.1 using $p_0 = p_1 = 1/2$ and implemented the three information-retrieving MCMC averages detailed in Section 3.2 in addition to the conventional Metropolis average (M). Estimates of $F_1 - F_{\frac{1}{2}}$ and $F_1 - F_0$ have been computed using

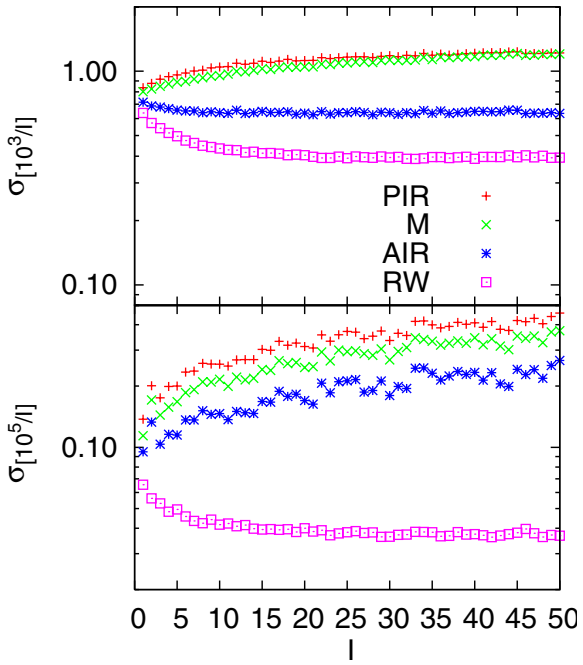

Fig. 4. Standard deviations, $\sigma_{[10^3/I]}$ and $\sigma_{[10^5/I]}$ computed using 10^4 and 10^3 estimates of $F_1 - F_{\frac{1}{2}}$, respectively, plotted as a function of I for the various information-retrieving methods.

blocks consisting of N/I ($N = 10^3$ or 10^5) and 10^5 webs, respectively. Figures 4 and 5a display the standard deviations, $\sigma_{10^3/I}$, and, $\sigma_{10^5/I}$, of $F_1 - F_{\frac{1}{2}}$ and the normalized standard deviations, $[\sigma_{10^5}]/I$, of $F_1 - F_0$, obtained with varying I. All standard deviations have then been computed using $B = 10^3$ estimates, except for $\sigma_{10^3/I}$ that used $B = 10^4$ estimates of $F_1 - F_{\frac{1}{2}}$. We observe that the PIR method yields the worst standard deviations. They are even larger than the ones given by the conventional Metropolis average (M). This feature illustrates the corollary by Jourdain and Delmas [9] stating that information-retrieving does not result, in general, in variance reduction (also demonstrated for the more general case $I \geq 1$). We also observe in Figures 4 and 5a that the RW method is the best for all I values, and that the AIR method exhibits an intermediate behaviour with standard deviations systematically better than the ones of the conventional method. The intermediate efficiency of the AIR method with respect to the PIR and RW methods stems from the fact that the AIR estimates are the exponential averages of the PIR and RW estimates.

The additional simulations displayed in Figure 5b have been carried out using the alternating procedure, which consists in alternately generating the trial paths forward and backward. Interestingly, this procedure results in standard deviations systematically smaller than the ones obtained with the parallel shooting procedure (see Fig. 5a). This is because the latter procedure exhibits larger correlating effects: repeatedly generating paths in the same direction indeed favours sequences of webs with similar actions, which results in increasing the statistical covariance.

To explain the origin of the detrimental effect of information retrieving from web sampling, we have extracted the action differences defined in equation (52) and constructed histograms of the quantities $(A^\theta_\theta - A^\theta_{\frac{1}{2}})/(\theta - \frac{1}{2})$ with increasing I values. Figure 6 displays the two series of histograms related to the RW method, with control parameters such that $(\theta, \alpha) = (0, 0)$ or $(\theta, \alpha) = (1, 1)$, while Figure 7 displays the two series of histograms related to the PIR method, with control parameters such that $(\theta, \alpha) = (0, 1)$ or $(\theta, \alpha) = (1, 0)$. In both Figures 6 and 7, the work histograms D_0, $D_{\frac{1}{2}}$ and D_1 have been plotted for comparison. We observe that all web histograms become narrower and narrower with increasing I and are shifted from distribution $D_{\frac{1}{2}}$ towards the distribution in which trial paths are generated (D_0 or D_1). The fraction of webs such that

$$\frac{A^\alpha_\theta(\boldsymbol{\pi}) - A^\alpha_{\frac{1}{2}}(\boldsymbol{\pi})}{\theta - \frac{1}{2}} \leq \frac{F^\alpha_\theta - F^\alpha_{\frac{1}{2}}}{\theta - \frac{1}{2}} \tag{59}$$

is equal to the histogram portion of Figures 6 and 7 that is located on the left of $2(F_1 - F_{\frac{1}{2}})$, for $\theta = 1$, or on the right of $2(F_{\frac{1}{2}} - F_0)$, for $\theta = 0$. This portion is called overlapping area. The contribution of the overlapping area to the free energy difference should exactly counter-balance the contribution of the non-overlapping area. Now, if the overlapping area is significantly smaller than the non-overlapping area, the standard deviation of the former contribution will be estimated using few data per blocks and will thus be significantly larger than the one of the latter contribution. The total precision and standard deviation of the free energy estimates will thus be dictated by the size of the two overlapping areas that are involved. We have thus collected the overlapping areas as a function of I from Figures 6 and 7 and plotted them in Figure 8. We observe that, with increasing I, the overlapping areas increase for the RW method and decrease for the PIR method, which well correlates with the observed beneficial or detrimental effect of information-retrieving.

Concerning the RW method, we also observe slow decreases of the standard deviations with increasing I. To understand this feature, we have investigated the correlating effects related to the Metropolis rejections: when averaging over the Markov chain of paths, a work quantity may appear several times in the successive residence weights, which results in increasing the standard deviations. Additional calculations have been carried out in order to quantify these correlating effects. We have first computed the standard deviations for blocks of size N/I and values of N ranging from 10^2 to 10^5. The same procedure was then repeated, but one out of twenty residence weights were used in the average, which required performing a simulation 20 times longer. This way of proceeding almost cancels the correlations between residence weights since the acceptance rate increases with I from 12% at $I = 1$ (see Fig. 9b). Figure 9a displays the obtained standard deviations. We observe that eliminating the correlations

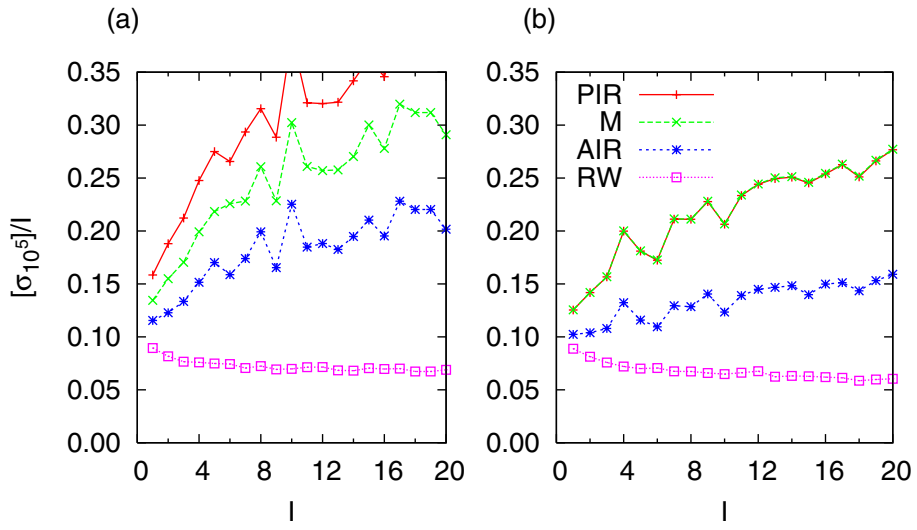

Fig. 5. Normalized standard deviations, $[\sigma_{10^5}]/I$ computed from $B = 10^3$ estimates and for the traditional Metropolis average (M), the AIR method, the PIR method and the RW method. Results using (a) the parallel shooting procedure ($p_0 = p_1 = \frac{1}{2}$), and (b) the alternating procedure.

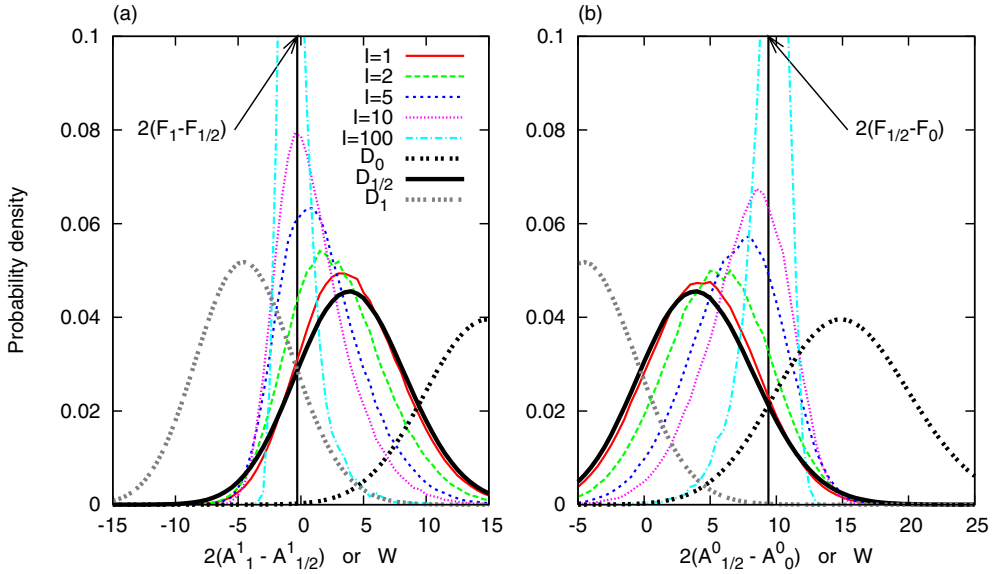

Fig. 6. Probability distributions of quantity $2[A^\theta_\theta(\boldsymbol{\pi}) - A^\theta_{\frac{1}{2}}(\boldsymbol{\pi})]$ (twice the residence-weight logarithms) for: (a) top webs and (b) bottom webs.

substantially decreases the standard deviation but only for small block sizes $N/I \leq 100$. Hence correlating effects only arise when a small number of paths can be afforded in the calculations which happens when system (0) and (1) strongly differ from each other, thus requiring to invest most of the computational resources on few paths with a very slow switching rate. In this case, it is beneficial to generate the trial paths in parallel, since the acceptance rate would increase (see Fig. 9b) and the detrimental correlating effects would be reduced at the same time (see Fig. 9a). Dispatching the generation of the trial moves on a parallel computer architecture should be done since the resulting gain of efficiency would largely compensate for the communication costs between processors.

5.3 Landau free energies

We have observed that information-retrieving is beneficial in computations of free energy differences when both the target ensemble and the ensemble in which trial paths are generated coincide. This behaviour results from the fact that the unselected trial moves contain an information relevant to the thermodynamic quantity to estimate. To complete our study, we now investigate the effect of information-retrieving when the sampled distribution coincides with the targeted distribution but differs from the distribution in which trial moves are generated. This situation arises in parallel tempering computations of a Landau free energy. With the present toy model, it amounts

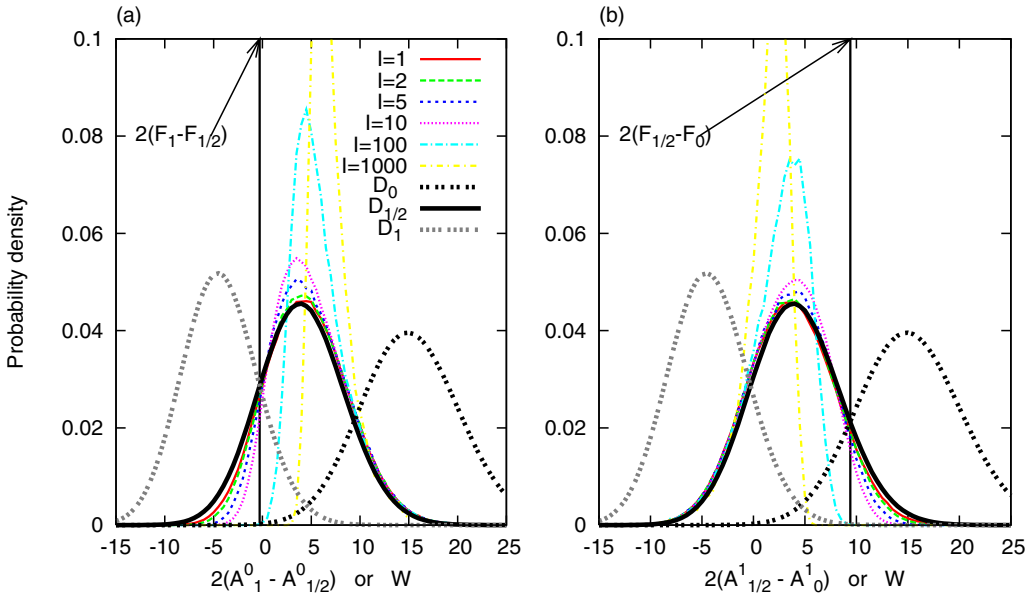

Fig. 7. Probability distribution of quantity $2[A_\theta^{1-\theta}(\boldsymbol{\pi}) - A_{\frac{1}{2}}^{1-\theta}(\boldsymbol{\pi})]$: (a) bottom webs and (b) top webs.

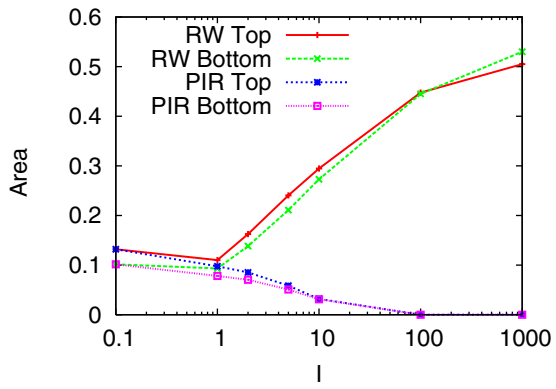

Fig. 8. Overlapping areas for the RW and PIR methods and for bottom and top webs, as a function of I. The values for $I = 0.1$ corresponds to the overlapping areas of the corresponding work histogram (D_0 or D_1).

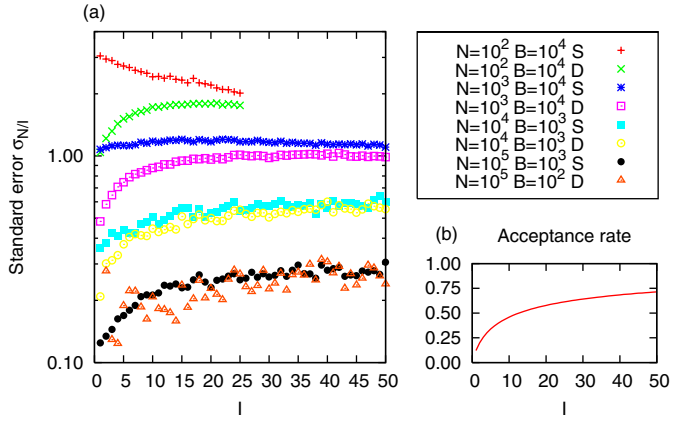

Fig. 9. (a) Standard deviations, $\sigma_{[N/I]}$, of $F_1 - F_0$ estimates, plotted as a function of I and for the RW method and various values of N: B is the block number, D indicates the decorrelating procedure and S indicates standard; (b) acceptance rate as a function of I.

to compute a Landau free energy in ensemble $\mathcal{Z}_{\frac{1}{2}}$, while proposing replica exchanges between ensembles $\mathcal{Z}_0$ and $\mathcal{Z}_{\frac{1}{2}}$ or between ensembles $\mathcal{Z}_{\frac{1}{2}}$ and $\mathcal{Z}_1$. The exponential of the Landau energy corresponds to a probability density defined with respect to an order parameter. Such a density can be estimated from a direct MCMC average. Considering that the internal energy is the order parameter, we have computed the probability density $P_{\frac{1}{2}}(E)$ by averaging the configurational function $A(\mathbf{r}) = \delta(\mathcal{E}_{\frac{1}{2}}(\mathbf{r}) - E)$ using equation (40) with $\theta = \frac{1}{2}$. Trial configurations are generated in both $\mathcal{Z}_0$ and $\mathcal{Z}_1$. This case study partly reproduces the computational set-up of reference [23] which combines information-retrieving and parallel tempering for computing a Landau free energy.

Figure 10 displays the energy histogram obtained with the AIR method and the conventional method (M). We observe that information-retrieving allows to explore a larger portion of the histogram, in agreement with the behaviour reported in reference [23], even though only rough estimates are obtained in the histogram tails. Indeed, information retrieving is observed to have a moderate effect (even if always beneficial) on the standard deviations, in contrast to study [23] that reported a considerable speed-up of the calculations. However, our investigation differs from the investigation of Coluzza et al. [23] in the fact that the latter one introduced an umbrella (or auxiliary) potential that was optimized on-the-fly

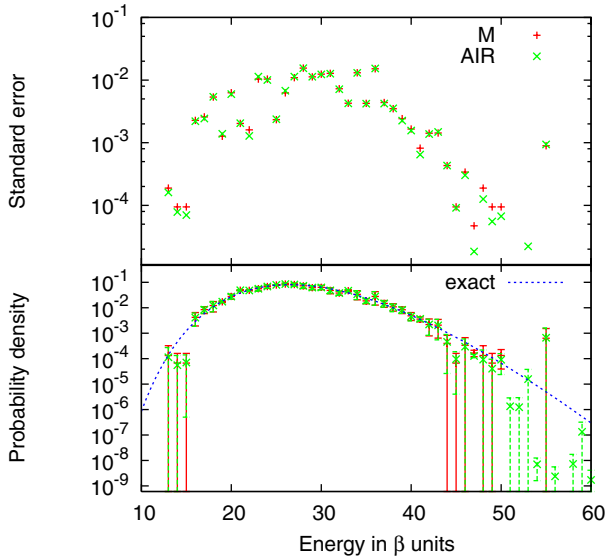

Fig. 10. Bottom panel: probability distribution of the potential energy $\mathcal{E}_{\frac{1}{2}}$. The exact analytical form is $\mathrm{P}(E) = \exp\left[\beta(E_0 - E)\right] \times (E - E_0)^{m/2-1}\beta^{m/2}/((m/2-1)!)$ where $m = 50$ and $E_0 = 2.52$. Error bars correspond to the standard deviations of the top panel.

using an adaptive procedure: flat-sampling along the order parameter was thus insured. In the light of the calculations displayed in Figure 10, we argue that owing to additional estimates along the order parameter that are made possible by retrieving all available information, the auxiliary potential is constructed faster and, subsequently, the adaptive procedure converges considerably faster. This argument may also apply to explain a similar speed-up reported in a simulation study [24] investigating the relevance of information-retrieving, using density-of-states sampling [25] instead of parallel tempering. Hence, information-retrieving is systematically found to be of relevance in *non-equilibrium* techniques involving either path-sampling or adaptive-sampling.

6 Alternative derivation

In this section, we give an alternative derivation of web sampling and information-retrieving averages. The objective is to clarify and correct the approach that was proposed in reference [5,6] to justify the residence weight method.

We start by noticing that, in the course of a Monte Carlo simulation, the path $\mathbf{z}_{i_s,n}$ that is selected at the n-th step corresponds to the path that would be selected at the $n+1$-st step of the Monte Carlo simulation that constructs the reverse Markov chain. We have $\mathbf{z}_{0,n+1} = \mathbf{z}_{i_s,n}$. Let then assume that the symmetric selecting probabilities of equation (21) are used, and let also decompose the forward and backward selecting symmetric probabilities as a function of the web conditional probabilities

$$p_R(\mathbf{z}_{0,n+1}|\boldsymbol{\pi}_n) = \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_{0,n+1})}{\widetilde{Z}_{\frac{1}{2}}} \frac{P_{\mathrm{cond}}(\boldsymbol{\pi}_{n+1}|\mathbf{z}_{0,n+1})}{(I+1)\Phi(\boldsymbol{\pi}_{n+1})}$$

$$p_R(\mathbf{z}_{0,n+1}|\boldsymbol{\pi}_{n+1}) = \frac{\mathcal{K}_{\frac{1}{2}}(\mathbf{z}_{0,n+1})}{\widetilde{Z}_{\frac{1}{2}}} \frac{P_{\mathrm{cond}}(\boldsymbol{\pi}_{n}|\mathbf{z}_{0,n+1})}{(I+1)\Phi(\boldsymbol{\pi}_{n})},$$

where a mean web probability density is introduced

$$\Phi(\boldsymbol{\pi}) = \frac{1}{I+1}\sum_{i=0}^{I} \Phi(\boldsymbol{\pi}|\mathbf{z}_i). \tag{60}$$

From these expressions, it immediately follows that the detailed balance equation

$$\Phi(\boldsymbol{\pi}_n)p_R(\mathbf{z}_{0,n+1}|\boldsymbol{\pi}_n)P_{\mathrm{cond}}(\boldsymbol{\pi}_{n+1}|\mathbf{z}_{0,n+1}) = \Phi(\boldsymbol{\pi}_{n+1})p_R(\mathbf{z}_{0,n+1}|\boldsymbol{\pi}_{n+1})P_{\mathrm{cond}}(\boldsymbol{\pi}_{n}|\mathbf{z}_{0,n+1}) \tag{61}$$

is obeyed in the web ensemble. This implies that the Markov chain of webs $\{\boldsymbol{\pi}_n\}$ is distributed according to the probability density Φ. Now, the $I+1$ equivalent web-averages of equation (29) can be averaged as follows

$$\langle B\rangle^{\star}_{\frac{1}{2}} = \frac{1}{I+1}\int \sum_{i=0}^{I} B(\mathbf{z}_i)\Phi(\boldsymbol{\pi}|\mathbf{z}_i)\mathcal{D}\boldsymbol{\pi}, \tag{62}$$

$$= \frac{1}{I+1}\int \sum_{i=0}^{I} B(\mathbf{z}_i)\left[\Phi(\boldsymbol{\pi}|\mathbf{z}_i)/\Phi(\boldsymbol{\pi})\right]\Phi(\boldsymbol{\pi})\mathcal{D}\boldsymbol{\pi}.$$

Estimating the web-average of equation (62) from the Markov chain $\{\boldsymbol{\pi}_n\}$ requires to re-weigh by dividing with the generated web weights $\Phi(\boldsymbol{\pi}_n)$, which yields

$$\langle B\rangle^{\star}_{\frac{1}{2}} = \frac{1}{N(I+1)}\sum_{n=1}^{N}\sum_{n=0}^{I}\left[B(\mathbf{z}_{i,n})\frac{\Phi(\boldsymbol{\pi}_n|\mathbf{z}_{i,n})}{\Phi(\boldsymbol{\pi}_n)}\right]. \tag{63}$$

The Monte Carlo averages of equation (63) obtained for the web-sampling scheme is strictly identical to equation (37).

The approach presented here that considers detailed balance in the web ensemble itself is equivalent to the one given in reference [5,6] that involved a weighted detailed balance equation with symmetric selecting probabilities $p_R(\mathbf{z}_i|\boldsymbol{\pi})$. However, the generalization of the weighted detailed balance equation that was derived for the Metropolis rule [12] is not correct, because the web probability density that would be defined from the residence weights would depend on the result of the Metropolis procedure. The residence weights that were then derived omitted a contribution and depended on acceptance or rejection. Paraphrasing a remark made by Boulougouris and Frenkel in the [10,11], the crucial point here is that the residence weights are independent of the algorithm used in the sampling process. Hence, the symmetrical form used in the present study and in reference [5,6] should always be used.

Since erroneous residence weights were used in references [12,16], we wish to quantify the systematic error that was introduced. Considering the present toy model and assuming the estimates normally distributed, one obtains [4.55677; 4.55833] for the 68% confidence interval of the mean, calculated using $N = 10^3$ estimates and $B = 10^6$ webs per estimate. The correct value, $\Delta F = 4.5580$, lies in this interval hence the error is hardly detectable. This is because the first error introduced in the numerator compensates for the second one introduced in the denominator. Hence, if we set $k_1/k_0 = 1.5$ (instead of 1.2) in order make the work-distributions more asymmetric, we find [10.12313, 10.13475] for a 68%-confidence interval obtained using $N = 10^3$ and $B = 10^4$. The exact value $\Delta F = 10.13662$ is clearly outside the interval. Now, in the model system used in reference [16], the D_0 and D_1 work-histograms are even more symmetrical than the ones of our toy model with $k_1/k_0 = 1.2$. Moreover, they are shifted to a high value $\beta\Delta F = -1457.4 \pm 0.2$. As a result, the relative error that was introduced is completely undetectable and is of order 10^{-6}.

Related to the comment above is the fact that when the path-sampling scheme with the Metropolis rule (Eq. (20)) is used, no detailed balance condition is satisfied in the web ensemble. The webs $\boldsymbol{\pi}_n$ associated to the generated paths $\mathbf{z}_{0,n}$ are simply distributed with probability $\Phi(\boldsymbol{\pi}_n|\mathbf{z}_{0,n})$. Since permuting the paths of the webs belonging to the constructed chain (performing $\{\boldsymbol{\pi}_n\} \to \{\bar{\boldsymbol{\pi}}_n^i\}$) does not affect the MCMC estimate of equation (62), the web chain seems to be distributed with respect to any probability density $\Phi(\boldsymbol{\pi}|\mathbf{z}_i)$, and, consequently, with respect to the mean probability density (Eq. (60)) as well, although it is not the case. This feature indeed illustrates the well-known statement that imposing detailed balance with respect to the probability density of a given ensemble is not a necessary condition to insure that the MCMC estimates converge in this ensemble.

7 Concluding remarks

In this article, statistical averages of thermodynamic quantities were formulated in an ensemble consisting of webs, defined as the sets of paths linked by a parallel stochastic procedure. MCMC averages can be taken in this web ensemble which allows to include information relative to Monte Carlo trial moves that have been discarded by the Metropolis acceptance procedure. The various MCMC averages that have been proposed have been implemented for estimating a free energy difference. Variance reduction with respect to the conventional sampling procedure is only observed when the unselected moves contain an information relevant to the quantity to compute, i.e. when the wasted work quantities strongly contribute to the free energy difference. In the present study, this requirement is satisfied by the residence weight method. Moreover, we numerically observe that, with this information-retrieving method, parallelization of the Metropolis procedure does not result in an increase of the standard deviation. This feature indicates a straightforward way to parallelize algorithms dedicated to the computation of free energy differences.

Finally, it is worth noting that the residence weights correspond to Rosenbluth factors involving path quantities. They present a formal similarity with the Rosenbluth factors that were derived and incorporated into the non-equilibrium work method by Wu and Kofke [26] in order to improve its computational efficiency. The RW method differs from the method of Wu and Kofke in the fact that two types of Rosenbluth factors are involved instead of only one type. This feature makes it possible to retrieve the relevant information from the two possible switching directions between system (1) and system (0). Indeed, a first type of Rosenbluth factor gives access to the $F_1 - F_{\frac{1}{2}}$ difference of free energy, while the second type to the $F_0 - F_{\frac{1}{2}}$ difference, in both case considered with respect to the intermediate system (1/2). Information-retrieving averages may thus be generalized to allow for a bidirectional implementation of the various multi-stage Rosenbluth-sampled techniques proposed by Wu and Kofke [26] and also of the interactive particle system techniques proposed by Rousset and Stoltz [27,28]. Alternatively, information-retrieving averages could have been implemented for computing the reverse free energy differences $F_{\frac{1}{2}} - F_1$ and $F_{\frac{1}{2}} - F_0$, where system (1/2) is considered with respect to systems (0) and (1), as it is done in overlap sampling [29]. Hence, the web ensemble formalism introduced in the present study might be further considered for improving existing free energy methods.

I am grateful to the European Science Foundation and the SimBioMa program for funding the workshop "Sampling paths in Molecular Simulation" that stimulated discussions on the present topic. Fruitful discussions with J.-M. Delmas, D. Frenkel, B. Jourdain and with participants of the workshop, G. Adjanor, D. Kofke, M. Rousset, G. Stoltz are gratefully acknowledged.

References

1. J.-M. Lanore, Rad. Eff. **22**, 153 (1974)
2. A.B. Bortz, M.H. Kalos, J.L. Lebowitz, J. Comp. Phys. **17**, 10 (1975)
3. M. Athènes, P. Bellon, G. Martin, Phil. Mag. A **76**, 527 (1997)
4. D.R. Mason, R.E. Rudd, A.P. Sutton, Computer Physics Communications **160**, 140 (2004)
5. M. Athènes, Phys. Rev. E **66**, 016701 (2002)
6. M. Athènes, Phys. Rev. E **66**, 046705 (2002)
7. D. Frenkel, *Proc. Natl. Acad. Sci. USA* **101**, 17571 (2004)
8. D. Cerpeley, C.G.V. K.M.H. Phys. Rev. B **16**, 3081 (1977)
9. J.-F. Delmas, B. Jourdain, *Does waste-recycling really improve Metropolis-Hastings Monte Carlo algorithm?* `arXiv:math/0611949v2` `http://fr.arxiv.org/abs/math/0611949v2` (submitted)
10. G.C. Boulougouris, D. Frenkel, J. Chem. Theory and Computation **1**, 389 (2005)

11. D. Frenkel, in *Lecture Notes in Physics Computer simulations in condensed matter systems: from materials to chemical biology,* edited by M. Ferrario, G. Ciccotti, K. Binder (Springer-Berlin, 2006), Vol. **703**, pp. 127–137
12. M. Athènes, Eur. Phys. J. B **38**, 651 (2004)
13. C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997)
14. G. Adjanor, M. Athènes, F. Calvo, Eur. Phys. J. B **53**, 47 (2005)
15. Ytreberg, D. Zuckermann, J. Chem. Phys. **120**, 10876 (2004)
16. G. Adjanor, M. Athènes, J. Chem. Phys. **123**, 234104 (2006)
17. H. Oberhofer, C. Dellago, P. Geissler, J. Phys. Chem. B **109**, 6902 (2005)
18. F.M. Ytreberg, R.H. Swendsen, D.M. Zuckermann, J. Chem. Phys. **125**, 184114 (2006)
19. W. Lechner, C. Dellago, J. Stat. Mech. P04001 (2007)
20. C. Dellago, P. Bolhuis, F. Csajka, D. Chandler, J. Chem. Phys. **108**, 1964 (1998)
21. C. Dellago, P. Bolhuis, P. Geissler, Adv. Chem. Phys. **123**, 1 (2002)
22. C.H. Bennett, J. Comp. Phys. **22**, 245 (1976)
23. I. Coluzza, D. Frenkel, Chem. Phys. Chem **6**, 1779 (2005)
24. M. Chopra, J.J. de Pablo, J. Chem. Phys. **124**, 114102 (2006)
25. F. Wang, D. Landau, Phys. Rev. Lett. **86**, 2050 (2001)
26. D. Wu, D. Kofke, J. Chem. Phys. **122**, 204104 (2005)
27. G. Stoltz, M. Rousset, J. Stat. Phys. **123**, 1251 (2006)
28. G. Stoltz, J. Comput. Phys., (in press, 2007)
29. N. Lu, T.B. Woolf, D. Kofke, Phys. Rev. E **69**, 057702 (2004)