JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Nonequilibrium Candidate Monte Carlo Simulations with Configurational Freezing Schemes

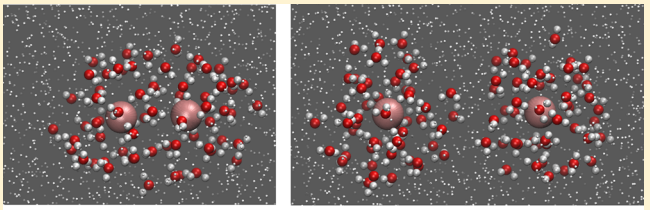
Edoardo Giovannelli,[†] Cristina Gellini,[†,‡] Giangaetano Pietraperzia,[†,‡] Gianni Cardini,[†,‡] and Riccardo Chelli*[,†,‡]

[†]Dipartimento di Chimica, Università di Firenze, Via della Lastruccia 3, I-50019 Sesto Fiorentino, Italy

[‡]European Laboratory for Nonlinear Spectroscopy (LENS), Via Nello Carrara 1, I-50019 Sesto Fiorentino, Italy

Supporting Information

ABSTRACT: Nonequilibrium Candidate Monte Carlo simulation [Nilmeier et al., *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1009−E1018] is a tool devised to design Monte Carlo moves with high acceptance probabilities that connect uncorrelated configurations. Such moves are generated through nonequilibrium driven dynamics, producing candidate configurations accepted with a Monte Carlo-like criterion that preserves the equilibrium distribution. The probability of accepting a candidate configuration as the next sample in the Markov chain basically depends on the work performed on the system during the nonequilibrium trajectory and increases with decreasing such a work. It is thus strategically relevant to find ways of producing nonequilibrium moves with low work, namely moves where dissipation is as low as possible. This is the goal of our methodology, in which we combine Nonequilibrium Candidate Monte Carlo with Configurational Freezing schemes developed by Nicolini et al. (*J. Chem. Theory Comput.* **2011**, *7*, 582−593). The idea is to limit the configurational sampling to particles of a well-established region of the simulation sample, namely the region where dissipation occurs, while leaving fixed the other particles. This allows to make the system relaxation faster around the region perturbed by the finite-time switching move and hence to reduce the dissipated work, eventually enhancing the probability of accepting the generated move. Our combined approach enhances significantly configurational sampling, as shown by the case of a bistable dimer immersed in a dense fluid.



Relaxing particles around dissipation region makes easier nonequilibrium Monte Carlo moves

1. INTRODUCTION

Equilibrium configurational sampling of chemical and physical systems is numerically realized through Monte Carlo (MC) and molecular dynamics simulations.[1] In the simplest MC scheme, particles are moved one at a time, while equations of motion enforced in molecular dynamics simulations provide a simultaneous evolution of the whole system. Nevertheless, the two methods are comparable in terms of computational efficiency. These techniques can produce highly correlated samples, which can slow the convergence of computed thermodynamical properties. An advantage of MC sampling in tackling this problem is its considerable flexibility in devising new types of moves.[1,2] For example, configurational mixing can be improved making use of cleverly constructed sets of moves, such as nonlocal moves, based on a priori knowledge of the system.

In this respect, Nilmeier and co-workers developed a Nonequilibrium Candidate Monte Carlo (NCMC) approach,[3] extending the Metropolis MC[4] method to nonequilibrium moves aimed at generating candidate configurations for equilibrium simulations. Unlike conventional MC simulations, the acceptance probability, established by the detailed balance condition, is determined by the nonequilibrium work performed in a finite time interval, rather than by the difference of potential energy before and after the attempted move. Configurational transitions, that in a normal MC simulation are rare because of high energy barriers or simply for the large viscosity of the sample, are forced by the NCMC switching process, which generates candidates with greater acceptance probabilities. Whereas generating finite-time switching trajectories incurs in additional cost in NCMC simulations, driving some degrees of freedom while allowing others to evolve naturally can lead to enhancements in acceptance probabilities due to a significant reduction of the structural correlation times.

A test of the NCMC methodology was performed on a bistable dimer in which compact and extended configurations are separated by a high energy barrier.[3] A dense solvent cage around the dimer increases rejection of attempted moves, preventing a direct transition between the two minimum energy configurations. Transition rate has been enhanced by means of NCMC, which converts the transition between compact and extended configurations in a nonequilibrium process. Overlap of solvent particles with the dimer is avoided alternating a series of perturbation steps that carry on the transition, with propagation steps that equilibrate the system during the nonequilibrium path, eventually reducing dissipation. It is worth noting the importance of selecting collective

Received: April 21, 2014
Published: August 18, 2014

   dx.doi.org/10.1021/ct500340b | *J. Chem. Theory Comput.* 2014, 10, 4273−4283

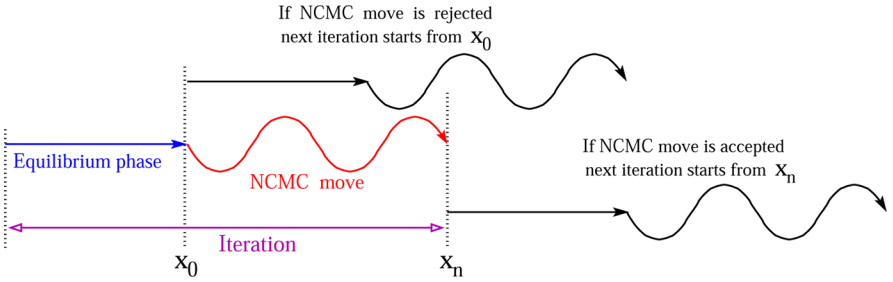


Figure 1. Schematic representation of a NCMC simulation protocol. An *iteration* (in magenta) consists of an initial *equilibrium phase* (in blue) and of a nonequilibrium switching phase, called NCMC move (in red). The NCMC move starts and ends at system configurations $\mathbf{x}_0$ and $\mathbf{x}_n$, respectively. If the NCMC move is accepted then the next iteration begins from the configuration $\mathbf{x}_n$. Upon rejection, the system is brought back to the configuration $\mathbf{x}_0$, where the new iteration starts. This procedure is repeated a given number of times during the simulation.

coordinates such that unbound degrees of freedom (i.e., not directly involved in the collective coordinate definition) guarantee a quick relaxation of the system along the generated driven trajectory.

In this article, we present a method that allows for a significant improvement of NCMC by enhancing structural relaxation around the space domain where the nonequilibrium driven transition occurs. This approach is based on the direct application of a technique, called Configurational Freezing[5,6] (CF), to NCMC simulations. CF has been developed in the context of free energy calculations via nonequilibrium work theorems.[7−10] In fast-switching nonequilibrium simulations, the system is mainly perturbed around the reaction site, and hence, dissipation is localized in that region of space.[5] On the basis of such an assumption, only propagation MC moves involving particles near the reaction site may significantly reduce dissipation. Accordingly, in CF, a particle is chosen for a trial move with probability dependent on its distance from the reactive centers. In such a way, we define a reaction domain, called mobility region. Inside the mobility region, the probability of selecting a particle for a trial move may be constant (uniform particle selection scheme[5]) or may decrease with the distance of the particle from the reactive centers (nonuniform particle selection scheme[6]). Outside the mobility region, no particle can be selected for a trial move. Detailed balance condition imposes rejection of moves that drive the particles out of the mobility region. Therefore, under CF moves, no particle traffic is allowed through the mobility region surface. Although sampling is limited to a restricted phase-space region, no approximation is introduced in the application of nonequilibrium work theorems (see discussion in ref 5). With respect to conventional MC simulations, CF enriches the sampling of configurational states in space regions near the reactive site, lowering dissipation of the process. Among the approaches developed to improve efficiency of free energy calculations,[11−16] CF and its molecular dynamics analog, called Dynamical Freezing,[17] presents the advantage of a low dependence of performances on the sample size.

In summary, the target of CF is to reduce the work made on the system during an externally driven process. In fact, lowering dissipation is fundamental to achieve accurate free energy estimates by work evaluation along nonequilibrium paths.[18,19] As stated above, in NCMC simulations, such a work lowering may lead to high transition rates along the driven collective coordinate, because acceptance probability is proportional to $\min(1, \exp(-\beta W))$, where W is the work performed on the system during the driven trajectory. This ultimately leads to an enhanced sampling of the system configurations along the collective coordinate.

Following Nilmeier and co-workers,[3] the technique resulting from combining CF and NCMC, from now on CF-NCMC, is tested on a bistable dimer immersed in a medium. We will present results illustrating the different sampling behavior of standard MC, NCMC, and CF-NCMC simulation schemes.

The outline of the article follows. NCMC and CF methods are first presented separately in sections 2.1 and 2.2, while their combination, CF-NCMC, is presented in section 2.3. The test system and the simulation details are described in section 3. The results of our tests are illustrated and discussed in section 4, while concluding remarks are given in section 5.

2. THEORY

2.1. Nonequilibrium Candidate Monte Carlo. Here, we introduce the NCMC method for the equilibrium sampling of a system in a given thermodynamic state. A more detailed description of NCMC, within a generalized-ensemble simulation framework, can be found in ref 3. In NCMC, a collective coordinate connecting states separated by high energy barriers is defined with the aim of enhancing the importance sampling of these states through nonequilibrium MC moves along the collective coordinate itself. Configurational transitions are thus promoted using noninstantaneous nonequilibrium MC moves accepted with a probability that preserves the equilibrium distribution. Examples of such transitions are the folding/unfolding of a solvated biopolymer, changes in a set of torsional angles of a molecule, or the formation/destruction of a molecular complex.

Although the general form of NCMC presented here holds for both molecular dynamics and MC simulation protocols,[3] our interest will be addressed to MC sampling as the natural framework of CF schemes (see later in section 2.2). For simplicity, let us consider a system at constant volume and temperature, whose potential energy $U(\mathbf{x})$ is a function of the coordinates $\mathbf{x}$ of the N particles. According to the Boltzmann distribution, the probability of a configuration $\mathbf{x}$ corresponds to $\rho(\mathbf{x}) = Z^{-1} \exp[-\beta U(\mathbf{x})]$, where β is the inverse temperature and $Z = \int \exp[-\beta U(\mathbf{x})] d\mathbf{x}$ is the partition function. NCMC is realized by carrying on a simulation made of a series of iterations, each consisting of an equilibrium sampling phase alternated with a nonequilibrium switching phase (referred to as NCMC phase or NCMC move in the following). A schematic view of the simulation protocol is reported in Figure 1. The equilibrium phase is performed with a conventional MC protocol, while in the nonequilibrium phase the system is externally escorted from an initial to a final configuration

dx.doi.org/10.1021/ct500340b | J. Chem. Theory Comput. 2014, 10, 4273−4283

according to some protocol Λ. For each iteration, the final configuration of the equilibrium phase, say $\mathbf{x}_0$, corresponds to the starting configuration of the nonequilibrium phase, at the end of which a configuration $\mathbf{x}_n$ is reached. Finally, an acceptance probability is computed and used to decide if the candidate configuration $\mathbf{x}_n$ can or cannot be accepted as the next sample in the Markov chain, namely as the initial sample of the equilibrium phase of the next iteration. If the candidate configuration $\mathbf{x}_n$ is not accepted then the new equilibrium phase begins with initial configuration $\mathbf{x}_0$ (the acceptance−rejection criterion is also represented in the scheme of Figure 1). The protocol Λ of the nonequilibrium process is uniquely determined by the starting configuration and defines an alternating pattern of perturbation and propagation kernels as follows: $\Lambda = (T_1, S_1, T_2, S_2, ..., T_n, S_n)$. In general, the perturbation kernel, $T_i \equiv T(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i^*)$, can consist of either stochastic or deterministic changes of the collective coordinate.[3] In the present work, it is build by deterministic changes of the collective coordinate defined as a geometrical relation between a subset of the system degrees of freedom. In principle, such a coordinate can be defined arbitrarily, but it is basically designed on the basis of the slow degrees of freedom to be sampled with NCMC. In the most simple cases, the collective coordinate can be, for example, an interatomic distance or a torsion angle, but even more fancy definitions, such as the coordination number of a particle or molecule, can be adopted. Due to the action of the perturbation kernel, the system evolves from a relaxed $\mathbf{x}_{i-1}$ configuration to a perturbed $\mathbf{x}_i^*$ configuration. The propagation kernel, $S_i \equiv S(\mathbf{x}_i^* \rightarrow \mathbf{x}_i)$, allows the set of unbound degrees of freedom, not entering the collective coordinate, to freely evolve, so that the system relaxes from the $\mathbf{x}_i^*$ to the $\mathbf{x}_i$ configuration. The degrees of freedom entering the collective coordinate remain fixed during the propagation steps. The specific design of the kernels T_i and S_i employed in the present study is given in sections 2.1.1 and 2.3, respectively. The rationale of introducing propagation kernels is to lower the work dissipated along the nonequilibrium processes, to eventually increase the acceptance probability of candidate configurations. The system evolution under the nonequilibrium phase of each iteration can be described by the following trajectory X:

$$X = (\mathbf{x}_0 \overset{T_1}{\rightarrow} \mathbf{x}_1^* \overset{S_1}{\rightarrow} \mathbf{x}_1 \rightarrow \cdots \rightarrow \mathbf{x}_{n-1} \overset{T_n}{\rightarrow} \mathbf{x}_n^* \overset{S_n}{\rightarrow} \mathbf{x}_n) \tag{1}$$

The reversed trajectory $\tilde{X}$ is generated by the protocol $\tilde{\Lambda}$, obtained by reversing the order in which the perturbation and propagation steps are applied:

$$\tilde{X} = (\mathbf{x}_n \overset{\tilde{S}_n}{\rightarrow} \mathbf{x}_n^* \overset{\tilde{T}_n}{\rightarrow} \mathbf{x}_{n-1} \rightarrow \cdots \rightarrow \mathbf{x}_1 \overset{\tilde{S}_1}{\rightarrow} \mathbf{x}_1^* \overset{\tilde{T}_1}{\rightarrow} \mathbf{x}_0) \tag{2}$$

where $\tilde{T}_i \equiv T^{-1}(\mathbf{x}_i^* \rightarrow \mathbf{x}_{i-1})$ and $\tilde{S}_i \equiv S(\mathbf{x}_i \rightarrow \mathbf{x}_i^*)$. The starting configuration of the nonequilibrium phase will determine the direction of the protocol. To ensure that, under application of a protocol Λ, the stationary distribution $\rho(\mathbf{x})$ is preserved, a strict pairwise form of the detailed balance is enforced[3]

$$\rho(\mathbf{x}_0)\mathcal{P}(X|\mathbf{x}_0, \Lambda)P(\Lambda|\mathbf{x}_0)A(\mathbf{x}_0 \rightarrow \mathbf{x}_n) = \rho(\mathbf{x}_n)\mathcal{P}(\tilde{X}|\mathbf{x}_n, \tilde{\Lambda}) \times P(\tilde{\Lambda}|\mathbf{x}_n)A(\mathbf{x}_n \rightarrow \mathbf{x}_0) \tag{3}$$

where $A(\mathbf{x}_0 \rightarrow \mathbf{x}_n)$ is the NCMC acceptance probability of trajectory X, $\mathcal{P}(X|\mathbf{x}_0, \Lambda)$ denotes the probability of generating the trajectory X given the initial configuration $\mathbf{x}_0$ using the protocol Λ and $P(\Lambda|\mathbf{x}_0)$ is the probability of generating the protocol Λ given the initial configuration $\mathbf{x}_0$, such that there exists a reverse protocol $\tilde{\Lambda}$ with $P(\tilde{\Lambda}|\mathbf{x}_n) > 0$. The quantities $A(\mathbf{x}_n \rightarrow \mathbf{x}_0)$, $\mathcal{P}(\tilde{X}|\mathbf{x}_n, \tilde{\Lambda})$ and $P(\tilde{\Lambda}|\mathbf{x}_n)$, related to the reverse direction, have analogous meaning. According to the protocols Λ and $\tilde{\Lambda}$ of eqs 1 and 2, the probabilities $\mathcal{P}(X|\mathbf{x}_0, \Lambda)$ and $\mathcal{P}(\tilde{X}|\mathbf{x}_n, \tilde{\Lambda})$ are defined as[3,8]

$$\mathcal{P}(X|\mathbf{x}_0, \Lambda) = \prod_{i=1}^{n} T_i S_i \tag{4}$$

$$\mathcal{P}(\tilde{X}|\mathbf{x}_n, \tilde{\Lambda}) = \prod_{i=1}^{n} \tilde{T}_i \tilde{S}_i \tag{5}$$

Thus, from the probabilistic point of view, kernels S_i and T_i represent conditional probabilities related to single propagation and perturbation steps, respectively. Exploiting eqs 4 and 5 into eq 3 and assuming a canonical stationary distribution $\rho(\mathbf{x})$, the ratio of the acceptance probabilities for the trajectory X and its conjugated $\tilde{X}$ is

$$\frac{A(\mathbf{x}_0 \rightarrow \mathbf{x}_n)}{A(\mathbf{x}_n \rightarrow \mathbf{x}_0)} = e^{-\beta[U(\mathbf{x}_n)-U(\mathbf{x}_0)]}\frac{P(\tilde{\Lambda}|\mathbf{x}_n)}{P(\Lambda|\mathbf{x}_0)}\prod_{i=1}^{n}\frac{\tilde{T}_i\tilde{S}_i}{T_iS_i} \tag{6}$$

Application of the Metropolis criterion,[4] allows to recover $A(\mathbf{x}_0 \rightarrow \mathbf{x}_n)$ as

$$A(\mathbf{x}_0 \rightarrow \mathbf{x}_n) = \min\left(1, e^{-\beta[U(\mathbf{x}_n)-U(\mathbf{x}_0)]}\frac{P(\tilde{\Lambda}|\mathbf{x}_n)}{P(\Lambda|\mathbf{x}_0)}\prod_{i=1}^{n}\frac{\tilde{T}_i\tilde{S}_i}{T_iS_i}\right) \tag{7}$$

Equation 7 represents the probability under which the final configuration $\mathbf{x}_n$, generated by the nonequilibrium process, will be accepted as starting configuration of the equilibrium phase of next iteration. In the case of rejection, the new iteration will start from the $\mathbf{x}_0$ configuration.

2.1.1. Perturbation Kernel. In general, perturbation steps consist of a driven evolution of the collective coordinate $\zeta(\mathbf{x})$, which depends on the atomic coordinates $\mathbf{x}$ of the system. Without loss of generality let suppose that, during a single perturbation step $T(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i^*)$, the collective coordinate undergoes an increase of $\Delta\zeta$.[20] The perturbation kernel $T(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i^*)$ can thus be written as

$$T(\mathbf{x}_{i-1} \rightarrow \mathbf{x}_i^*) = \delta(\zeta(\mathbf{x}_i^*) - \zeta(\mathbf{x}_{i-1}) - \Delta\zeta)J^{-1}(\mathbf{x}_i) \tag{8}$$

where $J(\mathbf{x}_i)$ is a normalization factor such that $\int T(\mathbf{x}_{i-1} \rightarrow \mathbf{x})\, d\mathbf{x} = 1$:

$$J(\mathbf{x}_i) = \int \delta(\zeta(\mathbf{x}) - \zeta(\mathbf{x}_{i-1}) - \Delta\zeta)d\mathbf{x} = \int \delta(\zeta(\mathbf{x}) - \zeta(\mathbf{x}_i))d\mathbf{x} \tag{9}$$

In the reverse direction, the perturbation kernel moves the collective coordinate backward. This allows to write $T^{-1}(\mathbf{x}_i^* \rightarrow \mathbf{x}_{i-1})$ as

$$T^{-1}(\mathbf{x}_i^* \rightarrow \mathbf{x}_{i-1}) = \delta(\zeta(\mathbf{x}_{i-1}) - \zeta(\mathbf{x}_i^*) + \Delta\zeta)J^{-1}(\mathbf{x}_{i-1}) \tag{10}$$

with

$$J(\mathbf{x}_{i-1}) = \int \delta(\zeta(\mathbf{x}) - \zeta(\mathbf{x}_{i-1}))d\mathbf{x} \tag{11}$$

In the special case of taking the distance d between two particles as collective coordinate, the normalization factor is

$4\pi d^2$, while it is 1 if the perturbation is applied along a fixed direction. Exploiting eqs 8, 9, and 10, we can recover the contribution of the perturbation kernels to the acceptance probability of eq 7:

$$\prod_{i=1}^{n} \frac{\tilde{T}_i}{T_i} = \frac{J(\mathbf{x}_n)}{J(\mathbf{x}_0)} \tag{12}$$

2.2. Monte Carlo Simulations with Configurational Freezing. Here, we give an overview of CF, remanding to refs 5 and 6 for a more exhaustive illustration of the method. The CF technique was originally devised to improve the efficiency of fast-switching MC simulations in computing free energy differences or potentials of mean force[21,22] via nonequilibrium work theorems.[7,9] The basic idea of CF is that in fast-switching simulations the system perturbation is confined about a limited space domain and that concentrating sampling in this domain may decrease dissipation and hence improve free energy calculations. In a broad sense, we can imagine such restricted region of space as a sort of reaction site and the fast-switching process as a localized (driven) chemical reaction. Upon the above well-founded assumption, we figure out that most molecules far from the reaction site (typically, but not necessarily only, the solvent molecules) may persist in a state of equilibrium and hence contribute marginally to dissipation mechanisms. This implies that the effect of the dynamics of these molecules on the performances of nonequilibrium methods is negligible. In spite of this, molecules at equilibrium affect the overall cost of the computation because, in any case, interatomic pair energies must be calculated to evolve the system. On such a basis, enhancing importance sampling in the reaction site of fast-switching MC simulations may decrease the dissipated work and, in the perspective of NCMC simulations, increase the acceptance probability of an attempted nonequilibrium move. This eventually leads to a faster diffusion along the collective coordinate employed in NCMC simulations with evident benefits.[3]

Operatively, CF works as in the following. A particle is chosen for a trial move only if its distance from the reaction site is smaller than an established value, whereas selection of the other particles is skipped. Once a particle is selected, only moves that leave the particle within the threshold distance from the reaction site are accepted with a probability that preserves the detailed balance condition, the other trial moves being rejected. With respect to conventional nonequilibrium MC simulations, the CF algorithm enriches the sampling of configurational states in space regions with high dissipation. However, as shown in ref 5, limiting the sampling to a restricted phase-space region does not introduce approximations in the method. Actually, the tunability of CF in terms of choice of the sampled phase space is used to optimize the calculation by reducing dissipation.

As stated above, CF is based on trial moves of particles localized into a spatial region encompassing the reaction site. Therefore, the geometrical definition of the reaction site is a key point of the method. The simplest approach is to define the reaction site as the union of spheres (called mobility spheres) whose centers lie on one or more "strategic" atoms of the system, denoted as *reactive centers* in the following. Even an immaterial point, such as the center of mass of a molecule, could be chosen as reactive center. Once the criterion for defining the mobility region is established, the probability of selecting a particle for a trial move inside this region is equal to N_{in}^{-1}, where N_{in} is the number of particles inside the mobility region. This was the first implementation of CF presented in ref 5. In a later CF-based approach,[6] a particle selection protocol with nonuniform probability was implemented together with the original CF algorithm in a unified scheme. In this improved scheme, the probability of selecting a particle (inside the mobility region) for a trial move is dependent on the distance between the particle and the reactive centers: particles closer to reactive centers are selected with higher probability and vice versa.

In CF-based as well as in conventional nonequilibrium simulations, the fast-switching process is realized alternating perturbation and propagation kernels. In CF, propagation is performed according to the protocol explained above.

2.3. Propagation Kernel under Configurational Freezing Schemes. In the nonequilibrium phase of a NCMC simulation, each perturbation step is alternated with a propagation step during which the collective coordinate $\zeta(\mathbf{x})$ is constrained to a given value and the degrees of freedom not entering perturbation kernel move toward a relaxed configuration by dissipating heat.

In the conventional approach, the propagation step typically consists of one or more attempted moves, each applied to one randomly selected particle by keeping fixed the remaining $N - 1$ particles. If the i-th propagation step is made of a single-particle move involving the k-th particle, it can be schematically represented as

$$\mathbf{x}_i^* \xrightarrow{S_i} \mathbf{x}_i = (\mathbf{r}_{i,1}^*, ..., \mathbf{r}_{i,k}^*, ..., \mathbf{r}_{i,N}^*) \xrightarrow{S_i} (\mathbf{r}_{i,1}^*, ...\mathbf{r}_{i,k}, ..., \mathbf{r}_{i,N}^*) \tag{13}$$

where $\mathbf{r}_{i,k}^*$ and $\mathbf{r}_{i,k}$ denote the positions of the k particle before and after the attempted move, respectively. From now on, the notations $\mathbf{x}_i^*$ and $\mathbf{x}_i$ will denote the configurations of eq 13. As explained in section 2.2, the underlying idea of CF is that the probability of selecting the k particle depends on its distance from established reactive centers. Only particles located inside a mobility region about the reactive centers can be chosen with nonzero probability for attempting a propagation move. The volume including such mobility region, denoted with $\mathcal{D}$, can be defined in various ways.[5] According to a self-adapting criterion, $\mathcal{D}$ may correspond to the union of fixed-radius spheres centered on reactive centers, chosen to follow the configurational changes induced by the perturbation steps. With such a definition, the mobility region can change in shape and size during the process, because of the evolution of reactive centers. Large values of the sphere-radius may guarantee to effectively encompass the dissipation domain. However, a very large radius will include particles unaffected by the external perturbation, thus making the method more ineffective. Due to the mobility of the reactive centers, the list of mobile and frozen particles (inside and outside the mobility region, respectively) must be periodically updated during the nonequilibrium process. Once the mobility region is established through the choice of the reactive centers and their associated radii, the probability of selecting a generic k particle varies in agreement with a normalized function of the atomic coordinates, $\pi(\mathbf{x}_i^*)$, generally decreasing with increasing the distance between the k particle located in $\mathbf{r}_{i,k}^*$ and the reactive centers. Typically, $\pi(\mathbf{x}_i^*)$ assumes the form

$$\pi(\mathbf{x}_i^*) = \begin{cases} d^{-\nu}(\mathbf{r}_{i,k}^*)\left(\sum_{\mathbf{r}_{i,m}^* \in \mathcal{D}} d^{-\nu}(\mathbf{r}_{i,m}^*)\right)^{-1} & \mathbf{r}_{i,k}^* \in \mathcal{D} \\ 0 & \mathbf{r}_{i,k}^* \notin \mathcal{D} \end{cases} \quad (14)$$

where $d(\mathbf{r}_{i,k}^*)$ is the distance of the particle k from the nearest reactive center in the configuration $\mathbf{x}_i^*$ and ν is a non-negative real number (note that the probability $\pi(\mathbf{x}_i^*)$ depends on the positions of all the particles inside the mobility region at step i). By choosing $\nu = 0$, a *uniform particle selection scheme*[5] is enforced, according to which the probability of selecting a particle within the mobility region is constant. By choosing $\nu > 0$, we instead apply the *nonuniform particle selection scheme*,[6] in which the above probability is modulated by the distances $d(\mathbf{r}_{i,k}^*)$.

The propagation kernel $S(\mathbf{x}_i^* \rightarrow \mathbf{x}_i)$ related to an attempted move involving the particle k, corresponds to the product between the probability $\pi(\mathbf{x}_i^*)$ of selecting the k particle at the position $\mathbf{r}_{i,k}^*$ of the configuration $\mathbf{x}_i^*$, the probability $w(\mathbf{x}_i|\mathbf{x}_i^*)$ to assign the final position $\mathbf{r}_{i,k}$ to the k particle once it has been selected at position $\mathbf{r}_{i,k}^*$ from configuration $\mathbf{x}_i^*$, and the probability $\alpha(\mathbf{x}_i^* \rightarrow \mathbf{x}_i)$ to accept the produced move:

$$S(\mathbf{x}_i^* \rightarrow \mathbf{x}_i) = \pi(\mathbf{x}_i^*) w(\mathbf{x}_i|\mathbf{x}_i^*) \alpha(\mathbf{x}_i^* \rightarrow \mathbf{x}_i) \quad (15)$$

Analogously, we can write

$$S(\mathbf{x}_i \rightarrow \mathbf{x}_i^*) = \pi(\mathbf{x}_i) w(\mathbf{x}_i^*|\mathbf{x}_i) \alpha(\mathbf{x}_i \rightarrow \mathbf{x}_i^*) \quad (16)$$

The probabilities $w(\mathbf{x}_i|\mathbf{x}_i^*)$ and $w(\mathbf{x}_i^*|\mathbf{x}_i)$ are equal for attempted moves keeping the particle k inside fixed-radius spheres centered in $\mathbf{r}_{i,k}^*$ and $\mathbf{r}_{i,k}$, respectively (indeed, this is the usual way of realizing single-particle moves in MC simulations). As no particles external to the mobility region can be chosen for a MC move, detailed balance condition enforces null probability of accepting MC moves leading a selected particle outside the mobility region, namely

$$\alpha(\mathbf{x}_i^* \rightarrow \mathbf{x}_i) = 0, \quad \text{if} \quad \mathbf{r}_{i,k}^* \in \mathcal{D} \quad \text{and} \quad \mathbf{r}_{i,k} \notin \mathcal{D} \quad (17)$$

All in all, only attempted MC moves leaving a particle, selected according to eq 14, inside the mobility region have a nonzero probability of being accepted. For these moves, detailed balance condition is applied as

$$\rho(\mathbf{x}_i^*) S(\mathbf{x}_i^* \rightarrow \mathbf{x}_i) = \rho(\mathbf{x}_i) S(\mathbf{x}_i \rightarrow \mathbf{x}_i^*) \quad (18)$$

Exploiting eqs 15 and 16 into eq 18 and applying the Metropolis criterion, we obtain the acceptance probability for an attempted propagation move

$$\alpha(\mathbf{x}_i^* \rightarrow \mathbf{x}_i) = \min\left(1, \frac{\pi(\mathbf{x}_i)}{\pi(\mathbf{x}_i^*)} e^{\beta[U(\mathbf{x}_i^*) - U(\mathbf{x}_i)]}\right) \quad \mathbf{r}_{i,k}^* \in \mathcal{D} \quad \text{and} \quad \mathbf{r}_{i,k} \in \mathcal{D} \quad (19)$$

Equations 17 and 19 are the acceptance probabilities enforced in propagation steps of the nonequilibrium NCMC phase. Equation 18 allows us to recover an expression for the contribution of propagation kernels to the acceptance probability of candidate configurations (eq 7):

$$\prod_{i=1}^{n} \frac{\tilde{S}_i}{S_i} = \prod_{i=1}^{n} \frac{\rho(\mathbf{x}_i^*)}{\rho(\mathbf{x}_i)} = \prod_{i=1}^{n} e^{\beta[U(\mathbf{x}_i) - U(\mathbf{x}_i^*)]} = e^{\beta Q} \quad (20)$$

where

$$Q = \sum_{i=1}^{n} [U(\mathbf{x}_i) - U(\mathbf{x}_i^*)] \quad (21)$$

is the heat absorbed by the system along the nonequilibrium trajectory X. The products in eq 20 are over all the propagation steps, each of which may involve a different particle. Recognizing the work made on the system along the trajectory X to be

$$W = \sum_{i=1}^{n} [U(\mathbf{x}_i^*) - U(\mathbf{x}_{i-1})] = U(\mathbf{x}_n) - U(\mathbf{x}_0) - Q \quad (22)$$

the acceptance probability of candidate configurations (eq 7) can be rewritten as follows

$$A(\mathbf{x}_0 \rightarrow \mathbf{x}_n) = \min\left(1, e^{-\beta W} \frac{P(\tilde{\Lambda}|\mathbf{x}_n)}{P(\Lambda|\mathbf{x}_0)} \frac{J(\mathbf{x}_n)}{J(\mathbf{x}_0)}\right) \quad (23)$$

where eqs 12 and 20 have been used. It is worth noting that the probabilities of selecting the protocols Λ and $\tilde{\Lambda}$ are not imposed by the physics of the system but can be freely specified to optimize the efficiency of the method. The most trivial approach is to impose protocols Λ and $\tilde{\Lambda}$ so that $P(\Lambda|X_0) = P(\tilde{\Lambda}|X_n) = 1$ and eq 23 simplifies.

Equation 23 highlights the dependence of the acceptance probability of a candidate configuration from the work performed on the system during the NCMC phase. In this respect, the ability shown by CF of decreasing the dissipative contribution to the work, and hence the total work W, in fast-switching processes is expected to enhance the acceptance probability of candidate configurations.

3. SYSTEM AND SIMULATION METHODS

Following Nilmeier and co-workers,[3] CF-NCMC has been tested on a bistable dimer immersed in a dense fluid. The dimer consists of a pair of particles that interact through a double-well potential whose expression is

$$U_{\text{bond}}(r) = \frac{16h}{r_0^4}(r - r_0)^2(r - 2r_0)^2 \quad (24)$$

$U_{\text{bond}}(r)$ has two minima at $r = r_0$ and $r = 2r_0$, corresponding to compact and extended configurations of the dimer, respectively. Such minima are separated by a barrier of height h located at the distance $r = 3/2\ r_0$. Dimer particles also interact via a Weeks−Chandler−Andersen (WCA) soft repulsive potential[23] (this is a variant with respect to the model of ref 3),

$$U_{\text{wca}}(r) = \begin{cases} 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] + \varepsilon & r \le r_{\text{wca}} \\ 0 & r > r_{\text{wca}} \end{cases} \quad (25)$$

where $r_{\text{wca}} = r_0 = 2^{1/6}\sigma$, with $\sigma = 3.4$ Å and $\varepsilon/k_B = 120$ K. The solvent is made of particles interacting each other and with the dimer particles via the nonbonded WCA interaction of eq 25. Three samples formed by 216, 1000, and 2000 particles, including dimer, have been simulated at constant volume and temperature, under cubic periodic boundary conditions. Reduced density corresponds to $\rho\sigma^3 = 0.96$ and the reduced temperature is $k_BT/\varepsilon = 0.824$. Simulations with energy barrier h of $5k_BT$ and $10k_BT$ have been carried out. The energy barrier and the number of particles will be specified case by case in

dx.doi.org/10.1021/ct500340b | J. Chem. Theory Comput. 2014, 10, 4273−4283

section 4. Three types of simulations have been carried out: a conventional MC (cnv-MC) simulation, standard NCMC (std-NCMC) simulations as formulated by Nilmeier and co-workers[3] and NCMC simulations supplied with CF (CF-NCMC simulations).

std-NCMC and CF-NCMC simulations have been performed enforcing 5000 iterations, each consisting of an equilibrium MC simulation phase of 10^5 single-particle MC steps[24] followed by a NCMC move (with or without CF). The NCMC move, aimed at generating a candidate configuration for the next iteration, is produced by forcing the dimer particles to increase or decrease their distance r by Δr along a fixed direction in an established number n of single-particle MC steps. Such a perturbation is enforced according to the following scheme

$$\Delta r = \begin{cases} 0, & r < \frac{1}{2} r_0 \\ +r_0, & \frac{1}{2} r_0 \le r \le \frac{3}{2} r_0 \\ -r_0, & \frac{3}{2} r_0 < r \le \frac{5}{2} r_0 \\ 0, & r > \frac{5}{2} r_0 \end{cases} \tag{26}$$

Equation 26 is the protocol of transition between compact and extended dimer configurations (from now on denoted r-transition protocol), which depends on the dimer extension r in the final configuration of the equilibrium phase. It is easy to realize that the above r-transition protocol promotes the barrier crossing/recrossing for the dimer system, as the energy barrier falls at $^3/_2 r_0$. We point out that a NCMC move is skipped when $\Delta r = 0$ according to eq 26. Each NCMC move is performed with a fixed number n of single-particle MC steps. The distance r can in principle be varied deterministically by $\Delta r/n$ at every MC step, to reach the value $r + \Delta r$ at the end of the NCMC move. This scheme is however quite inefficient in fast-switching moves, because it requires updating dimer-solvent energy contributions upon a small change of the dimer position (i.e., $\Delta r/n$). Therefore, we have chosen to vary r every 100 MC steps by a quantity of $100\Delta r/n$. NCMC simulations with various lengths of the nonequilibrium phase, namely with different numbers of MC steps n, have been performed: $n = a \times 10^5$, with $a = 2,3,4,5$. We will specify n when discussing the results. In all MC simulations, the single-particle trial MC moves are enforced by keeping the new particle position at random inside a cube of side-length 0.6 Å, centered at the current particle position.

In CF-NCMC simulations, NCMC moves (nonequilibrium phase) have been made according to the uniform particle selection scheme,[5] namely by setting $\nu = 0$ in eq 14. This choice is dictated by the consideration that this scheme, though less performing than the nonuniform particle selection counterpart,[6] is simpler to be implemented in a MC simulation code. The mobility region has been defined as the union of the spheres centered on the dimer particles, both with radius 10 Å. Only in this space domain solvent particles can be selected for a trial MC move during the nonequilibrium phase of the iterations.

As stated above, the driven transition of r by Δr occurs along a fixed direction. This eliminates the Jacobian contribution from the acceptance probability of eq 23, as $J(\mathbf{x}_0) = J(\mathbf{x}_n)$. Moreover, considering eq 26, we note that the choice of an extension or contraction driven transition is uniquely determined by the current positions of the dimer particles and in particular by their distance. Extension is attempted when $^1/_2 r_0 \le r \le {}^3/_2 r_0$, while contraction is attempted when $^3/_2 r_0 \le r \le {}^5/_2 r_0$. No transition is attempted when the r distance falls out of such intervals. This implies that the conditional probabilities $P(\Lambda|\mathbf{x}_0)$ and $P(\tilde{\Lambda}|\mathbf{x}_n)$ of generating the two protocols are equal, because both protocols are generated when r is found inside intervals of equal size, namely r_0. Thus, in our model, the Metropolis criterion applied to a NCMC move (eq 23) can be simplified as

$$A(\mathbf{x}_0 \to \mathbf{x}_n) = \min(1, e^{-\beta W}) \tag{27}$$

The acceptance probability of a NCMC move depends only on the work W performed on the system along the nonequilibrium path, which is computed via eq 22.

The capability of cnv-MC, std-NCMC, and CF-NCMC simulations of generating a correct configurational sampling has been evaluated through estimates of the probability distribution of the dimer extension r. The reference probability distribution has been computed very accurately by means of an umbrella sampling simulation.[25] Such a simulation adopts a modified bonded potential acting between the dimer particles, designed to remove the energy barrier between compact and extended configurations:

$$U_{us}(r) = k_B T\{[a(r - r_0)^2 - \ln(r/r')]\Theta(r_0 - r) - \ln(r/r')\Theta(r - r_0)\Theta(r' - r) + a(r - r')^2\Theta(r - r')\} \tag{28}$$

where $r' = 2.05\ r_0$, $a = 5.556$ Å^{-2}, and $\Theta(x)$ is the Heaviside step function. An umbrella sampling simulation with 216 particles, including dimer, lasting 10^{10} MC steps has been carried out.

4. ILLUSTRATIVE CALCULATIONS

In this section, we report the outcomes of cnv-MC, std-NCMC, and CF-NCMC simulations to illustrate the relative performances of the three methods. However, as the relative performances of cnv-MC and std-NCMC have been thoroughly discussed in ref 3 and we are interested essentially in comparing CF-NCMC with std-NCMC, we will focus the comparative analysis on these latter simulation schemes. Thus, in order to simplify the discussion, a limited amount of data will be reported for the cnv-MC simulation. The complete set of data is however available upon request to the corresponding author.

The main difference between standard NCMC[3] (std-NCMC) and its CF variant (CF-NCMC) lies in the way of realizing the propagation steps of finite-time switching moves. By adopting CF, it is possible to enhance sampling around the reactive site, generating less dissipative switching trajectories and eventually greater probabilities of accepting candidate configurations (eq 23). This improves sampling along the chosen collective coordinate, which is the target of NCMC. In this respect, the benefits of using NCMC with respect to conventional simulations and of combining CF with NCMC schemes are well represented in Figure 2, where we compare the time evolution of the dimer extension resulting from std-NCMC, CF-NCMC, and cnv-MC simulations. Calculations have been performed on a sample of 216 WCA particles, in which the dimer particles interact through eqs 25 and 24 supplied with a $10k_BT$ energy barrier. In both NCMC simulations, each iteration consists of an equilibrium phase of

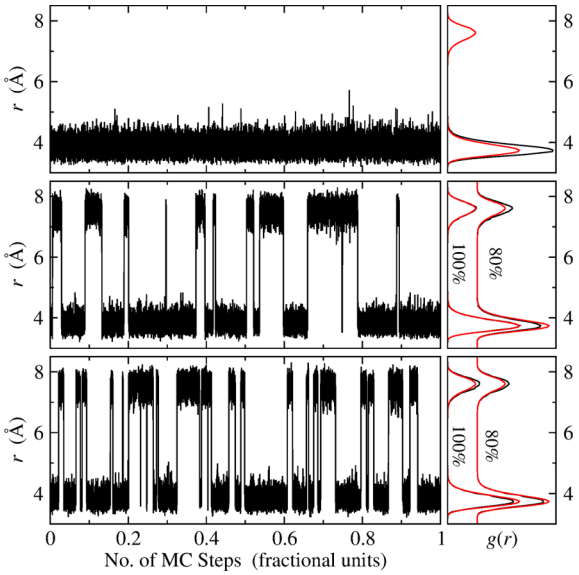


Figure 2. Left panels: evolution of the dimer extension r as a function of the number of MC steps (in fractional units) obtained from cnv-MC, std-NCMC, and CF-NCMC simulations (top, middle and bottom panels, respectively) of a 216-particle sample supplied with a double-well potential having $10k_BT$ energy barrier. Right panels: black lines are estimates of the probability distribution $g(r)$ of the dimer extension computed from cnv-MC, std-NCMC, and CF-NCMC simulations (top, middle, and bottom panels, respectively). Red lines correspond to the reference $g_{us}(r)$ obtained from the umbrella sampling simulation. For std-NCMC and CF-NCMC, the curves labeled with 80% represent the $g(r)$ computed removing the last 20% of simulated trajectory. The curves labeled with 100% represent the $g(r)$ computed with the whole trajectories.

10^5 MC steps and a switching, or NCMC, phase of 2×10^5 MC steps, so that the total number of MC steps is 1.5×10^9 (5000 iterations per simulation). To get comparable CPU elapsed time, the cnv-MC simulation has been carried out with the same number of MC steps. Estimations of the $g(r)$ function (i.e., the probability distribution of the dimer extension r) are also reported in Figure 2 and compared with the exact distribution recovered from an umbrella sampling simulation (see section 3 for details). In the cnv-MC simulation (Figure 2, top left panel), we do not observe any hopping event, or r-transition, due to the high energy barrier compared to thermal energy. In spite of the large energy barrier, numerous r-transitions are instead observed in std-NCMC (Figure 2, middle left panel) and CF-NCMC (Figure 2, bottom left panel) simulations, clearly induced by NCMC moves. Considering the large difference between NCMC and conventional methods in sampling the r space, the much better performances of the former technique in reproducing the $g(r)$ function are not surprising. In our context, it is important to note the higher hopping rate generated by CF-NCMC with respect to std-NCMC: 37 against 22 r-transitions, respectively. Nevertheless, the enhanced hopping rate obtained through CF-NCMC does not lead to a more accurate $g(r)$ with respect to std-NCMC. This can be ascribed to the fact that 22 r-transitions for std-NCMC are already sufficient to get satisfactory estimates of $g(r)$. We stress however that, reducing the number of iterations by only 10%, std-NCMC and CF-NCMC give practically congruent $g(r)$s, while reducing the number of iterations by 20% makes the CF-NCMC estimate more accurate than the std-NCMC counterpart (see right panels of Figure 2). In fact, the relative errors with respect to the umbrella sampling distribution $g_{us}(r)$, that is, $[g(r)-g_{us}(r)]/g_{us}(r)$, related to the peaks at r = 3.75 Å and r = 7.6 Å, are −0.11 and 0.26 for std-NCMC and −0.08 and 0.14 for CF-NCMC.

When the dimer double-well potential presents a lower energy barrier, the sampling powers of NCMC and cnv-MC methods become comparable. This can be seen in the Supporting Information figure, where we report the outcomes of simulations performed by enforcing a double-well potential with $5k_BT$ energy barrier. However, also using a flatter dimer potential, differences between the three methods may still emerge if the dimer is immersed in a large amount of solvent. This can be appreciated in Figure 3, where we draw the time

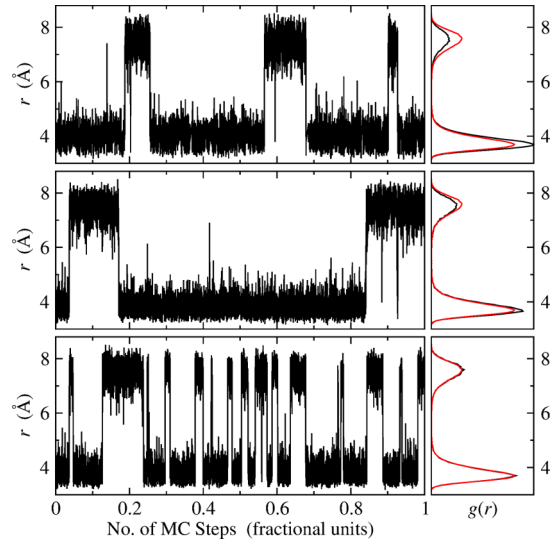


Figure 3. Left panels: evolution of the dimer extension r as a function of the number of MC steps (in fractional units) obtained from cnv-MC, std-NCMC, and CF-NCMC simulations (top, middle and bottom panels, respectively) of a 1000-particle sample supplied with a double-well potential having $5k_BT$ energy barrier. Right panels: black lines are estimates of the probability distribution $g(r)$ of the dimer extension computed from cnv-MC, std-NCMC, and CF-NCMC simulations (top, middle, and bottom panels, respectively). Red lines correspond to the reference $g_{us}(r)$ obtained from the umbrella sampling simulation.

evolutions of the dimer extension and the related $g(r)$ functions obtained from std-NCMC, CF-NCMC, and cnv-MC simulations of a 1000-particle sample with $5k_BT$ energy barrier. In such a case, the hopping rates gained from std-NCMC and cnv-MC appear much more ineffective than that obtained from CF-NCMC. As a matter of fact, this ineffectiveness does not seem to affect significantly the $g(r)$ function, which is quite well reproduced, especially by the std-NCMC simulation. However, the behavior of the time evolution of r recovered from the std-NCMC simulation (Figure 3, middle left panel) points to a quite accidental agreement. The good $g(r)$ estimate provided by CF-NCMC is instead well supported by the high number of r-transitions (Figure 3, bottom left panel). We also notice that, among the three r-transitions observed in the std-NCMC simulation, none arises from accepting candidate configurations, all hopping events being spontaneous. As we will see below, this situation does not change substantially if we enhance the acceptance probability of NCMC moves by making them slower, that is, by increasing the number of MC steps in the nonequilibrium phase from 2×10^5 to 5×10^5.

dx.doi.org/10.1021/ct500340b | J. Chem. Theory Comput. 2014, 10, 4273−4283

This means that, in the sample of 1000 particles, std-NCMC scheme cannot improve significantly the sampling power of the simulation. The aspects discussed above, though quite irrelevant for the present model system, may be of paramount importance in large-size systems of biochemical or biophysical interest.

The reason why hopping rates are comparable in std-NCMC and cnv-MC simulations of the 1000-particle sample (Figure 3) is clearly ascribed to low acceptance probabilities of NCMC moves in the former simulation, which are, in turn, correlated to poor sampling around the dissipation domain (located around the reactive site). In fact, most of MC moves performed during the nonequilibrium phase are employed to sample portions of system far away from the dimer and hence irrelevant for the dissipation mechanism. This evidently leads to an increase of the work and therefore to a dramatic decrease of the work-dependent acceptance probabilities. As stated above, in the std-NCMC simulation realized to get the data of Figure 3, none of the 5000 NCMC moves has been able to produce a r-transition. This implies that the std-NCMC simulation virtually corresponds to a cnv-MC simulation with a "ghost" phase (2×10^5 MC steps per iteration) in which the system does not evolve. It is thus evident that, in situations of negligible hopping rate, using std-NCMC is largely unfavorable. Even the strategy of increasing the acceptance probabilities by producing much slower NCMC moves could reveal ineffective because of the excessive computational cost. As a matter of fact, in view of applying std-NCMC to large systems, these results are not encouraging at all.

The hopping rate is instead enormously enhanced by adopting CF-NCMC, also in many-particle systems, as we can infer from Figure 3, and especially from Figure 4, where we report the total number of r-transitions and the number of accepted NCMC moves (induced r-transitions) for a series of std-NCMC and CF-NCMC simulations differing in number of particles and in nonequilibrium-phase length. The numbers of r-transitions obtained from cnv-MC simulations of 1.5×10^9 MC steps (1000 and 2000 particle samples) are also reported for comparison. The improvement obtained by CF-NCMC with respect to std-NCMC and cnv-MC ranges from one to 2 orders of magnitude, in dependence of the nonequilibrium-phase length and of the number of particles. In Figure 4, we note that, for CF-NCMC simulations, the number of accepted NCMC moves is almost systematically greater than the total number of r-transitions. This depends on how a r-transition is defined. Indeed, while the number of accepted NCMC moves is a well-defined quantity within the NCMC protocol, some arbitrariness is introduced in defining "r-transition". In our case, a transition is counted when the dimer extension r crosses the position of the energy barrier ($r = {}^3/_2 r_0$; see eq 24) and the system remains in the new configuration for an established number of MC steps (see Supporting Information for a rigorous definition). Therefore, in the total number of r-transitions we do not include fast-rate recrossing of the energy barrier. In CF-NCMC, the increase of hopping rate with the number of MC steps per NCMC move is due to the higher reversibility of the NCMC moves which makes lower the dissipated work and consequently greater the acceptance probabilities (eq 27). In this respect, it is worth noting that, in the std-NCMC simulation of the 1000-particle sample, the length of the NCMC phase has a marginal effect on the hopping rate, while it is almost ineffective in the 2000-particle sample. This definitely points to a low efficiency of std-NCMC in these systems. A further important aspect to be remarked is the role of the number of solvent particles, which does not alter the behavior of CF-NCMC. This unexpected behavior becomes clearer considering that the size of the reactive domain in CF-NCMC does not depend on the number of solvent particles, because, during the NCMC moves, the propagation steps always occur in an almost constant volume around the dimer particles (this volume corresponds to the union of spheres with fixed radius centered on the dimer particles).

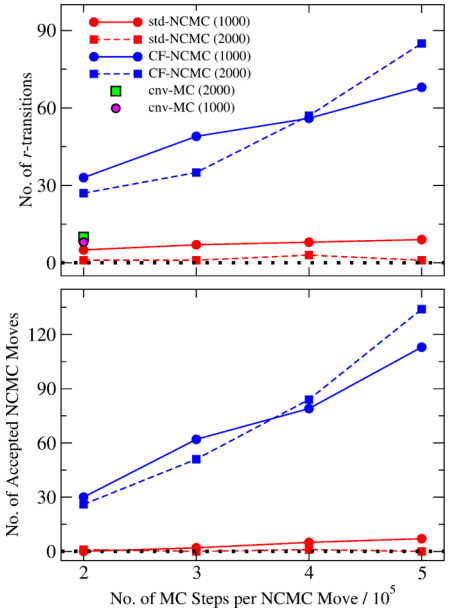


Figure 4. Number of r-transitions (top panel) and number of accepted NCMC moves (bottom panel) as a function of the number of MC steps per NCMC move, obtained from std-NCMC and CF-NCMC simulations of 1000 and 2000-particle samples (see the legend). Data for cnv-MC simulations are also reported. The total number of MC steps of cnv-MC simulations corresponds to that of NCMC simulations realized with 2×10^5 MC steps per NCMC move. The black dots represent the baseline. Lines are guides for the eyes.

Therefore, combining noninstantaneous MC dimer extension/contraction moves, which modify the dimer extension by $\Delta r = \pm r_0$, with CF schemes allows for a thorough sampling of the collective coordinate. The results on the time series of the dimer extension (Figures 2 and 3) already give a clear picture on the possible advantages obtainable by implementing CF in NCMC schemes. It is evident that changes in hopping rate may significantly affect the time needed to reach convergence in sampling the populations of compact and extended dimer configurations and, more generally, to get an accurate evaluation of the $g(r)$ probability distribution. This has been partially seen in Figure 3. A more detailed information on the $g(r)$ estimates obtained from various simulation setups for 1000 and 2000-particle systems is provided in Figures 5 and 6, respectively. In both systems, CF-NCMC outperforms systematically std-NCMC and cnv-MC. This is also quantified in Figure 7, where we report the root-mean-square deviations of the NCMC $g(r)$ estimates from the $g_{us}(r)$ distribution (by umbrella sampling simulation), averaged over the whole r domain. The performances of CF-NCMC relative to std-NCMC and cnv-MC are enhanced in larger samples, basically due to the nonlocal character of sampling in the latter methods. Such a nonlocality leads to greater dissipation because the

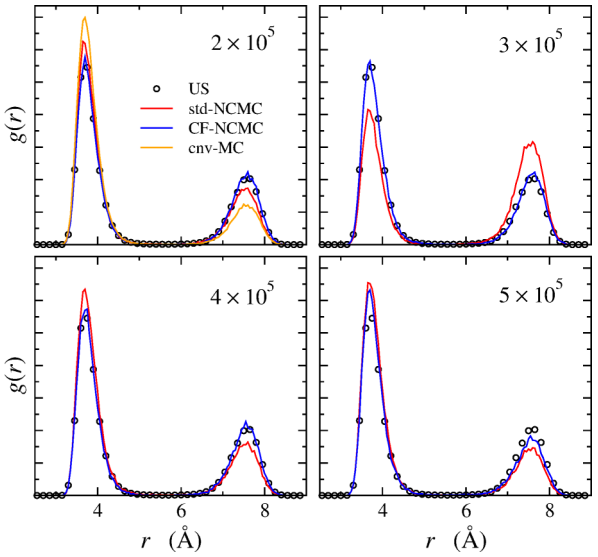


Figure 5. Estimates of the probability distribution $g(r)$ as a function of the dimer extension r (Å) in the 1000-particle system. Estimates obtained from CF-NCMC (blue lines), std-NCMC (red lines), and cnv-MC (orange lines) simulations are compared with the reference distribution $g_{us}(r)$ (US; open circles) obtained from the umbrella sampling simulation (see section 3 for details). Panels refer to data obtained from simulation settings differing in the number of MC steps per NCMC move (see numbers in the panels).

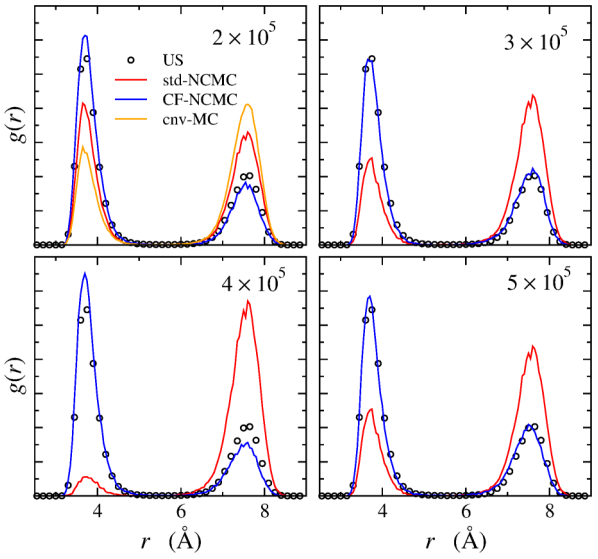


Figure 6. Estimates of the probability distribution $g(r)$ as a function of the dimer extension r (Å) in the 2000-particle system. Estimates obtained from CF-NCMC (blue lines), std-NCMC (red lines), and cnv-MC (orange lines) simulations are compared with the reference distribution $g_{us}(r)$ (US; open circles) obtained from the umbrella sampling simulation (see section 3 for details). Panels refer to data obtained from simulation settings differing in the number of MC steps per NCMC move (see numbers in the panels).

system is not allowed to relax fastly. In std-NCMC, the global effect of slow relaxation is to decrease the acceptance probability of NCMC moves thus making the diffusion of the system through the r subspace slower. In the case of cnv-MC, instead, the r-transitions are simply reduced for poorer sampling around the reactive site.

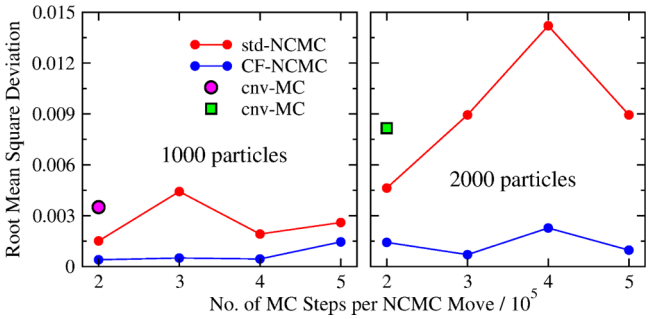


Figure 7. Root-mean-square deviations of CF-NCMC, std-NCMC, and cnv-MC probability distributions $g(r)$ from the reference $g_{us}(r)$, obtained from the umbrella sampling simulation. The root-mean-square deviations are reported as a function of the number of MC steps per NCMC move. The total number of MC steps of cnv-MC simulations corresponds to that of NCMC simulations realized with 2 × 10^5 MC steps per NCMC move. Data are reported for 1000-particle and 2000-particle systems (left and right panels, respectively). Lines are guides for the eyes.

Contrarily to expectations, the root-mean-square deviations of Figure 7 do not show any clear improvement of the $g(r)$ estimations when decreasing the velocity of NCMC moves. This means that still fast NCMC moves may allow for accurate evaluations. This is an extremely positive aspect, because it suggests that even fast NCMC moves can provide effective acceptance probabilities. However, we must acknowledge that this may not be a truly general behavior and could in principle depend on the nature of the system under consideration and on the definition of the collective coordinate.

The previous analysis provides a faithful picture of the accuracy of CF-NCMC and std-NCMC in recovering $g(r)$, which is the primary quantity to evaluate the sampling power of NCMC simulations. A supplemental comparative analysis relies on the precision of the methods that has been computed by exploiting the block average procedure. In particular, an average $g(r)$ has been determined using 20 independent estimates, each being computed from blocks of dimer configurations taken sequentially from the simulation (each block is formed by configurations uniformly sampled in an interval of 250 iterations). The statistical error is thus defined as a function of r and corresponds to the standard deviation of the 20 estimates from the average. In Figure 8, we report the relative error of std-NCMC and CF-NCMC simulations, which corresponds to the ratio between the standard deviation and the average $g(r)$ defined above. Different panels refer to simulations of a 1000-particle sample with NCMC phases of variable length. Data for the 2000-particle sample are not reported because very similar trends have been obtained. As a general behavior of the error curves, we observe large values in the central and side regions of the r domain, where the $g(r)$ function nearly vanishes. Therefore, these regions are of little significance to our aim. The most interesting regions are those corresponding to high values of $g(r)$, associated with the compact and extended configurations of the dimer. By focusing on such regions, we note the systematic higher precision of CF-NCMC with respect to std-NCMC. This allows us to conclude that CF-NCMC outperforms std-NCMC from both accuracy and precision points of view.

If on one side, combining NCMC with CF may lead to significant improvement of sampling along the collective coordinate, on the other side an additional CPU cost of CF-

dx.doi.org/10.1021/ct500340b | J. Chem. Theory Comput. 2014, 10, 4273−4283

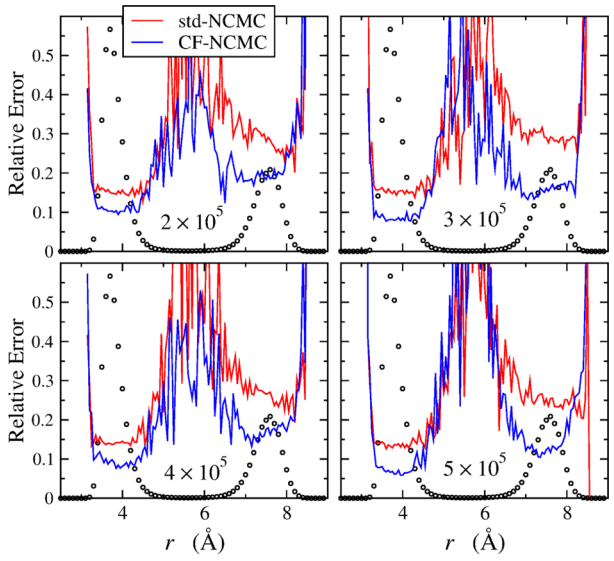


Figure 8. Relative error of the probability distribution $g(r)$ (see text for definition) as a function of the dimer extension r (Å) obtained from CF-NCMC and std-NCMC simulations of the 1000-particle system. Panels refer to data obtained from simulation settings differing in the number of MC steps per NCMC move (see numbers in the panels). The probability distribution $g_{us}(r)$, computed from the umbrella sampling simulation, is displayed to highlight the relevant r ranges (open circles).

NCMC with respect to std-NCMC and cnv-MC is expected. Such a cost is introduced by periodical updates of the list of mobile particles, namely the particles belonging to the mobility region. In fact, as the reactive centers are moved upon application of a perturbation step, the mobility region changes in shape and size.[5] Therefore, the list of mobile particles must be updated by checking the distances between the reactive centers and the other particles. This procedure is particularly expensive in CF-NCMC simulations, especially if a perturbation step is alternated with a single propagation step during a NCMC move. For this reason, it is generally convenient to alternate a perturbation step with a series of propagation steps, realized consecutively without updating the list of mobile particles. In our CF-NCMC simulations a perturbation step is alternated with 100 propagation steps. Another computationally expensive task, but of secondary importance with respect to the previous one, occurs during each propagation step of the NCMC phase. In such a case, we need to establish if the particle has been moved outside the mobility region upon application of the propagation step. This requires calculation of the distance between the particle and the reactive centers. To quantify the additional CPU time of CF-NCMC simulations, we have compared the CPU elapsed times of std-NCMC and CF-NCMC simulations (of a 1000-particle sample) taking a cnv-MC simulation as reference. For our simulation setting, namely enforcing iterations made of an equilibrium phase of 10^5 MC steps and a NCMC phase of 2×10^5 MC steps, the additional CPU time required by CF-NCMC is quite low. In fact, the ratio between the CPU elapsed times of CF-NCMC and cnv-MC simulations is only 1.159. It is remarkable that such a time is almost identical to that of the std-NCMC simulation (the ratio between std-NCMC and cnv-MC elapsed times is 1.156). This is ascribed to the low rate with which a perturbation step is applied in NCMC simulations. In fact, by increasing the rate of perturbation steps from 1 every 100 propagation steps (as in our simulations) to 1 every 10 propagation steps, the ratio related to CF-NCMC increases to 1.298. By enforcing a rate of 1 perturbation step every 1 propagation step, the ratio increases further, becoming 2.345. In any case, our outcomes suggest that also using a small number of perturbation steps, namely a low perturbation rate, may allow to obtain highly performing CF-NCMC simulations.

5. CONCLUDING REMARKS

We have presented a Monte Carlo simulation scheme in which candidate configurations of a system are generated through a finite-time process that actively drives the system out of equilibrium. The nonequilibrium switching moves are applied to established degrees of freedom, or collective configurational coordinates, whose sampling we are interested in. Our approach arises from integrating the nonequilibrium candidate Monte Carlo (NCMC) method by Nilmeier and co-workers[3] to the configurational freezing technique developed by some of us.[5,6] The capability of configurational freezing in enhancing the system relaxation around the domain where the nonequilibrium process is active, the so-called reactive site, allows to decrease the dissipated work produced during the fast switching trajectories and ultimately to increase the acceptance probabilities of the finite-time nonequilibrium moves. This leads to a significant improvement of sampling of the target collective coordinate, which could be extremely poor via conventional MC or simple NCMC simulations. In spite of the large efficiency of the configurational freezing variant of NCMC proposed here, we also recognize possible drawbacks of the general NCMC simulation protocol. In particular, NCMC may have severe sampling problems when the potential of mean force (PMF) along the chosen collective coordinate presents very large variations, especially between regions of the collective coordinate among which switching occurs. This is schematically represented in Figure 9, left panel. The shadow areas represent

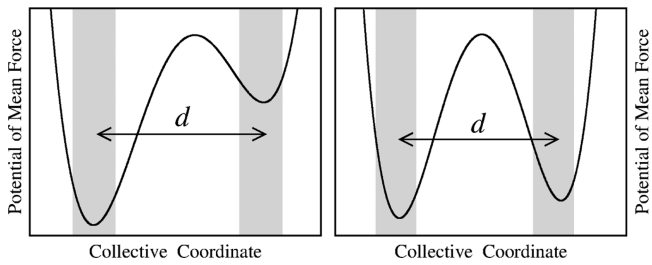


Figure 9. Left and right panels show the potential of mean force as a function of the collective coordinate for two systems. Shadow areas represent the domains where switching of the collective coordinate can be activated (by a quantity d in the right or left direction). NCMC method is well suited for systems in which wells of the potential of mean force have similar depths (right panel situation).

the domains in which a NCMC move is attempted. Specifically, we mean that, as the system is found in a shadow area, a NCMC move by a quantity d is attempted toward the other shadow area. The large difference of PMF in the two areas implies that the work performed on the system during left to right switching is very large, while it is very small in the opposite direction. Consequently, the acceptance probability for moving the system from left to right is much smaller than the acceptance probability of the reverse move. This would lead to a much greater sampling of the left area as it should on the basis of the free energy difference between the two areas.

Unluckily, this would have negative effects on sampling efficiency of NCMC based schemes, including our configurational freezing variant. This problem could be in principle tackled by using methodologies based on serial generalized ensemble schemes combined with finite-time nonequilibrium moves.[26] However, even if these methods may provide a uniform sampling of the collective coordinate, they would need some procedure to recover the sampling weights (namely the free energy difference between different regions of the collective coordinate), which may not be a simple task.[27−29] The configurational freezing NCMC methodology is therefore suitable for systems where the PMF along the collective coordinate does not present free energy minima with very different depths (see, e.g., Figure 9, right panel). Furthermore, another important aspect of NCMC is to select efficient nonequilibrium proposals, which requires a priori information, at least qualitative, about the physical system, especially about the location of the relevant energy barriers. Even if this could turn out to be difficult in most cases, in other situations designing nonequilibrium proposals could be suggested from the simple symmetry of the problem. This is just the case of torsional coordinates in flexible molecules. In any case, the selection of effective transition protocols in NCMC is a problem similar to that of choosing good reaction coordinates in other sampling methods[30] and is an object of active research.[3]

■ ASSOCIATED CONTENT

Supporting Information

(Figure) Evolution of the dimer extension r and related $g(r)$ obtained from cnv-MC, std-NCMC, and CF-NCMC simulations of a 216-particle sample supplied with a double-well potential having $5k_BT$ energy barrier. (Section 1) Definition of transition between compact and extended dimer configurations. (Section 2) Definition of root-mean-square deviation of the probability distribution $g(r)$ from the reference $g_{us}(r)$, obtained from the umbrella sampling simulation. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Email: riccardo.chelli@unifi.it.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by European Union Contract No. RII3-CT-2003-506350 and by the Italian Ministero dell'Istruzione, dell'Università e della Ricerca (No. PRIN 2010-2011).

■ REFERENCES

(1) Frenkel, D.; Smit, B. *Understanding Molecular Simulations: From Algorithms to Applications*; Academic Press: San Diego, 2002.
(2) Liu, J. S. *Monte Carlo Strategies in Scientific Computing*; Springer, San Diego, 2nd Ed., 2002.
(3) Nilmeier, J. P.; Crooks, G. E.; Minh, D. D. L.; Chodera, J. D. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, E1009−E1018.
(4) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. N.; Teller, E. J. *Chem. Phys.* **1953**, *21*, 1087−1092.
(5) Nicolini, P.; Frezzato, D.; Chelli, R. *J. Chem. Theory Comput.* **2011**, *7*, 582−593.
(6) Chelli, R. *J. Chem. Theory Comput.* **2012**, *8*, 4040−4052.
(7) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(8) Crooks, G. E. *J. Stat. Phys.* **1998**, *90*, 1481−1487.
(9) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361−2366.
(10) Chelli, R. *J. Chem. Phys.* **2009**, *130*, 054102.
(11) Schmiedl, T.; Seifert, U. *Phys. Rev. Lett.* **2007**, *98*, 108301.
(12) Lechner, W.; Oberhofer, H.; Dellago, C.; Geissler, P. L. *J. Chem. Phys.* **2006**, *124*, 044113.
(13) Ytreberg, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2004**, *120*, 10876−10879.
(14) Sun, S. X. *J. Chem. Phys.* **2003**, *118*, 5769−5775.
(15) Geissler, P. L.; Dellago, C. *J. Phys. Chem. B* **2004**, *108*, 6667−6672.
(16) Wu, D.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 204104.
(17) Nicolini, P.; Chelli, R. *Phys. Rev. E* **2009**, *80*, 041124.
(18) Chelli, R.; Gellini, C.; Pietraperzia, G.; Giovannelli, E.; Cardini, G. *J. Chem. Phys.* **2013**, *138*, 214109.
(19) Giovannelli, E.; Gellini, C.; Pietraperzia, G.; Cardini, G.; Chelli, R. *J. Chem. Phys.* **2014**, *140*, 064104.
(20) We note that the increase $\Delta\zeta$ may be different for the various perturbation steps. Here, for simplicity, we assume a fixed $\Delta\zeta$.
(21) Kirkwood, J. G. *J. Chem. Phys.* **1935**, *3*, 300−313.
(22) McQuarrie, D. A. *Statistical Mechanics*; HarperCollins Publishers: New York, 1976.
(23) Weeks, J. D.; Chandler, D.; Andersen, H. C. *J. Chem. Phys.* **1971**, *54*, 5237−5247.
(24) From now on, "MC step" is synonymous with "MC trial move".
(25) Torrie, G. M.; Valleau, J. P. *J. Comput. Phys.* **1977**, *23*, 187−199.
(26) Dirks, R. M.; Xu, H.; Shaw, D. E. *J. Chem. Theory Comput.* **2012**, *8*, 162−171.
(27) Chelli, R. *J. Chem. Theory Comput.* **2010**, *6*, 1935−1950.
(28) Chelli, R.; Signorini, G. *J. Chem. Theory Comput.* **2012**, *8*, 830−842.
(29) Chelli, R.; Signorini, G. *J. Chem. Theory Comput.* **2012**, *8*, 2552−2552.
(30) Nicolini, P.; Frezzato, D.; Gellini, C.; Bizzarri, M.; Chelli, R. *J. Comput. Chem.* **2013**, *34*, 1561−1576.