Konrad-Zuse-Zentrum
für Informationstechnik Berlin

Takustraße 7
D-14195 Berlin-Dahlem
Germany

ALEXANDER FISCHER CHRISTOF SCHÜTTE
PETER DEUFLHARD FRANK CORDES

Hierarchical Uncoupling-Coupling of Metastable Conformations

ZIB-Report 01-03 (February 2001)

Hierarchical Uncoupling-Coupling of Metastable Conformations

Alexander Fischer[1], Christof Schütte[1], Peter Deuflhard[1,2], and Frank Cordes[2]

[1] Freie Universität Berlin, Institut für Mathematik II, Arnimallee 2–6, 14195 Berlin, Germany
Email: alexander.fischer@math.fu-berlin.de

[2] Konrad-Zuse-Zentrum, Takustraße 7, 14195 Berlin, Germany

Abstract. Uncoupling-coupling Monte Carlo (UCMC) combines uncoupling techniques for finite Markov chains with Markov chain Monte Carlo methodology. UCMC aims at avoiding the typical metastable or trapping behavior of Monte Carlo techniques. From the viewpoint of Monte Carlo, a slowly converging long-time Markov chain is replaced by a limited number of rapidly mixing short-time ones. Therefore, the state space of the chain has to be hierarchically decomposed into its metastable conformations. This is done by means of combining the technique of conformation analysis as recently introduced by the authors, and appropriate annealing strategies. We present a detailed examination of the uncoupling-coupling procedure which uncovers its theoretical background, and illustrates the hierarchical algorithmic approach. Furthermore, application of the UCMC algorithm to the n-pentane molecule allows us to discuss the effect of its crucial steps in a typical molecular scenario.

Keywords: almost invariant sets, bridge sampling, metastability, hierarchical annealing, hybrid Monte Carlo, n-pentane molecule, ratio of normalizing constants, reweighting, uncoupling-coupling.

Mathematical Subject Classification: 60J22, 65C05, 65C40, 82B80

1 Introduction

Many problems in statistical physics can be stated as the computation of thermodynamical integrals $\mathrm{E}_f(g) = \int g(x)f(x)\,dx$ of a function or observable g w.r.t. a density f [1,7]. The widely used Markov chain Monte Carlo (MCMC) methodology provides a flexible and general framework for approximations of such expectation values by averaging over the realization of an appropriate Markov chain with invariant density f generated by the Monte Carlo algorithm.

Usually, application of MCMC to biomolecular systems has to tackle the *trapping problem*, i.e., the Markov chain remains for a very long time in one part of the state space before it moves on to another part. Such undesirable behavior of the Markov chain is caused by *metastable sets*—also called *modes* or *conformations*—in the state space, between which transitions are extremely rare. There exists a huge literature addressing the trapping problem [1,6,7]. Especially the so-called extended ensemble methods [18] like simulated tempering [20] or multicanonical algorithms [17,30] which are essentially based on reweighting techniques [8] are gaining significant popularity.

We herein present an alternative approach, the uncoupling-coupling scheme (UCMC), which has recently been introduced by the first author in [9]. The UCMC scheme combines the reweighting technique with the hierarchical decomposition of the state space into its metastable sets. The key idea of UCMC is to regard metastable sets as *almost invariant sets* w.r.t. some propagation operator corresponding to the Markov chain. Furthermore it combines aspects from simulated annealing approaches in optimization [19], aggregation-disaggregation techniques [28] and stochastic complementation techniques [22] for finite state space Markov chains. A hierarchical annealing structure is also used in the macrostate dissection approach for thermodynamical integrals [3]. UCMC essentially differs from these approaches by the consequent iterative decomposition into a hierarchy of almost invariant sets.

It has been shown recently that these almost invariant sets are strongly connected to the spectral structure of the propagation operator [4,26], and that it is even possible for a wide range of problem classes to identify almost invariant sets by computing the dominant eigenvalues of the propagation operator [5]. Typically the actual number of metastable sets is small for biomolecules [1], though the corresponding state space is high-dimensional. Even for such high dimensions the computational identification of almost invariant sets becomes possible in a hierarchical way by means of *parameter embedding* through the algorithm presented in [12] which the interested reader may also find in this volume.

Whenever the m dominant almost invariant sets are identified, significantly improved convergence properties are achieved by *uncoupling*, i.e., by parallel simulation of n independent chains, each one restricted to one of the almost invariant sets. Subsequently, the information lost in the uncoupling step, i.e., weighting factors between the almost invariant sets, is reconstructed by means of the stationary distribution of an appropriate coupling matrix C. In order to design an efficient algorithmic scheme, the uncoupling-coupling step is embedded into a hierarchical annealing structure, which naturally leads to bridge sampling techniques for computing the entries of the coupling matrix. Due to the independent Markov chains emerging in UCMC, its implementation is well suited for parallel computation.

2 Metastability in Markov Chain Monte Carlo

The paradigm of MCMC methods is to sample from a probability density f and use the output of a Markov chain to compute expectation values w. r. t. that density. To set the notation, let $\Omega \subset \mathbf{R}^d$ be the state space and f the density under consideration with $f > 0$.

Usually, f is defined in terms of an *unnormalized density* $\hat{f}$ via

$$f(x) = \frac{\hat{f}(x)}{Z_{\hat{f}}}, \quad \text{where} \quad Z_{\hat{f}} = \int_{\Omega} \hat{f}(x)\, dx$$

where $Z_{\hat{f}}$ denotes the *normalizing constant* of $\hat{f}$. In most cases one is interested in the canonical or Boltzmann density $\hat{f} = \exp(-\beta V)$ with inverse temperature β for some potential energy function $V : \Omega \to \mathbf{R}$.

The evolution of a Markov chain $\mathcal{X} = (X_k)$ with state space Ω is defined by a *stochastic transition function* $K : \Omega \times \Omega \to \mathbf{R}$, where $K(x, A)$ is the probability density to move from x to the set A in one step [23]. We call f an *invariant density* of the Markov chain given by K, if

$$f(y) = \int_{\Omega} K(x,y) f(x)\, dx \tag{1}$$

holds for all $y \in \Omega$.

In the Metropolis-Hastings algorithm a transition function K which satisfies (1) is realized by first defining an arbitrary but *irreducible* transition kernel $q(x,y)$ together with the *acceptance function*

$$\alpha(x,y) = \begin{cases} \min\left(1, \frac{q(y,x)\, f(y)}{q(x,y)\, f(x)}\right) & \text{for } q(x,y) > 0 \\ 1 & \text{otherwise} \end{cases} . \tag{2}$$

In α only ratios of the form $f(y)/f(x)$ have to be computed, which is feasible even if the normalizing constant $Z_{\hat{f}}$ is unknown.

Based on q and α we define K as the sum of two contributions,

$$K(x,y) = k(x,y) + r(x)\delta(x-y),$$

where the absolutely continuous part k is given by

$$k(x,y) = \begin{cases} q(x,y)\alpha(x,y) & \text{if } x \neq y \\ 0 & \text{otherwise} \end{cases}$$

and the singular component by $r(x) = 1 - \int k(x,y)\, dy$.

With this K one step in the realization of the Markov chain from the state $X_k = x$ consists of: *a*) propose some y distributed according to $q(x,y)$, *b*) accept this step by setting $X_{k+1} = y$ with probability $\alpha(x,y)$ or *c*) reject the proposal leaving $X_{k+1} = x$.

The construction of K guarantees that the associated Markov chain $\mathcal{X}$ is irreducible—provided that q is irreducible—and that for all $x, y \in \Omega$ the detailed balance condition

$$f(x)\, k(x,y) = f(y)\, k(y,x) \tag{3}$$

holds (for details, see e.g. [29]). Due to (3) K is called *reversible* w.r.t. f. If we further assume that $\mathcal{X}$ is aperiodic—which is guaranteed whenever $r > 0$—we can state that f is the *unique* invariant density of $\mathcal{X}$.

A realization $\{x_k\}$ of $\mathcal{X}$ with sample points $k = 1, \ldots, N$ now enables us to calculate expectation values $\mathrm{E}_f(g) = \int_\Omega g(x)f(x)\,dx$ w.r.t. f by using the estimator

$$\mathrm{E}_f^N(g) = \frac{1}{N}\sum_{k=1}^{N} g(x_k), \tag{4}$$

which converges to $\mathrm{E}_f(g)$ for $N \to \infty$. Altogether, we can say that MCMC is a method that allows to sample from f without knowledge of the normalizing constant $Z_{\hat{f}}$.

For two sets $A, B \subseteq \Omega$ the *transition probability* between A and B within an ensemble distributed w.r.t. the density f and during one step of the Markov chain is given by

$$\kappa(A,B) = \frac{1}{\int_A f(x)\,dx}\int_A\int_B K(x,y)f(x)\,dx\,dy. \tag{5}$$

Discretization. Using this, we can easily discretize the Markov chain given by K. This is done by *coarse graining* with an arbitrary box decomposition of the phase space Ω into m disjoint sets $B_1, \ldots, B_m \subset \Omega$ with $\cup B_j = \Omega$. Based on this box decomposition, we introduce the new finite phase space $\tilde{\Omega} = \{B_1, \ldots, B_m\}$ and define the transition function $\tilde{K}$ on $\tilde{\Omega}$ via

$$\tilde{K}(B_k, B_l) \;=\; \kappa(B_k, B_l). \tag{6}$$

The finite dimensional Markov chain defined by $\tilde{K}$ again is reversible w.r.t. its invariant density $\tilde{f}$ given by $\tilde{f}(B_k) = \int_{B_k} f(x)dx$. Whenever f is unique for K, $\tilde{f}$ is also unique for $\tilde{K}$.

2.1 Propagator

In the following we want to understand the global behavior of a Markov chain via the eigenmodes of its associated *propagator* P. This propagator is defined in terms of the transition function K by

$$Pu(y) = \int_\Omega k(x,y)u(x)\,dx + r(y)u(y). \tag{7}$$

P describes the propagation of a phase space density with one step of the Markov chain. One can show that the reversibility of K w.r.t. f implies that its spectrum $\sigma(P)$ is real-valued. More exactly, we have $\sigma(P) \subseteq [-1, 1]$, and the largest eigenvalue is $\lambda = 1$, for which f is an eigenfunction, i.e. $Pf = f$. Assume that we order the eigenvalues of P w.r.t. their modulus, such that we have $\lambda_1 = 1 > \lambda_2 \geq \ldots$. Then, the approximation of expectation values via the Markov chain $\mathcal{X}$ goes with a geometric rate:

$$|\mathrm{E}_f(g) - \mathbf{E}^{x_0} \frac{1}{N} \sum_{k=1}^{N} g(x_k)| \leq C\,|\lambda_2|^N, \qquad (8)$$

where E^{x_*} denotes the expectation over all realization of the chain starting in $x_* \in \Omega$. The smaller $|\lambda_2|$, the faster a good sampling of the density f is achieved. A Markov chain is called *rapidly mixing* whenever $|\lambda_2| \ll 1$.

Recall that the spectrum $\sigma(P)$ consists of two disjoint parts: the discrete spectrum containing all isolated eigenvalues of P with finite multiplicity, and a continuous part, the so-called essential spectrum. In the case considered herein, this essential spectrum is contained in some interval $[-\rho, \rho]$ with ρ being the supremum of the rejection function r over Ω. We assume that $\rho \ll \lambda_2$, which typically is the case; therefore an inference of the essential spectrum with eigenvalues in the vicinity of $\lambda_1 = 1$ will not occur in the following.

Discretization of P. Next, assume that we discretize the Markov chain w.r.t. some box decomposition $B_1, \ldots, B_m \subset \Omega$ resulting in the transition function $\tilde{K}$ given in (6). Then the phase space is finite and the propagator P becomes an $m \times m$ propagation matrix $\tilde{P}$ which simply is the column stochastic matrix with entries $P_{lk} = \tilde{K}(B_k, B_l) = \kappa(B_k, B_l)$.

2.2 Metastability

If λ_2 is close to $\lambda_1 = 1$, we often find that the reason for the undesirably slow convergence is that the Markov chain remains for a long time in a *metastable* region—also called *mode* or *conformation*—of the phase space, before it moves on to another one. We will call a set A *metastable* or *almost invariant* under our Markov chain, if the transition probability from A to itself is close to one, i.e., if $\kappa(A, A) \approx 1$.

We herein will exploit the following observation concerning metastability: If there are n eigenvalues close to $\lambda_1 = 1$ (including λ_1 itself) and a significant spectral gap to all remaining eigenvalues, then there also are n disjoint metastable subsets and vice versa [22,27].

If this is the case, the chain is rapidly mixing *within* the corresponding metastable subsets and the undesirably slow overall convergence results from the rareness of transitions between these metastable sets.

The close connection between a separated cluster of dominant eigenvalues and the existence of metastable subsets has another very important algorithmic consequence: it has been shown that one can *identify* the n metastable subsets only on basis of the *eigenvectors* associated with the n dominant eigenvalues [26,27]. This insight leads to a significantly general identification algorithm [5] used for the detection of biomolecular conformations. For illustration of its key idea we introduce an appropriate example in Sect. 2.3 below.

2.3 Illustrative Example

Now and in Sect. 3, our construction will be illustrated by means of the *n-butane* molecule. The United-Atoms representation [24] is used to set up a *separated Hamiltonian* $\mathcal{H}(p,x) = \mathcal{T}(p) + \mathcal{V}(x)$, where the kinetic energy $\mathcal{T}(p)$ depends only on the generalized momenta p and the potential energy $\mathcal{V}(x)$ only on the coordinates x. Although the potential $\mathcal{V}$ is 12-dimensional, the overall structure and dynamical properties of n-pentane is mainly determined by its torsion angle $\theta = \theta(x)$. For means of illustration, we therefore choose θ as our single conformational degree of freedom and select a decomposition of the range $[0, 2\pi]$ of θ into 23 disjoint boxes $\Theta_k \subset [0, 2\pi]$, $k = 1, \ldots, 23$. These boxes define disjoint boxes

$$B_k \;=\; \{x \in \Omega : \quad \theta(x) \in \Theta_k\}, \quad k = 1, \ldots, 23$$

which decompose the entire phase space $\Omega \subset \mathbf{R}^{12}$.

For the MCMC sampling we use hybrid Monte Carlo (HMC) [2,6,26] which has become a widely used method over the last decade for computing expectation values (mainly thermodynamic observables) in molecular systems [1]. It perfectly fits into the framework of MCMC discussed herein. Under weak assumptions on the potential, the canonical density given by $\hat{f} = \exp(-\beta\mathcal{V})$ is its unique invariant density. Let K denote the transition function of HMC for the inverse temperature β, and let $\tilde{K}$ be its discretization w.r.t. the box decomposition designed in the last paragraph. The propagation matrix $\tilde{P}$ associated with $\tilde{K}$ is illustrated in Fig. 1, whereas the eigenvectors of the three dominant eigenvalues are given in Fig. 2.

Fig. 2 also illustrates the key idea of the algorithm for identifying almost invariant sets via these eigenvectors: For each state $j = 1, \ldots, 23$, we denote by $s_j \in \{+,-\}^3$ the 3-tupel of signs of the jth components in each of the three eigenvectors, the so-called sign combinations. The forth subfigure in Fig. 2 shows that there are only three different sign combinations, and that all states j with the same s_j belong to the same metastable set. Thus, the metastable sets can be identified as sets of states with identical sign combination. This idea can be generalized substantially, and is the key idea of the identification algorithm presented in [5].

3 The Uncoupling Step

Assume now, that we are in a situation with n disjoint metastable subsets. If it is true, that the chain is rapidly mixing *within* each metastable subset and that the undesirably slow overall convergence results from the weak coupling

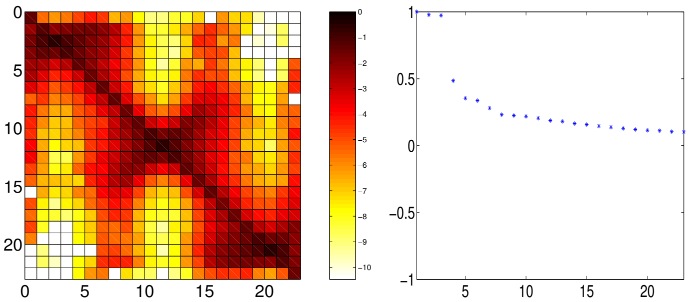

Fig. 1. Left: Illustration of the entries of the propagation matrix $\tilde{P}$ defined in the text above. Intensity of entries due to logarithmic scale. Right: Ordered spectrum of $\tilde{P}$ with a cluster of three eigenvalues close to $\lambda = 1$ and a significant gap to all remaining ones.

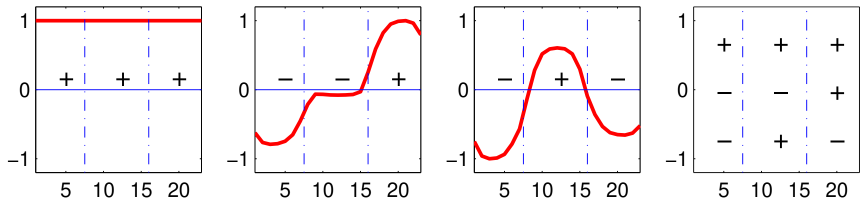

Fig. 2. Subfigure 1–3: Right eigenvectors of the three dominant eigenvalues of the propagation matrix $\tilde{P}$ shown in Fig. 1. Subfigure 4: The three metastable subsets are characterized by three different sign combinations of these eigenvectors.

between this metastable sets, then uncoupling of the metastable sets should lead to n rapidly mixing uncoupled chains.

3.1 Restricted Sampling

Assume that we know the n disjoint metastable sets $A_1, \ldots, A_n$ of our Markov chain, and that we now want to sample separately in each A_l, for $l = 1, \ldots, n$. Then, for each l we define a restricted Markov kernel K_l from K on A_l by setting

$$K_l(x, y) = k_l(x, y) + r_l(x)\delta(x - y) \tag{9}$$

with

$$k_l(x, y) = \begin{cases} q(x, y)\alpha(x, y) & \text{if } x \neq y \text{ and } y \in A_l \\ 0 & \text{otherwise} \end{cases}$$

and

$$r_l(x) = 1 - \int k_l(x, y)\, dy.$$

Clearly, detailed balance still holds, so that K_l is again a reversible Markov kernel. Now, let $\hat{f}_l = \mathbf{1}_{A_l}\hat{f}$ be the restricted unnormalized density on A_l, with $\mathbf{1}_A$ denoting the indicator function on A, i. e., $\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ otherwise. Then, under the assumption, that K_l is irreducible, $f_l = \hat{f}_l / Z_{\hat{f}_l}$ is the unique invariant density of K_l.

We denote by P_l the corresponding propagator of K_l. If we assume that A_l is metastable and that it cannot be subdivided further into two or more almost invariant sets, then we can state the following: The second largest eigenvalue λ_2 of P_l is substantially less than 1, otherwise there would exist a decomposition into two or more metastable subsets. As a consequence, due to $\lambda_2 \ll 1$, the corresponding Markov chain $\mathcal{X}_l$ is rapidly mixing.

For the restricted Markov kernel K_l the detailed balance condition (3) still holds for all $x, y \in A_l$; therefore the density f_l is a scalar multiple of the correct global density f of the unrestricted Markov chain. Thus, we can regain the global density via

$$f = \sum_{l=1}^{n} \pi_l f_l \tag{10}$$

in terms of the local densities f_k. Only the scalars π_l, $l = 1, \ldots, n$, are unknowns which represent the neglected coupling between the sets A_l.

But before we go into the details of the coupling algorithm for the computation of these weights π_l, we want to give illustrations of the steps taken so far.

3.2 Illustration of Restricted Sampling

For ease of presentation we will now illustrate this procedure in a *finite* dimensional situation. To this end, let $\tilde{P}$ again denote the propagation matrix of HMC for the simple n-butane molecule associated with the box discretization given in Sec. 2.3. Let us denote the associated chain by $\mathcal{X}$ in the following. Moreover, let A_1, A_2, A_3 be three disjoint subsets which we want to be uncoupled. The resulting restricted sampling corresponds to the propagation matrix $\tilde{P}_{\text{restr}}$ with entries

$$\tilde{P}_{\text{restr},kl} = \begin{cases} \tilde{P}_{kl}, & k, l \in A_i \text{ for } i \in \{1,2,3\} \text{ and } k \neq l \\ 0, & k \in A_i, l \in A_j \text{ for } i, j \in \{1,2,3\} \text{ and } i \neq j \\ \tilde{P}_{ll} + \sum\limits_{i \notin A_j} \tilde{P}_{il}, & k = l \in A_j \end{cases}$$

Consequently, if we assume the boxes for each subset A_i to be in a successive order, $\tilde{P}_{\text{restr}}$ has block-diagonal form, i.e., the associated restricted sampling chain consists of three uncoupled Markov chains. The three stochastic matrices $\tilde{P}_l$ on the block diagonal of $\tilde{P}_{\text{restr}}$ are the propagation matrices associated with these uncoupled chains.

Figure 3 illustrates the situation when the sets $A_1 = \{1,\ldots,7\}, A_2 = \{8,\ldots,16\}, A_3 = \{17,\ldots,23\}$ are *good* approximations of the three metastable sets of $\mathcal{X}$. The right hand part of Fig. 3 shows the ordered spectra of the three uncoupled propagation matrices P_l. We observe that the largest second eigenvalues of all three P_l indeed are substantially less than 1, i.e., the three restricted sampling chains in fact are rapidly mixing.

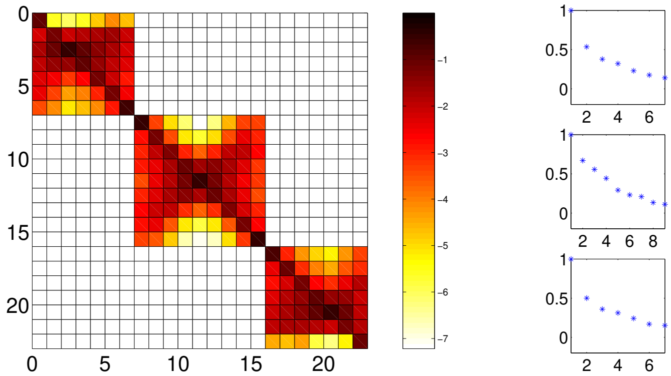

Fig. 3. Left: Illustration of the entries of the propagation matrix $\tilde{P}_{\mathrm{restr}}$ (as defined in the text above) for a good choice of A_1, A_2, A_3. Intensity of entries due to logarithmic scale. Right: Ordered spectrum of $\tilde{P}$.

Figure 4 shows the invariant density f of the unrestricted chain in comparison to the invariant densities f_l of the three resulting restricted chains. As a consequence of (10), we observe $f/f_l = \text{const} = \pi_l$ on each subset A_l, $l = 1, 2, 3$.

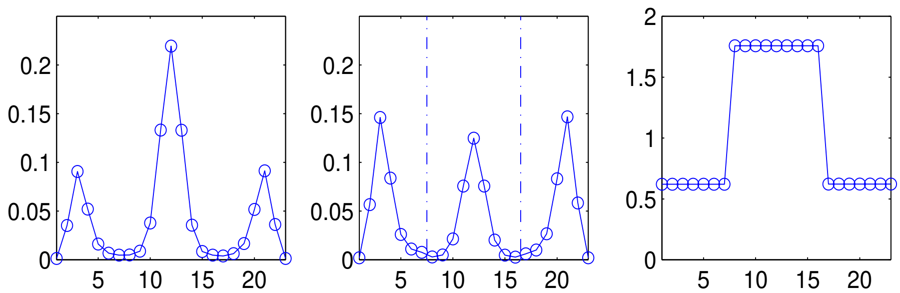

Fig. 4. Left: Invariant density f. Center: Invariant densities f_l, $l = 1, 2, 3$ in the three metastable sets A_l (as in Fig. 3). Right: Quotients $\pi_l = f/f_l$ in the three sets A_l.

Figure 5 now illustrates *bad* approximations of the three metastable sets of $\mathcal{X}$, namely $A_1 = \{1, \ldots, 9\}, A_2 = \{10, \ldots, 15\}, A_3 = \{16, \ldots, 23\}$. Again, the right hand part of Fig. 3 shows the ordered spectra of the three uncoupled propagation matrices P_l. We observe that the largest second eigenvalue now is much closer to 1, so that at least the first of the three restricted sampling chains is slowly mixing.

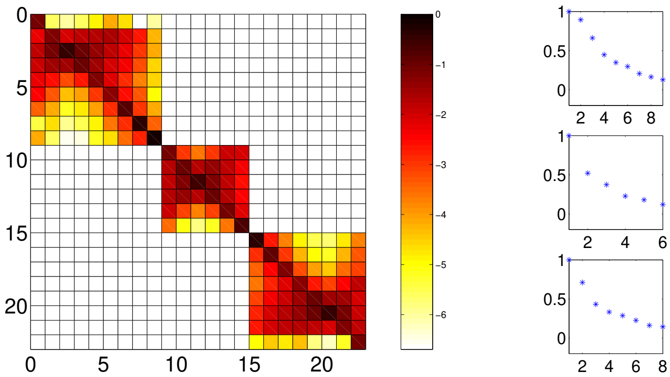

Fig. 5. Left: Illustration of the entries of the propagation matrix $\tilde{P}_{\mathrm{restr}}$ for a bad choice of A_1, A_2, A_3. Intensity of entries due to logarithmic scale. Right: Ordered spectrum of $\tilde{P}$.

3.3 Metastable Sets in Many Dimensions

In Sects. 2.2 and 2.3 we saw how to identify the metastable sets of Markov chains with low-dimensional and/or finite state space via dominant eigenvectors of the associated propagator. In this section we will show how to deal with high-dimensional state spaces. We will see that one may extract good approximations of the metastable subsets of a certain Markov chain $\mathcal{X}$ (if they exist) from good $\mathcal{X}$-samplings of its invariant density f.

This seems to be a vicious circle: we want to use good approximations of the metastabe sets to find faster sampling strategies by uncoupling-coupling techniques, but we also want to start the construction of such good approximations by assuming that some reliable sampling already is available. In Sec. 3.4 below, we will see that we are able to avoid this supposed circle by exploiting an appropriate annealing strategy.

Suppose that $\{x_k\}_{k=1,\ldots,L}$ is the realization of a Markov chain $\mathcal{X}$ acting on the state space Ω and corresponding to a propagator P. In order to identify the metastable sets we have to compute the dominant eigenvectors of P, and therefore discretize P. Hence, we need an appropriate box discretization of Ω where the number of boxes should not be so large that it prevents us from computing the dominant eigenvectors of the discretization matrix $\tilde{P}$ of P. In order to design such a box discretization we first identify the cluster structure of the sampling, i.e., we cluster $\{x_k\}$ into a sufficiently large number of similarity clusters. This can be done, e.g., via the algorithm presented in [12] using self-organizing maps. These clusters define our discretization boxes such that the algorithm works as follows:

(1) Use the techniques presented in [12] to cluster $\{x_k\}_{k=1,\ldots,L}$ into M similarity clusters $C_1, \ldots, C_M$. (M need not be defined in advance, but emerges during the algorithmic process.)
(2) Assemble the discretization matrix $\tilde{P} \in \mathbf{R}^{M \times M}$ of P with entries

$$\tilde{P}_{kl} = \frac{\#\{x_j \in C_l \text{ and } x_{j+1} \in C_k\}}{\#\{x_j \in C_l\}}$$

(3) Compute metastable sets of $\tilde{P}$ via its dominant eigenvectors using the algorithm presented in [5].

3.4 Annealing

If we start the simulation with a Markov chain for the normalized density f of interest, it will typically get trapped in one of its metastable sets—which is exactly what we want to avoid. One way out of this undesirable occurrence is to exploit the embedding of the unnormalized density $\hat{f}$ out of a family of densities $\hat{f}(\beta)$. Here β is the embedding parameter, e.g., the inverse temperature in the usual case of the Gibbs densities $\hat{f}(\beta) = \exp(-\beta\mathcal{V})$. Let β_* be such that the density of interest is $\hat{f} = \hat{f}(\beta_*)$. Correspondingly, we have a family of Monte Carlo Markov chains $\mathcal{X}(\beta)$, and a family of propagation operators $P(\beta)$.

Consider now the special case of the inverse temperature β and $\hat{f}(\beta) = \exp(-\beta\mathcal{V})$: By increasing temperature, i.e., decreasing β, the density $\hat{f}(\beta_*)$ transforms into smoother ones with less metastable regions. This means that the second eigenvalue $\lambda_2(\beta)$ of $P(\beta)$ moves away from the largest eigenvalue $\lambda_1(\beta) = 1$, i.e., the spectral gap increases with decreasing β, and we get better and better convergence properties for the chains $\mathcal{X}(\beta)$.

For reasonably small β (but sufficiently bounded away from $\beta = 0$), the chain $\mathcal{X}(\beta)$ will still exhibit almost invariant sets with a significantly reduced degree of metastability compared to that of the chain $\mathcal{X}(\beta_*)$. However, the almost invariant sets of $\mathcal{X}(\beta)$ will be reasonable approximations of almost invariant sets of $\mathcal{X}(\beta_*)$ [27]. This is due to the fact that the form of the dominant eigenvectors of the $P(\beta)$ is only mildly sensitive to changes in β (in contrast to the drastic effect of such changes on $\lambda_2(\beta)$) [25].

We can nicely see this when returning to the n-butane example of Sect. 2.3. Let the inverse temperature β be associated with a temperature of 300 K. Based on simulation of $\mathcal{X}(\beta)$, the identification algorithm explained in Sect. 3.3

results in the almost invariant sets A_1, A_2, A_3 already shown in Fig. 2. Let $\mathcal{X}_j$ denote the restricted chains associated with the densities $\mathbf{1}_{A_j} \hat{f}(\beta_*)$, $j = 1, 2, 3$, for the low temperature $\beta_* \sim 100\,\mathrm{K}$ of interest. Then, the second eigenvalues λ_j of the propagators associated with the $\mathcal{X}_j$ are

j	1	2	3
λ_j	0.328	0.372	0.307

which illustrates that these restricted chains are rapidly mixing as desired. In comparison, an estimate for the second eigenvalue of the unrestricted chain $\mathcal{X}(\beta_*)$ yields $\lambda_2(\beta_*) \geq 0.999$, indicating extremely slow mixing.

4 The Coupling Step

In the coupling step we will show that it is possible to regain information about a global density $f = \sum_k \pi_k f_k$ in terms of densities f_k without sampling f itself, by setting up a coupling matrix C with π as its stationary distribution. This together with the decomposition from the uncoupling step allows us to formulate the algorithmic hierarchical annealing scheme in Sec. 5.

4.1 The Coupling Matrix C

Now suppose that arbitrary unnormalized densities $\hat{f}_0, \ldots, \hat{f}_n$ are given and that we are interested in the global unnormalized density

$$\hat{f} = \sum_{k=0}^{n} \hat{f}_k \, . \tag{11}$$

For some set $A \subset \Omega$ we introduce the notation $\mu(A) = \int_A \hat{f}(x)dx$. To obtain information about the density f corresponding to $\hat{f}$, it is sufficient to know the ratios of normalizing constants $\pi_k = Z_{\hat{f}_k}/Z_{\hat{f}}$, because then we can reconstruct f from the f_k's due to (11) by

$$\sum_{k=0}^{n} \pi_k \, f_k = \sum_{k=0}^{n} \frac{Z_{\hat{f}_k}}{Z_{\hat{f}}} \frac{\hat{f}_k}{Z_{\hat{f}_k}} = \frac{\hat{f}}{Z_{\hat{f}}} = f. \tag{12}$$

Let us denote by $A_k = \mathrm{supp}(\hat{f}_k)$ the support of the densities $\hat{f}_k$ in the state space Ω. Furthermore, let us assume in the following that each A_i is connected to any A_j in the sense that there exists a sequence of sets

$$A_i = A_{l_1}, A_{l_2}, \ldots, A_{l_{k-1}}, A_{l_k} = A_j, \text{ such that } \mu(A_{l_r} \cap A_{l_{r+1}}) > 0, \tag{13}$$

for $r = 1, \ldots, k-1$. This condition obviously is not satisfied if we assume the $\hat{f}_k$ to be the restricted densities of Sect. 3.1. However, it makes sense in the context of annealing, e.g., with $\hat{f}_0 = \hat{f}(\beta)$, and $\hat{f}_j = \mathbf{1}_{A_j} \hat{f}(\beta_*)$, with A_j, j=1, …, n being the almost invariant sets of $\mathcal{X}(\beta)$, cf. Sect. 3.4. Now, the densities $\hat{f}_0, \ldots, \hat{f}_n$ automatically satisfy our above conditions, in particular the connectivity condition (13).

Definition of the Coupling Matrix. Returning to the general case, we need to design an algorithm which allows to compute the weights π_k in (12) (or at least approximations of them) without directly referring to $Z_{\hat{f}}$. We therefore define the *coupling matrix* $C = (c_{ij}) \in \mathrm{Mat}_{n+1\times n+1}$ by

$$c_{ij} = \begin{cases} \frac{1}{n+1}\,\frac{Z_{\phi_{ij}\hat{f}_i}}{Z_{\hat{f}_i}}\,\min\left(1, \frac{Z_{\phi_{ji}\hat{f}_j}}{Z_{\phi_{ij}\hat{f}_i}}\right) & \text{for } i \neq j \text{ and } \mu(A_i \cap A_j) > 0 \\ 0 & \text{for } i \neq j \text{ and } \mu(A_i \cap A_j) = 0 \\ 1 - \sum_{k=0(k\neq i)}^{n} c_{ik} & \text{else} \end{cases} \tag{14}$$

where $\phi_{ij} = \mathbf{1}_{A_i \cap A_j}$ denotes the common support of the densities $\hat{f}_i$ and $\hat{f}_j$.

Properties of the Coupling Matrix. Obviously, C is a stochastic matrix, because for $i \neq j$ we have $0 \leq c_{ij} \leq 1/(n+1)$, while due to the diagonal entries the sum of each row is 1. The Markov chain corresponding to C is also aperiodic, simply because $c_{ii} \geq 1/n+1$ for each diagonal entry. Condition (13) guarantees that for any two $i, j \in \{0, \ldots, n\}$ there is a path from the state i to the state j in C, which makes C irreducible in addition. The key point in the construction of C is that

$$(\pi_1, \ldots, \pi_n) = \frac{1}{Z_{\hat{f}}}(Z_{\hat{f}_0}, \ldots, Z_{\hat{f}_n})$$

is the unique stationary distribution due to the aperiodicity and irreducibility of C. This follows immediately from the detailed balance condition

$$\pi_i\,\frac{Z_{\phi_{ij}\hat{f}_i}}{Z_{\hat{f}_i}}\,\min\left(1, \frac{Z_{\phi_{ji}\hat{f}_j}}{Z_{\phi_{ij}\hat{f}_i}}\right) = \pi_j\,\frac{Z_{\phi_{ji}\hat{f}_j}}{Z_{\hat{f}_j}}\,\min\left(1, \frac{Z_{\phi_{ij}\hat{f}_i}}{Z_{\phi_{ji}\hat{f}_j}}\right), \tag{15}$$

which moreover shows that C is reversible.

Expectation Values. Suppose that we can compute expectation values for the f_k's, e.g., by restricted sampling. Moreover assume that we know the correct weights π of $f = \sum_k \pi_k f_k$ via the stationary distribution of C. Then, we are able to compute expectation values w.r.t. f, which are now given by

$$\mathrm{E}_f(g) = \mathrm{E}_{\sum_k \pi_k f_k}(g) = \sum_k \pi_k \int_{A_k} g(x) f_k(x)\, dx. \tag{16}$$

Thus, the remaining bottleneck now is to find an algorithm for efficiently computing an approximation $\tilde{C}$ of C. Therein, we will have to approximate the ratio of normalizing constants, which as opposed to approximation of the normalizing constants itself can at least in principle be computed efficiently (see Sec. 4.2). In fact, like in the Metropolis algorithm we replace a direct computation of $Z_{\hat{f}}$ by the computation of ratios of normalizing constants between the $Z_{\hat{f}_k}$'s.

4.2 Bridge Sampling

We herein continue to use the notation of the last subsection. In the following, let i and j always be two indices satisfying $\mu(A_i \cap A_j) > 0$. Then we have to approximate the ratio of normalizing constants $Z_{\phi_{ij}\hat{f}_i}/Z_{\phi_{ji}\hat{f}_j}$. We could try to sample $\hat{f}_i$ and $\hat{f}_j$ directly (with good convergence rates both) and compute the approximation from the generated sampling data. Unfortunately, because of the high dimensionality of Ω, the overlap between these two samplings would in general be to small to extract a statistically reliable approximation of the ratio of normalizing constants.

Therefore, we have to use so-called *bridge densities* to compute the desired ratios of normalizing constants [14,21]. A generic choice for a bridge density on the set $A_i \cap A_j$ is given by

$$\hat{f}_{ij} = \sigma \mathbf{1}_{A_i \cap A_j}\, \hat{f}_i + (1-\sigma)\mathbf{1}_{A_i \cap A_j}\, \hat{f}_j \tag{17}$$

for some $\sigma \in [0,1]$; however, more elaborate bridge densities are in use depending on the specific application. By mixing both densities into f_{ij}, we expect to satisfy in particular:

(i) The Markov chain $\mathcal{X}_{ij}$ corresponding to f_{ij} is rapidly mixing. This assumption is justified whenever we can guarantee rapid mixing for the chains $\mathcal{X}_i$ and $\mathcal{X}_j$ associated with f_i and f_j.
(ii) A simulation run of $\mathcal{X}_{ij}$ allows a statistical reasonable *reweighting* to the densities f_i and f_j; this presupposes that all important parts of the densities f_i and f_j got sampled by the simulation. This means that the reweighted data allow to approximate $Z_{\phi_{ij}\hat{f}_i}/Z_{\phi_{ij}\hat{f}_i}$.

4.3 Annealing Example

Let us again return to the situation with $\hat{f}_0 = \hat{f}(\beta)$, and $\hat{f}_j = \mathbf{1}_{A_j}\hat{f}(\beta_*)$, with A_j, j=1,…,m being the almost invariant sets of $\mathcal{X}(\beta)$, cf. Sect. 3.4. Suppose that β_* is the inverse temperature of interest, and that the inverse temperature $\beta < \beta_*$ gives us a chain $\mathcal{X}(\beta)$ with sufficient convergence properties. Since now $\hat{f} = \sum_j \hat{f}_j = \hat{f}(\beta) + \hat{f}(\beta_*)$ by construction, the last m weights π_j in (12) satisfy

$$\sum_{j=1}^{m} \pi_j \;=\; \sum_{j=1}^{m} \frac{Z_{\hat{f}_j}}{Z_{\hat{f}}} \;=\; \frac{Z_{\beta_*}}{Z_\beta + Z_{\beta_*}},$$

where Z_β and Z_{β_*} denotes the normalizing constants of $\hat{f}(\beta)$ and $\hat{f}(\beta_*)$. So, the reweighted coefficients

$$\pi_j^* \;=\; \pi_j / \sum_{j=1}^{m} \pi_j \tag{18}$$

are the correct weighting factors for the density $f(\beta_*) = \sum_{j=1}^{m} \pi_j^* f_j$ of interest.

By construction, we have $\mu(A_i \cap A_j) > 0$ only for $i = 0$ and $j = 1, \ldots, m$ or the reverse case. Since $\phi_{0j} = \phi_{j0} = \mathbf{1}_{A_j}$, we have to approximate the ratios $Z_{\phi_{0j} \hat{f}_0} / Z_{\phi_{j0} \hat{f}_j} = Z_{\mathbf{1}_{A_j} \hat{f}_0} / Z_{\hat{f}_j}$. Thus, the required bridge densities f_{0j} combine the chain $\mathcal{X}^{(\beta)}$ with the restricted chains $\mathcal{X}_j$ associated with $\hat{f}_j$, which all are rapidly mixing, since β was assumed to be small enough and the $\hat{f}_j$'s are restricted to the almost invariant sets A_j.

For the n-butane example of Sec. 2.3 a sampling for β given by a temperature of 300 K results in the three almost invariant sets already illustrated in Fig. 4. If we are interested in β_* given by a temperature of 100 K, the bridge sampling results in the following approximate 4×4 coupling matrix:

$$\tilde{C} = \begin{pmatrix} 0.9939 & 0.0001 & 0.0058 & 0.0001 \\ 0.2500 & 0.7500 & 0 & 0 \\ 0.2500 & 0 & 0.7500 & 0 \\ 0.2500 & 0 & 0 & 0.7500 \end{pmatrix},$$

with the stationary distribution

$$\tilde{\pi} = (0.97637, 0.00054, 0.02254, 0.00056)^T,$$

which implies due to (18):

$$(\tilde{\pi}_1^*, \tilde{\pi}_2^*, \tilde{\pi}_3^*) = (0.0228, 0.9535, 0.0237).$$

Let $\{x_k^{0j}, k = 1, \ldots, N_k\}$ denote the sampling of the bridge densities f_{0j}, $j = 1, 2, 3$, and let $\{\alpha_k^{0j}, k = 1, \ldots, N_k\}$ denote the reweighting factors of the f_{0j}-sampling to f_j. Then, due to (16), expectation values of a function g w.r.t. the density $f(\beta_*) = \hat{f}(\beta_*)/Z_{\hat{f}(\beta_*)}$ of interest can be approximated via

$$\mathrm{E}_{f(\beta_*)}(g) \approx \sum_{j=1}^{m} \tilde{\pi}_j^* \left(\sum_{k=1}^{N_j} \alpha_k^{0j} \; g\left(x_k^{0j}\right) \right), \tag{19}$$

If we use this to approximate the probabilities $p_j = \mathrm{E}_{f(\beta_*)}(\mathbf{1}_{A_j}) = \int_{A_j} f(\beta_*)\, dx$ to be in the sets A_j for the n-butane example, we get

$$p = (0.0228, 0.09535, 0.0237),$$

which we have to compare with the exact values

$$p = (0.0240, 0.9520, 0.0240),$$

which can be computed analytically in this simple example.

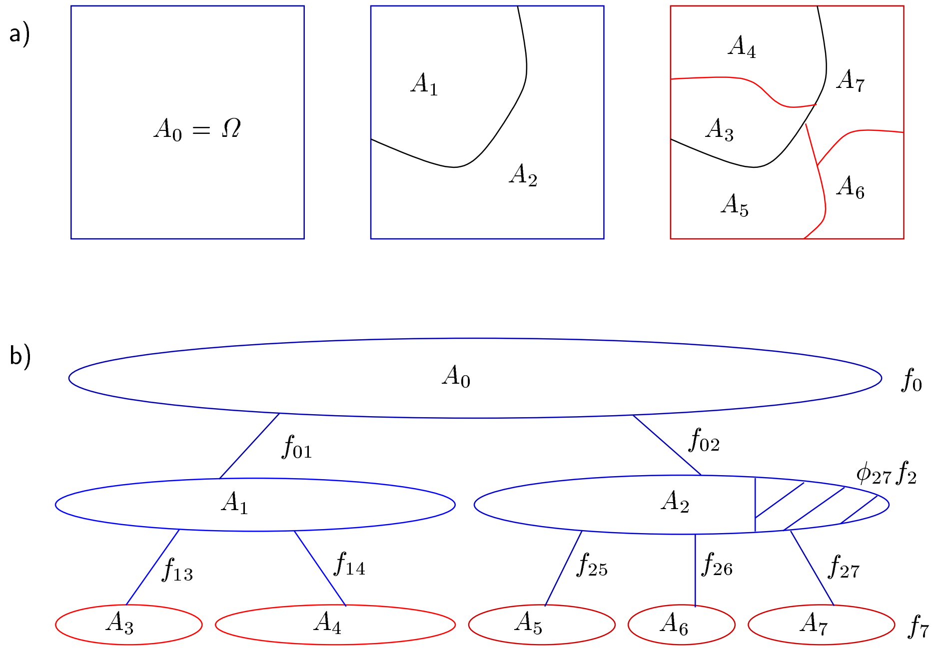

Fig. 6. a) Schematic plot of a three level hierarchical decomposition. An initial sampling of f_0 decomposes the state space $\Omega = A_0$ into two subsets A_1 and A_2, which get further subdivided into $\{A_3, A_4\}$ and $\{A_5, A_6, A_7\}$. The three levels are related to an annealing process; the top level is related to β_0, the intermediate level to $\beta_1 > \beta_0$, and the ground level to the inverse temperature $\beta_2 = \beta_* > \beta_1$ of interest. b) The same subdivision as in a), but now represented as a graph, where nodes correspond to the sets A_k. As an example, the density $\phi_{27} f_2$ corresponds to the hatched part of A_2; in UCMC neither $\phi_{27} f_2$ nor f_7 is sampled, but rather the bridge density f_{27}, which sufficiently encompasses the important parts of $\phi_{27} f_2$ and f_7. Additionally, the tree structure of the graph guarantees that the coupling matrix C is irreducible.

5 Hierarchical Annealing

Why should we restrict the algorithm to just two temperatures and only one decomposition of the phase space in almost invariant subsets? As already emphasized we have to expect that there is a *hierarchy of almost invariant sets*. Moreover, it will be much easier to choose "good" initial "high" temperatures, if we allow for a *hierarchy of temperatures* ranging from the high initial temperature down to the probably low temperature of interest. The presentation of all prior steps has been general enough to allow for a recursive, hierarchical k-level generalization of the two-level approach explained in the annealing example in the previous section. The resulting concept of "Hierarchical UCMC" is illustrated and explained in detail in Fig. 6.

In order to control the statistical error of the required samplings of the bridge densities, we have to control the simulation length of each sampling. Since appropriate simulation lengths may vary drastically, we use the convergence estimator described in [13,15] to automatically stop the simulation. For this estimator, multiple realizations of a Markov chain $\mathcal{X}_k$ are generated to compute estimates depending on the variances between these realizations.

In view of the various samplings of bridge densities the hierarchical approach has another benefit: Parameters for a bridge density can be directly extracted from its previous density in the hierarchy. Therefore, there is no need to perform preliminary simulations to adjust parameters as is typically the case for other techniques using extended ensembles [1,18]. Because we start UCMC with a standard Monte Carlo simulation at a sufficiently high temperature, the uncoupling step of the algorithm runs fully automatic until the density of interest is reached. After that, the approximation $\tilde{C}$ of the coupling matrix C is obtained from the uncoupling step by pure data analysis; the corresponding reweighting formulas and the approximation of ratios of normalizing constants can be found in [9].

6 Numerical Example

Biomolecules are an important application class of MCMC methods in statistical physics. There exists a wide range of MCMC algorithms, which are trying to tackle the problems and challenges of biomolecular systems [1]. Biomolecules are also well suited for the UCMC approach: They possess many conformational substates which can be clustered into only a few, extremely metastable ones [11]. In other words, metastable conformations consist themselves out of less metastable subsets. As an illustration of the hierarchical decomposition of conformational substates, we here apply UCMC to n-pentane. This molecule is still far below the complexity of proteins or nucleic acids, but the algorithmic properties can be discussed in more detail. We have used the all-atom Merck force field [16] for the energy representation which we also use for simulations on biomolecules. The n-pentane molecule consists of

5 C-atoms and 12 H-atoms, which results in $\Omega \subset \mathbf{R}^{51}$. The clustering of data by means of self-organizing box maps [12] is restricted to the two torsion angles defined by C-atoms only. To make use of the independence between all emerging Markov chains, all Markov chains on each level of the hierarchy run in parallel. For the initial sampling we use HMC [2,6] whereas the bridge densities are sampled with *adaptive temperature* HMC [10] (ATHMC), an enhancement of HMC where a bridge density between two adjacent temperatures is sampled by adapting the temperature of each HMC step according to the potential energy.

We illustrate UCMC by a three level hierarchy at temperatures 400 K, 200 K and 100 K corresponding to inverse temperatures β_0, β_1 and $\beta_2 = \beta_*$, respectively. The initial sampling is started at 400 K with state space $A_0 = \Omega$, a temperature at which HMC overcomes all conformational barriers. Yet, if we perform a clustering we observe the presence of metastable sets. In Table 1 the eigenvalues of the hereby discretized propagation operator $P(\beta_0)$ are given.

Table 1. Eigenvalues of the discretized operator $P(\beta)$ at different temperatures. In addition to the initial sampling at 400 K, we also give the spectra from long simulation runs of 2×10^5 steps at 200 K and 100 K. Note, that the spectrum at 200 K clearly indicates the effect of insufficient sampling due to increased metastability, which get even worse at 100 K.

	400 K	200 K	100 K
λ_1	1.0000	1.0000	1.0000
λ_2	0.9935	0.9998	0.9999
λ_3	0.9925	0.9997	-0.2455
λ_4	0.9914	0.9997	-0.2206
λ_5	0.9913	0.9994	0.2161
λ_6	0.9859	0.9992	-0.1878
λ_7	0.9855	0.2425	0.1766
λ_8	0.9254	0.1919	0.1716
λ_9	0.9009	0.1541	0.1670
λ_{10}	0.3380	-0.1494	0.1499

Figure 7 shows the bridge samplings in 7 metastable sets $A_1, \ldots, A_7$ obtained from the identification algorithm [5], where the spectral gap was detected between $\lambda_7 = 0.9855$ and $\lambda_8 = 0.9254$. Note that the eigenvalues $\lambda_8 = 0.9254$ and $\lambda_9 = 0.9009$ together with the extremely large spectral gap beyond λ_9 already indicates the emergence of two further metastable sets. Actually, these two metastable sets got uncoupled on the next level of the hierarchy (see Figure 7). This gives us the sets $A_8, \ldots, A_{16}$, in which again bridge densities were restarted.

Having completed the uncoupling step, an approximation $\tilde{C} \in \mathrm{Mat}_{17\times 17}$ of the coupling matrix C by data analysis was set up, which gave us the

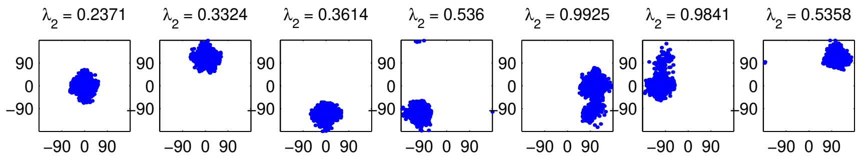

Fig. 7. Initial uncoupling of A_0 into 7 metastable sets $A_1, \ldots, A_7$ as a projection of the two torsion angles. The data points shown are the sampling points from the restricted Markov chains corresponding to the bridge densities between 400 K and 200 K, whereas the number above each set denotes the 2nd eigenvalue of the corresponding propagation operator. The sets A_5 and A_6 which possess 2nd eigenvalues close to one get further uncoupled into two sets each—their 3rd eigenvalue is 0.3591 and 0.3395, respectively—for the next hierarchical level.

invariant density $\tilde{\pi}$. Exemplarily for the coupling part of the algorithm, in Fig. 8 reweightings of the bridge density f_{04} to the densities f_0 and f_4 are shown. Finally, using (19) the restriction $\tilde{\pi}^* = (\tilde{\pi}_9, \ldots, \tilde{\pi}_{17})$ of the invariant density $\tilde{\pi}$ of $\tilde{C}$ enables us to compute any desired expectation values w.r.t. $f(\beta_*)$.

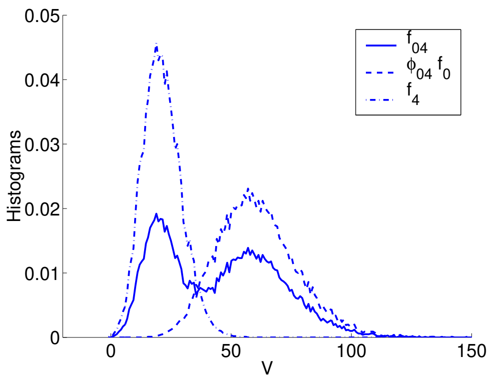

Fig. 8. Reweighting of the bridge density f_{04} to the densities $\phi_{04} f_0$ and f_4. The reweighted densities show the typical gaussian shape. Additionally, the overlap between f_{04} with each of the reweighted densities indicates a statistical reliable computation of the ratio of normalizing constant $Z_{\phi_{40}\hat{f}_4}/Z_{\phi_{04}\hat{f}_0}$.

Due to a symmetry in the Hamiltonian we can derive symmetric sets in the state space by discretizing each torsion angle into three boxes of 120°. If we denote by s_{ij} the probability to be in the corresponding $120° \times 120°$-box,

$S = (s_{ij})$ possesses the following symmetric structure:

$$S = \begin{pmatrix} d & b & c \\ b & a & b \\ c & b & d \end{pmatrix}.$$

Actually, these 9 sets also reflects roughly the 9 metastable sets $A_8, \ldots, A_{16}$ found by UCMC. Using (19) to get an approximation $\tilde{S}$ of S at 100 K yields

$$\tilde{S}_{\mathrm{UCMC}} = \begin{pmatrix} 3.97 \times 10^{-7} & 0.0123 & 4.09 \times 10^{-4} \\ 0.0118 & 0.9542 & 0.0117 \\ 2.50 \times 10^{-4} & 0.0095 & 1.57 \times 10^{-7} \end{pmatrix},$$

which is in good agreement with S, whereas the direct HMC simulation at 200 K and 100 K from Table 1 results in

$$\tilde{S}(\beta_1) = \begin{pmatrix} 0 & 0.098 & 0 \\ 0.168 & 0.558 & 0.104 \\ 8.17 \times 10^{-3} & 0.065 & 0 \end{pmatrix} \quad \text{and} \quad \tilde{S}(\beta_*) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.601 & 0 \\ 0 & 0.399 & 0 \end{pmatrix},$$

respectively, that once more demonstrates the insufficient samplings already observed in the spectra from Table 1.

Regarding the small probabilities to be within the metastable sets $\tilde{s}_{11}, \tilde{s}_{33}, \tilde{s}_{13}$ and $\tilde{s}_{31}$ of n-pentane, one appropriate strategy to bound the number of metastable sets for larger molecules would be to skip identified metastable sets during annealing whenever the probability to be in such a set is below a given threshold. However, we have seen that an uncoupling into all local minima only occurs if we proceed towards very low temperatures. For larger molecules one is normally interested in temperatures around 300 K where one would expect a reasonably small number of metastable sets composed of less metastable subsets.

Conclusion

In this article, we have worked out details of a hierarchical uncoupling-coupling Monte Carlo (UCMC) method. In a first step, we investigated the question of how to uncouple a reversible Markov chain hierarchically by decomposing its state space into metastable sets. In a second step, we illustrated the algorithmic scheme by applying it to the simple n-pentane problem, where everything is well-understood and which therefore can be used for testing. In this example, UCMC has shown to allow fast samplings in metastable sets as well as a correct reweighting. For the decomposition of the state space in the example, UCMC has been combined with self-organized box maps as presented also in this volume.

As a next step, we aim at applying UCMC to larger molecules. However, before we can take this step, we still need to gain a deeper understanding of the combination of UCMC with decompositions in higher dimensional state spaces.

Acknowledgment

This work has partially been supported by the European Regional Development Fund (ERDF).

References

1. B. J. Berne and J. E. Straub. Novel methods of sampling phase space in the simulation of biological systems. *Curr. Opinion in Struct. Biol.*, 7:181–189, 1997.
2. A. Brass, B. J. Pendleton, Y. Chen, and B. Robson. Hybrid Monte Carlo simulations theory and initial comparison with molecular dynamics. *Biopolymers*, 33:1307–1315, 1993.
3. B. W. Church, A. Ulitsky, and D. Shalloway. Macrostate dissection of thermodynamic Monte Carlo integrals. In D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar, editors, *Monte Carlo Methods in Chemical Physics*, volume 105 of *Advances in Chemical Physics*. J. Wiley & Sons, New York, 1999.
4. M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM J. Num. Anal.*, 36(2):491–515, 1999.
5. P. Deuflhard, W. Huisinga, A. Fischer, and C. Schütte. Identification of almost invariant aggregates in reversible nearly uncoupled Markov chains. *Lin. Alg. Appl.*, 315:39–59, 2000.
6. S. Duane, A. D. Kennedy, B. J. Pendleton, and D. Roweth. Hybrid Monte Carlo. *Phys. Lett. B*, 195(2):216–222, 1987.
7. D. M. Ferguson, J. I. Siepmann, and D. G. Truhlar, editors. *Monte Carlo Methods in Chemical Physics*, volume 105 of *Advances in Chemical Physics*. Wiley, New York, 1999.
8. A. M. Ferrenberg and R. H. Swendsen. Optimized Monte Carlo data analysis. *Phys. Rev. Lett.*, 63(12):1195–1197, 1989.
9. A. Fischer. An uncoupling-coupling technique for Markov chain Monte Carlo methods. Available as ZIB-Report 00-04 via `http://www.zib.de/bib/pub/pw`, 2000.
10. A. Fischer, F. Cordes, and C. Schütte. Hybrid Monte Carlo with adaptive temperature in mixed-canonical ensemble: Efficient conformational analysis of RNA. *J. Comput. Chem.*, 19(15):1689–1697, 1998.
11. H. Frauenfelder, S. G. Sligar, and P. G. Wolynes. The energy landscapes and motions of proteins. *Science*, 254:1598–1603, 1991.
12. T. Galliat, P. Deuflhard, R. Roitzsch, and F. Cordes. Automatic identification of metastable conformations via self-organized neural networks. Available as ZIB-Report 00-51 via `http://www.zib.de/bib/pub/pw`, 2000.
13. A. Gelman. Inference and monitoring convergence. In W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors, *Markov Chain Monte Carlo in Practice*, pages 131–143. Chapman & Hall, 1996.

14. A. Gelman and X.-L. Meng. Simulating normalizing constants: From importance sampling to bridge sampling to path sampling. *Statist. Sci.*, 13(2):163–185, 1998.
15. A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences (with discussion). *Statist. Sci.*, 7(4):457–511, 1992.
16. T. A. Halgren. Merck molecular force field I-V. *J. Comp. Chem.*, 17(5,6):490–641, 1996.
17. U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger. Molecular dynamics, Langevin and hybrid Monte Carlo simulations in a multicanonical ensemble. *Chem. Phys. Lett.*, 259:321–330, 1996.
18. Y. Iba. Extended ensemble Monte Carlo. ISM Research Memo. No.777, available via `http://xxx.lanl.gov/abs/cond-mat/0012323`, 2000.
19. P. J. M. v. Laarhoven and E. H. L. Aarts. *Simulated Annealing: Theory and Applications*. Reidel, Dordrecht, 1987.
20. E. Marinari and G. Parisi. Simulated tempering: a new Monte Carlo scheme. *Europhys. Lett.*, 19(6):451–458, 1992.
21. X.-L. Meng and W. H. Wong. Simulating ratios of normalizing constants via a simple identity: A theoretical exploration. *Statist. Sinica*, 6:831–860, 1996.
22. C. D. Meyer. Stochastic complementation, uncoupling Markov chains, and the theory of nearly reducible systems. *SIAM Rev.*, 31:240–272, 1989.
23. S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer, Berlin, 1993.
24. J.-P. Ryckaert and A. Bellemans. Molecular dynamics of liquid alkanes. *Faraday Discuss.*, 66:95–106, 1978.
25. C. Schütte. *Conformational Dynamics: Modelling, Theory, Algorithm, and Application to Biomolecules*. Habilitation Thesis, Fachbereich Mathematik und Informatik, Freie Universität Berlin, 1998. Available as ZIB-Report SC-99-18 via `http://www.zib.de/bib/pub/pw`.
26. C. Schütte, A. Fischer, W. Huisinga, and P. Deuflhard. A direct approach to conformational dynamics based on hybrid Monte Carlo. *J. Comput. Phys.*, 151:146–168, 1999.
27. C. Schütte, W. Huisinga, and P. Deuflhard. Transfer operator approach to conformational dynamics in biomolecular systems. In B. Fiedler, editor, *Ergodic Theory, Analysis, and Efficient Simulation of Dynamical Systems*. Springer, 2001.
28. H. A. Simon and A. Ando. Aggregation of variables in dynamic systems. *Econometrica*, 29(2):111–138, 1961.
29. L. Tierney. Markov chains for exploring posterior distributions (with discussion). *Ann. Statist.*, 22:1701–1762, 1994.
30. F. Yaşar, T. Çelik, B. A. Berg, and H. Meirovitch. Multicanonical procedure for continuum peptide models. *J. Comput. Chem.*, 21(14):1251–1261, 2000.