

EXTENDED ENSEMBLE MONTE CARLO*

YUKITO IBA

The Institute of Statistical Mathematics

E-mail: iba@ism.ac.jp

Received 13 April 2001

Revised 6 June 2001

“Extended Ensemble Monte Carlo” is a generic term that indicates a set of algorithms, which are now popular in a variety of fields in physics and statistical information processing. Exchange Monte Carlo (Metropolis-Coupled Chain, Parallel Tempering), Simulated Tempering (Expanded Ensemble Monte Carlo) and Multicanonical Monte Carlo (Adaptive Umbrella Sampling) are typical members of this family. Here, we give a cross-disciplinary survey of these algorithms with special emphasis on the great flexibility of the underlying idea. In Sec. 2, we discuss the background of Extended Ensemble Monte Carlo. In Secs. 3, 4 and 5, three types of the algorithms, i.e., Exchange Monte Carlo, Simulated Tempering, Multicanonical Monte Carlo, are introduced. In Sec. 6, we give an introduction to Replica Monte Carlo algorithm by Swendsen and Wang. Strategies for the construction of special-purpose extended ensembles are discussed in Sec. 7. We stress that an extension is not necessarily restricted to the space of energy or temperature. Even unphysical (unrealizable) configurations can be included in the ensemble, if the resultant fast mixing of the Markov chain offsets the increasing cost of the sampling procedure. Multivariate (multicomponent) extensions are also useful in many examples. In Sec. 8, we give a survey on extended ensembles with a state space whose dimensionality is dynamically varying. In the appendix, we discuss advantages and disadvantages of three types of extended ensemble algorithms.

Keywords: Extended Ensemble; Exchange Monte Carlo; Simulated Tempering; Multicanonical Monte Carlo; Replica Monte Carlo; Complexity Ladder; Bridge; Multivariate Extension.

1. Introduction

In this paper, we will give a survey on **Extended Ensemble Monte Carlo** algorithms,^a which are useful tools in computational physics and in the fields of

*IJMPC has now taken over the book series *Annual Reviews of Computational Physics* published before also by World Scientific.

^aWe choose “Extended Ensemble Monte Carlo” as a generic term to represent a family of algorithms, which we want to discuss here, e.g., Exchange Monte Carlo, Multicanonical Monte Carlo, etc. Another term, **Generalized Ensemble Monte Carlo**, is used by some authors. The term **Expanded Ensemble Monte Carlo**, which we use here in a more restricted meaning, might also be used in the generic meaning. However, the original definition¹ of Expanded Ensemble Monte Carlo seems not to cover Exchange Monte Carlo.

statistical information processing. Well-known algorithms in this family are **Exchange Monte Carlo** (Metropolis-Coupled Chain, Parallel Tempering),^{2–8} **Simulated Tempering** (Expanded Ensemble Monte Carlo),^{1,9} and **Multicanonical Monte Carlo** (Adaptive Umbrella Sampling).^{10–14} These approaches are characterized by *modification of ensembles* sampled by the algorithm. In this respects, they contrast with other attempts to overcome the limitation of conventional Dynamical Monte Carlo, i.e., *improved dynamics* that preserve original ensembles^{15,16} and *improved algorithms* that maintain original dynamics.¹⁷

These algorithms are useful for the studies of stochastic models in various fields of physics, e.g., spin models (Potts models,^{12,18} spin glass models,^{7,13,19–30} random field models,⁹ quantum spin models³¹), polymer models (lattice polymers,^{8,32} diblock copolymer,³³ lattice heteropolymers/proteins,^{34–38} off-lattice polymers,^{32,39–41} realistic protein/polypeptide models^{38,42–49}), models of molecules in vacuum or water,^{10,11,50} hard core fluid (solid),^{1,51–54} Lennard–Jones fluid,^{1,55,56} models of aqueous solution,^{57,58} Lennard–Jones clusters,⁵⁹ lattice gauge models,^{60,61} models of quantum gravity.⁶² They are also successfully used in statistical inference^{5,63,64} and combinatorics.⁶⁵ Our aim here is, however, not to give a list of references on this subject. Instead, we want to discuss basic ideas behind algorithms and show relations and differences among the algorithms.

An important issue in this paper is the great flexibility of the idea of extended ensemble, i.e., an extension is not necessary restricted to the space of energy or temperature. In fact, extensions in the space of arbitrary macroscopic variables are possible and useful (note that some authors already noticed this flexibility in the early stage of the development of extended ensemble methods, e.g., Lyubartsev *et al.*,¹ Kerler and Weber.¹⁸ See also the studies on Adaptive Umbrella Sampling algorithm^{10,11}). As we will discuss in Sec. 7, even unphysical (unrealizable) configurations can be included in the ensemble, if the fast mixing of the Markov chain offsets the increasing cost of the sampling procedure. Such an observation enables us a number of “special purpose” algorithms, which depend on specific properties of the model and the computational aims.

We are also careful to cross-disciplinary nature of this subject. The physicists are no more the only major users of dynamical Monte Carlo algorithms. Many algorithms have also been developed recently in various areas^{66–68} that are often overlooked by physicists. For example, Exchange Monte Carlo (Metropolis-Coupled Chain, Parallel Tempering) is independently discovered by computer scientists working for the fifth-generation computer project,^{2,3} a statistician,^{4,5} physicists,⁷ and the author.⁶

In Sec. 2, we discuss the background of Extended Ensemble Monte Carlo. In the following three sections, Secs. 3, 4 and 5, three types of algorithms, i.e., Exchange Monte Carlo, Simulated Tempering, Multicanonical Monte Carlo, are introduced. In Sec. 6, we give an introduction to Replica Monte Carlo algorithm by Swendsen and Wang,^{28,29} which interpolates Extended Ensemble Monte Carlo and Cluster Monte Carlo algorithms. Strategies for the construction of special-purpose extended

ensembles are discussed in Sec. 7. In Sec. 8, extended ensembles with a state space whose dimensionality is dynamically varying are discussed. In the appendix, we compare three types of extended ensemble algorithms and discuss their advantages and disadvantages.

This paper was originally written as a part of the Ph.D. thesis by the author, and then rewritten as an independent review paper. When I was writing the manuscript, I discovered several interesting surveys on this subject. For example, a lecture note by Marinari⁶⁹ gives a survey on this field including Exchange Monte Carlo. The book⁶⁸ provides a cross-disciplinary survey on the recent progress of Monte Carlo methods. Specifically, Berg¹⁴ in Ref. 68 gives a recent review of Multicanonical Monte Carlo and related topics. A review on the calculation of partition functions (normalizing constants) by thermodynamic integration and/or Extended Ensemble Monte Carlo from the viewpoint of statisticians is available in Ref. 70. Now, there are increasing references in this field, but I hope that this review gives fresh perspectives both for beginners and experts in this field.

2. From Natural Ensemble to Artificial Ensemble

Dynamical Monte Carlo algorithms are useful tools for sampling from non-Gaussian, highly multivariate distributions. They, however, often suffer from slow mixing of the Markov chains, or, in terms of physics, *slow relaxation*. Slow relaxation reduces the effective number of samples and sometimes leads to wrong results sensitive to initial states of the Markov chain. There are several different situations that lead to slow relaxation: (1) “**Critical slowing down**” near second order phase transition points; (2) “**Nucleation**” associated with first order phase transitions; (3) **Trapping in metastable states around local minima** in models with rugged energy landscapes. Difficulties of the category (3) are often encountered with “random frustrated systems” such as models of spin glasses, interacting spins in random fields, and heteropolymers. Slow mixing also appears in complex statistical inference problems, where models (i.e., likelihoods or priors, or both) are highly non-Gaussian.

In the period 1985–1995, a powerful strategy to overcome the difficulties of categories (2) and (3), Extended Ensemble Monte Carlo algorithms, has been introduced.^b Simulated Tempering (Expanded Ensembles), Exchange Monte Carlo (Metropolis-Coupled Chain, Parallel Tempering), Multicanonical Monte Carlo (Adaptive Umbrella Sampling) are well-known members of this family. While conventional Dynamical Monte Carlo algorithms simulate a Markov chain, whose invariant distribution is a given target distribution (e.g., a Gibbs distribution for statistical physics and a posterior distribution for Bayesian inference), Extended

^bIt is believed that they are not useful to fight against difficulties of the category (1), critical slowing down, to which Cluster Monte Carlo algorithms¹⁵ and Accelerated Hybrid algorithms¹⁶ are successfully applied.

Ensemble Monte Carlo algorithms sample from artificial ensembles that are constructed as extensions or compositions of the original ensembles. Fast mixing of the Markov chain in higher temperature (energy, etc.) components of the artificial ensembles greatly facilitate the mixing in other components. Averages over the original ensemble are calculated by marginalization (in Exchange Monte Carlo), conditional sampling (in Simulated Tempering), or, a **reweighting** procedure (in Multicanonical Monte Carlo). As we will represent in the next section, Sec. 3, they can be interpreted as extensions of Simulated Annealing algorithms for finite temperature simulations.

Acceleration of the relaxation is not the only aim of Extended Ensemble Monte Carlo. There are at least two more motivations for the introduction of artificial ensembles.

2.1. *Calculation of integrals or summations*

Calculation of multivariate integrals or multiple summations (e.g., free energy difference, marginal likelihood difference) is important in many applications, but cannot be directly done with conventional Dynamical Monte Carlo algorithms. Extended Ensemble Monte Carlo methods are particularly suitable for the calculations of these quantities. As we will show in the following sections, Exchange Monte Carlo naturally gives samples from a set of distributions necessary for thermodynamic integration. In Multicanonical Monte Carlo, integrals are calculated by a reweighting formula. In both cases, we can enjoy the advantage of fast mixing with extended ensembles without additional computational resources. On the other hand, there are cases where the use of an extended ensemble is essential for the calculation of integrals, as we will discuss in the section on Multicanonical Monte Carlo, Sec. 5.

2.2. *Efficient sampling of “rare events”*

Extended Ensemble Monte Carlo methods are also suitable for the calculation of the frequencies of configurations with small probability in a given ensemble. For example, we can use them for the calculation of the free energy surface as a function of one- or two-macroscopic variables. The relative error of the computation is proportional to $1/\sqrt{M}$, where M is the frequency of independent visits to configurations with a set of values of the macroscopic variables. Thus, with a conventional Monte Carlo algorithm, the histogram in the log-scale contains large noise in low probability regions. On the other hand, we can compute free energy surface with much more uniform accuracy with Extended Ensemble Monte Carlo methods. We will give some comments on the implementation of this idea in Sec. 7.

Calculation of free energy difference and free energy surface by artificial ensembles is especially stressed in the studies on **Expanded Ensemble Monte Carlo**^{1,32,57,58} and **Adaptive Umbrella Sampling**.^{10,11,50} In this context, Extended Ensemble Monte Carlo is considered as a descendant of **Umbrella Sampling** algorithms for the calculation of free energy, which was introduced by Torrie

and Valleau⁷¹ in 1970s. In the original form of Umbrella Sampling algorithms, artificial ensembles were also used, but systematic ways for the construction of ensembles had not been implemented. Implementation of such a procedure characterizes Extended Ensemble Monte Carlo algorithms developed later.

An important feature shared by Extended Ensemble Monte Carlo and Umbrella Sampling is that they are *not* designed for the direct simulation of *natural phenomena*. Although conventional Dynamical Monte Carlo methods themselves are something between *simulation of physics* and *numerical methods*, they are still strongly motivated by simulation of physical dynamics — this is a reason why the use of single-spin flip algorithms that directly sample Gibbs distributions has persisted for half a century. On the other hand, Extended Ensemble Monte Carlo are free from such restrictions. It does not mean that insight into the *physics* (or *mathematics*, *statistics*, etc.) of the problems are unnecessary. On the contrary, they are essential for the construction of artificial ensembles for efficient computation. Moreover, the performance of simulation with an extended ensemble can be regarded as a measure of our understanding of the underlying physical phenomena. Our belief is:

If an algorithm based on a physical picture is efficient, it supports the validity of the picture; if we understand the physics, we can write an efficient algorithm.

This manifestation is also applicable to other “artificial” algorithms, say, Cluster Monte Carlo¹⁵ or Hybrid Monte Carlo with acceleration.¹⁶

3. Exchange Monte Carlo

A useful way for the search of ground states of complex models is **Simulated Annealing** algorithm.^{72,73} The term “annealing” indicates that we start simulation at a high temperature and gradually decrease temperature to zero. Then, there are more chances of escaping from shallow local minima and reaching a deep local minimum, or, if we are lucky enough, attaining the global minimum of the energy function. Instead of the inverse temperature β , we can use an arbitrary parameter λ to interpolate an “easy” problem (a problem with a smooth landscape) to the original problem (a problem with a rugged landscape).

Simulated Annealing is, however, no more than a prescription for *optimization*, i.e., it is useful for the computation of ground states of a system but does not exactly give finite temperature properties of the system. Then, a question arises: *can we extend it for the sampling from multivariate distributions*, e.g., sampling from Gibbs distribution at finite temperatures? A naive method to achieve this purpose is to start the simulation at a high temperature and gradually decrease the temperature to the target temperature and keep it constant through the rest of the simulation, where we measure the required quantities. This method has, however, a fatal weak point that the annealing is useful for escaping from shallow local minima but *not* for accelerating jumping between deep metastable states.

To facilitate such jumping, we should not monotonically decrease the temperature but make it “*up and down*” alternately. However, at a first glance, any attempt to change the temperature by external programs, periodic or stochastic, seems to violate the *detailed balance condition*, which is a foundation of Dynamical Monte Carlo algorithms.

Here, we introduce **Exchange Monte Carlo algorithm**^{7,20} (**Metropolis-Coupled Chain algorithm**,^{4,5} **Time-homogeneous Parallel Annealing**² (see also Ref. 3), **Multiple Markov Chain algorithm**,⁸ **Parallel Tempering**⁶⁹) as a solution of the dilemma. The algorithm seems to have been independently discovered by several different groups of authors in the period of 1990–1994^{2,4,6,7} and, as a result, has a variety of different names.^c

Consider a set of the distributions $\{p_k(\mathbf{x})\}$ with different parameters^d $\{\lambda_k\}$, $k = 1, \dots, K$ and assume that the parameters are ordered as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_K$. An example is a family of the Gibbs distributions defined with inverse temperatures $\{\lambda_k\} = \{\beta_k\}$,

$$p_k(\mathbf{x}) = \frac{\exp(-\beta_k E(\mathbf{x}))}{Z(\beta_k)}, \quad (1)$$

where $Z(\beta_k)$ is the partition function of the system. If we denote the variables of k th system (k th *replica*) as $\mathbf{x}_k$, the simultaneous distribution $\tilde{p}$ of $\{\mathbf{x}_k\}$ is written as:

$$\tilde{p}(\{\mathbf{x}_k\}) = \prod_k p_k(\mathbf{x}_k). \quad (2)$$

We introduce two types of update with which the simultaneous distribution $\tilde{p}$ of Eq. (2) is invariant. First, we consider conventional updates in a replica k that satisfy the detailed balance condition for the corresponding factor $p_k(\mathbf{x}_k)$, e.g., local spin flips in a replica. In addition, we define a *replica exchange* between replicas, which have neighboring values λ_k and λ_{k+1} of the parameter. In this step, candidates of new configurations $\tilde{\mathbf{x}}_k$ and $\tilde{\mathbf{x}}_{k+1}$ are defined by the exchange of configurations of the replicas: $\tilde{\mathbf{x}}_k = \mathbf{x}_{k+1}$ and $\tilde{\mathbf{x}}_{k+1} = \mathbf{x}_k$. If we give the acceptance probability of the replica exchange flip by $\max\{1, r\}$ with r defined by:

$$r = \frac{p_k(\tilde{\mathbf{x}}_k)p_{k+1}(\tilde{\mathbf{x}}_{k+1})}{p_k(\mathbf{x}_k)p_{k+1}(\mathbf{x}_{k+1})} = \frac{p_k(\mathbf{x}_{k+1})p_{k+1}(\mathbf{x}_k)}{p_k(\mathbf{x}_k)p_{k+1}(\mathbf{x}_{k+1})}, \quad (3)$$

^c**J-Walk**^{74,75} algorithm also uses multiple copies of systems. In this method, however, the propagation of configurations is unidirectional, i.e., restricted to that from a higher temperature to a lower temperature. As a result, the J-Walk algorithm does not exactly produce samples from the original distribution, unless the correlation between samples from the auxiliary simulations at a high temperature is negligible (or erased by some off-line manipulation). Another algorithm closely related to Exchange Monte Carlo is **Replica Monte Carlo**²⁸ developed by Swendsen and Wang in 1986, which we will discuss in Sec. 6.

^dA note for the applications to statistical information processing: the “parameter” here often corresponds to a “hyperparameter” when hierarchical Bayesian models are treated by the algorithm.

the simultaneous distribution $\tilde{p}$ of Eq. (2) is invariant under the transition. That is, the detailed balance condition for the simultaneous distribution $\tilde{p}$ is satisfied. When $\{p_k(\mathbf{x}_k)\}$ is a family of the Gibbs distributions, Eq. (1), defined with inverse temperatures $\{\beta_k\}$, we can express the ratio r as:

$$r = \exp((\beta_k - \beta_{k+1}) \cdot (E(\mathbf{x}_k) - E(\mathbf{x}_{k+1}))). \quad (4)$$

The averages taken over each factor $p_k(\mathbf{x}_k)$ *exactly* reproduce the desired averages at the value $\lambda = \lambda_k$ of the parameter, because the transitions defined by the algorithm do not change the simultaneous distribution, Eq. (2). On the other hand, the states of the replicas are effectively propagated from high temperatures to lower temperatures through replica exchanges and the mixing of Markov chain is facilitated by the fast relaxation at higher temperatures (or, in general, at the values of the parameter λ with which the mixing of the Markov chain is fast and the entropy of the distribution is large).

A problem is the choice of the points $\{\beta_k\}$ or $\{\lambda_k\}$. A naive way is that uses a set of points with *regular spacing* that contains sufficiently high temperatures (or, the values of the parameter, where the mixing is fast and the entropy of the distribution is high). Of course, we should use the interval $|\lambda_{k+1} - \lambda_k|$ that gives a sufficiently large frequency of replica exchange. A more sophisticated way is to allow points of variable spacing and choose them as the exchange rates are uniform in the prescribed range of the temperature (parameter). Although the naive method and some tuning *by hand* is often enough for practical applications, it is instructive to see how we can determine the spacing with this criterion⁷ (discussions on related subjects from the viewpoint of computational statistics are found in Refs. 70 and 76). From Eq. (3), the average $\overline{\log r}$ of the logarithm of the ratio r that determines the exchange rate of the neighboring replicas is:

$$\overline{\log r} = \sum_{\mathbf{x}_k} \sum_{\mathbf{x}_{k+1}} p_k(\mathbf{x}_k) p_{k+1}(\mathbf{x}_{k+1}) \cdot \log \left\{ \frac{p_k(\mathbf{x}_{k+1}) p_{k+1}(\mathbf{x}_k)}{p_k(\mathbf{x}_k) p_{k+1}(\mathbf{x}_{k+1})} \right\}, \quad (5)$$

which is expressed as:

$$\overline{\log r} = - \left\{ \sum_{\mathbf{x}} p_k(\mathbf{x}) \log \frac{p_k(\mathbf{x})}{p_{k+1}(\mathbf{x})} + \sum_{\mathbf{x}} p_{k+1}(\mathbf{x}) \log \frac{p_{k+1}(\mathbf{x})}{p_k(\mathbf{x})} \right\}. \quad (6)$$

The expression in the braces $\{ \}$ is a “symmetrized” Kullback–Leibler divergence $D(p_k \| p_{k+1}) + D(p_{k+1} \| p_k)$ between p_k and p_{k+1} . When $\lambda_k \sim \lambda_{k+1}$, it is approximated by:

$$\overline{\log r} \sim -I(\lambda_k) \cdot (\lambda_{k+1} - \lambda_k)^2, \quad (7)$$

with

$$I(\lambda_k) = - \sum_{\mathbf{x}} p_k(\mathbf{x}) \frac{\partial^2 \log p_{\lambda}(\mathbf{x})}{\partial \lambda^2} \bigg|_{\lambda=\lambda_k} = - \left\langle \frac{\partial^2 \log p_{\lambda}(\mathbf{x})}{\partial \lambda^2} \right\rangle_{\lambda_k}, \quad (8)$$

where $\langle \cdots \rangle_{\lambda_k}$ is average over the distribution $p_k(\mathbf{x})$. $I(\lambda_k)$ is called **Fisher information** in statistics. For the case of Gibbs distributions, Eq. (1), it is related to **susceptibility** $\sigma_E^2 = -d\langle E \rangle_\beta / d\beta$ to inverse temperature β ,^e and **specific heat** $C = (\beta^2/N)\sigma_E^2$ per system size N as:

$$I(\beta_k) = \left. \frac{\partial^2 \log Z(\beta)}{\partial \beta^2} \right|_{\beta=\beta_k} = \sigma_E^2 = N \cdot \frac{C}{\beta_k^2}. \quad (9)$$

Thus, the interval $|\lambda_{k+1} - \lambda_k|$ that gives reasonable and uniform replica exchange rate is given by:

$$I(\lambda_k) \cdot |\lambda_{k+1} - \lambda_k|^2 \sim 1. \quad (10)$$

From the condition, Eq. (10), we have an expression of the density $Q(\lambda)$ of points $\{\lambda_k\}$:

$$Q(\lambda) \propto \sqrt{I(\lambda)}, \quad (11)$$

which is also written as:

$$Q(\beta) \propto \sqrt{\sigma_E^2} = \sqrt{\frac{N \cdot C}{\beta^2}}, \quad (12)$$

for Gibbs distributions. The expression in Eq. (12) gives two important results. First, it shows that the number of replicas that is required for Exchange Monte Carlo increases with $\sqrt{N}$ when the specific heat is constant. It is easy to understand the result when we note that the exchange is caused by fluctuation of the energies of the replicas. Another observation from Eq. (12) is that a larger number of replicas is required in the region where σ_E^2 or $I(\lambda)$ takes larger values, say, near the critical points of second order transitions.

The rest of the problem is how to determine the density without preliminary knowledge on the specific heat or Fisher information. The answer²⁰ is that we can most easily do it through step-by-step tuning of the number and/or positions of the points $\{\beta_k\}$ or $\{\lambda_k\}$. After finishing the tuning process, we perform the simulation for the sampling of desired quantities with fixed values of all simulation parameters. This idea is an example of a central strategy in Extended Ensemble Monte Carlo algorithms:

Learn (or tune) the optimal value of parameters of the algorithm by a step-by-step manner in preliminary runs.

This strategy is more important in Simulated Tempering and Multicanonical Monte Carlo discussed in the following sections.

Finally, we discuss some concrete results obtained by Exchange Monte Carlo and show how it works in real problems in physics. A field where Exchange Monte

^eThe notation σ_E^2 indicates that it coincides with the variance of the energy $\langle E^2 \rangle_\beta - \langle E \rangle_\beta^2$.

Carlo is effectively used in studies on spin glasses. By using Exchange Monte Carlo, we can explore large systems considerably below T_c ,^{7,22–24} typically, at $T \sim 0.7T_c$ and system size $\sim 16^3$ for the 3-dim $\pm J$ Ising spin glass model. With these calculations, as well as with the use of novel methods for the analysis of the data, we are approaching the nature of the spin-glass phase of models with short-ranged interaction, which is a long-standing query in this field. Exchange Monte Carlo is also successfully used for the study of spin glass models with continuous spins, e.g., 3-dim Heisenberg spin glass models. Hukushima and Kawamura²¹ reported a strong evidence of chiral glass transition for the model, as well as peculiar properties of the transition. Another, potentially important field of the application of Exchange Monte Carlo is simulation of protein models. An application to protein folding is already found in Hirose *et al.*³ (1992). Recent developments and applications to realistic peptide models are described in Refs. 43, 44, 77–80, as well as attempts to combine Exchange Monte Carlo with Multicanonical Monte Carlo.

4. Simulated Tempering

Here, we discuss **Simulated Tempering algorithm**,^{9,25,39} an algorithm closely related to Exchange Monte Carlo algorithm. A similar method called **Expanded Ensemble**^{1,32,57,58} was introduced by Lyubartsev *et al.* almost at the same time, whose main aim is the calculation of free energy. In this approach, the dilemma of temperature up-down and detailed balance is resolved by treating the temperature itself as a *dynamical variable* updated in Monte Carlo simulations. That is, we construct Markov chains whose state vector is a direct product (**original states, temperature**). Although the following arguments are easily generalized for an arbitrary family of distributions $\{p_\lambda(\mathbf{x})\}$ with a parameter λ ,¹ here and hereafter, we discuss a family of Gibbs distributions:

$$p(\mathbf{x}) = \frac{\exp(-\beta E(\mathbf{x}))}{Z(\beta)}, \quad (13)$$

with a parameter β , where $Z(\beta)$ is the partition function. When we treat the inverse temperature β as a dynamical variable, the distribution $p(\mathbf{x}, \beta)$ in the extended space $(\mathbf{x}, \beta)$ is represented as:

$$p(\mathbf{x}, \beta) \propto \exp(-\beta E(\mathbf{x}) + g(\beta)). \quad (14)$$

Here, we have introduced an arbitrary function $g(\beta)$ of β that controls the distribution of β . With this definition, it is not difficult to construct Markov chains to sample from $p(\mathbf{x}, \beta)$. We simply regard $-\beta E(\mathbf{x}) + g(\beta)$ as the energy of the extended system and simulate it with ordinary updates of $\mathbf{x}$ *plus* Metropolis update of the inverse temperature β . Here and hereafter, we often restrict the value of β to discrete values $\{\beta_k\}$. With this restriction, the function $g(\beta)$ is determined by the finite set of the values $\{g_k\} = \{g(\beta_k)\}$, $k = 1, \dots, K$. It is a convenient property for the implementation of the algorithm. Then, the distribution $p(\mathbf{x}, \beta)$ in Eq. (14)

is represented as:

$$p(\mathbf{x}, \beta_k) = \frac{\exp(-\beta_k E(\mathbf{x}))}{Z(\beta_k)} \cdot \tilde{\pi}_k, \quad (15)$$

$$\tilde{\pi}_k \propto \exp(g_k + \log Z(\beta_k)), \quad \sum_k \tilde{\pi}_k = 1. \quad (16)$$

Note that the $\log Z(\beta_k)$ term in Eq. (16) comes from the normalization constant (partition function) of each component.

If we use the samples of $\mathbf{x}$ at a value of $\beta = \beta_k$, we exactly recover the canonical average at $\beta = \beta_k$, because $p(\mathbf{x}, \beta)$ conditioned on β reduces to the original canonical distributions by its definition. The problem is the choice of the function $g(\beta)$. Without a proper choice of $g(\beta)$, the value of β gets stuck around an uncontrolled value and there is *no samples* available at the desired values of β . A naive choice $g_k \equiv 0$ usually gives unsatisfactory results. A reasonable way is to take

$$g_k = -\log Z(\beta_k). \quad (17)$$

With this choice, the marginal probability $\sum_{\mathbf{x}} p(\mathbf{x}, \beta_k) = \tilde{\pi}_k$ of β is the uniform distribution on a given set $\{\beta_k\}$ of β . This means that the temperature varies in a stochastic way in a prescribed range and the proportion of the time that it stays at a value of $\beta = \beta_k$ is independent of β_k in a sufficiently long run.

But, *how can we know the value of $\log Z(\beta)$ prior to the simulation?* Any algorithm that requires the values of $Z(\beta)$ as inputs appears unrealistic because $Z(\beta)$ is usually an unknown quantity that is computed through the Monte Carlo simulation. Here, we can use the idea of optimizing the parameters of simulation in the preliminary run. That is, the algorithm learns the optimal value of $\{g_k\}$ by the iteration of preliminary runs. Here, we will not discuss the way²⁵ of tuning further. Instead, we give an account of similar techniques for multicanonical algorithm in the next section, Sec. 5. Note that the optimal *spacing* of $\{\beta_k\}$ can also be estimated in the preliminary runs to enable sufficiently frequent change of β .

The analogy with Exchange Monte Carlo algorithm is clear. The change of parameter(s) β (or, in general, γ) in Simulated Tempering algorithm corresponds to the “replica exchange” procedure in Exchange Monte Carlo algorithm. The mixture distribution $\sum_{k=1}^K p(\mathbf{x}, \beta_k)$ in Simulated Tempering has a direct correspondence to the simultaneous distribution, Eq. (2), in Exchange Monte Carlo, when the set of inverse temperatures of replicas $\{\beta_k\}$ in Exchange Monte Carlo is the same as $\{\beta_k\}$ in Simulated Tempering. Formally, we can write the correspondence as:

$$\sum_{k=1}^K p(\mathbf{x}_1, \beta_k) \Leftrightarrow \sum_{\mathbf{x}_2, \mathbf{x}_3, \dots, \mathbf{x}_K} \sum_{s=0}^{K-1} p_1(\mathbf{x}_{P_s 1}) \cdot p_2(\mathbf{x}_{P_s 2}) \cdots p_K(\mathbf{x}_{P_s K}), \quad (18)$$

where P_s is a cyclic shift operator $k \rightarrow \text{mod}(k - 1 + s, K) + 1$ with a shift s . It is also easy to show that the rate of temperature flip in Simulated Tempering is the same order as the exchange rate in Exchange Monte Carlo with the same $\{\beta_k\}$.

In this sense, Exchange Monte Carlo algorithm is a parallel version of Simulated Tempering (thus, called “Parallel Tempering” by some authors).

In the continuum limit, which is better for conceptual arguments, the mixture distribution $\sum_k p(\mathbf{x}, \beta_k)$ in Eq. (18) is represented as^f:

$$\int p(\mathbf{x}, \beta) d\beta = \int \frac{\exp(-\beta E(\mathbf{x}))}{Z(\beta)} \cdot \tilde{\pi}(\beta) d\beta, \quad (19)$$

where

$$\tilde{\pi}(\beta) \propto \exp(g(\beta) + \log Z(\beta) + Q(\beta)). \quad (20)$$

Here, $Q(\beta)$ represent the relative number of the points $\{\beta_k\}$ between $\beta < \beta_k < \beta + d\beta$ in both of Exchange Monte Carlo and Simulated Tempering. Consider the case $g(\beta) = -\log Z(\beta)$ in Eq. (20). If we use $Q(\beta) \propto \sqrt{I(\beta)}$ that gives the uniform exchange rate for Exchange Monte Carlo and the uniform *temperature flip* rate for Simulated Tempering, the mixing distribution $\tilde{\pi}(\beta)$ coincides with **Jeffreys’ prior**⁸¹:

$$\int p(\mathbf{x}, \beta) d\beta \propto \int \frac{\exp(-\beta E(\mathbf{x}))}{Z(\beta)} \cdot \sqrt{I(\beta)} d\beta. \quad (21)$$

Simulated Tempering and related methods are successfully used for various problems in physics and statistics. For example, a random field Ising model on a 10^3 lattice at low temperatures is treated in the original paper of Marinari and Parisi.⁹ Studies^{32,33,40,41,57,82} based on the idea of expanded ensembles¹ will also be discussed in Secs. 7 and 8. Geyer and Thompson⁵ discussed an application to statistical inference on propagation of genes of rare recessive disease on a pedigree. The problem is computation of probability distributions of career status of genes over a large pedigree from observed data, for which conventional Dynamical Monte Carlo suffers from slow mixing and nonergodicity of dynamics. By using a version of Simulated Tempering (see Sec. 7 of this survey), they successfully treated problems that contain thousands of individuals.

In most situations, however, Exchange Monte Carlo is more convenient than Simulated Tempering. Some of the advantages of Exchange Monte Carlo are discussed in Sec. 9. An exception occurs when the dimension of the system is so large that it is impossible to store a number of replicas in the memory of our computer. In this case, Simulated Tempering has an obvious advantage. Simulated Tempering may also be useful for the conceptual studies on extended ensembles.

5. Multicanonical Monte Carlo

The third method is **Multicanonical Monte Carlo** algorithm.^{12–14,83} Methods based on a similar idea are also known as **Adaptive Umbrella Sampling**

^fThe notation $\tilde{\pi}(\beta)$ is motivated by the analogy to Bayesian statistics. In fact, the distribution $\tilde{\pi}(\beta)$ is at times called a “pseudo prior” by statisticians.⁵ It is formally regarded as a prior distribution for a (hyper)parameter β , but determined for the convenience of the computation.

algorithm.^{10,11,50} In their original forms, Multicanonical Monte Carlo deals with extensions in the space of the energy, while Adaptive Umbrella Sampling focused on extensions in the space of a reaction coordinate. Now, they are being merged and we can freely construct algorithms optimal for our purpose, which will be discussed in Sec. 7.

Unlike the algorithms discussed in the previous sections, multicanonical algorithm deals only with an exponential family of distributions. First, we discuss the case of a family of Gibbs distributions, Eq. (13), with different inverse temperatures β . The *density of state* $D(E)$ on the energy axis is defined so that the number of the state $\mathbf{x}$ satisfying $E < E(\mathbf{x}) < E + dE$ is $D(E)dE$. The partition function (the normalization constant) at inverse temperature β is written as:

$$Z(\beta) = \int \exp(-\beta E) D(E) dE. \quad (22)$$

Multicanonical algorithm is defined as Dynamical Monte Carlo sampling with the weight proportional to $D(E(\mathbf{x}))^{-1}$ instead of the original canonical weight $\exp(-\beta E(\mathbf{x}))$. The distribution defined with this weight is called **multicanonical ensemble**. With the definition of $D(E)$ and $D(E) \cdot D(E)^{-1} = 1$, it is easy to see that the marginal distribution of E becomes constant within the region where $D(E) \neq 0$, i.e., the energy E of the system takes the values in $E \sim E + dE$ with an equal chance in a long simulation.⁵ This results in a random walk in the energy space. It is similar to the random walk on the temperature axis in Simulated Tempering, but there is no explicit temperature variable in Multicanonical Monte Carlo. When we introduce **microcanonical ensemble** with an energy E_0 defined by:

$$p_{E_0}(\mathbf{x}) = \frac{\delta(E(\mathbf{x}) - E_0)}{D(E_0)}, \quad (23)$$

multicanonical ensemble $p_{\text{mul}}(\mathbf{x}) \propto 1/D(E(\mathbf{x}))$ is considered as the mixture

$$p_{\text{mul}}(\mathbf{x}) = \int dE_0 p_{E_0}(\mathbf{x}) \tilde{\pi}(E_0) \quad (24)$$

of microcanonical ensembles with the uniform pseudo prior $\tilde{\pi}(E_0) = \text{const.}$ for the energy. In this sense, multicanonical ensemble is a special case of Expanded Ensemble,¹ whose components are microcanonical distributions.

How can we recover the canonical averages from the simulation with the weight $D(E(\mathbf{x}))^{-1}$? A **reweighting formula**,

$$\langle A(\mathbf{x}) \rangle_\beta = \frac{\sum_j A(\mathbf{x}^j) \cdot \exp(-\beta E(\mathbf{x}^j)) \cdot D(E(\mathbf{x}^j))}{\sum_j \exp(-\beta E(\mathbf{x}^j)) \cdot D(E(\mathbf{x}^j))}, \quad (25)$$

⁵Note that $D(E)$ is a severely varying function of a macroscopic variable E and the choice of $D(E(\mathbf{x}))^{-1}$ as a weight severely penalized the appearance of the states $\mathbf{x}$ with a large value of $D(E(\mathbf{x}))$, which are usually nearly random “high temperature” configurations. Multicanonical sampling corresponds to random selection of $\mathbf{x}$ with a fixed value of E after random sampling of the energy E . The point is that it is very different from random sampling of $\mathbf{x}$, which gives almost surely a sample of large E .

gives an answer, which gives a method for the reconstruction of the canonical average $\langle A(\mathbf{x}) \rangle_\beta$ at β of an arbitrary quantity $A(\mathbf{x})$. Here, the summation $\sum_j$ is taken over the samples $\{\mathbf{x}^j\}$ generated by the simulation of a Markov chain, whose invariant densities is $D(E(\mathbf{x}))^{-1}$. The Eq. (25) means that each observation $A(\mathbf{x}^j)$ is multiplied by the factor $D(E(\mathbf{x}^j))$ that cancels the weight $D^{-1}(E(\mathbf{x}^j))$ used in the simulation and reweighted by the canonical weight $\exp(-\beta E(\mathbf{x}^j))$. We can also introduce a reweighting formula for the normalization constant (partition function) $Z(\beta)$ as:

$$\frac{Z(\beta)}{V} = \frac{\sum_j \exp(-\beta E(\mathbf{x}^j)) \cdot D(E(\mathbf{x}^j))}{\sum_j D(E(\mathbf{x}^j))}, \quad (26)$$

where V equals to the total volume of the configuration space or the total number of the configurations (e.g., 2^N for N binary variables). For earlier studies on *reweighting* in Dynamical Monte Carlo methods, see Refs. 84–86.

Now, we will discuss an essential part of the algorithm: *how to sample with the weight $D(E(\mathbf{x}))^{-1}$ without prior knowledge on $D(E)$* . This is a basic problem, because $D(E)$ is the kind of the quantities, which we want to calculate by the simulation, just as $Z(\beta)$ in Simulated Tempering algorithm. The answer is, again, step-by-step learning procedure, which we will discuss in detail here. Note that we do not need to know the value $D(E)$ exactly. An approximation $\tilde{D}(E)$ to $D(E)$ in a region $E_{\min} < E < E_{\max}$, with which we can safely apply the reweighting procedure, Eqs. (25) and (26), is enough for our purpose. Note also that multiplication of a constant factor to $\tilde{D}(E)$ does not change the result. Paying an attention to these remarks, we introduce an iterative procedure (**preliminary runs**) to approximate $D(E)$, starting from $D^0 \equiv \text{const.}$ (or some initial guess). Here and hereafter, we assume that the energy E takes discrete values $\{E_k\}$, ($k = 1, \dots, K$), and $\tilde{D}(E)$ are represented by the values $\{\tilde{D}_k\} = \{\tilde{D}(E_k)\}$.

- (1) Simulation and Histogram Construction: simulate a Markov chain with the weight $D^t(E(\mathbf{x}))^{-1}$ and record the frequencies h_k^t that the energy E takes the value E_k for each k (**Histogram Construction**). Here, we can use arbitrary dynamics with which the density $D^t(E(\mathbf{x}))^{-1}$ is invariant.
- (2) Update the weight: define new values of the weight by:

$$\frac{1}{D_k^{t+1}} := \frac{1}{D_k^t} \cdot \frac{1}{h_k^t + \epsilon}, \quad (27)$$

or, equivalently,

$$\log D_k^{t+1} := \log D_k^t + \log(h_k^t + \epsilon). \quad (28)$$

Here, ϵ is a constant, which is added to remove the divergence with $h_k^t = 0$ (i.e., no visit to $E = E_k$). For example, we can use $\epsilon = 1$.

- (3) Normalization:

$$\log D_k^{t+1} := \log D_k^{t+1} - \frac{1}{K} \sum_k \log D_k^{t+1}. \quad (29)$$

This normalization procedure is added for the convenience of monitoring convergence and not essential (adding a constant to all $\log D_k^{t+1}$ does not change the simulation).

(4) Set $t := t + 1$.

After we find an appreciate weight $\{\tilde{D}_k\}$ with the iteration of the preliminary runs, a **measurement run (production run)** is performed. In the measurement run, we collect samples with a fixed $\{\tilde{D}_k\}$ and apply reweighting formulae, Eqs. (25) and (26), for the calculation of canonical averages, where $D(E_k)$ is substituted for its approximation $\tilde{D}_k$.

This simple iterative procedure, sometimes referred to as (the learning stage of) the **entropic sampling** method⁸⁷ is sufficient for many practical problems. When the system is very large or has a continuous energy spectrum, we should replace histogram construction by a more sophisticated density estimation method. Another problem of the above-mentioned procedure is that it is sensitive to the fluctuation of frequencies of visits to E_k . Some authors proposed estimators of $\tilde{D}_k$ based on the ratio h_k^t/h_{k+1}^t of the neighboring frequencies (or the frequencies of the transition $E_k \rightarrow E_{k'}$,^{51,83}) for the improvement of the performance in the tuning stage. “Flat Histogram Monte Carlo” (see Refs. 88 and 89 and references therein) can also be considered as an efficient way to realize Multicanonical Ensemble, although it has a different origin⁹⁰ and its own perspective.

When we deals with systems with quenched disorders, we usually do not have prior knowledge of the upper and lower bounds of E . In such cases, we apply the iterative procedure assuming a sufficiently wide region of $E_{\min} < E < E_{\max}$. Then, we conclude there is no energy level at E_k , if $h_k = 0$ even with a sufficiently large value of the weight $1/D_k^t$ and with a long run of the simulation. In principle, we can neglect regions that do not contribute required canonical averages with Eq. (25), but should be careful to include a sufficiently high energy region (or, in general, a high entropy region) to facilitate the relaxation. Determination of lower energy bound (as well as higher energy bound if entropy is also small there) is often the most time-consuming part of the algorithm. If the length of each run of preliminary simulation is not sufficient, we often observe “oscillation” of the histogram at the extremes of the energy band, i.e., h_k that takes a small value in t th simulation becomes large in $(t + 1)$ th simulation, and, again become small in $(t + 2)$ th step, etc. and it does not converge.

We emphasize that a single simulation with an approximately multicanonical weight $1/\tilde{D}(E(\mathbf{x}))$ is enough to obtain canonical averages at any β . It is because a random walk with the weight $1/\tilde{D}(E(\mathbf{x}))$ covers whole range of the energy and we can collect information at all possible values of energy using them. An examination of the reweighting formulae also shows that the efficiency of the reweighting procedure relies on the flatness of the marginal distribution of the energy in the range of $E_{\min} < E < E_{\max}$. The flatness ensures that the weight $\exp(-\beta E)\tilde{D}(E)$ of a sample in reweighting formula (25) is a Gaussian-like distribution with the width

$\propto 1/\sqrt{N}$. Thus, the number of the samples that contribute to a canonical average is proportional to $1/\sqrt{N}$ for any value of β . For the purpose of a comparison, consider the sampling with the canonical weight $\exp(-\beta' E(\mathbf{x}))$ with a temperature $1/\beta'$ and reweighting to the temperature $1/\beta$ ($\beta > \beta'$).⁸⁴ The corresponding reweighting formula,

$$\langle A(\mathbf{x}) \rangle_\beta = \frac{\sum_j A(\mathbf{x}^j) \exp((\beta' - \beta) \cdot E(\mathbf{x}^j))}{\sum_j \exp((\beta' - \beta) \cdot E(\mathbf{x}^j))}, \quad (30)$$

is formally valid for any pair β' and β , but not useful except when β' are close to β . In this case, the marginal distribution of the energy of samples is a Gaussian-like distribution with the width $\propto 1/\sqrt{N}$, while the factor $\exp((\beta' - \beta) \cdot E)$ is quickly decreasing function of E with a decay constant $\propto N$. As a result, an exponentially small fraction of samples contribute to the required average for a large system size N .

So far, we discussed multicanonical algorithm with Gibbs distributions. It is easy to extend it to a one-parameter exponential family,

$$p_\lambda(\mathbf{x}) = \frac{\exp(\beta \cdot E(\mathbf{x}) + \lambda \cdot f(\mathbf{x}))}{Z(\lambda)}, \quad (31)$$

parameterized by λ . We can choose any physical quantity as a function $f(\mathbf{x})$, e.g., volume, magnetization, dihedral angle, radius of gyration, and replica overlap. To define the algorithm, we use $D(f(\mathbf{x}))$ instead of $D(E(\mathbf{x}))$, where $D(f)df$ is defined as the number of the states that satisfy $f < f(\mathbf{x}) < f + df$. We can also consider multi-dimensional extensions for a multi-parameter exponential family,^{35,36,46,50,91–93}

$$p_\lambda(\mathbf{x}) = \frac{\exp\left(\sum_l \lambda_l \cdot f_l(\mathbf{x})\right)}{Z(\boldsymbol{\lambda})}, \quad (32)$$

where $Z(\boldsymbol{\lambda}) = Z(\lambda_1, \lambda_2, \dots, \lambda_L)$ is the partition function. In this case, we define a simultaneous density $D(f_1, f_2, \dots, f_L)$ and estimate it by multivariate histogram construction with preliminary runs. Of course, we can treat β as one of the parameters λ_l . It seems, however, not possible to extend multicanonical algorithm beyond exponential family that is determined by a set of sufficient statistics $\{f_l(\mathbf{x})\}$. In terms of physics, multicanonical ensemble is defined using extensive quantities $\{f_l(\mathbf{x})\}$ conjugated to a set of “external force”s $\{\lambda_l\}$ and cannot be generalized to cases to which we cannot specify such quantities. At this point, it differs from Exchange Monte Carlo or Simulated Tempering, which can apply to any family of distributions $p_\lambda(\mathbf{x})$.^h

On the other hand, *there are cases that are well treated by Multicanonical Monte Carlo, but not by the other two methods.* Typical examples are provided by models

^hAn example of nonexponential family of distributions is Cauchy’s distributions $p_\lambda(x) = (\lambda/\pi) \cdot 1/(x^2 + \lambda^2)$ parameterized by the scale parameter λ .

with first order phase transition with latent heat. In these cases, there is a region on the energy axis that cannot be covered by a Gibbs distribution, i.e., for any value of inverse temperature β , there is negligible chance of finding a sample $\mathbf{x}$ with a value of $E(\mathbf{x})$ in the region. In the case of multicanonical Monte Carlo, the ensemble constructed by the iterative learning procedure contains samples with these missing values of the energy, which make the algorithm work as we expect. On the other hand, Exchange Monte Carlo and Simulated Tempering do not work in these cases, because any mixture of canonical ensembles (Gibbs distributions) contains little portion of samples with an energy in the gap region. In a physical interpretation that applies to models of liquids and lattice spin models, multicanonical “energy” $-\log D(E(\mathbf{x}))$ contains an artificial correction term to the interfacial tension of a droplet that makes the critical radius of a droplet zero and also controls the growth of a droplet after nucleation. While it seems very difficult to design such a term by hand, multicanonical algorithm automatically learns a term with desired properties with a histogram construction (or, some alternative iterative tuning methods).ⁱ

When there is no first order transition and resultant “phase coexistence regions”, *how can we relate a multicanonical ensemble to a mixture of Gibbs distributions?* At first sight, it may be natural to expect that the mixture with $\tilde{\pi}(\lambda) \propto \sqrt{I(\lambda)}$ (Jeffreys’ mixture) approximates well the corresponding multicanonical ensembles. It is, however, *not* true. In fact, the choice $\tilde{\pi}(\lambda) \propto I(\lambda)$, which gives larger weight to high specific heat regions, provides a better approximation to the multicanonical ensemble. We illustrate the result with a simple example of binomial distribution,

$$p_p(n) = {}_N C_n \cdot p^n (1-p)^{N-n}, \quad (33)$$

which is expressed as an exponential family,

$$p_\lambda(n) = \frac{\exp(\lambda \cdot n)}{Z(\lambda)}, \quad \lambda = \log \frac{p}{1-p}, \quad (34)$$

with a parameter λ . For this model, $Z(\lambda) = 1/{}_N C_n \cdot 1/(1-p)^N$, $I(\lambda) = Np(1-p)$, $I(p) = N/(p(1-p))$ and $d\lambda/dp = 1/(p(1-p))$. By using them, it is easy to show that the Jeffreys’ prior $\tilde{\pi}(\lambda)$ of λ is given by:

$$\tilde{\pi}(\lambda)d\lambda \propto \sqrt{I(\lambda)}d\lambda = \sqrt{I(p)}dp = \frac{\sqrt{N}dp}{\sqrt{p(1-p)}}. \quad (35)$$

On the other hand, the mixture with

$$\tilde{\pi}(\lambda)d\lambda \propto I(\lambda)d\lambda = Ndp \quad (36)$$

ⁱIn some cases, we need additional techniques to estimate the weights for multicanonical calculation in preliminary runs. In the work¹² of Berg and Neuhaus that deals with a Potts model, approximate weights found in smaller systems is used as an initial guess of the corresponding weights in larger systems.

gives the uniform density on n axis, i.e., it is a multicanonical ensemble. It is easily confirmed with the identity,

$$\int p_p(n) dp = \int {}_N C_n \cdot p^n (1-p)^{N-n} dp = \frac{1}{N+1}, \quad (37)$$

whose right-hand side does not contain n .

Let us discuss some of the typical results obtained by Multicanonical Monte Carlo. One of the attractive applications is found in the field of protein folding.^{38,42,43} An illustrative example of the ability of avoiding local optima by Multicanonical Monte Carlo is shown in Figs. 2 and 3 of Ref. 43, where Multicanonical Monte Carlo efficiently realizes α -helical conformations expected by laboratory experiments while conventional Monte Carlo fails. An advantage of Extended Ensemble Monte Carlo is, however, more clear in the examples, where fluctuation among the structures are significant. Such examples are also found in literatures, for example, Refs. 43 and 48. In Ref. 48, a β -hairpin peptide in explicit water (139 peptide atoms and 1060 water molecules) is simulated, and it is observed that the peptide takes a number of different conformations even at a physiological temperature, which are classified into several groups by the authors.

As we have discussed in this section, an advantage of Multicanonical Monte Carlo over other Extended Ensemble Monte Carlo algorithms is that it enables the treatment of first-order transitions. In this respects, the paper¹² by Berg and Neuhaus already gave an impressive example, i.e., simulation of 10-state Potts model up to the size 100^2 . Multicanonical Monte Carlo and its variants are also useful for the simulation of liquids and gas⁵⁶ (see also Sec. 8 of this paper), where we encounter classical examples of first-order transitions.

6. Replica Monte Carlo

Replica Monte Carlo algorithm^j by Swendsen and Wang^{28,29} is one of the pioneering works on Dynamical Monte Carlo algorithms that use multiple copies of the system. In fact, it includes Exchange Monte Carlo algorithm as a limit and precedes any study on Exchange Monte Carlo referred in this paper. However, *cluster dynamical aspect* of the algorithm is highly stressed in the original representation²⁸ and it seems not trivial to extract Exchange Monte Carlo algorithm from it. Cluster identification using a pair of replicas is a really ingenious and attractive idea itself, but it severely restricts the application of the algorithm when we persist in it.

In this section, we give an introduction to cluster dynamics of Replica Monte Carlo algorithm and clarify the relation between Replica Monte Carlo and Exchange Monte Carlo. It seems that there has been no concrete attempt to generalize the idea of cluster dynamics in Replica Monte Carlo beyond Ising models, although some suggestions are given in Ref. 28. Thus, we restrict our attention

^jNote that a similar term “**Replica Exchange Monte Carlo**” is sometimes used as a synonym of “Exchange Monte Carlo”.

to Ising models with inhomogeneous couplings $\{J_{ij}\}$. For this class of models, the Gibbs distribution is written as:

$$p(\{x_i\}) = \frac{\exp(\beta \sum_{(ij)} J_{ij} x_i x_j)}{Z(\beta)}, \quad (38)$$

where $x_i \in \{\pm 1\}$ is a Ising spin variable that is defined on the vertex i of a graph G (e.g., a square lattice) and $Z(\beta)$ is the partition function. The summation $\sum_{(ij)}$ runs over the all pairs (ij) , where i and j are neighboring on G , i.e., the edge $(i, j) \in G$. Consider a set of the Gibbs distributions, Eq. (38), defined with temperatures $\{\beta_k\}$. We denote the spin variables of k th system as $\{x_i\}^k$. Then, the simultaneous distribution $\tilde{p}$ of $\{\{x_i\}^k, k = 1, \dots, K\}$ is written as:

$$p_k(\{x_i\}^k) = \frac{\exp(\beta_k \sum_{(ij)} J_{ij} x_i^k x_j^k)}{Z(\beta_k)}, \quad (39)$$

$$\tilde{p}(\{\{x_i\}^k\}) = \prod_k p_k(\{x_i\}^k), \quad (40)$$

where $Z(\beta_k)$ is the partition function of the k th system (k th replica).

So far, the construction is the same as the one for Exchange Monte Carlo. In Replica Monte Carlo by Swendsen and Wang, nonlocal *cluster update* is used as well as usual single-spin flip update in each replica. It is defined for a pair of replicas, which have neighboring values of temperatures β_k and β_{k+1} . To define clusters in a pair of configurations $\{x_i\}^k$ and $\{x_i\}^{k+1}$, we introduce variables $t_i = x_i^k x_i^{k+1}$ and define an equivalence relation $\equiv$ among the vertices of G : If $(i, j) \in G$ and $t_i = t_j$, then $i \equiv j$. We define *clusters* as the equivalence classes with the relation $\equiv$. Note that there are two types of clusters distinguished by the values of t_i . In this paper, we call them *parallel clusters* ($t_i = 1$) and *anti-parallel clusters* ($t_i = -1$), respectively. Then, we define *cluster flip* as simultaneous flips of spins in a cluster in both replicas. That is, we choose a cluster c defined with configurations $\{x_i\}^k$, $\{x_i\}^{k+1}$ and generate candidates of new configurations $\{\tilde{x}_i\}^k$ and $\{\tilde{x}_i\}^{k+1}$ defined by: $\tilde{x}_i^k = -x_i^k$ if $i \in c$, else $\tilde{x}_i^k = x_i^k$; $\tilde{x}_i^{k+1} = -x_i^{k+1}$ if $i \in c$, else $\tilde{x}_i^{k+1} = x_i^{k+1}$. With this $\{\tilde{x}_i\}^k$ and $\{\tilde{x}_i\}^{k+1}$, the acceptance probability of the cluster flip is given by $\max\{1, r\}$, where,

$$r = \frac{p_k(\{\tilde{x}_i\}^k) p_{k+1}(\{\tilde{x}_i\}^{k+1})}{p_k(\{x_i\}^k) p_{k+1}(\{x_i\}^{k+1})}. \quad (41)$$

When we define the boundary ∂c of a cluster c as the set of edges (i, j) that satisfy $j \in c$ and $i \notin c$, the logarithm of the ratio r is expressed as:

$$\log r = -2 \cdot (\beta_k - \beta_{k+1}) \cdot \sum_{(i,j) \in \partial c} J_{ij} x_i^k x_j^k. \quad (42)$$

This expression gives $r \sim 1$ for $\beta_k \sim \beta_{k+1}$, i.e., a cluster flip is accepted with a high probability for a pair of replicas with sufficiently close temperatures, even when $|\beta_k J_{ij}|$ are large. These cluster flips share some characteristics with **crossover** in

Genetic algorithms in the sense that *they generate a new configuration from two existing configurations*. However, in contrast to random crossover that is rarely accepted with a large cluster exchange, cluster flips in Replica Monte Carlo are designed to realize high acceptance ratio while satisfying a detailed balance condition.

Let us discuss connections to Exchange Monte Carlo algorithm introduced in Sec. 3. In the case of Exchange Monte Carlo, we also consider the simultaneous distribution $\tilde{p}$ of Eq. (2) and define *replica exchange* between replicas, which have neighboring values of temperatures β_k and β_{k+1} . The identification of clusters is, however, not necessary in Exchange Monte Carlo. Instead, we define candidates of new configurations $\{\tilde{x}_i\}^k$ and $\{\tilde{x}_i\}^{k+1}$ by the exchange of configurations of replicas: for all i , $\tilde{x}_i^k = x_i^{k+1}$ and $\tilde{x}_i^{k+1} = x_i^k$. Then, the acceptance probability of the cluster flip is given by $\max\{1, r\}$ with r defined by Eq. (3). The explicit form of $\log r$ for the present model is given by:

$$\log r = -2 \cdot (\beta_k - \beta_{k+1}) \cdot \sum_{(i,j)} J_{ij} \cdot (x_i^k x_j^k - x_i^{k+1} x_j^{k+1}). \quad (43)$$

Although the implementation of Exchange Monte Carlo does not require the cluster identification procedure, it is useful for our purpose to rewrite it with the language of the clusters defined in Replica Monte Carlo. We define the sets $c^+ = \bigcup_{m \in M_+} c_m$ and $c^- = \bigcup_{m \in M_-} c_m$ as joint sets of parallel and anti-parallel clusters, respectively. Here, $M_{\pm}$ indicates the set of indices of parallel/anti-parallel clusters. Then, the *replica exchange* is equivalent to *flipping of the joint of all anti-parallel clusters* c^- : $\tilde{x}_i^k = -x_i^k$ if $i \in c^-$, else $\tilde{x}_i^k = x_i^k$; $\tilde{x}_i^{k+1} = -x_i^{k+1}$ if $i \in c^-$, else $\tilde{x}_i^{k+1} = x_i^{k+1}$. Thus, Eq. (43) is written as:

$$\log r = -2 \cdot (\beta_k - \beta_{k+1}) \cdot \sum_{(i,j) \in \partial c^-} J_{ij} x_i^k x_j^k, \quad (44)$$

where the boundary $\partial c^{\pm}$ of $c^{\pm}$ is defined as the sets (i, j) that $i \in c^{\pm}$ and $j \notin c^{\pm}$ (with this definition, $\partial c^+ = \partial c^- = \bigcup_{m \in M_+} \partial c_m = \bigcup_{m \in M_-} \partial c_m$). Note that the flip of the set c^+ of all parallel clusters is essentially equivalent to the flip of c^- for an Ising model with up-down symmetry. It reduces to the exchange of the replicas when the flip of all spins in replicas $\tilde{x}_i^k = -\tilde{x}_i^k$ and $\tilde{x}_i^{k+1} = -\tilde{x}_i^{k+1}$ is added after the cluster flip.

The analogy between Eq. (42) and Eq. (44) is obvious. If there are only two clusters, one is parallel and the other is anti-parallel, both algorithms give essentially the same dynamics. In this case, Exchange Monte Carlo has an advantage, because it does not require a cluster identification procedure. On the other hand, with this observation, it is not difficult to construct a family of algorithms that interpolate Replica Monte Carlo and Exchange Monte Carlo.^k In these algorithms,

^kIn the original paper,²⁸ there is no specification on the dynamics for cluster flips in Replica Monte Carlo. In this sense, Replica Monte Carlo virtually contains these interpolations. However, there seems no explicit suggestion on naive multi-spin flips in the references of Replica Monte Carlo (a comment on the use of percolation representation for cluster update is found in Ref. 29).

we construct clusters just the same way as that in Replica Monte Carlo, but flip more than one cluster simultaneously with the restriction of *only the same types of clusters can be flipped at one time*. That is, we generate a new candidate of configurations by the flip of the union of a set of parallel clusters, or, a set of anti-parallel clusters. It is easy to see that Replica Monte Carlo and Exchange Monte Carlo is regarded as two extremes of the algorithm, where the number of clusters updated in a single trial is *one* and *maximum*, respectively. Again, there is a discontinuity in the performance between Exchange Monte Carlo and the Exchange Monte Carlo-like limit of generalized Replica Monte Carlo, which is caused by the cost of the cluster identification procedure.

A weakness of the cluster identification procedure in Replica Monte Carlo is that it is not easy to generalize it for complicated models, e.g., lattice or off-lattice protein models. Another weak point might be found in the way of defining clusters itself. In Replica Monte Carlo, a cluster is defined with a pair of replicas that is *mutually independent*, possibly coming from very different regions of the phase space. Whether the clusters identified by such a way have adequate properties will depend on the model to be examined. On the other hand, Exchange Monte Carlo algorithm without cluster dynamics is much more robust and has been applied to a variety of models in various fields.

In the literature, two-dimensional $\pm J$ Ising spin glass models seem to be the only example with which the efficiency of Replica Monte Carlo is quantitatively discussed.^{28,30} A recent study by Houdayer³⁰ treated the model on $12^2 \sim 100^2$ lattices by a modification of Replica Monte Carlo.¹ The paper reports that the algorithm performs much better than Exchange Monte Carlo for the model and thermal equilibrium is attained even at very low temperatures. Although we should be careful to check outputs from such a complicated algorithm, the reported results, which are averaged over 100–400 realizations of disorder, is very attractive and likely to be decisive one for this model. Replica Monte Carlo is also applied to higher dimensional spin glass models.²⁹ The performance of cluster dynamics, however, seems not a remarkable one for these models. Houdayer³⁰ argued that the inability of cluster dynamics for three-dimensional spin glass models is a consequence of a mismatch between the structure of phase space of the model and the definition of a cluster in the algorithm.

7. Designing Special Purpose Ensembles

In the previous sections, we introduced three algorithms, Exchange Monte Carlo, Simulated Tempering and Multicanonical Monte Carlo. We also discuss cluster dynamics in Replica Monte Carlo. In all cases, an ensemble with extension in the temperature or energy axis is useful in the sense that they have straightforward

¹Houdayer's algorithm uses multiple series of replicas with the same set of temperatures. Cluster dynamics is defined only with pairs of replicas with same temperatures, while Exchange Monte Carlo is applied to replicas with different temperatures.

applications to various problems in statistical physics. On the other hand, we can introduce extensions specially designed for a target distribution and our computational aims. The use of such ensembles is already discussed in the pioneering paper on Expanded Ensembles by Lyubartsev *et al.*¹ and in the studies on Adaptive Umbrella Sampling^{10,11} (and also in the earlier studies on Umbrella sampling, where the weight is manually determined). Here, we discuss designing principle and utility of such “special purpose” ensembles.

7.1. Complexity ladder

To obtain a fast mixing Markov chain and evaluate the free energy (multiple integrals), it is reasonable to construct an ensemble composed of a sequence of distributions that interpolates a “complex/unknown” distribution to a “simple/known” distribution.^{1,5,70,94} Wong⁹⁴ called such a structure as a “**complexity ladder**”. Here, the “simple/known” components should have sufficient entropy to obtain the required diversity of paths to the “complex/unknown” states (the ferromagnetic state is an example of the state simple but does not have sufficient entropy). The canonical distributions with different temperatures give an example of such structures, where components of higher temperatures correspond to simpler components. At the infinite temperature, it reduces to a known distribution that described a completely disordered state. Another example is provided by “Multi-System-Size”-type ensembles, which consists of distributions of different system size N . Here, components with small N fill the role of simple components.^m We will discuss them further in Sec. 8. We can proceed more along this way. For example, we can introduce an extended ensemble with “soft spin”s. At an extreme, its component is a distribution with discrete (or constrained) variables, say, Ising spins or rigid plane rotators. At the other extreme, its component reduces to a Gaussian distribution, whose samples are easily generated with Cholesky decomposition of the covariance matrix. This method is not implemented yet, but might be useful for some hard problems in the study of spin glass and combinatorics.

7.2. Bridge

A way to design artificial ensemble for efficient computation is the inclusion of *nonphysical configurations* (states prohibited in the original problem).^{1,5} The above-mentioned soft-spin algorithm is regarded as an example. Another example is ensembles for lattice polymers, which contain conformations that violate the *self-avoiding condition*.^{1,32,35,36} Specifically, Multi-Self-Overlap Ensemble (MSOE) introduced by Iba, Chikenji and Kikuchi³⁵ had a remarkable success in the calculation of ground states and thermodynamic properties of lattice protein models.³⁶ Similar

^mIt is clear that $N = 0$ component has not sufficient entropy. The intermediate states, however, can have enough entropy at moderate temperatures. Algorithms with Multi-System-Size ensemble will not be efficient or biased at very low temperatures, where these states have not enough entropy.

approaches for off-lattice polymers with truncated Lennard–Jones cores are discussed in Refs. 32, 36, 95 and 96. The other examples are ensembles that relax *hard core condition* of hard core fluid (solid)^{1,52,53,n} and ensembles for gene-propagation analysis (pedigree analysis)⁵ that contain configurations violating *Mendel’s law* of genealogy.

These ensembles often result in great enhancement of the efficiency, because “**bridges**”^o provided by the nonphysical configurations give a lot of additional paths between the configurations that are separated in the original problem. An instructive example of such “stepping stones” is shown in Fig. 1 of Ref. 36. Of course, a drawback of such an approach is that we “lose” the nonphysical samples. It is thus necessary for the improvement in the mixing rate to be large enough to overcome this loss.

For Multi-Self-Overlap ensemble for the HP model of lattice heteropolymers, this requirement is checked by Iba *et al.*³⁵ For a chain of length 56 with highly degenerate ground states, simulation with Multi-Self-Overlap ensemble produces more independent samples than a conventional Multicanonical Monte Carlo within the same number of MCS, despite the loss of nonphysical samples that violate the self-avoiding condition. Chikenji *et al.*³⁶ successfully applied Multi-Self-Overlap Monte Carlo to problems biophysically more interesting, e.g., generation of the lowest energy state of a chain of length 100 and calculation of thermodynamic properties of a protein-like chain of length 42. Exploration of thermal states of such a long heterogeneous chain have rarely been reported in this field. Based on the results of these calculations, Chikenji *et al.* proposed a hypothesis on the relation between the order of the phase transition and ground state degeneracy.

Besides slow mixing of “mathematically correct” algorithms, genuine nonergodicity of dynamics, i.e., lack of the connectivity of the graph defined by the transition matrix, is often an annoying problem in Dynamical Monte Carlo. It is not always easy to prove the ergodicity of a given Markov chain (we should carefully check unexpected appearance of isolated configurations. See Fig. 2.10–2.12 in Ref. 99 for examples of such configurations in self-avoiding walk simulations). The introduction of unphysical states as bridges provides a simple way to resolve the difficulty of non-ergodicity. Examples are seen in the studies on self-avoiding lattice polymers,^{32,35,36} pedigrees,⁵ and Latin squares.⁹⁸

The inclusion of the forbidden states as bridges is a natural idea to improve relaxation and has been used in pre-extended-ensemble stages.^{98,100,101} It is, however, not always easy to set the penalty for putting adequate part of samples into

ⁿHard core fluid is also treated by a multicanonical-type extension in the space of *volume* occupied by the fluid,⁵¹ which is equivalent to an extension with *varying diameter of hard cores* in athermal models.

^oI borrow the keyword **bridge** from Lin and Thompson,⁹⁷ while the term **stepping-stones** have appeared in Refs. 36 and 98. **Catalytic Tempering** algorithm proposed in Ref. 96 is also based on a similar idea.

“bridges” without the idea of learning in preliminary runs (or the use of multiple replicas). Now, we can use any of three types of implementations for this purpose, i.e., Exchange Monte Carlo, Simulated Tempering, and Multicanonical Monte Carlo. Extended Ensemble Monte Carlo with nonphysical bridges is also considered as a finite temperature version of constrained optimization algorithm by Geman *et al.*¹⁰²

7.3. Calculation of a free energy surface

An important motivation to special purpose ensembles is the observation of rare events and calculation of free energy surfaces. When we want to calculate free-energy surface as a function of one- or two-macroscopic variables, we can use an ensemble extended in the dimensions of these variables.^{10,11,50,91,92,103} For example, Sheto *et al.*⁹¹ design an ensemble for the study of the free-energy surface of an Ising model on the energy-magnetization plane. Similar approaches are extensively used in the studies with Adaptive Umbrella Sampling^{10,11,50} for the calculation of free energy as a function of reaction coordinates (e.g., dihedral angles of polypeptides). **Multioverlap ensemble**²⁶ designed for the calculation of the distribution of “replica overlap” between two independent samples is also based on a modification of this strategy.

Note that the extension required for the measurement often does not provide large entropy states that are necessary for fast mixing of the Markov chain. For this reason, some of the calculations by Adaptive Umbrella Sampling or Multioverlap ensemble might possibly be affected by the slow mixing of the Markov chain. It is often covered by fast tunneling between the states that have extreme values of the reaction coordinate, but such tunneling does not ensure unbiased sampling from all metastable states. Multidimensional extensions provide a way to circumvent this difficulty. We will discuss them in the next paragraph.

7.4. Multivariate extensions

To implement special purpose ensembles, the idea of ensembles extended in multidimensions (multivariate/multicomponent extensions^{35–37,46,50,80,91–93}) plays an important role. Since the use of too many dimensions is not realistic, bivariate or three-variate extensions are usually most useful. For example, in the case of Multi-Self-Overlap ensemble for lattice polymers,^{35,36} we use an ensemble defined by uniform density on two-dimensional space (degree of self-overlap, energy). It seems essential for heteropolymers with attractive interactions to include the extension in the space of the energy, because the relaxation of the self-avoiding constraint often causes collapse of polymers at low temperature.

As we have already remarked in the previous paragraph, two-dimensional extension is especially useful for extended ensembles for the measurement of rare events (see Sec. 2). That is, two-dimensional extensions, say, (an axis for the measurement,

the axis of the temperature) or (an axis for the measurement, the axis of the energy) improve the efficiency (and safety) of the algorithms for the calculation of free energy surfaces on the axis of measurement. Examples are found in Refs. 37, 80, 91–93.

For example, Chikenji and Kikuchi³⁷ explored the entropy density of a lattice protein model (a Gō model) using an extension of Multi-Self-Overlap ensemble defined by the uniform density in a three-dimensional space spanned by (the degree of self-overlap, the number of native contacts, the number of total contacts). Their motivation is the study of the curious folding mechanism of β -lactoglobulin, where α -helix rich intermediates tentatively appears before it finally folds into a stable β -sheet rich conformation. Entropy density on the space (the number of native contacts, the number of total contacts) calculated by the method vividly illustrates the role of entropy in the folding.

How can we construct ensemble with multi-dimensional extensions? We have already discussed it for the case of multicanonical ensembles (Sec. 5). For Exchange Monte Carlo and Simulated Tempering, the introduction of a multiparameter family is also straightforward. In the case of Exchange Monte Carlo,^{20,35,80} a two-dimensional version of the simultaneous distribution of replicas is expressed as:

$$\tilde{p}(\{\mathbf{x}_{kj}\}) = \prod_{(k,j) \in G} p_{kj}(\mathbf{x}_{kj}), \quad (45)$$

where the values of two parameters are indexed by k and j . There is a degree of freedom in the choice of a Graph G that defines the way of the extension. A way is the use of a “lattice type” configuration of (k, j) in the two-dimensional parameter space.⁸⁰ Another possibility is the use of a “quasi one-dimensional” configuration where (k, j) are points on a curve in the parameter space.²⁰ The former corresponds to two-dimensional multicanonical algorithms. The latter saves computational resources, but it is not easy to tune the large degree of freedom of setting a curve in the two-dimensional space.

8. Multi-System-Size Ensembles

As we have discussed in the previous sections, we can freely choose the space of extension in Extended Ensemble Monte Carlo. An interesting possibility is the use of an extension in a “space of size of the system” (or, in general, in a “space of dimensionality of the state space”), which corresponds to sampling from a mixture of systems of different sizes. A simulation that realizes such an ensemble is considered as a “growth/diminish (construction/destruction) method” for sampling a probability distribution. Note that monotonic growth of the system is prohibited by the detailed-balance condition.

In Iba (1999),²⁷ I discuss a simulation of spin glass with an ensemble extended in the space of system-size, Multi-System-Size Ensemble. After I completed the

work, I came to know references with similar ideas in various fields of physics and statistical sciences. Here, we will give a *cross-disciplinary* survey on this subject.

We begin with simulations of fluids with a varying number of particles. We know an ensemble corresponding to such a simulation. It is **grandcanonical ensemble** found in any textbook of statistical physics. It is interesting to introduce multicanonical or other types of extensions in the space of particle number (particle density, chemical potential) as a generalization of grandcanonical ensemble. While the number of particles fluctuates in a limited range in conventional grandcanonical ensemble, it can fluctuate in an arbitrary wide range, say, zero to 1000 in the extended ensemble. In the literature, Lyubartsev *et al.*¹ already discussed it in the context of free energy calculation. An application to Lennard–Jones fluid is found in Wilding,⁵⁵ which explored the subcritical coexistence region of Lennard–Jones fluid by a multicanonical ensemble defined by the uniform density on the space of the particle density.

There is a variation on this idea. When we are interested in the calculation of the chemical potential, an ensemble that gives an interpolation between size N and size $N + 1$ systems is often required. It is precisely realized by an extended ensemble that contains components corresponding to systems with *partial decoupling* between a particle and other N particles. Examples of such a “ghost particle” (ghost polymer) method are found in Refs. 57 and 82. In the former study,⁵⁷ solvation free energies of methane and alkali halide ions are calculated, and the results are compared to experimental data. These ensembles are also considered as examples of the extended ensembles that consist of nonphysical systems. Similar idea with an umbrella method is also found in Ref. 104.

Another application in physics, where extension in the space of system-size is naturally introduced is simulation of polymers. It is considered as a variation of grandcanonical simulation of self-avoiding walk,⁹⁹ where monomers are added and removed at the both ends of a polymer chain. In the conventional grandcanonical simulation, the length of polymers fluctuates around an equilibrium length. Introducing the idea of extended ensembles, we can systematically enhance the fluctuation and perform practical calculations for intricate models, such as self-avoiding walks in a restricted geometry or with complex interactions between monomers. For example, an adaptive, multicanonical-like method for self-avoiding walk is given by Grassberger,¹⁰⁵ while grandcanonical simulation with a length dependent chemical potential is already seen in earlier works, e.g., Ref. 106. In Ref. 105, self-avoiding walks on a lattice with random obstacles (i.e., randomly chosen excluded sites) up to the length 100 are efficiently generated by the method, and it is shown that the universality class of the problem is different from the one of the corresponding uniform lattice. Escobedo and de Pablo^{40,41} reported a systematic study of expanded ensemble simulation of polymers, where the length of polymers are dynamically changed. Their method is recently applied to diblock copolymer.³³ It might be interesting to point out that a preliminary work on multicanonical ensemble¹⁰⁷ also treated a size-variable system, an ensemble of random surfaces.

Extended Ensemble Monte Carlo algorithms with a state space of varying dimension is also of current interest in statistical sciences. It is natural to apply them to the problems with a built-in sequential structure, say, time-series modeling and gene-propagation analysis. On the other hand, they have an interesting motivation in statistics, i.e., they are useful for the simulation of a (posterior) distribution over the space of models with different number of parameters^{108,109} (here a “parameter” means an element of a model to be estimated from data.).^p After a pioneering comment by Wong,¹¹⁰ Wong and Liang⁹⁴ gave applications of their idea in various types of the problems, including a Traveling Salesman Problem. Liu and Sabbati⁶³ extensively discussed applications of “**Simulated Sintering**” algorithm in statistics.^q

Finally, we will touch on a related idea, extended ensemble that consists of systems of *variable types of dynamical variables*. Kerler and Weber¹⁸ discussed an extended ensemble simulation of Potts model, where the number q of possible states of a spin (“colors”) is dynamically changed in a run. They implement the idea in combination with cluster dynamics. A corresponding situation in statistics is found in Richardson and Green,¹¹² which deals with simulation-based Bayesian classification of objects to an unknown number of clusters.^r They treated problems like “How many Gaussian components are identified in a given experimental data?”, and designed a Monte Carlo procedure for computation of probabilities. In their work, the cluster to which each object belongs is indicated by the state of a Potts spin corresponding to the object. Then, dynamical changes of the number q of the clusters corresponds to variation of the number q of the states of Potts spins.

9. Summary and Future Perspectives

In this paper, we review three types of Extended Ensemble Monte Carlo algorithm, i.e., Exchange Monte Carlo, Simulated Tempering, and Multicanonical Monte Carlo and discuss approaches with special purpose ensembles. We also give a guide to extended ensembles with variable dimension of state spaces and Replica Monte Carlo algorithm. Throughout the paper, the possibility of various type of extensions is stressed. They are not only useful for the calculation of free energy, but also efficient for acceleration of the relaxation. Our idea is summarized in the following “correspondence principle”:

^pWe note that models with a large number of parameters do not necessarily show better performance with finite number of data. It is rather evident when we consider extreme cases, such as fitting of 10 data points by 9th order polynomials and classification of 100 samples into 100 clusters. Thus, we need to select a model (or mix models) with an appropriate number of parameters.

^qIn these works, a “Dynamic Weighting” technique proposed in Refs. 94 and 111 is used, which does not belong to the class of Extended Ensemble Monte Carlo defined in this paper.

^rThe work does not seem to use an iterative learning procedure to improve the sampling scheme. In this sense, it is not Extended Ensemble Monte Carlo defined in this paper. A reason that we refer to the study here is that it provides a good example of the application of Dynamical Monte Carlo in computational statistics.

If we have an annealing strategy for searching ground states, we can design an Extended Ensemble Monte Carlo algorithm for sampling from the corresponding distribution.

With this principle, we can translate optimization algorithms to algorithms for the calculation of thermodynamic properties. Note that this is only true for simulated annealing-type algorithms, and not applicable to more intricate/sophisticated optimization algorithms, e.g., algorithms with “genetic crossover” or with other heuristics that violate detailed-balance, methods based on the use of the correspondence between ground states of different systems. The principle, however, still provides a useful guideline for the construction of sampling schemes.

With Extended Ensemble Monte Carlo algorithms, we can attack difficult problems, where conventional Dynamical Monte Carlo algorithms are too slow even with the most powerful computers. Up to now, the most significant applications are found in computational physics and statistical information processing. But I believe that Extended Ensemble Monte Carlo is also a key to any field that requires sampling from complex distributions and estimation of the entropy. The introduction of Dynamical Monte Carlo — in 1950s for physics and in 1990s for statistics — gave great impacts on these fields. I hope that Extended Ensemble Monte Carlo will give *second impacts* on the study of the fields, where we are interested in the properties of probabilistic distributions and *large deviation* from nonweighted averages, including computational physics and statistics as special examples.

Appendix. Comparison of the Methods

In this appendix, we discuss issues on relative advantages of the three (types of) algorithms, Exchange Monte Carlo (Metropolis-Coupled Chain, Parallel Tempering), Simulated Tempering (Expanded Ensemble Monte Carlo), and Multicanonical Monte Carlo (Adaptive Umbrella). The results are summarized in the following table:

Subjects	Comment #	Exchange Monte Carlo [Sec. 3]	Simulated Tempering [Sec. 4]	Multicanonical Monte Carlo [Sec. 5]
First Order Trans.	1	×	×	○
NonExp. Family	2	○	○	×
Replica Overlap	3	○	×	×
Learning Speed	4	○	?	?
Molecular Dynamics	5	○	○	○
Step Size Control	6	○	○	?

While the symbol $\bigcirc$ indicate that the algorithm has an advantage on the subject, $\times$ means that the algorithm has severe disadvantage on the subject. The symbols $?$ means “still controversial”. The number in “comment #” column indicates the item number (#) of the discussion on the corresponding subject.

We will give remarks on the subjects in the table in the following:

(1) **First Order Transition**

As we have discussed in the previous sections, a remarkable advantage of Multicanonical Monte Carlo is that it can treat systems with first-order-like transitions.

(2) **NonExponential Family**

On the other hand, nonexponential family of distributions is not suitable for multicanonical-type treatment, as we remarked in Sec. 5. While the significance of nonexponential family is not clear in statistical physics, they are often important in the applications in statistical sciences.

(3) **Calculation of Replica Overlap**

In the study of statistical physics of random systems, it is often required to calculate the distribution of a quantity defined with two independent samples from a Gibbs distribution. An example of such quantities is “replica overlap” q , which is defined as an overlap of mutually independent samples $\mathbf{x}$ and $\mathbf{x}'$ from a given distribution. An easy way to compute such a quantity is to simulate two independent Markov chains S, S' and use a pair $(\mathbf{x}, \mathbf{x}')$ of samples where $\mathbf{x}$ and $\mathbf{x}'$ are sampled from S and S' , respectively. Then, independence of $\mathbf{x}$ from $\mathbf{x}'$ is assured even with slow mixing of the Markov chains S and S' .

This simple method, however, does not work well when we use Simulated Tempering, because the states in two chains have different values of temperature for most part of the simulation — they coincide with each other with probability $1/K$ when the number of discretized temperatures is K , which results in severe waste of samples. The situation is essentially the same when we use multicanonical-type algorithms, or, when any parameter is used for the construction of the extended ensemble. For Simulated Tempering, we can use two copies of the system with a common temperature variable, but it lowers the performance of the algorithm. A few other methods have been proposed up to now, but all of them have some drawbacks, e.g., they cause slow relaxation of system (Berg and Celik¹⁹) or introduce additional complexity (multioverlap ensemble²⁶).

On the other hand, calculation of the overlap q and other quantities defined with two independent samples from the original distribution is straightforward with Exchange Monte Carlo algorithm. We just simulate two independent Markov chains each of which consists of K replicas with the same set of temperature $\{\beta_k\}$. Then, we calculate and record the overlap of two replicas with the same temperature whose time evolution is governed by mutually

independent Markov chains. This is a significant advantage of Exchange Monte Carlo algorithm.

(4) **Learning Speed in Preliminary Runs**

Exchange Monte Carlo algorithm seems to show better performance and less complexity in the learning phase. It does not require simultaneous tuning of the strength of the penalty and discretization of the temperature required in Simulated Tempering. The tuning of the values of the temperature (parameter) of replicas is still required, but we can enjoy a benefit from the use of the exchange rate between replicas.

On the other hand, Simulated Tempering and Multicanonical Monte Carlo have a handicap in the learning stage, if we use the naive method of tuning based on the frequency of the visits to a temperature or an energy. It is because a random walk on the temperature or energy axis causes fluctuation of the visiting frequencies, which directly induces instability of the calculated weights.

Some authors,^{25,51} however, proposed tuning methods based on transition probabilities, which will reduce the instability of this type. Recent development of Flat Histogram Monte Carlo and related algorithms^{88–90} can also improve the efficiency of the learning stage. On the other hand, experiences on difficult cases suggest that the most difficult part of the tuning phase is often the determination of several weights near ground states of the system (and ground states themselves). We need more experiences and carefully designed experiments to evaluate these factors — this is the reason why we give “?” to this subject.

(5) **Molecular Dynamics, Hybrid Monte Carlo, Langevin Equation**

Molecular Dynamics is a useful tool for the simulation of continuous systems, say, simulation of realistic protein models, even when we are interested only in equilibrium properties. Specifically, combinations of Dynamical Monte Carlo and Molecular Dynamics (**Hybrid Monte Carlo**) are convenient tools for the sampling from Gibbs distributions. There are also methods based on Langevin Equation, which can be regarded as a version of Hybrid Monte Carlo.

Here, we consider how to combine these methods with the idea of Extended Ensemble Monte Carlo. At first sight, Exchange Monte Carlo and Simulated Tempering have an advantage, because implementation is quite straightforward.^{32,39,57,59,61,78} For example, the addition of an exchange procedure is enough for the combination of Exchange Monte Carlo and Hybrid Monte Carlo, where the states of replicas are swapped with fixed values of the corresponding demons (i.e., momentum part of Hamiltonian).^s

On the other hand, multicanonical-type implementation requires the estimation of the derivative $d \log D(E)/dE$ of the logarithm of the density of state,

^sSugita and Okamoto⁷⁸ have proposed a different method, where demons and replicas are exchanged simultaneously with rescaling of momenta of the demons. Another approach is introducing an exchange procedure to microcanonical ensembles in which demons are integrated out.⁵⁹

which causes additional complexity in the tuning stage. However, studies by several authors^{11,45,46,49,50,60,113} have proved that it is not difficult despite the apparent difficulty.

(6) Step-Size Control

For continuous systems, step-size of trial moves (or, in general, the distribution of sizes and directions of trial moves) is an important factor in the fast mixing of the Markov chain governed by Metropolis dynamics. The optimal step-size depends on temperature and other parameters of the target distribution. There is no established way for the determination of optimal step-size, but there are some “rules of thumb”, e.g., step-size with moderate acceptance ratio (say, $\sim 50\%$) usually gives good results.

For Exchange Monte Carlo and Simulated Tempering, step-size can depend upon the temperature (or any parameter used for the construction of extended ensemble). It does not spoil the detailed balance condition because flips in a replica (flips with a fixed temperature) are separated from replica exchanges (temperature changes) in these algorithms.

It is not the case with Multicanonical Monte Carlo. If we use energy dependent step-size in a multicanonical simulation of a continuous system, it usually gives wrong results, because it results in violation of the detailed balance condition. Introduction of additional weights that compensate the bias caused by energy dependent step-size seems not easy.

This disadvantage, however, can be solved by using a “patchwork” of ensembles proposed by several authors.[†] For example, consider an ensemble composed by several multicanonical-type ensembles, each of which is defined on an interval on the energy axis. These intervals are assumed to partly overlap each other and we use a method like Exchange Monte Carlo for the integration of them. In this setting, we can safely use different step sizes in different components.

Acknowledgement

I would be grateful to Dr. K. Hukushima, Prof. M. Kikuchi, Mr. G. Chikenji, Prof. Y. Okamoto, Prof. H. Nakamura and Dr. A. Doucet for giving me kind advices and teaching references.

References

1. A. P. Lyubartsev, A. A. Martsinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, *J. Chem. Phys.* **96**, 1776 (1992).

[†]Hansmann⁷⁷ discussed a patchwork of Tsallis ensembles. A patchwork of locally multicanonical ensembles is discussed by Sugita and Okamoto.^{43,79} The present author (Y. Iba) has also proposed a version of Exchange Monte Carlo algorithm, where canonical ensemble is replaced by the ensemble defined by $E < E_{th}$ (unpublished). Similar ensembles naturally appear, when demons (momentum part of Hamiltonian) are integrated out from microcanonical ensembles.⁵⁹

2. K. Kimura and K. Taki, in *Proceedings of the 13th IMACS World Congress on Computation and Applied Mathematics (IMACS'91)*, Vol. 2, eds. R. Vichnevetsky and J. J. H. Miller (1991), pp. 827–828.
3. M. Hirose, R. J. Feldmann, D. Rawn, M. Ishikawa, M. Hoshida, and G. Micheals, in *Proceedings of the International Conference on Fifth Generation Computer Systems (ICOT, 1992)*, p. 300.
4. C. J. Geyer, in *Computing Science and Statistics: Proceedings of the 23rd Symposium on the Interface*, ed. E. M. Keramidas (Interface Foundation, Fairfax Station, Va., 1991), p. 156.
5. C. J. Geyer and E. A. Thompson, *J. Amer. Stat. Assoc.* **90**, 909 (1995).
6. Y. Iba, *Proceedings of the Institute of Statistical Mathematics* **41**, 65 (1993) [in Japanese. A translation to English is included in the Ph.D. thesis by Y. Iba (2000, Osaka Univ.)]; Y. Iba, *Bussei Kenkyu (Kyoto)* **60**, 677 (1993) [in Japanese].
7. K. Hukushima, H. Takayama, and K. Nemoto *Int. J. Mod. Phys. C* **7**, 337 (1996); K. Hukushima and Y. Nemoto, *J. Phys. Soc. Jpn.* **65**, 1604 (1996); K. Hukushima, H. Takayama, and H. Yoshino, *J. Phys. Soc. Jpn.* **67**, 12 (1998).
8. M. C. Tesi, E. J. Janse van Rensburg, E. Orlandini, and S. G. Whittington, *J. Stat. Phys.* **82**, 155 (1996); S. G. Whittington, *Monte Carlo Methods, Fields Inst. Commun.*, Vol. 26, ed. N. Madras (American Mathematical Society, 2000), pp. 131–141.
9. E. Marinari and G. Parisi, *Europhys. Lett.* **19**, 451 (1992).
10. M. Mezei, *J. Comp. Phys.* **68**, 237 (1987).
11. R. W. W. Hooft, B. P. van Eijck, and J. Kroon, *J. Chem. Phys.* **97**, 6690 (1992).
12. B. A. Berg and T. Neuhaus, *Phys. Lett. B* **267**, 249 (1991); B. A. Berg and T. Neuhaus, *Phys. Rev. Lett.* **68**, 9 (1992).
13. B. A. Berg and T. Celik, *Phys. Rev. Lett.* **69**, 2292 (1992).
14. B. A. Berg, *Monte Carlo Methods, Fields Inst. Commun.*, Vol. 26, ed. N. Madras (American Mathematical Society, 2000), pp. 1–24.
15. R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **58**, 86 (1987); U. Wolff, *Phys. Rev. Lett.* **62**, 361 (1989); S. Liang, *Phys. Rev. Lett.* **69**, 2145 (1992).
16. A. S. Kronfeld, *Progress of Theoretical Physics Supplement*, No. 111, 293 (1993).
17. M. A. Novotny, *Annual Reviews of Computational Physics*, Vol. IX, ed. D. Stauffer (World Scientific, Singapore, 2001), p. 153.
18. W. Kerler and A. Weber, *Phys. Rev. B* **47**, 563 (1993).
19. B. A. Berg and T. Celik, *Int. J. Mod. Phys. C* **3**, 1251 (1992).
20. K. Hukushima, *Phys. Rev. E* **60**, 3606 (1999).
21. K. Hukushima and H. Kawamura, *Phys. Rev. E* **61**, R1008 (2000).
22. E. Marinari, G. Parisi, and J. J. Ruiz-Lorenzo, *Spin Glasses and Random Fields*, ed. A. P. Young (World Scientific, Singapore, 1998), pp. 59–98.
23. E. Marinari, O. C. Martin, and F. Zuliani, “Equilibrium valleys in spin glasses at low temperature,” [cond-mat/0103534](#).
24. G. Hed, E. Domany, M. Palassini, and A. P. Young, “Correlated spin domains generate hierarchical structure in state space in short range spin glasses,” [cond-mat/0104264](#).
25. W. Kerler and P. Rehberg, *Phys. Rev. E* **50**, 4220 (1994).
26. B. A. Berg and W. Janke, *Phys. Rev. Lett.* **80**, 4771 (1998); W. Janke, B. A. Berg, and A. Billoire, *Ann. Phys. (Leipzig)* **7**, 544 (1998).
27. Y. Iba, *J. Phys. Soc. Jpn.* **68**, 1057 (1999).
28. R. H. Swendsen and J.-S. Wang, *Phys. Rev. Lett.* **57**, 2607 (1986).
29. J.-S. Wang and R. H. Swendsen, *Phys. Rev. B* **37**, 7745 (1988); J.-S. Wang and R. H. Swendsen, *Phys. Rev. B* **38**, 4840 (1988); J.-S. Wang and R. H. Swendsen, *Phys. Rev. B* **38**, 9086 (1988).

30. J. Houdayer, "A cluster Monte Carlo algorithm for two-dimensional spin glasses," *cond-mat/0101116*.
31. T. V. Kuznetsova and P. N. Vorontsov-Velyaminov, *J. Phys. Condens. Matter* **5**, 717 (1993); A. W. Sandvik, *Phys. Rev. Lett.* **80**, 5196.
32. P. N. Vorontsov-Velyaminov, A. V. Broukhno, T. V. Kuznetsova, and A. P. Lyubartsev, *J. Phys. Chem.* **100**(4), 1153 (1996).
33. Q. Wang, Q. Yan, P. F. Nealey, and J. J. de Pablo, *J. Chem. Phys.* **112**, 450 (2000).
34. N. Urakami and M. Takasu, *J. Phys. Soc. Jpn.* **65**, 2694 (1996); N. Urakami and M. Takasu, *Molecular Simulation* **19**, 63 (1997).
35. Y. Iba, G. Chikenji, and M. Kikuchi, *J. Phys. Soc. Jpn.* **67**, 3327 (1998).
36. G. Chikenji, M. Kikuchi, and Y. Iba, *Phys. Rev. Lett.* **83**, 1886 (1999).
37. G. Chikenji and M. Kikuchi, *Proc. Natl. Acad. Sci.* **97**, 14273 (2000).
38. U. H. E. Hansmann and Y. Okamoto, *Current Opinion in Structural Biology* **9**, 177 (1999).
39. A. Irbäck, C. Peterson, and F. Potthast, *Phys. Rev. E* **55**, 860 (1997).
40. F. A. Escobedo and J. J. de Pablo, *J. Chem. Phys.* **103**, 2703 (1995).
41. F. A. Escobedo and J. J. de Pablo, *J. Chem. Phys.* **105**, 4391 (1996).
42. U. H. E. Hansmann and Y. Okamoto, *J. Comput. Chem.* **14**, 1333 (1993); Y. Okamoto, *Recent Res. Devel. Pure Appl. Chem.* **2**, 1 (1998).
43. A. Mitsutake, Y. Sugita, and Y. Okamoto, *Biopolymers (Peptide Science)* **60**(2), 96 (2001).
44. Y. Sugita and Y. Okamoto, "Free-energy calculations in protein folding by generalized-ensemble algorithms," *cond-mat/0102296*.
45. N. Nakajima, J. Higo, A. Kidera, and H. Nakamura, *Chem. Phys. Lett.* **278**, 297 (1997).
46. N. Nakajima, *Chem. Phys. Lett.* **288**, 319 (1998).
47. S. Ono, N. Nakajima, J. Higo, and H. Nakamura, *Chem. Phys. Lett.* **312**, 247 (1999).
48. J. Higo, O. V. Galzitskaya, S. Ono, and H. Nakamura, *Chem. Phys. Lett.* **337**, 169 (2001).
49. C. Bartels and M. Karplus, *J. Phys. Chem. B* **102**, 865 (1998).
50. C. Bartels and M. Karplus, *J. Comput. Chem.* **18**, 1450 (1997).
51. G. R. Smith and A. D. Bruce, *Europhys. Lett.* **34**, 91 (1996); G. R. Smith and A. D. Bruce, *Phys. Rev. E* **53**, 6530 (1996).
52. A. D. Bruce, N. B. Wilding, and G. J. Ackland, *Phys. Rev. Lett.* **79**, 3002 (1997).
53. S. Mase, J. Möller, D. Stoyan, R. P. Waagepetersen, and G. Döge, "Packing densities and simulated tempering for hard core Gibbs point processes," preprint (1999).
54. N. B. Wilding and A. D. Bruce, *Phys. Rev. Lett.* **85**, 5138 (2000).
55. N. B. Wilding, *Phys. Rev. E* **52**, 602 (1995).
56. N. B. Wilding, "Computer simulation of fluid phase transitions," *cond-mat/0103248*, to appear in *Amer. J. Phys.*
57. A. P. Lyubartsev, O. K. Førrisdahl, and A. Laaksonen, *J. Chem. Phys.* **108**, 227 (1998).
58. A list of papers on free energy calculation by expanded ensembles is found in the following web site: <http://www.fos.su.se/physical/sasha/freeen.html>
59. F. Calvo, J. P. Neirotti, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.* **112**, 10350 (2000).
60. G. Arnold, K. Schilling, and Th. Lippert, *Phys. Rev. D* **59**, 054509 (1999) ("054509" is an article number used for the identification of a paper in the journal).
61. B. Joó, B. Pendleton, S. M. Pickles, Z. Sroczynski, A. C. Irving, and J. C. Sexton, (UKQCD Collaboration), *Phys. Rev. D* **59**, 114501 (1999) ("114501" is an article number used for the identification of a paper in the journal).

62. T. Hotta, T. Izubuchi, and J. Nishimura, *Nucl. Phys. B* **531**, 446 (1998).
63. J. S. Liu and C. Sabatti, *Bayesian Statistics 6*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith (Clarendon Press, Oxford, 1999; Oxford University Press, New York, 1999), pp. 389–413.
64. K. E. Andersen, S. P. Brooks, and M. B. Hansen, *Inverse Problems* **17**, 121 (2001).
65. K. Pinn and C. Wiecekowsky, *Int. J. Mod. Phys. C* **9**, 541 (1998).
66. A good reference on the applications of dynamical Monte Carlo algorithms in statistical sciences is: W. R. Gilks, S. Richardson, and D. J. Spiegelhalter (ed.), *Markov Chain Monte Carlo in Practice* (Chapman and Hall, London, 1996).
67. R. M. Neal, *Bayesian Learning for Neural Networks, Lecture Notes in Statistics 118* (Springer-Verlag, 1996).
68. N. Madras (ed.), *Monte Carlo Methods, Fields Inst. Commun.*, Vol. 26 (American Mathematical Society, 2000).
69. E. Marinari, *Advances in Computer Simulation*, eds. J. Kertész and I. Kondor (Springer-Verlag, 1998), p. 50, *Lecture Note in Physics* (cond-mat/9612010).
70. A. Gelman and X.-L. Meng, *Stat. Sci.* **13**, 163 (1998).
71. G. M. Torrie and J. P. Valleau, *Chem. Phys. Lett.* **28**, 578 (1974); G. M. Torrie and J. P. Valleau, *J. Comp. Phys.* **23**, 187 (1977).
72. S. Kirkpatrick, C. D. Gelatt Jr., and M. P. Vecchi, *Science* **220**, 671 (1983).
73. S. Geman and D. Geman, *IEEE Trans. Pattern Analysis and Machine Intelligence* **6**, 721 (1984).
74. D. D. Frantz, D. L. Freeman, and J. D. Doll, *J. Chem. Phys.* **93**, 2769 (1990).
75. A. Matro, D. L. Freeman, and R. Q. Topper, *J. Chem. Phys.* **104**, 8690 (1996).
76. N. Ogawa and S. Eguchi, *Proceedings of the Institute of Statistical Mathematics* **45**, 377 (1997) [in Japanese].
77. U. H. E. Hansmann, *Chem. Phys. Lett.* **281**, 140 (1997).
78. Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **314**, 141 (1999).
79. Y. Sugita and Y. Okamoto, *Chem. Phys. Lett.* **329**, 261 (2000).
80. Y. Sugita, A. Kitao and Y. Okamoto, *J. Chem. Phys.* **113**, 6042 (2000).
81. P. M. Lee, *Bayesian Statistics: An Introduction* (Oxford University Press, New York, 1989), p. 90; T. Leonard and J. S. J. Hsu, *Bayesian Methods: An Analysis for Statisticians and Interdisciplinary Researchers* (Cambridge University Press, Cambridge, UK, 1999), p. 134.
82. N. B. Wilding and M. Müller, *J. Chem. Phys.* **101**, 4324 (1994).
83. G. R. Smith and A. D. Bruce, *J. Phys. A* **28**, 6623 (1995).
84. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **61**, 2635 (1988).
85. A. M. Ferrenberg and R. H. Swendsen, *Phys. Rev. Lett.* **63**, 1195 (1989).
86. R. H. Swendsen, J.-S. Wang, and A. M. Ferrenberg, *The Monte Carlo Method in Condensed Matter Physics, Topics in Applied Physics*, Vol. 71, 2nd edition, ed. K. Binder (Springer-Verlag, Berlin, 1995), pp. 75–91.
87. J. Lee, *Phys. Rev. Lett.* **71**, 211 (1993) [2353 (erratum)].
88. J.-S. Wang, *Physica A* **281**, 147 (2000).
89. J.-S. Wang, “Efficient Monte Carlo simulation methods in statistical physics,” cond-mat/0103318.
90. P. M. C. de Oliveira, T. J. P. Penna, and H. J. Herrmann, *The Eur. Phys. J. B* **1**, 205 (1998).
91. I. Shteto, J. Linares, and F. Varret, *Phys. Rev. E* **56**, 5128 (1997).
92. J. Higo, N. Nakajima, H. Shirai, A. Kidera, and H. Nakamura, *J. Comput. Chem.* **18**, 2086 (1997).
93. N. Hatano and J. E. Gubernatis, *Prog. Theoret. Phys. Supplement*, No. 138, 442 (2000).

94. W. H. Wong and F. Liang, *Proc. Natl. Acad. Sci.* **94**, 14220 (1997).
95. A. Bunker and B. Dünweg, *Phys. Rev. E* **63**, 016701-(1-10) (2000) (“016701” is an article number used for the identification of a paper in the journal).
96. G. Stolovitzky and B. J. Berne, *Proc. Natl. Acad. Sci.* **97**, 11164 (2000).
97. S. Lin and E. A. Thompson, *J. R. Stat. Soc. B* **55**, 81 (1993); in discussion on Gibbs sampler and other MCMC methods.
98. M. T. Jacobson and P. Matthews, *J. Combinatorial Designs* **4**, 405 (1996).
99. A. D. Sokal, *Monte Carlo and Molecular Dynamics Simulations in Polymer Science*, ed. K. Binder (Oxford University Press, New York, Oxford, 1995), pp. 47–124.
100. Earlier reports on simulations of lattice proteins with a relaxed self-avoiding condition are found in the following papers. Ideas for the systematic control of the penalty to self-overlapped conformations are not found in these works. E. Shakhnovich, G. Farztdinov, A. M. Gutin, and M. Karplus, *Phys. Rev. Lett.* **67**, 1665 (1991); A. Šali, E. Shakhnovich, and M. Karplus, *J. Mol. Biol.* **235**, 1614 (1994).
101. An idea of treating *van der Waals radius* as a dynamical variable in fluid/polymer simulation is already discussed in 1993 by Liu and Berne. Z. Liu and B. J. Berne, *J. Chem. Phys.* **99**, 6071 (1993).
102. D. Geman, S. Geman, C. Graffigne, and P. Dong, *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**, 609 (1990).
103. B. A. Berg, U. Hansmann, and T. Neuhaus, *Phys. Rev. B* **47**, 497 (1993).
104. K. Ding and J. P. Valleau, *J. Chem. Phys.* **98**, 3306 (1993).
105. P. Grassberger, *J. Phys. A* **26**, 1023 (1993).
106. E. J. Janse van Rensburg and S. G. Whittington, *J. Phys. A* **24**, 3935 (1991).
107. B. Baumann, *Nucl. Phys. B* **285**, [FS19], 391 (1987).
108. P. J. Green, *Biometrika* **82**, 711 (1995).
109. C. Andrieu, P. M. Djurić, and A. Doucet, *Signal Processing (special issue on MCMC for Signal Processing)* **81**, 19 (2001).
110. A discussion to the paper by Besag *et al.*: W. H. Wong, *Statist. Sci.* **10**, 52 (1995); See also the rejoinder to the discussions: J. Besag, P. Green, D. Higdon, and K. Mengersen, *Statist. Sci.* **10**, 58 (1995).
111. F. Liang and W. H. Wong, *Phys. Lett. A* **252**, 257 (1999).
112. S. Richardson and P. J. Green, *J. R. Statist. Soc. B* **59**, 731 (1997) [with discussion].
113. U. H. E. Hansmann, Y. Okamoto, and F. Eisenmenger, *Chem. Phys. Lett.* **259**, 321 (1996).