Generalization of the Fortuin-Kasteleyn-Swendsen-Wang representation and Monte Carlo algorithm

Robert G. Edwards and Alan D. Sokal
Department of Physics, New York University, 4 Washington Place, New York, New York 10003
(Received 7 January 1988)

We give a simple explanation of the Swendsen-Wang algorithm for Potts models in terms of a joint model of Potts spin variables interacting with bond occupation variables. We then show how to generalize this representation, as well as the corresponding Monte Carlo algorithm, to arbitrary models. We give initial results of tests of the new algorithm on the two-dimensional XY model.

Two decades ago, Fortuin and Kasteleyn[1-4] found an interesting mapping between the Potts model[5] and a correlated bond-percolation model called the random-cluster model. Very recently, Swendsen and Wang[6] have exploited this mapping to devise an extraordinarily efficient Monte Carlo algorithm for Potts models, which appears[6,7] to have far less critical slowing down than the standard single-spin-update algorithms.[8] In this paper we give a simple explanation of the Fortuin-Kasteleyn representation and the Swendsen-Wang algorithm in terms of a joint model of interacting Potts spins and bond occupation variables. We then show how this representation, as well as the corresponding Monte Carlo algorithm, can be generalized to *any* model.[9] We conclude by giving initial results of tests of the new algorithm on the two-dimensional XY model.

Consider a finite-volume q state Potts model defined by the probability distribution

$$d\mu_{\rm Potts}(\{\sigma\}) = Z_{\rm Potts}^{-1} \exp\left[\sum_{(ij)} J_{ij}(\delta_{\sigma_i,\sigma_j}-1)\right] d\mu_0(\{\sigma\}) \tag{1a}$$

$$= Z_{\rm Potts}^{-1} \prod_{(ij)} [(1-p_{ij})+p_{ij}\delta_{\sigma_i,\sigma_j}]\, d\mu_0(\{\sigma\}) , \tag{1b}$$

where $J_{ij} \geq 0$ for all i,j (ferromagnetism), $p_{ij} \equiv 1-\exp(-J_{ij})$, and $d\mu_0$ is the counting measure. Consider next a random-cluster model[1-4] with parameter q, i.e., a model of bond occupation variables $n_{ij}=0,1$ with the probability distribution

$$d\mu_{\rm RC}(\{n\}) = Z_{\rm RC}^{-1} \left[\prod_{(ij):n_{ij}=1} p_{ij}\right] \left[\prod_{(ij):n_{ij}=0} (1-p_{ij})\right] \times q^{\mathcal{C}(\{n\})} d\mu_0(\{n\}) , \tag{2}$$

where $\mathcal{C}(\{n\})$ is the number of connected components (including one-site components) in the graph whose edges are the bonds having $n_{ij}=1$, and $d\mu_0$ is again counting measure. Finally, let us define a joint model,[10] which for lack of imagination we call the Fortuin-Kasteleyn-Swendsen-Wang (FKSW) model, having q-state Potts spins σ_i at the sites and occupation variables n_{ij} on the bonds, with joint probability distribution

$$d\mu_{\rm FKSW}(\{\sigma\},\{n\}) = Z_{\rm FKSW}^{-1} \prod_{(ij)} [(1-p_{ij})\delta_{n_{ij},0} + p_{ij}\delta_{n_{ij},1}\delta_{\sigma_i,\sigma_j}] \times d\mu_0(\{\sigma\})\, d\mu_0(\{n\}) . \tag{3}$$

By performing explicitly in (3) the summation over either the $\{\sigma\}$ or the $\{n\}$ variables, it is easy to verify the following facts about the FKSW model.

(i) $Z_{\rm Potts} = Z_{\rm RC} = Z_{\rm FKSW}$.

(ii) The marginal distribution on the Potts variables $\{\sigma\}$ is precisely the Potts model $d\mu_{\rm Potts}$.

(iii) The marginal distribution on the bond occupation variables $\{n\}$ is precisely the random-cluster model $d\mu_{\rm RC}$.

(iv) The conditional distribution of the $\{n\}$ given the $\{\sigma\}$ is as follows: independently for each bond (ij), one sets $n_{ij}=0$ in case $\sigma_i \neq \sigma_j$, and sets $n_{ij}=0,1$ with probability $1-p_{ij}, p_{ij}$, respectively, in case $\sigma_i=\sigma_j$.

(v) The conditional distribution of the $\{\sigma\}$ given the $\{n\}$ is as follows: independently for each connected cluster, one sets all the spins σ_i in the cluster to the same value, chosen equiprobably from the set $\{1,\ldots,q\}$.

The joint model (3) makes it easy to understand both the Fortuin-Kasteleyn representation and the Swendsen-Wang Monte Carlo algorithm. Indeed, the Fortuin-Kasteleyn identities[1,2,4] are immediate consequences of facts (ii), (iii), and (v). The Swendsen-Wang algorithm simulates the joint model (3) by alternately applying the conditional distributions (iv) and (v)—that is, by alternately generating new bond occupation variables (independent of the old ones) given the spins, and new spin variables (independent of the old ones) given the bonds. (Both of these operations can be carried out in a computer time of order volume.[11-13]) The joint model (3) can also be used to derive Griffiths-type correlation inequalities for Potts models.[4,14,15]

Consider now an *arbitrary* statistical-mechanical model with dynamical variables $\{\phi\}$ and probability distribution

$$d\mu(\{\phi\}) = Z^{-1} \left[\prod_b W_b(\{\phi\})\right] d\mu_0(\{\phi\}) , \tag{4}$$

where the Boltzmann weight is decomposed as a product of terms W_b which we assume normalized to $0 \leq W_b \leq 1$, and $d\mu_0$ is some *a priori* measure (which is usually a prod-

 ©1988 The American Physical Society

uct measure, though it need not be). In most applications the terms W_b in (4) would be associated with sites, links, or plaquettes of a regular lattice, but they need not have any geometric significance at all. We now introduce, for each b, a *real-valued* variable κ_b taking values in the interval [0,1], and use the trivial identity $W_b = \int_0^1 \theta(W_b - \kappa_b) d\kappa_b$. This leads us to define a joint model of the variables $\{\phi\}$ and $\{\kappa\}$, with probability distribution

$$d\mu_{\text{joint}}(\{\phi\},\{\kappa\}) = Z_{\text{joint}}^{-1} \left[\prod_b \theta(W_b(\{\phi\}) - \kappa_b) \right] \times d\mu_0(\{\phi\}) \prod_b d\kappa_b \ , \tag{5}$$

where $d\kappa_b$ is Lebesgue measure on [0,1]. It is immediately obvious that the joint model (5) has the following properties.

(i′) $Z = Z_{\text{joint}}$.

(ii′) The marginal distribution on the $\{\phi\}$ variables is precisely the original model $d\mu$.

(iv′) The conditional distribution of the $\{\kappa\}$ given the $\{\phi\}$ is as follows: independently for each "bond" b, κ_b is a real number uniformly distributed in the interval $[0, W_b(\{\phi\})]$.

(v′) The conditional distribution of the $\{\phi\}$ given the $\{\kappa\}$ is the restriction of the *a priori* measure $d\mu_0$ to the set of $\{\phi\}$ satisfying the constraints $W_b(\{\phi\}) \geq \kappa_b$ for all b.

The analogy with (1)–(3) is clear. Indeed, the Potts model is the special case of (4) and (5) in which the partial Boltzmann factors W_b take only the two values 1 and $w_{b,\min}$, so that all relevant information in the real-valued variables κ_b is, in fact, contained in the dichotomous variables $n_b \equiv \theta(\kappa_b - w_{b,\min})$.

The ferromagnetic Potts case does, however, have simplifying features not possessed by the general case. First, the marginal distribution on the $\{\kappa\}$ variables is not, in general, simple. [The weight factor $\mathcal{W}(\{\kappa\})$ for this distribution is precisely the partition function of the conditional ensemble (v′). But this partition function is not usually expressible in closed form.] Second, the conditional distribution of $\{\phi\}$ given $\{\kappa\}$ is not, in general, amenable to static Monte Carlo simulation—that is, there does not, in general, exist a simple (e.g., order-volume) algorithm to generate *independent* samples from this distribution. Nevertheless, a legitimate Monte Carlo algorithm for simulating the joint distribution (5) can be based on alternately applying the conditional distribution (iv′) and *any updating process which leaves invariant the conditional distribution* (v′). In what follows, we shall use the term *pure Swendsen-Wang* (SW) *algorithm* to denote one which uses independent samples from the conditional distribution (v′), and *extended* SW *algorithm* to denote one which uses a more general updating strategy. The critical slowing down of an extended SW algorithm will be a combination (product?) of the critical slowing down in the corresponding pure SW algorithm along with the critical slowing down, if any, in the auxiliary updating process.

Consider, for example, a ferromagnetic XY model with probability distribution

$$d\mu(\{\theta\}) = Z^{-1} \prod_{(ij)} \exp\{\beta[\cos(\theta_i - \theta_j) - 1]\} \prod_i d\theta_i \ , \tag{6}$$

where the first product runs over nearest-neighbor pairs (ij). Then the conditional distribution of $\{\theta\}$ given $\{\kappa\}$ is simply the Haar measure $\prod_i d\theta_i$ restricted to the set of $\{\theta\}$ satisfying the constraints $|\theta_i - \theta_j| \leq \Delta_{ij}$, where

$$\Delta_{ij} \equiv \begin{cases} \arccos\left[1 + \dfrac{1}{\beta} \ln \kappa_{ij}\right] & \text{if } \kappa_{ij} > \exp(-2\beta) \ , \\ \pi & \text{if } \kappa_{ij} \leq \exp(-2\beta) \ . \end{cases} \tag{7}$$

This is a zero-temperature step model[16] with random $\{\Delta\}$ (bond disorder). Unfortunately, we know of no efficient way to generate *independent* samples from the step-model distribution. Nevertheless, it is easy to devise heat-bath and multigrid[17–19] algorithms which leave the step-model

TABLE I. Susceptibility χ (first data row) and autocorrelation time $\tau_{\text{int},\mathcal{M}^2}$ (second data row) for two-dimensional XY model on 16×16 lattice. Standard error is shown in parentheses. The numbers m_1 and m_2 are the numbers of heat-bath sweeps before and after coarsening. Multigrid with $\gamma = 0$ is equivalent to $m_1 + m_2$ sweeps of the heat-bath method. Note that $\beta_{\text{critical}} \approx 1.13$ (Ref. 19).

	Extended Swendsen-Wang		Multigrid ($m_1 = 1, m_2 = 1$)	
β	$n = 20$	$n = 40$	$\gamma = 0$ (Heat bath)	$\gamma = 2$
0.88	57.3 (0.6)	56.4 (0.7)	57.3 (0.6)	55.7 (0.7)
	5.4 (0.4)	5.4 (0.5)	7.9 (0.5)	2.6 (0.2)
0.92	76.4 (0.6)	75.0 (0.6)	75.4 (0.6)	75.4 (0.7)
	5.2 (0.4)	5.5 (0.4)	8.2 (0.6)	2.5 (0.2)
0.96	92.8 (0.5)	92.4 (0.6)	92.5 (0.5)	93.8 (0.6)
	5.0 (0.3)	6.1 (0.4)	6.1 (0.4)	2.0 (0.1)
1.00	106.9 (0.4)	106.6 (0.5)	106.2 (0.5)	106.2 (0.5)
	3.6 (0.2)	4.0 (0.3)	6.5 (0.4)	1.7 (0.1)
1.04	117.3 (0.4)	117.6 (0.3)	117.5 (0.4)	117.4 (0.4)
	3.4 (0.2)	2.2 (0.1)	4.9 (0.3)	1.3 (0.1)
1.13	134.2 (0.4)		134.4 (0.2)	134.3 (0.3)
	1.6 (0.1)		3.2 (0.1)	0.8 (0.1)

TABLE II. Same as Table I, on 32×32 lattice.

	Extended Swendsen-Wang			Multigrid ($m_1=1, m_2=1$)	
β	$n=20$	$n=40$	$n=80$	$\gamma=0$ (Heat bath)	$\gamma=2$
0.92	135.8 (4.2)	138.7 (4.6)	141.3 (3.0)	142.4 (2.2)	139.6 (3.0)
	10.4 (1.8)	9.4 (1.7)	11.0 (1.3)	30.5 (2.6)	5.4 (0.7)
0.96	219.4 (6.2)	236.7 (5.8)	227.9 (3.4)	223.0 (2.9)	227.1 (3.9)
	17.5 (3.8)	13.3 (2.8)	11.4 (1.4)	42.3 (4.3)	6.9 (0.9)
1.00	312.7 (5.3)		308.9 (3.2)	311.4 (2.1)	309.2 (3.0)
	15.3 (3.1)		11.5 (1.4)	27.0 (2.2)	5.0 (0.6)
1.04	373.8 (1.4)	372.9 (1.6)		372.8 (1.6)	373.3 (2.2)
	6.2 (0.4)	8.1 (0.6)		20.5 (1.4)	3.6 (0.4)
1.13	459.1 (1.1)			460.0 (0.9)	459.8 (0.9)
	1.4 (0.1)			11.0 (0.6)	1.1 (0.1)

distribution invariant. Our experiments on multigrid algorithms for the two-dimensional step model based on standard geometric coarsening ($2^n\times2^n$ blocks with piecewise-constant injection) were not particularly successful: significant critical slowing-down was observed, in fact more than is found in a straight multigrid simulation of the XY model.[19] We think, however, that more efficient multigrid algorithms for the step model can be devised, based on nonstandard coarsenings which are adapted to the particular pattern of coefficients $\{\Delta\}$ observed on each grid.[20] In order to study the dynamic critical behavior of SW-type algorithms for the two-dimensional XY model independently from the problem of devising a good updating procedure for the step model, we made runs in which a standard multigrid updating for the random step model is applied n times for each fixed set of $\{\kappa\}$. As $n\to\infty$ this simulates the performance of the pure SW algorithm.

The results of these simulations are shown in Tables I–III. We have also included data[19] on heat-bath and multigrid algorithms, for purposes of comparison. Here χ is the susceptibility per lattice site, and $\tau_{\text{int},\mathcal{M}^2}$ is the integrated autocorrelation time for the squared magnetization (which is one of the slowest modes[21]), defined by

$$\tau_{\text{int},A}=\tfrac{1}{2}\sum_{t=-\infty}^{\infty}\rho_{AA}(t)\,, \tag{8}$$

where $\rho_{AA}(t)$ is the normalized time-autocorrelation function for the observable A. Error bars were determined by standard procedures of statistical time-series analysis.[22]

It can be seen that the SW and multigrid algorithms behave for the XY model in a qualitatively similar way: there is significant critical slowing down (with roughly the same dynamic critical exponent as in the heat-bath algorithm) on the high-temperature side of finite-size-shifted criticality, but no critical slowing down whatsoever on the low-temperature side. Apparently the SW and multigrid algorithms are effective in creating spin waves, but not in creating widely separated vortex-antivortex pairs. Unfortunately, for the SW algorithm we do not have a good physical understanding of why this is the case.

From a practical point of view, the performance of the (extended) SW algorithm for the XY model is mildly disappointing: equally good results can be obtained with

TABLE III. Same as Table I, on 64×64 lattice.

	Extended Swendsen-Wang			Multigrid ($m_1=1, m_2=1$)	
β	$n=20$	$n=40$	$n=80$	$\gamma=0$ (Heat bath)	$\gamma=2$
0.92	163.5 (4.8)	165.6 (9.1)		167.2 (4.7)	156.2 (4.9)
	10.6 (1.2)	13.2 (2.8)		53.5 (6.1)	5.6 (0.7)
0.96	376.6 (13.6)	389.6 (21.1)		371.7 (9.0)	370.1 (13.1)
	25.0 (4.5)	29.3 (8.2)		122.5 (14.9)	11.8 (2.1)
1.00	781.1 (36.5)	756.0 (28.2)	803.5 (22.7)	793.2 (12.4)	799.3 (19.7)
	54.4 (20.8)	31.5 (9.2)	24.0 (6.1)	147.0 (19.5)	17.5 (3.8)
1.04	1137.5 (16.8)	1164.5 (13.2)	1146.0 (16.0)	1156.6 (7.9)	1174.4 (8.6)
	17.0 (3.6)	12.3 (2.2)	16.2 (3.4)	91.9 (9.7)	5.3 (0.6)
1.13	1572.4 (3.4)			1581.6 (5.0)	1575.8 (3.5)
	1.3 (0.1)			34.5 (3.1)	1.4 (0.1)

the multigrid algorithm, at a vastly reduced computational cost. But from a physical point of view we find these results encouraging: they show that the (extended) SW algorithm lies in a different dynamic universality class from the standard single-spin-update algorithms. This fact, combined with the extraordinary performance of the SW algorithm for Potts models,[6,7] suggests that appropriate generalizations of the Swendsen-Wang idea could lead to significant reductions in critical slowing down. What is needed, therefore, is a good physical understanding of the dynamic properties of the SW algorithm. More generally, one would like to understand how the introduction of auxiliary variables (like our $\{\kappa\}$) affects the dynamic critical behavior. In our view, further work, both numerical and theoretical, on a variety of models is now in order.[23]

This work was begun while one of us (A.D.S.) was a visitor at the Scuola Normale Superiore, Pisa. We would like to thank Robert Swendsen for communicating his results piror to publication, and Richard Brower, Jonathan Goodman, David Ritchie, Jian-Sheng Wang, Ulli Wolff, Jean Zinn-Justin, and Dan Zwanziger for helpful conversations and correspondence. This work was supported in part by National Science Foundation Grants Nos. PHY-8413569 and DMS-8705599 and the John Von Neumann Supercomputer Center Grant No. NAC-705.

[1]P. W. Kasteleyn and C. M. Fortuin, J. Phys. Soc. Jpn. **26** (Suppl.), 11 (1969).

[2]C. M. Fortuin and P. W. Kasteleyn, Physica (Utrecht) **57**, 536 (1972).

[3]C. M. Fortuin, Physica (Utrecht) **58**, 393 (1972).

[4]C. M. Fortuin, Physica (Utrecht) **59**, 545 (1972).

[5]F. Y. Wu, Rev. Mod. Phys. **54**, 235 (1982); **55**, 315(E) (1983).

[6]R. H. Swendsen and J.-S. Wang, Phys. Rev. Lett. **58**, 86 (1987).

[7]R. G. Edwards and A. D. Sokal (in preparation).

[8]G. F. Mazenko and O. T. Valls, Phys. Rev. B **24**, 1419 (1981); J. K. Williams, J. Phys. A **18**, 49 (1985); R. B. Pearson, J. L. Richardson, and D. Toussaint, Phys. Rev. B **31**, 4472 (1985); S. Wansleben and D. P. Landau, J. Appl. Phys. **61**, 3968 (1987); S. Jain, J. Phys. A **20**, L607 (1987).

[9]A different generalization of the Swendsen-Wang algorithm has been proposed by U. Wolff, Phys. Rev. Lett. **60**, 1461 (1988); Nucl. Phys. B [FS] (to be published). His algorithm is simpler than ours in that it simulates a dilute version of the original model; however, it has the disadvantage that the transformation involved is not exact, and the resulting systematic error is uncontrolled. (Alternatively, his algorithm can be thought of as simulating a well-defined variant action whose relationship to the original model is uncontrolled.)

[10]This joint model is the special case $(p_B)_{ij} = 1 - \exp(-J_{ij})$ of a correlated site-bond percolation model: see A. Coniglio and W. Klein, J. Phys. A **13**, 2775 (1980); A. Coniglio and F. Peruggi, *ibid.* **15**, 1873 (1982).

[11]See, for example, E. M. Reingold, J. Nievergelt, and N. Deo, *Combinatorial Algorithms: Theory and Practice* (Prentice-Hall, Englewood Cliffs, NJ, 1977), Chap. 8.

[12]J. Hoshen and R. Kopelman, Phys. Rev. B **14**, 3438 (1976).

[13]R. G. Edwards and A. D. Sokal (in preparation).

[14]M. Aizenman, J. T. Chayes, L. Chayes, and C. M. Newman, J. Phys. A **20**, L313 (1987); J. Stat. Phys. **50**, 1 (1988).

[15]A. D. Sokal (in preparation).

[16]A. J. Guttmann and G. S. Joyce, J. Phys. C **6**, 2691 (1973); M. N. Barber, J. Phys. A **16**, 4053 (1983); I-H. Lee and R. E. Shrock, Phys. Rev. B **36**, 3712 (1987).

[17]J. Goodman and A. D. Sokal, Phys. Rev. Lett. **56**, 1015 (1986). Very similar ideas were proposed independently by A. Brandt, D. Ron, and D. J. Amit, in *Multigrid Methods II*, edited by W. Hackbusch and U. Trottenberg (Lecture Notes in Mathematics, Vol. 1228) (Springer, Berlin, 1986), Sec. 7.1.

[18]J. Goodman, A. D. Sokal, and D. Zwanziger (in preparation).

[19]R. G. Edwards, J. Goodman, A. D. Sokal, and D. Zwanziger (in preparation).

[20]Similar ideas arise in applying multigrid to partial differential equations with rapidly varying coefficients and in the so-called algebraic multigrid (AMG).

[21]In the heat-bath algorithm, the magnetization vector $\mathcal{M}$ has a vastly longer autocorrelation time.

[22]M. B. Priestley, *Spectral Analysis and Time Series* (Academic, London, 1981), Vols. 1 and 2, Chaps. 5–7. We used a self-consistent truncation window of width $6\tau_{\text{int},A}$. For more details, see N. Madras and A. D. Sokal, J. Stat. Phys. **50**, 109 (1988), Appendix C.

[23]Other generalizations of the SW algorithm can be found in D. Kandel *et al.*, Phys. Rev. Lett. **60**, 1591 (1988); F. Niedermayer, Eötvös University report, 1988 (unpublished).