Revista Brasileira de Física, Vol. 21, nº 4, 1991

Monte Carls Determination of the Convergence Time of Two Cluster-Flfp Algorithms in the Psing Model

Alejandro *C*. Frery
INPE — DPI, Av. dos Astronautas, 1758, São José dos Campos, 12201, SP - Brasil

Received June 20, 1991; in final form September 9, 1991

Abstract We use the maximum pseudolikehood estimator as a stopping rule for two cluster-flip algorithms in the Ising model. Swendsen-Wang and Wolff dynamics are compared by means of the number of iterations and CPU time required to achieve convergence for different values of the atractivity parameter β.

Monte Carlo techniques are a powerful method for investigating problems in Statistical physics[1,2]. The advent of fast and relatively inexpensive computing machines, with enormous memory capabilities, has helped the study of thermodynamic systems at phase transitions and lattice gauge theory calculations, to name a few applications[3].

One of the problems that arise when Monte Carlo techniques are used is known as *critical slowing down.* It is related to the fact that, for some models, there is a discontinuity of the thermodynamic properties in different parameter regions producing high correlations between observations. This reduces, in a known fashion, the effective size of the considered[4] sample. A model where this occurs is the system of interacting particles known as the Potts model (also as the Potts-Strauss model or as the Multilevel Logistic model), which has the well known Ising model as the simplest non-trivial case.

Aiming at the solution of the critical slowing down problem, Swendsen and Wang proposed in 1987, c. f. Ref. 5, a novel technique for the simulatioa of this model. Their idea is based upon a mapping between the Potts model and a percolation model, due to Kasteleyn and Fortuin, and allows the updating of

more than one spin per iteration. That dynamic differs from the more classical dynamics (Metropolis, Gibbs sampler[6]) in that these allow the change of only one site per iteration. It is successfully used for the understanding of the metastable behaviour[7,8,9] and of the geometrical features of systems of interacting particles[10].

This cluster technique was then modified by Wolff in 1989[11] and generalized to other cases: the x – y and O(n) nonlinear- a two-dimensional models[12]. Wolff's technique has also inspired experiments with similar dynamics[13], in dimensions greater than 2 [14], and other problems such as Z_2 lattice gauge and ϕ^4 theories[15].

Most of the literatre related to these algorithms deals with the Monte Carlo evaluation of certain characteristic quantities for different models (integrated autocorrelation times in the two-dimensional two- and three- states Potts model, for instance[13]). In this work, we are interested in *previous* quantities for both dynamics: the required number of iterations and CPU times to achieve convergence in the two- dimensional Ising model, when the initial measure is a collection of independent identically distributed Bernoulli(1/2) random variables. To our knowledge, there are no published works determining these values.

The results presented here are part of a series of experiences to be carried out on systems of interacting particles, bearing in mind their application to digital image segmentation, restoration and classification. A precise knowledge of the convergence properties of the used algorithms is paramount for these applications.

In previous works[16,17] we checked that, for the 128 x 128 free boundary Ising model the Swendsen-Wang dynamic is up to 100 times faster, in terms of convergence time, than the Gibbs sampler. This improvement is quite evident for supercritical values of the parameter, i. e. $\beta > \beta_c = \ln(1+\sqrt{2})$ and mainly due to the capability of the Swendsen-Wang algorithm to escape quite rapidly from metastable configurations that trap the evolution of the Gibbs sampler (a spin-flip dynamic).

The Wolff algorithm is supposed to decorrelate successive iterations faster than the Swendsen-Wang, and this assertion makes sense only when these iterations are obtained **after** the convergence has been achieved. In this work, our aim is to compare the convergence properties of both dynamics via a Monte Carlo

experiment. We show numerical evidence that the Swendsen-Wang dynamic is much faster than the Wolff dynamic for the studied 128×128 free boundary Ising model, in the considered range of the parameter $\beta \in [0.7; 1.2]$ and under the afore-mentioned initial measure.

Consider the usual finite size free boundary ferromagnetic Ising model characterized by the state space $\Xi = \prod_{s \in S} \Xi_s = \{-1, +1\}^{\#S}$, with $S = (1, \ldots, N) \times \{1, \ldots, N)$ and local characteristics

$$\Pr(X_s = x_s | X_{S \setminus \{s\}} = x_{S \setminus \{s\}}) = \frac{\exp\left\{\beta \frac{1+x_s}{2} \Sigma_s\right\}}{1 + \exp\left\{\beta \Sigma_s\right\}},$$

$$\Sigma_s = \sum_{\|s-t\|=1} x_t \forall s \in S. \tag{1}$$

$\beta = 1/T > 0$ is the reciproca1 of the temperature. Negative values of β could also be considered, yielding the *antiferromagnetic* Ising model. We shall refer to the $\beta = 0$ model as the *infinite temperature* model. Notice that, in this last case, the variables $[X_s]_{s \in S}$ are a collection of independent identically distributed Bernoullis with $\Pr(X_s = +1) = \Pr(X_s = -1) = 1/2$.

It can be seen in Ref. 18 that given $x \in \Xi$, a good estimator of β is the *maximum pseudo-likehood estirnator* defined as $\widehat{\beta}(x) = \sup_{\beta}^{-1} \prod_{s \in W \subset S} \Pr(X_s = x_s | X_{\partial_s} = x_{\partial_s})$, where the *neighbourhood* sets, $\partial_s \subset S \setminus \{s\}$, satisfy the Markov condition $\Pr(X_s = x_s | X_{S \setminus \{s\}} = x_{S \setminus \{s\}}) = \Pr(X_s = x_s | X_{\partial_s} = x_{\partial_s})$. In this case, it is given by the zero of the following expression:

$$\begin{aligned} 2(C_4^1 - C_{-4}^1) + (C_2^1 - C_{-2}^1) + 2(C_{-4}^{-1} + C_4^1) \frac{\exp\{-4\widehat{\beta}\}}{1 + \exp\{-4\widehat{\beta}\}} \\ + (C_{-2}^{-1} + C_2^1) \frac{\exp\{-2\widehat{\beta}\}}{1 + \exp\{-2\widehat{\beta}\}} - (C_2^{-1} + C_2^1) \frac{\exp\{2\widehat{\beta}\}}{1 + \exp\{2\widehat{\beta}\}} \\ - 2(C_4^{-1} + C_{-4}^1) \frac{\exp\{4\widehat{\beta}\}}{1 + \exp\{4\widehat{\beta}\}} = 0, \end{aligned} \tag{2}$$

where we have written $C_k^{\ell} = \#\{x_w : w \in W | x_w = \ell, \Sigma = k)$, $\Sigma_w = \sum_{\|w-t\|=1} x_t$ and $W = (2, \ldots, N - 1) \times \{2, \ldots, N - 1)$. For details on the behaviour of this estimator see Ref. 19.

The considered dynamics are known to converge in distribution to $\Pr(X = x)$, the unique joint distribution of the random vector $X = [X_s]_{s \in S}$ induced by the local characteristics (1) given by

$$\Pr(X = x) = Z_\beta^{-1} \exp\left\{2\beta \sum_{\|s-t\|=1} x_s x_t\right\}, \tag{3}$$

where $Z_\beta = \sum_{x \in \Xi} \exp\{2\beta \sum_{\|s-t\|=1} x_s x_t\}$ is the partition function. A rigorous proof of the Swendsen-Wang algorithm convergence can be seen in Refs. 20, 21. Both algorithms act iteratively on successive vectors $x(k) = [x_s(k)]_{s \in S}$ in such a way that $\Pr(X(k) = x) \longrightarrow \Pr(X = x)$ as $k \to \infty$ for every $x(0) \in \Xi$. They can be described, for $\beta \geq 0$, by the iteration: let $k = 0$ and change $x(k)$ to $x(k+1)$ as follows:

1. Place a bond between all equal-valued nearest neighbours of $x(k)$ with probability $1 - \exp\{-\beta\}$.
2. Construct C, the set of all connected components of the graph obtained in Step 1.
3. Form $C' \subset C$, the set of connected components that will change their state.
4. Change $x_s(k+1) = (-1)^{\mathbf{1}_{\{s \in C'\}}} x_s(k)$ for every $s \in S$.
5. Go to Step 1 if some condition is not satisfied.

The algorithms differ in Step 3.: how to form C', the set of connected components that will change of state. The specifications are as follows:

Suppose that in the kth iteration there are $L(k)$ connected components in the graph induced by Step 1 above, i. e. $C = \{\lambda_1, \ldots, \lambda_{L(k)}\}$; now, observe $U_1, \ldots, U_{L(k)}$, a sequence of $L(k)$ independent identically distributed $[0;1]$ uniform random variables. Swendsen-Wang constructs C' as

$$C'_{\mathrm{SW}} = \{\lambda_j \in C : U_j < 1/2,\ 1 \leq j \leq L(k)\}.$$

For the other algorithm, observe $U_1, \ldots, U_{\#S}$, a sequence of $\#S$ independent identically distributed $[0;1]$ uniform random variables and form C' as

$$C'_{\mathrm{Wolff}} = \{\lambda_j \in C : x_m \in \lambda_j \quad \text{such that} \quad U_m = \sup_{1 \leq i \leq \#S} \{U_i\}\}.$$

In words, the Swendsen-Wang dynamics allows the flipping of (all the states of) every connected component per iteration with probability $1/2$, while the Wolff dynamic always switches (al the states of) only one connected component every iteration. Another interpretation for the Wolff construction consists in saying that the connected component $\lambda_j \in \mathcal{C}'_{\text{Wolff}}$ is chosen among the elements of C with probabilities proportional to their sizes.

As previously stated, $\Pr(X(k) = \text{x}) \longrightarrow \Pr(X = x)$ as $\text{k} \rightarrow \infty$ for every $x(0) \in \Xi$ and, among the available measures of this convergence, we chose to work with $\widehat{\beta}(x(k))$. The initial distribution from which $x(0)$ is sampled was the infinite temperature model, i. e. the uniform measure over Ξ, where

$$\Pr(X(0) = x(0)) = (\#\Xi)^{-1} \forall x(0) \in \Xi.$$

Starting from one such $x(0)$, we say that the convergence has been achieved in the kth iteration if $\widehat{\beta}(x(k)) \geq \beta$, and define

$$k^* = \inf\{k \in \mathbb{N}: \widehat{\beta}(x(k)) \geq \beta\};$$

our previous experiences strongly suggest that this is a good stopping rule for the iterative procedure when $X(0)$ is as defined above, yielding quite stable measurements of $\widehat{\beta}(x(k))$, $|\sum_{s \in S} x_s(k)|$, $K_1 \sum_{\|s-t\|=1} x_s(k) x_t(k)$ and $K_2 \sum_{\|s-t\|=10} x_s(k) x_t(k)$; these last three quantities are the estimates of, respectively, the absolute mean magnetization M_β, and the short and long range correlations.

In other words, we propose the use of $\widehat{\beta}(x(k))$ as a measure of $|\Pr X - \Pr(X(k))|$ when $X(0)$ is the infinite temperature model and when $x(k)$ is transformed into $x(k+1)$ using either Swendsen-Wang or Wolff dynamics. This measure fails to detect lack of convergence for the same initial measure when the used dynamic is the Gibbs Sampler algorithm; see results in Refs. 16, 17.

Figure 1 (Figure 2 respectively) shows the Monte Carlo means and standard deviation of k* for 101 (11) values of β in the range $[0.7; 1.2]$, as obtained with 100 independent runnings of the Swendsen-Wang (Wolff) algorithm for every value of β. The Table below summarizes 11 of the values shown in the Figures for direct

comparison. t^* is the required CPU seconds to achieve k* in a SUN SPACK 2 station.

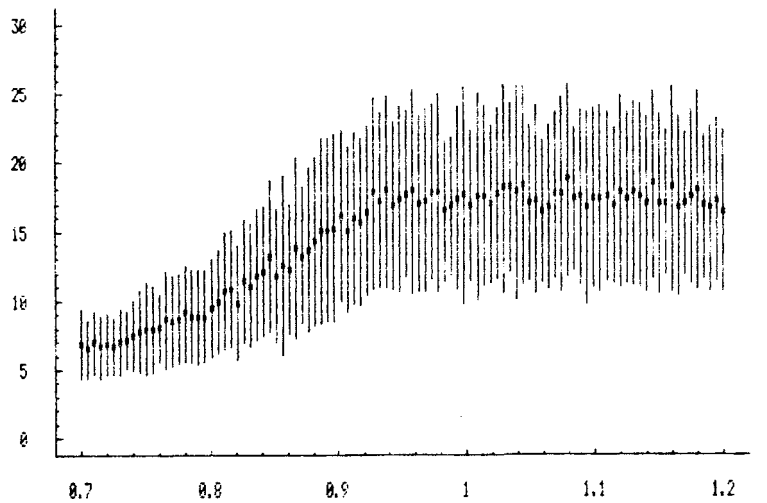

Fig. 1 – Monte Carlo results for k^* and the Swendsen-Wang algorithm.

Table

β	$\langle k^* \rangle$(STD) -SW-	$\langle t^* \rangle$(STD) -SW-	$\langle k^* \rangle$(STD) -W-	$\langle t^* \rangle$(STD) -W-
0.70	6.79(2.49)	0.49(0.21)	11447.21(4874.93)	377.72(299.84)
0.75	7.99(3.36)	0.54(0.20)	7960.50(1933.68)	185.58(46.60)
0.80	9.52(3.58)	0.67(0.25)	4921.27(1061.16)	230.16(49.94)
0.85	11.82(4.86)	0.91(0.45)	3015.48(435.82)	155.78(58.34)
0.90	16.12(6.19)	1.14(0.44)	1163.37(109.90)	65.11(13.82)
0.95	17.76(5.94)	1.26(0.43)	644.27(62.68)	35.17(2.92)
1.00	16.89(5.43)	1.41(0.63)	459.74(58.39)	26.33(2.66)
1.05	17.32(6.79)	1.38(0.53)	365.13(44.88)	34.13(5.91)
1.10	17.44(6.69)	1.54(0.58)	306.78(46.57)	28.08(3.60)
1.15	17.16(5.22)	1.38(0.55)	262.10(40.64)	24.62(2.80)
1.20	17.84(6.99)	1.49(0.61)	228.06(35.85)	26.24(4.00)

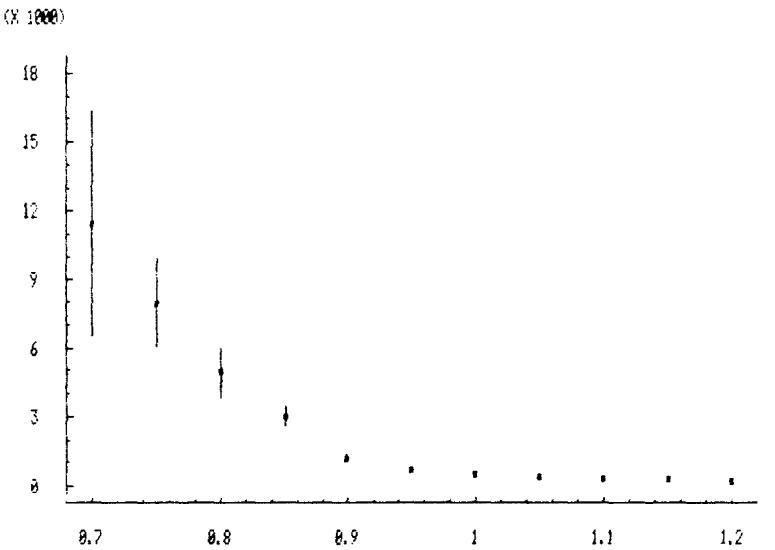

Fig. 2 – Monte Carlo results for k* and the Wolff algonthm.

In order to make a fair comparison between dynamics, we chose to work with a recursive algorithm[22] that returns the connected components of a graph in such a way that it is possible to construct either all the components or only one component. The first option was used for the Swendsen-Wang dynamic while the second was used for the Wolff dynamic.

Conclusions

Most of the papers cited in the References report a superior behaviour of Wolff's algorithm and use the integrated autocorrelation times as the companson scale. Also, in the cited literature, one iteration is considered, for both dynamics, when at *least* $\#S$ particles have been flipped.

We were concerned witb the associated times of the proposed measure of $|\Pr(X) - \Pr(X(k))|$ and one iteration, for us, was defined in a sligthly different manner: we find this definition more natural for the applications we bear in mind. Within this context, Swendsen-Wang's algorithm surpasses Wolff's.

Our results are different, mainly because the expected number of spins flipped by our definition of iteration is, for Wolff dynamic given approximately by $\sqrt{\#S}\,\langle M_\beta^2\rangle$ and for Swendsen-Wang, by $\#S/2$, c. f. Ref. **14.** The analytic expression for $\langle M_\beta\rangle$ given in ref. 23

$$\langle M_\beta\rangle = \begin{cases} 0 & \text{if } \beta < \beta_c \\ \sqrt{\sqrt{\dfrac{[1+\exp\{-2\beta\}]\sqrt{1-6\exp\{-2\beta\}+\exp\{-4\beta\}}}{1-\exp\{-2\beta\}}}} & \text{if } \beta \geq \beta_c, \end{cases}$$

implies that most of Wolff dynamic's iterations will be spent in the flipping of a few sites, for the subcritical case, increasing the required time to achieve convergence; as β increases the Wolff dynamic becomes more efficient. Since the number of sites flipped per iteration by the Swendsen-Wang dynamic is approximately constant, the different number of iterations required to achieve convergence does not change so dramatically when β changes.

Admitting the claim that the Wolff dynamic decorrelates successive samples of the Potts model faster than the Swendsen-Wang dynamic, c. f. Ref. 12, 14 along with the results presented in this paper, a convenient set up for a Monte Carlo evaluation of some quantity $\langle\psi(X)\rangle$, X being the Potts model, would be: i) achieve convergence using the Swendsen-Wang dynamic, and ii) estimate $\langle\psi(X)\rangle$ using $\langle\widehat{\psi(X)}\rangle$ given by

$$\langle\widehat{\psi(X)}\rangle = N^{-1}\sum_{1\leq k\leq N}\psi\Big(x(Kk^*+\ell k)\Big),$$

where k* is the quantity studied in this paper, K $>$ 1 is a safety integer factor, ℓ a convenient fixed lag, $x(0)$ an initial sample, $x(1),\ldots,x(Kk^*)$ iterations with the Swendsen-Wang dynarnic and $x(Kk^*+1),\ldots,x(Kk^*+\ell N)$ iterations with the Wolff dynamic. In other words: iterate Kk* times with the Swendsen-Wang algorithm to achieve convergence and then use N samples taken by skipping $\ell-1$ iterations of the Wolff algorithm.

This procedure should ensure a good approximation of a sample of N independent identically distributed observations of $\psi(X)$, X being the Potts model.

Acknowledgment

The author wishes to thank the anonymous referee that contributed to the readability of this work and to Maria Eulália Vares (IMPA-Brazil) for suggesting the study of these dynamics.

1. **K.** Binder, (ed): ***Monte Carlo Methods in Statistical Physics,*** 2nd ed. Springer-Verlag, Berlin. **1986.**
2. K. Binder, ***Applications of the Monte Carlo Method in Statistical Physics,*** 2nd ed. Springer-Verlag, Berlin. **1987.**
3. R. Swendsen, Comp. Phys. Comm. 65, **281 (1991).**
4. **H.** Müller-Krumbhaar, and K. Binder, J. Statist. Phys. 8, **1 (1972).**
5. R. Swendsen, and J. Wang, Phys. Rev. Lett. 58, **86 (1987).**
6. S. Geman, and D. Geman, I. E. E. E. Trans. **PAMI-6, 721 (1984).**
7. T. Ray, and P. Tamayo, J. Statist. Phys. **60, 851 (1990).**
8. E. Miranda, Physica A-175, **235 (1991).**
9. M. D'Onorio De Meo, D. Heermann, and K. Binder, J. Statist. Phys. 60, **585 (1990).**
10. E. Miranda, Physica A-175, **229 (1991).**
11. **U.** Wolff, Phys. Rev. Lett. 62, **361 (1989).**
12. R. Edwards, and A. Sokal, Phys. Rev. D-40, **1374 (1989).**
13. C. Baillie, and P. Coddington, Phys. Rev. B-43, **10617 (1991).**
14. U. Wolff, Phys. Lett. B-228, **379 (1989).**
15. R. Ben-Av, **D.** Kandel, E. Kantznelson, P. Lauwers, and S. Solomon, J. Statist. Phys. 58, **125 (1990).**
16. **O.** Bustos, and A. Frery, Markovian Fields as Limits of Markov Chains. ***Proc. Second Intern. Conf. on Use of Computers in Scientific and Technical Research* 1: 85–97.** Argentina. **1991.**
17. Frery, A. and M. Marengoni: The Swendsen-Wang Dynamic in the Simulation of the Ising Model. ***Proc. Second Intern. Conf. on Use of Computers in Scientific and Technical Research 1:*** **98–107.** Argentina. **1991.**

18. J. Jensen, and L. J. Møller: Pseudolikehood for Exponential Family Models of Spatial Processes. Research Report; Department of Theoretical Statistics, Arhus University. **1989.**

19. A. Frery, Markovian Fields: Theory, Simulation and Parameter Estimation. Submitted to ***REBRAPE*. 1991.**

20. F. Martinelli, E. Olivieri, and E. Scoppola, J. Statist. Phys. 62, **117 (1991).**

21. F. Martinelli, E. Olivieri, and E. Scoppola, J. Statist. Phys. 62, **135 (1991).**

22. E. Reingold, J. Nievergelt, and N. Deo, ***Combinatorial Algorithms. Theory and Practice.*** Prentice-Hall. **1977.**

23. K. Huang, ***Statistical Mechanics.*** John Wiley & Sons, Inc. **1963.**

Resumo

Utilizamos o estimador de máxima pseudo-verossimilhança como regra de parada para dois algoritmos de simulação do modelo de Ising. Comparamos as dinâmicas de Swendsen-Wang e de Wolff, utilizando o número de iterações e o tempo de CPU necessários para atingir a convergência para diferentes valores do parâmetro de atratividade β.