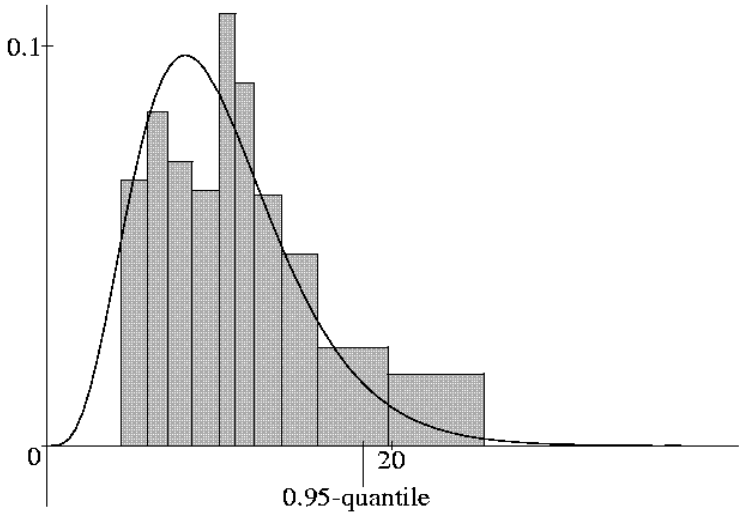
UNIVERSITÄT POTSDAM
Institut für Mathematik

Time-to-Coalescence for Interacting Particle Systems: Parallel versus Sequential Updating

Pierre-Yves Louis

Jean-Baptiste Rouquier



Mathematische Statistik und
Wahrscheinlichkeitstheorie

Universität Potsdam – Institut für Mathematik

Mathematische Statistik und Wahrscheinlichkeitstheorie

Time-to-Coalescence for Interacting Particle Systems: Parallel versus Sequential Updating

Pierre-Yves Louis,

Institut für Mathematik der Universität Potsdam
e-mail: louis@math.uni-potsdam.de

Jean-Baptiste Rouquier

Université de Lyon, LIP, ÉNS Lyon
46 allée d'Italie, 69364 Lyon, France
e-mail: jean-baptiste.rouquier@ens-lyon.org

Preprint 2009/03

April 2009

Impressum

Herausgeber: Mathematische Statistik und Wahrscheinlichkeitstheorie
am Institut für Mathematik

Adresse: Universität Potsdam
PF 60 15 53
14415 Potsdam
Telefon:
Fax: +49-331-977 1500
E-mail: +49-331-977 1578
neisse@math.uni-potsdam.de

ISSN 1613-3307

Time-to-Coalescence for Interacting Particle Systems: Parallel versus Sequential Updating

Pierre-Yves Louis,*
louis@math.uni-potsdam.de
Institut für Mathematik, Universität Potsdam,
Am neuen Palais,10 – SansSouci,
D-14 469 Potsdam, Germany

Jean-Baptiste Rouquier,
Université de Lyon, LIP, ENS Lyon
46 allée d'Italie, 69364 Lyon, France

April 2009

Abstract

Studying the influence of the updating scheme for MCMC algorithm on spatially extended models is a well known problem. For discrete-time interacting particle systems we study through simulations the effectiveness of a synchronous updating scheme versus the usual sequential one. We compare the speed of convergence of the associated Markov chains from the point of view of the time-to-coalescence arising in the coupling-from-the-past algorithm. Unlike the intuition, the synchronous updating scheme is not always the best one. The distribution of the time-to-coalescence for these spatially extended models is studied too.

Keywords Interacting particle systems, Probabilistic Cellular Automata, Coupling From The Past, Time to coalescence, Markov Chain Monte Carlo, Discrete time Glauber dynamics, Extreme value distribution

AMS MSC 2000 Classification
60K35 Interacting random processes; statistical mechanics type models; percolation theory
82C20 Dynamic lattice systems (kinetic Ising, etc.) and systems on graphs
68W10 Parallel algorithms

*Corresponding author

Contents

List of Figures

List of Tables

1 Introduction

The question of sampling from a given probability distribution arises for several purposes, for instance when one wants to find some coloring of a given graph or for dimer models as for statistics purposes. Since the cardinality of the space under consideration may be – still finite but – huge, direct computations are not the useful and general ones. The Markov-Chain-Monte-Carlo (MCMC) approach consists in realizing the considered probability distribution as asymptotic-in-time state from some Markov process. There are some canonical ways to do it: the *Metropolis-Hastings algorithm* and the *Gibbs sampler*. These algorithms are nowadays known from all practitioners and their study is related to some huge literature, from practical to theoretical aspects (see for instance [Gilks et al.(1996)Gilks, Richardson, and Spiegelhalter, Roberts and Rosenthal(2004), Diaconis and Saloff-Coste(1998)]).

The main limiting aspect of this approach is: the probability is only reached as an asymptotic state. From this point, studies develop two main directions. Either: one states results to control the error between the steady state we are interested in and the Markov chain realized for a time large enough (in some sense to be defined, so called *burn-in time*); or: through some modification of the algorithm, one finds a way to sample exactly from the steady state. Propp and Wilson proposed [Propp and Wilson(1996)] such a modification using the idea of the reversed sequence, see Subsection 2.1.

This modification is called *exact sampling* or *perfect simulation*. A new practical problem arising within this approach is the termination of this algorithm in some reasonable machine-time. In this article, we are interested in this *running-time* – also called *time-to-coalescence* for some reasons to be explained further – for interacting particle systems.

An *interacting particle system* (IPS) is a (Markov) stochastic process $(\sigma(t))_{t\in I}$ ($I = \mathbb{N}$ or $\mathbb{R}^+$) on some product space S^G where S is the so-called spin-space and G is a graph. The time-evolution of the stochastic process $(\sigma_k(t))_{t\in I}$ associated to a site $k \in G$ is defined according to the neighbor processes $(\sigma_{k'}(t))_{t\in I}$ ($k' \in V_k$) where $V_k \subset G$ is the (finite) set of neighbors of k in G. It means we are dealing with stochastic processes $(\sigma_k(t))_{t\in I}$ interacting on some spatial structure G. These stochastic processes were motivated as mathematical models for the dynamical evolution of the statistical mechanics equilibrium models. The physical systems "states" are then modeled as probability measure on S^G and studied as steady states of interacting particle systems. The heart of the study of such probabilistic objects lies in the relations between the limit increasing size finite graph G and an infinite graph (called *thermodynamics limit*). That is a motivation for considering state spaces with huge cardinality.

A whole theory has been developed for the case when the time-parameter is continuous ($I = \mathbb{R}^+$). See for instance [Liggett(1985)] for the traditional basics. The *probabilistic cellular automata* (PCA) are a class of such interacting particle systems with discrete time-parameter. Their specificity is the *synchronous updating*: between time n and time $n + 1$ all the values $(\sigma_k(n))$, $k \in G$, are modified according to some probabilistic rules. They contrast to the continuous-time interacting particle systems where during an infinitesimal amount of time δt the value of only one (possibly finitely many) site may be updated. In this case the values of the sites are updated one after the other, it is thus a (one site) *sequential updating*. A general theory for PCA like the one for continuous-time processes is still to be developed (see [Louis(2002)] and references therein). Nevertheless, cellular automata (with deterministic updating) and their probabilistic counterpart (PCA) are well known in the computer science and used from practitioners (see for instance the applications part (Part. 2) of the proceedings [Yacoubi et al.(2006)Yacoubi, Chopard, and Bandini]).

Since the time is discrete, PCA may be considered as a class of Markov chains. To exemplify their specificity inside the MCMC theory, let us remark that quite differently to the usual Hastings-Metropolis algorithm practice, for a given probability measure ν on S^G, there

is no general recipe to construct a PCA for which ν is stationary. In particular, there exists Gibbs measures on $S^{\mathbb{Z}^2}$ such that no PCA admits them as stationary reversible measures (see Theorem 4.2 in [Dawson(1974)]).

Since a few years, an important literature has been dedicated to the study of convergence rates to equilibrium for Markov chains, see [Cowles and Rosenthal(1998)]. The main question is how fast the burn-in time (also called *mixing time*) grows with the characteristic size of the model. See [David A. Levin(2008)]. Another related questions is the influence of the updating scheme for models with a spatial extension like IPS or the Gibbs sampler. See for instance [Dyer et al.(2006b)Dyer, Goldberg, Jerrum, and Martin, Randall and Tetali(2000)].

In this article, we compare the speed of convergence of IPS with synchronous or sequential updating by considering their time-to-coalescence. See [Murdoch and Rosenthal(2000)]. How is it distributed? How does the distribution depend on the updating? We want to quantify in that sense the effectiveness of sequential and parallel updating. Theoretical results are stated and simulations are developed to confirm them and analyze the situation further. Simulations are contradicting the intuition: synchronous updating is not always better than sequential updating. We quantify this fact.

This article is organized as follow. We first present – Section 2 – the definitions and the framework. In Section 4 we study the distribution of the time-to-coalescence and its evolution in connection with dynamics parameters. In Section 5 we compare the effectiveness of both updating schemes (parallel and sequential) from the point of view of the time-to-coalescence. This work is based on simulations, whose details are developed in Section 3.

2 Backgrounds

2.1 CFTP algorithm, running-time T_r

Let $(E, \mathcal{E})$ be a finite probability space and ν a probability measure on it. There is an irreducible aperiodic Markov transition kernel $P(\cdot,\cdot)$ on $E \times E$ such that ν is its unique *stationary measure i.e.* $\nu P = \nu$ where $\nu P(dy) = \int P(x, dy)\nu(dx)$. ν is called *steady state*. The CFTP algorithm uses P to draw a random sample according to the distribution ν.

Let $(X_n(\rho))_{n\geq 0}$ denote the Markov chain whose transition kernel is $P(\cdot,\cdot)$ and whose starting condition X_0 is drawn according to ρ. If ρ is a Dirac measure δ, *i.e.* if $\rho = \delta_z$ for some $z \in E$, then the starting condition is deterministic and we write the Markov chain $(X_n(z))_{n\geq 0}$. Recall that, on a finite state space, irreducible aperiodic Markov transition kernel are *ergodic*: for any starting distribution ρ, the chain converges in law towards the stationary measure ν.

We consider Markov chains as iterated random functions, as introduced in [Diaconis and Freedman(1999)]: this constructive approach is better for simulation purposes. Iterated random functions are a special case of discrete-time dynamical system. A *discrete-time dynamical system* (f, m) is defined by

- a probability space Θ,
- a sequence $(\theta_n)_{n\geq 1}$ of Θ-valued independent identically distributed random variables, with the probability measure m as distribution for θ_n
- a measurable application $f : \Theta \times E \to E$. We use the notation $f_\theta(x) := f(x, \theta)$.

The Markov chain $(X_n)_{n\geq 0}$ is then recursively constructed by independently choosing X_0 and $(\theta_n)_{n\geq 1}$, and defining for $n \geq 0$

$$X_{n+1} := f(\theta_{n+1}, X_n).$$

which is equivalent to

$$X_n := f_{\theta_n} \circ \ldots \circ f_{\theta_1}(X_0). \tag{1}$$

It is always possible (see for instance Chapter 2 in [Benaïm and El Karoui(2003)]) to represent a given transition kernel $P(\cdot,\cdot)$ as a discrete-time dynamical system (f,m).

The Propp-Wilson coupling from the past algorithm (CFTP) makes use of the following sequence $(Y_n)_{n\geq 0}$, also called *backward iteration*:

$$Y_n = f_{\theta_1} \circ \ldots \circ f_{\theta_n}(X_0).$$

Remark that for any $n \geq 0$, Y_n has the same distribution as X_n. Nevertheless the random sequences $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ are not equal in distribution. In particular $(Y_n)_{n\geq 0}$ is *not* a Markov chain. For a fixed $n \in \mathbb{N}$, the usual interpretation is to consider $(Y_k)_{0\leq k\leq n}$ as the chain starting at time $-n$ with initial distribution $\mathcal{L}(X_0)$ and evolving from time $-n+k-1$ $(k \geq 1)$ to time $-n+k$ according to the transition probability kernel $P(\cdot,\cdot)$. Let $(\xi^{-n}_{-n+k})_{k\geq 0}$ be the Markov chain defined by $\xi^{-n}_{-n+k} := f_{\theta_{n-k+1}}(\xi^{-n}_{-n+k-1})$. Note the equality in distribution on the path-space: $\mathcal{L}((X_k)_{k\geq 0}) = \mathcal{L}((\xi^{-n}_{-n+k})_{k\geq 0})$. The distribution at time 0 from ξ^{-n} is then the one from Y_n *i.e.* we get $\xi^{-n}_0 = Y_n$. Note that we use the transition kernel $P(\cdot,\cdot)$ and not the reversed kernel w.r.t. ν.

Let T_r be *the running-time* of the CFTP-algorithm

$$T_r := \inf\{n \ : \ Y_n \equiv \text{constant}\}$$

or equivalently $T_r := \inf\{k : \exists c \in E,\ \forall z \in E \text{ s.t. } \xi^{-k}_{-k} = z, \quad \xi^{-k}_0 = c\} \in \mathbb{N}$. Use the usual convention $\inf \emptyset = \infty$. It is a stopping-time with respect to the natural filtration.

We say that a random dynamical system (f,m) is *monotone* if E is a partially ordered set and f_θ is a monotone application from E to E (m-almost surely) *i.e.* $\forall x \preceq y,\ f_\theta(x) \preceq f_\theta(y)$. We say that the kernel $P(\cdot,\cdot)$ is *attractive* (or monotone) if $\forall x, y \in E, \quad P(x,\cdot) \preceq P(y,\cdot)$ where the order on the set of probability measures $\mathcal{M}_1(E)$ is given by: for any $\mu_1, \mu_2 \in \mathcal{M}_1(E)$, we say $\mu_1 \preceq \mu_2$ if $\forall f : E \to \mathbb{R}$ monotone, $\mu_1(f) \leq_{\mathbb{R}} \mu_2(f)$ where $\mu_1(f) = \int f(x)\mu_1(dx)$. If E is a totally ordered space, the existence of such a representation of the Markov dynamics thanks to a monotone random dynamical system is equivalent to the transition kernel being attractive. If E is only a partially ordered space, the existence of some representation of the Markov kernel through a monotone random dynamical system is a much stronger assumption (called complete monotonicity). We refer to [Dai Pra et al.(2006)Dai Pra, Louis, and Minelli] for a detailed analysis.

Let us consider E a partially ordered set and assume there is a representation of the considered Markov kernel as monotone random dynamical system. Let $E_{\mathbf{X}}$ be the set of maximal and minimal elements of E. It follows easily that

$$T_r = \inf\{k : \exists c \in E,\ \forall z \in E_{\mathbf{X}} \text{ s.t. } \xi^{-k}_{-k} = z, \quad \xi^{-k}_0 = c\}.$$

The monotonicity makes CFTP a very efficient algorithm since it reduces the number of paths to control, from $|E|$ to $|E_{\mathbf{X}}|$. Pay attention that the same realizations of the θ_i random variables are used for all the possible starting conditions.

Moreover, using the monotonicity and the irreducibility of the transition kernel $P(\cdot,\cdot)$, it is easy to prove that the stopping-time T_r is a.s. finite. With these assumptions, the distribution of $\xi^{-T_c}_0 = Y_{T_c}$ is the steady state ν of the transition kernel $P(\cdot,\cdot)$. Hence the name *exact sampling*. See [Häggström(2002)] for a more detailed presentation.

2.2 Convergence to equilibrium and time-to-coalescence T_c

Let us denote with T_c the *time-to-coalescence* or *coupling time*:

$$T_c := \inf\{n \ : \ \exists c \in E,\ \forall z \in E, \quad X_n(z) = c\}.$$

Since the same realizations of the θ_i random variables are used for all the possible starting conditions, we get a coupling of the $X_n(z)$ for all $z \in E$ on the same probability space. Moreover, the monotonicity of the coupling implies the *coalescence property*: if two paths meet, they will remain equal forever a.s. This implies that it is equivalent to check the coalescence before time 0 or at time 0. Using the time-translation invariance, it is then clear that $\mathcal{L}(T_c) = \mathcal{L}(T_r)$. Note that this is true only for the time, not the state: indeed, X_{T_c} does not necessarily have distribution ν.

Let us recall the control of the distance to equilibrium of the Markov chain thanks to the tail of the time-to-coalescence distribution. The *total variation distance* between probability measures μ_1, μ_2 on E is defined by:

$$||\mu_1 - \mu_2||_{\text{TV}} := \max_{A \in \mathcal{E}} |\mu_1(A) - \mu_2(A)|.$$

It is easy to state

$$||\mu_1 - \mu_2||_{\text{TV}} = \frac{1}{2} \sum_{x \in E} |\mu_1(x) - \mu_2(x)|.$$

The following result gives the control in total variation distance of the distance to equilibrium of the Markov kernel by the tail of the time-to-coalescence. For the proof, see Th. 1.2 Chap. 4 in [Brémaud(1999)] or for a more systematic study [Peres(2005)].

Theorem 1 (source?)**.** *Let $(X_n^1)_n$ and $(X_n^2)_n$ be two Markov chains with values on some finite state space E, associated to the same irreducible aperiodic Markov kernel $P(\cdot,\cdot)$, defined on the same probability space such that* coalescence *holds a.s. It holds:*

$$||\mathbb{P}_{X_n^1} - \mathbb{P}_{X_n^2}||_{TV} \leq \mathbb{P}(T_c > n).$$

In particular,

$$||\rho P^n - \nu||_{TV} \leq \mathbb{P}(T_c > n)$$

where ν is the stationary measure and ρ any starting distribution.

This result motivates our approach to quantitatively compare the speed of convergence of different transition kernel in term of their respective time-to-coalescence. It motivates as well the study of the distribution of the time-to-coalescence, since any inequality relating the tail and the moments, like the Markov inequality, will be useful (see examples developed in [Peres(2005)]).

Moreover, let us recall the following relation between the time-to-coalescence and the mixing-time:

Theorem 2 (see Th. 2 in [Randall and Tetali(2000)])**.**

$$\tau(\varepsilon) \leq 6T_c(1 + \log(\frac{1}{\varepsilon})),$$

where the mixing-time $\tau(\varepsilon)$ *is defined through*

$$\tau(\varepsilon) := \max_{\sigma \in E} \min\{n \ : \forall n' \geq n, \quad ||\delta_x P^{n'} - \nu||_{TV} \leq \varepsilon\}.$$

2.3 Read-once modification of the CFTP Algorithm

The first version of the CFTP algorithm require to store de random bits used (the values of $(\theta_n)_{n\geq 1}$), because each backward iteration must use the *same* $(\theta_n)_{n\geq 1}$ and not draw them again. While it is often sufficient to only store the seed of a pseudo random number generator, this fact was still considered as a drawback for the algorithm. Thus, Wilson [Wilson(2000)] suggested a modification that allows one to use a read-once source of randomness, i.e. not having to store the random bits. It is the running time of this version that we analyse here.

2.4 Sequential updating: Glauber dynamics

Before introducing the Markovian dynamics we are dealing with, let us define Gibbs measures. They are strongly related in both cases (sequential and parallel updating dynamics) to the time-asymptotics. Moreover, as will be exemplified through the simulations, when the phase transition phenomenon occurs, there is a change of behavior for the coupling time T_c.

2.4.1 Gibbs measures on lattice

From now on, we consider $E = S^G$ where S is a finite totally ordered space and $G = \mathbb{Z}^d$ or $G = \Lambda \Subset \mathbb{Z}^d$ (*finite* subset). A *configuration* σ is an element of E, its coordinates are written $(\sigma_k)_{k\in G}$. We note $-$ (resp. $+$) the minimum (resp. maximum) from S, and $-$ (resp. $+$) the configurations for which $\forall k \in G,\ \sigma_k = -$ (resp. $\sigma_k = +$).

A *Gibbs measure* is a natural measure related to some potential. A set of functions $(\varphi_A)_{A\subset G}$ is said to be a *potential* if for all A, $\varphi_A : S^G \to \mathbb{R}$ depends only on the sites in A, *i.e.* $\forall \sigma \in S^G, \varphi(\sigma) = \varphi(\sigma_A)$ where $\sigma_A = (\sigma_k)_{k\in A}$. We define the associated *Hamiltonian* $(H_\Lambda)_{\Lambda\Subset\mathbb{Z}^d}$ by

$$\forall \sigma \in S^G,\ H_\Lambda(\sigma) = \sum_{A\ :\ A\cap\Lambda\neq\emptyset} \varphi_A(\sigma).$$

We assume here for simplicity that φ is a *finite range* potential *i.e.*

$$\exists C > 0,\ \forall A \text{ s.t. } |A| \geq C,\ \varphi_A \equiv 0, \tag{2}$$

Example The well-known *Ising-potential* is given by

$$\varphi_A(\sigma) := \begin{cases} -\beta J_{i,j}\sigma_i\sigma_j & \text{when } A = \{i,j\}, \\ \beta h \sigma_i & \text{when } A = \{i\}, \\ 0 & \text{otherwise.} \end{cases}$$

The associated Ising Hamiltonian is then

$$H_\Lambda(\sigma) = -\beta \sum_{i\sim j,\ \{i,j\}\cap\Lambda\neq\emptyset} J_{i,j}\sigma_i\sigma_j - \beta h \sum_{k\in\Lambda}\sigma_k.$$

The model is parametrized through the "inverse-temperature" parameter $\beta > 0$, the magnetization $h \in \mathbb{R}$ and the (finite range) pair-interaction parameter $J_{i,j}$ between the sites i, j. Here only nearest neighbors can interact: $J_{i,j} > 0 \Rightarrow ||i - j||_1 = 1$.

For a fixed $\Lambda \Subset \mathbb{Z}^d$ (it is called a *finite volume*), we write Λ^c its boundary. The *finite-volume* Gibbs measure μ^τ_Λ, with *boundary condition* $\tau \in S^{\Lambda^c}$, is a probability measure on S^Λ. It is defined by:

$$\forall \sigma_\Lambda \in S^\Lambda,\ \mu^\tau_\Lambda(\sigma_\Lambda) := \frac{1}{\mathcal{Z}^\tau_\Lambda} e^{-H_\Lambda(\sigma_\Lambda\tau_{\Lambda^c})}$$

where the configuration $(\sigma_\Lambda\tau_{\Lambda^c})_k$ is σ_k if $k \in \Lambda$, τ_k otherwise. The important structural relation verified by this family of measures is:

$$\forall \Lambda' \subset \Lambda \Subset \mathbb{Z}^d, \forall \tau, \eta_{\Lambda\setminus\Lambda'}, \sigma_\Lambda, \quad \mu^\tau_\Lambda(\sigma_{\Lambda'}|\eta_{\Lambda\setminus\Lambda'}) = \mu^{\tau_{\Lambda^c}\eta_{\Lambda\setminus\Lambda'}}_{\Lambda'}(\sigma_{\Lambda'}).$$

Since S is finite and the potential is finite range, one proves that there is (*at least*) one measure μ on the infinite space $S^{\mathbb{Z}^d}$ such that

$$\forall \Lambda \Subset \mathbb{Z}^d,\ \mu(\ .\ |\ \tau_{\Lambda^c}) = \mu^{\tau_{\Lambda^c}}_\Lambda(\ .\ ).$$

Such a probability measure is called *infinite-volume Gibbs measure* with respect to the potential φ. When more than one probability measure fulfill these conditions, this is called *phase transition*. For more details and motivation on the theory of Gibbs measure, we refer to the standard statistical mechanics literature, for instance [Georgii(1988)].

2.4.2 Discrete-time Glauber dynamics

Let Λ be a finite subset of $\mathbb{Z}^d$ and for simplicity $S = \{-1, +1\}$. We now describe the *Metropolis-Hastings algorithm*, which gives a canonical way to associate an irreducible aperiodic Markov kernel to a given probability measure μ_Λ^τ such that this measure its only stationary measure.

Define the transition Markov kernel P_Λ^τ on S^Λ for $\sigma \neq \eta$ by

$$P_\Lambda^\tau(\eta, \sigma) := Q_\Lambda(\eta_\Lambda \tau_{\Lambda^c}, \sigma) R(\eta_\Lambda \tau_{\Lambda^c}, \sigma), \tag{3}$$

where

- Q_Λ is any Markov kernel on S^Λ such that $Q(\eta, \sigma) > 0$ implies $Q(\sigma, \eta) > 0$;
- R is given for $\sigma \neq \eta$ by

$$R(\eta, \sigma) = \begin{cases} h(\frac{\mu(\sigma)Q(\sigma,\eta)}{\mu(\eta)Q(\eta,\sigma)}) \text{ if } Q(\eta, \sigma) \neq 0 \\ 0 \text{ otherwise,} \end{cases}$$

- $h :]0, +\infty[\to]0, 1[$ is such that $h(u) = uh(\frac{1}{u})$.

Finally, normalize $P_\Lambda^\tau(\eta, \eta)$ so that P_Λ^τ is a Markov kernel on S^Λ. Q is called *proposal* probability kernel, R is the *acceptance/rejection* probability kernel.

We assume moreover that

- Q_Λ is any symmetric Markov kernel such that $Q_\Lambda(\eta, \sigma) > 0$ if there is at most one site $k \in \Lambda$ such that $\eta_k \neq \sigma_k$;
- R is given by

$$R(\eta, \eta^k) = \frac{e^{-H_\Lambda^\tau(\eta^k)}}{e^{-H_\Lambda^\tau(\eta^k)} + e^{-H_\Lambda^\tau(\eta)}} \tag{4}$$

  where η^k is the configuration η with site k flipped: $\forall j \neq k$, $(\eta^k)_j := \eta_j$ and $(\eta^k)_k := -\eta_k$.

Here we made the choice $h(u) := \frac{u}{1+u}$ (sometimes called Barker-rate). Another common choice is the Metropolis-rate: $h(u) = \min(1, u)$.

The *detailed balance condition*

$$P_\Lambda^\tau(\eta, \sigma)\mu_\Lambda^\tau(\eta) = P_\Lambda^\tau(\sigma, \eta)\mu_\Lambda^\tau(\sigma)$$

holds. It means that μ_Λ^τ is a reversible measure w.r.t. P_Λ^τ so it is a stationary one.

If we make the particular choice

$$Q_\Lambda(\eta, \eta^k) := \frac{1}{|\Lambda|} \tag{5}$$

we get a *random choice sequential updating (spin flip) dynamics*. It is the discrete time equivalent of the Glauber dynamics. A Glauber dynamics is a continuous-time Markov process on S^G where jumps occur at rate (4) between configurations differing in one site (see [Liggett(1985), Martinelli(1999)] for more details). As noticed in [Dyer et al.(2006a)Dyer, Goldberg, and Jerrum], the term Glauber dynamics does not have a precise agreed meaning. It means in general: local (see (2)) single sites updates performed in random sequence. Remark that unlike its discrete-time version, the continuous-time usual form can not be generalized on $S^{\mathbb{Z}^d}$ to a (fully) parallel dynamics. This motivates PCA processes as models for the parallel updating.

In the particular case of the Ising model with $h = 0$ and nearest-neighbor interaction (*i.e.* $J_{i,j} = 1$ if $||i - j|| = 1$, 0 otherwise), the acceptance/rejection probability becomes

$$R(\eta, \eta^k) = \frac{e^{\beta \sum_{j \sim k} \eta_k \eta_j}}{2 \cosh(\beta \sum_{j \sim k} \eta_j)}. \tag{6}$$

Remark Since $S = \{-1, +1\}$,

$$R(\eta, \eta^k) = \mu_{\{k\}}^{\eta_{\{k\}^c}}(-\eta_k) = \mu_\Lambda(-\eta_k|\eta_{\{k\}^c}), \tag{7}$$

where μ denotes the Gibbs measure on S^Λ.

Since the one-site conditional distribution is used, it is the spin-flip version of the usual *Gibbs sampler* also called *heat-bath* dynamics. We refer to [Sokal(1989)] for more details and to [Frigessi et al.(1993)Frigessi, Hwang, di Stefano, and Sheu] for a comparison of the velocity of convergence to equilibrium in term of the second-largest eigenvalue in absolute value between the Gibbs-Sampler and the Hastings-Metropolis algorithm.

2.4.3 Deterministic sequential updating scheme

For the Gibbs sampler, it is a natural question (see [Fishman(1996)]) to study the role played by the updating scheme: random or systematic scans (also called systematic sweep). In [Dyer et al.(2006a)Dyer, Go the authors study Glauber dynamics for graph coloring and compare the effectiveness of random and systematic scan in term of mixing-time of the processes. In [Diaconis and Ram(2000)] the authors give "the first analysis of a systematic scan version of the Metropolis algorithm" (sic). According to these works, it seem that random update and systematic scan are comparable in term of the velocity of convergence. But proof is still lacking except in special cases like some considered therein. Our article is to be considered as a first step in extending these questions towards a comparison between single updating and synchronous updating.

2.5 Parallel updating: PCA

We now want to apply the updating rule $R(\eta, \eta^k)$ synchronously. We use the Markov transition kernel

$$\wp_\Lambda^\tau(\eta, \sigma) = \prod_{k\in\Lambda} p_k(\sigma_k|\eta)$$

where the (local) updating rule is given by

$$p_k(\sigma_k|\eta) = \mu_{\{k\}}^{\eta_{\{k\}^c}}(\sigma_k). \tag{8}$$

The stationary measure is not the finite-volume Gibbs measure any more but differ with some additional boundary term. We refer to the detailed analysis made in [Dai Pra et al.(2002)Dai Pra, Louis, and Rœll (Proposition 3.1) for the precise expression of the stationary measure ν_Λ^τ of $\wp_\Lambda^\tau$. The updating rule (8) may also be written

$$p_k(\sigma_k|\eta) = \frac{1}{2}\big(1 + \sigma_k \tanh(\beta \sum_{j\sim k} \eta_j)\big). \tag{9}$$

PCA are cellular automata where the local updating rule is probabilistic. There is a huge literature about cellular automata, from computer science to applications: we refer to the literature referred in [Louis(2002)]. The family we deal with here satisfies the property $\forall k \in \mathbb{Z}^d, \forall \eta, \forall s \in S,\ p_k(s|\eta) > 0$. Such PCA are called *purely stochastic PCA*.

Note that here the value $\sigma_k(n)$ is not involved in the updating rule $p_k(\cdot|\sigma(n))$ used to determine the value $\sigma_k(n+1)$. If it were the case, the sequential time counterpart would still be the Hastings-Metropolis one (6) but not coinciding with the Gibbs sampler (7) anymore. We do not consider this case here.

2.6 Definition of the coupling

Since we want to run the CFTP algorithm for PCA dynamics as well as Glauber dynamics, we are looking for a representation as a monotone dynamical system which has a product structure too. For an IPS with local updating, PCA or Glauber dynamics, the monotonicity of the transition kernel is equivalent to require

$$\forall \sigma \preceq \eta \in E,\ \forall k \in \Lambda,\ p_k(\cdot|\sigma) \preceq p_k(\cdot|\eta) \tag{10}$$

in the sense of the ordering between measures on S.

We recall the following result (see Th. 2.4 in [Louis(2005)] for details) whose proof is given here for algorithmic purposes.

Proposition 3. *Let P be the transition kernel on S^Λ ($\Lambda \Subset \mathbb{Z}^d$ or $= \mathbb{Z}^d$) of a discrete-time Glauber dynamics or PCA dynamics. Assume the spin space S is totally ordered and P is monotone. Then there is a representation as monotone random dynamical system.*

Proof. Let $(\sigma_k(n))_{k\in\mathbb{Z}^d}$ be a PCA on $E = S^\Lambda$ with transition probability $\wp = \prod p_k$. We define

$$\Theta = \times_{k\in\Lambda}[0,1], \quad m = \otimes_{k\in\Lambda}\lambda_{|_{[0,1]}}$$

where $\lambda_{|_{[0,1]}}$ is the uniform distribution on $[0,1]$. It means we consider a sequence of independent identically distributed random variables $(\theta(n))_{n\geq 0}$ such that

$$\forall n \geq 0, \quad \theta(n) = \otimes_{k\in\Lambda}\theta_k(n),$$

where the random variables $(\theta_k(n))_{k\in\Lambda}$ are independent, uniformly distributed on $[0,1]$. Let $f : \Theta \times E \to E$ be defined through

$$f(\theta,\sigma) = \otimes_{k\Lambda}\ g_k(\theta_k,\sigma)$$

where $\theta = \times_{k\in\Lambda}\ \theta_k \in \Theta$, and $\forall k \in \Lambda,\ g_k : [0,1] \times E \to E$ be local updating functions to be defined. Let $F_k(\cdot,\sigma)$ ($k \in \Lambda$, $\sigma \in E$) be the probability distribution function of $p_k(\cdot|\sigma)$

$$F_k(s,\sigma) = \sum_{s'\preceq s} p_k(s'|\sigma),\ (s \in S).$$

Let us define the local updating functions g_k ($k \in \Lambda$) through

$$g_k(u,\sigma) = F_k(\cdot,\sigma)^{(-1)}(u)$$

where $u \in [0,1]$, $\sigma \in E$ and $(F_k)^{(-1)}$ denotes the Lévy probability transform (generalized inverse probability transform) of the F_k distribution function

$$(F_k)^{(-1)}(u) = \inf_{\preceq}\ \{s \in S : F_k(s) \geq u\}, \quad u \in]0,1[,$$

where $\preceq$ denotes the ordering on S. It is easy to check that (Θ, f, m) is a representation of $\wp$ as random dynamical system: $\forall n \geq 0$, $\forall k \in \Lambda$:

$$\sigma_k(n+1) = g(\theta_k(n+1), \sigma(n)).$$

Let $(\sigma_k(n))_{k\in\mathbb{Z}^d}$ be a Glauber dynamics on $E = S^\Lambda$ with transition kernel P as defined in (3) with (4) and (5). In that case, let Θ, m, $(\theta(n))_n$ be defined as previously and $\tilde{f} : \Theta \times \Lambda \times E \to E$ be defined through

$$\tilde{f}(\theta,\mathcal{K},\sigma) = g_{\mathcal{K}}(\theta_{\mathcal{K}},\sigma) \otimes_{k\in\Lambda,\ k\neq\mathcal{K}} \mathrm{Id}_k(\sigma)$$

where $\mathcal{K} \in \Lambda$ and $\mathrm{Id}_k(\sigma) = \sigma_k$. Define $\mathrm{unif}(\Lambda)$ be the uniform probability measure on Λ. It is easy to check that $(\Theta \times \Lambda, \tilde{f}, m \times \mathrm{unif}(\Lambda))$ is a representation of the Glauber dynamics P as random dynamical system.

Note that the local updating functions g_k are applied either synchronously for the PCA or sequentially with a uniform random choice for the Glauber dynamics. Further note that the monotonicity of $\wp$ (resp. P) means, $\forall k \in \Lambda$,

$$\forall \sigma \preceq \eta \in E,\ F_k(\cdot, \sigma) \geq F_k(\cdot, \eta)$$

or

$$\forall \sigma \preceq \eta \in E,\ \forall u \in [0,1],\ g_k(u, \sigma) \preceq g_k(u, \eta).$$

So the functions $f(\theta, \cdot)$ (resp. $\tilde{f}(\theta, \mathcal{K}, \cdot)$) are monotone (on S^Λ with the ordering $\preceq$).

□ □

We are here more precisely interested in the local updating rule on $S = \{-1, +1\}$ given by (9) which is the same as (6). It is easy to check that this rule is monotone: (10) holds. As noted in Subsection 2.1, the monotonicity of the representation allows to run CFTP when controlling only two copies of the same dynamics. We consider here the two copies $(\sigma^-(n))_{n\geq 0}$ (resp. $(\sigma^+(n))_{n\geq 0}$) such that $\forall k \in \Lambda,\ \sigma_k^-(0) = -1$ (resp. $\forall k \in \Lambda,\ \sigma_k^+(0) = +1$). Since for the coupling we are using the same random variables $\theta_k(n)$ for all the copies of the dynamics, the transition probability q_k on $(S^\Lambda \times S^\Lambda) \times (S \times S)$ may be written too

$$\left\{ \begin{array}{l|ll} q_k\big((-1,-1) & (\sigma^-(n), \sigma^+(n))\big) & = 1 - p_k(+1 \mid \sigma^+(n)) \\ q_k\big((-1,+1) & (\sigma^-(n), \sigma^+(n))\big) & = p_k(+1 \mid \sigma^+(n)) - p_k(+1 \mid \sigma^-(n)) \\ q_k\big((+1,-1) & (\sigma^-(n), \sigma^+(n))\big) & = 0 \\ q_k\big((+1,+1) & (\sigma^-(n), \sigma^+(n))\big) & = p_k(+1 \mid \sigma^-(n)). \end{array} \right. \tag{11}$$

Note that, for any k, $\sigma^-_{V_k}(n) \equiv \sigma^+_{V_k}(n)$ implies a.s. $\sigma_k^-(n+1) = \sigma_k^+(n+1)$ where $V_k = \{j : j \sim k\}$. So this coupling has the *coalescence property*: $\sigma^-(n) = \sigma^+(n)$ implies a.s. $\sigma^-(n+1) = \sigma^+(n+1)$.

For the coupling of the PCA dynamics, this updating rule q_k is applied to all k between time n and $n+1$. It means the coupled process is a PCA too but on the product space $S^\Lambda \times S^\Lambda$.

For the coupling of sequential dynamics, the updating rule q_k is applied to one randomly uniformly chosen site k between time n and $n+1$. Note this site is the same for both the coupled processes. So each coupled process is using first (5) and then (6). It means we are coupling in an order preserving way the Glauber dynamics introduced in Section 2.4.2.

3 Simulations

Simulating those cellular automata is straightforward: simply have an array of cells, each cell storing a spin (represented by a boolean). For parallel updating, we need two copies of the array: to update all cells "at once", we read the states in one copy and write the result of the transition rule in the other. The running time is measured as the exact number of steps needed to reach coalescence, not the wall-clock time. We considered only periodic boundary condition. We simulated arrays of 1000×1000 cells. For some figures (if the time-to-coalescence was too long for the choice of the other parameters, in particular β), we used only 200×200 cells.

We needed one month of computation on a cluster of 16 Pentium IV Xeon biprocessor 2.6Ghz. The source code for the program used is available at cimula.sf.net.

Since we are interested in the distribution of T_c and not in the realization of an unbiased steady state, we do not run the CFTP algorithm but simulate the dynamics in a "forward

iteration" manner (like (1)) and find realizations of the coupling-time T_c. I.e., the algorithm compute a biased sample (which we discard immediately), but its running time exactly follows the distribution of the time-to-coalescence.

Since the parallel dynamics performs $|\Lambda|$ updates in one step time, we multiply the coupling time of the sequential dynamics with $|\Lambda|$ to compensate.

4 Distribution of the time-to-coalescence

We now analyze the results of simulations of the time-to-coalescence for the local updating rule (9) on $S = \{-1, +1\}$ applied sequentially (Glauber dynamics) or synchronously (PCA).

The distribution of the time-to-coalescence is important for the time needed to run the CFTP-algorithm. As was explained in Section 2.2, the tail of the distribution gives a control on the speed of convergence of the Markov chain (Mixing-time).

Let us define T_k the time-to-coalescence at site $k \in \Lambda$:

$$T_k = \inf\{n \ : \forall m \geq n, \ \ \sigma_k^-(m) = \sigma_k^+(m)\}. \tag{12}$$

Remark $\sigma_k^-(n) = \sigma_k^+(n)$ does *not* imply that this equality holds for all the successive times. The time-to-coalescence for the entire volume Λ is

$$T_c = \max_{k \in \Lambda} T_k. \tag{13}$$

Let T_c^{par} denote the coalescence time defined in (13) with respect to the PCA dynamics defined through (9). As remarked previously, since one site is updated for one time-unit in the sequential updating case versus $|\Lambda|$ in the parallel updating scheme, we renormalize this quantity for the time-to-coalescence T_c^{seq}:

$$T_c^{\text{seq}} = |\Lambda| \max_{k \in \Lambda} T_k. \tag{14}$$

4.1 Case $\beta = 0$: theoretical result

The case $\beta = 0$ (in equation (9)) means spatial and temporal independence between the sites. Take $V_k = \{k\}$. According to the coupling (11), the definition (12) becomes

$$T_k = \inf\{n \ : \ \sigma_k^-(n) = \sigma_k^+(n)\}.$$

In that case, one can identify the infinitely-many interacting sites asymptotics of the time-to-coalescence's distribution.

First, recall that a real valued random variable G is said to have a *Gumbel*-distribution if its distribution function is

$$x \mapsto e^{-e^{-x}}.$$

Such a distribution plays a central role in the theory of extreme values. Recall the following Theorem:

Theorem 4 (Th. 3.2.3 in [Embrechts et al.(2008)Embrechts, Klüppelberg, and Mikosch]). *Let $(X_n)_{n\geq 1}$ be a sequence of independent identically distributed random variables. Let $M_n := \max_{1\leq k\leq n} X_k$. Assume there exist a sequence $(a_n)_{n\geq 1}$ of positive real numbers, a sequence $(b_n)_{n\geq 1}$ of real numbers and a random variable G on $\mathbb{R}$ such that*

$$\frac{M_n - b_n}{a_n} \xrightarrow[n\to\infty]{\mathcal{L}} G. \tag{15}$$

Then, either G has a Dirac mass as distribution, or its distribution is one of the three following extreme values *distributions corresponding respectively to the cumulative distribution functions*

$$F_1(x) = e^{-e^{-x}}, \quad F_2(x) = e^{-\frac{1}{x^a}}\ \mathbb{1}_{\mathbb{R}^+}, \quad F_3(x) = \mathbb{1}_{\mathbb{R}^+} + e^{-\frac{1}{(-x)^a}}\ \mathbb{1}_{\mathbb{R}_*^-}.$$

Note that the assumption (15) is not true for any distribution. For instance, the Poisson distribution does not admits a rescaling such that the maximum would converge towards a non trivial distribution. We can now state

Proposition 5. *When $\beta = 0$, $T_c^{par} \log 2 - \log|\Lambda|$ converges in distribution towards a Gumbel distributed random variable G as $|\Lambda|$ goes to $+\infty$.*

Proof. The random variables T_k are independent and for any $k \in \Lambda$ T_k is the distribution of the first success in a Bernoulli trials with probability $\frac{1}{2}$ which is a geometric distribution with parameter $\frac{1}{2}$.

Let $x \in \mathbb{R}$, $k_0 \in \Lambda$. It follows:

$$\begin{aligned}\mathbb{P}(T_c^{\text{par}} \log 2 - \log|\Lambda| \le x) &= \left(\mathbb{P}(T_{k_0} \le \frac{x + \log|\Lambda|}{\log 2})\right)^{|\Lambda|} \\ &= \left(1 - \frac{1}{2^{\lfloor \frac{x+\log|\Lambda|}{\log 2} \rfloor}}\right)^{|\Lambda|} \sim_{|\Lambda|\to\infty} \left(1 - \frac{e^{-x}}{|\Lambda|}\right)^{|\Lambda|}\end{aligned}$$

where $\lfloor y \rfloor$ denotes the integer part of the real y. So the probability distribution function of $(\log 2)\ T_c^{\text{par}} - \log|\Lambda|$ converges with $|\Lambda|$ going to ∞ towards F_1. □ □

Proposition 6. *When $\beta = 0$, $T_c^{seq} \log 2 - \log|\Lambda|$ converges in distribution towards a Gumbel distributed random variable G as $|\Lambda|$ goes to $+\infty$.*

Proof. In the sequential updating case, the waiting time until coalescence will be distributed like a geometric distribution with parameter $\frac{1}{2|\Lambda|}$. □ □

4.2 Case $0 < \beta < \beta_c$: simulation results

Assume now $\beta > 0$. The sites are interacting. Since the random variables $(T_k)_{k\in\Lambda}$ are not independent anymore, it is difficult to prove the validity of an analogue result as Proposition 5. Some generalizations of the Theorem 4 are known when the sequence of random variables is stationary and a weak mixing condition holds (see for instance Th.2.1. in [Bovier(2005)]). So we may expect the Proposition 5 to hold for small values of β. Nevertheless, proof is still to be written. A first step in this direction are the results of the simulations. We state that the time-to-coalescence distribution is still "near" to a Gumbel distribution for $|\Lambda|$ large, whose parameters are estimated. Thus we can statistically analyze how the parameter β of the dynamics influences this distribution. As expected, there is a critical value β_c of β such that approaching it, the time-to-coalescence explodes. This will be analyzed in the following Section 6.1.

Let us first define the whole family of the Gumbel distributions. It is a family of probability distributions parametrized through $q \in \mathbb{R}$ and $s \in \mathbb{R}^+$. The probability density function is given by

$$f(x) = e^{-\frac{x-q}{s}}\ e^{-e^{-\frac{x-q}{s}}}\ \mathbb{1}_{\mathbb{R}^+}(x). \tag{16}$$

We refer to the Chapter 21 in [Johnson et al.(1994)Johnson, Kotz, and Balakrishnan] for details. If G is a random variable with this distribution, let us only recall

$$\begin{aligned}\mathbb{E}(G) &= q + \gamma s \text{ with } \gamma \text{ the Euler-constant } (\gamma \simeq 0.577), \\ \mathbb{V}\text{ar}(G) &= \frac{\pi^2}{6} s^2.\end{aligned} \tag{17}$$

As presented in Section 3 we consider here the nearest neighbor Ising potential φ with $h = 0$. We run simulations on one hand for the associated dynamics with parallel updating scheme (PCA) and on the other hand for the dynamics with a systematic sequential updating scheme (Glauber dynamics). For the PCA is means that $k \notin V_k$. It is indeed a very precise model but it remains a PCA dynamics and since much more detailed results are known about it, the analyze will gain in accuracy. For instance, it was proven in [Dai Pra et al.(2002)Dai Pra, Louis, and Rœlly] that $\beta_c = \log(\frac{1+\sqrt{2}}{2}) \simeq 0.44$.

Let us synthesize results we stated from these simulations. We only present here the parallel updating case, results for the sequential updating are similar. A comparative analysis is in Section 5.

Claim 7. *With $|\Lambda|$ large, $0 < \beta < \beta_c$,*

1. *the distribution of T_c is approximated by a generalized extreme value distribution. For small β, it is a Gumbel distribution. For β near to β_c^-, it is a Fréchet distribution;*

2. *the parameter β in this interval does* not *modify the nature of the distribution, only its associated parameters.*

The coefficients $\hat{q}, \hat{s}$ are estimated through the usual estimators for expectation and variance and solving the Equations (17) (so called moments method). The maximum likelihood estimation gives similar results. The Figures 1 shows the histogram of T_c with the density of the approximated Gumbel distribution. The Figure 2 shows the qq-plot graph, which means the quantile of the empirical distribution of T_c versus the quantile of the approximated Gumbel distribution. The linear dependence confirms the hypothesis of the Claim 7. The graphical representation of the sample's empirical probability distribution function leads to the same result.

Statistical tests were made to justify these graphical results. The chi square test, the Anderson-Darling test and the Kolmogorov-Smirnov test give the same conclusion. For small postive values of β, say $0 < \beta < 0.2$, these tests advice to reject the null-hypothesis, that for Λ large, the distribution of T_c is good fitted with an estimated Gumbel distribution (for the usual significance levels $\alpha = 0.01$ and $\alpha = 0.05$). For bigger values of β, $0.2 < \beta < 0.44 < \beta_c$, all these tests indicate not to reject the null-hypothesis. For $0.44 < \beta < \beta_c$, the null-hypothesis is rejected. Nevertheless, for these values of β, statistical test indicate not to reject the null-hypothesis that a Fréchet distribution is a good approximation. The results are presented in Table 1.

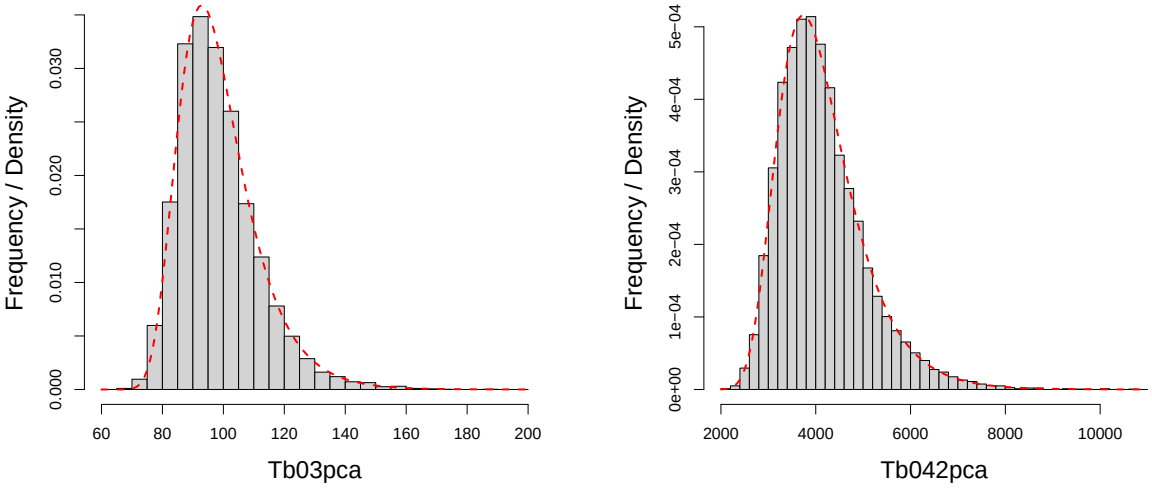
The distribution of T_c we are looking for when $|\Lambda|$ is large (but finite in these simulations) is discrete on the contrary to the Gumbel distribution. Nevertheless, we considered the Kolmogorov-Smirnov statistical test (D). For the size of the samples we have, it is considered as powerful as the chi squared test (see section 6 p. 19 in [Ricci(2005)]). Results are going in fact the same way as the chi-square and the Anderson-Darling tests.

For $0 < \beta < 0.2$, the rejection of the null-hypothesis comes from the discreteness of T_c's distribution. Several values arise with a high frequency in the samples. On the contrary, for β larger, the range of the sample's values is bigger, so values are less repeated and the discreteness becomes somehow negligeable.

5 Influence of the updating rule

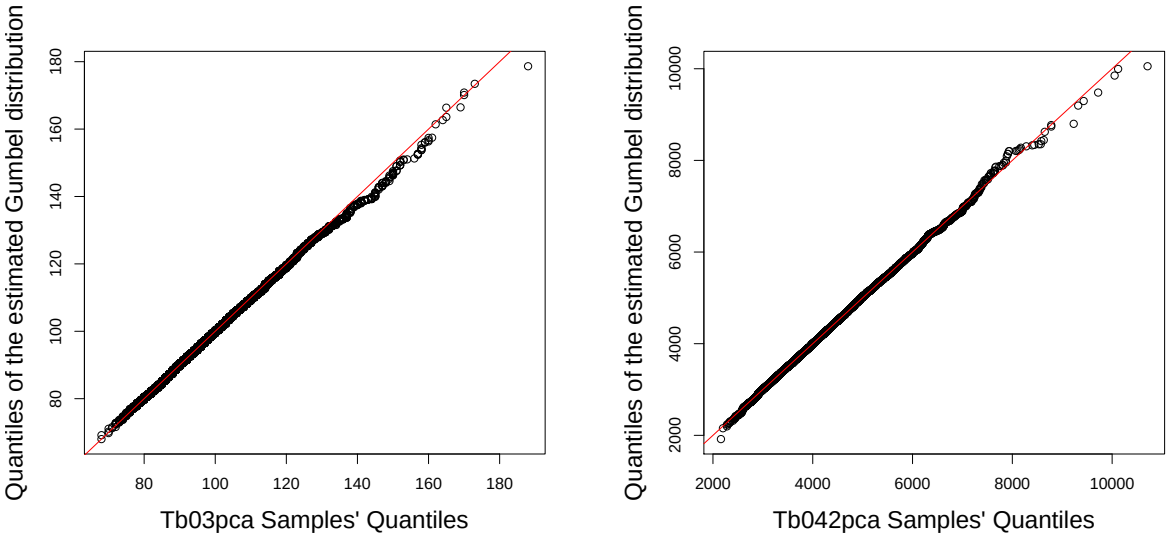
5.1 Influence of the updating rule on the distribution of T_c

Developing for T_c^{seq} the same statistical analysis as before, we find the results summarized in the Table 2.



(a) $\beta = 0.3$ (b) $\beta = 0.42$

Figure 1: Histogram of T_c and the density of the approximated Gumbel distribution, sample of size 10,000.



(a) $\beta = 0.3$ (b) $\beta = 0.42$

Figure 2: Histogram of T_c and the density of the approximated Gumbel distribution, sample of size 10,000.

The null-hypothesis is that the sample is Gumbel distributed, with parameters estimated. With a level of significance from $\alpha = 0.01$, we reject the null-hypothesis for $0 < \beta < 0.2$, not reject for $0.25 < \beta < 0.435$ and reject for $0.438 < \beta < \beta_c$. For $0.438 < \beta < \beta_c$ others test with "adequation to a Fréchet distribution" as null-hypothesis suggest not to reject. So we claim:

Claim 8. *For $\beta < \beta_c$, the updating scheme (parallel/sequential) does* not *modify the nature of the distribution, only its associated parameters.*

In the Figures 4, the estimated Gumbel parameters $\hat{q}$, $\hat{s}$ are compared

β	0	0.05	0.1	0.15	0.2	0.25
$\hat{s}$	0.00	0.67	1.05	1.64	2.74	4.95
$\hat{q}$	1.00	6.96	11.40	17.90	28.64	48.52
$100D$	9.64	27.82	17.08	11.56	7.43	4.37
p.value	$< 1e-04$	$< 1e-04$	$< 1e-04$	$< 1e-04$	$< 1e-04$	0.0436

β	0.3	0.35	0.4	0.41	0.42	0.425	0.43	0.435	0.438	0.439	0.440	0.4405	β_c^-
$\hat{s}$	11	28	166	306	697	1204	2844	10988	45533	836450	191275	332480	381848
$\hat{q}$	93	227	1078	1814	3687	6042	11793	32382	86818	144605	281736	400076	492848
$100D$	2.6	2.7	2.6	1.8	1.9	1.8	2.02	2.57	5.05	4.54	4.74	6.79	5.36
p.value	0.48	0.44	0.5	0.9	0.8	0.9	0.81	0.52	0.01218	0.0323	0.0225	0.00019	0.00644

Table 1: Estimated coefficients $\hat{q}, \hat{s}$ of a Gumbel distribution fitting to the sample of T_c^{par}, influence of the parameter β. Kolmogorov-Smirnov Statistics D. Samples of size 1000. Lattice's size: $200 * 200$.

β	0	0.05	0.1	0.15	0.2	0.25
$\hat{s}$	1.08	1.29	1.62	2.39	3.26	5.39
$\hat{q}$	11.09	13.46	17.05	22.69	32.43	50.69
$100D$	16.91	15.18	11.76	7.73	7.00	5.01
p.value	$< 1e-04$	$< 1e-04$	$< 1e-04$	$< 1e-04$	0.00011	0.0132

β	0.3	0.35	0.4	0.41	0.42	0.425	0.43	0.435	0.438	0.439	0.440	0.4405
$\hat{s}$	11	27	158	310	699	1317	2851	9887	40761	75756	179204	276696
$\hat{q}$	91	211	959	1571	3181	5138	9690	25562	61423	97189	166655	246785
$100D$	2.45	2.39	1.95	2.54	2.04	2.39	2.08	2.89	7.47	7.62	11.01	11.52
p.value	0.58	0.62	0.84	0.54	0.80	0.62	0.78	0.37	$< 1e-04$	$< 1e-04$	$< 1e-04$	$< 1e-04$

Table 2: Estimated coefficients $\hat{q}, \hat{s}$ of a Gumbel distribution fitting to the sample of T_c^{seq}, influence of the parameter β. Kolmogorov-Smirnov Statistics D. Samples of size 1000. Lattice's size: $200 * 200$.

5.2 Parallel versus sequential

For $\beta < \beta_c$, drawing the mean of $T_c^{\text{par}}/Tseqc$ sampled values in function of $\beta_c - \beta$ leads to the Figure 5. For a region of β away from 0 and β_c, note it is claiming the existence of a power law relation

$$\exists \gamma > 0,\ \overline{T_c} = \frac{1}{(\beta_c - \beta)^\gamma}.$$

Moreover, it is underlining the fact that for $\beta < 0.25$, the relation $\overline{T_c^{\text{par}}} < \overline{T_c^{\text{seq}}}$ holds between the mean time-to-coalescence. For $\beta > 0.3$, $\overline{T_c^{\text{par}}} > \overline{T_c^{\text{seq}}}$ holds. See the Figure 6. It means, for small values of β (situation not so far from the independence case between the site), the PCA dynamics is quicker than the sequential one, when considering the CFTP algorithm. For larger values of $\beta < \beta_c$, unlike the intuition, the PCA dynamics is slower.

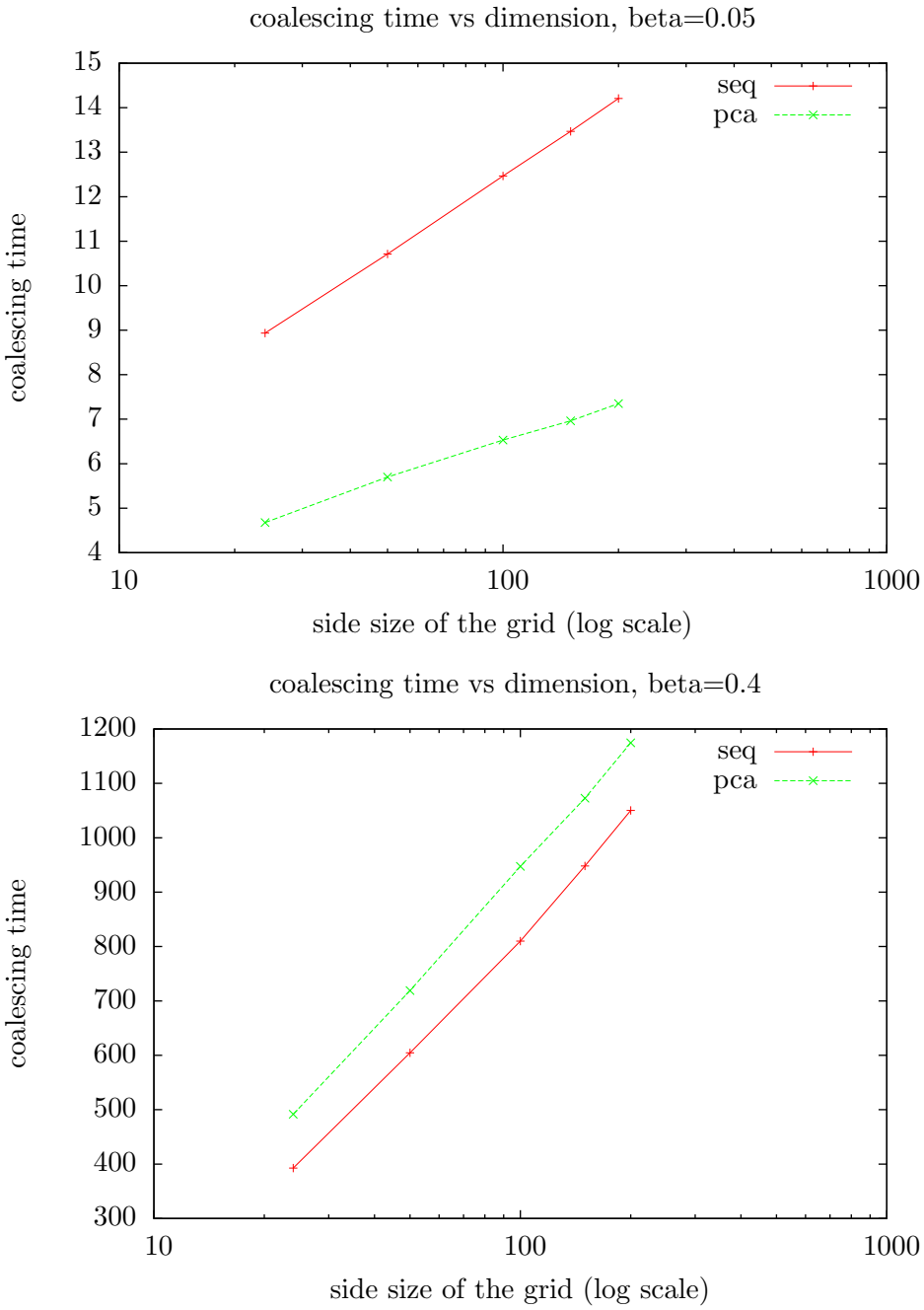


Figure 3: Influence of the number of cells on the time-to-coalescence. Note the different scales.

6 Conclusion

6.1 Phase transition

From the reference [Dai Pra et al.(2002)Dai Pra, Louis, and Rœlly], we know that the critical parameter β_c delimits two region for the set of the Gibbs measures: uniqueness of the infinite-volume Gibbs measure and existence of several Gibbs measure (so called *phase transition regime*) for the potential φ naturally associated to the PCA dynamics. The critical value $\beta_c = \log(\frac{1+\sqrt{2}}{2})$ is the same as the one for the Ising potential. We know precisely how the PCA dynamics with $h = 0$ and nearest-neighbor Ising potential behaves – in finite volume as well as infinite volume. It is based on a precise analysis of the relations between the static – study of the Gibbs measures

associated to a given potential – and the dynamics – ergodicity of the dynamics, characterization from the stationary (respectively reversible) probability measures. Let us briefly recall, for the infinite volume PCA dynamics (*i.e.* on $S^{\mathbb{Z}^d}$ or the infinitely many interacting sites asymptotics) the critical value β_c characterize two dynamical regimes too:

- for $0 < \beta < \beta_c$, the PCA dynamics is ergodic and converges (exponentially fast) towards a unique reversible state, which is also the unique (infinite volume) Gibbs measure associated to the Ising potential;
- for $\beta > \beta_c$ there are several reversible probability measures, thus the dynamics is *not ergodic*.

Since the PCA dynamics is purely probabilistic, the finite volume PCA-dynamics is always ergodic and converges towards a unique stationary measure. When considering periodic boundary conditions, this measure is also the unique finite-volume Gibbs measure with respect to the chosen Ising potential. Nevertheless, the critical value β_c also plays a role there: when approaching β_c for below, we state with the simulations that the running-time from the CFTP algorithm explodes. When considering the systematic sequential updating dynamics, similar phenomenon holds too. See Figure 7.

The critical parameter β_c related to two different behaviors of the infinitely many interacting sites asymptotic dynamics plays a very important role for the finitely many interacting dynamics too. It bounds two regions for the distribution of the CFTP algorithm's running-time: in the non phase transition regime $\beta < \beta_c$, we claimed T_c has an extreme value distribution whereas in the phase transition regime $\beta > \beta_c$, the distribution of T_c is unknown. The relations (13) and (14) holds for any value of β. Nevertheless, for $\beta < \beta_c$ the random variables T_k are near from independence enough to ensure the some renormalisation towards an extreme value distribution like emphasized with Th. 4. For $\beta > \beta_c$, the algorithm has to wait until the coalescence takes place for the majority of the sites. It explains the explosion observed.

References

[Benaïm and El Karoui(2003)] M. Benaïm and N. El Karoui. *Promenade aléatoire (chaînes de Markov, algorithmes, martingales)*. Éditions École Polytechnique, Palaiseau, 2003. Cours École Polytechnique, Palaiseau.

[Bovier(2005)] A. Bovier. Extreme values of random processes. Lecture Notes Technische Universität Berlin, 2005.

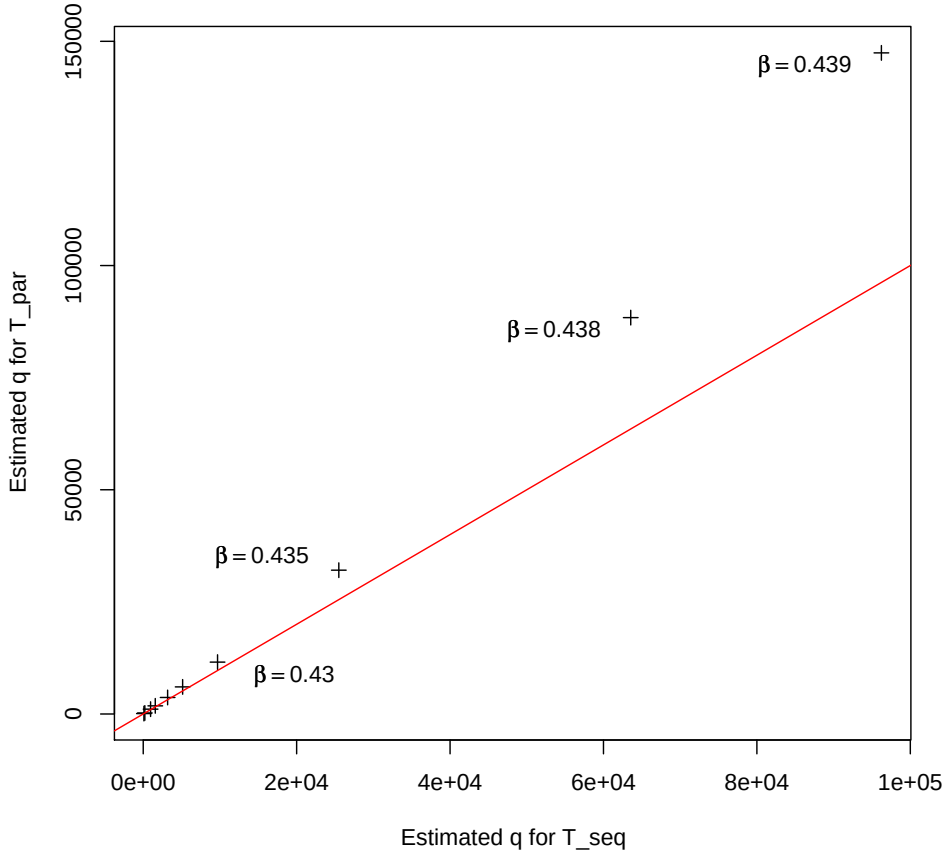
[Brémaud(1999)] P. Brémaud. *Markov chains*. Springer-Verlag, New York, 1999. ISBN 0-387-98509-3. Gibbs fields, Monte Carlo simulation, and queues.

[Cowles and Rosenthal(1998)] M.K. Cowles and J.S Rosenthal. A simulation approach to convergence rates for markov chain monte carlo algorithms. *Stat. Comput.*, 8(2):115–124, 1998. ISSN 0960-3174.

[Dai Pra et al.(2002)Dai Pra, Louis, and Rœlly] P. Dai Pra, P.-Y. Louis, and S. Rœlly. Stationary measures and phase transition for a class of probabilistic cellular automata. *ESAIM : Probability and Statistics*, 6:89–104, 2002.

[Dai Pra et al.(2006)Dai Pra, Louis, and Minelli] P. Dai Pra, P.-Y. Louis, and I. Minelli. Monotonicity and complete monotonicity for continuous-time Markov chains. *C. R. Math. Acad. Sci. Paris*, 342(12):965–970, 2006. ISSN 1631-073X.

[David A. Levin(2008)] Elizabeth L. Wilmer David A. Levin, Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.



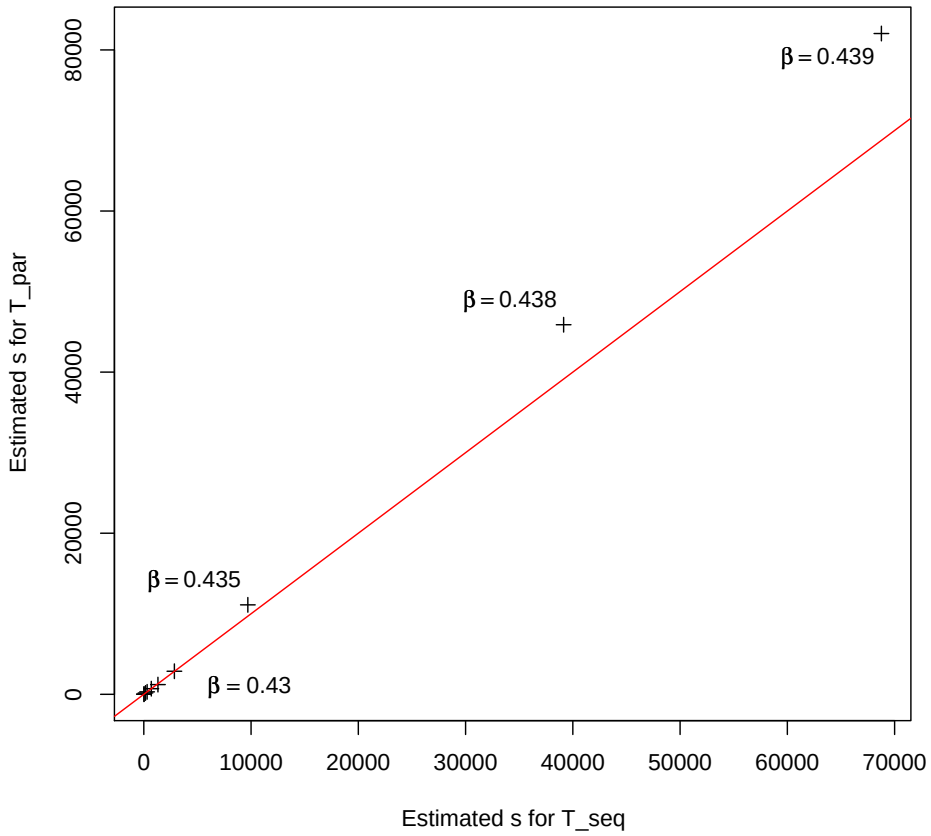


Figure 4: $\hat{q}$ (resp. $\hat{s}$) for T_c^{par} vs. $\hat{q}$ (resp. $\hat{s}$) for T_c^{seq}, parametrized by β. The values of β are listed in Table 1. For comparison, the red line is $y = x$.

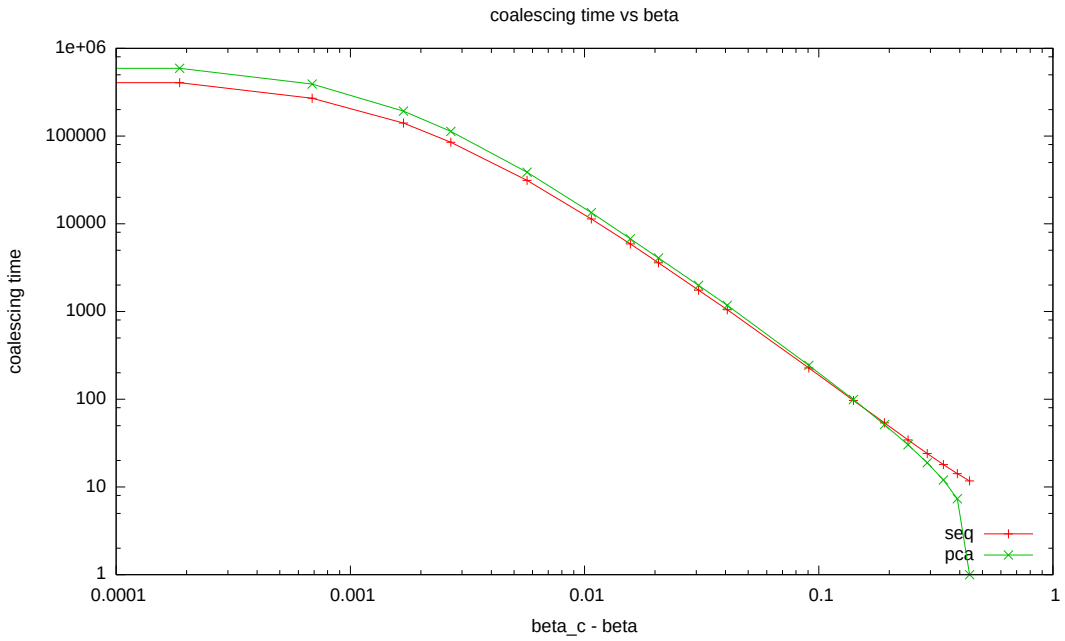


Figure 5: Dependence of T_c in β; sequential and parallel updating

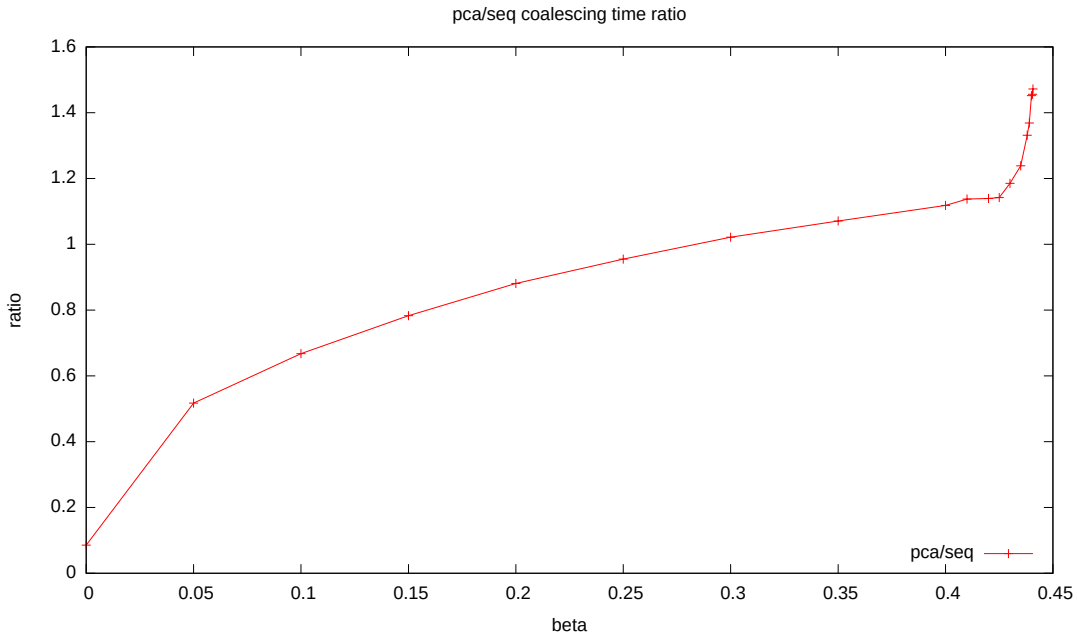


Figure 6: Ratio $\overline{T_c^{\mathrm{par}}}$ on $\overline{T_c^{\mathrm{seq}}}$ for $\beta < \beta_c$

(a) parallel updating, $T_c^{\text{par}} = f(\beta)$

(b) sequential updating, $T_c^{\text{seq}} = f(\beta)$

Figure 7: Boxplot from T_c^{par} and T_c^{seq} when $0 < \beta < \beta_c$, samples of size 1000, lattice $200 * 200$ (log-scale in y).

[Dawson(1974)] D. A. Dawson. Synchronous and asynchronous reversible Markov systems. *Canad. Math. Bull.*, 17(5):633–649, 1974.

[Diaconis and Freedman(1999)] P. Diaconis and D. Freedman. Iterated random functions. *SIAM Rev.*, 41(1):45–76 (electronic), 1999. ISSN 0036-1445.

[Diaconis and Ram(2000)] P. Diaconis and A. Ram. Analysis of systematic scan metropolis algorithms using iwahori-hecke algebra techniques. *Michigan Math. J.*, 48(1):157–190, 2000.

[Diaconis and Saloff-Coste(1998)] P. Diaconis and L. Saloff-Coste. What do we know about the Metropolis algorithm? *J. Comput. System Sci.*, 57(1):20–36, 1998. ISSN 0022-0000. 27th Annual ACM Symposium on the Theory of Computing (STOC'95) (Las Vegas, NV).

[Dyer et al.(2006a)Dyer, Goldberg, and Jerrum] M. Dyer, L.A. Goldberg, and M. Jerrum. Systematic scan for sampling colorings. *Ann. Appl. Prob.*, 16(1):185–230, 2006a.

[Dyer et al.(2006b)Dyer, Goldberg, Jerrum, and Martin] M. Dyer, L.A. Goldberg, M. Jerrum, and R. Martin. Markov chain comparison. *Proba. Surv.*, 3:89–111, 2006b.

[Embrechts et al.(2008)Embrechts, Klüppelberg, and Mikosch] P. Embrechts, C. Klüppelberg, and Th. Mikosch, editors. *Modelling Extremal Events*. Springer, 2008. ISBN 978-3-54060931-5.

[Fishman(1996)] G.S. Fishman. Coordinate selection rules for gibbs sampling. *Ann. Appl. Probab.*, 6(2):444–465, 1996.

[Frigessi et al.(1993)Frigessi, Hwang, di Stefano, and Sheu] A. Frigessi, C.-R. Hwang, P. di Stefano, and S.-J. Sheu. Convergence rates of the gibbs sampler, the metropolis algorithm and other single-site updating dynamics. *J.R. Statis. Soc. B*, 55(1):205–219, 1993.

[Georgii(1988)] H.-O. Georgii. *Gibbs measures and phase transitions*. Walter de Gruyter & Co., Berlin, 1988. ISBN 0-89925-462-4.

[Gilks et al.(1996)Gilks, Richardson, and Spiegelhalter] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, editors. *Markov chain Monte Carlo in practice*. Interdisciplinary Statistics. Chapman & Hall, London, 1996. ISBN 0-412-05551-1.

[Häggström(2002)] O. Häggström. *Finite Markov chains and algorithmic applications*, volume 52 of *London Mathematical Society Student Texts*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-81357-3; 0-521-89001-2.

[Johnson et al.(1994)Johnson, Kotz, and Balakrishnan] N. L. Johnson, S. Kotz, and N. Balakrishnan. *Continuous univariate distributions. Vol. 1*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1994. ISBN 0-471-58495-9. A Wiley-Interscience Publication.

[Liggett(1985)] T. M. Liggett. *Interacting particle systems*. Springer-Verlag, New York, 1985. ISBN 0-387-96069-4.

[Louis(2002)] P.-Y. Louis. *Automates Cellulaires Probabilistes : mesures stationnaires, mesures de Gibbs associées et ergodicité*. PhD thesis, Université de Lille 1 and Politecnico di Milano, september 2002. URL http://tel.ccsd.cnrs.fr/documents/archives0/00/00/22/45/index_fr.html.

[Louis(2005)] P.-Y. Louis. Increasing coupling of probabilistic cellular automata. *Statist. Probab. Lett.*, 74(1):1–13, 2005. ISSN 0167-7152.

[Martinelli(1999)] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on probability theory and statistics (Saint-Flour, 1997)*, volume 1717 of *Lecture Notes in Math.*, pages 93–191. Springer, Berlin, 1999.

[Murdoch and Rosenthal(2000)] D.J. Murdoch and J.S. Rosenthal. Efficient use of exact samples. *Stat. Comput.*, 10(3):237–243, 2000. ISSN 0960-3174.

[Peres(2005)] Yuval Peres. Mixing for markov chains and spin systems. Lecture Notes, 2005. URL http://www.stat.berkeley.edu/~peres/ubc.pdf.

[Propp and Wilson(1996)] J. G. Propp and D. B. Wilson. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Algorithms*, 9(1-2):223–252, 1996.

[Randall and Tetali(2000)] Dana Randall and Prasad Tetali. Analyzing Glauber dynamics by comparison of Markov chains. *J. Math. Phys.*, 41(3):1598–1615, 2000. ISSN 0022-2488.

[Ricci(2005)] V. Ricci. Fitting distributions with r. Notes, R-project, 2005. URL http://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf.

[Roberts and Rosenthal(2004)] G.O. Roberts and J.S. Rosenthal. General state space markov chains and MCMC algorithms. *Probab. Surv.*, 1:299–320, 2004. ISSN 1549-5787.

[Sokal(1989)] A.D. Sokal. Monte Carlo Methods in Statistical Mechanics: Foundations and New Algorithms. *Cours de Troisième Cycle de la Physique en Suisse Romande, Lausanne*, 105: 133, 1989.

[Wilson(2000)] D. B. Wilson. How to couple from the past using a read-once source of randomness. *Random Structures Algorithms*, 16(1):85–113, 2000. ISSN 1042-9832.

[Yacoubi et al.(2006)Yacoubi, Chopard, and Bandini] Samira El Yacoubi, Bastien Chopard, and Stefania Bandini, editors. *Cellular Automata, 7th International Conference on Cellular Automata, for Research and Industry, ACRI 2006, Perpignan, France, September 20-23, 2006, Proceedings*, volume 4173 of *Lecture Notes in Computer Science*, 2006. Springer. ISBN 3-540-40929-7.