

# Sampling the Fermi statistics and other conditional product measures<sup>1</sup>

A. Gaudillièr<sup>a</sup> and J. Reygner<sup>b</sup>

<sup>a</sup>LATP, Université de Provence, CNRS, 39 rue F. Joliot Curie, 13013 Marseille, France. E-mail: [gaudilli@cmi.univ-mrs.fr](mailto:gaudilli@cmi.univ-mrs.fr)

<sup>b</sup>CMAP, École Polytechnique, Route de Saclay, 91120 Palaiseau, France. E-mail: [julien.reygner@polytechnique.edu](mailto:julien.reygner@polytechnique.edu)

Received 27 November 2009; revised 28 June 2010; accepted 15 July 2010

---

**Abstract.** Through a Metropolis-like algorithm with single step computational cost of order one, we build a Markov chain that relaxes to the canonical Fermi statistics for  $k$  non-interacting particles among  $m$  energy levels. Uniformly over the temperature as well as the energy values and degeneracies of the energy levels we give an explicit upper bound with leading term  $km \ln k$  for the mixing time of the dynamics. We obtain such construction and upper bound as a special case of a general result on (non-homogeneous) products of ultra log-concave measures (like binomial or Poisson laws) with a global constraint. As a consequence of this general result we also obtain a disorder-independent upper bound on the mixing time of a simple exclusion process on the complete graph with site disorder. This general result is based on an elementary coupling argument, illustrated in a simulation appendix and extended to (non-homogeneous) products of log-concave measures.

**Résumé.** En définissant un algorithme de type Metropolis dont le coût de chaque pas est d'ordre 1, nous construisons une chaîne de Markov dont la mesure d'équilibre est donnée par la statistique de Fermi canonique pour  $k$  particules sans interaction parmi  $m$  niveaux d'énergie. Uniformément en la température, ainsi qu'en les énergies et capacités des différents niveaux d'énergie, nous donnons une majoration explicite et de terme dominant  $km \ln k$  du temps de mélange de la dynamique. Nous obtenons cette construction et cette majoration comme cas particulier d'un résultat général sur les produits (non homogènes) de mesures ultra log-concaves (comme les lois binômiales ou de Poisson) sous une contrainte globale. Ce résultat général fournit aussi une majoration indépendante du désordre pour le temps de mélange du processus d'exclusion simple sur le graphe complet en potentiel aléatoire. Il découle d'un argument de couplage élémentaire, est illustré dans une appendice de simulations et étendu aux produits (non homogènes) de mesures log-concaves.

MSC: 60J10; 82C44

Keywords: Metropolis algorithm; Markov chain; Sampling; Mixing time; Product measure; Conservative dynamics

---

## 1. From the Fermi statistics to general conditional products of log-concave measures

### 1.1. Sampling the Fermi statistics

Given two positive integers  $k$  and  $m$ , given a non-negative real number  $\beta$ , given  $m$  real numbers  $v_1, \dots, v_m$  and given  $m$  integers  $n_1, \dots, n_m$  such that

$$n := \sum_{j=1}^m n_j \geq k, \tag{1.1}$$

---

<sup>1</sup>The whole research for this paper was done at the Dipartimento di matematica dell'università Roma Tre, during A.G.'s post-doc and J.R.'s research internship. This work was supported by the European Research Council through the "Advanced Grant" PTRELESS 228032.

the canonical Fermi statistics at *inverse temperature*  $\beta$  for  $k$  non-interacting particles among the  $m$  *energy levels*  $1, \dots, m$ , with *energy values*  $v_1, \dots, v_m$  and *degeneracies*  $n_1, \dots, n_m$  is the conditional probability measure on

$$\mathcal{X}_{k,m} := \{(k_1, \dots, k_m) \in \mathbb{N}^m : k_1 + \dots + k_m = k\} \tag{1.2}$$

given by

$$\nu := \mu(\cdot | \mathcal{X}_{k,m}) \tag{1.3}$$

with  $\mu$  the product measure on  $\mathbb{N}^m$  such that

$$\mu(k_1, \dots, k_m) := \frac{1}{Z} \prod_{j=1}^m \binom{n_j}{k_j} \exp\{-\beta k_j v_j\}, \tag{1.4}$$

$$Z := \sum_{k_1, \dots, k_m} \prod_{j=1}^m \binom{n_j}{k_j} \exp\{-\beta k_j v_j\} \tag{1.5}$$

with  $\binom{n_j}{k_j} = 0$  whenever  $k_j > n_j$ . In other words,  $\nu$  is a (non-homogeneous) product of binomial laws in  $k_1, \dots, k_m$  with the global constraint

$$k_1 + \dots + k_m = k \tag{1.6}$$

and we can write

$$\nu(k_1, \dots, k_m) = \frac{1}{Q} \prod_{j=1}^m e^{\phi_j(k_j)}, \quad (k_1, \dots, k_m) \in \mathcal{X}_{k,m}, \tag{1.7}$$

where the  $\phi_j$  are defined by

$$\phi_j : k_j \in \mathbb{N} \mapsto -\beta k_j v_j + \ln \binom{n_j}{k_j} \in \mathbb{R} \cup \{-\infty\} \tag{1.8}$$

and  $Q$  is such that  $\nu$  is a probability measure.

The first aim of this paper is to describe an algorithm that simulates a sampling according to  $\nu$  in a time that can be bounded from above by an explicit polynomial in  $k$  and  $m$ , uniformly over  $\beta$ ,  $(v_j)_{1 \leq j \leq m}$  and  $(n_j)_{1 \leq j \leq m}$ . The reason why we prefer a bound in  $k$  and  $m$  rather than in the ‘volume’ of the system  $n = \sum_j n_j$ , will be clarified later.

A first naive (and wrong) idea to do so consists in choosing the position (the energy level) of a first, second, ... and eventually  $k$ th particle in the following way. First choose randomly the position of the first particle according to the exponential weights associated with the ‘free entropies’ of the empty sites, that is choose level  $j$  with a probability proportional to  $\exp\{-\beta v_j + \ln n_j\}$ . Then decrease by 1 the degeneracy of the chosen energy level and repeat the procedure to choose the position of the second, third, ... and eventually  $k$ th particle. It is easy to check that, doing so, the final distribution of the occupation numbers  $k_1, \dots, k_m$  associated with the different energy levels, that is of the numbers of particles placed in each level, is in general not given by  $\nu$  as soon as  $k$  is larger than one. But it turns out that this naive idea can be adapted to build an efficient algorithm to perform approximate samplings under the Fermi statistics.

Very classically, the fast sampling performed by the algorithm we will build will be obtained by running a Markov chain  $X$  with transition matrix  $p$  on  $\mathcal{X}_{k,m}$  and with equilibrium measure  $\nu$ . The efficiency of the algorithm will be measured through the bounds that we will be able to give on the *mixing time*  $t_\varepsilon$ , defined for any positive  $\varepsilon < 1$  by

$$t_\varepsilon := \inf\{t \geq 0 : d(t) \leq \varepsilon\}, \tag{1.9}$$

$$d(t) := \max_{\eta \in \mathcal{X}_{k,m}} \|p^t(\eta, \cdot) - \nu\|_{\text{TV}}, \tag{1.10}$$

where  $\|\cdot\|_{\text{TV}}$  stands for the *total variation distance* defined for any probability measures  $\nu_1$  and  $\nu_2$  on  $\mathcal{X}_{k,m}$  by

$$\|\nu_1 - \nu_2\|_{\text{TV}} := \max_{A \subset \mathcal{X}_{k,m}} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sum_{\eta \in \mathcal{X}_{k,m}} |\nu_1(\eta) - \nu_2(\eta)|. \quad (1.11)$$

As a consequence, estimating mixing times is not the only one issue of this paper, building a ‘good’ Markov chain is part of the problem.

As far as that part of the problem is concerned, we propose to build a Metropolis-like algorithm that uses the ‘free energies’ of the naive approach to define a conservative dynamics. Assuming that at time  $t \in \mathbb{N}$  the system is in some configuration  $X_t = \eta$  in  $\mathcal{X}_{k,m}$  with  $\nu(\eta) > 0$ , and defining for any  $\eta = (k_1, \dots, k_m)$  and any distinct  $i$  and  $j$  in  $\{1; \dots; m\}$

$$\eta^{ij} := (k'_1, \dots, k'_m) \quad \text{with } k'_s = \begin{cases} k_s & \text{for } s \in \{1; \dots; m\} \setminus \{i; j\}, \\ k_i - 1 & \text{if } s = i, \\ k_j + 1 & \text{if } s = j, \end{cases} \quad (1.12)$$

the configuration at time  $t + 1$  will be decided as follows:

- choose a *particle* with uniform probability (it will stand in a given level  $i$  with probability  $k_i/k$ ),
- choose an *energy level* with uniform probability (a given level  $j$  will be chosen with probability  $1/m$ ),
- with  $i$  the level where stood the chosen particle and  $j$  the chosen energy level, extract a uniform variable  $U$  on  $[0; 1)$  and set  $X_{t+1} = \eta^{ij}$  if  $i \neq j$  and

$$U < \exp\{-\beta v_j + \ln(n_j - k_j) + \beta v_i - \ln(n_i - (k_i - 1))\}, \quad (1.13)$$

$X_{t+1} = \eta$  if not.

In other words, denoting by  $[a]^+ = (a + |a|)/2$  the positive part of any real number  $a$  and with

$$\psi_j : k_j \in \{0; \dots; n_j\} \mapsto -\beta v_j + \ln(n_j - k_j) \in \mathbb{R} \cup \{-\infty\}, \quad j \in \{1; \dots; m\}, \quad (1.14)$$

for any distinct  $i$  and  $j$  in  $\{1; \dots; m\}$

$$P(X_{t+1} = \eta^{ij} | X_t = \eta) = p(\eta, \eta^{ij}) = \frac{k_i}{k} \frac{1}{m} \exp\{-[\psi_i(k_i - 1) - \psi_j(k_j)]^+\} \quad (1.15)$$

and

$$P(X_{t+1} = \eta | X_t = \eta) = p(\eta, \eta) = 1 - \sum_{i \neq j} p(\eta, \eta^{ij}). \quad (1.16)$$

**Remark.** In order to avoid any ambiguity in (1.15) in the case  $k_i = 0$ , we set  $\psi_i(-1) = +\infty$  (even though the algorithm we described does not require any convention for  $\psi_i(-1)$ ).

This Markov chain is certainly irreducible and aperiodic. To prove that it relaxes to  $\nu$  we will check the reversibility of the process with respect to  $\nu$ . Then we will have to estimate the mixing time of the process. We will carry out both the tasks in a more general setup.

## 1.2. A general result

For any function  $f : \mathbb{N} \rightarrow \mathbb{R}$  we define

$$\nabla^+ f : x \in \mathbb{N} \mapsto \nabla_x^+ f := f(x + 1) - f(x), \quad (1.17)$$

$$\nabla^- f : x \in \mathbb{N} \setminus \{0\} \mapsto \nabla_x^- f := f(x - 1) - f(x), \quad (1.18)$$

$$\Delta f : x \in \mathbb{N} \setminus \{0\} \mapsto \Delta_x f := \nabla_x^+ f + \nabla_x^- f = -\nabla_x^-(\nabla^+ f) \quad (1.19)$$

and we say that a measure  $\gamma$  on the integers

$$\gamma : x \in \mathbb{N} \mapsto e^{\phi(x)} \in \mathbb{R}_+ \tag{1.20}$$

with  $\phi : \mathbb{N} \rightarrow \mathbb{R} \cup \{-\infty\}$ , is log-concave if  $\mathbb{N} \setminus \gamma^{-1}(\{0\})$  is an interval of the integers and

$$\gamma(x)^2 \geq \gamma(x-1)\gamma(x+1), \quad x \in \mathbb{N} \setminus \{0\}, \tag{1.21}$$

i.e., if  $\nabla^+ \phi$  is non-increasing, or, equivalently,  $-\Delta \phi$  is non-negative (with the obvious extension of the previous definitions to such a possibly non-finite  $\phi$ ). The measure  $\mu$  defined in (1.4) is a product of log-concave measures and the canonical Fermi statistics is such a product measure normalized over the condition (1.6).

**N.B.** From now on, and except for explicit mentioning of additional hypotheses, we will only assume that the probability measure  $\nu$  we want to sample is a product of log-concave measures normalized over the global constraint (1.6), i.e., that  $\nu$  is a probability on  $\mathcal{X}_{k,m}$  that can be written in the form (1.7) with non-increasing  $\nabla^+ \phi_j$ 's.

In this more general setup we will often refer to the indices  $j$  in  $\{1; \dots; m\}$  as *sites* rather than energy levels of the system.

Actually the  $e^{\phi_j}$ 's of the Fermi statistics are much more than log-concave measures. They are *ultra log-concave* measures according to the following definition by Pemantle [14] and Liggett [12].

**Definition 1.2.1.** A measure  $\gamma : \mathbb{N} \rightarrow \mathbb{R}_+$  is ultra log-concave if  $x \mapsto x! \gamma(x)$  is log-concave.

In other words  $e^{\phi_j}$  is ultra log-concave if and only if

$$\psi_j := \nabla^+ \phi_j + \ln(1 + \cdot) \tag{1.22}$$

is non-increasing (for the Fermi statistics observe that so are the  $\phi_j$ 's and that (1.22) is consistent with (1.14)).

For birth and death processes that are reversible with respect to ultra log-concave measures, Caputo, Dai Pra and Posta [6] proved modified log-Sobolev inequalities and stronger convex entropy decays, both giving good upper bounds on the mixing time of the processes. Johnson [8] proved also easier Poincaré inequalities that give weaker bounds on the mixing times. We refer to [1,13,15], for an introduction to this classical functional inequality approach to convergence to equilibrium and we note that for such birth and death processes the ultra-log concavity hypotheses allowed for a Bakry–Émery like approach (see [2]) to derive (modified) log-Sobolev inequalities. Actually, [6] was an attempt to extend this celebrated analysis to Markov processes with jumps (see also [9] for a more geometric perspective). But it turns out, as we will soon review, that beyond the case of birth and death processes the role of ultra log-concavity to follow the Bakry–Émery line in the discrete setup is still unclear.

In this paper we will not follow the functional analysis approach to control mixing times. To bound from above the mixing time of a Markov chain that is reversible with respect to a conditional product of ultra log-concave measures, we will follow (and recall in Section 2) the non-less classical, and, in this case, elementary, probabilistic approach via coalescent coupling. We will prove:

**Theorem 1.** *If  $\nu$  derives from a product of ultra log-concave measures, then the Markov chain with transition matrix  $p$  defined by*

$$p(\eta, \eta^{ij}) = \frac{k_i}{k} \frac{1}{m} \exp\{-[\psi_i(k_i - 1) - \psi_j(k_j)]^+\},$$

$$\eta = (k_1, \dots, k_m) \in \mathcal{X}_{k,m} \setminus \nu^{-1}\{0\},$$

$$i \neq j \in \{1; \dots; m\}, \tag{1.23}$$

$$p(\eta, \eta) = 1 - \sum_{i \neq j} p(\eta, \eta^{ij}) \tag{1.24}$$

is reversible with respect to  $\nu$  and, for any positive  $\varepsilon < 1$ , its mixing time  $t_\varepsilon$  satisfies

$$t_\varepsilon \leq km \ln(k/\varepsilon). \quad (1.25)$$

**Proof.** See Section 2. □

The most relevant point of Theorem 1 with respect to the previous results we know stands in the uniformity of the upper bound above the disorder of the system (except for the ultra log-concavity hypothesis on  $e^{\phi_j}$  in each  $j$ ). In particular and as far as the Fermi statistics is concerned, our estimate does not depend on the temperature, and, more generally it is independent from the energy values as well as the level degeneracies.

To illustrate this fact let us start with the case  $n_j = 1$  for all  $j$ . In this case our dynamics is a simple exclusion process with site disorder. Caputo [4,5] proved Poincaré inequalities for such processes, in their continuous time version, assuming a uniform lower (and upper) bound on general transition rates, while Caputo, Dai Pra and Posta [6], looking at particular rates for the process and still assuming moderate disorder – that is, uniform lower and upper bounds on these rates – proved a modified log-Sobolev inequality. For the particular choice of rates they made, the upper bound on the mixing time implied by [6] could not hold in a strong disorder context (for example with  $k = 1$ ,  $m = 3$ ,  $v_1 = v_2 = 0$ ,  $v_3 > 0$  and  $\beta \gg 1$ ). Our uniformity over the disorder of the system depends then strongly on our particular choice for the transition probabilities. As it is often the case with Markov processes on discrete state space, the details of the dynamics are not less important than the properties of its equilibrium measure.

Under the moderate disorder hypotheses of [4–6], however, the upper bound on the mixing time implied by [6] and suggested by [4,5] is better than our upper bound in Theorem 1 (by a factor of order  $k$ ). But we will see that introducing such moderate disorder hypotheses in our arguments directly improves our result by a factor of order  $m \geq k$  (see the last remark at the end of Section 2).

To close the discussion on simple exclusion processes we note that those of the arguments in [5] that do not depend on the disorder suggest an upper bound on the mixing time of order  $m^2 \ln m$ . Theorem 1 improves this estimate when  $k$  is small with respect to  $m$ .

As far as more general conditional non-homogeneous product of log-concave measures are concerned, we stress once again that Theorem 1 gives a uniform bound over the disorder that can be improved by a factor of order  $m$  by adding moderate disorder hypotheses (see the last remark at the end of Section 1.4 and our [Simulation appendix](#)). Then, such product measures can be equilibrium measures of zero-range processes with a continuous time generator as in [3,6]:

$$\mathcal{L}_\eta f := \frac{1}{m} \sum_{i \neq j} c_i(k_i) (f(\eta^{ij}) - f(\eta)), \quad (1.26)$$

where the  $c_j : k_j \in \mathbb{N} \mapsto [0, +\infty)$  are such that  $c_j(0) = 0$  and  $c_j(k_j) > 0$  for  $k_j > 0$ . Indeed, such a process is reversible with respect to  $\nu$  provided that

$$\forall j \in \{1; \dots; m\}, \forall k_j \in \mathbb{N}, \quad \phi_j(k_j) = \sum_{0 < l \leq k_j} \ln \frac{1}{c_j(l)}. \quad (1.27)$$

Boudou, Caputo, Dai Pra and Posta [3] proved a Poincaré inequality for such a process, assuming that there exists a positive  $c$  such that

$$\forall j \in \{1; \dots; m\}, \forall k_j \in \mathbb{N}, \quad c_j(k_j + 1) - c_j(k_j) \geq c. \quad (1.28)$$

This is a moderate disorder hypothesis that implies the log-concavity of the  $\mu_j$ . To go to modified log-Sobolev inequalities, i.e., to good mixing time estimates rather than simple gap estimates, there is the additional hypothesis in [6] that there exists a non-negative  $\delta < c$  such that

$$\forall j \in \{1; \dots; m\}, \forall k_j \in \mathbb{N}, \quad c_j(k_j + 1) - c_j(k_j) \leq c + \delta. \quad (1.29)$$

Clearly there are ultra log-concave measures that do not satisfy (1.29): to have an ultra log-concave measure one needs a strongly decreasing  $\nabla^+ \phi_j$ , i.e., a strongly increasing  $c_j$ . Conversely, it is not true that (1.28) and (1.29) imply ultra log-concavity, i.e.,

$$\forall j \in \{1; \dots; m\}, \forall k_j \geq 1, \quad \ln \frac{k_j + 1}{c_j(k_j + 1)} \leq \ln \frac{k_j}{c_j(k_j)}. \tag{1.30}$$

However, elementary algebra shows that (1.28) together with (1.29) implies

$$\forall j \in \{1; \dots; m\}, \forall k_j \geq 1, \quad \ln \frac{k_j + 1/2}{c_j(k_j + 1)} \leq \ln \frac{k_j}{c_j(k_j)} \tag{1.31}$$

and this strangely looks like (1.30). Therefore we said that the role played by ultra log-concavity is still unclear along the functional analysis line of research in the discrete setup. We conclude stressing once again that in this paper we will follow a different line, that our uniform bound on the mixing time comes from an elementary coupling argument and that we do not need any moderate disorder hypothesis.

### 1.3. Interpolating between sites and particles

It seems that today available techniques are such that the less log-concavity we have, the more homogeneity we need to control the convergence to equilibrium. Staying to the papers mentioned above, without ultra log-concavity or at least something that looks like ultra log-concavity we only have Poincaré inequalities for non-homogeneous product of log-concave measures, and without log-concavity we have modified log-Sobolev inequalities for homogeneous product measures only (see [6]). In addition, the only result we know for a conservative dynamics in (weakly) disordered context and with an equilibrium measure that is a product of measures that are not log-concave is that of Landim and Noronha Neto [10] for the (continuous) Ginzburg–Landau process.

We will see that all the ideas of the proof of Theorem 1 can be extended to deal with a large class of conditional product of log-concave measures that are not ultra log-concave. To do so, let us define

$$\delta := \max \left\{ \lambda \in [0; 1]: \forall j \in \{1; \dots; m\}, \forall k_j > 0, -\Delta_{k_j} \phi_j \geq \lambda \ln \frac{1 + k_j}{k_j} \right\}. \tag{1.32}$$

In other words,  $\delta$  is the largest real number in  $[0; 1]$  for which all the

$$\psi_j^{[\delta]} := \nabla^+ \phi_j + \delta \ln(1 + \cdot), \quad j \in \{1; \dots; m\}, \tag{1.33}$$

are non-increasing. Denoting by  $a \wedge b$  the minimum of two real numbers  $a, b$  and defining

$$l_\delta := k^\delta (k \wedge m)^{1-\delta} \tag{1.34}$$

we will prove:

**Theorem 2.** *If  $\delta > 0$  then the Markov chain with transition matrix  $p$  defined by*

$$p(\eta, \eta^{ij}) = \frac{k_i^\delta}{l_\delta} \frac{1}{m} \exp\{-[\psi_i^{[\delta]}(k_i - 1) - \psi_j^{[\delta]}(k_j)]^+\},$$

$$\eta = (k_1, \dots, k_m) \in \mathcal{X}_{k,m} \setminus v^{-1}\{0\},$$

$$i \neq j \in \{1; \dots; m\}, \tag{1.35}$$

$$p(\eta, \eta) = 1 - \sum_{i \neq j} p(\eta, \eta^{ij}) \tag{1.36}$$

*is reversible with respect to  $v$  and, for any positive  $\varepsilon < 1$ , its mixing time  $t_\varepsilon$  satisfies*

$$t_\varepsilon \leq \frac{(k \wedge m)^{1-\delta}}{\delta} km \ln(k/\varepsilon). \tag{1.37}$$

**Remark 1.** By Hölder’s inequality, if  $\delta > 0$  then

$$\sum_{i=1}^m k_i^\delta = \sum_{i=1}^m k_i^\delta \mathbb{1}_{\{k_i \neq 0\}}^{1-\delta} \leq \left( \sum_{i=1}^m k_i \right)^\delta \left( \sum_{i=1}^m \mathbb{1}_{\{k_i \neq 0\}} \right)^{1-\delta} \leq l_\delta \tag{1.38}$$

and this ensures that (1.35)–(1.36) define a probability matrix.

**Remark 2.** As far as this can make sense in our discrete setup, we note that the hypothesis  $\delta > 0$  is slightly weaker than a “uniform strict log-concavity hypothesis” (see (1.32)).

**Proof of Theorem 2.** See Section 3. □

For  $\delta = 1$  the transition matrix represents an algorithm starting with a uniform choice of a particle. For  $\delta = 0$  Theorem 2 is empty but (1.35)–(1.36) still define a Markov chain  $X$  that can be seen as a particular version of a discrete state space non-homogeneous Ginzburg–Landau process. In this case the transition matrix represents an algorithm that starts with a uniform choice of a (non-empty) site. The case  $0 < \delta < 1$  can be seen as an interpolation between uniform choices of site and particle. More precisely, assuming that at time  $t \in \mathbb{N}$  the system is in some configuration  $X_t = \eta = (k_1, \dots, k_m)$  in  $\mathcal{X}_{k,m} \setminus v^{-1}\{0\}$ , the configuration at time  $t + 1$  can be decided as follows.

- Choose a site  $i$  or no site at all with probabilities proportional to  $k_i^\delta$  and  $l_\delta - \sum_i k_i^\delta$ .
- If some site  $i$  was chosen, then proceed as in the previously described algorithm using the functions  $\psi_j^{[\delta]}$  instead of the  $\psi_j$ ’s, if not, then set  $X_{t+1} = \eta$ .

#### 1.4. Last remarks and original motivation

First, we note that as long as one wants bounds that are uniform over the disorder, Theorem 1 gives the right order for the mixing time: for the Fermi statistic in the very low temperature regime with, for example,  $v_1 < v_2 < \dots < v_m$  and  $n_1, n_2 \geq k$ , the equilibrium measure will be concentrated on  $(k, 0, \dots, 0)$  while, starting from  $(0, k, 0, \dots, 0)$ , the system will reach the ground state in a time of order  $km \ln k$  for large  $k$  (this is a coupon-collector estimate). This fact is illustrated in our [Simulation appendix](#).

Next, we observe that, with our definitions,  $p(\eta, \eta)$  can often be close to one, especially in strong disorder situations or when the right- and left-hand sides in (1.38) are far from each other. If one would like to use these results to perform practical simulations, then it could be useful to note that the computational time would still be improved by implementing an algorithm that at each step simulates, for a given configuration  $\eta$  on the trajectory of the Markov chain, the elapsed time before the particle reach a *different* configuration (this is a geometric time) and choose this configuration  $\eta' \neq \eta$  according to the (easy to compute) associated law. It would then be enough to stop the algorithm as soon as the total simulated time goes beyond the mixing time (and then return the last configuration, that the system occupied at the mixing time) rather than waiting for the original algorithm to make a step number equal to the mixing time.

Turning back to the first naive and wrong idea, it is interesting to note that it can easily be modified to determine the most probable states of the system, i.e., the configurations  $\eta^* = (k_1^*, \dots, k_m^*)$  in  $\mathcal{X}_{k,m}$  such that

$$\sum_{j=1}^m \phi_j(k_j^*) = \max_{k_1 + \dots + k_m = k} \sum_{j=1}^m \phi_j(k_j). \tag{1.39}$$

One can prove, using the concavity of the  $\phi_j$ ’s, that the most probable configurations for the system with  $k$  particles can be obtained from the most probable configurations  $(k'_1, \dots, k'_m)$  for the system with  $k - 1$  particles simply adding one particle where the corresponding gain in ‘free energy’ is the highest, that is in  $j^*$  such that

$$\nabla_{k_{j^*}'}^+ \phi_{j^*} = \max_j \nabla_{k'_j}^+ \phi_j. \tag{1.40}$$

As a consequence one can place the particles one by one, each time maximizing this free energy gain, to build the most probable configuration.

Then, as a referee pointed out, since the sampling problem is a trivial one in the Poissonian case (when all the  $\mu_j$ 's are Poisson measures, so that  $\nu$  is nothing but a multinomial law  $M$ ), one can ask about the expected time needed to perform a rejection sampling with respect to the multinomial case. It is given by  $\max_{\eta} \nu(\eta)/M(\eta)$ . If we take the Fermi statistics with  $m = k$  and all the  $n_j$  equal to 1, then we find (optimizing on the multinomial parameters by a geometric/arithmetic mean comparison)  $k^k/k! \sim e^k/\sqrt{2\pi k}$ . Of course, the sampling problem for that Fermi statistics is also a trivial one. But if we take  $\beta = 0$  and all the  $n_j$  equal to 2, it is not anymore a trivial problem and we find an expected time for the rejection sampling that is logarithmically equivalent to  $(e/2)^k$ .

Turning back to the Fermi statistics we now explain why we were interested in bounds in  $k$  and  $m$  rather than in the volume  $n$ . Iovanella, Scoppola and Scoppola defined in [7] an algorithm to individuate cliques (i.e., complete subgraphs) with  $k$  vertices inside a large Erdős–Reyni random graph with  $n$  vertices. Their algorithm requires to perform repeated approximate samplings of Fermi statistics in volume  $n$ , with  $k$  particles and  $m = 2k + 1$  energy levels. Now, the key observation is that the largest cliques in Erdős–Reyni graphs with  $n$  vertices are of order  $\ln n$ , so that  $k$  and  $m$  in this problem are logarithmically small with respect to  $n$ . Before Theorem 1 the samplings for their algorithm were done by running simple exclusion processes with  $k$  particles on the complete graph (with site disorder) of size  $n$ . Such processes converge to equilibrium in a time of order  $kn \ln k$ . Now the samplings are done in a time of order  $km \ln k \sim 2k^2 \ln k$ , and that was the original motivation of the present work.

Finally we note that our bound in  $k$  and  $m$  is not only useful when  $k$  is small with respect to  $n$  but also when  $k$  is close to  $n$ . In this case we can define a dynamics on the  $n - k$  vacancies rather than on the  $k$  particles.

## 2. Proof of Theorem 1

In this section we assume that  $\nu$  is a measure on  $\mathcal{X}_{k,m}$  deriving from a product of ultra log-concave measures, which means that we can write  $\nu(k_1, \dots, k_m) = (1/Q) \prod_j e^{\phi_j(k_j)}$  and, for all  $j \in \{1; \dots; m\}$ ,  $\psi_j$  defined by (1.22) is non-increasing.

### 2.1. Reversibility

We first prove that the transition matrix  $p$  defined by (1.23) and (1.24) is reversible with respect to the measure  $\nu$ . Let  $\eta = (k_1, \dots, k_m) \in \mathcal{X}_{k,m} \setminus \nu^{-1}\{0\}$  and let  $i \neq j \in \{1; \dots; m\}$ . We have  $p(\eta, \eta^{ij}) \neq 0$  if and only if  $\nu(\eta^{ij}) \neq 0$  and, in that case,

$$\frac{\nu(\eta^{ij})}{\nu(\eta)} = \frac{e^{\phi_i(k_i-1)+\phi_j(k_j+1)}}{e^{\phi_i(k_i)+\phi_j(k_j)}} = \exp\{\nabla_{k_i}^- \phi_i + \nabla_{k_j}^+ \phi_j\} \tag{2.1}$$

while

$$\frac{p(\eta^{ij}, \eta)}{p(\eta, \eta^{ij})} = \frac{k_j + 1}{k_i} \exp\{-[\psi_j(k_j) - \psi_i(k_i - 1)]^+ + [\psi_i(k_i - 1) - \psi_j(k_j)]^+\} \tag{2.2}$$

$$= \exp\{\psi_i(k_i - 1) - \psi_j(k_j) + \ln(k_j + 1) - \ln(k_i)\} \tag{2.3}$$

$$= \exp\{\nabla_{k_i-1}^+ \phi_i - \nabla_{k_j}^+ \phi_j\} \tag{2.4}$$

$$= \exp\{-\nabla_{k_i}^- \phi_i - \nabla_{k_j}^+ \phi_j\} \tag{2.5}$$

so that

$$\nu(\eta)p(\eta, \eta^{ij}) = \nu(\eta^{ij})p(\eta^{ij}, \eta). \tag{2.6}$$

## 2.2. A few words about the coupling method

In order to upper bound the mixing time of the Markov chain with transition matrix  $p$ , we will use the coupling method. Given a Markov chain  $(X^1, X^2)$  on  $\mathcal{X}_{k,m} \times \mathcal{X}_{k,m}$ , we say it is a (Markovian) *coupling* for the dynamics if both  $X^1$  and  $X^2$  are Markov chains with transition matrix  $p$ . Given such a coupling, we define the *coupling time*  $\tau_{\text{couple}}$  as the first (random) time for which the chains meet, that is

$$\tau_{\text{couple}} := \inf\{t \geq 0: X_t^1 = X_t^2\}. \quad (2.7)$$

In this work, every coupling will also satisfy the condition

$$t \geq \tau_{\text{couple}} \Rightarrow X_t^1 = X_t^2. \quad (2.8)$$

Then, it is a well-known fact that for all  $t \geq 0$ ,

$$d(t) \leq \max_{\eta, \theta \in \mathcal{X}_{k,m}} P(\tau_{\text{couple}} > t | X_0^1 = \eta, X_0^2 = \theta), \quad (2.9)$$

where  $d(t)$  is defined by (1.10). A proof of this fact as well as an exhaustive introduction to mixing time theory can be found in [11].

In the proof of both Theorems 1 and 2 we will build a coupling for which there exists a function  $\rho$  that measures in some sense a ‘distance’ between  $X_t^1$  and  $X_t^2$  and from which we will get a bound on the mixing time thanks to the following proposition.

**Proposition 2.2.1.** *Let  $(X^1, X^2)$  be a coupling for a Markov chain with transition matrix  $p$ . We assume that the coupling satisfies the property (2.8). Let  $\rho: \mathcal{X}_{k,m} \times \mathcal{X}_{k,m} \rightarrow \mathbb{N}$  such that  $\rho(\eta, \theta) = 0$  if and only if  $\eta = \theta$ . If  $M$  is the maximum of  $\rho$  and if there exists  $\alpha > 1$  such that, for all  $t \geq 0$ ,*

$$E(\rho(X_{t+1}^1, X_{t+1}^2) | X_t^1, X_t^2) \leq \left(1 - \frac{1}{\alpha}\right) \rho(X_t^1, X_t^2), \quad (2.10)$$

then for all  $\varepsilon > 0$  the mixing time  $t_\varepsilon$  of the dynamics is upper bounded by  $\alpha \ln(M/\varepsilon)$ .

**Proof.** Remark that (2.10) actually means that  $(\rho(X_t^1, X_t^2)(1 - 1/\alpha)^{-t})_{t \in \mathbb{N}}$  is a super-martingale. Taking the expectation we get

$$E(\rho(X_{t+1}^1, X_{t+1}^2)) \leq \left(1 - \frac{1}{\alpha}\right) E(\rho(X_t^1, X_t^2)). \quad (2.11)$$

As a consequence

$$E(\rho(X_t^1, X_t^2)) \leq \left(1 - \frac{1}{\alpha}\right)^t E(\rho(X_0^1, X_0^2)) \leq M e^{-t/\alpha}. \quad (2.12)$$

Since we assume  $\rho(\eta, \theta) = 0$  if and only if  $\eta = \theta$  and (2.8),

$$P(\tau_{\text{couple}} > t) = P(\rho(X_t^1, X_t^2) > 0) = P(\rho(X_t^1, X_t^2) \geq 1). \quad (2.13)$$

From Markov’s inequality and (2.12) we deduce

$$P(\tau_{\text{couple}} > t) \leq E(\rho(X_t^1, X_t^2)) \leq M e^{-t/\alpha}. \quad (2.14)$$

Since the upper bound in (2.14) is uniform in  $X_0^1$  and  $X_0^2$ , according to (2.9) we get

$$d(t) \leq M e^{-t/\alpha}. \quad (2.15)$$

Thus, given  $\varepsilon > 0$ , if  $t \geq \alpha \ln(M/\varepsilon)$  then  $d(t) \leq \varepsilon$ , so that  $t_\varepsilon \leq \alpha \ln(M/\varepsilon)$ .  $\square$

### 2.3. A colored coupling

In order to prove Theorem 1, we introduce a dynamics on the set of the possible distributions of the  $k$  particles in the  $m$  energy levels. Therefore we define the set of the configurations of  $k$  distinguishable particles in  $m$  energy levels

$$\Omega := \{\omega: \{1; \dots; k\} \rightarrow \{1; \dots; m\}\} \quad (2.16)$$

and for every such distribution  $\omega$ , we define  $\xi(\omega) = (k_1, \dots, k_m) \in \mathcal{X}_{k,m}$  by

$$\forall i \in \{1; \dots; m\}, \quad k_i := \sum_{x=1}^k \mathbb{1}_{\{\omega(x)=i\}}. \quad (2.17)$$

We will couple two dynamics  $\omega^1$  and  $\omega^2$  on  $\Omega$ , then we will work with the coupling  $(\xi(\omega^1), \xi(\omega^2))$ . We will build the coupling  $(\omega^1, \omega^2)$  thanks to the coloring we now introduce.

At step  $t$ , a *red–blue coloring* of  $(\omega_t^1, \omega_t^2)$  is a couple of functions  $C^1, C^2: \{1; \dots; k\} \rightarrow \{\text{blue}; \text{red}\}$  such that for all energy level  $i \in \{1; \dots; m\}$ , the number of blue particles in the level  $i$  is the same in both distributions, i.e., if  $|A|$  refers to the cardinality of the set  $A$ ,

$$|\{x: C^1(x) = \text{blue}, \omega_t^1(x) = i\}| = |\{x: C^2(x) = \text{blue}, \omega_t^2(x) = i\}| \quad (2.18)$$

and an energy level cannot contain red particles in both distributions

$$C^1(x) = \text{red} \Rightarrow \forall y, \quad (\omega_t^2(y) = \omega_t^1(x) \Rightarrow C^2(y) = \text{blue}), \quad (2.19)$$

$$C^2(x) = \text{red} \Rightarrow \forall y, \quad (\omega_t^1(y) = \omega_t^2(x) \Rightarrow C^1(y) = \text{blue}). \quad (2.20)$$

As a consequence of (2.18), there exists a one-to-one correspondence

$$\Phi: \{x: C^1(x) = \text{blue}\} \rightarrow \{x: C^2(x) = \text{blue}\} \quad (2.21)$$

such that for all  $x$ ,  $\omega_t^2(\Phi(x)) = \omega_t^1(x)$ . Since the number of blue particles is the same in both distributions, so is the number of red particles. Moreover,  $\xi(\omega_t^1) = \xi(\omega_t^2)$  if and only if all particles are blue. Then we are willing to provide a coupling  $(\omega^1, \omega^2)$  for which the number of red particles is non-increasing. This number does not depend on the red–blue coloring and at any time  $t$  we will call it  $\rho_t$ . Writing  $\xi(\omega_t^1) = (k_1^1, \dots, k_m^1)$  and  $\xi(\omega_t^2) = (k_1^2, \dots, k_m^2)$  we have the identities

$$\rho_t = \frac{1}{2} \sum_{i=1}^m |k_i^1 - k_i^2| = \sum_{i=1}^m [k_i^1 - k_i^2]^+ = \sum_{i=1}^m [k_i^2 - k_i^1]^+. \quad (2.22)$$

We are now ready to build our coupling  $(\omega^1, \omega^2)$ . Given  $(\omega_t^1, \omega_t^2)$  and a red–blue coloring  $(C^1, C^2)$  at step  $t$  (such a coloring certainly exists for any couple of distributions  $(\omega_t^1, \omega_t^2)$ ), let  $x^1$  be a uniform random integer in  $\{1; \dots; k\}$ .

- If  $C^1(x^1) = \text{blue}$ : then we set  $x^2 = \Phi(x^1)$  where  $\Phi$  is provided by (2.21).
- If  $C^1(x^1) = \text{red}$ : let  $x^2$  be a uniform random integer in  $\{x: C^2(x) = \text{red}\}$ .

**Lemma 2.3.1.** *The random integer  $x^2$  has a uniform distribution on the set  $\{1; \dots; k\}$ .*

**Proof.** We write

$$P(x^2 = y) = \sum_{x=1}^k P(x^2 = y | x^1 = x) P(x^1 = x) \quad (2.23)$$

$$= \sum_{x: C^1(x)=\text{blue}} \mathbb{1}_{\{y=\Phi(x)\}} \frac{1}{k} + \sum_{x: C^1(x)=\text{red}} \frac{\mathbb{1}_{\{C^2(y)=\text{red}\}}}{\rho_t} \frac{1}{k} \quad (2.24)$$

$$= \frac{\mathbb{1}_{\{C^2(y)=\text{blue}\}}}{k} + \frac{\mathbb{1}_{\{C^2(y)=\text{red}\}}}{\rho_t} \frac{\rho_t}{k} \quad (2.25)$$

$$= \frac{1}{k}. \quad (2.26)$$

□

We then choose an energy level  $j$  uniformly in  $\{1; \dots; m\}$  and we write  $\xi(\omega_t^1) = (k_1^1, \dots, k_m^1)$  and  $\xi(\omega_t^2) = (k_1^2, \dots, k_m^2)$ .

- If  $C^1(x^1) = \text{blue}$ : then  $C^2(x^2) = \text{blue}$ , and  $x^1, x^2$  are in the same energy level  $i := \omega_t^1(x^1) = \omega_t^2(x^2)$ . Let  $a \neq b \in \{1; 2\}$  such that  $\psi_i(k_i^a - 1) - \psi_j(k_j^a) \leq \psi_i(k_i^b - 1) - \psi_j(k_j^b)$ . Then,

$$p_a := \exp\{-[\psi_i(k_i^a - 1) - \psi_j(k_j^a)]^+\} \quad (2.27)$$

$$\geq \exp\{-[\psi_i(k_i^b - 1) - \psi_j(k_j^b)]^+\} =: p_b. \quad (2.28)$$

Let  $U$  be a uniform random variable on  $[0; 1)$ .

- If  $U < p_b$ : then we set  $\omega_{t+1}^a(x) = \omega_t^a(x)$  for all  $x \neq x^a$ ,  $\omega_{t+1}^a(x^a) = j$ ,  $\omega_{t+1}^b(x) = \omega_t^b(x)$  for all  $x \neq x^b$  and  $\omega_{t+1}^b(x^b) = j$ . Then  $\xi(\omega_{t+1}^a) = (\xi(\omega_t^a))^{ij}$  and  $\xi(\omega_{t+1}^b) = (\xi(\omega_t^b))^{ij}$ , and for any red-blue coloring of  $(\omega_{t+1}^1, \omega_{t+1}^2)$  the number of red particles  $\rho_{t+1}$  remains the same (since both particles  $x^1$  and  $x^2$  have moved together).
  - If  $p_b \leq U < p_a$ : then we set  $\omega_{t+1}^a(x) = \omega_t^a(x)$  for all  $x \neq x^a$ ,  $\omega_{t+1}^a(x^a) = j$  and  $\omega_{t+1}^b(x) = \omega_t^b(x)$  for all  $x$ . Then  $\xi(\omega_{t+1}^a) = (\xi(\omega_t^a))^{ij}$  and  $\xi(\omega_{t+1}^b) = \xi(\omega_t^b)$ . The only situation in which the number of red particles could increase is the following:  $k_j^a \geq k_j^b$  and  $k_i^a \leq k_i^b$ . Since  $\psi_i$  and  $\psi_j$  are non-increasing, this would imply  $p_a \leq p_b$  that contradicts  $p_b \leq U < p_a$ . Then,  $\rho_{t+1} \leq \rho_t$  for any red-blue coloring of  $(\omega_{t+1}^1, \omega_{t+1}^2)$ .
  - If  $U \geq p_a$ : then we set  $\omega_{t+1}^a = \omega_t^a$  and  $\omega_{t+1}^b = \omega_t^b$ . Then  $\xi(\omega_{t+1}^a) = \xi(\omega_t^a)$ ,  $\xi(\omega_{t+1}^b) = \xi(\omega_t^b)$  and  $\rho_{t+1} = \rho_t$ .
  - If  $C^1(x^1) = \text{red}$ : then  $C^2(x^2) = \text{red}$ , and we define  $i^1 := \omega_t^1(x^1)$  and  $i^2 := \omega_t^2(x^2)$ . Note that according to (2.19),  $i^1 \neq i^2$ . Let then  $V^1, V^2$  be two independent uniform random variables on  $[0; 1)$ .
    - If  $V^1 < \exp\{-[\psi_{i^1}(k_{i^1}^1 - 1) - \psi_j(k_j^1)]^+\}$  then we set  $\omega_{t+1}^1(x) = \omega_t^1(x)$  for all  $x \neq x^1$  and  $\omega_{t+1}^1(x^1) = j$  and then  $\xi(\omega_{t+1}^1) = (\xi(\omega_t^1))^{ij}$ ; otherwise we leave  $\omega_{t+1}^1 = \omega_t^1$ .
    - If  $V^2 < \exp\{-[\psi_{i^2}(k_{i^2}^2 - 1) - \psi_j(k_j^2)]^+\}$  then we set  $\omega_{t+1}^2(x) = \omega_t^2(x)$  for all  $x \neq x^2$  and  $\omega_{t+1}^2(x^2) = j$  and then  $\xi(\omega_{t+1}^2) = (\xi(\omega_t^2))^{ij}$ ; otherwise we leave  $\omega_{t+1}^2 = \omega_t^2$ .
- Whether red particles move or not, the number of red particles cannot increase, so it is clear that  $\rho_{t+1} \leq \rho_t$ .

We conclude

**Proposition 2.3.2.**  $(\rho_t)_{t \in \mathbb{N}}$  is a non-increasing process.

and claim

**Proposition 2.3.3.** The process  $(\xi(\omega^1), \xi(\omega^2))$  is a coupling for the Markov chain with transition matrix  $p$ .

**Proof.** Writing  $\xi(\omega_t^1) = \eta$  and given  $i, j \in \{1; \dots; m\}$ ,

$$\begin{aligned} P(\xi(\omega_{t+1}^1) = \eta^{ij}) &= P(\xi(\omega_{t+1}^1) = \eta^{ij} | C^1(x^1) = \text{blue}) P(C^1(x^1) = \text{blue}) \\ &\quad + P(\xi(\omega_{t+1}^1) = \eta^{ij} | C^1(x^1) = \text{red}) P(C^1(x^1) = \text{red}). \end{aligned} \quad (2.29)$$

We have

$$P(C^1(x^1) = \text{red}) = 1 - P(C^1(x^1) = \text{blue}) = \frac{\rho_t}{k} \quad (2.30)$$

and

$$\begin{aligned} P(\xi(\omega_{t+1}^1) = \eta^{ij} | C^1(x^1) = \text{blue}) \\ = P(\omega_t^1(x^1) = i | C^1(x^1) = \text{blue}) \times \frac{1}{m} \times P(U < p_1) \end{aligned} \tag{2.31}$$

$$= \frac{k_i^1 \wedge k_i^2}{(k - \rho_t)m} \exp\{-[\psi_i(k_i^1 - 1) - \psi_j(k_j^1)]^+\}, \tag{2.32}$$

$$\begin{aligned} P(\xi(\omega_{t+1}^1) = \eta^{ij} | C^1(x^1) = \text{red}) \\ = P(\omega_t^1(x^1) = i | C^1(x^1) = \text{red}) \times \frac{1}{m} \times P(V^1 < p_1) \end{aligned} \tag{2.33}$$

$$= \frac{[k_i^1 - k_i^2]^+}{\rho_t m} \exp\{-[\psi_i(k_i^1 - 1) - \psi_j(k_j^1)]^+\} \tag{2.34}$$

which finally leads to

$$P(\xi(\omega_{t+1}^1) = \eta^{ij}) = \frac{k_i^1}{km} \exp\{-[\psi_i(k_i^1 - 1) - \psi_j(k_j^1)]^+\} = p(\eta, \eta^{ij}). \tag{2.35}$$

Then,  $P(\xi(\omega_{t+1}^1) = \eta^{ij} | \omega_t^1, \omega_t^2)$  depends on  $\xi(\omega_t^1)$  only, which means that  $\xi(\omega^1)$  is a Markov chain. Besides, according to (2.35) its transition matrix is  $p$ . Finally, by Lemma 2.3.1, the same is true for  $\xi(\omega_2)$ .  $\square$

From now on, we will write  $X_t^1 := \xi(\omega_t^1)$  and  $X_t^2 := \xi(\omega_t^2)$ . The previous proposition ensures that  $(X^1, X^2)$  is a coupling for the dynamics with transition matrix  $p$ .

#### 2.4. Estimating the coupling time

We will use Proposition 2.2.1. Since  $(\rho_t)_{t \in \mathbb{N}}$  is non-increasing and  $\rho_t = 0$  if and only if  $X_t^1 = X_t^2$ , all we have to do is to estimate from below the probability of the event  $\{\rho_{t+1} < \rho_t\}$ .

**Proposition 2.4.1.** *If at step  $t$  of the coupled dynamics  $(\omega^1, \omega^2)$  we assume that red particles have been chosen, i.e.,  $C^1(x^1) = \text{red} = C^2(x^2)$ , then there is a choice of  $j \in \{i^1; i^2\}$  for which the number of red particles decreases with probability 1, where  $i^1$  (resp.  $i^2$ ) still refers to  $\omega_t^1(x^1)$  (resp.  $\omega_t^2(x^2)$ ).*

**Proof.** If the inequalities

$$\exp\{-[\psi_{i^1}(k_{i^1}^1 - 1) - \psi_{i^2}(k_{i^2}^1)]^+\} < 1, \tag{2.36}$$

$$\exp\{-[\psi_{i^2}(k_{i^2}^2 - 1) - \psi_{i^1}(k_{i^1}^2)]^+\} < 1 \tag{2.37}$$

hold together, then  $\psi_{i^1}(k_{i^1}^1 - 1) > \psi_{i^2}(k_{i^2}^1)$  and  $\psi_{i^2}(k_{i^2}^2 - 1) > \psi_{i^1}(k_{i^1}^2)$ . Besides, since  $C^1(x^1) = \text{red}$ , according to (2.19)  $k_{i^1}^1 > k_{i^1}^2$  from which we get  $k_{i^1}^1 - 1 \geq k_{i^1}^2$  and, since  $\psi_{i^1}$  is non-increasing,  $\psi_{i^1}(k_{i^1}^1 - 1) \leq \psi_{i^1}(k_{i^1}^2)$ . Likewise, since  $C^2(x^2) = \text{red}$  we have  $\psi_{i^2}(k_{i^2}^2 - 1) \leq \psi_{i^2}(k_{i^2}^1)$ . We finally may write

$$\psi_{i^1}(k_{i^1}^1 - 1) > \psi_{i^2}(k_{i^2}^1) \geq \psi_{i^2}(k_{i^2}^2 - 1) > \psi_{i^1}(k_{i^1}^2) \geq \psi_{i^1}(k_{i^1}^1 - 1) \tag{2.38}$$

which is absurd. As a result, either (2.36) or (2.37) is false. For instance let us assume that (2.36) is false, then if  $j = i^2$ , with probability 1 we have  $\omega_{t+1}^1(x) = \omega_t^1(x)$  for all  $x \neq x^1$ ,  $\omega_{t+1}^1(x^1) = i^2 \neq \omega_t^1(x^1)$  and  $\omega_{t+1}^2 = \omega_t^2$ . Then, the number of red particles for any red–blue coloring of  $(\omega_{t+1}^1, \omega_{t+1}^2)$  is exactly  $\rho_{t+1} = \rho_t - 1$ .  $\square$

It follows

**Corollary 2.4.2.** *At step  $t$ , if  $\rho_t > 0$ , the probability for  $\rho_{t+1}$  to be  $\rho_t - 1$  is at least  $\rho_t/km$ .*

Consequently, and owing to the fact that  $\rho_t$  cannot increase, we have the inequality  $E(\rho_{t+1} - \rho_t | X_t^1, X_t^2) \leq -P(\rho_{t+1} = \rho_t - 1 | X_t^1, X_t^2)$  from which we deduce

$$E(\rho_{t+1} | X_t^1, X_t^2) \leq \left(1 - \frac{1}{km}\right) \rho_t. \tag{2.39}$$

Then we can apply Proposition 2.2.1 to  $\rho(X_t^1, X_t^2) = \rho_t$  with  $M = k$  and  $\alpha = km$ , which finally proves Theorem 1.

**Remark.** Adding “moderate disorder hypotheses” we can gain a lot, depending on the specific model we consider. For example in the case of our simple exclusion dynamics, i.e., for the Markov chain associated with the Fermi statistics when all the  $n_j$ ’s are equal to 1, we can gain a factor of order  $m$ . We can indeed assume that  $k/m \leq 1/2$  (if not, we consider a dynamics on the vacancies rather than on the particles as mentioned in Section 1.4) and, if, for example,  $\nu$  derives from an homogeneous product measure, i.e.,  $\psi_i = \psi_j$  for all  $i$  and  $j$ , then there are much more than one choice for  $j$  for which the number of red particles will decrease with probability 1 when red particles were chosen: any of the more than  $m/2$  vacant sites in  $X_t^1$  or  $X_t^2$  will do the job. In this case we gain a factor  $m/2$ , first in Corollary 2.4.2, then in the final estimate. More generally, if the functions  $|\psi_i - \psi_j|$  are uniformly bounded (this is a moderate disorder hypothesis), then, when red particles are chosen together with one of these more than  $m/2$  vacant sites, the probability that the number of red particles decrease is bounded away from zero: once again we gain a factor of order  $m$ .

### 3. Proof of Theorem 2

We now work with the dynamics defined by (1.35)–(1.36) and we assume  $\delta > 0$ . First, the reversibility of this dynamics with respect to  $\nu$  still holds, with exactly the same computation as in Section 2.1. However, it is no longer possible to work with an underlying process  $\omega_t \in \Omega$  since the factor  $k_i^\delta$  cannot stand for a number of particles as soon as  $\delta < 1$ . Therefore we need to adapt the coupling  $(X^1, X^2)$  directly on  $\mathcal{X}_{k,m}$ .

#### 3.1. Generalizing the previous coupling

At step  $t$ , let us assume  $X_t^1 = (k_1^1, \dots, k_m^1) \in \mathcal{X}_{k,m}$  and  $X_t^2 = (k_1^2, \dots, k_m^2) \in \mathcal{X}_{k,m}$ . We define the following sets:

$$R_1 := \{i \in \{1; \dots; m\}: k_i^1 > k_i^2\}, \tag{3.1}$$

$$R_2 := \{i \in \{1; \dots; m\}: k_i^2 > k_i^1\}, \tag{3.2}$$

$$B := \{i \in \{1; \dots; m\}: k_i^1 = k_i^2\} \tag{3.3}$$

and the following quantities:

$$w_1 := \sum_{i \in R_1} (k_i^1)^\delta, \quad w'_1 := \sum_{i \in R_2} (k_i^1)^\delta, \tag{3.4}$$

$$w_2 := \sum_{i \in R_2} (k_i^2)^\delta, \quad w'_2 := \sum_{i \in R_1} (k_i^2)^\delta, \tag{3.5}$$

$$w_B := \sum_{i \in B} (k_i^1)^\delta = \sum_{i \in B} (k_i^2)^\delta. \tag{3.6}$$

Finally we define

$$\rho_t := \frac{1}{2} \sum_{i=1}^m |k_i^1 - k_i^2| = \sum_{i=1}^m [k_i^1 - k_i^2]^+ = \sum_{i=1}^m [k_i^2 - k_i^1]^+. \tag{3.7}$$

Keeping in mind the previous coloring,  $R_1$  (resp.  $R_2$ ) is the set of sites in which there are red particles for the first (resp. the second) configuration,  $B$  is the set of sites in which there are only blue particles or no particles for both configurations,  $w_1$  (resp.  $w_2$ ) is proportional to the probability to choose a site for the first (resp. the second) configuration in which there are red particles,  $w'_1$  (resp.  $w'_2$ ) is proportional to the probability to choose a site for the first (resp. the second) configuration in which there are only blue particles while there are red particles in the second (resp. the first) configuration,  $w_B$  is proportional to the probability to choose a site in which there are only blue particles for both configurations, and  $l_\delta - (w_B + w_1 + w'_1)$  (resp.  $l_\delta - (w_B + w_2 + w'_2)$ ) is proportional to the probability not to choose any site for the first (resp. the second) configuration, and it can be positive as soon as  $\delta < 1$ . Finally,  $\rho_t$  still stands for the number of red particles and it is clear that  $\rho_t = 0$  if and only if  $X_t^1 = X_t^2$ .

**N.B.** In the remaining part of this subsection, we assume  $w_1 + w'_1 \geq w_2 + w'_2$  in order to not overload the notations and not increase the number of cases to investigate. Obviously, the case  $w_1 + w'_1 \leq w_2 + w'_2$  is exactly symmetric.

Let  $I$  be a uniform random variable on  $[0; l_\delta)$ .

(i) If  $I < w_B$ : then there exists a unique  $i \in B$  such that

$$\sum_{i' \in B; i' < i} (k_{i'}^1)^\delta \leq I < \sum_{i' \in B; i' \leq i} (k_{i'}^1)^\delta \tag{3.8}$$

and we set  $i^1 = i^2 = i$ .

**Remark.** We then have, for all  $i \in \{1; \dots; m\}$ ,

$$P(i^1 = i | (i)) = \mathbb{1}_{\{i \in B\}} \frac{(k_i^1)^\delta}{w_B}, \tag{3.9}$$

$$P(i^2 = i | (i)) = \mathbb{1}_{\{i \in B\}} \frac{(k_i^1)^\delta}{w_B} = \mathbb{1}_{\{i \in B\}} \frac{(k_i^2)^\delta}{w_B}. \tag{3.10}$$

(ii) If  $w_B \leq I < w_B + w'_1$ : then there exists a unique  $i \in R_2$  such that

$$\sum_{i' \in R_2; i' < i} (k_{i'}^1)^\delta \leq I - w_B < \sum_{i' \in R_2; i' \leq i} (k_{i'}^1)^\delta \tag{3.11}$$

and we set  $i^1 = i^2 = i$ .

**Remark.** We then have, for all  $i \in \{1; \dots; m\}$ ,

$$P(i^1 = i | (ii)) = \mathbb{1}_{\{i \in R_2\}} \frac{(k_i^1)^\delta}{w'_1}, \tag{3.12}$$

$$P(i^2 = i | (ii)) = \mathbb{1}_{\{i \in R_2\}} \frac{(k_i^1)^\delta}{w'_1}. \tag{3.13}$$

(iii) If  $w_B + w'_1 \leq I < w_B + w'_1 + w_1$ : then there exists a unique  $i \in R_1$  such that

$$\sum_{i' \in R_1; i' < i} (k_{i'}^1)^\delta \leq I - w_B - w'_1 < \sum_{i' \in R_1; i' \leq i} (k_{i'}^1)^\delta. \tag{3.14}$$

We set  $i^1 = i$  and we define  $u := I - w_B - w'_1 - \sum_{i' \in R_1; i' < i} (k_{i'}^1)^\delta$ , so that  $0 \leq u < (k_i^1)^\delta$ . Notice that since  $i \in R_1$ ,  $k_i^1 > k_i^2$ . If  $u < (k_i^2)^\delta$  then we set  $i^2 = i$ . Otherwise, for all  $i' \in R_2$  we write  $\bar{k}_{i'} := \sum_{\ell \in R_2; \ell < i'} (k_\ell^1)^\delta$  and we denote by  $T$  the disjoint union of intervals

$$T := \left\{ \bigcup_{i' \in R_2} [\bar{k}_{i'} + (k_{i'}^1)^\delta; \bar{k}_{i'} + (k_{i'}^2)^\delta] \right\} \cup [w_2 + w'_2; w_1 + w'_1]. \quad (3.15)$$

Let  $I'$  be a uniform random variable on  $T$ . If there exists  $i' \in R_2$  such that  $I' \in [\bar{k}_{i'} + (k_{i'}^1)^\delta; \bar{k}_{i'} + (k_{i'}^2)^\delta]$  then we set  $i^2 = i'$ . Else we do not define  $i^2$ .

**Remark.** We then have, for all  $i \in \{1; \dots; m\}$ ,

$$P(i^1 = i | \text{(iii)}) = \mathbb{1}_{\{i \in R_1\}} \frac{(k_i^1)^\delta}{w_1} \quad (3.16)$$

and for all  $i' \in \{1; \dots; m\}$ ,

$$\begin{aligned} P(i^2 = i' | \text{(iii)}) &= \mathbb{1}_{\{i' \in R_1\}} \frac{(k_{i'}^2)^\delta}{w_1} \\ &+ \mathbb{1}_{\{i' \in R_2\}} \sum_{i \in R_1} \frac{(k_i^1)^\delta - (k_i^2)^\delta}{w_1} \times \frac{(k_{i'}^2)^\delta - (k_{i'}^1)^\delta}{\lambda(T)}, \end{aligned} \quad (3.17)$$

where

$$\lambda(T) := \sum_{i' \in R_2} ((k_{i'}^2)^\delta - (k_{i'}^1)^\delta) + (w_1 + w'_1) - (w_2 + w'_2) \quad (3.18)$$

$$= w_1 - w'_2 \quad (3.19)$$

so that

$$P(i^2 = i' | \text{(iii)}) = \mathbb{1}_{\{i' \in R_1\}} \frac{(k_{i'}^2)^\delta}{w_1} + \mathbb{1}_{\{i' \in R_2\}} \frac{(k_{i'}^2)^\delta - (k_{i'}^1)^\delta}{w_1}. \quad (3.20)$$

(iv) If  $I \geq w_B + w_1 + w'_1$  (this case cannot occur when  $\delta = 1$ ): then we do not define  $i^1$  and  $i^2$ .

Before going ahead with the definition of our coupling we note, as a direct consequence of our remarks in (i), (ii), (iii) and of the fact that  $I$  has a uniform distribution:

**Proposition 3.1.1.** For all  $i \in \{1; \dots; m\}$ ,  $P(i^1 = i | X_t^1, X_t^2) = (k_i^1)^\delta / l_\delta$  and  $P(i^2 = i | X_t^1, X_t^2) = (k_i^2)^\delta / l_\delta$ .

We then choose an integer  $j \in \{1; \dots; m\}$  with uniform law and we distinguish once again between our four previous cases.

(i) If  $i^1 \in B$ : then  $i^2 = i^1$ , we just write  $i^1 = i^2 = i$ . Then  $k_i^1 = k_i^2$ . Thus, let  $a, b \in \{1; 2\}$  such that  $a \neq b$  and  $k_j^a \leq k_j^b$ . Since both  $\psi_i^{[\delta 1]}$  and  $\psi_j^{[\delta 1]}$  are non-increasing,

$$p_a := \exp\{-[\psi_i^{[\delta 1]}(k_i^a - 1) - \psi_j^{[\delta 1]}(k_j^a)]^+\} \quad (3.21)$$

$$\geq \exp\{-[\psi_i^{[\delta 1]}(k_i^b - 1) - \psi_j^{[\delta 1]}(k_j^b)]^+\} =: p_b. \quad (3.22)$$

Let  $U$  be a uniform random variable on  $[0; 1)$ .

– If  $U < p_b$ : we set  $X_{t+1}^a = (X_t^a)^{ij}$  and  $X_{t+1}^b = (X_t^b)^{ij}$ .

- If  $p_b \leq U < p_a$ : we set  $X_{t+1}^a = (X_t^a)^{ij}$  and  $X_{t+1}^b = X_t^b$ .
- If  $p_a \leq U$ : we set  $X_{t+1}^a = X_t^a$  and  $X_{t+1}^b = X_t^b$ .

In any of these cases, we certainly have  $\rho_{t+1} = \rho_t$ .

(ii) If  $i^1 \in R_2$ : then  $i^2 = i^1$ , we just write  $i^1 = i^2 = i$ . Let  $a, b \in \{1; 2\}$  such that  $a \neq b$  and  $\psi_i^{[\delta]}(k_i^a - 1) - \psi_j^{[\delta]}(k_j^a) \leq \psi_i^{[\delta]}(k_i^b - 1) - \psi_j^{[\delta]}(k_j^b)$ , so that

$$p_a := \exp\{-[\psi_i^{[\delta]}(k_i^a - 1) - \psi_j^{[\delta]}(k_j^a)]^+\} \quad (3.23)$$

$$\geq \exp\{-[\psi_i^{[\delta]}(k_i^b - 1) - \psi_j^{[\delta]}(k_j^b)]^+\} =: p_b. \quad (3.24)$$

Let  $U$  be a uniform random variable on  $[0; 1)$ .

- If  $U < p_b$ : we set  $X_{t+1}^a = (X_t^a)^{ij}$  and  $X_{t+1}^b = (X_t^b)^{ij}$ .
- If  $p_b \leq U < p_a$ : we set  $X_{t+1}^a = (X_t^a)^{ij}$  and  $X_{t+1}^b = X_t^b$ .
- If  $p_a \leq U$ : we set  $X_{t+1}^a = X_t^a$  and  $X_{t+1}^b = X_t^b$ .

In the last case we obviously have  $\rho_{t+1} = \rho_t$ . In the first case the particles move together and  $\rho_{t+1} = \rho_t$ . In the second case the number of red particles could increase only if  $k_i^a \leq k_i^b$  and  $k_j^a \geq k_j^b$ , but, since  $\psi_i^{[\delta]}$  and  $\psi_j^{[\delta]}$  are non-increasing, this would contradict  $p_a > p_b$ . As a consequence we have  $\rho_{t+1} \leq \rho_t$  in all the three cases.

(iii) If  $i^1 \in R_1$ : there are three cases for  $i^2$ . Either  $i^2 = i^1 = i$  and this case is the symmetric of (ii). Or  $i^2$  is randomly chosen in  $R_2$ , and we define

$$p_1 := \exp\{-[\psi_{i^1}^{[\delta]}(k_{i^1}^1 - 1) - \psi_j^{[\delta]}(k_j^1)]^+\}, \quad (3.25)$$

$$p_2 := \exp\{-[\psi_{i^2}^{[\delta]}(k_{i^2}^2 - 1) - \psi_j^{[\delta]}(k_j^2)]^+\}. \quad (3.26)$$

Or else  $i^2$  is not defined, and we set  $p_2 := 0$  still defining  $p_1$  by (3.25). Let then  $V^1, V^2$  be independent uniform random variables on  $[0; 1)$ .

- If  $V^1 < p_1$ : we set  $X_{t+1}^1 = (X_t^1)^{i^1 j}$ , else we set  $X_{t+1}^1 = X_t^1$ .
- If  $V^2 < p_2$ : we set  $X_{t+1}^2 = (X_t^2)^{i^2 j}$ , else we set  $X_{t+1}^2 = X_t^2$ .

In the first case we have  $\rho_{t+1} \leq \rho_t$  as previously. In the last two cases we also have  $\rho_{t+1} \leq \rho_t$  since only particles from  $R_1$  in the first configuration and from  $R_2$  in the second configuration can move.

(iv) If  $i_1$  and  $i_2$  are not defined: then we simply set  $(X_{t+1}^1, X_{t+1}^2) = (X_t^1, X_t^2)$  and we have  $\rho_{t+1} = \rho_t$ .

In any of the previous cases, once  $i^1, i^2$  and  $j$  have been defined, the probability for  $X_{t+1}^1$  (resp.  $X_{t+1}^2$ ) to be  $(X_t^1)^{i^1 j}$  (resp.  $(X_t^2)^{i^2 j}$ ) is  $\exp\{-[\psi_{i^1}^{[\delta]}(k_{i^1}^1 - 1) - \psi_j^{[\delta]}(k_j^1)]^+\}$  (resp.  $\exp\{-[\psi_{i^2}^{[\delta]}(k_{i^2}^2 - 1) - \psi_j^{[\delta]}(k_j^2)]^+\}$ ). Thus, according to Proposition 3.1.1, the fact that  $j$  is uniformly chosen in  $\{1; \dots; m\}$  and our study on the variation of  $\rho$  we conclude:

**Proposition 3.1.2.** *The process  $(X^1, X^2)$  is a coupling for the dynamics defined by (1.35)–(1.36) and such that  $(\rho_t)_{t \in \mathbb{N}}$  is non-increasing.*

### 3.2. Estimating the coupling time

Similarly to the proof of Theorem 1 we will use Proposition 2.2.1: since  $(\rho_t)_{t \in \mathbb{N}}$  is non-increasing it will be enough to give a lower bound for the probability of  $\{\rho_{t+1} < \rho_t\}$ .

**Proposition 3.2.1.** *If at step  $t$  of the coupled dynamics  $(X^1, X^2)$ , we assume that “red particles have been chosen,” i.e.,  $i^1 \in R_1$  and  $i^2 \in R_2$ , then, there is a choice of  $j \in \{i^1; i^2\}$  for which the number of red particles decreases with probability 1.*

**Proof.** Assuming  $i^1 \in R_1$  and  $i^2 \in R_2$  yields  $k_{i^1}^1 > k_{i^1}^2$  and  $k_{i^2}^1 < k_{i^2}^2$ . Using exactly the same argument as for Proposition 2.4.1 we prove that either  $\exp\{-[\psi_{i^1}^{[\delta]}(k_{i^1}^1 - 1) - \psi_{i^2}^{[\delta]}(k_{i^2}^1)]^+\} = 1$  or  $\exp\{-[\psi_{i^2}^{[\delta]}(k_{i^2}^2 - 1) - \psi_{i^1}^{[\delta]}(k_{i^1}^2)]^+\} = 1$ . Eventually, if one red particle in some configuration moves to a site with a red particle in the other configuration, then both particles turn blue and the number of red particles decreases by one.  $\square$

**Corollary 3.2.2.** *At step  $t$ , the probability for  $\rho_{t+1}$  to be  $\rho_t - 1$  is at least  $\delta k^{\delta-1} \rho_t / ml_\delta$ .*

**Proof.** The probability to choose  $i^1 \in R_1$  and  $i^2 \in R_2$  is

$$P(i^1 \in R_1, i^2 \in R_2) = \sum_{i' \in R_2} \sum_{i \in R_1} P(i^2 = i' | i^1 = i) \times P(i^1 = i) \quad (3.27)$$

$$= \sum_{i' \in R_2} \sum_{i \in R_1} \frac{(k_i^1)^\delta - (k_i^2)^\delta}{(k_i^1)^\delta} \frac{(k_{i'}^2)^\delta - (k_{i'}^1)^\delta}{\lambda(T)} \times \frac{(k_i^1)^\delta}{l_\delta} \quad (3.28)$$

$$= \frac{1}{l_\delta} \sum_{i' \in R_2} ((k_{i'}^2)^\delta - (k_{i'}^1)^\delta). \quad (3.29)$$

Since, for any concave function  $f: \mathbb{R}_+ \rightarrow \mathbb{R}$  and any  $s \in \mathbb{N} \setminus \{0\}$ ,  $(z_1, \dots, z_s) \in \mathbb{R}_+^s \mapsto f(\sum_i z_i) - \sum_i f(z_i)$  is non-increasing in all its  $s$  variables (as a consequence of the slope inequalities), by concavity of  $z \mapsto z^\delta$  and using the fact that, for all  $i' \in R_2$ ,  $k_{i'}^2 > k_{i'}^1$  we get

$$P(i^1 \in R_1, i^2 \in R_2) \geq \frac{1}{l_\delta} \left[ \left( \sum_{i' \in R_2} k_{i'}^2 \right)^\delta - \left( \sum_{i' \in R_2} k_{i'}^1 \right)^\delta \right] \quad (3.30)$$

$$= \frac{1}{l_\delta} \left[ \left( \rho_t + \sum_{i' \in R_2} k_{i'}^1 \right)^\delta - \left( \sum_{i' \in R_2} k_{i'}^1 \right)^\delta \right]. \quad (3.31)$$

Using the same property of concave functions on  $\mathbb{R}^+$  (with  $s = 2$ ) and the fact that  $\sum_{i' \in R_2} k_{i'}^1 \leq k - \rho_t$ , then using once again the concavity of  $z \mapsto z^\delta$ , we write

$$P(i^1 \in R_1, i^2 \in R_2) \geq \frac{1}{l_\delta} [k^\delta - (k - \rho_t)^\delta] \geq \frac{\delta k^{\delta-1} \rho_t}{l_\delta} \quad (3.32)$$

so that, by the previous proposition,

$$P(\rho_{t+1} = \rho_t - 1 | X_t^1, X_t^2) \geq \frac{\delta k^{\delta-1} \rho_t}{ml_\delta}. \quad (3.33)$$

$\square$

As a consequence

$$E(\rho_{t+1} | X_t^1, X_t^2) \leq \left( 1 - \frac{\delta}{km(k \wedge m)^{1-\delta}} \right) \rho_t \quad (3.34)$$

and, by Proposition 2.2.1, this finally proves Theorem 2.

## Simulation appendix

In this appendix we report on simulations for the temporal evolution of  $d(t)$  defined in (1.10) as well as the dependency in  $\beta$  of the mixing time  $t_\varepsilon$  defined in (1.9) with  $\varepsilon = 0.1$  in the context of the Fermi statistics  $\nu$  associated with  $k = 50$  particles among  $m = 20$  energy levels with two different profiles.

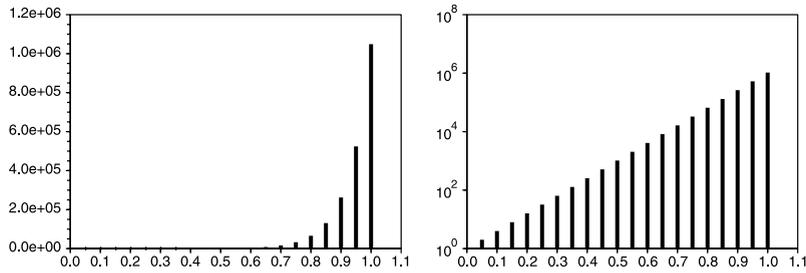


Fig. 1. Degeneracies of the energy levels for case 1, in semi-logarithmic representation for the second picture.

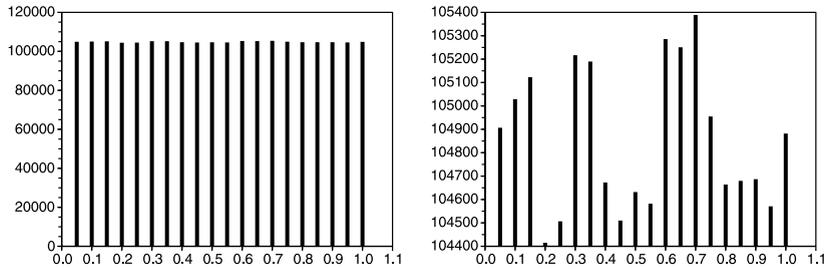


Fig. 2. Degeneracies of the energy levels for case 2, with shifted and rescaled axis for the second picture.

- In the first case we chose  $v_j = j/20$  and  $n_j = 2^j$  for  $1 \leq j \leq m$  so that  $n = \sum_j 2^j = 2^{21} - 2$ . See Fig. 1.
- In the second case we chose the same  $v_j$ 's, the same volume  $n = 2^{21} - 2$  and we chose the degeneracies according to a multinomial law with parameters  $n, 1/20, \dots, 1/20$ . See Fig. 2.

We were then interested in

$$d(t) = \max_{\eta \in \mathcal{X}_{k,m}} \|p^t(\eta_0, \cdot) - \nu\|_{\text{TV}} = \max_{\eta \in \mathcal{X}_{k,m}} \frac{1}{2} \sum_{\eta \in \mathcal{X}_{k,m}} |p^t(\eta_0, \eta) - \nu(\eta)| \tag{A.1}$$

for  $t \geq 0$ . The cardinality of  $\mathcal{X}_{k,m}$  being very large (equal to  $\binom{69}{19} \simeq 5 \times 10^{16}$  in case 2), one faces three difficulties with such a formula:

- One cannot compute  $\nu(\eta)$  and  $p^t(\eta_0, \eta)$  for given  $\eta$  and  $\eta_0$ .
- The sum in (A.1) contained too many terms to be computed.
- It is not possible to try all the possible  $\eta_0$  before taking the maximum in (A.1).

We addressed the first difficulty by replacing  $p^t(\eta_0, \cdot)$  by the empirical measure

$$\mu_t^N := \frac{1}{N} \sum_{l=1}^N \delta_{X_t^l} \tag{A.2}$$

computed with a large number  $N$  of simulated independent copies  $(X_t^l)_{t \in \mathbb{N}}$  of our Markov chain starting from  $\eta_0$  and, using Theorem 1, by replacing  $\nu$  by

$$\nu^N := \frac{1}{2T} \sum_{t=T+1}^{2T} \mu_t'^N + \mu_t''^N \tag{A.3}$$

with  $T := \lfloor km \ln(k/\varepsilon) \rfloor$  and where  $\mu_t'^N$  and  $\mu_t''^N$  are independently computed like  $\mu_t^N$ , starting from two different configurations. We addressed the second difficulty by looking at a coarse grained version  $\tilde{\mathcal{X}}_{k,m}$  of  $\mathcal{X}_{k,m}$  through the free entropy  $\phi$ . More precisely, with  $I^N = [\phi_-, \phi_+]$  the smallest interval containing the support of  $\phi(\nu^N) = \nu^N \circ \phi^{-1}$

we divided  $I^N$  into  $M$  intervals of length  $\Delta = (\phi_+ - \phi_-)/M$ , we extended this partition of  $I^N$  into a partition of  $\mathbb{R}$  in intervals of length  $\Delta$  and grouped in a same class the configurations  $\eta \in \mathcal{X}_{k,m}$  that fall into a same interval. (For possible cases where  $\phi_- = \phi_+$ , we used a partition of  $\mathbb{R}$  into intervals of length  $10^{-6}$ .) With  $\bar{\phi}$  the canonical projection from  $\mathcal{X}_{k,m}$  to  $\bar{\mathcal{X}}_{k,m}$  we could then compute  $\|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV}$  rather than  $\|\mu_t^N - v^N\|_{TV}$ . Note that

$$\|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV} \leq \|\mu_t^N - v^N\|_{TV}, \tag{A.4}$$

and we also have, for large  $M$  and as soon as  $\phi$  is an injection from  $\mathcal{X}_{k,m}$  to  $\mathbb{R}$ ,

$$\|\mu_t^N - v^N\|_{TV} = \|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV} + o(1). \tag{A.5}$$

As far as the third difficulty was concerned we underestimated

$$\bar{d}^N(t) = \max_{\eta_0 \in \mathcal{X}_{k,m}} \|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV} \tag{A.6}$$

by making a guess on a particular configuration  $\eta_0$  for which  $\|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV}$  could be of the order of  $\bar{d}^N(t)$ . We simply took an  $\eta_0$  for which we could have say, for all  $\beta$ , that it was “very far from typical configurations.” We first took  $\eta_0 = (0, \dots, 0, k)$  since typical configurations are concentrated on the low energy levels when  $\beta$  is large and have their occupation numbers  $k_j$  distributed like the degeneracies  $n_j$  when  $\beta$  is small. Since this last point tends to show that in case 1 our specific  $\eta_0$  is “not so far” from typical equilibrium configurations (we will come back on this point later), we also considered a different  $\eta_0$ , that one for which the high energy levels are empty and the low energy levels are saturated or as close to saturation as possible:

$$\eta_0 = \begin{cases} (2, 4, 8, 16, 20, 0, \dots, 0) & \text{in case 1,} \\ (50, 0, \dots, 0) & \text{in case 2.} \end{cases} \tag{A.7}$$

Then we simply took the maximum of the two quantities  $\|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV}$  computed, in each case, with these two specific choices.

Summing up, we approximated  $d(t)$  by

$$\bar{d}_0^N(t) = \max_{\eta_0 \in \mathcal{X}_{k,m}^0} \|\bar{\phi}(\mu_t^N) - \bar{\phi}(v^N)\|_{TV}, \tag{A.8}$$

where  $\mathcal{X}_{k,m}^0$  is the two configuration set described above. We chose  $N = 1024$  and  $M = \sqrt{N} = 32$  and we plotted for different temperatures a graphical representation of the law of  $\bar{\phi}(v^N)$  and the temporal evolution of  $\bar{d}_0^N(t)$  for  $0 \leq t \leq 2T$ . See Figs 3 to 7 for case 1 and Figs 8 to 12 for case 2.

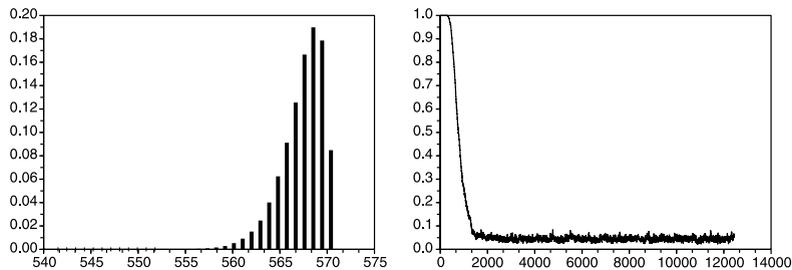


Fig. 3. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 1 and  $\beta = 0$ .

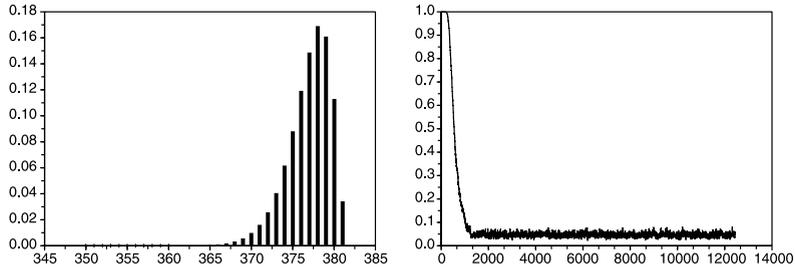


Fig. 4. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 1 and  $\beta = 4$ .

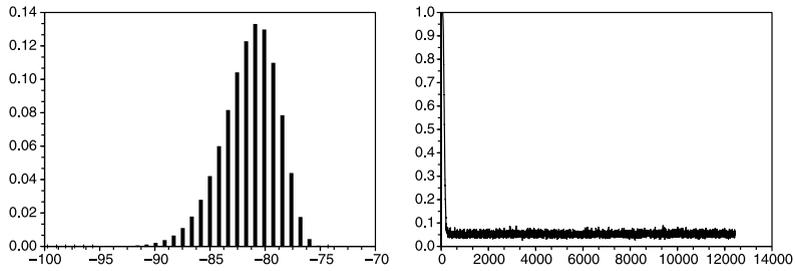


Fig. 5. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 1 and  $\beta = 16$ .

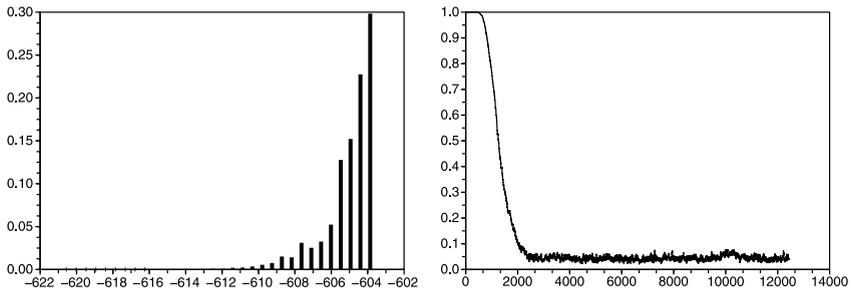


Fig. 6. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 1 and  $\beta = 63$ .

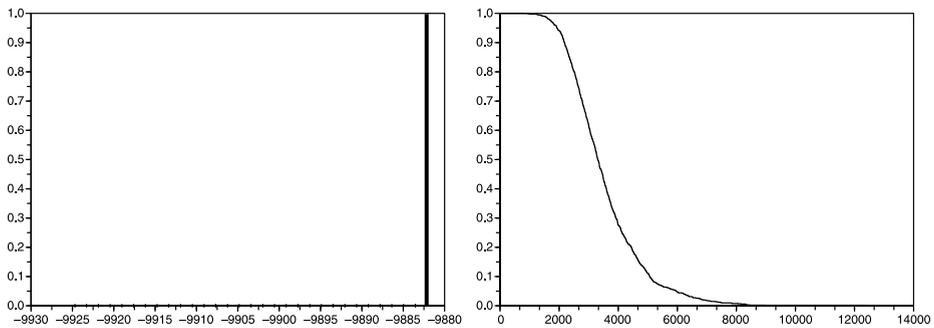


Fig. 7. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 1 and  $\beta = 1000$ .

For each temperature we get in each case an estimation  $\hat{t}_\varepsilon$  of  $t_\varepsilon$ . Estimating  $t_\varepsilon$  in this way for 51 temperatures ( $\beta = 0$  included) we show in Fig. 13 the graphical representations of the simulated dependency of  $\hat{t}_\varepsilon$  on  $\beta > 0$  together with our upper bound from Theorem 1.

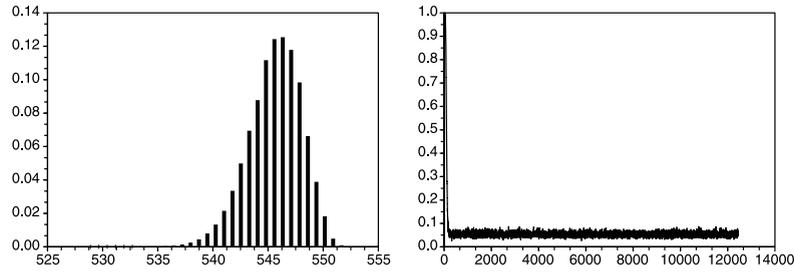


Fig. 8. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 2 and  $\beta = 0$ .

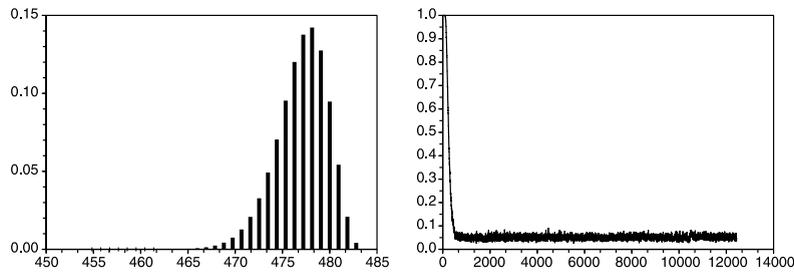


Fig. 9. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 2 and  $\beta = 4$ .

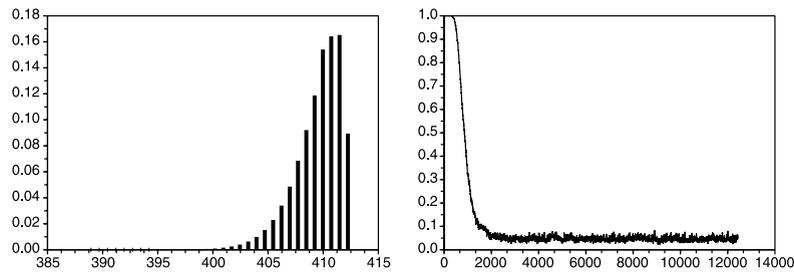


Fig. 10. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 2 and  $\beta = 16$ .

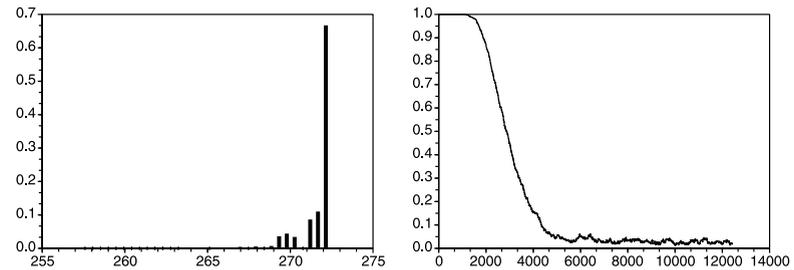


Fig. 11. Probability distribution function of  $\bar{\phi}(v^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 2 and  $\beta = 63$ .

We can now conclude with a few comments. First, with such a procedure,  $t_\varepsilon$  is essentially underestimated. Indeed, except for the replacement of  $p^t(\eta_0, \cdot)$  and  $v$  by  $\mu_t^N$  and  $v^N$  our successive approximations underestimated  $d(t)$  then  $t_\varepsilon$ . As a consequence we recover with case 2 that  $km \ln(k/\varepsilon)$  is the best upper bound on  $t_\varepsilon$  that is uniform over the disorder.

Second, we note that  $\beta$  can be seen as a tuning parameter for the disorder. We recover that our theoretical estimate on  $t_\varepsilon$  can be improved by a factor of order  $m$  in weak disorder situations.

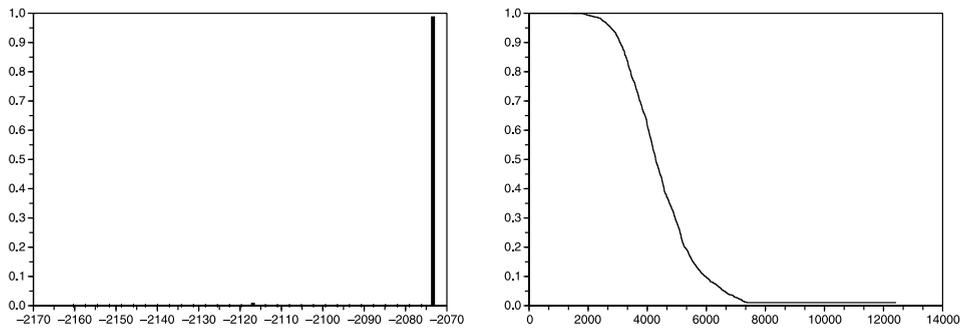


Fig. 12. Probability distribution function of  $\bar{\phi}(\nu^N)$  and temporal evolution of  $\bar{d}_0^N(t)$  for case 2 and  $\beta = 1000$ .

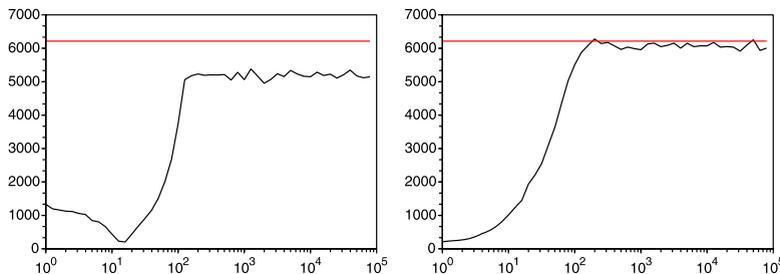


Fig. 13. Dependency of  $\hat{t}_\varepsilon$  on  $\beta$  in case 1 and 2 respectively (semi-logarithmic representation). The horizontal line stands for our theoretical upper bound.

Last, it is worth to note that case 1 suggests that  $t_\varepsilon$  is not generally decreasing with the temperature. In our preliminary simulations we only considered  $\eta_0 = (0, \dots, 0, 50)$  to estimate the mixing time. Then we feared that our non-monotonic estimation could be an artefact of our simulation. This led us to introduce the alternative saturated low level starting configuration, confirming in that way the non-monotonicity.

## Acknowledgment

J.R. is very grateful to everyone who made his stay in Rome possible and pleasant.

## References

- [1] C. Ané, S. Blachère, D. Chafaï, P. Fougères, I. Gentil, F. Malrieu, C. Roberto and G. Scheffer. *Sur les Inégalités de Sobolev Logarithmiques. Panoramas et Synthèses* **10**. Soc. Math. France, Paris, 2000. [MR1845806](#)
- [2] D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de Probabilités, XIX, 1983/84* 177–206. *Lecture Notes in Math.* **1123**. Springer, Berlin, 1985. [MR0889476](#)
- [3] A. S. Boudou, P. Caputo, P. Dai Pra and G. Posta. Spectral gap inequalities for interacting particle systems via a Bochner type inequality. *J. Funct. Anal.* **232** (2006) 222–258. [MR2200172](#)
- [4] P. Caputo. Spectral gap inequalities in product spaces with conservation laws. In *Stochastic Analysis on Large Scale Interacting Systems* 53–88. T. Funaki and H. Osada (Eds). *Adv. Stud. Pure Math.* **39**. Math. Soc. Japan, Tokyo, 2004. [MR2073330](#)
- [5] P. Caputo. On the spectral gap of the Kac walk and other binary collision processes. *ALEA Lat. Am. J. Probab. Math. Stat.* **4** (2008) 205–222. [MR2429910](#)
- [6] P. Caputo, P. Dai Pra and G. Posta. Convex entropy decay via the Bochner–Bakry–Emery approach. *Ann. Inst. H. Poincaré Probab. Statist.* **45** (2009) 734–753. Available at [arXiv:0712.2578](#). [MR2548501](#)
- [7] A. Iovanela, B. Scoppola and E. Scoppola. Some spin glass ideas applied to the clique problem. *J. Stat. Phys.* **126** (2007) 895–915. [MR2311889](#)
- [8] O. Johnson. Bounds on the Poincaré constant of ultra log-concave random variables. Preprint, Univ. Bristol, 2008. Available at [arXiv:0801.2112](#).
- [9] A. Joulin and Y. Ollivier. Curvature, concentration, and error estimates for Markov chain Monte Carlo. *Ann. Probab.* **38** (2010) 2418–2442. Available at [arXiv:0904.1312](#). [MR2683634](#)

- [10] C. Landim and J. Noronha Neto. Poincaré and logarithmic Sobolev inequality for Ginzburg–Landau processes in random environment. *Probab. Theory Related Fields* **131** (2005) 229–260. [MR2117953](#)
- [11] D. A. Levin, Y. Peres and E. L. Wilmer. *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI, 2008. [MR2466937](#)
- [12] T. M. Liggett. Ultra logconcave sequences and negative dependence. *J. Combin. Theory Ser. A* **79** (1997) 315–325. [MR1462561](#)
- [13] R. Montenegro and P. Tetali. *Mathematical Aspects of Mixing Times in Markov Chains. Foundations and Trends in Theoretical Computer Science* **1:3**. M. Sudan (Ed.). Now Publishers, Boston–Delft, 2006. [MR2341319](#)
- [14] R. Pemantle. Towards a theory of negative dependence, *J. Math. Phys.* **41** (1997) 1371–1390. [MR1757964](#)
- [15] L. Saloff-Coste. Lectures on finite Markov chains. In *Lectures on Probability Theory and Statistics, Saint-Flour 1996* 301–413. *Lecture Notes in Math.* **1665**. Springer, Berlin, 1997. [MR1490046](#)