Journal of Statistical Physics, Vol. 125, No. 1, October 2006 (© 2006)
DOI: 10.1007/s10955-006-9175-y

Rigorous Inequalities Between Length and Time Scales in Glassy Systems

Andrea Montanari[1] and Guilhem Semerjian[2]

Received March 1, 2006; accepted July 9, 2006
Published Online: October 4, 2006

Glassy systems are characterized by an extremely sluggish dynamics without any simple sign of long range order. It is a debated question whether a correct description of such phenomenon requires the emergence of a large correlation length. We prove rigorous bounds between length and time scales implying the growth of a properly defined length when the relaxation time increases. Our results are valid in a rather general setting, which covers finite-dimensional and mean field systems.

As an illustration, we discuss the Glauber (heat bath) dynamics of p-spin glass models on random regular graphs. We present the first proof that a model of this type undergoes a purely dynamical phase transition not accompanied by any thermodynamic singularity.

KEY WORDS: glass transition, Glauber dynamics, correlation time, correlation length

1. INTRODUCTION

A broad class of liquids show a dramatic slowing down of their dynamics as the temperature is lowered (or the density increased) without any simple sign of long range order setting in Ref. 1. In particular, the static structure factor $S(\vec{k}) = \langle \rho_{\vec{k}} \rho_{-\vec{k}} \rangle$ is hardly distinguishable from the one of a liquid, even if the system has become a solid from a dynamical point of view. We shall generically refer to systems displaying analogous phenomena as 'glassy.'

The above features are reproduced in mode-coupling theory, as well as in some mean-field models, where the slowing down is promoted to a real *dynamical phase transition* without any simple *static* signature[2] In both these theories, only

[1] Laboratoire de Physique Théorique de I'Ecole Normale Supérieure (UMR 8549, Unité Mixte de Recherche du CNRS et de 1' ENS); e-mail: montanar@lpt.ens.fr
[2] Dipartimento di Fisica, CNR-INFM (UdR and SMC centre), Università di Roma "La Sapienza."

short range correlations are taken into account, thus leading to the idea that they are the only responsible for the slow dynamics.

Physical commonsense suggests however that, in systems with finite range interactions, only cooperative effects on large length scales may lead to a large relaxation time. This expectation has been recently substantiated by introducing a 'dynamical' correlation length in terms of a (dynamical) four points correlation function[3–7] (see Refs. 8, 9 for recent developments). In this paper we confirm this line of thought through a rigorous argument. We define a time scale τ, and length scale ℓ appropriate for glassy systems and prove that they satisfy the inequalities

$$C_1\ell \le \tau \le \exp\{C_2\ell^d\}, \tag{1}$$

where $C_{1,2}$ are two constants to be specified below, and d is the system spatial dimension. This implies, in particular, that glassy systems are characterized by a large length scale, increasing as the temperature is lowered. The solution to the conundrum is that ℓ is defined in terms of *point to set* instead of *point to point* correlations (which are probed, for instance, in scattering experiments). It is worth mentioning that the definition of ℓ is purely statical (it does not depend on the dynamics, as long as this satisfies a few conditions), and that it is closely related to the *Gedanken* experiment discussed in Ref. 10.

The physical intuition behind (1) is elementary. The lower bound follows from the observation that in order for the system to relax, information must be spread through at least one correlation length, and this cannot happen quicker than ballistically. For the upper bound, one argues that, without much harm, the system can be cut into boxes of size ℓ, and that, within each box, the relaxation time cannot be larger than exponential in the volume.

Both these arguments apply to a much larger family of models than the ones defined on d-dimensional lattices. Here we consider systems on general factor graphs with bounded degree (see Sec. 2.1 for a definition). In this case, the ℓ^d factor in the upper bound of Eq. (1) must be replaced by the volume of a ball of radius ℓ in the graph. There are two motivations for considering such generalizations. On one hand, sparse random graphs allow to define mean field theory, while retaining some features (a bounded number of neighbors and a locality structure) of finite dimensional models. On the other hand, they appear in a variety of random combinatorial and constraint satisfaction problems, ranging from coloring[11] to K-satisfiability.[12] In this context, one recurrent question is whether configurations (or solutions) can be sampled efficiently using the Markov Chain Monte Carlo method. Understanding the relaxation time scale for Glauber dynamics is relevant for this question. In this context, point to set correlations and their relation to relaxation times were first considered in Ref. 13 and subsequently studied in Ref. 14. In the following pages, we shall analyze one prototypical example of such system: the p-spin model (also known as XOR-SAT) on random regular graphs.

The relation between correlations in space and time (or spatial and temporal mixing) is indeed a well established subject in probability and mathematical physics, with many beautiful results (see for instance[15,16] and references therein). There are several reasons for presenting a new variation on this classical theme: (i) Most of the mathematical literature deals with translation invariant systems on finite dimensional lattices. (ii) The focus there is on 'global' characterizations of the space and time correlations, such as the spectral gap of the Markov generator, or the log-Sobolev constant. These are hardly accessible to experiments or numerical simulations, and in fact are not considered in the physics literature. (iii) The dichotomy between vanishing and not-vanishing gap (or log-Sobolev constant) which is crucial there, is not that important for glasses. In many such systems, the gap vanishes well above the glass transition due to Griffiths singularities, or to the existence of metastable states that are not relevant when the dynamics is initiated with a random initial condition. (iv) The spatial mixing hypothesis used in mathematics requires decay of correlations uniform over the boundary conditions. This is probably too restrictive, especially in models defined on sparse graphs, for which the boundary of a domain can scale like its volume.

The treatment presented here is essentially self contained, and based on elementary combinatorial arguments (in part inspired by Ref. 17. In Sec. 2 we provide our definitions of length and time scales and state precisely our main results. In the same Section we discuss qualitatively possible generalizations. As an illustration of the main results we consider in Sec. 4 the p-spin glass model on random graphs. In order to carry on our analysis, we prove the first rigorous bounds on the relaxation time of a model in this class. In particular, these imply the occurrence of a purely dynamical phase transition as the temperature is lowered. Section 3 contains a few alternative definitions of length scale and a discussion of their equivalence. The proofs of the main results are contained in Secs. 5 and 6, with the most technical parts relegated in a series of Appendices.

2. DEFINITIONS AND MAIN RESULT

2.1. General Graphical Models: Equilibrium Distribution

Factor graphs[18] are a convenient language to describe a large class of statistical mechanics models. A factor graph $G \equiv (V, F, E)$ is a bipartite graph with two types of vertices: variable nodes (also called sites in the following) $V \ni i, j, k, \ldots$ and function nodes $F \ni a, b, c, \ldots$ Edges are ordered pairs (i, a), with $i \in V$ and $a \in F$. The number of variable nodes in G will be denoted by $N \equiv [V]$, and we shall identify $V = \{1, \ldots, N\}$. Given $i \in V$ (respectively $a \in F$), its *neighborhood* $\partial i(\partial a)$ is defined as the set of function nodes a (variable nodes i) such that $(i, a) \in E$. We assume that the graph has *bounded degree*, ie. that $|\partial i|, |\partial a| \leq \Delta$ for some $\Delta > 0$.

The *distance* d_{ij} between two nodes $i, j \in V$ is the length (number of function nodes encountered along the path) of the shortest path joining i to j. Given a non-negative integer r and a node $i \in V$, the *ball* of radius r around i, $\mathbf{B}_i(r)$, is the subset of variable nodes j with $d_{ij} \leq r$. With a slight abuse of notation, $\mathbf{B}_i(r)$ will sometimes be the sub-graph induced by these nodes. Finally, given $U \subseteq V$, we let $\partial U \subseteq V \backslash U$ be the subset of variable nodes at unit distance from U.

We deal with models with discrete variables, taking values in the finite set $\mathcal{X}$. A *configuration* is a vector $x = (x_1, \ldots, x_N), x_i \in \mathcal{X}$. For $A \subseteq V$, we write $x_A \equiv \{x_i : i \in A\}$. In order to lighten the notation we shall write $x_{\sim i,r}$ instead of $x_{V \backslash \mathbf{B}_i(r)}$ for the configuration *outside* the ball of radius r around i.

A probability distribution over configurations is defined by introducing one *compatibility function* ψ_{node} for each node in G. These are non-negative functions, with $\psi_a : \mathcal{X}^{\partial a} \to \mathbb{R}$ at function nodes and $\psi_i : \mathcal{X} \to \mathbb{R}$ at variable nodes. We define

$$\mu(x) = \frac{1}{Z} \prod_{a \in F} \psi_a(x_{\partial a}) \prod_{i \in V} \psi_i(x_i). \tag{2}$$

Compatibility functions can also be used to define conditional probabilities. Let $U \subseteq V$ and $F(U) \subseteq F$ be the subset of function nodes having at least one neighbor in U. If $y = (y_1, \ldots, y_N)$ is a configuration, we define

$$\mu_U^y(x) = \frac{1}{Z_U^y} \prod_{a \in F(U)} \psi_a(x_{\partial a}) \prod_{i \in U} \psi_i(x_i), \tag{3}$$

if x coincides with y on $U^c \equiv V \backslash U$ (i.e. $x_{U^c} = y_{U^c}$) and $\mu_U^y(x) = 0$ otherwise. The normalization constant Z_U^y ensures that $\sum_x \mu_U^y(x) = 1$.

Since we allow for vanishing compatibility functions (hard core interactions), the above expressions could be a *priori* ill defined. In order to avoid this, and to simplify our treatment, we shall restrict ourselves to *permissive* interactions (the definition we give here is very close to the one of Ref. 17. This means that for each site i there exists $x_i^* \in \mathcal{X}$ such that, given any non-empty $U \subseteq V$, if $x_i = x_i^*$ for each $i \in U$, then the right hand side of Eq. (3) is strictly positive regardless of the assignment of x on ∂U. In particular, this must be the case when the subset is a single vertex $U = \{i\}$. We denote by $\mu_0 > 0$ a lower bound on the conditional probability for the corresponding state to be x_i^*. In other words, we ask that $\mu_i^y(x_i^*) \geq \mu_0$ for all i and y.

2.2. Glauber Dynamics

The dynamics is specified as a single spin flip, continuous-time Markov process, irreducible, aperiodic and satisfying detailed balance with respect to the distribution (2). More precisely, for each variable node i, a set of transition rates

$\kappa_i^x(\xi) \geq 0$, with $\xi \in \mathcal{X}$ and $\sum_\xi \kappa_i^x(\xi) = 1$ is specified. Each variable node $i \in V$ is associated to a clock, whose ringing times are distributed according to independent rate-one Poisson processes. When the clock at site i rings, a new value $\xi \in \mathcal{X}$ is drawn from the distribution $\kappa_i^x(\xi)$, x being the current configuration. The new configuration x' coincides everywhere with x but on i, where $x'_i = \xi$. In order to verify detailed balance, the following condition must be satisfied by the transition rates:

$$\mu(x)\kappa_i^x(x'_i) = \mu(x')\kappa_i^{x'}(x_i). \tag{4}$$

for any two configurations x and x' that coincide everywhere but on i.

In our treatment we shall make two assumptions on the transition rates.

- *Locality*. The transition rates $\kappa_i^x(\cdot)$ depend on the current configuration x only through x_j, with $d_{ij} \leq 1$. Although its precise form could be modified (and somewhat relaxed), this is a crucial physical requirement: the Markov dynamics must be local with respect to the underlying graph G.
- *Permissivity*. Let x_i^* the state permitted at node i regardless of the configuration at $V\backslash\{i\}$. Then, there exists $\kappa_0 > 0$ such that $\kappa_i^x(x_i^*) \geq \kappa_0$ independently of i and x. This assumption can somewhat be relaxed at the expense of some technical difficulties.

A well known example of transition rates satisfying these conditions is given by the so-called 'heat-bath rule':

$$\kappa_i^x(\xi) = \mu_i^x(\xi). \tag{5}$$

It will be also useful to define the Markov dynamics on a subset of the vertices $U \subseteq V$, with boundary condition $y \in \mathcal{X}^V$. By this we mean that the initial configuration agrees with y on $V\backslash U$, and that variables outside U are 'frozen': each time the clock rings at a site $i \in V\backslash U$, the configuration is left unchanged.

We shall sometimes use the notation $\langle\cdot\rangle$ for averages with respect to the equilibrium distribution μ or the Markov process defined above, with initial condition distributed according to μ. We also denote by $\langle\cdot\rangle_U^y$ averages with respect to the subset in U with boundary y (i.e. either with respect to the distribution μ_U^y or with respect to the Markov process on U with initial condition distributed according to μ_U^y).

2.3. Length and Time Scales and Their Relation

Both our definitions of length and time scales depend on a parameter ε. This should be thought of as some fixed small number (let's say 0.01), that provides a cut-off for distinguishing 'highly correlated' from 'weakly correlated' degrees of freedom. The physical idea is that, near a glass transition, the order of magnitude of the resulting length and time scales should be roughly independent of ε as long as

this is smaller than a characteristic value (the 'Edwards-Anderson' or 'ergodicity breaking' parameter).

Consider a vertex i in G. Let $f(x_i)$ be a function of the variable at i, and $F(x_{\sim i,r})$ a function of the variables whose distance from i are larger than r. The covariance $\langle f(x_i)F(x_{\sim i,r})\rangle - \langle f(x_i)\rangle\langle F(x_{\sim i,r})\rangle$ measures the degree of correlation between the observables f and F. In order to quantify the degree of correlation of x_i and $x_{\sim i,r}$, it makes sense to consider the 'most correlated' observables and define

$$G_i(r) \equiv \sup_{f,F} \left| \langle f(x_i)F(x_{\sim i,r})\rangle - \langle f(x_i)\rangle\langle F(x_{\sim i,r})\rangle \right|, \tag{6}$$

where the sup is taken over all functions such that $|f(x_i)|, |F(x_{\sim i,r})| \le 1$ for any x. The correlation length $\ell_i(\varepsilon)$ of vertex i is defined as the smallest integer ℓ such that $G_i(r) \le \varepsilon$ for all $r \ge \ell$. In formulae

$$\ell_i(\varepsilon) \equiv \min\{\ell \ge 0 \qquad \text{s.t.} \quad G_i(r) \le \varepsilon \forall r \ge \ell\}. \tag{7}$$

If no such ℓ exists, $\ell_i(\varepsilon)$ is set by convention to the maximum distance from i of a vertex in G.

The time scale definition is completely analogous to the above one. We let

$$C_i(t) = \sup_f |\langle f(x_i(0))f(x_i(t))\rangle - \langle f(x_i(0))\rangle\langle f(x_i(t))\rangle|, \tag{8}$$

the sup being taken over functions of the variable at i such that $|f(x_i)| \le 1$ for any x_i. Then we let

$$\tau_i(\varepsilon) = \inf\{\tau \ge 0 \quad \text{s.t.} \quad C_i(t) \le \varepsilon \forall t \ge \tau\}. \tag{9}$$

This expression is well defined, and the resulting $\tau_i(\varepsilon)$ is always finite. In fact, if we define $C_i^f(t) \equiv \langle f(x_i(0))f(x_i(t))\rangle - \langle f(x_i(o))\rangle\langle f(x_i(t))\rangle$, the spectral representation of the transition probabilities,[19] implies $C_i^f(t) = \sum_{l=2}^n e^{-\lambda_l t} B_l(i, f)$. Here $0 < \lambda_2 \le \lambda_3 \le \ldots$ are the eigenvalues of the Markov generator, the B_l are non-negative coefficients, and n is the number of configurations of the system. In consequence this correlation function is positive, decreasing and has a vanishing limit when $t \to \infty$.

Both definitions given here admit several essentially equivalent variants which can be helpful depending on the circumstances; we shall discuss some of them in Sec. 3. We can now state our main result, whose proof can be found in Secs. 5 and 6.

Theorem 1. *Under the hypothesis presented in this Section*

$$C_1\ell_i(|\mathcal{X}|\sqrt{2\varepsilon}) \le \tau_i(\varepsilon) \le 1 + A\exp\{C_2|\mathrm{B}_i(\ell_i(\varepsilon/2))|\}, \tag{10}$$

where $A = \log(\frac{4}{\varepsilon})$, $C_1 = 1/2e\Delta^2$, $C_2 = -\log(\kappa_0(1-e^{-1}))$, *and the lower bound holds under the assumption* $\ell_i(|\mathcal{X}|\sqrt{2\varepsilon}) > \log_2(2/\varepsilon)$.

Let us stress that Ref. 13 (Theorem 1.5) proves a very general result that is closely related to the above lower bound. Despite the fact that both the statements and the proofs are quite similar, we think that our formulation can be better suited for physics applications. The authors of Ref. 13 provide in fact a bound on the spectral gap, a quantity that is hardly accessible in experiments or simulations. Further it does not correspond to an interesting time scale when studying the glass transition. In fact, the relaxation time defined as the inverse spectral gap becomes exponential in the system size well above the dynamical glass transition (roughly speaking, at the appearence of the first metastable states). This can be shown, for instance, in the example of Sec. 4.

2.4. Problems and Generalizations

Let us briefly discuss Theorem 1 and a few related research directions. As shown in Sec. 4, this theorem can be only marginally improved in the general setting described above. It would therefore be interesting to determine additional conditions under which the upper/lower bounds are in fact closer to the actual correlation time.

Recently there has been a considerable interest in a 'dynamical' length scale ξ_4 defined through four point correlation functions.[9] It would be interesting to understand whether a general relation holds between ξ_4, and the length ℓ defined in Eq. (7). Within mean field models undergoing a discontinuous glass transition, it has been argued that they are in fact closely related.[20,21]

The notion of 'growing length scale' is useful also when the initial condition for the Markov dynamics is not drawn from the equilibrium measure[3] $\mu(x)$. Of particular interest is the case of a uniformly random initial condition $x \in \mathcal{X}^N$ (a 'quench from infinite temperature'). The definition (7) can be generalized to this setting if the expectations in Eq. (6) are taken with respect to the measure at some time $t \geq 0$. The resulting length will depend on time: $\ell_i(\varepsilon, t)$, and is expected to increase with t, starting from $\ell_i(\varepsilon, 0) = 0$ until it reaches the equilibrium value $\ell_i(\varepsilon, \infty) = \ell_i(\varepsilon)$. The disagreement percolation technique used to prove the lower bound in Theorem 1 can be generalized to this case, cf. Sec. 5, to show that $\ell_i(\varepsilon, t) \leq \tilde{C}t$.

Finally, it would be interesting to consider more realistic models for glasses, e.g. off-lattice particle systems with Langevin dynamics. We guess that similar

[3] The question was posed to us by Giulio Biroli.

ideas to the ones proposed here can be useful in that case, although at the expenses of several technical difficulties.

3. ALTERNATIVE DEFINITIONS OF LENGTH SCALE

The basic physical idea in the definition of length scale is to look at correlations between a point (a vertex) and the whole set of variables at distance larger than r from it. Equation (6) provides one measure of these correlations. Here we shall define four alternative measures $G_i^{(n)}(r)$, $n = 1, \dots, 4$. For any such measure, we can define a length scale exactly as in Sec. 2.3,

$$\ell_i^{(n)}(\varepsilon) \equiv \min\{\ell \geq 0 \quad \text{s.t.} \quad G_i^{(n)}(r) \leq \varepsilon \forall r \geq \ell\}. \tag{11}$$

The question which we shall address shortly is to what extent these definitions are equivalent.

The first two definitions express the idea that two random variables are weakly correlated if their joint distribution is (approximately) factorized. With an abuse of notation we denote by $\mu(x_i, x_{\sim i,r})$ the joint distribution of x_i and $x_{\sim i,r}$ and by $\mu(x_i)$, $\mu(x_{\sim i,r})$ their marginal distributions. Then we define

$$G_i^{(1)}(r) \equiv \sup_{x_i, x_i'} \sum_{x_{\sim i,r}} \left| \mu(x_{\sim i,r}|x_i) - \mu(x_{\sim i,r}|x_i') \right|, \tag{12}$$

$$G_i^{(2)}(r) \equiv \sum_{x_i, x_{\sim i,r}} \left| \mu(x_i, x_{\sim i,r}) - \mu(x_i)\mu(x_{\sim i,r}) \right|. \tag{13}$$

One inconvenient of the definition of $G_i(r)$, cf. Eq. (6), as well as of $G_i^{(1)}(r)$ and $G_i^{(2)}(r)$ is that they are difficult to evaluate. Equation (6), for instance, requires the optimization with respect to F which is, in general, a function of $\Theta(N)$ variables. Given a function f of the variable at i, we let $\widehat{f}$ be the function of $x_{\sim i,r}$ obtained by taking the expectation of f with respect to the conditional distribution corresponding to the boundary condition $x_{\sim i,r}$ outside $\mathbf{B}_i(r)$. In formulae: $\widehat{f}(x_{\sim i,r}) = \langle f(x_i) \rangle^x_{\mathbf{B_i(r)}}$.
We then define

$$G_i^{(3)}(r) = \sup_f \left| \langle f(x_i) \widehat{f}(x_{\sim i,r}) \rangle - \langle f(x_i) \rangle \langle \widehat{f}(x_{\sim i,r}) \rangle \right|, \tag{14}$$

where, again, the sup is taken over the functions f such that $|f(x_i)| \leq 1$ for any x. Note that one can always decompose an arbitrary function $f(x_i)$ in terms of the $|\mathcal{X}|$ indicator functions $f_\xi(x_i) = \mathbb{I}(x_i = \xi)$, hence $G_i^{(3)}(r)$ can be computed with a finite number of covariance estimations. The correlation function $G_i^{(3)}(r)$ was already used in Ref. 14 to bound the spectral gap of Ising and hard core models on trees.

A suggestive interpretation of the last definition is provided by the following procedure. Generate a reference configuration x according to the distribution μ, and evaluate $f(x_i)$ on it. Then 'freeze' everything is outside $\mathbf{B}_i(r)$ and generate a new configuration x_i' inside, according to the conditional distribution $\mu^x_{\mathbf{B_i(r)}}$, (i.e. with the boundary condition given by the frozen variables). Evaluate $f(x_i')$ on the new configuration. The correlation function $G_i^{(3)}(r)$ is (the sup over f of) the covariance between $f(x_i)$ and $f(x_i')$. This procedure was indeed discussed in Ref. 10.

A last definition consists in considering the mutual information[22] between x_i and $x_{\sim i,r}$:

$$G_i^{(4)}(r) \equiv I(X_i; X_{\sim i,r}). \tag{15}$$

Recall that, given two discrete random variables X and Y with distribution $p(x, y)$, their mutual information is defined as $I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)}$. A pleasing interpretation follows from the general principles of information theory. Suppose that a configuration x is generated according to the distribution μ, but only $x_{\sim i,r}$ is revealed to you. In a very precise sense, $G_i^{(4)}(r)$ measures how much information you would have about x_i.

It turns out that the length scales extracted from $G^{(1)}(r), \ldots, G_i^{(4)}(r)$ convey essentially the same information as $\ell_i(\varepsilon)$. More precisely, they differ only by a redefinition of the parameter ε and a rescaling.

Proposition 1. *Let $\ell_i^{(n)}(\varepsilon), n = 1, \ldots, 4$ be defined as in Eq. (11), and $\mu_* \equiv \min_{x_i \in \mathcal{X}} \mu(x_i)$. Then*

$$\ell_i^{(2)}(|\mathcal{X}|\varepsilon) \le \ell_i(\varepsilon) \le \ell_i^{(2)}(\varepsilon), \tag{16}$$

$$\ell_i^{(2)}(\varepsilon) \le \ell_i^{(1)}(\varepsilon) \le \ell_i^{(2)}\left(\mu_* \frac{\varepsilon}{2}\right), \tag{17}$$

$$\ell_i^{(2)}(|\mathcal{X}|\sqrt{\varepsilon}) \le \ell_i^{(3)}(\varepsilon) \le \ell_i(\varepsilon), \tag{18}$$

$$\ell_i^{(2)}(\sqrt{2\varepsilon}) \le \ell_i^{(4)}(\varepsilon) \le \ell_i^{(2)}\left(\frac{\varepsilon\mu_*}{1-\mu_*}\right). \tag{19}$$

The proof of this proposition can be found in Appendix A.

Notice that all the changes of argument in the above expressions amount to a finite rescaling in ε, apart from the upper bounds in Eqs. (17) and (19). If μ_* is bounded away from zero, also these are finite rescalings and the length scales $\ell_i(\varepsilon), \ell_i^{(1)}(\varepsilon), \ldots, \ell_i^{(4)}(\varepsilon)$ are all equivalent. In the proof of Theorem 1 we shall only use the upper bound in Eqs. (16) and (17).

4. A MEAN FIELD EXAMPLE: THE *P*-SPIN MODEL ON RANDOM (HYPER) GRAPHS

In this Section we discuss a particular example exhibiting glassy behavior. Our objective is twofold. First of all, we want to check to what extent Theorem 1 can be improved over. Second, we want to show that the idea of diverging correlation can be of relevance even for mean field systems, as soon as the underlying factor graph is sparse.

We consider a model of N Ising spins $\sigma = (\sigma_1, \dots, \sigma_N)$, $\sigma_i \in \{+1, -1\}$ (for historical reasons, we use here σ_i, instead of x_i to denote the i-th variable), defined through the Boltzmann distribution

$$\mu(\sigma) = \frac{1}{Z(\beta)} e^{-\beta E(\sigma)}, \tag{20}$$

$$E(\sigma) = -\sum_{a=1}^{M} J_a \prod_{i \in \partial a} \sigma_i \equiv -\sum_{a=1}^{M} J_a \sigma_{i_1(a)} \cdots \sigma_{i_p(a)}. \tag{21}$$

Here we think of $i \in \{i, \dots, N\}$ as the variable nodes of a factor graph G, and of $a \in \{1, \dots, M\}$ as its function nodes. Further, we assume G to be a random factor graph with degree l at variable nodes, and p at function nodes.[4] A random factor graph from this ensemble will also be referred to as a 'random l-regular hypergraph' (function nodes being identified with hyperedges joining the neighboring variable nodes). Finally, $\beta = 1/T$ is the inverse temperature and the J_a's are i.i.d. random variables, uniform in $\{+1, -1\}$.

In the following it will be always understood that $p, l \geq 3$. It is in fact expected that for $p = 2$ the model undergoes a spin glass transition without a dynamical phase transition, and for $l = 2$, no phase transition at all occurs at finite temperature.

As the variables σ_i can only take two opposite values, the general definitions given in Sec. 2.3 simplify somewhat. For instance

$$C_i(t) = \langle \sigma_i(0)\sigma_i(t) \rangle - \langle \sigma_i \rangle^2, \tag{22}$$

$$G_i^{(3)}(r) = \left| \langle \sigma_i \langle \sigma_i \rangle_{\mathrm{B}_i(r)}^{\sigma} \rangle - \langle \sigma_i \rangle^2 \right|. \tag{23}$$

The phase diagram of this model is expected to be characterized by two phase transitions[23]: a dynamical phase transition at temperature T_{d}, and a statical one at $T_{\mathrm{c}} < T_{\mathrm{d}}$ (see also Refs. 20, 21 for a detailed study of the dynamics in the case

[4] More precisely G is distributed according to the corresponding *configuration model*. To sample a factor graph from this ensemble, each of the N variable nodes is attributed l sockets, and each of the M function nodes, p sockets. The $Nl = Mp$ sockets on the two sides are then matched according to a uniformly random permutation over Nl elements. Multiple edges are removed, if they occur an even number of times, and replaced by a single edge, in the opposite case.

of uniformly drawn random hypergraphs). At high temperature $T > T_{\rm d}$, Glauber dynamics is fast and the relaxation time τ does not depend on the system size N. Note that, in the thermodynamic limit, any finite neighborhood of any spin is a regular hypertree with high probability. Therefore as far as the τ_i's are finite, they converge in probability to a deterministic value τ.

Below $T_{\rm d}$, the τ_i's become instead exponentially large in N. Heuristic estimates on the exponential rate can be expressed in terms of free-energy barriers, computed from the so-called quenched potential.[24,25] In Ref. 26, it was shown that the spectral gap is exponentially small[5] at low enough temperature for some values of p and l.

Remarkably, the free energy remains analytic in the interval $(T_{\rm c}, \infty)$, and a true thermodynamic phase transition takes place only at $T_{\rm c}$. If $p \geq l$, $T_{\rm c} = 0$, and there is no thermodynamic phase transition.

Assuming that this picture is correct (below we shall partially confirm it), Theorem 1 implies that the correlation lengths ℓ_i are finite for $T > T_{\rm d}$. On the other hand, for $T < T_{\rm d}$, the ℓ_i's are necessarily divergent in the system size. Since the size of a ball of radius r is bounded as $|\mathsf{B}_i(r)| \leq l(p-1)^r(l-1)^r$, a moment of thought shows that $\ell \geq C \log N$. On the other hand, ℓ cannot be larger than the graph diameter, whence $\ell = \Theta(\log N)$.

In the high temperature phase $T > T_{\rm d}$, the correlation function $G^{(3)}(r)$ can be computed recursively, using a non-rigorous approach that exploits the locally tree-like structure of the factor graph. Let us sketch the procedure here and refer to Refs. 21, 27 for a detailed description. We consider a rooted tree factor graph with R generations $\mathsf{T}_*(R)$ defined as follows. For $R = 0$, $\mathsf{T}_*(R)$ is a single node (the root). For any $R \geq 0$, one first defines $\mathsf{T}_*(R)'$ by joining $(l-1)$ copies of $\mathsf{T}_*(R)$ at the root, and then joins $(p-1)$ copies of $\mathsf{T}_*(R)'$ to a common function node a of degree p to obtain $\mathsf{T}_*(R+1)$. The remaining variable node adjacent to a is the new root. One then generate a configuration σ on this graph according to the Boltzmann weight (20), 'freezes' it from generation r on, and consider the conditional $\mu_{i,r}(\cdot) = \mu(\sigma_i = \cdot | \sigma_{\sim i,r})$, i being the root node. When $\sigma_{\sim i,r}$ is generated randomly according to the above procedure, $\mu_{i,r}(\cdot)$ can be considered as a random variable. Knowing its distribution allows to compute $G^{(3)}(r)$ (it turns out that this distribution does not depend upon R). Thanks to the tree structure, a recursive distributional equation can be written for $\mu_{i,r}(\cdot)$, and solved numerically using a sampling ('population dynamics') technique.

In Fig. 1, left frame, we plot the results of such a computation for $p = l = 3$ and a few values of T. As the temperature is lowered towards $T_{\rm d}$, $G^{(3)}(r)$ develops a plateau of length diverging as $(T - T_{\rm d})^{-1/2}$. As a consequence, for any ε smaller

[5] One interesting feature is that the temperature $T_{\rm gap}$ below which the gap becomes exponentially small, is higher than $T_{\rm d}$. In the temperature range $T_{\rm d} < T < T_{\rm gap}$, the slow modes correspond to metastable states that are 'not seen' if the dynamics starts from a random or equilibrated initial condition.

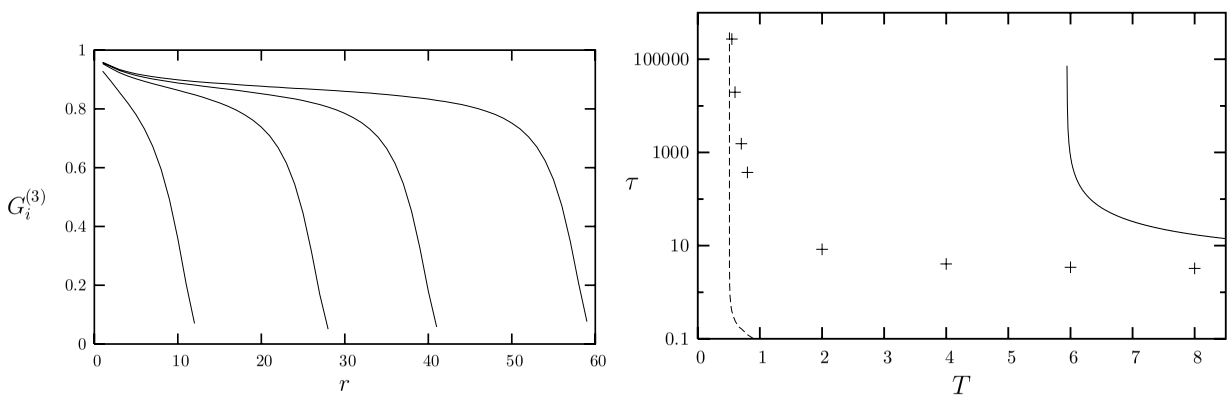

Fig. 1. Numerical results for the 3-spin model on random 3-regular hypergraphs at high temperature. Left: the spatial correlation function $G_i^{(3)}(r)$, from left to right $T = 0.6, 0.53, 0.52, 0.515$. Right: correlation times. Symbols are obtained by Monte Carlo simulations of the heat bath dynamics, the solid line is the rigorous upper bound of Proposition 2, the dashed line corresponds to the lower bound from Theorem 1.

than the value of $G^{(3)}(r)$ on the plateau, $\ell^{(3)}(\varepsilon) \sim (T - T_{\rm d})^{-1/2}$. In order to check whether the lower bound of Theorem 1 is optimal, we estimated the time τ from Monte Carlo simulations of the heat bath dynamics on large ($N = 10^6$) samples. We averaged over several samples and checked that τ_i is approximately independent of i. Our data are presented in the right frame of Fig. 1 and compared with the lower bound of Theorem 1 with $\ell^{(3)}$ evaluated via the recursive method. The correlation time turns out to diverge algebraically at $T_{\rm d}$: $\tau \sim (T - T_{\rm d})^{-\gamma}$. Fitting the data, we get $\gamma \approx 3.2 > 1/2$.

In Fig. 2 we plot the exponential growth rate of the correlation time obtained by extensive numerical simulations on small ($N = 100$) systems (see also Ref. 28 for an analogous study in the fully connected limit). As above, we consider here the case $l = 3$, $p = 3$. This is compared with a rigorous lower bound stated in Proposition 2 below, and with the upper bound obtained from Theorem 1, using $|\mathsf{B}_i(r)| \le N$.

The above discussion of the behavior of correlation times was largely based on the analogy with the fully connected spherical model,[2] and an on heuristic arguments. Here we confirm rigorously several elements of this picture. For the sake of definiteness, we shall consider Glauber dynamics with the heat bath rule

$$\kappa_i^{\sigma}(\sigma_i) = \frac{1}{2}(1 + \sigma_i \tanh \beta h_i(\sigma)), \tag{24}$$

where $h_i(\sigma) = \sum_{a \in \partial i} J_a \prod_{j \in \partial a \setminus i} \sigma_j$ is the 'local field' acting on the i-th spin. Then we have the following result, whose proof is deferred to Appendices C and D.

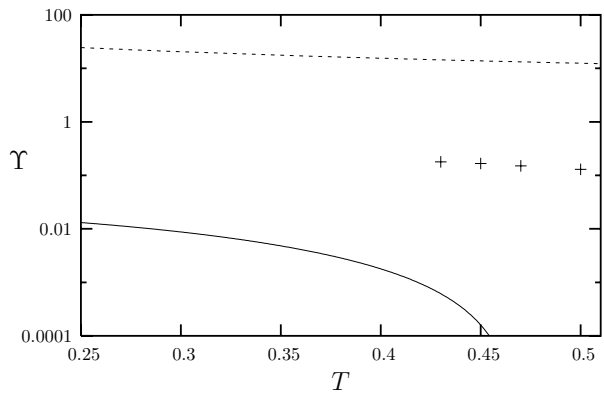

Fig. 2. Correlation times for the 3-spin model on random 3-regular graphs at low temperature. Here we plot $\Upsilon = (\log \overline{\tau})/N$ versus temperature ($\overline{\tau}$ being uniform average of τ_i over the site i). The continuous line is the rigorous lower bound from Proposition 2. The dashed line is the upper bound from Theorem 1. Symbols correspond to numerical results from Monte Carlo simulations.

Proposition 2. *Let $\tau_i(\varepsilon)$ be the correlation time for spin σ_i in a p-spin model on a random l-regular hypergraph with $p, l \geq 3$. Let $T^{\rm fast}_{p,l} = (\text{arctanh}(1/l(p-1)))^{-1}$, and $T^{\rm barr}_{p,l}$, $T^{\rm ann}_{p,l}$ be the temperatures defined in Appendix D.*

If $T > T^{\rm fast}_{p,l}$, then $\tau_i(\varepsilon) \leq (1/\kappa)\log(1/\kappa\varepsilon)$, where $\kappa \equiv (1 - l(p-1)\tanh\beta)$.

If $T^{\rm ann}_{p,l} < T < T^{\rm barr}_{p,l}$, then there exists constants q_* and $\Upsilon > 0$ such that, for any $1/4 > \delta > 0$, $\tau_i(\varepsilon) \geq e^{N[\Upsilon-\delta]}$ for at least $N(q_* - \delta - \varepsilon)$ spins σ_i with high probability.

The analysis in Appendix D also provides rather explicit expressions for the temperatures $T^{\rm barr}_{p,l}$, $T^{\rm ann}_{p,l}$ as well as for q_* and Υ. The numerical values of some of these constants are reported in Table 1for a few values of p and l.

We are now in position to discuss to what extent Theorem 1 can be improved (here we focus on the large ℓ, large τ behavior) without loosing in the generality of its hypotheses. In the high temperature phase $\tau \sim (T - T_{\rm d})^{-\gamma}$, while $\ell \sim (T - T_{\rm d})^{-1/2}$, and the analogy with fully connected models suggests $\gamma \geq 1$ quite generically.[2] Therefore we expect that the lower bound can be improved at most to $\tau \geq \ell^{\zeta}$ with, probably, $\zeta = 2$.

At low temperature $\tau = \exp\{\Theta(N)\}$ while Theorem 1 implies $\tau \leq \exp\{CN\}$. This is optimal apart from a possible improvement in the exponential rate. Our conclusion is that, without further assumptions on the system, Theorem 1 can be improved at most to $C_1\ell_i^{\zeta} \leq \tau_i \leq \exp\{C_2|\mathsf{B}_i(\ell)|\}$.

Let us finally observe that Proposition 2, together with Lemma 4 imply that the model (21) undergoes a purely dynamical phase transition. This is the first time such a behavior is proved for a model of this family (Glauber dynamics on a sparse graph spin model). In finite dimensional models, we expect that such a sharp dynamical transition cannot occur. Nevertheless, it would be interesting to understand whether the correlation time of glassy systems generically undergoes a

Table I. Various Characteristic Temperatures for the *p*-Spin Model on Random *l*-Regular Hypergraphs: $T_{\rm ann}$ is an Upper Bound on the Static Transition Temperature $T_{\rm c}$; $T_{\rm d}$ is the Dynamical Transition Obtained by a Cavity Calculation; $T_{\rm barr}$ and $T_{\rm fast}$ are, Respectively, Lower and Upper Bounds on $T_{\rm d}$

p	l	$T_{\rm ann}$	$T_{\rm barr}$	$T_{\rm d}$	$T_{\rm fast}$
3	3	0	0.470124	0.510	5.944027
3	4	0.854138	0.687684	0.753	7.958158
3	5	1.113214	0.849507	0.935	9.966577
4	3	0	0.376808	0.410	8.962840
4	4	0	0.575513	0.625	11.972171
4	5	0.771325	0.724693	0.785	14.977751

crossover from a polynomial ($\tau \sim \ell^z$) to an exponential ($\tau \sim \exp\{\Theta(\ell^\psi)\}$) relation with the correlation length.

5. PROOF OF THE LOWER BOUND IN THEOREM 1

For pedagogical reasons we first recall some definitions and well-known facts from probability theory, that will be repeatedly used in this and the following proofs. Consider a finite set S and two probability measures $p^{(1)}$, $p^{(2)}$ defined on it. The total variation distance between these two measures is defined as

$$\|p^{(1)} - p^{(2)}\|_{\rm TV} \equiv \frac{1}{2}\sum_x |p^{(1)}(x) - p^{(2)}(x)|. \tag{25}$$

One easily checks that this defines indeed a distance, in particular it vanishes if and only if $p^{(1)} = p^{(2)}$, and that the two following characterizations are equivalent:

$$\|p^1 - p^{(2)}\|_{\rm TV} = \max_{T\subset S} |p^{(1)}(T) - p^{(2)}(T)| = 1 - \sum_x \bar{p}(x), \tag{26}$$

where we defined $\overline{p}(x) = \min[p^{(1)}(x), p^{(2)}(x)]$.

Another very useful form of the total variation distance is defined in terms of couplings of the two measures, that is to say joint distributions $q(x_1, x_2)$ whose marginals are $p^{(1)}$ and $p^{(2)}$: $\sum_{x_2} q(x_1, x_2) = p^{(1)}(x_1)$ and $\sum_{x_1} q(x_1, x_2) = p^{(2)}(x_2)$.

Considering the random variable $(X_1, X_2) \in S^2$ drawn from such a coupling, one can easily prove that

$$\|p^{(1)} - p^{(2)}\|_{\rm TV} \le \mathbb{P}[X_1 \neq X_2]. \tag{27}$$

Moreover one can construct an optimal (or greedy) coupling that achieves the bound. We thus have

$$\|p^{(1)} - p^{(2)}\|_{\rm TV} = \min_q \mathbb{P}[X_1 \neq X_2]. \tag{28}$$

We can now start the proof of the lower bound in Theorem 1 (a presentation in a restricted setting is provided in Ref. 21). This is based on a 'disagreement percolation' argument first used in Ref. 29 (for recent applications, see Refs. 13, 30). More precisely, for a given variable node i and a positive integer r, we construct a Markov process $(x^{(1)}(t), x^{(2)}(t))$ on $\mathcal{X}^N \times \mathcal{X}^N$ in the following way:

- at the initial time $t = 0$, a configuration $x(0)$ is drawn from the law μ, and imposed to the two trajectories, $x^{(1)}(0), x^{(2)}(0) = x(0)$.
- Each variable node owns an independent rate-one Poisson clock. When the clock at j rings, say at time t:
  - if $j \notin \mathsf{B}_i(r)$, $x_j^{(2)}(t)$ is replaced by ξ, drawn from $\kappa_j^{x^{(2)}(t)}(\xi)$, and $x^{(1)}$ remains unchanged.

– if $j \in \mathsf{B}_i(r)$, we draw (ξ_1, ξ_2) from the greedy coupling between $\kappa_j^{x^{(1)}(t)}$ and $\kappa_j^{x^{(2)}(t)}$, and replace $(x_j^{(1)}(t), x_j^{(2)}(t))$ by (ξ_1, ξ_2).

$(x^{(1)}(t), x^{(2)}(t))$ is a coupling of two Markov processes: viewing them separately, $x^{(2)}(t)$ is the original dynamics we are interested in, whereas $x^{(1)}(t)$ is the dynamics on $\mathsf{B}_i(r)$ (cf. Sec. 2.2 for a definition) with boundary condition $x_{\sim i,r}(0)$.

Consider now a function $f(x_i)$ with $|f| \le 1$. Conditioning on $x_{\sim i,r}(0)$, we can apply the observation stated in Sec. 2.3 on the dynamics of $x^{(1)}_{\mathsf{B}_i(r)}(t)$ to obtain

$$\widehat{f}(x_{\sim i,r}(0))^2 \le \left\langle f\big(x_i^{(1)}(0)\big) f\big(x_i^{(1)}(t)\big)\right\rangle_{i,r}^{x(0)}. \tag{29}$$

Averaging over $x_{\sim i,r}(0)$, this yields

$$\langle f(x_i(0))\widehat{f}(x_{\sim i,r}(0))\rangle \le \left\langle f\big(x_i^{(1)}(0)\big) f\big(x_i^{(1)}(t)\big)\right\rangle. \tag{30}$$

We now introduce the indicator function

$$\mathbb{I}(t) = \begin{cases} 1 & \text{if } x_i^{(1)}(t) = x_i^{(2)}(t) \\ 0 & \text{otherwise} \end{cases}, \tag{31}$$

in terms of which we can rewrite the above inequality as

$$\begin{aligned} \langle f(x_i)\widehat{f}(x_{\sim i,r})\rangle &\le \left\langle f\big(x_i^{(2)}(0)\big) f\big(x_i^{(2)}(t)\big)\mathbb{I}(t)\right\rangle + \left\langle f\big(x_i^{(1)}(0)\big) f\big(x_i^{(1)}(t)\big)(1-\mathbb{I}(t))\right\rangle \\ &\le \left\langle f\big(x_i^{(2)}(0)\big) f\big(x_i^{(2)}(t)\big)\right\rangle + 2\langle 1-\mathbb{I}(t)\rangle. \end{aligned} \tag{32}$$

In the last step we used the fact that f is bounded in absolute value by 1.

A disagreement percolation argument implies the bound

$$\langle 1-\mathbb{I}(t)\rangle \le \left(\frac{e\Delta^2 t}{r+1}\right)^{r+1} \tag{33}$$

Indeed, $\langle 1-\mathbb{I}(t)\rangle$ is the probability that $x^{(1)}$ and $x^{(2)}$ disagree on the variable node i at time t. The coupling between $x^{(1)}(t)$ and $x^{(2)}(t)$ has been defined in such a way that at initial time the two configurations agree on all variable nodes, and that a disagreement on a node $j \in \mathsf{B}_i(r)$ can appear when j is updated only if at least one of the neighbors of j already bears a disagreeing assignment. In other words, disagreement has to appear in $V\backslash\mathsf{B}_i(r)$ and propagates towards i. More formally, let us call a disagreement path α an ordered list of variable nodes $\alpha = (i_r, \dots, i_1, i_0 = i)$ such that two successive nodes are adjacent in G, and with $d_{i i_k} = k$. A path α is said to percolate if there exists a sequence of times $0 < t_1 < \cdots < t_{r+1} < t$ such that the clock of the node i_k rings at time t_{r+1-k}. The quantity $\langle 1-\mathbb{I}(t)\rangle$ is thus bounded by the probability of the event "at least one disagreement path has percolated," which is itself smaller than the product of the number of such paths by the probability $p_{r+1}(t)$ for a given path to percolate. As a given clock rings in the infinitesimal time interval $[t_k, t_k + dt]$ for the first

time since t_{k-1} with probability $e^{-(t_k - t_k - 1)}dt$,

$$p_{r+1}(t) = \int_0^t dt_1 e^{-t_1} \int_{t_1}^t dt_2 e^{-(t_2 - t_1)} \cdots \int_{t_r}^t dt_{r+1} e^{-(t_{r+1} - t_r)} = \sum_{s=r+1}^{\infty} \frac{e^{-t} t^s}{s!} \quad (34)$$

$$\leq \frac{t^{r+1}}{(r+1)!} \leq \left(\frac{et}{r+1}\right)^{r+1}. \quad (35)$$

The number of disagreement paths can be bounded by $\Delta^{2(r+1)}$, hence Eq. (33).

Subtracting $\langle f(x_i)\rangle^2$ from both sides of Eq. (32), and taking the supremum over $|f| \leq 1$, we obtain

$$G_i^{(3)}(r) \leq C_i(t) + 2\left(\frac{e\Delta^2 t}{r+1}\right)^{r+1}. \quad (36)$$

Setting $t = \tau_i(\varepsilon)$ and calling $r_* = \max(2e\Delta^2\tau_i(\varepsilon), \log_2(2/\varepsilon))$, we have

$$G_i^{(3)}(r) \leq 2\varepsilon \qquad \forall r \geq r_*, \quad (37)$$

hence an upperbound on $\ell_i^{(3)}(2\varepsilon)$, which can be translated into the form stated in Theorem 1 using the inequalities of Proposition 1. □

6. PROOF OF THE UPPER BOUND IN THEOREM 1

The upper bound is proved by viewing the dynamics inside $\mathsf{B}_i(r)$ as the dynamics of a 'reduced' model whose degrees of freedom are only the ones in $\mathsf{B}_i(r)$, and on which the exterior acts as a time-dependent boundary condition. On one hand, time correlations inside the system decay in a finite time (since the system is finite). On the other, if r is large enough, the time-dependent boundary condition does not affect the behavior in the center of $\mathsf{B}_i(r)$.

Let us begin by defining the Markov dynamics on a subset of the vertices $U \subseteq V$ with time dependent boundary conditions $\{y(t)\}_{t\geq 0}$. This means that one is given a sequence of times $t_0 = 0 < t_1 < \cdots < t_n < \cdots$ with $t_n \to \infty$ as $n \to \infty$, and of configurations $y_0, y_1, \ldots$. The chain is initialized in a configuration x distributed according to $\mu_U^{y_0}$. In each time interval $[t_n, t_{n+1})$, $n = 0, 1, \ldots$, one runs the chain with boundary condition y_n. Then, at time t_{n+1}, the configuration outside U is changed from y_n to y_{n+1}. Averages with respect to this process will be denoted by $\langle\cdot\rangle_U^{\{y\}}$.

It is convenient to state separately the following estimate on time decay of correlations for this dynamics (for the proof, see Appendix B).

Lemma 1. *Let f and g be two functions of $x \in \mathcal{X}^N$, such that $|f(x)|, |g(x)| \le 1$ for any x. Then*

$$\left| \langle f(x(0))g(x(t))\rangle_U^{\{y\}} - \langle f(x)(0))\rangle_U^{\{y\}} \langle g(x(t))\rangle_U^{\{y\}} \right| \le 2e^{-(t-1)/\tau_U^*}, \tag{38}$$

where $\tau_U^* = \exp\{A|U\}$, and $A = -\log(\kappa_0(1-e^{-1}))$.

Let us now turn to the actual proof. Fix a vertex i, a positive integer r, and consider a function f of x_i, with $|f(x_i)| \le 1$ for all x_i's. By conditioning, we can write

$$\langle f(x_i(0))f(x_i(t))\rangle = \mathbb{E}_{\{y\}}\left[\langle f(x_i(0))f(x_i(t))\rangle_{i,r}^{\{y\}}\right], \tag{39}$$

where $\mathbb{E}_{\{y\}}$ denotes expectation with respect to the process $\{y(t)\}_{t\ge 0}$ distributed according to the (stationary) Markov chain on G, and we used the shorthand $\langle\cdot\rangle_{i,r}^{\{y\}}$ for $\langle\cdot\rangle_{\mathsf{B}_i(r)}^{\{y\}}$. Lemma 1 implies

$$\langle f(x_i(0))f(x_i(t))\rangle \le \mathbb{E}_{\{y\}}\left[\langle f(x_i(0))\rangle_{i,r}^{\{y\}}\langle f(x_i(t))\rangle_{i,r}^{\{y\}}\right] + 2e^{-(t-1)/\tau^*(i,r)}, \tag{40}$$

where $\tau^*(i,r)$ is a shorthand for $\tau^*_{\mathsf{B}_i(r)}$. The expectation on the right hand side can be simplified by conditioning on the initial condition $y(0)$:

$$\begin{aligned}
&\mathbb{E}_{\{y\}}\left[\langle f(x_i(0))\rangle_{i,r}^{\{y\}}\langle f(x_i(t))\rangle_{i,r}^{\{y\}}\right] \\
&\quad = \mathbb{E}_{y(0)}\left\{\mathbb{E}_{\{y(t>0)\}}\left[\langle f(x_i(0))\rangle_{i,r}^{\{y\}}\langle f(x_i(t))\rangle_{i,r}^{\{y\}}|y(0)\right]\right\} &(41)\\
&\quad = \mathbb{E}_{y(0)}\left\{\langle f(x_i(0))\rangle_{i,r}^{y(0)}\mathbb{E}_{\{y(t>0)\}}\left[\langle (x_i(t))\rangle_{i,r}^{\{y\}}|y(0)\right]\right\} &(42)\\
&\quad = \mathbb{E}_{y(0)}[f(y_i(0))F(y_{-\sim i,r}(0))], &(43)
\end{aligned}$$

where we defined

$$F(y_{\sim i,r}(0)) = \mathbb{E}_{y(0)}\left[\mathbb{E}_{\{y(t>0)\}}\left[\langle f(x_i(t))\rangle_{i,r}^{\{y\}}|y(0)\right]\middle| y_{\sim i,r(0)}\right]. \tag{44}$$

Since f is uniformly bounded by 1, $|F(x_{\sim i,r})| \le 1$ for all x as well. Moreover one easily shows that $\langle F(x_{\sim i,r})\rangle = \langle f(x_i)\rangle$. Subtracting $\langle f(x_i)\rangle^2$ from Eq. (40) and using the definition of the spatial correlation function, cf. Eq. (6), we obtain

$$C_i^f(t) \le G_i(r) + 2e^{-(t-1)/\tau^*(i,r)}. \tag{45}$$

Taking the supremum over f and setting $r = \ell_i(\varepsilon/2)$ implies the upper bound of Theorem 1.

APPENDIX A: PROOF OF PROPOSITION 1

In this appendix we denote simply by x the variable x_i, by y the 'far apart' variables $x_{\sim i,r}$ and by $\mu(x, y)$ (respectively $\mu(x)$, $\mu(y)$) their joint distribution (respectively, marginal distributions). Finally, we omit the arguments i and r from the correlation functions $G_i(r)$, $G_i^{(1)}(r)$, $\dots$ $G_i^{(4)}(r)$. Proving Proposition 1 amounts to deriving the following inequalities between these functions:

$$\frac{1}{|\mathcal{X}|} G^{(2)} \le G \le G^{(2)}, \tag{46}$$

$$G^{(2)} \le G^{(1)} \le \frac{2}{\mu_*} G^{(2)}, \tag{47}$$

$$\left(\frac{1}{|\mathcal{X}|} G^{(2)}\right) \le G^{(3)} \le G, \tag{48}$$

$$\frac{1}{2} G^{(2)^2} \le G^{(4)} \le \left(\frac{1}{\mu_*} - 1\right) G^{(2)}. \tag{49}$$

Proofs of similar statements can be found repeatedly in the literature. We refer in particular to Ref. 31 for a general presentation, and to Ref. 32 that deals with the tree reconstruction problem which is closely related to the theme of this paper. We collect nevertheless the proofs here for the sake of self-containedness. Notice that the upper bounds in Eqs. (47) and (49) become trivial if $\mu_* = 0$. We shall therefore assume, without loss of generality, $\mu_* > 0$.

(46), lower bound

In the definition of G, take[6] $f(x) = \mathbb{I}(x = x_*)$, and $F(y) = \text{sign}[\mu(y|x_*) - \mu(y)]$. Then

$$G \ge \sum_{x,y} f(x)F(y)\mu(x)[\mu(y|x) - \mu(y)] = \mu(x_*) \sum_{y} |\mu(y|x_*) - \mu(y)|. \tag{50}$$

The thesis follows by choosing x_* which maximizes the last expression.

(46), upper bound

We have

$$G = \sup_{f,F} \left| \sum_{x,y} f(x)F(y)[\mu(x, y) - \mu(x)\mu(y)] \right| \le \sum_{x,y} |\,[\mu(x, y) - \mu(x)\mu(y)]|, \tag{51}$$

[6] We denote by $\mathbb{I}(A)$ the indicator function for the property A.

that is what was claimed.

(47), lower bound

We have

$$G^{(2)} = \sum_{x,y} \mu(x) \left| \sum_{x'} \mu(x')[\mu(y|x) - \mu(y|x')] \right| \leq \sum_{x,x'} \mu(x)\mu(x') \sum_{y} |\mu(y|x) - \mu(y|x')|, \tag{52}$$

And the last expression is upper bounded as $\sup_{x,x'} \sum_y |\mu(y|x) - \mu(y|x')| \equiv G^{(1)}$.

(47), upper bound

We start by noticing that

$$G^{(2)} = \sum_{x} \mu(x) \sum_{y} |\mu(y|x) - \mu(y)| \geq \mu_* \sup_{x} \|\mu_{Y|X}(\cdot|x) - \mu_Y(\cdot)\|_1, \tag{53}$$

where we introduced the standard notation for L^1 norm and used subscripts to precise which variable we are considering the distribution of. By triangular inequality $\|\mu_{Y|X}(\cdot|x_1) - \mu_{Y|X}(\cdot|x_2)\|_1 \leq 2 \sup_x \|\mu_{Y|X}(\cdot|x) - \mu_Y(\cdot)\|_1$ for any x_1, x_2. The thesis follows by taking the sup over x_1, x_2.

(48), lower bound

Take $f(x) = \mathbb{I}(x = x_*)$, and therefore $\widehat{f}(y) = \mu(x_*|y)$. Then

$$G^{(3)} \geq \sum_{x,y} f(x)\widehat{f}(y)[\mu(x, y) - \mu(x)\mu(y)] = \sum_{y} \frac{1}{\mu(y)}[\mu(x_*, y) - \mu(x_*)\mu(y)]^2. \tag{54}$$

By maximizing the right hand side over x_*, we obtain

$$G^{(3)} \geq \frac{1}{|\mathcal{X}|} \sum_{x,y} \frac{1}{\mu(y)}[\mu(x, y) - \mu(x)\mu(y)]^2 \tag{55}$$

$$= \frac{1}{|\mathcal{X}|^2} \left(\sum_{x,y} \frac{1}{\mu(y)} \mu(y)^2 \right) \left(\sum_{x,y} \frac{1}{\mu(y)} [\mu(x, y) - \mu(x)\mu(y)]^2 \right)$$

$$\geq \frac{1}{|\mathcal{X}|^2} \left(\sum_{x,y} |\mu(x, y) - \mu(x)\mu(y)| \right)^2, \tag{56}$$

where the last step followed from Cauchy-Schwarz inequality.

(48), upper bound

Trivial: take $F = \widehat{f}$.

(49), lower bound

We notice that $I(X;Y) = D(p\|q)$ where p and q are distributions on the pair $z = (x, y)$ defined by $p(x, y) = \mu(x, y)$ and $q(x, y), = \mu(x)\mu(y)$, and $D(p\|q) = \sum_z p(z)\log[p(z)/q(z)]$. Defining $P_\lambda(z) = (1-\lambda)q(z) + \lambda p(z)$, by elementary calculus

$$D(p\|q) = \int_0^1 (1-\lambda)\sum_z \frac{1}{p_\lambda(z)}[p(z)-q(z)]^2 d\lambda. \tag{57}$$

By Cauchy-Schwarz (applied to the scalar product with weight $1/p_\lambda(z)$) we have $\sum_z \frac{1}{p_\lambda(z)}[p(z)-q(z)]^2 \geq (\sum_z |p(z)-q(z)|)^2$ and the thesis follows.

(49), upper bound

We use $p_\lambda(z) \geq (1-\lambda)q(z)$ in Eq. (57), and get

$$G^{(4)} \leq \sum_z \frac{1}{q(z)}[p(z)-q(z)]^2 \leq \sup_z \left|\frac{p(z)}{q(z)} - 1\right| \sum_z |p(z)-q(z)|. \tag{58}$$

The thesis follows by noticing that $p(z)/q(z) = \mu(x|y)/\mu(x) \leq 1/\mu_*$.

APPENDIX B: PROOF OF LEMMA 1

Lemma 1 is a well known elementary result for Markov chains with time-independent boundary conditions (see for instance[15] for a functional-analytic argument). We present here an independent and self-contained proof for the general case. We start by restating it in a slightly stronger form. For this purpose, we need to define the dynamics on $U \subseteq V$, with time-dependent boundary condition $\{y(t), t \geq 0\}$ and generic initial distribution ν. This is defined exactly as the process with time-dependent boundary condition introduced in Sec. 6, but for the fact that the initial state is distributed according to ν, instead of $\mu_U^{y(0)}$. It is understood that $\nu(x) = 0$ unless $x_{V\setminus U} = y(0)_{V\setminus U}$.

Lemma 2. *Let $\nu_t^{(1)}$ and $\nu_t^{(2)}$ the distributions at time t for the dynamics on $U \subseteq V$, with the same time-dependent boundary condition $\{y(t), t \geq 0\}$, and initial*

distributions (respectively), $\nu^{(1)}$ and $\nu^{(2)}$. Then

$$\left\| \nu_t^{(1)} - \nu_t^{(2)} \right\|_{\rm TV} \le e^{-(t-1)/\tau_U^*} \left\| \nu^{(1)} - \nu^{(2)} \right\|_{\rm TV}, \tag{59}$$

where $\tau_U^* = \exp\{A|U|\}$, and $A = -\log(k_0(1-e^1))$.

Proof: Denote by $\{x^{(\alpha)}(t), t \ge 0\}, \alpha \in \{1,2\}$ the two processes. We construct a coupling of these two processes, similar to the one of Sec. 5:

- at the initial time $t = 0$, $x^{(1)}(0)$ and $x^{(2)}(0)$ are drawn from the greedy coupling of $\nu^{(1)}$ and $\nu^{(2)}$, hence $\mathbb{P}[x^{(1)}(0) \ne x^{(2)}(0)] = \|\nu^{(1)} - \nu^{(2)}\|_{\rm TV}$.
- When the clock at $j \in U$ rings, say at time t, we draw (ξ_1, ξ_2) from the greedy coupling between $\kappa_j^{x^{(1)}(t)}$ and $\kappa_j^{x^{(2)}(t)}$, and replace $(x_j^{(1)}(t), x_j^{(2)}(t))$ by (ξ_1, ξ_2).

Consider now two times t and $t' = t + \Delta t > t$. If the two processes coincide at a given time, the definition of the coupling implies that they will coincide at all subsequent times. Therefore

$$\begin{aligned} \mathbb{P}\left[x^{(1)}(t') \ne x^{(2)}(t')\right] &= \left\{1 - \mathbb{P}\left[x^{(1)}(t')\right.\right. \\ &= \left.\left. x^{(2)}(t') | x^{(1)} \ne x^{(2)}(t)\right]\right\} \mathbb{P}\left[x^{(1)}(t) \ne x^{(2)}(t)\right]. \end{aligned} \tag{60}$$

The conditional probability appearing in the last expression can be lower bounded by the probability of a particular event implying $x^{(1)}(t') = x^{(2)}(t')$ irrespective of $x^{(1)}(t), x^{(2)}(t)$. The event is defined as follows. Each variable in U tries at least one flip during the time interval $[t, t')$ (this happens with probability $(1 - e^{-\Delta t})^{|U|}$). Furthermore, the last time a flip is attempted on each of the spins, it brings the two coupled processes to coincide on it (this happens with probability at least k_0 for each spin). We have therefore

$$\mathbb{P}\left[x^{(1)}(t') = x^{(2)}(t') | x^{(1)}(t) \ne x^{(2)}(t)\right] \ge [\kappa_0(1 - e^{-\Delta t})]^{|U|}. \tag{61}$$

Finally consider the time interval $[0, t]$ and split it into $[t/\Delta t]$ sub-intervals of size Δt (plus, eventually a sub-interval of smaller size). By repeatedly applying Eq. (60), recalling that $(1 - x) \le e^{-x}$ and $\|\nu_t^{(1)} - \nu_t^{(2)}\|_{\rm TV} \le \mathbb{P}[x^{(1)}(t) \ne x^{(2)}(t)]$, we get

$$\left\| \nu_t^{(1)} - \nu_t^{(2)} \right\|_{\rm TV} \le \exp\left\{ -[\kappa_0(1 - e^{-\Delta t})]^{|U|} \left\lfloor \frac{t}{\Delta t} \right\rfloor \right\} \left\| \nu^{(1)} - \nu^{(2)} \right\|_{\rm TV} \tag{62}$$

The thesis is proved by taking $\Delta t = 1$, and noticing that $\lfloor t \rfloor \ge t - 1$. □

We next show that this result implies Lemma 1. We have

$$|\langle f(x(0)) g(x(t)) \rangle_U^{\{y\}} - \langle f(x(0)) \rangle_U^{\{y\}} \langle g(x(t)) \rangle_U^{\{y\}}| \tag{63}$$

$$= \left| \sum_{x,x'} \mathbb{P}\{x(0) = x\} f(x) g(x') [\mathbb{P}\{x(t) = x' | x(0) = x\} - \mathbb{P}\{x(t) = x'\}] \right| \quad (64)$$

$$\leq \sum_{x} \mathbb{P}\{x(0) = x\} \sum_{x'} |\mathbb{P}\{x(t) = x' | x(0) = x\} - \mathbb{P}\{x(t) = x'\}|. \quad (65)$$

The sum over x' in the last expression is $2\|\nu_t^{(1)} - \nu_t^{(2)}\|_{\mathrm{TV}}$, for two processes of initial conditions $\nu^{(1)}(x') = \mathbb{I}(x = x')$ and $\nu^{(2)} = \mu_U^{\nu(0)}$. The proof is completed by applying Lemma 2, with $\|\nu^{(1)} - \nu^{(2)}\|_{\mathrm{TV}} \leq 1$.

APPENDIX C: HIGH-TEMPERATURE UPPER BOUND FOR THE *p*-SPIN MODEL

In this Appendix we prove the first part of Proposition 2: at high enough temperature, the correlation time $\tau_i(\varepsilon)$ is, with high probability, finite.

We begin by recalling that the temperature $\beta_{p,l}^{\mathrm{fast}} = 1/T_{p,l}^{\mathrm{fast}}$ appearing in the statement of Proposition 2 is defined as the largest value of β such that

$$(p-1) l \tanh \beta \leq 1. \quad (66)$$

We shall also use the notation $\rho(\sigma, \tau)$ to denote the Hamming distance (number of different spins) between two configurations σ and τ. The proof makes use of the following crucial result.

Lemma 3. *Consider the p-spin model, cf. Eqs. (20) and (21), at inverse temperature $\beta < \beta_{p,l}^{\mathrm{fast}}$, and let $\mu_{\pm}^{(i)}$ be the Boltzmann measure, conditioned to $\sigma_i = \pm 1$. Then there exists a coupling of $\mu_+^{(i)}$ and $\mu_-^{(i)}$ such that, if (σ, τ) is a pair of configurations distributed according to such a coupling, their expected Hamming distance is bounded as*

$$\langle \rho(\sigma, \tau) \rangle \leq \frac{1}{1 - l(p-1)\tanh \beta}. \quad (67)$$

Proof: Without loss of generality, we set $i = 0$.

Given a coupling ν between $\mu_+^{(0)}$ and $\mu_-^{(0)}$, and a pair of configurations σ, τ distributed according to ν, let $p_j(\nu)$ be the probability that $\sigma_j \neq \tau_j$ under ν. By definition, $p_0(\nu) = 1$. We claim that, given $j \in V$, $j \neq 0$, it is possible to construct another coupling ν' between $\mu_+^{(0)}$ and $\mu_-^{(0)}$ in such a way that

$$\begin{cases} p_j(\nu') \leq \tanh \beta \sum_{a \in \partial j} \sum_{k \in \partial a \setminus j} p_k(\nu), \\ p_k(\nu') = p_k(\nu) \forall k \neq j \end{cases} \quad (68)$$

We shall denote in the following T_j the mapping $\nu' = \mathsf{T}_j \nu$. The coupling ν' can be defined through the following sampling procedure. First sample σ and τ from ν. Then draw σ'_j, τ'_j by coupling in a greedy fashion the conditional distributions $\mu(\sigma_j|\sigma_{\sim j})$, and $\mu(\tau_j|\tau_{\sim j})$ (to be explicit, we shall denote these conditional distributions as $\mu_j(\cdot|\sigma_{\sim j})$, and $\mu_j(\cdot|\tau_{\sim j})$ in the following). Finally set $\sigma'_k = \sigma_k$ and $\tau'_k = \tau_k$ for all nodes $k \neq j$, and define ν' to be the joint distribution of σ' and τ'. Obviously $p_k(\nu') = p_k(\nu)$ for all $k \neq j$. Moreover

$$p_j(\nu') = \sum_{\sigma,\tau} \nu(\sigma,\tau) \|\mu_j(\cdot|\sigma_{\sim j}) - \mu_j(\cdot|\tau_{\sim j})\|_{\mathrm{TV}}. \tag{69}$$

Notice that $\mu_j(\cdot|\sigma_{\sim j})$ depends on $\sigma_{\sim j}$ only through $\sigma_a \equiv \{\sigma_k : k \in \partial a \backslash j\}$, with $a \in \partial_j$. Denote by $a_1, \dots, a_l$ the indices[7] of function nodes which are neighbors of j and define $\sigma^{(t)} t = 0, \dots, l$ in such a way that $\sigma^{(t)}_{a_s} = \sigma_{a_s}$ for $s \le t$ and $\sigma^{(t)}_{a_s} = \tau_{a_s}$ for $s > t$ (in particular $\sigma^{(0)}_{a_s} = \tau_{a_s}$ and $\sigma^{(l)} = \sigma$). Then

$$p_j(\nu') \le \sum_{\sigma,\tau} \nu(\sigma,\tau) \sum_{t=1}^{l} \|\mu_j(\cdot|\sigma^{(t)}_{\sim j}) - \mu_j(\cdot|\sigma^{(t-1)}_{\sim j})\|_{\mathrm{TV}} \le \tag{70}$$

$$\le \sum_{\sigma,\tau} \nu(\sigma,\tau) \sum_{t=1}^{l} \mathbb{I}(\sigma_{a_t} \neq \tau_{a_t}) \tanh\beta = \tanh\beta \sum_{t=1}^{l} \mathbb{P}_\nu(\sigma_{a_t} \neq \tau_{a_t}). \tag{71}$$

Here we used the fact that

$$\left\|\mu_j\left(\cdot\,|\sigma^{(t)}_{\sim j}\right) - \mu_j\left(\cdot\,|\sigma^{(t-1)}_{\sim j}\right)\right\|_{\mathrm{TV}} = \frac{1}{2}\left|\tanh\left(\beta \sum_{a\in\partial j} J_a \prod_{k\in\partial a\backslash j} \sigma^{(t)}_k\right)\right.$$
$$\left. - \tanh\left(\beta \sum_{a\in\partial j} J_a \prod_{k\in\partial a\backslash j} \sigma^{(t-1)}_k\right)\right|, \tag{72}$$

and $|\tanh a - \tanh b| \le 2 \tanh(|a-b|/2)$. Applying the union bound to Eq. (71), we get Eq. (68).

We shall now construct another mapping $\nu' = \mathsf{T}\nu$ by combining the elementary T_j. Defining $\nu^{(0)} = \nu$, and ordering arbitrarily the variable nodes

[7] Within the configuration model is possible (although with probability vanishing as $N \to \infty$) that the variable node j has less than l neighboring function nodes. Although the proof remains valid in this case, we shall not consider it explicitly in order to lighten the notation.

$1, \dots, N-1$, we construct recursively the couplings $\nu^{(1)}, \dots, \nu^{(N-1)}$ with

$$\nu^{(j)} = \begin{cases} \mathsf{T}_j \nu^{(j-1)} & \text{if } p_j(\mathsf{T}_j \nu^{(j-1)}) \le p_j(\nu^{(j-1)}) \\ \nu^{(j-1)} & \text{otherwise} \end{cases}, \quad \text{for } j = 1, \dots, N-1. \tag{73}$$

Finally we let $\nu^{(N-1)} \equiv \mathsf{T}\nu$. A moment of thought shows that

$$p_j(\mathsf{T}_\nu) \le \tanh\beta \sum_{a \in \partial j} \sum_{k \in \partial a \setminus j} p_k(\nu), \tag{74}$$

for all $j \neq 0$.

To conclude the proof, denote by $\bar{\rho}(\nu)$ the expectation of $\rho(\sigma, \tau)$ when σ and τ are distributed according to the coupling ν, which can also be rewritten as $\bar{\rho}(\nu) = \sum_j p_j(\nu)$. By summing over $j \neq 0$ Eq. (74), recalling that $p_0(\nu) = 1$ and that each node k is the neighbor of at most $l(p-1)$ modes j, we get

$$\overline{\rho}(\mathsf{T}_\nu) \le 1 + l(p-1)\tanh\beta\,\bar{\rho}(\nu). \tag{75}$$

Since $l(p-1)\tanh\beta < 1$, a coupling achieving (67) can be constructed by iterating a sufficient number of times the transformation $\nu \mapsto \mathsf{T}\nu$ from an arbitrary initial coupling. In fact the sequence of distributions $\mathsf{T}^n\nu$ admits a subsequential limit because the space of distributions over a finite set is compact. □

Notice that the above proof is in fact closely related to the proof of Dobrushin uniqueness condition[33] as described, for instance, in Ref. 34. For arguments of this type we also refer to Ref. 35.

We are now in position of proving the upper bound in Proposition 2. Consider two processes $\{\sigma^{(+)}(t)\}_{t\ge 0}$ and $\{\sigma^{(-)}(t)\}_{t\ge 0}$ evolving according to the Glauber dynamics with initial conditions distributed accordingly to $\mu_+^{(i)}$ (for $\sigma^{(+)}(0)$), and $\mu_-^{(i)}$ (for $\sigma^{(-)}(0)$). In other words, $\{\sigma^{(+)}(t)\}$ (respectively $\{\sigma^{(-)}(t)\}$) is the stationary Glauber dynamics conditioned to $\sigma_i^{(+)}(0) = +1$ (respectively, $\sigma_i^{(-)} = -1$). Then it is easy to show that

$$C_i(t) = \frac{1}{2}\left(1 - m_i^2\right)\left[\langle \sigma_i^{(+)}(t)\rangle - \langle \sigma_i^{(-)}(t)\rangle\right], \tag{76}$$

where $m_i = \langle \sigma_i \rangle$ is the expectation of σ_i, with respect to the (unconditional) Boltzmann measure.

Given an arbitrary coupling of the two processes $\{\sigma^{(+)}(t)\}$, and $\{\sigma^{(-)}(t)\}$, the correlation function is obviously bounded as

$$C_i(t) \le \langle \rho\left(\sigma^{(+)}(t), \sigma^{(-)}(t)\right)\rangle. \tag{77}$$

We construct such a coupling as follows. The initial conditions $\sigma^{(+)}(0)$, and $\sigma^{(-)}(0)$ are chosen according to a coupling of the conditional distributions $\mu_\pm^{(i)}$ that

achieves the bound (67). The joint dynamics is defined using 'path coupling.'[19,36] This construction only requires to define the evolution of $\sigma^{(+)}(t)$ and $\sigma^{(-)}(t)$ when they differ in a single position. If this is the case, we use the greedy coupling of the update probabilities, as in the proof of Lemma 1, cf. Appendix B. From a pair of configurations $(\sigma^+, \sigma^{(-)})$ with $\rho(\sigma^{(+)}, \sigma^{(-)}) = 1$, the expected (with respect to the greedy coupling). Hamming distance after one spin update can be upper bounded by $1 - (\kappa/N)$, with $\kappa \equiv (1 - l(p-1)\tanh\beta)$. Standard path coupling arguments allow to extend this bound for an arbitrary number of spin updates. Let us denote by U_t the number of spin flips between times 0 and t (i.e. a Poisson variable of mean Nt), and by $\mathbb{E}_U$ the corresponding expectation:

$$C_i(t) \le \langle \rho(\sigma^{(+)}(0), \sigma^{(-)}(0)) \rangle \mathbb{E}_U \left(1 - \frac{\kappa}{N}\right)^{U_t} \le \kappa^{-1} \exp\{-\kappa t\}, \tag{78}$$

which clearly implies the thesis.

APPENDIX D: LOW-TEMPERATURE LOWER BOUND FOR THE *p*-SPIN MODEL

In this Appendix we prove the second part of Proposition 2 on the correlation time of the p-spin model on random regular hypergraphs. Throughout the Appendix we denote by $Q_{\sigma,\tau} = N^{-1}\sum_{i=1}^N \sigma_i \tau_i$ the normalized overlap of configurations σ and τ. We further let $Z(\beta)$ be the partition function, i.e. the normalization constant in Eq. (20), and $Z(q;\beta)$ the *constrained* partition function

$$Z(q;\beta) \equiv \sum_{\sigma^{(1)},\sigma^{(2)}} e^{-\beta E(\sigma^{(1)}) - \beta E(\sigma^{(2)})} \mathbb{I}(Q_{\sigma^{(1)},\sigma(2)} = q), \tag{79}$$

where $q \in \{-1, -1+2/N, \dots, 1-2/N, 1\}$.

It is convenient to state a few preliminary results. We start by computing the expectation of $Z(\beta)$ and $Z(\beta, q)$. It is straightforward to get

$$\mathbb{E}Z(\beta) = 2^N (\cosh\beta)^M = e^{N\phi(\beta)} \tag{80}$$

where we defined $\phi(\beta) = \log 2 + \frac{l}{p}\log\cosh\beta$. The expected constrained partition function is only slightly more involved

$$\begin{aligned}\mathbb{E}Z(q;\beta) &= 2^N \binom{N}{m}\binom{Nl}{ml}^{-1} \operatorname{coeff}\big[((\cosh 2\beta)p_+(x) + p_-(x))^M, x^{(N-m)l}\big] \\ &\doteq e^{N\phi(q;\beta)}\end{aligned} \tag{81}$$

where we denoted by $\doteq$ identity to the leading exponential order, $m \equiv N(1+q)/2$, and $p_\pm(x) = (1+x)^p \pm (1-x)^p/2$. Using Haymann (or saddle-point) method,

one obtains the exponential growth rate

$$\phi(q;\beta) = \log 2 - (l-1)h\left(\frac{1-q}{2}\right) + \frac{l}{p}\log((\cosh 2\beta)p_+(z_q) + p_-(z_q)) - l\frac{1-q}{2}\log z_q, \tag{82}$$

where $h(x) = -x\log x - (1-x)\log(1-x)$ is the entropy function and z_q is the unique non-negative solution of the equation

$$z_q \frac{(\cosh 2\beta)p'_+(z_q) + p'_-(z_q)}{(\cosh 2\beta)p_+(z_q) + p_-(z_q)} = p\frac{1-q}{2}. \tag{83}$$

The function $\phi(q;\beta)$ is straightforwardly evaluated numerically.

Let us underline some of its properties. One easily shows that for all temperatures, it has a local maximum at $q = 0$, with $\phi(0;\beta) = 2\phi(\beta)$. Moreover, for $\beta = 0$, the function reduces to $\phi(q;0) = \log 2 + h((1-q)/2)$, for which $q = 0$ is the global maximum. We define $\beta^{\rm ann}_{p,l}$ as the largest value of β such that for any $\beta' < \beta$, $\phi(q;\beta')$ achieves its global maximum at $q = 0$. Notice that, if $p \geq l$, then $\beta^{\rm ann}_{p,l} = \infty$.

Let us now turn to the behaviour around $q = 1$. A simple calculation shows that $\phi(q = 1;\beta) = \phi(2\beta)$. Let us define $U(q;\beta) \equiv \phi(q = 1;\beta) - \phi(q;\beta)$, and $U(q;\infty)$ its limit as $\beta \to \infty$. Taking afterwards the limit $q \to 1$, one shows that $U(1-\delta;\infty) = -(l-2)(\delta\log\delta)/4 + O(\delta)$.

We define the *annealed free energy barrier*

$$\Upsilon(\beta) \equiv \sup_{\alpha\in(1/2,1)} [(1-\alpha)\Upsilon_0(\beta) - \widetilde{\phi}(\beta,\alpha)], \tag{84}$$

$$\Upsilon_0(\beta) \equiv \sup_{q\in(0,1]} U(q;\beta), \tag{85}$$

$$\widetilde{\phi}(\beta,\alpha) \equiv \alpha\phi((2-\alpha^{-1})\beta) + (1-\alpha)\phi(2\beta) - \phi(\beta). \tag{86}$$

From the behaviour of $U(q;\infty)$ in the limit $q \to 1$ one can deduce that, if $l \geq 3$, $\Upsilon_0(\beta)$ is strictly positive for β large enough. On the other hand, $\widehat{\phi}(\beta,\alpha)$ is non-negative because $\phi(\beta)$ is convex. From the low temperature expansion $\phi(\beta) = (1-l/p)\log 2 + (l/p)\beta + O(e^{-2\beta})$ it follows that $\widehat{\phi}(\beta,\alpha) \to 0$ as $\beta \to \infty$. We proved therefore that, if $l \geq 3$, there exists $\beta < \infty$ such that $\Upsilon(\beta') > 0$ for $\beta' > \beta$. We call $\beta^{\rm barr}_{p,l}$ the smallest β with such a property. The numerical estimation of the associated temperatures (inverse of β) for a few values of (p,l) are given in Table 1.

These computations have several immediate consequences. The first is that they yield the free energy in the thermodynamic limit.

Lemma 4. *Consider the model (20), (21) with $p \geq 3$ and $l \geq 2$, and let $\phi(\beta)$ be given as above. Define the expected free energy density as $\phi_N(\beta) \equiv N^{-1}\mathbb{E}\log Z(\beta)$, and assume $\beta < \beta^{\rm ann}_{p,l}$. Then $\phi_N(\beta) \to \phi(\beta)$ as $N \to \infty$. Furthermore, for any $\delta > 0$ there exists $C(\delta) > 0$ such that:*

$$\mathbb{P}\{|\log Z(\beta) - N_{\phi_N}(\beta)| \geq N\delta\} \leq 2e^{-NC(\delta)}. \tag{87}$$

Proof: The statement $\phi_N(\beta) \to \phi(\beta)$ follows from the second moment method applied to the random variable $Z(\beta)$ (notice that $\mathbb{E}Z(\beta)^2 \doteq \exp\{N \sup_q \phi(q;\beta)\}$). Equation (87) can then be proved through standard concentration inequalities.[37] □

Lemma 5. *Consider the model (20), (21) with $p \geq 3$ and $l \geq 2$, and let $\phi(q:\beta)$ be defined as in Eqs. (82) and (83). Then, for any $\beta > 0$ and any $\delta > 0$, and any N large enough:*

$$\mathbb{P}\{Z(q;\beta) \geq e^{N[\phi(q;\beta)+\delta]}\} \leq e^{-N\delta/2}. \tag{88}$$

Proof: This is just Markov inequality applied to the random variable $Z(q;\beta)$, noting that $\lim(1/N)\log\mathbb{E}Z(q;\beta) = \phi(q;\beta)$. □

Lemma 6. *Consider the model (21) with $p \geq 3$ and $l \geq 2$, and $\beta < \beta^{\rm ann}_{p,l}$. Let $\sigma^{(1)}$ and $\sigma^{(2)}$ be two i.i.d. configurations drawn from the Boltzmann distribution (20), and denote by Q_{12} be their overlap. Then, for any $\delta > 0$, there exist constants C_1 and $C_2 > 0$ such that, for all N large enough*

$$\mathbb{P}\{|\langle Q_{12}\rangle| \geq N\delta\} \leq C_1(\delta)e^{-NC_2(\delta)}. \tag{89}$$

Proof: First notice that, by the two previous Lemmas, there exist constants $C_1(\delta)$ and $C_2(\delta)$ such that, with probability at least $1 - C_1(\delta)e^{-NC_2(\delta)}$ the following happens: (i) $Z(q;\beta) \leq e^{N[\phi(q;\beta)+\delta^2]}$ for any $q \in \{-1, -1+2/N, \dots, 1-2/N, 1\}$, and (ii) $Z(\beta) \geq e^{N[\phi(\beta)-\delta^2]}$. Under these conditions, for any $\xi > 0$ we have

$$\langle Q_{12}\rangle \leq \xi + \mathbb{P}\{|Q_{12}| \geq \xi\} = \xi + \frac{1}{Z(\beta)^2}\sum_{|q|\geq\xi} Z(q;\beta) \tag{90}$$

$$\leq \xi + e^{3N\delta^2}\sum_{|q|\geq\xi} \exp\{N[\phi(q;\beta) - 2\phi(\beta)]\}, \tag{91}$$

with the sum being restricted to $q \in \{-1, -1+2/N, \dots, 1-2/N, 1\}$. Recall that $q = 0$ is a stationary point of $\phi(q;\beta)$ with $\phi(0,\beta) = 2\phi(\beta)$. A little calculus

shows that $\phi''(0;\beta) < 0$ for any $p \geq 3$. As a consequence for any $T > T_{p,l}^{\rm ann}$ there exists $\alpha > 0$ such that $\phi(q;\beta) \leq 2\phi(\beta) - \alpha q^2$. Therefore

$$|\langle Q_{12}\rangle| \leq \xi + 2Ne^{3N\delta^2 - N_{\alpha\xi^2}}. \tag{92}$$

By taking $\xi = 2\delta/\sqrt{\alpha}$, we have $|\langle Q_{12}\rangle| \leq 4\delta/\sqrt{\alpha}$ for any N large enough. The thesis follows by rescaling δ. □

The crucial step is made in the next Lemma, which estimates the global correlation function

$$\widetilde{C}(t) = \frac{1}{N}\sum_{i=1}^{N}\langle\sigma_i(0)\sigma_i(t)\rangle. \tag{93}$$

Lemma 7. *Consider the model (20), (21) with $p \geq 3, l \geq 2$ and $\beta_{p,l}^{\rm barr} \leq \beta \leq \beta_{p,l}^{\rm ann}$. Let $\Upsilon = \Upsilon(\beta)$ be the associated annealed free energy barrier, and q_* be the largest value of q at which the sup in Eq. (85) is achieved. Then, for any $\delta \in (0, 1/4]$, and $t > (4/N)\log(2/\delta)$, one has $\widetilde{C}(t) \geq q_* - \delta - te^{-N[\Upsilon-\delta]}$ with high probability.*

Proof: Consider two equilibrium trajectories $\{\sigma^{(1)}(t) : t \geq 0\}, \{\sigma^{(2)}(t) : t \geq 0\}$ evolving independently according to the stationary Glauber dynamics for the model (20), (21). In particular, at any time $\sigma^{(1)}(t)$ and $\sigma^{(2)}(t)$ are distributed independently according to the equilibrium distribution μ. We further let $Q(t) \equiv Q_{\sigma^{(1)}(t),\sigma^{(2)}(t)}$. Throughout the proof, we denote by E and P, respectively, expectation and probability with respect to this process (not to be confused with expectation and probability with respect to the graph and energy function).

It is elementary that

$$\widetilde{C}(2t) = \sum_{\sigma}\mu_\beta(\sigma)\mathsf{E}\left[Q(t)\sigma^{(1)}(0) = \sigma^{(2)}(0) = \sigma\right]. \tag{94}$$

where, for future convenience, we specified the temperature β at which the Boltzmann measure must be considered. Let us denote by I_σ the event $\{\sigma^{(1)}(0) = \sigma^{(2)}(0) = \sigma\}$. Moreover, for any $q \in \{-1, -1 + 2/N, \ldots, 1 - 2/N, 1\}$, we denote by $\mathcal{A}_{q,t}$ the event that there exists a time $s \in [0, t]$ such that $Q(s) = q$. Clearly

$$\begin{aligned}
\mathsf{E}[Q(t)|I_\sigma] &= \mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma\}\mathsf{E}[Q(t)|I_\sigma, \mathcal{A}_{q,t}] \\
&\quad + (1 - \mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma\})\mathsf{E}[Q(t)|I_\sigma, \overline{\mathcal{A}}_{q,t}] &(95)\\
&\geq -\mathsf{P}\{\mathcal{A}_{q,t|I_\sigma}\} + q(1 - \mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma) &(96)\\
&= q - (1+q)\mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma\}. &(97)
\end{aligned}$$

Denote by U_t the total number of spin flips in the two configurations up to time t (this is a Poisson random variable of mean $2Nt$). Then

$$\mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma\} \le \mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma, U_t \le 4Nt\} + \mathsf{P}\{U_t > 4Nt\} \tag{98}$$

$$\le \mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma, U_t \le 4Nt\} + e^{-Nt/2}. \tag{99}$$

If we call $\mathcal{U}$ the event $\{U_t \le 4Nt\}$, we found

$$\widetilde{C}(2t) \ge q - 2\left\{\sum_\sigma \mu_\beta(\sigma)\mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma, \mathcal{U}\} + e^{-Nt/2}\right\} \tag{100}$$

Call $F(\sigma) \equiv \mathsf{P}\{\mathcal{A}_{q,t}|I_\sigma, \mathcal{U}\}$, and choose $\alpha \in (1/2, 1)$. By Hölder inequality

$$\sum_\sigma \mu_\beta(\sigma)F(\sigma) = \sum_\sigma \mu_{2\beta}(\sigma)\left(\frac{\mu_\beta(\sigma)}{\mu_{2\beta(\sigma)}}\right)F(\sigma) \tag{101}$$

$$\le \left\{\sum_\sigma \mu_{2\beta}(\sigma)\left(\frac{\mu_\beta(\sigma)}{\mu_{2\beta}(\sigma)}\right)^{1/\alpha}F(\sigma)\right\}^\alpha \left\{\sum_\sigma \mu_{2\beta}(\sigma)F(\sigma)\right\}^{1-\alpha} \tag{102}$$

$$= \frac{Z(\beta(2-\alpha^{-1}))^\alpha Z(2\beta)^{1-\alpha}}{Z(\beta)}\left\{\sum_\sigma \mu_{2\beta}(\sigma)F(\sigma)\right\}^{1-\alpha}, \tag{103}$$

where, in passing from (102) to (103), we used the fact that $F(\sigma) \le 1$ and the definition of $Z(\beta)$. A moment of thought reveals that

$$\sum_\sigma \mu_{2\beta}(\sigma)F(\sigma) = \mathsf{P}\{\mathcal{A}_{q,t}|Q(0) = 1, \mathcal{U}\} \le \frac{\mathsf{P}\{\mathcal{A}_{q,t}|U_t \le 4Nt\}}{\mathsf{P}\{Q(0) = 1\}} \tag{104}$$

$$\le 4Nt\frac{\mathsf{P}\{Q(t) = q|U_t \le 4Nt\}}{\mathsf{P}\{Q(0) = 1\}} \le 8Nt\frac{\mathsf{P}\{Q(t) = q\}}{\mathsf{P}\{Q(0) = 1\}}. \tag{105}$$

(the last inequality follows from the fact that $\mathsf{P}(U_t \le 4Nt) \ge 1/2$ for $t \ge \frac{2}{N}\log 2$). Next notice that $\mathsf{P}\{Q(t) = q\} = Z(q;\beta)/Z(\beta^2)$, and $\mathsf{P}\{Q(0) = 1\} = Z(2\beta)/Z(\beta)^2$. Putting the various terms together, we obtain

$$\widetilde{C}(2t) \ge q - 2e^{-Nt/2} - 2(8Nt)^{1-\alpha}\frac{Z(\beta(2-\alpha^{-1}))^\alpha Z(q;\beta)^{1-\alpha}}{Z(\beta)}. \tag{106}$$

Using Lemmas 4 and 5, the product of the partition functions can be upper bounded by $\exp[-N((1-\alpha)U(q;\beta) - \widehat{\phi}(\beta,\alpha) - \delta/2)]$, for any $\delta > 0$, with high probability. With the hypothesis of the lemma $8Nt > 1$, we can thus replace $(8Nt)^{1-\alpha}$ by $8Nt$ in the inequality. Moreover this prefactor $8N$ will be smaller than $\exp[N\delta/2]$ for large enough N. We will also have $2e^{-Nt/2} \le \delta$. Fixing q in such a way to achieve the sup in Eq. (85) and optimizing over $\alpha \in (1/2, 1)$ completes the proof. □

We can now prove the low temperature lower bound in Proposition 2. Consider the sum of the local correlation functions

$$C(t) = \frac{1}{N}\sum_{i=1}^{N} C_i(t) = \frac{1}{N}\sum_{i=1}^{N}[\langle\sigma_i(t)\sigma_i(0)\rangle - \langle\sigma_i(t)\rangle\langle\sigma_i(0)\rangle]. \tag{107}$$

One has $C(t) = \widetilde{C}(t) - \langle Q_{12}\rangle$, where Q_{12} is the normalized overlap of two i.i.d. configurations $\sigma^{(1)}$ and $\sigma^{(2)}$ distributed according to the Boltzmann measure (20), (21). Lemma 6 implies that, for any $\delta > 0$, $C(t) \geq \widetilde{C}(t) - \delta$ with high probability. By Lemma 7, we have therefore

$$C(t) \geq q_* - 2\delta - te^{-N[\Upsilon-\delta]}. \tag{108}$$

Fix $t_* = e^{N[\Upsilon-2\delta]}$. Then, for N large enough, we have $C(t_*) \geq q_* - 3\delta$. If $f(\varepsilon)$ is the fraction of sites i such that $\tau_i(\varepsilon) > t_*$, then $C(t_*) \leq f(\varepsilon) + \varepsilon$, hence $f(\varepsilon) \geq q_* - 3\delta - \varepsilon$. The thesis follows by rescaling δ.

ACKNOWLEDGMENTS

A.M. would like to thank Elchanan Mossel for his careful reading of the manuscript and a stimulating mail exchange. A.M. is also grateful to the Newton Institute for Mathematical Sciences (Cambridge, UK) for its kind hospitality during the completion of this work. G.S. is supported by EVERGROW, integrated project No. 1935 in the complex systems initiative of the Future and Emerging Technologies directorate of the IST Priority, EU Sixth Framework.

REFERENCES

1. J.-L. Barrat, M. Feigelman, J. Kurchan and J. Dalibard, (eds.), *Slow Relaxations and Non-quilibrium Dynamics in Condensed Matter*, Proceedings of the LXXVII Les Houches Summer School (Springer, Berlin, 2003).
2. J.-P. Bouchaud, L. F. Cugliandolo, J. Kurchan and M. Mézard, Out of equilibrium dynamics in spin glasses and other glassy systems, in A. P. Young (ed.), *Spin Glasses and Random Fields* (World Scientific, Singapore, 1997).
3. C. Donati, S. C. Glotzer and P. H. Poole, *Phys. Rev. Lett.* **82**:5064 (1999).
4. C. Bennemann, C. Donati, J. Baschnagel and S. C. Glotzer, *Nature* **399**:246 (1999).
5. S. Franz and G. Parisi, *J. Phys: Condens. Matter* **12**:6335 (2000).
6. C. Donati, S. Franz, S. C. Glotzer and G. Parisi, *J. Non-Cryst. Sol.* **307–310**:215 (2002).
7. E. Weeks, J. C. Crocker, A. C. Levitt, A. Schfield and D. A. Weitz, *Science* **287**:627 (2000).
8. J.-P. Bouchaud and G. Biroli, *Europhys. Lett.* **67**:21 (2004).
9. L. Berthier, G. Biroli, J.-P. Bouchaud, L. Cipelletti, D. El Masri, D. L'Hôte, F. Ladieu and M. Pierno, *Science* **310**:1797 (2005).
10. G. Biroli and J.-P. Bouchaud, *J. Chem. Phys.* **121**:7347 (2004).
11. R. Mulet, A. Pagnani, M. Weigt and R. Zecchina, *Phys. Rev. Lett.* **89**:268701 (2002).
12. M. Mézard, G. Parisi and R. Zecchina, *Science* **297**:812 (2002).
13. N. Berger, C. Kenyon, E. Mossel and Y. Peres, *Prob. Theory Rel. Fields* **131**:311 (2005).

14. F. Martinelli, A. Sinclair and D. Weitz, *Comm. Math. Phys.* **250**:301 (2004).
15. A. Guionnet and B. Zegarlinski, Lectures on logarithmic sobolev inequalities. *Séminaire de Probabilites de Strasbourg* **36**:1 (2002).
16. F. Martinelli, Lectures on Glauber dynamics for discrete spin models, in *Lectures on probability theory and statistics, Saint-Flour 1997* (Springer, Berlin, 1999).
17. M. Dyer, A. Sinclair, E. Vigoda and D. Weitz, *Rand. Struc. Alg.* **24**:461 (2004).
18. F. R. Kschischang, B. J. Frey and H.-.A. Loeliger, *IEEE Trans. Inform. Theory* **47**:498 (2001).
19. D. Aldous and J. A. Fill, *Reversible Markov Chains and Random Walks on Graphs*, book in preparation. Draft available at http://www.stat.berkeley.edu/users/aldous/RWG/book.html.
20. A. Montanari and G. Semerjian, *Phys. Rev. Lett.* **94**:247201 (2005).
21. A. Montanari and G. Semerjian, On the dynamics of the glass transition on Bethe lattices, cond-mat/0509366, *J. Stat. Phys.* **124**(1):103 (2006).
22. T. M. Cover and J. A. Thomas, *Elements of Information Theory* (Wiley Interscience, New York, 1991).
23. S. Franz, M. Mézard, F. Ricci-Tersenghi, M. Weigt and R. Zecchina, *Europhys. Lett.* **55**:465 (2001).
24. S. Franz and G. Parisi, *J. Phys. I (France)* **5**:1401 (1995).
25. S. Franz, *Europhys. Lett.* **73**:492 (2006).
26. A. Montanari and C. Toninelli, unpublished.
27. M. Mézard and A. Montanari, Reconstruction on trees and spin glass transition, cond-mat/0512295.
28. A. Billoire, L. Giomi and E. Marinari, *Europhys. Lett.* **71**:824 (2005).
29. J. van den Berg, *Comm. Math. Phys.* **152**:161 (1993).
30. T. P. Hayes and A. Sinclair, A general lower bound for mixing of single site dynamics on graphs, FOCS 2005 and math.PR/0507517.
31. A. L. Gibbs and F. E. Su, *Inter. Stat. Rev.* **70**:419 (2002).
32. E. Mossel, *Ann. Appl. Probab.* **11**:285 (2001).
33. R. L. Dobrushin, *Theory Prob. Appl.* **13**:197 (1968).
34. H. O. Georgii, *Gibbs Measures and Phase Transitions* (de Gruyter, Berlin, 1988).
35. D. Weitz, *Rand. Struc. Alg.* **27**:445 (2005).
36. R. Bubley and M. Dyer, *Proceedings of the 38th Symposium on the Foundations of Computer Science*, IEEE, 223 (1997).
37. M. Talagrand, *Spin Glasses: Cavity and Mean Field Models* (Springer, New York, 2003).