Journal of Statistical Physics, Vol. 75, Nos. 3/4, 1994

A Dynamical Phase Transition in a Caricature of a Spin Glass

Arnoldo Frigessi[1] and Frank den Hollander[2]

Received July 8, 1993; final November 10, 1993

This paper studies the rate of convergence to equilibrium of Glauber dynamics (Gibbs Sampler) for a system of N Ising spins with random energy (at inverse temperature $\beta > 0$). For each of the 2^N spin configurations the energy is drawn independently from the values 0 and $-\log N$ with probabilities $1 - N^{-\gamma}$, resp. $N^{-\gamma}$ ($\gamma > 0$), and is kept fixed during the evolution. The main result is an estimate of the coupling time of two Glauber dynamics starting from different configurations and coupled via the same updating noise. As $N \to \infty$ the system exhibits two dynamical phase transitions: (1) at $\gamma = 1$ the coupling time changes from polynomial ($\gamma > 1$) to stretched exponential ($\gamma < 1$) in N; (2) if $\gamma < 1$, then at $\beta = \gamma$ the "almost coupling time" [i.e., the first time that the two dynamics are within distance $o(N)$] changes from polynomial ($\beta < \gamma$) to stretched exponential ($\beta > \gamma$) in N. The techniques used to control the randomness in the coupling are static and dynamic large-deviation estimates and stochastic domination arguments.

KEY WORDS: Gibbs Sampler; coupling; Glauber dynamics; random medium; large deviations.

INTRODUCTION

Glauber dynamics is an important tool in the simulation of random-energy spin systems, e.g., spin glasses,[9,13] neural nets,[2] and Bayesian models in image analysis.[11] The *Gibbs Sampler* has been particularly popular because of its simple updating rule: at each step one spin is drawn randomly and is updated according to the local conditional Gibbs measure. The dynamics is reversible and converges to the Gibbs measure associated with the energy.

[1] Laboratorio di Statistica, Università di Venezia, and Istituto per le Applicazioni del Calcolo-C.N.R., 00161 Rome, Italy.

[2] Mathematisch Instituut, Universiteit Utrecht, P.O. Box 80.010, 3508TA Utrecht, The Netherlands.

0022-4715/94/0500-0585$07.00/0

Knowledge of the rate of convergence is important in order to understand equilibrium and to devise stopping rules for the simulation. One way of probing this rate is to study the *coupling time* of two Glauber dynamics starting from different configurations and coupled via the same updating noise. The *coupling inequality* gives an upper bound on the rate of convergence to zero of the total variation between the two probability laws. The upper bound is in terms of the tail of the distribution of the coupling time.[1,8] This approach is pursued by Derrida *et al.*[3,5,6] in a series of papers which present heuristic arguments and numerical simulations.

Recently Cassandro *et al.*[4] made a rigorous analysis of the coupling time for a very simple random-energy model, where each configuration can take only two energy values 0 and $-\log N$ drawn independently with probability $1-N^{-\gamma}$, resp. $N^{-\gamma}$ ($\gamma>0$). N is the number of spins in the system. They consider two types of dynamics: (1) the uniform sampler; (2) the Gibbs Sampler (both at inverse temperature $\beta>0$). A spin is updated according to the Boltzmann factor associated with: (1) the energy of the old configuration; (2) the energy difference between the old and the new configuration (i.e., at each step a random spin is updated according to the local conditional Gibbs measure). The two initial configurations are chosen with all spins up, resp. all spins down, i.e., at maximal distance N. The low-energy configurations are energy valleys that slow down the motion of the dynamics. Three different regimes are established as $N\to\infty$:

I: $\gamma>1$ (weak disorder). The energy is constant in large parts of the configuration space. The coupling time is $N\log N$, as in the case of homogeneous energy.

II: $\gamma<1$, $\beta<\gamma$ (strong disorder, high temperature). The energy is sufficiently random to slow down the coupling. Still the system is fast in escaping the low-energy configurations. The two components get close together, within distance εN for any $\varepsilon>0$, in a time of order N.

III: $\gamma<1$, $\beta>\gamma$ (strong disorder, low temperature). The system is slow in escaping the low-energy configurations. The distance between the two components stays large, above δN for any $\delta<1/2$, during a time at least $N^{1+\varepsilon}$ for some $\varepsilon=\varepsilon(\delta)>0$.

Thus the coupling exhibits a *dynamical phase transition* at $\beta=\gamma$ provided $\gamma<1$, i.e., a crossover between regimes II and III.

In the present paper we study the Gibbs Sampler version of the Cassandro *et al.* random-energy model. We extend their results by establishing the following:

A. There is a *drastic change of time scale* between regimes II and III.

Namely, we prove that in regime III the distance needs a time of order $\exp(KN^{1-\gamma})$ to drop from $\frac{1}{2}N$ to δN for any $\delta < 1/3$ and some $K = K(\delta) > 0$. The drastic crossover between regimes II and III, from polynomial to stretched-exponential time scale, indicates *extremely slow convergence to equilibrium* at low temperature.

B. In regime II the distance drops to $N^{1-(\gamma-\beta)}$ in a time of order N and then is blocked for a long time. That is, there appears an *intermediate distance scale* at which the two components "almost meet" after a relatively short time, but then suddenly lose track of each other and need a very long time to finally hit each other.

Section 1.5 gives heuristic explanation of A and B.

Result A: In regime II the average sojourn time of a component in a low-energy point is much smaller then the average time it runs between low-energy points. Therefore most of the time the two components are moving. In regime III precisely the reverse is true. Therefore most of the time either both components are stuck or one of the components is stuck and the other is moving. In the latter case the two components easily lose track of each other because the dimension N of the space is very large.

Result B: In regime II, when the distance between the two configurations is small the updating noise has only a small probability to pick a spin coordinate where the two configurations do not yet coincide, take away the separation between the spins, and thereby decrease the distance. On the other hand, it has a large probability to pick a spin coordinate where the two configurations do coincide. Since there is a positive probability (in the inhomogeneous medium) that this will lead to a separation of the spins, the distance may also increase. (The separation probability is zero in the homogeneous medium.) There is an intermediate distance scale where these two effects balance.

Glauber dynamics other than the Gibbs Sampler are interesting as well. The analysis in this paper can also be carried out, for instance, for the Metropolis algorithm,[12] leading essentially to the same results. Comparison of different dynamics is interesting because some dynamics are known to converge faster to equilibrium than others in the case of homogeneous energy.[10]

The layout of this paper is as follows. In Section 1 we define the model, state the theorems, and give a heuristic explanation. In Section 2 we set up the mathematical framework and introduce the *mean-field* version of the coupled system, where the dynamics evolve according to "effective medium" transitions. In Sections 3 and 4 we prove our estimates of the coupling time for the mean-field model. In Section 5 we show how these estimates carry over to the random-energy dynamics.

Coupling of random processes in random media is an area with very few rigorous results. The reader of ref. 4 and the present paper will appreciate the complexity of the problem, even for a toy model like the one under consideration. We hope that the mathematical ideas and techniques developed here will be useful in the study of other, more realistic situations. The fact that even this simple model exhibits two interesting dynamical phase transitions seems to indicate what type of behavior can be expected when running Glauber dynamics on real models.

1. MODEL AND THEOREMS

1.1. Random Energy

Consider a system of N Ising spins and let $\Sigma = \{\sigma = (\sigma_1, ..., \sigma_N) \in \{-, +\}^N\}$ be the set of configurations. Assign to each $\sigma \in \Sigma$ independently a random energy $H(\sigma)$, taking the values 0 and $-\log N$ with probabilities $1 - N^{-\gamma}$ and $N^{-\gamma}$, respectively, where $\gamma > 0$ is a density parameter. The function $H: \Sigma \to \{0, -\log N\}$ will be called the *random medium*. Its probability law, the corresponding product measure on Σ, is denoted by $\bar{P}$. Think of Σ as the N-dimensional 2-cube. Configurations $\sigma \in \Sigma$ with $H(\sigma) = 0$ will be called *white sites* (W), those with $H(\sigma) = -\log N$ *black sites* (B).

1.2. Random Dynamics

Given H, let $(\sigma(t))_{t \geq 0}$ be the discrete-time Markov chain on Σ evolving according to the following rule. At times $t = 1, 2, ...$ draw independently two random variables $i(t)$ and $u(t)$ uniformly from $\{1, ..., N\}$, resp. $(0, 1]$. Update $\sigma(t)$ to $\sigma(t+1)$ given by

$$\begin{aligned} \sigma(t+1) &= \sigma(t) && \text{if} \quad u(t) > p(\sigma(t), i(t)) \\ &= \sigma^{i(t)}(t) && \text{if} \quad u(t) \leqslant p(\sigma(t), i(t)) \end{aligned} \tag{1.1}$$

where σ^i is the spin configuration obtained from σ by flipping spin i, and

$$p(\sigma, i) = (1 + e^{\beta[H(\sigma^i) - H(\sigma)]})^{-1} \tag{1.2}$$

with $\beta > 0$ the inverse temperature. Denote by $\tilde{P}_\mu^H$ the measure of $(\sigma(t))_{t \geq 0}$ on the trajectory space $\Sigma^{\mathbb{N}}$ for the given H when $\sigma(0)$ is drawn from Σ according to μ. This process is the usual *Glauber dynamics*, sometimes called the *Gibbs Sampler*. It is ergodic and reversible with respect to the Gibbs measure

$$\pi_H(\sigma) = Z_H^{-1} \exp[-\beta H(\sigma)] \tag{1.3}$$

where Z_H is the normalizing constant. An equivalent way of doing the dynamics is to update

$$\begin{aligned}\sigma(t+1) &= \sigma^{i(t)+}(t) \qquad \text{if} \quad u(t) > q(\sigma(t), i(t)) \\ &= \sigma^{i(t)-}(t) \qquad \text{if} \quad u(t) \leqslant q(\sigma(t), i(t))\end{aligned} \tag{1.4}$$

where σ^{i+} and σ^{i-} are the spin configurations obtained from σ by putting $\sigma_i = +$, resp. $\sigma_i = -$, and

$$q(\sigma, i) = (1 + e^{\beta\sigma_i[H(\sigma^i) - H(\sigma)]})^{-1} \tag{1.5}$$

This formulation is more convenient for the coupling.

1.3. Coupling

Given H, let

$$(x(t))_{t \geqslant 0} = (\sigma(t), \xi(t))_{t \geqslant 0} \tag{1.6}$$

be the discrete-time Markov chain on Σ^2 obtained by running two Glauber dynamics $\sigma(t)$ and $\xi(t)$ according to (1.4) and (1.5) with the *same* random variables $i(t)$ and $u(t)$. This will be called *coupled Glauber dynamics*. Denote by $\tilde{P}^H_{\mu \times \nu}$ the measure of $(x(t))_{t \geqslant 0}$ on the trajectory space $(\Sigma^2)^{\mathbb{N}}$ for the given H when $x(0)$ is drawn from Σ^2 according to $\mu \times \nu$.

Let

$$\tau = \inf\{t \geqslant 0 \colon \sigma(t) = \xi(t)\} \tag{1.7}$$

be the first hitting time of the two components. The well-known *coupling inequality* bounds the total variation between $\tilde{P}^H_\mu$ and $\tilde{P}^H_\nu$,[8]

$$\frac{1}{2} \sum_{\eta \in \Sigma} |\tilde{P}^H_\mu(\sigma(t) = \eta) - \tilde{P}^H_\nu(\xi(t) = \eta)| \leqslant \tilde{P}^H_{\mu \times \nu}(\tau > t) \tag{1.8}$$

Thus, if $\nu = \pi_H$ given by (1.3), then the tail of the distribution of τ gives an upper bound on the rate of convergence of the Glauber dynamics to equilibrium starting from μ.

Inequality (1.8) is valid for any coupling, but its strength depends on the specific type. There always exists an *optimal coupling* that achieves equality, but this coupling is generally non-Markovian and usually intractable. It is believed that most natural couplings roughly achieve equality.[1,7,8] The above coupling, (1.4)–(1.6), should be no exception. In this paper we study the r.h.s. of (1.8) for $\mu = \nu = \pi_H$. The choice of starting both components in equilibrium is technically convenient but not essential.

Qualitatively our estimates should not depend on μ and ν as long as the two components start well apart. So, from now on we assume that, given H, $\sigma(0)$ and $\xi(0)$ are drawn independently according to π_H. We suppress H and the starting measure from the notation. Throughout the paper we use the symbol P to denote the *joint* probability measure for the random medium and the coupled Glauber dynamics.

1.4. Theorems

Define the Hamming distance

$$d(\sigma, \xi) = \sum_{i=1}^{N} 1\{\sigma_i \neq \xi_i\} \tag{1.9}$$

Define the *distance process*

$$(d(t))_{t \geqslant 0} = (d(x(t)))_{t \geqslant 0} \tag{1.10}$$

on the state space $D = \{0, 1,..., N\}$. Clearly

$$P(\tau > t) = P(d(t) \neq 0) \tag{1.11}$$

because once the two components meet the coupling keeps them together afterward. The trouble with the distance process is that it is non-Markovian because the random medium H is not constant.

Theorems 1, 4, and 5 below give estimates of τ in the limit as $N \to \infty$. Different behavior is found in the following three regimes:

I: $\gamma > 1$.

II: $\gamma < 1$, $\beta < \gamma$.

III: $\gamma < 1$, $\beta > \gamma$.

The following theorem is proved in ref. 4.

Theorem 1. In regime I, for any $\varepsilon > 0$

$$\lim_{N \to \infty} P\left(\left|\frac{\tau}{N \log N} - 1\right| < \varepsilon\right) = 1 \tag{1.12}$$

Consider also the first time when $d(t)$ drops down to level $M > 0$, i.e.,

$$\tau(M) = \inf\{t \geqslant 0\colon d(t) = M\} \tag{1.13}$$

[Note that our choice of starting point $(\sigma(0), \xi(0))$ has the property that $d(0)/N \to 1/2$ with P-probability 1 as $N \to \infty$]. The following four theorems are our main results. Theorems 2 and 3 apply to the *mean-field version* of our model (defined in Section 2.2), where the coupled dynamics evolve according to "effective medium" transitions, i.e., with a 1-step transition kernel that is the *average over H* of the 1-step transition kernel controlling the coupled dynamics defined in (1.6). We denote by P^{mf} the corresponding probability measure on the trajectory space $(\Sigma^2)^{\mathbb{N}}$. Theorems 4 and 5, on the other hand, apply to the *random medium dynamics* corresponding to P.

Theorem 2. In regime II:

(i) For any $\delta > 1$ there exist $0 < K_1 < K_2 < \infty$ such that

$$\lim_{N \to \infty} P^{mf}(K_1 N < \tau(\delta N^{1-(\gamma-\beta)}) < K_2 N) = 1 \tag{1.14}$$

(ii) For any $\delta < 1$ there exist $0 < K_1 < K_2 < \infty$ such that

$$\lim_{N \to \infty} P^{mf}(e^{K_1 N^{1-\gamma}} < \tau(\delta N^{1-(\gamma-\beta)}) < e^{K_2 N^{1-\gamma}}) = 1 \tag{1.15}$$

Theorem 3. In regime III, for any $\delta < 1/3$ there exist $0 < K_1 < K_2 < \infty$ such that

$$\lim_{N \to \infty} P^{mf}(e^{K_1 N^{1-\gamma}} < \tau(\delta N) < e^{K_2 N^{1-\gamma}}) = 1 \tag{1.16}$$

Theorem 4. In regime II:

(i) (a) If $2\gamma - \beta < 1$, then the same result as in Theorem 2(i) holds under P.

(b) If $2\gamma - \beta \leqslant 1$, then for any $\varepsilon > 0$ there exists $0 < K_2 < \infty$ such that $\lim_{N \to \infty} P(\tau(N^{\gamma+\varepsilon}) < K_2 N)) = 1$.

(ii) If $2\gamma - \beta < 1$, then for any $\delta < 1$ there exists $0 < K_1 < \infty$ such that $\lim_{N \to \infty} P(e^{K_1 N^{1-(2\gamma-\beta)}} < \tau(\delta N^{1-(\gamma-\beta)})) = 1$

Theorem 5. In regime III, the same result as in Theorem 3 holds under P.

In regime III, Theorem 5 provides us with a complete extension of Theorem 3 showing that the coupling time is stretched exponential. In regime II, however, Theorem 4 is weaker than Theorem 2. Namely, when $2\gamma - \beta < 1$ we again find a sudden slowdown at an intermediate distance scale, but when $2\gamma - \beta \geqslant 1$ we lose most of the information. The reason for this is a breakdown of certain large-deviation estimates in Section 5. We have no doubt that the restrictions are artificial, but the technical complications needed to remove them are substantial.

1.5. Discussion

In order to heuristically explain our results, consider first the homogeneous medium where $H(\sigma)=0$ for all $\sigma\in\Sigma$ (which corresponds to $\gamma=\infty$). In this case $d(t)$ is Markov with transition probabilities

$$\begin{aligned} P(d(t+1)=d\,|\,d(t)=d)&=1-\frac{d}{N}\\ P(d(t+1)=d-1\,|\,d(t)=d)&=\frac{d}{N}\end{aligned}\tag{1.17}$$

Indeed, $q(\sigma, i)=1/2$ for all σ and i [recall (1.5)], and so at time t the coupling forces the spins with index $i(t)$ in the configurations $\sigma(t)$ and $\xi(t)$ to become identical [recall (1.4)]. Spins already identical stay identical: $d(t)/N$ is the probability that $i(t)$ picks spins that are not yet identical. Equation (1.17) says that $d(t)$ is a death chain. Standard estimates yield $\tau/N\log N\to 1$ with P-probability 1 as $N\to\infty$. The $N\log N$ scale comes from

$$E(\tau)=E\left(\sum_{d=1}^{d(0)}\frac{N}{d}\right)\sim NE(\log d(0))\sim N\log\frac{N}{2}\tag{1.18}$$

In order to understand the dynamics in the random medium it will be helpful to draw *coupling diagrams* as shown in Fig. 1. The first line is for the σ-component and lists the spin value $\sigma_{i(t)}(t)$ and the colors $H(\sigma(t))$ and $H(\sigma^{i(t)}(t))$. The interval $(0, 1]$ is split into two parts, separated at $q(\sigma(t), i(t))$ which is computable from (1.5). According to (1.4), if $u(t)$ falls to the left of $q(\sigma(t), i(t))$, then the spin $i(t)$ is set to $-$, otherwise to $+$. This allows us to read off the transition probabilities for the spin $i(t)$. The second line is for the ξ-component. By reading the two lines together we can read off the transition probabilities for the distance $d(t)$.

For instance, if $H(\sigma)=0$ for all $\sigma\in\Sigma$, then the diagram is as shown in Fig. 2, from which (1.17) easily follows by considering the four possible values for $(\sigma_{i(t)}(t), \xi_{i(t)}(t))$.

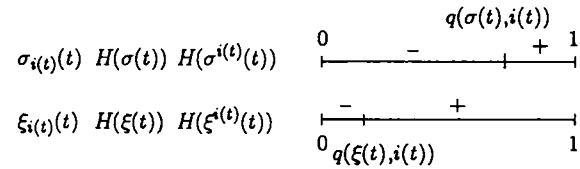

Fig. 1

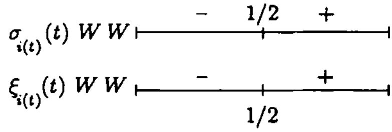

Fig. 2

Regime I ($\gamma > 1$). Theorem 1 states that when $\gamma > 1$ the same qualitative behavior is found in the random medium as in the homogeneous medium. This can be explained as follows. In addition to the distance process (1.10), define the *color pair process*

$$(c(t))_{t \geqslant 0} = (H(\sigma(t)), H(\xi(t)))_{t \geqslant 0} \tag{1.19}$$

on the state space $C^2 = \{W, B\}^2$. (Recall from Section 1.1 that the energy levels 0 and $-\log N$ are called white and black.) Each site $\sigma \in \Sigma$ has N neighbors and each neighbor has probability $N^{-\gamma}$ to be black. Therefore, when $\gamma > 1$ a typical site has only white neighbors and the average time needed for either of the components to hit a black site for the first time is of the order $N^\gamma \gg N \log N$. This implies that typically $c(t) = WW$ up to a time large enough for the components to meet, as if H were constant and $d(t)$ were decreasing according to (1.17).

Regimes II and III ($\gamma < 1$). The situation is drastically different when $\gamma < 1$. In this case typically one of the components hits a black site after time $N^\gamma \ll N \log N$ and gets stuck. If the other component is still white, then the corresponding diagram at this stage is as shown in Fig. 3 (most neighbors are white), with

$$\begin{aligned} q(\sigma(t), i(t)) &= N^\beta/(N^\beta + 1) \qquad \text{if} \quad \sigma_{i(t)}(t) = - \\ &= 1/(N^\beta + 1) \qquad \text{if} \quad \sigma_{i(t)}(t) = + \end{aligned} \tag{1.20}$$

or the same diagram with σ and ξ reversed. Hence $c(t) \in \{BW, WB\}$ for a time of order N^β. Consequently, one component being stuck and the other being mobile, *the two components tend* to *drift apart because Σ is very high dimensional.* From Fig. 3 we compute that for large N the transition probabilities for $d(t)$ are roughly given by

$$\begin{aligned} P(d(t+1) = d-1 \mid d(t) = d) &\approx \frac{1}{2}\frac{d}{N} \\ P(d(t+1) = d \mid d(t) = d) &\approx \frac{1}{2} \\ P(d(t+1) = d+1 \mid d(t) = d) &\approx \frac{1}{2}\left(1 - \frac{d}{N}\right) \end{aligned} \tag{1.21}$$

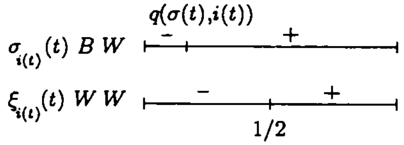

Fig. 3

so $d(t)$ indeed tends to drift in the direction of $N/2$. Thus we see that the presence of the black sites is felt in that it tends to increase the distance (as opposed to decrease when the components are both on white sites). Of course, it is also possible that both components get stuck simultaneously in a black site for a time of order N^β. In that case $c(t) = BB$, but $d(t)$ does not change.

Hence we conclude that when $\gamma < 1$ there is a *competition* between $c(t) = WW$ and $c(t) \in \{BW, WB, BB\}$, which corresponds to $d(t)$ competing between (1.17) driving it down and (1.21) driving it up. This explains why in regimes II and III the coupling is slower than in regime I. To explain where the difference between Theorems 2 and 3 (resp. Theorems 4 and 5) comes from, we argue as follows.

Regime II ($\gamma < 1$, $\beta < \gamma$). If $\beta < \gamma$, then $N^\beta \ll N^\gamma$ and so much more time is spent on white sites than on black sites. Hence, $c(t) = WW$ occurs more often than $c(t) \in \{BW, WB, BB\}$, so that $d(t)$ mostly follows (1.17). As a result, $d(t)$ drops below εN for any $\varepsilon > 0$ in a time of order N, just as in the homogeneous medium where

$$E(\tau(\varepsilon N)) = E\left(\sum_{d=\varepsilon N+1}^{d(0)} \frac{N}{d}\right) \sim NE\left(\log \frac{d(0)}{\varepsilon N}\right) \sim N \log \frac{1}{2\varepsilon} \tag{1.22}$$

Theorem 2 says that $d(t)$ in fact drops down to level $N^{1-(\gamma-\beta)}$ in a time of order N and then gets blocked. To explain why, we observe that some sort of *balance is struck* at this level, namely: (1) the average decrease of $d(t)$ is $N^{-(\gamma-\beta)}N^\gamma = N^\beta$ over the time intervals of length N^γ where $c(t) = WW$ [as in (1.17)]; (2) the average increase is N^β over the time intervals of length N^β where $c(t) \in \{BW, WB, BB\}$ [as in (1.21)].

It is only after $d(t)$ drops below level $N^{1-(\gamma-\beta)}$ that the coupling slows down considerably and that a very long time is needed for $d(t)$ to move down further. Indeed, starting from $\delta N^{1-(\gamma-\beta)}$, $d(t)$ decreases on the average by $\delta N^{-(\gamma-\beta)}N^\gamma = \delta N^\beta$ when following (1.17) over an average time N^γ. On the other hand, $d(t)$ increases on the average by N^β when following (1.21) over an average time N^β. Hence there is a drift upward when $\delta < 1$. In order for $d(t)$ to drop a total of $\varepsilon N^{1-(\gamma-\beta)}$ the system has to essentially wait until it happens to spend a time $\varepsilon N^{1-(\gamma-\beta)}/N^{-(\gamma-\beta)} = \varepsilon N$ doing (1.17). This event occurs after a time of order $(1-N^{-\gamma})^{\varepsilon N} \approx \exp(\varepsilon N^{1-\gamma})$, on a rough scale.

Regime III ($\gamma < 1$, $\beta > \gamma$). If $\beta > \gamma$, then $N^{\beta} \gg N^{\gamma}$ and so now much more time is spent on black sites than on white sites. Hence, $c(t) \in \{BW, WB, BB\}$ occurs more often than $c(t) = WW$, so that $d(t)$ mostly follows (1.21). The coupling is extremely slow and Theorem 3 shows that a stretched-exponential length of time is needed for $d(t)$ to even drop below δN for any $\delta < 1/3$.

To understand the estimate in Theorem 3, observe that if $\beta > \gamma$, then in order for $d(t)$ to drop a total of εN the system has to essentially wait until it happens to do a white run with $c(t) = WW$ [as in (1.17)] over a time interval of length εN. Since $1 - N^{-\gamma}$ typically is the probability that a white site is hit when a component moves, it takes a time of order $(1-N^{-\gamma})^{\varepsilon N} \approx \exp(\varepsilon N^{1-\gamma})$ for this run to occur, on a rough scale.

2. MARKOV STRUCTURE AND MEAN-FIELD MODEL

In this section we set up the probabilistic framework and introduce the mean-field version of the coupled dynamics.

2.1. Markov Structure

Suppose H is fixed. Define the *noncoincidence set* of $\sigma, \xi \in \Sigma$ by

$$I(\sigma, \xi) = \{i \in \{1,..., N\}: \sigma_i \neq \xi_i\} \tag{2.1}$$

For $\sigma, \xi \in \Sigma$ and $a, b \in C$ define

$$\begin{aligned} V_1((\sigma, \xi), (a, b)) &= \#\{i \in I(\sigma, \xi): H(\sigma^i) = a, H(\xi^i) = b\} \\ V_2((\sigma, \xi), (a, b)) &= \#\{i \notin I(\sigma, \xi): H(\sigma^i) = a, H(\xi^i) = b\} \end{aligned} \tag{2.2}$$

i.e., the numbers of indices i inside, resp. outside, the noncoincidence set of σ and ξ with the property that the colors a and b are seen when the spins indexed i are flipped. These numbers control the dynamics as follows.

We shall be monitoring the *process of color pair and distance* on the state space $C^2 \times D$ [recall (1.10) and (1.19)]

$$(r(t))_{t \geq 0} = (c(t), d(t))_{t \geq 0} \tag{2.3}$$

The problem is that this process is non-Markovian because H is not

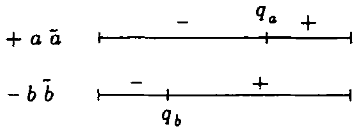

Fig. 4

constant. In order to describe its behavior, we consider the conditional transition probability [recall (1.6)]

$$P\big(r(t+1)=r' \mid r(t)=r,\, x(t)=x\big) \tag{2.4}$$

The first question we ask is what information contained in $x(t)=x\in\Sigma^2$ is needed to compute (2.4). In other words, what additional information besides $r(t)=r$ determines the transition to $r(t+1)=r'$? The answer comes in two lemmas. Define

$$I(t)=I(\sigma(t),\, \xi(t)) \tag{2.5}$$

and

$$\tilde{c}(t)=(H(\sigma^{i(t)}(t)),\, H(\xi^{i(t)}(t))) \tag{2.6}$$

Lemma 1. Given $r(t)$, the transition $r(t)\to r(t+1)$ only depends on $\tilde{c}(t)$ and on whether $i(t)\in I(t)$ or $i(t)\notin I(t)$.

Proof. If $i(t)\in I(t)$, then $(\sigma_{i(t)}(t),\, \xi_{i(t)}(t))=(+,-)$ or $(-,+)$. Suppose $(+,-)$. Write

$$\begin{aligned} c(t)&=(a,b)\\ \tilde{c}(t)&=(\tilde{a},\tilde{b})\\ q(\sigma(t),\, i(t))&=q_a\\ q(\xi(t),\, i(t))&=q_b \end{aligned} \tag{2.7}$$

Then we have the diagram shown in Fig. 4 (we pick the case $q_a>q_b$ as an example). If $r=((a,b),d)$, then we can read off

$$\begin{aligned} r'&=((\tilde{a},b),d-1) && \text{w.p.}\quad q_b\\ &=((\tilde{a},\tilde{b}),d) && \text{w.p.}\quad q_a-q_b\\ &=((a,\tilde{b}),d-1) && \text{w.p.}\quad 1-q_a \end{aligned} \tag{2.8}$$

Suppose $(-,+)$. Then we have the diagram shown in Fig. 5. Inspection shows that we obtain the same scheme as (2.8). We thus see that the coupling is symetric for $(+,-)$ and $(-,+)$.

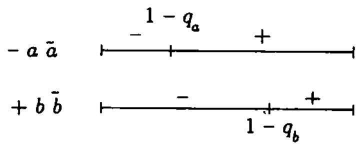

Fig. 5

A similar argument holds if $i(t) \notin I(t)$. The coupling is symmetric for $(+, +)$ and $(-, -)$. ∎

Next we define the following family of transition kernels on $C^2 \times D$, indexed by $k = 1, 2$ and $\tilde{c} \in C^2$:

$$\begin{aligned} P_1^{\tilde{c}}(r, r') &= P\big(r(t+1) = r' \mid r(t) = r,\, i(t) \in I(t),\, \tilde{c}(t) = \tilde{c}\big) \\ P_2^{\tilde{c}}(r, r') &= P\big(r(t+1) = r' \mid r(t) = r,\, i(t) \notin I(t),\, \tilde{c}(t) = \tilde{c}\big) \end{aligned} \tag{2.9}$$

According to Lemma 1, these kernels control the transition $r(t) \to r(t+1)$.

Lemma 2. For all x, r, and r'

$$P\big(r(t+1) = r' \mid r(t) = r,\, x(t) = x\big) = \sum_{k=1,2} \sum_{\tilde{c} \in C^2} \frac{1}{N} V_k(x, \tilde{c})\, P_k^{\tilde{c}}(r, r') \tag{2.10}$$

Proof. Condition on $i(t) \in I(t)$, $\tilde{c}(t) = \tilde{c}$, resp. on $i(t) \notin I(t)$, $\tilde{c}(t) = \tilde{c}$. Apply Lemma 1 and note that by (2.2)

$$\begin{aligned} P\big(i(t) \in I(t),\, \tilde{c}(t) = \tilde{c} \mid x(t) = x\big) &= \frac{1}{N} V_1(x, \tilde{c}) \\ P\big(i(t) \notin I(t),\, \tilde{c}(t) = \tilde{c} \mid x(t) = x\big) &= \frac{1}{N} V_2(x, \tilde{c}) \quad ∎ \end{aligned} \tag{2.11}$$

The difference between $k = 1$ and $k = 2$ is that $d(t)$ cannot increase when $i(t)$ is inside the noncoincidence set and cannot decrease when $i(t)$ is outside. The transition kernels $P_k^{\tilde{c}}(r, r')$ are tabulated in Table I. There are eight kernels indexed by $k = 1, 2$ and $\tilde{c} \in C^2 = \{WW, BW, WB, BB\}$, and each kernel corresponds to four rows of the table. Each row can be computed from a coupling diagram. Take, for example, the row corresponding to $k = 1$, $\tilde{c} = BW$, and $r = WWd$. Draw the diagram shown in Fig. 6. As

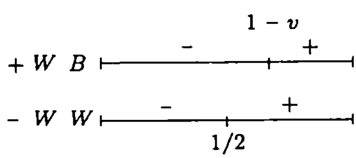

Fig. 6

Table IA. The Transition Kernel $P_1^{\tilde{c}}(r, r')$

$\tilde{c}$	$r \backslash r'$	WWd	$WWd-1$	BWd	$BWd-1$	WBd	$WBd-1$	BBd	$BBd-1$
WW	WWd		1						
	BWd		v	$\frac{1}{2}-v$	$\frac{1}{2}$				
	WBd		v			$\frac{1}{2}-v$	$\frac{1}{2}$		
	BBd				v		v	$1-2v$	
BW	WWd		v	$\frac{1}{2}-v$	$\frac{1}{2}$				
	BWd				1				
	WBd		v						$1-v$
	BBd				v			$\frac{1}{2}-v$	$\frac{1}{2}$
WB	WWd		v			$\frac{1}{2}-v$	$\frac{1}{2}$		
	BWd		v						$1-v$
	WBd						1		
	BBd						v	$\frac{1}{2}-v$	$\frac{1}{2}$
BB	WWd				v		v	$1-2v$	
	BWd				v			$\frac{1}{2}-v$	$\frac{1}{2}$
	WBd						v	$\frac{1}{2}-v$	$\frac{1}{2}$
	BBd								1

explained in Section 1.5, the first line displays the spin value $\sigma_{i(t)}(t)$, the colors $H(\sigma(t))$ and $H(\sigma^{i(t)}(t))$, and a graph for the value of $q(\sigma(t), i(t))$, with the abbreviation $v = (1 + N^{\beta})^{-1}$. The second line does the same for the ξ-component. The coupling uses a common value of $u(t)$ in the two graphs. The diagram allows us to read off the corresponding entries in Table I:

(a) w.p. 1/2 only $\sigma(t)$ flips, yielding:
$(-, -)$, $c(t+1) = BW$, $d(t+1) = d(t) - 1$.

(b) w.p. $1/2 - v$ both $\sigma(t)$ and $\xi(t)$ flip, yielding:
$(-, +)$, $c(t+1) = BW$, $d(t+1) = d(t)$.

(c) w.p. v only $\xi(t)$ flips, yielding:
$(+, +)$, $c(t+1) = WW$, $d(t+1) = d(t) - 1$.

We remark that the symmetry in the coupling produces the same entries when $+$ and $-$ change places. Fig. 7 shows one more diagram, corre-

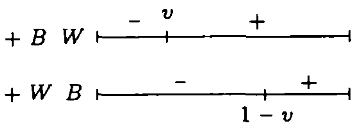

Fig. 7

Table IB. The Transition Kernel $P_2^{\tilde{c}}(r, r')$

$\tilde{c}$	$r \backslash r'$	WWd	$WWd-1$	BWd	$BWd-1$	WBd	$WBd-1$	BBd	$BBd-1$
WW	WWd	1							
	BWd	v		$\frac{1}{2}$	$\frac{1}{2}-v$				
	WBd	v				$\frac{1}{2}$	$\frac{1}{2}-v$		
	BBd	v						$1-v$	
BW	WWd	v		$\frac{1}{2}$	$\frac{1}{2}-v$				
	BWd			1					
	WBd			v		v			$1-2v$
	BBd			v				$\frac{1}{2}$	$\frac{1}{2}-v$
WB	WWd	v				$\frac{1}{2}$	$\frac{1}{2}-v$		
	BWd			v		v			$1-2v$
	WBd					1			
	BBd					v		$\frac{1}{2}$	$\frac{1}{2}-v$
BB	WWd	v						$1-v$	
	BWd			v				$\frac{1}{2}$	$\frac{1}{2}-v$
	WBd					v		$\frac{1}{2}$	$\frac{1}{2}-v$
	BBd							1	

sponding to $k=2$, $\tilde{c}=WB$, and $r=BWd$. The reader is asked to check the resulting entries in Table I.

To summarize this section, Lemma 2 and Table I describe how the process $(r(t))_{t \geqslant 0}$ defined in (2.3) evolves. The dependence on the underlying position $(x(t))_{t \geqslant 0}$, defined in (1.6), is through the x-dependence of the numbers $V_k(x, \tilde{c})$ defined in (2.2), which depend on the random medium H.

2.2. Mean-Field Model

The *mean-field approximation* consists of replacing $V_k(x, \tilde{c})$ in (2.10) by its average value $\bar{E}V_k(x, \tilde{c})$ over the random medium. For every $\sigma, \xi \in \Sigma$ and $a, b \in C$ such that $d(\sigma, \xi) \neq 0$, N we have

$$\begin{aligned} \bar{E}V_1((\sigma, \xi), (a, b)) &= d(\sigma, \xi)\, p(a)\, p(b) \\ \bar{E}V_2((\sigma, \xi), (a, b)) &= (N - d(\sigma, \xi))\, p(a)\, p(b) \end{aligned} \tag{2.12}$$

with the abbreviation $p(W) = 1 - N^{-\gamma}$ and $p(B) = N^{-\gamma}$. This follows by noting that the coloring is i.i.d. and that $\# I(\sigma, \xi) = d(\sigma, \xi)$. Thus, in the mean-field model the process $(r(t))_{t\geqslant 0}$ defined in (2.3) is Markov with state space $C^2\times D$ and with transition kernel

$$Q(r, r') = \sum_{\tilde{c}\in C^2} p(\tilde{c})\left\{\frac{d}{N}P_1^{\tilde{c}}(r, r') + \left(1-\frac{d}{N}\right)P_2^{\tilde{c}}(r, r')\right\} \tag{2.13}$$

between the states $r=(c, d)$ and $r'=(c', d')$, where we use the abbreviation $\tilde{c}=(\tilde{a}, \tilde{b})$ and $p(\tilde{c})=p(\tilde{a})\,p(\tilde{b})$. Of the original dependence on $x(t)$ in (2.10) only the dependence on the distance $d(t)=d(x(t))$ is left over in (2.13).

Table II gives the entries $Q(r, r')$ computed from Table I via (2.13). We use the abbreviations

$$\Delta=\frac{d}{N}, \qquad v=(1+N^{\beta})^{-1}, \qquad w=N^{-\gamma} \tag{2.14}$$

The last column in Table II lists $Q(r, r')$ to *leading order* in v and w (as $N\to\infty$), which means all terms up to and including order $N^{-\beta}$ and $N^{-\gamma}$. Note that some zeros in the last column correspond to transitions with probability $o(N^{-\beta}\wedge N^{-\gamma})$.

In the analysis in Sections 3 and 4 we shall at first ignore all transitions that do not appear in the leading-order column of Table II. At the end of Sections 3.1 and 4.1 we shall see that these transitions can be easily taken care of as small *perturbations* of the leading-order transitions.

2.3. Stochastic Domination for Mean-Field Model

Table II has too many entries to see clearly what is going on. The main idea in our approach to estimate coupling times is to sandwich the distance process $(d(t))_{t\geqslant 0}$ between two processes $(d_*(t))_{t\geqslant 0}$ and $(d^*(t))_{t\geqslant 0}$ on the same state space D but with a more accessible evolution. The construction will be done in such a way that if we define, as in (1.13),

$$\begin{aligned}\tau_*(M)&=\inf\{t\geqslant 0: d_*(t)=M\}\\ \tau^*(M)&=\inf\{t\geqslant 0: d^*(t)=M\} \qquad (0\leqslant M\leqslant N)\end{aligned} \tag{2.15}$$

then for N sufficiently large and for appropriate choice of M

$$\tau_*(M)\prec\tau(M)\prec\tau^*(M) \tag{2.16}$$

[The symbol $\prec$ means *stochastically smaller*, i.e., $X\prec Y$ when $P(X>k)\leqslant P(Y>k)$ for all k. This notion is equivalent to the existence of an ordered coupling, i.e., $X\prec Y$ iff there exists a coupling such that $P(X\leqslant Y)=1$.] By

Table II. Mean-Field Markov Kernel $Q(r, r')$

c	c'	$d'-d$	$Q(r, r')$	Leading order
WW	WW	$+1$	0	0
		0	$(1-\Delta)[vw(2-w)+(1-w)^2]$	$(1-\Delta)(1-2w)$
		-1	$\Delta[2vw(1-w)+(1-w)^2]$	$\Delta(1-2w)$
	BW/WB	$+1$	$(1-\Delta)(\frac{1}{2}-v)\,w(1-w)$	$\frac{1}{2}(1-\Delta)w$
		0	$\frac{1}{2}w(1-w)-\Delta vw(1-w)$	$\frac{1}{2}w$
		-1	$\Delta[vw^2+\frac{1}{2}w(1-w)]$	$\frac{1}{2}\Delta w$
	BB	$+1$	0	0
		0	$(1-v)\,w^2-\Delta vw^2$	0
		-1	0	0
BB	WW	$+1$	0	0
		0	$(1-\Delta)\,v(1-w)^2$	$(1-\Delta)v$
		-1	0	0
	BW/WB	$+1$	0	0
		0	$(1-\Delta)\,vw(1-w)$	0
		-1	$\Delta v(1-w)$	Δv
	BB	$+1$	$(1-\Delta)(\frac{1}{2}-v)\,2w(1-w)$	$(1-\Delta)w$
		0	$1-v(1-w)^2-w(1-w)-\Delta[v(1-w^2)+w^2]$	$1-(1+\Delta)v-w$
		-1	Δw	Δw
BW	WW	$+1$	0	0
		0	$(1-\Delta)\,v(1-w)^2$	$(1-\Delta)v$
		-1	$\Delta v(1-w)$	Δv
	BW	$+1$	$(1-\Delta)(\frac{1}{2}-v)(1-w)^2$	$(1-\Delta)(\frac{1}{2}-v-w)$
		0	$vw+\frac{1}{2}(1-w^2)-\Delta[vw^2+(v+w)(1-w)]$	$\frac{1}{2}-\Delta(v+w)$
		-1	$\Delta[vw^2+\frac{1}{2}(1-w^2)]$	$\frac{1}{2}\Delta$
	WB	$+1$	0	0
		0	$(1-\Delta)\,vw(1-w)$	0
		-1	0	0
	BB	$+1$	$(1-\Delta)[(\frac{1}{2}-v)\,w^2+(1-2v)\,w(1-w)]$	$(1-\Delta)w$
		0	$\frac{1}{2}w^2-\Delta vw^2$	0
		-1	$\Delta[(1-v)\,w(1-w)+\frac{1}{2}w^2]$	Δw
WB			By symmetry	

subsequently estimating $\tau_*(M)$ and $\tau^*(M)$ we shall be able to prove lower and upper bounds for $\tau(M)$.

In Sections 3 and 4 we give the proofs of Theorems 2 and 3 for the mean-field coupling defined by (2.13). The proofs are based on a sequence of technical lemmas involving coupling arguments (Sections 3.1 and 4.1) and large-deviation estimates (Sections 3.2 and 4.2). Much of the work lies here. *The mean-field coupling has all the essential ingredients and is complex enough to warrant a separate treatment.* In Section 5 we shall see how to handle the random medium described by (2.10) and how to prove Theorems 4 and 5. This will involve some additional large-deviation estimates, but most of the work in Sections 3 and 4 will carry over.

3. PROOF OF LOWER BOUNDS FOR MEAN-FIELD MODEL

3.1. Lower Stochastic Domination

Referring to the kernel Q in Table II, we first note that by symmetry all paths through BW have the same probability when rerouted through WB. Hence we may identify BW and WB into a single state BW/WB and define a new kernel Q^1 on $\{WW, BB, BW/WB\} \times D$ by

$$\begin{aligned} Q^1(r, (BW/WB, \cdot)) &= 2Q(r, (BW, \cdot)) \quad \text{if } r = (WW, \cdot), (BB, \cdot) \\ Q^1((BW/WB, \cdot), r') &= 2Q((BW, \cdot), r') \quad \text{if } r' = (WW, \cdot), (BB, \cdot) \\ Q^1((BW/WB, \cdot), (BW/WB, \cdot)) &= Q((BW, \cdot), (BW, \cdot)) \\ &\quad + Q((BW, \cdot), (WB, \cdot)) \\ Q^1(r, r') &= Q(r, r') \quad \text{otherwise.} \end{aligned} \tag{3.1}$$

We next proceed with the first lower stochastic domination. Namely, we skip all the time spent by the process without changing. For this we define a new kernel Q^2 on $\{WW, BB, BW/WB\} \times D$ by

$$\begin{aligned} Q^2(r, r') &= \frac{Q^1(r, r')}{1 - Q^1(r, r)} \quad &\text{if } r' \neq r \\ &= 0 \quad &\text{if } r' = r \end{aligned} \tag{3.2}$$

Obviously

$$\tau(M) = \tau^1(M) \geqslant \tau^2(M) \tag{3.3}$$

In Table III we display the transition probabilities $Q^2(r, r')$ up to *leading order* as $N \to \infty$.

In Sections 3.1.1 and 3.1.2 we treat regimes III and II, respectively. We introduce a new Markov process $r^3(t) = (c^3(t), d^3(t))$ which has the property that $d^3(t)$ decreases faster than $d^2(t)$ in a certain interval. The

Table III. The Markov Kernel $Q^2(r, r')$

c	c'	$d'-d$	$Q^2(r, r')$
WW	WW	-1	$\dfrac{\Delta(1-2w)}{\Delta+2w(1-\Delta)}$
	BW/WB	$+1$	$\dfrac{(1-\Delta)w}{\Delta+2w(1-\Delta)}=: p_3$
		0	$\dfrac{w}{\Delta+2w(1-\Delta)}=: p_2$
		-1	$\dfrac{\Delta w}{\Delta+2w(1-\Delta)}=: p_1$
BB	WW	0	$\dfrac{(1-\Delta)v}{(1+\Delta)v+w}=: p_{10}$
	BW/WB	-1	$\dfrac{2\Delta v}{(1+\Delta)v+w}=: p_9$
	BB	$+1$	$\dfrac{(1-\Delta)w}{(1+\Delta)v+w}$
		-1	$\dfrac{\Delta w}{(1+\Delta)v+w}=: p_{11}$
BW/WB	WW	0	$\dfrac{(1-\Delta)v}{\frac{1}{2}+\Delta(v+w)}$
		-1	$\dfrac{\Delta v}{\frac{1}{2}+\Delta(v+w)}=: p_8$
	BW	$+1$	$\dfrac{(1-\Delta)(\frac{1}{2}-v-w)}{\frac{1}{2}+\Delta(v+w)}=: p_5$
		-1	$\dfrac{\frac{1}{2}\Delta}{\frac{1}{2}+\Delta(v+w)}=: p_4$
	BB	$+1$	$\dfrac{(1-\Delta)w}{\frac{1}{2}+\Delta(v+w)}=: p_7$
		-1	$\dfrac{\Delta w}{\frac{1}{2}+\Delta(v+w)}=: p_6$

comparison between $d^2(t)$ and $d^3(t)$ is obtained via a *series of coupling diagrams*. In Section 3.2 we shall use $d^3(t)$ to prove lower bounds for $\tau^2(M)$. Together with (3.3) this will give the lower bounds in Theorems 2 and 3.

While reading Section 3.1 the reader should keep in mind the heuristic picture described in Section 1.5.

3.1.1. Regime III ($\gamma < 1$, $\beta > \gamma$). Pick $0 < \delta < 1/2$, $\varepsilon > 0$, and let $r^3(t)$ be the Markov process on $\{1, 2\} \times \mathbb{Z}$ with the following kernel:

$$
\begin{aligned}
Q^3((1,d),(1,d-1)) &= 1 - \frac{2w}{\delta} \\
q_1 := Q^3((1,d),(2,d-1)) &= \frac{2w}{\delta} \\
Q^3((2,d),(1,d-1)) &= 2(1+\varepsilon)w && \text{if } 2\gamma > \beta \\
&= \frac{v}{w} && \text{if } 2\gamma \leqslant \beta \\
q_2 := Q^3((2,d),(2,d-1)) &= [1 - 2(1+\varepsilon)w]\delta && \text{if } 2\gamma > \beta \qquad (3.4) \\
&= \left(1 - \frac{v}{w}\right)\delta && \text{if } 2\gamma \leqslant \beta \\
q_3 := Q^3((2,d),(2,d+1)) &= [1 - 2(1+\varepsilon)w](1-\delta) && \text{if } 2\gamma > \beta \\
&= \left(1 - \frac{v}{w}\right)(1-\delta) && \text{if } 2\gamma \leqslant \beta \\
& Q^3 \text{ zero otherwise}
\end{aligned}
$$

Figure 8 gives the graph of the above transitions. The + and − attached to the arrows denote the change of distance in each transition. The starting point of the process is chosen

$$
\begin{aligned}
r^3(0) &= (1, \delta N) && \text{if } 2\gamma > \beta \\
&= (1, \delta N) && \text{if } 2\gamma \leqslant \beta, \quad c^2(0) = WW, BW/WB \\
&= (2, \delta N) && \text{if } 2\gamma \leqslant \beta, \quad c^2(0) = BB \qquad (3.5)
\end{aligned}
$$

The distinction between the two cases $2\gamma > \beta$ and $2\gamma \leqslant \beta$ is a technical necessity for the proof. The parameters δ, ε are built in to accommodate small perturbations later on (recall the remark made at the end of Section 2.2).

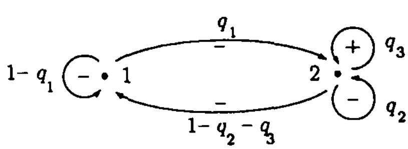

Fig. 8

Lemma 3. For N sufficiently large, as long as

$$\frac{d^2(t)}{N} \in [\delta_1, \delta_2] \qquad \text{with} \quad 0 < \delta_1 < \delta_2 < \delta \tag{3.6}$$

there exists a coupling of $r^2(t)$ and $r^3(t)$ such that

$$d^2(t+1) - d^2(t) \geqslant d^3(t+1) - d^3(t) \qquad \text{a.s.} \tag{3.7}$$

Proof of Lemma 3 for $2\gamma > \beta$. We construct a coupling such that (c^2, c^3) can only appear in the combinations

$$(c^2, c^3) \in \{(WW, 1), (BW/WB, 1), (BB, 1), (BW/WB, 2)\} \tag{3.8}$$

The transition probabilities of the coupled process (r^2, r^3) are given by four diagrams showing the transitions from each of the four pairs (c^2, c^3) in (3.8). The first diagram is shown in Fig. 9, with

$$p_1 + p_2 + p_3 \geqslant q_1 \tag{3.9}$$

The first line is for the r^2-component and shows the transitions from WW to either BW/WB or WW, where $-$, 0, $+$ indicates the variation in the distance d. The probabilities p_1, p_2, p_3 are taken from Table III. The second line is for the r^3-component and shows the transitions from 1 to either 1 or 2, where $-$ indicates that the distance d^3 drops. The probability q_1 comes from (3.4). The coupling is achieved by using the same random variable, drawn uniformly from (0, 1], for both lines.

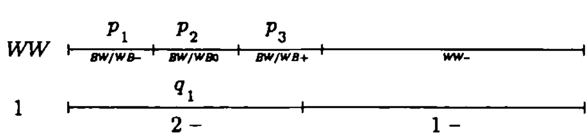

Fig. 9

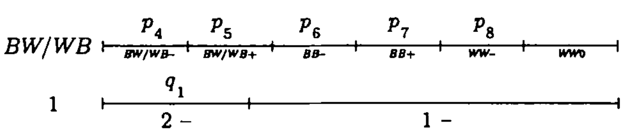

Fig. 10

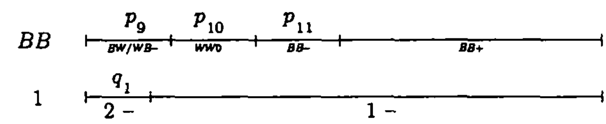

Fig. 11

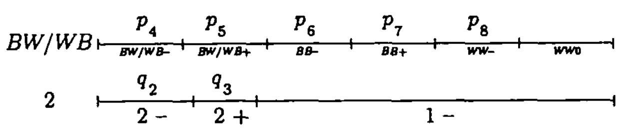

Fig. 12

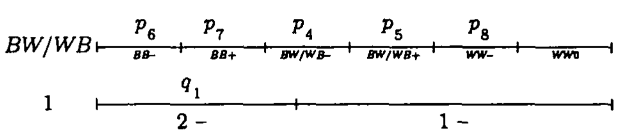

Fig. 13

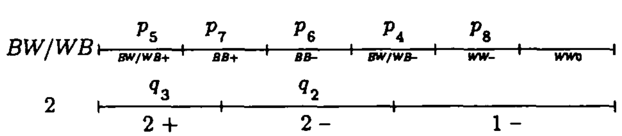

Fig. 14

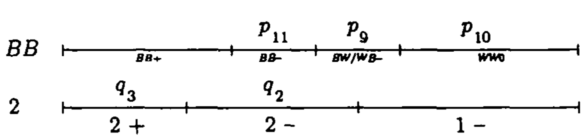

Fig. 15

The other three diagrams are given in Figs. 10–12, with

$$p_4 + p_5 \geqslant q_1 \tag{3.10}$$

$$p_9 \geqslant q_1 \tag{3.11}$$

$$\begin{aligned} p_4 &\leqslant q_2 \\ p_4 + p_5 &\geqslant q_2 + q_3 \end{aligned} \tag{3.12}$$

The conditions (3.9)–(3.12) guarantee that (c^2, c^3) stays on the four combinations given in (3.8), and that the variation in (d^2, d^3) satisfies (3.7). It remains to check that (3.6) implies (3.9)–(3.12) for N sufficiently large. This is easily verified from (3.4) and Table III by using that $2\gamma > \beta > \gamma > 0$ is the same as $w^2 \ll v \ll w \ll 1$. ∎

Proof of Lemma 3 for $2\gamma \leqslant \beta$. This time the four allowed combinations of (c^2, c^3) are

$$(c^2, c^3) \in \{(WW, 1), (BW/WB, 1), (BW/WB, 2), (BB, 2)\} \tag{3.13}$$

The first diagram is the same as Fig. 9 and again requires (3.9). The other three diagrams are given in Figs. 13–15, with

$$p_6 + p_7 \leqslant q_1 \tag{3.14}$$

$$\begin{aligned} p_4 + p_5 + p_6 + p_7 &\geqslant q_1 \\ p_5 + p_7 &\geqslant q_3 \\ p_5 + p_6 + p_7 &\leqslant q_2 + q_3 \\ p_4 + p_5 + p_6 + p_7 &\geqslant q_2 + q_3 \end{aligned} \tag{3.15}$$

$$\begin{aligned} p_{10} &\leqslant 1 - q_2 - q_3 \\ p_9 + p_{10} &\geqslant 1 - q_2 - q_3 \\ p_9 + p_{10} + p_{11} &\leqslant 1 - q_3 \end{aligned} \tag{3.16}$$

Again, (3.9) and (3.14)–(3.16) guarantee that (c^2, c^3) stays on the four combinations given in (3.13), and that the variation in (d^2, d^3) satisfies (3.7). One easily verifies that (3.6) implies (3.9) and (3.14)–(3.16) for N sufficiently large using that $\beta > \gamma > 0$ is the same as $v \ll w \ll 1$.

The condition $2\gamma \leqslant \beta$ is not needed at this point, but will be in Section 3.2. ∎

3.1.2. Regime II ($\gamma < 1$, $\beta < \gamma$). Pick $0 < \delta < 1/2$, $\varepsilon > 0$ and let $r^3(t)$ be given by the following kernel:

$$
\begin{aligned}
&Q^3((1,d),(1,d-1)) = 1 - \frac{2v}{\delta} \\
q_1 :=\ &Q^3((1,d),(2,d-1)) = \frac{2v}{\delta} \\
&Q^3((2,d),(1,d-1)) = 2(1+\varepsilon)v && \text{if} \quad \gamma < 2\beta \\
q_2 :=\ &Q^3((2,d),(2,d-1)) = [1-2(1+\varepsilon)v]\,\frac{\delta w}{v} && \text{if} \quad \gamma < 2\beta \qquad (3.17)\\
q_3 :=\ &Q^3((2,d),(2,d+1)) = [1-2(1+\varepsilon)v]\cdot\left(1-\frac{\delta w}{v}\right) && \text{if} \quad \gamma < 2\beta \\
&Q^3 \text{ zero otherwise}
\end{aligned}
$$

The starting point is

$$r^3(0) = \left(1, \frac{\delta w}{v}\right) \qquad \text{if} \quad \gamma < 2\beta \tag{3.18}$$

Lemma 4. For N sufficiently large, as long as

$$\frac{d^2(t)}{N} \in [\delta_1, \delta_2]\,\frac{w}{v} \qquad \text{with} \quad 0 < \delta_1 < \delta_2 < \delta \tag{3.19}$$

there exists a coupling of $r^2(t)$ and $r^3(t)$ such that

$$d^2(t+1) - d^2(t) \geqslant d^3(t+1) - d^3(t) \qquad \text{a.s.} \tag{3.20}$$

Proof. The coupling is the same as in Figs. 9–12, and again one checks that (3.9)–(3.12) are satisfied under (3.19) using that $0 < \beta < \gamma < 2\beta$ is the same as $v^2 \ll w \ll v \ll 1$. ∎

Remark. The case $\gamma \geqslant 2\beta$ presents a small complication. Namely, the transition probability p_{10} from BB to WW (see Table III) exceeds all other transition probabilities and therefore we cannot find a coupling with a process of the type in Fig. 8. The difficulty sits in Fig. 11: p_{10} is so big that the combination $(c^2, c^3) = (WW, 2)$ cannot be avoided, making (3.20) false. It is possible to get around this complication by the following modification of Fig. 8: When the process goes from 2 to 1 the distance decreases not by 1 but by a random variable Z^* which takes into account the total drop in distance while the process resides in BB. All that will be needed later on in

Section 3.2 is that Z^* has an exponential tail with a mean that is bounded as $N \to \infty$. With this modification it is possible to construct a coupling satisfying $d^3(t) \leqslant d^2(t)$ a.s. as long as (3.19) holds. The coupling diagrams are somewhat messy and we refrain from spelling out the details.

To close this section we recall that we have so far ignored all transitions in Table II with probability $o(v \wedge w)$. However, these can now be easily incorporated because the coupling diagrams in Figs. 9–15 are flexible. Indeed, such perturbations correspond to including some extra intervals in the diagrams, all of which are of smaller order than the smallest interval drawn so far. The parameters δ, ε in (3.4) and (3.17) were built in to show that these intervals cannot affect Lemmas 3 and 4. *Thus, Lemma* 3 *and* 4 *are valid with the perturbations included*, i.e., for the fourth column of Table II. The skeptical reader is asked to check how, for instance, Fig. 9 modifies.

3.2. Proof of Lower Bounds in Theorems 2 and 3

Lemmas 3 and 4 provide us with a lower stochastic domination of the distance component $d^2(t)$ in the process $r^2(t)$ defined in Table III in terms of the distance component $d^3(t)$ in the simpler process $r^3(t)$ defined in (3.4) and (3.17) (see also Fig. 8). For $M' < M$ define the crossing probabilities

$$\begin{aligned} P^2(M', M) &= P^2(d^2(s) \in (M', M) \text{ for } s \in (\tau^2(M), \tau^2(M'))) \\ P^3(M', M) &= P^3(d^3(s) \in (M', M) \text{ for } s \in (\tau^3(M), \tau^3(M'))) \end{aligned} \tag{3.21}$$

i.e., the probability that $d^2(t)$ and $d^3(t)$ after hitting M continue to drop to M' without returning to M. Here P^2 and P^3 denote the measures on the trajectory space $(\Sigma^2)^{\mathbb{N}}$ under the kernel Q^2, resp. Q^3.

Lemma 5. For N sufficiently large and $0 < \delta_1 < \delta_2 < \delta$, in regime III

$$P^2(\delta_1 N, \delta_2 N) \leqslant P^3(\delta_1 N, \delta_2 N) \tag{3.22}$$

and in regime II

$$P^2(\delta_1 N^{1-(\gamma-\beta)}, \delta_2 N^{1-(\gamma-\beta)}) \leqslant P^3(\delta_1 N^{1-(\gamma-\beta)}, \delta_2 N^{1-(\gamma-\beta)}) \tag{3.23}$$

Proof. Immediate from Lemmas 3 and 4. ∎

Counting the number of excursions of $d^2(t)$ to the left of M before hitting $M' < M$, we have

$$\begin{aligned} &\tau^2(M') \succ R \\ &P(R > k) = [1 - P^2(M', M)]^k \qquad (k \geqslant 1) \end{aligned} \tag{3.24}$$

The following two lemmas combined with (3.3) and (3.22)–(3.24) complete the proof of the lower bounds in Theorems 2 and 3.

Lemma 6. In regime III, for any $0<\delta_1<\delta_2<1/3$ there exists $K>0$ such that

$$\lim_{N\to\infty} e^{KN^{1-\gamma}}P^3(\delta_1 N, \delta_2 N)=0 \tag{3.25}$$

Lemma 7. In regime II, for any $0<\delta_1<\delta_2<1$ there exists $K'>0$ such that

$$\lim_{N\to\infty} e^{K'N^{1-\gamma}}P^3(\delta_1 N^{1-(\gamma-\beta)}, \delta_2 N^{1-(\gamma-\beta)})=0 \tag{3.26}$$

Proof of Lemmas 6 and 7. Return to (3.4). When $c^3(t)$ enters 1, $d^3(t)$ starts to decrease. The total decrease before it exits 1 is

$$Z_1= -(Y_1+\cdots+Y_K) \tag{3.27}$$

with

$$\begin{gathered} K, Y_1, Y_2,\ldots \text{ independent} \\ P(Y_i=1)=1 \\ P(K>k)=(1-q_1)^k \qquad (k\geqslant 1) \end{gathered} \tag{3.28}$$

When $c^3(t)$ enters 2, $d^3(t)$ may move either up or down. The total decrease before it exits 2 is

$$Z_2=Y_1+\cdots+Y_K \tag{3.29}$$

with

$$\begin{gathered} K, Y_1, Y_2,\ldots \text{ independent} \\ P(Y_i=-1)=\frac{q_2}{q_2+q_3}, \qquad P(Y_i=+1)=\frac{q_3}{q_2+q_3} \\ P(K>k)=(q_2+q_3)^k \qquad (k\geqslant 1) \end{gathered} \tag{3.30}$$

Consider now $-Z_1+Z_2-2$, the increase of $d^3(t)$ as $c^3(t)$ makes one cycle in $\{1,2\}$. We want to think of this sum as a single step in a random jump process on $\mathbb{Z}$. More precisely, define

$$S(t)=W_1+\cdots+W_t \qquad (t\geqslant 0) \tag{3.31}$$

with

$$W_1, W_2,\ldots \text{ i.i.d. with the same distribution as } -Z_1+Z_2-2 \tag{3.32}$$

Compute [recall q_1, q_2, q_3 defined in (3.4) and (3.17)]

$$\begin{aligned} E(Z_1) &= \frac{1}{q_1} && (3.33)\\ &= \frac{\delta}{2w} \sim \frac{\delta}{2} N^{\gamma} && \text{if } \beta > \gamma \\ &= \frac{\delta}{2v} \sim \frac{\delta}{2} N^{\beta} && \text{if } \beta < \gamma < 2\beta \\ E(Z_2) &= \frac{1}{1-q_2-q_3} \frac{q_3-q_2}{q_2+q_3} \\ &= \frac{1-2\delta}{2(1+\varepsilon)w} \sim \frac{1-2\delta}{2(1+\varepsilon)} N^{\gamma} && \text{if } 2\gamma > \beta > \gamma \\ &= (1-2\delta)\frac{w}{v} \sim (1-2\delta) N^{\beta-\gamma} && \text{if } 2\gamma \leqslant \beta \\ &= \frac{1}{2(1+\varepsilon)v}\left(1-\frac{2\delta w}{v}\right) \sim \frac{1}{2(1+\varepsilon)} N^{\beta} && \text{if } \beta < \gamma < 2\beta \end{aligned}$$

In all three cases $E(W_1) = E(-Z_1+Z_2-2)$ is positive and large as $N \to \infty$. This means that $S(t)$ is a random walk with positive drift.

Next we proceed separately in the three cases:

(i) $2\gamma > \beta > \gamma$ (regime III): Divide $\mathbb{Z}$ into blocks of size N^{γ} and monitor $S(t)$ on this block scale, i.e., consider $\tilde{S}(t) = \lfloor S(t)/N^{\gamma} \rfloor$. The process $\tilde{S}(t)$ is a random walk on $\mathbb{Z}$ with positive drift when $\delta < 1/3$ [by (3.33)] and with exponentially bounded tails [by (3.27)–(3.30)]. For $S(t)$ to go from 0 to $-(\delta_2-\delta_1)N$ is the same as for $\tilde{S}(t)$ to go from 0 to $-(\delta_2-\delta_1)N^{1-\gamma}$. This has probability $\exp(-KN^{1-\gamma})$ by a standard large-deviation estimate, which in turn yields (3.25).

(ii) $2\gamma \leqslant \beta$ (regime III): Apply the same argument. Since $\beta - \gamma \geqslant \gamma$, this time the right tail of the random walk decays slower than exponential. However, this does not affect the estimate for the probability of $\tilde{S}(t)$ to run left from 0 to $-(\delta_2-\delta_1)N^{1-\gamma}$, and so again (3.25) holds.

(iii) $\beta < \gamma < 2\beta$ (regime II): Scale $\tilde{S}(t) = \lfloor S(t)/N^{\beta} \rfloor$. For $S(t)$ to go from 0 to $-(\delta_2-\delta_1)N$ is the same as for $\tilde{S}(t)$ to go from 0 to $-(\delta_2-\delta_1)N^{1-\gamma}$. The same estimate holds as before, yielding (3.26).

The case not listed in (3.33) is:

(iv) $\gamma \geqslant 2\beta$ (regime II): See the remark below Lemma 4. The same argument as in (iii) is still valid after replacing $-Z_1+Z_2-2$ in (3.32) by $-Z_1+Z_2-(1+Z^*)$, because Z^* has an exponential tail and EZ^* is bounded as $N \to \infty$. ∎

Combine (3.3) with (3.24)–(3.26) to get the lower bounds in Theorems 2 and 3.

4. PROOF OF UPPER BOUNDS FOR MEAN-FIELD MODEL

4.1. Upper Stochastic Domination

In Sections 4.1.1 and 4.1.2 we treat regimes III and II, respectively. We first consider only the leading-order column in Table II.

4.1.1. Regime III ($\gamma < 1$, $\beta > \gamma$). Pick $0 < \delta < 1/2$, $\varepsilon > 0$ and let $r^3(t)$ be the Markov process with the following kernel:

$$\begin{aligned} Q^3((1,d),(1,d-1)) &= 1-\frac{2w}{\delta} \\ q_1 := Q^3((1,d),(2,N)) &= \frac{2w}{\delta} \\ Q^3((2,N),(2,N)) &= 1-2(1-\varepsilon)v \\ q_2 := Q^3((2,N),(1,N)) &= 2(1-\varepsilon)v \\ Q^3 \text{ zero otherwise} & \end{aligned} \tag{4.1}$$

This chain keeps the distance fixed at the maximal value N while in state 2, and allows it to go down stepwise only while in state 1. The starting point is

$$r^3(0)=(2,N) \tag{4.2}$$

Lemma 8. For N sufficiently large, as long as

$$\frac{d^2(t)}{N} \geqslant \delta_1 > \delta \tag{4.3}$$

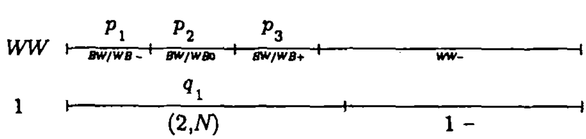

Fig. 16

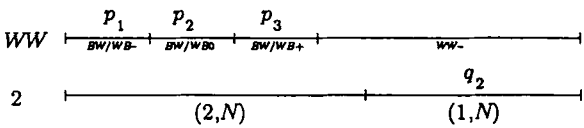

Fig. 17

there exists a coupling of $r^2(t)$ and $r^3(t)$ such that

$$d^2(t) \leqslant d^3(t) \qquad \text{a.s.} \tag{4.4}$$

Proof. The four allowed combinations of (c^2, c^3) are

$$(c^2, c^3) \in \{(WW, 1), (WW, 2), (BW/WB, 2), (BB, 2)\} \tag{4.5}$$

The four corresponding diagrams are given in Figs. 16–19, with

$$p_1 + p_2 + p_3 \leqslant q_1 \tag{4.6}$$

$$p_1 + p_2 + p_3 \leqslant 1 - q_2 \tag{4.7}$$

$$p_4 + p_5 + p_6 + p_7 \leqslant 1 - q_2 \tag{4.8}$$

$$p_{10} \geqslant q_2 \tag{4.9}$$

As before, (4.6)–(4.9) guarantee that (c^2, c^3) stays on the four combinations given in (4.5), and that (4.4) is satisfied. One verifies that (4.6)–(4.9) hold for N sufficiently large using that $\beta, \gamma > 0$ is the same as $v, w \ll 1$. The condition $\beta > \gamma$ is not needed. ∎

4.1.2. Regime II ($\gamma < 1$, $\beta < \gamma$). Pick $\delta, \varepsilon > 0$ and let $r^3(t)$ be given by the following kernel:

$$\begin{aligned} Q^3((1, d), (1, d-1)) &= 1 - \frac{2v}{\delta} \\ q_1 := Q^3((1, d), (2, d+1)) &= \frac{2v}{\delta} \\ Q^3((2, d), (2, d+1)) &= 1 - 2(1-\varepsilon)v \\ q_2 := Q^3((2, d), (1, d+1)) &= 2(1-\varepsilon)v \\ Q^3 \text{ zero otherwise} & \end{aligned} \tag{4.10}$$

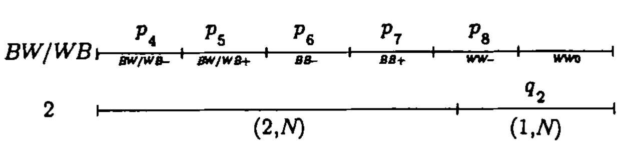

Fig. 18

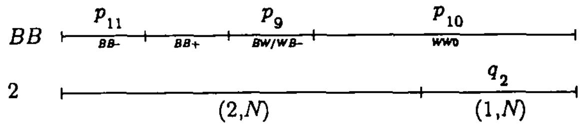

Fig. 19

The starting point is

$$r^3(0) = (2, \delta N^{1-(\gamma-\beta)}) \tag{4.11}$$

Lemma 9. For N sufficiently large, as long as

$$\frac{d(t)}{N} \geqslant \delta_1 \frac{w}{v} \qquad \text{with} \quad \delta_1 > \delta \tag{4.12}$$

there exists a coupling of $r^2(t)$ and $r^3(t)$ such that

$$d^2(t+1) - d^2(t) \leqslant d^3(t+1) - d^3(t) \qquad \text{a.s.} \tag{4.13}$$

Proof. The four allowed combinations of (c^2, c^3) are the same as in (4.5). The diagrams are the same as in Figs. 16–19, but with $(1, N)$, $(2, N)$ replaced by $(1, d+1)$, $(2, d+1)$. The conditions (4.6)–(4.9) are implied by (4.12) for N sufficiently large because $0 < \beta < \gamma$ is the same as $w \ll v \ll 1$. ▮

To close this section we recall the remark made at the end of Section 3.1. All transitions that we have ignored by considering the leading-order column in Table II can now be easily incorporated. *Lemmas* 8 *and* 9 *are valid with the perturbations included.*

4.2. Proof of Upper Bounds in Theorems 2 and 3

Lemmas 8 and 9 provide us with an upper stochastic domination of the distance component $d^2(t)$ in the process $r^2(t)$ defined in Table III. Return to Section 3.1. Our original process $r(t)$ had been reduced to the kernels Q^1 in (3.1) and Q^2 in (3.2). In going from Q^1 to Q^2, however, we ignored the time spent by the process without changing. This was compatible with lower stochastic domination [see (3.3)], but it is not with upper stochastic domination. We proceed as follows.

From Table II we see that the average waiting time in the color pairs *WW*, *BW/WB*, and *BB* is

$$\begin{aligned} WW: &\quad \frac{1}{\Delta + 2w(1-\Delta)} \leqslant \frac{1}{2w} \\ BW/WB: &\quad \frac{1}{1/2 + \Delta(v+w)} \leqslant 2 \\ BB: &\quad \frac{1}{(1+\Delta)v + w} \leqslant \frac{1}{v+w} \end{aligned} \tag{4.14}$$

The largest one being bounded above by $1/w \sim N^{\gamma}$, let us define

$$T \text{ independent of } r^2(t) \text{ with } P(T > k) = (1-w)^k \qquad (k \geqslant 1) \tag{4.15}$$

Then we have

$$\tau(M) = \tau^1(M) \prec T\tau^2(M) \tag{4.16}$$

The factor T compensates for the waiting time that we ignored in going from Q^1 to Q^2. Since $T = O(N^{\gamma}) = o(N)$ as $N \to \infty$, it cannot compete with the estimates for $\tau^2(M)$ that we shall derive below.

Regime III ($\gamma < 1$, $\beta > \gamma$). From Lemma 8 we get

$$\tau^2(\delta_1 N) \leqslant \tau^3(\delta_1 N) \tag{4.17}$$

The upper bound in Theorem 3 follows by combining (4.16)–(4.17) with the following estimate:

Lemma 10. For any $\delta_1 < 1/2$ there exists $K > 0$ such that

$$\lim_{N \to \infty} P^3(\tau^3(\delta_1 N) < e^{KN^{1-\gamma}}) = 1 \tag{4.18}$$

Proof. The only way the chain in (4.1) can reach the distance $\delta_1 N$ is by staying in state 1 for at least $(1-\delta_1)N$ steps in a row. This has probability

$$\left(1 - \frac{2}{\delta N^{\gamma}}\right)^{(1-\delta_1)N} \approx \exp\left(-\frac{2(1-\delta_1)}{\delta} N^{1-\gamma}\right) \tag{4.19}$$

Since $\tau^3(\delta_1 N)$ is at most $(1-\delta_1)N$ times the number of trials before this happens, the claim follows for $K > 2(1-\delta_1)/\delta$. ∎

Regime II ($\gamma<1$, $\beta<\gamma$). From Lemma 9 we get

$$\tau^2(\delta_1 N^{1-(\gamma-\beta)}) \leqslant \tau^3(\delta_1 N^{1-(\gamma-\beta)}) \tag{4.20}$$

The upper bounds in Theorem 2(i, ii) follow by combining (4.16) and (4.20) with the following estimates:

Lemma 11. (i) For any $\delta_1>1$ there exists $K>0$ such that

$$\lim_{N\to\infty} P^3\big(\tau^3(\delta_1 N^{1-(\gamma-\beta)})<KN\big)=1 \tag{4.21}$$

(ii) For any $\delta_1<1$ there exists $K'>0$ such that

$$\lim_{N\to\infty} P^3\big(\tau^3(\delta_1 N^{1-(\gamma-\beta)})<e^{K'N^{1-\gamma}}\big)=1 \tag{4.22}$$

Proof. The argument is similar to that in Section 3.2. When $c^3(t)$ enters 1, $d^3(t)$ starts to decrease. The total decrease while in 1 is Z_1 given by (3.27)–(3.28) with q_1 defined in (4.10). When $c^3(t)$ enters 2, $d^3(t)$ starts to increase. The total increase while in 2 is Z_2 given by (3.29)–(3.30) with q_2+q_3 replaced by $1-q_2$ defined in (4.10). We have

$$\begin{aligned} E(Z_1) &= \frac{1}{q_1}=\frac{\delta}{2v}\sim\frac{\delta}{2}N^{\beta} \\ E(Z_2) &= \frac{1}{q_2}=\frac{1}{2(1-\varepsilon)v}\sim\frac{1}{2(1-\varepsilon)}N^{\beta} \end{aligned} \tag{4.23}$$

The total increase of $d^3(t)$ as $c^3(t)$ makes one cycle in $\{1,2\}$ equals $-Z_1+Z_2-2$, while the total cycle time equals Z_1+Z_2+2.

Let

$$S(t)=\tfrac{1}{2}N+W_1+\cdots+W_t \qquad (t\geqslant 0) \tag{4.24}$$

with

$$W_1, W_2,\ldots \text{ i.i.d. with the same distribution as } -Z_1+Z_2-2 \tag{4.25}$$

The random walk $S(t)$ is the distance $d^3(t)$ observed at the completion times of a cycle. The starting point $d^3(0)=\frac{1}{2}N$ corresponds to $d(0)\sim\frac{1}{2}N$ ($N\to\infty$) in the original model (see below (1.13)). From (4.23) we see that $E(W_1)$ is large, negative when $\delta>1/(1-\varepsilon)$ and large, positive when $\delta<1/(1-\varepsilon)$. We now prove parts (i) and (ii) of Lemma 11:

(i) Pick $\delta>1/(1-\varepsilon)$. Then $E(W_1)\sim -c_1N^{\beta}$ for some $c_1>0$. Hence $\tilde{S}(t)=\lfloor S(t)/N^{\beta}\rfloor$ is a random walk with negative drift and with exponentially bounded tails. For $S(t)$ to go from $\frac{1}{2}N$ to $\delta_1 N^{1-(\gamma-\beta)}$ is the same as for $\tilde{S}(t)$ to go from $\frac{1}{2}N^{1-\beta}$ to $\delta_1 N^{1-\gamma}$. This takes $\tilde{S}(t)$ a number of steps that is of order $N^{1-\beta}$. Because $E(Z_1+Z_2+2)\leqslant c_2 N^{\beta}$ for some $c_2>0$, which bounds the mean time involved with a single step of $\tilde{S}(t)$, the claim follows by picking $1<\delta<\delta_1$ and ε sufficiently small.

(ii) Pick $\delta<\delta_1<1<\delta_2<1/(1-\varepsilon)$. Then $E(W_1)\sim c_1 N^{\beta}$ for some $c_1>0$. Each excursion of $d^3(t)$ to the right of $\delta_2 N^{1-(\gamma-\beta)}$ takes a time of order N, by (4.21) just proved. For $S(t)$ to go from $\delta_2 N^{1-(\gamma-\beta)}$ to $\delta_1 N^{1-(\gamma-\beta)}$ is the same as for $\tilde{S}(t)$ to go from $\delta_2 N^{1-\gamma}$ to $\delta_1 N^{1-\gamma}$. This takes $\tilde{S}(t)$ a number of steps that is of order $\exp(KN^{1-\gamma})$, because now $\tilde{S}(t)$ has a positive drift. Because $E(Z_1+Z_2+2)\leqslant c_2 N^{\beta}$ is a polynomial factor, the claim follows. ▮

5. RANDOM MEDIUM

In this section we have to face the problem of showing that the process $(r(t))_{t\geqslant 0}$ defined by (2.3) and evolving according to (2.4) is *well approximated by the mean-field version* evolving according to (2.13), in such a way that we can carry over the estimates of Sections 3 and 4. The main idea is that, as the dimension N of the configuration space Σ tends to infinity, *the random medium H tends to be almost uniform*, i.e., for most $\sigma, \xi \in \Sigma$ the numbers of neighboring color pairs defined in (2.2) are very close to their average values given in (2.12). To make this idea precise involves some large-deviation estimates for H, which are given in Section 5.1. These have to be quite sharp, because to prove Theorems 4 and 5 we need to follow the distance process $(d(x(t)))_{t\geqslant 0}$ over a stretched-exponential length of time. In Section 5.2 we prove a key lemma showing that the coupled dynamics sees an almost uniform medium over a stretched-exponential length of time. In Section 5.3 we use this lemma to prove Theorems 4 and 5.

5.1. Static Large-Deviation Estimate for *H*

Recall the notation introduced in Sections 1.1–1.2 and 2.1–2.2.

Lemma 12. Fix $0<\delta<1/2$ and $x\in\Sigma^2$ such that $\delta N\leqslant d(x)\leqslant (1-\delta)N$. For every $\varepsilon>0$ there exists $K_1=K_1(\delta,\varepsilon)>0$ such that

$$\bar{P}\big(|V_k(x,c)-\bar{E}V_k(x,c)|\leqslant \varepsilon N^{1-\gamma} \text{ for } k=1,2 \text{ and } c=WW, BW, WB\big) \geqslant 1-e^{-K_1 N^{1-\gamma}} \tag{5.1}$$

Proof. Consider $k=1$ and fix c. By the Markov inequality

$$\bar{P}\big(V_1(x,c)>\bar{E}V_1(x,c)+\varepsilon N^{1-\gamma}\big) \leqslant \inf_{\lambda>0} \exp\{-\lambda[\bar{E}V_1(x,c)+\varepsilon N^{1-\gamma}]\}\,\bar{E}(\exp[\lambda V_1(x,c)]) \tag{5.2}$$

By the independence of colors

$$\bar{E}(\exp[\lambda V_1(x, c)]) = [1 - p(c) + p(c)\, e^{\lambda}]^{d(x)} \tag{5.3}$$

For $0 < \lambda \leqslant \lambda_0$ sufficiently small

$$1 - p + pe^{\lambda} \leqslant e^{p\lambda + p(1-p)\lambda^2/2} \qquad (p \in [0, 1]) \tag{5.4}$$

By substituting $\bar{E}V_1(x, c) = d(x)\, p(c)$ and using the restriction $d(x) \geqslant \delta N$ we get

$$\bar{P}(V_1(x, c) > \bar{E}V_1(x, c) + \varepsilon N^{1-\gamma}) \\ \leqslant \inf_{\lambda \leqslant \lambda_0} \exp\{-\lambda\varepsilon N^{1-\gamma} + \tfrac{1}{2}p(c)[1 - p(c)]\,\lambda^2\,\delta N\} \tag{5.5}$$

Note that

$$\begin{aligned} p(c)[1 - p(c)] &\sim 2N^{-\gamma} \qquad \text{if} \quad c = WW \\ &\sim N^{-\gamma} \qquad \text{if} \quad c = BW, WB \end{aligned} \tag{5.6}$$

Pick $\lambda = 4\varepsilon/\delta$ in (5.5) to arrive at the upper bound $\exp(-KN^{1-\gamma})$ with $K = 4\varepsilon^2/\delta$. This proves the upper large-deviation estimate for ε small enough such that $4\varepsilon/\delta \leqslant \lambda_0$, and hence for all $\varepsilon > 0$. The same proof works for $k = 2$ and relies on $N - d(x) \geqslant \delta N$.

The lower large-deviation estimate is obtained similarly after replacing λ by $-\lambda$. Inequality (5.1) summarizes 12 large deviations, namely for $k = 1, 2$ and $c = WW$, BW, WB in both directions. ∎

Remark. Lemma 12 makes no statement about $c = BB$. In fact, since $p(BB) \sim N^{-2\gamma}$, the estimate for this color pair has the weaker exponent $N^{1-2\gamma}$. However, we shall not need this case.

5.2. Dynamic Large-Deviation Estimate for $x(t)$

For $\varepsilon > 0$ define the set

$$A_\varepsilon = \{x \in \Sigma^2 : |V_k(x, c) - \bar{E}V_k(x, c)| \leqslant \varepsilon N^{1-\gamma} \text{ for } k = 1, 2 \\ \text{and } c = WW, BW, WB\} \tag{5.7}$$

Lemma 12 shows that $\bar{P}(x \notin A_\varepsilon) \leqslant \exp(-K_1 N^{1-\gamma})$ for any $x \in \Sigma^2$ with $\delta N \leqslant d(x) \leqslant (1-\delta)N$ and for some $K_1 > 0$. Our key lemma below shows that on a stretched-exponential time scale $x(t)$ does not get out of A_ε before $d(t) = d(x(t))$ hits either δN or $(1-\delta)N$.

Lemma 13. For every $\varepsilon > 0$ and $0 < \delta < 1/2$ there exists $K_2 = K_2(\delta, \varepsilon) > 0$ such that

$$\lim_{N \to \infty} P\big(x(t) \notin A_\varepsilon \text{ for some } t \leqslant e^{K_2 N^{1-\gamma}} \wedge \tau(\delta N) \wedge \tau((1-\delta)N)\big) = 0 \quad (5.8)$$

Proof. Fix ε, δ. The proof consists of several steps.

1. Let

$$E_t^1 = \{\tau(\delta N) \wedge \tau((1-\delta)N) > t\} \quad (5.9)$$

For any n we have

$$P(\exists 0 \leqslant t \leqslant n: E_t^1; x(t) \notin A_\varepsilon) \leqslant \sum_{t=0}^{n} P(E_t^1; x(t) \notin A_\varepsilon) \quad (5.10)$$

2. Define

$$\begin{aligned} l_t^\sigma &= \#\{1 \leqslant i \leqslant N: \sigma(s) = \sigma^i(t) \text{ for some } 0 \leqslant s < t\} \\ l_t^\xi &= \#\{1 \leqslant i \leqslant N: \xi(s) = \xi^i(t) \text{ for some } 0 \leqslant s < t\} \end{aligned} \quad (5.11)$$

i.e., the number of neighbors of $\sigma(t)$, resp. $\xi(t)$, visited prior to time t. Pick $C_1 > 0$ and let

$$E_t^2 = \{l_t^\sigma \vee l_t^\xi \leqslant C_1 N^{1-\gamma}\} \quad (5.12)$$

For any t we have

$$P(E_t^1; x(t) \notin A_\varepsilon) \leqslant P([E_t^2]^c) + P(E_t^1; E_t^2; x(t) \notin A_\varepsilon) \quad (5.13)$$

3. **Lemma 14.** For every $C_1 > 0$ there exist $C_2, C_3 > 0$ such that

$$P([E_t^2]^c) \leqslant e^{-C_2 N^{1-\gamma}} \text{ for } t \leqslant e^{C_3 N^{1-\gamma}} \quad (5.14)$$

The proof will be given below.

4. Write

$$P(E_t^1; E_t^2; x(t) \notin A_\varepsilon)$$

$$= \sum_{\{x:\, \delta N < d(x) < (1-\delta)N\}} P(E_t^1; E_t^2; x(t) = x)\, P(x \notin A_\varepsilon \mid E_t^1; E_t^2; x(t) = x) \quad (5.15)$$

We estimate the second factor.

5. Given E_t^2 and $x(t) = x = (\sigma, \xi)$ we have

$$\#\{1 \leqslant i \leqslant N: \sigma(s) \neq \sigma^i, \xi(s) \neq \xi^i \text{ for all } 0 \leqslant s < t\} \geqslant N - 2C_1 N^{1-\gamma} \quad (5.16)$$

i.e., there are at least $N-2C_1N^{1-\gamma}$ directions i in which the neighbors of σ and ξ *have not yet been visited.* By the independence of colors each such direction has an independent probability of contributing to the set $V_k(x, c)$ [recall (2.2)]. It therefore follows that when we pick $2C_1 < \frac{1}{2}\varepsilon$ [recall (5.7)]

$$P(x \in A_\varepsilon \mid E_t^1; E_t^2; x(t)=x) \geqslant P(x \in A_{\varepsilon/2}) \tag{5.17}$$

6. By combining (5.15) with (5.17) we get

$$P(E_t^1; E_t^2; x(t) \notin A_\varepsilon) \leqslant \sum_{\{x:\, \delta N < d(x) < (1-\delta)N\}} P(E_t^1; E_t^2; x(t)=x)\, P(x \notin A_{\varepsilon/2}) \tag{5.18}$$

According to Lemma 12, $P(x \notin A_{\varepsilon/2}) \leqslant \exp(-K_1 N^{1-\gamma})$ for all x such that $\delta N < d(x) < (1-\delta)N$. Hence

$$P(E_t^1; E_t^2; x(t) \notin A_\varepsilon) \leqslant e^{-K_1 N^{1-\gamma}} \tag{5.19}$$

7. By combining (5.13), (5.14), and (5.19) we get

$$P(E_t^1; x(t) \notin A_\varepsilon) \leqslant e^{-C_2 N^{1-\gamma}} + e^{-K_1 N^{1-\gamma}} \text{ for } t \leqslant e^{C_3 N^{1-\gamma}} \tag{5.20}$$

Substitute (5.20) into (5.10) and pick $n = \exp(K_2 N^{1-\gamma})$ with $K_2 < C_2 \wedge K_1$ to arrive at

$$\lim_{N \to \infty} P(\exists 0 \leqslant t \leqslant e^{K_2 N^{1-\gamma}}: E_t^1; x(t) \notin A_\varepsilon) = 0 \tag{5.21}$$

This is (5.8) in Lemma 13. ∎

Proof of Lemma 14. The proof consists of several steps.

1. Since the two components have the same initial distribution π_H [see (1.3) and Section 1.3], we have

$$P([E_t^2]^c) \leqslant 2P(l_t^\sigma > C_1 N^{1-\gamma}) \tag{5.22}$$

Since π_H is reversible under the dynamics (see Section 1.2), we have

$$P(l_t^\sigma > C_1 N^{1-\gamma}) = P(m_t^\sigma > C_1 N^{1-\gamma}) \tag{5.23}$$

where we define

$$m_t^\sigma = \#\{1 \leqslant i \leqslant N: \sigma(s) = \sigma^i(0) \text{ for some } 0 < s \leqslant t\} \tag{5.24}$$

i.e., the number of neighbors of $\sigma(0)$ visited up to time t.

2. Denote the number of black neighbors of σ by

$$B(\sigma) = \#\{1 \leqslant i \leqslant N: H(\sigma^i) = B\} \qquad (\sigma \in \Sigma) \tag{5.25}$$

Let

$$E_t^3 = \{|B(\sigma(s)) - N^{1-\gamma}| \leqslant \varepsilon N^{1-\gamma} \text{ for } 0 \leqslant s \leqslant t\} \tag{5.26}$$

By the stationarity of the color environment as seen relative to $\sigma(t)$ (which is a consequence of the stationarity of π_H under the dynamics) we have

$$\begin{aligned} P([E_t^3]^c) &\leqslant \sum_{s=0}^{t} P(|B(\sigma(s)) - N^{1-\gamma}| > \varepsilon N^{1-\gamma}) \\ &= (t+1)\, P(|B(\sigma(0)) - N^{1-\gamma}| > \varepsilon N^{1-\gamma}) \\ &\leqslant (t+1)\, e^{-C_4 N^{1-\gamma}} \end{aligned} \tag{5.27}$$

where the last inequality follows from Lemma 12 [note that $\sigma(0)$ is independent of $B(\sigma(0))$]. Pick $t = e^{C_3 N^{1-\gamma}}$ with $C_3 < \frac{1}{2} C_4$ to get

$$P([E_t^3]^c) \leqslant e^{-C_4 N^{1-\gamma}/2} \text{ for } t \leqslant e^{C_3 N^{1-\gamma}} \tag{5.28}$$

3. By combining (5.22), (5.23), and (5.28) we see that to get Lemma 14 it now suffices to prove that

$$P(m_t^\sigma > C_1 N^{1-\gamma}; E_t^3) \leqslant e^{-C_5 N^{1-\gamma}} \text{ for } t \leqslant e^{C_3 N^{1-\gamma}} \tag{5.29}$$

since this implies (5.14) with $C_2 < \frac{1}{2} C_4 \wedge C_5$.

4. To prove (5.29) we proceed as follows. Let $\hat{\sigma}(t)$ denote the Glauber dynamics observed at the jump times, i.e.,

$$\begin{aligned} \hat{\sigma}(u) &= \sigma(\tau_u) \qquad (u = 0, 1, 2,...) \\ \tau_0 &= 0 \\ \tau_{u+1} &= \inf\{t > \tau_u : \sigma(t) \neq \sigma(\tau_u)\} \end{aligned} \tag{5.30}$$

We have obviously

$$P(m_t^\sigma > C_1 N^{1-\gamma}; E_t^3) \leqslant P(\hat{m}_t^\sigma > C_1 N^{1-\gamma}; \hat{E}_t^3) \tag{5.31}$$

where $\hat{m}_t^\sigma$ and $\hat{E}_t^3$ are the same as in (5.24), resp. (5.26), but with $\sigma(t)$ replaced by $\hat{\sigma}(t)$. We estimate the r.h.s. of (5.31).

5. Each site η at distance k from $\hat{\sigma}(0)$ has exactly k neighbors $\eta_1,..., \eta_k$ that are at distance $k-1$ from $\hat{\sigma}(0)$ $(k < N/2)$. Since $\hat{E}_t^3$ implies

that $\hat{\sigma}(s)$ has at least $(1-\varepsilon)N^{1-\gamma}$ black neighbors for all $0 \leqslant s \leqslant t$, we can estimate

$$P(\hat{\sigma}(s+1) \in \{\eta_1,..., \eta_k\} \mid \hat{\sigma}(s) = \eta; \hat{E}_t^3) \leqslant \frac{k}{(1-\varepsilon)N^{1-\gamma} - k} \text{ for } k \leqslant (1-\varepsilon)N^{1-\gamma} \tag{5.32}$$

Indeed, the transition probability from η to $\{\eta_1,..., \eta_k\}$ is maximal when η is white and $\eta_1,..., \eta_k$ are all black (see Section 1.2). In that case at least $(1-\varepsilon)N^{1-\gamma} - k$ neighbors of η at distance $k+1$ from $\hat{\sigma}(0)$ are also black, which gives the upper bound in (5.32).

6. From (5.32) we see that the distance process $\hat{d}(t) = d(\hat{\sigma}(t), \hat{\sigma}(0))$ given $\hat{E}_t^3$ decreases more slowly than the birth–death chain $d_0(t)$ on the set $\{0, 1,..., \frac{1}{2}(1-\varepsilon)N^{1-\gamma}\}$ defined by

$$\begin{aligned} P(d_0(t+1) = k-1 \mid d_0(t) = k) &= \frac{2k}{(1-\varepsilon)N^{1-\gamma}} \\ P(d_0(t+1) = k+1 \mid d_0(t) = k) &= 1 - \frac{2k}{(1-\varepsilon)N^{1-\gamma}} \end{aligned} \tag{5.33}$$

Since obviously [recall (5.24)]

$$\hat{m}_t^\sigma \leqslant \#\{0 < s \leqslant t: \hat{d}(s) = 1\} \tag{5.34}$$

it follows that

$$P(\hat{m}_t^\sigma > C_1 N^{1-\gamma} \mid \hat{E}_t^3) \leqslant P(\#\{0 < s \leqslant t: d_0(s) = 1\} > C_1 N^{1-\gamma}) \tag{5.35}$$

[Note that both $\hat{d}(t)$ and $d_0(t)$ must return to 1 after hitting 0.] Finally we estimate the r.h.s. of (5.35).

7. Let $\rho_1, \rho_2,...$ denote the successive return times of $d_0(t)$ to 1. We have

$$\begin{aligned} P(\#\{0 < s \leqslant t: d_0(s) = 1\} > C_1 N^{1-\gamma}) &\leqslant P\left(\sum_{j=1}^{C_1 N^{1-\gamma}} \rho_j \leqslant t\right) \\ &\leqslant P(\rho_j \leqslant t \text{ for } 1 \leqslant j \leqslant C_1 N^{1-\gamma}) \\ &= [P(\rho_1 \leqslant t)]^{C_1 N^{1-\gamma}} \end{aligned} \tag{5.36}$$

where the last equality uses that the ρ_j are i.i.d. Now, it is a standard property that as $N \to \infty$

$$\rho_1/E(\rho_1) \Rightarrow \exp(1)$$
$$\frac{1}{N^{1-\gamma}} \log E(\rho_1) \to C_6 > 0 \tag{5.37}$$

[exp(1) denotes the exponential distribution with mean 1]. Hence

$$P(\rho_1 \leqslant e^{C_3 N^{1-\gamma}}) \leqslant \tfrac{1}{2} \qquad \text{for} \quad C_3 \text{ sufficiently small.} \tag{5.38}$$

8. Combine (5.31), (5.35), (5.36), and (5.38) to get (5.29) with $C_5 < C_1 \log 2$. ∎

5.3. Small Perturbation of Mean Field

Proof of Theorem 5. The point of Lemma 13 is that $(1/N)\, V_k(x, c)$ is very close to $(1/N)\, \bar{E}V_k(x, c)$ along the trajectory of $x(t)$ for a very long time. Indeed, as long as $x(t) \in A_\varepsilon$ we have

$$\left|\frac{1}{N} V_k(x(t), c) - \frac{1}{N} \bar{E}V_k(x(t), c)\right| \leqslant \varepsilon N^{-\gamma} \qquad \text{if} \quad c = WW, BW, WB$$
$$\leqslant 3\varepsilon N^{-\gamma} \qquad \text{if} \quad c = BB \tag{5.39}$$

The bound for $c = BB$ follows because the fluctuations add up to zero.

Now return to (2.10). The transition probabilities of $r(t)$ defined in (2.3) deviate from their mean-field value $Q(r, r')$ in (2.13) by not more than

$$\varepsilon N^{-\gamma} \sum_{k=1,2} \left(\sum_{\tilde{c} = WW, BW, WB} P_k^{\tilde{c}}(r, r') + 3P_k^{BB}(r, r') \right)$$

Hence the random medium transition kernel in (2.10) equals

$$Q(r, r') + 12\varepsilon N^{-\gamma} z_t(r, r') \tag{5.40}$$

where $(z_t)_{t \geqslant 0}$ is some random process on $C^2 \times D$ with the property

$$\sum_{r'} z_t(r, r') = 0, \qquad |z_t(r, r')| \leqslant 1 \tag{5.41}$$

The point is to think of the last term in (5.40) as a *random perturbation* of the mean-field model. We have little information about z_t. It is some functional of $x(t)$ and H, which we have chosen to ignore by contracting to

$r(t)$. However, since ε may be picked arbitrarily small, the random perturbation is of order εv ($\varepsilon \ll 1$), where $v = N^{-\gamma}$ is the parameter in Tables I–III. *Thus, by Lemma* 13, *the leading-order column in Table* II *essentially describes the perturbed process up to a time of order* $\exp(K_2 N^{1-\gamma})$, *as long as it does not hit distance* δN *or* $(1-\delta)N$. *Since all stochastic domination estimates in Sections* 3 *and* 4 *are flexible under small perturbations* (*i.e.*, v *by* εv *and* w *by* εw, *with* $\varepsilon \ll 1$), *it immediately follows that Theorem* 3 *carries over to Theorem* 5 *for the random medium.* Recall here the remarks made at the end of Sections 3.1 and 4.1. ∎

Proof of Theorem 4. The statements in Lemmas 12 and 13 require that the distance lies in the interval $[\delta N, (1-\delta)N]$ for some $0 < \delta < 1/2$. Therefore the results of Theorem 2 cannot be immediately carried over because they involve distances of order $N^{1-(\gamma-\beta)} = o(N)$. However, the weaker estimates of Theorem 4 can be obtained as follows.

Pick α such that $\gamma < \alpha < 1$. It is easily checked that Lemma 12 holds for any $x \in \Sigma^2$ such that $N^\alpha \leqslant d(x) \leqslant N - N^\alpha$ when in (5.1) we replace $\exp(-K_1 N^{1-\gamma})$ by $\exp(-K_1 N^{\alpha-\gamma})$. Similarly, Lemma 13 holds when in (5.8) we replace $\exp(K_2 N^{1-\gamma})$ by $\exp(K_2 N^{\alpha-\gamma})$ and δN by N^α. Therefore we know that the mean-field approximation works up to a time of order $\exp(K_2 N^{\alpha-\gamma})$. Now, if $2\gamma - \beta < 1$, then we can pick $\alpha = 1 - (\gamma - \beta)$ [which is compatible with $\gamma < \alpha < 1$]. It follows that Theorem 2(i) immediately carries over to Theorem 4(i)(a) because (1.14) involves a time scale of order $N = o(\exp[K_2 N^{1-(2\gamma-\beta)}])$. The mean-field approximation works up to a time of order $\exp[K_2 N^{1-(2\gamma-\beta)}]$, which explains Theorem 4(ii). If, on the other hand, $2\gamma - \beta \geqslant 1$, then no α satisfying $\alpha \leqslant 1 - (\gamma - \beta)$ [and compatible with $\gamma < \alpha < 1$] exists. All we can say is that, since the mean-field model drops below $N^{1-(\gamma-\beta)}$ in a time of order N, we must have $\tau(N^\alpha) < K_2 N$ for all $\alpha > \gamma \geqslant 1 - (\gamma - \beta)$. This explains Theorem 4(i)(b). ∎

ACKNOWLEDGMENTS

Part of the work in this paper was carried out in the winter of 1991 while A.F. was visiting the Department of Statistics and F.d.H. the Department of Mathematics at Rutgers University. We thank Yehuda Vardi and Joel Lebowitz for their kind hospitality. A.F. was supported by CNR–NATO Fellowship 215.24/01. F.d.H. was supported by the Royal Netherlands Academy of Arts and Sciences. We thank Antonio Galves for communicating to us a preprint of ref. 4 and for discussions on the present paper.

REFERENCES

1. D. Aldous, Random walk on finite groups and rapidly mixing Markov chains, in *Séminaire de Probabilités XVII, Lecture Notes in Mathematics*, No. 986 (Springer-Verlag, New York, 1983), pp. 2113–2297.
2. R. Azencott, Boltzmann machines: High order interactions and synchronous learning, in *Lecture Notes in Statistics*, P. Barone, A. Frigessi, and M. Piccioni, eds. (Springer-Verlag, New York, 1992), pp. 14–45.
3. M. N. Barber and B. Derrida, Dynamical phase transitions in the two-dimensional ANNNI model, *J. Stat. Phys.* **51**:877–891 (1988).
4. M. Cassandro, A. Galves, and P. Picco, Dynamical phase transitions in a disordered system: The study of a random walk model, *Ann. Inst. H. Poincaré* **55**:689–705 (1991); see also M. Cassandro, A. Galves, and P. Picco, preprint in preparation.
5. B. Derrida, Dynamical phase transitions and spin glasses, *Phys. Rep.* **184**:207–212 (1989).
6. B. Derrida and G. Weinbush, Dynamical phase transitions in 3-dimensional spin glasses, *Europhys. Lett.* **4**:657–662 (1987).
7. P. Diaconis, R. L. Graham, and J. A. Morrison, Asymptotic analysis of a random walk on a hypercube with many dimensions, *Random Structures Algorithms* **1**:51–72 (1990).
8. P. Diaconis, *Group Representation in Probability and Statistics* (Institute of Mathematical Statistics, Hayward, California, 1988).
9. Th. Eisele, On a third-order phase transition, *Commun. Math. Phys.* **90**:125–159 (1983).
10. A. Frigessi, C.-R. Hwang, S.-J. Sheu, and P. di Stefano, Convergence rates of the Gibbs Sampler, the Metropolis algorithm and other single-site updating dynamics, *J. R. Stat. Soc. B* **55**:205–220 (1993).
11. D. Geman, *Random Fields and Inverse Problems in Imaging* (Springer-Verlag, New York, 1990).
12. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.* **21**:1087–1092 (1953).
13. M. Mezard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (Springer-Verlag, New York, 1992).