

Asymptotic behavior and aging of a low temperature cascading 2-GREM dynamics at extreme time scales*

Luiz Renato Fontes[†] Véronique Gayrard[‡]

Abstract

We derive scaling limit results for the Hierarchical Random Hopping Dynamics for the cascading two-level GREM at low temperature at extreme time scales. It is known that in the cascading regime there are two static critical temperatures. We show that there exists a (narrow) set of *fine tuning* temperatures; when they lie below the static lowest critical temperature, three distinct *dynamical* phases emerge below the lowest critical temperature, with three different types of limiting dynamics depending on whether the temperature is (well) above or below, or at a fine tuning temperature, all of which are given in terms of K processes. We also derive scaling limit results for temperatures between the lowest and the highest critical ones, as well as aging results for all the limiting processes mentioned above, by taking a second small time limit.

Keywords: GREM; Random Hopping Dynamics; low temperature; fine tuning temperature; scaling limit; extreme time-scale; K process; aging; spin glasses.

AMS MSC 2010: 60K35; 82C44.

Submitted to EJP on April 30, 2019, final version accepted on November 15, 2019.

Supersedes arXiv:1801.08832.

1 Introduction

It is believed that *activated* dynamics of spin glasses, which occur at time-scales on which the process has time to escape deep valleys of a highly complex random energy landscape, exhibit aging at low temperature, the nature of which should be closely linked to the specific properties of this landscape [10], [9]. Over the past decades one type of aging behavior was isolated and identified, first in the REM [3], [4], [5], [20] whose complete dynamic phase diagram is now known [25], and later also in the p -spin SK models [2, 13] but for restricted time-scales at which the dynamics does not have time

*L.R.F. was supported in part by CNPq 311257/2014-3 and 305760/2010-6; FAPESP 2017/10555-0 and 09/52379-8.

[†]Universidade de São Paulo, Instituto de Matemática e Estatística, Rua do Matão 1010, 05508-090 São Paulo SP, Brazil. E-mail: lrfontes@usp.br

[‡]Aix Marseille Univ, CNRS, Centrale Marseille, I2M, Marseille, France.
E-mail: veronique.gayrard@math.cnrs.fr

to discover the full correlation structure of the random environment and, not being influenced by strong long-distance correlations, behaves essentially like a REM.

A question arose as to whether this *REM-like* behavior was truly universal or depended on the choice of dynamics. Indeed, all the papers quoted above study the same dynamics, the *Random Hopping dynamics*, whose trajectories do not depend on the random Hamiltonian, a choice that is clearly not satisfactory from the point of view of physics. Would a physically more realistic choice of dynamics change the nature of aging? The recent paper [24] answered in the negative for the key example of *Metropolis dynamics* of the REM, for all time-scales away from *extreme time-scales* at which the phase transition from aging to stationarity takes place. Nothing is known to date about Metropolis dynamics of the REM at extreme time-scales, whether from a mathematical, theoretical or numerical point of view [1], let alone Metropolis dynamics of p -spin SK models.

The present paper initiates the study of the aging dynamics of the GREM, a model with built-in hierarchical structure for which both the process of extremes and the low temperature Gibbs measure are fully understood [15, 14]. We consider here only the 2-GREM, which already enables us to identify new aging behaviors and an unforeseen mechanism, called *fine tuning mechanism*, while avoiding the heaviness inherent in the model with finitely many hierarchies. Furthermore, we restrict our analysis to extreme time-scales, where the process undergoes a transition from aging to equilibrium and, in doing so, visits the configurations in the support of the Gibbs measure. The parameters of the model and the temperature are chosen such that this support has a fully cascading structure, a setting that would be qualified as “2-step replica symmetry breaking” in physics. Shorter time-scales where the dynamics is aging will be considered in a follow-up paper. Finally, our choice of the dynamics is dictated by that of the time-scales. As already mentioned, Metropolis dynamics at extreme time-scales is out of reach. However, we do not study the plain Random Hopping dynamics considered in earlier literature but a hierarchical version, the *Hierarchical Random Hopping dynamics*, which is closer to Metropolis dynamics. Based on our knowledge of the REM, it is reasonable to expect that our choice of dynamics allows to correctly predict aging of Metropolis dynamics of the GREM near the phase transition.

For the most part of this paper, we will be concerned with the scaling limit of the dynamics for the time-scales and parameters specified above. Having obtained the limiting dynamics – given by K-processes appropriate to each regime that emerges in the analysis – (see Subsection 2.5), we then proceed to take a *further small time limit* for which aging results follow (see Subsection 1.3). Not only do we obtain a new, distinctive *GREM-like* aging behavior that goes beyond the known REM-like behavior, but we also isolate a new, temperature dependent *fine tuning* mechanism, that gives rise to three distinct aging regimes (corresponding to three distinct K-processes) which the dynamics can be tuned in by adjusting the temperature. This completely new fine tuning mechanism, and the rich aging picture that emerges were not predicted in the physics literature on the GREM dynamics [10], [29], [9].

1.1 The model

We now specify our setting. Let $\mathcal{V}_N = \{-1, 1\}^N$, $\mathcal{V}_{N_i} = \{-1, 1\}^{N_i}$, $\sigma = \sigma_1 \sigma_2$, $\sigma_i \in \mathcal{V}_{N_i}$, $i = 1, 2$, and make $N_1 = \lfloor pN \rfloor$ for some $p \in (0, 1)$ and $N_2 = N - N_1$; we view σ_i as the i -th *hierarchy* or *level* of σ . Given $a \in (0, 1)$, set

$$H_N(\sigma) = H_N^{(1)}(\sigma) + H_N^{(2)}(\sigma), \quad \sigma \in \mathcal{V}_N, \quad (1.1)$$

where, for $\sigma \in \mathcal{V}_N$

$$H_N^{(1)}(\sigma) = H_N^{(1)}(\sigma_1) = -\sqrt{aN}\Xi_{\sigma_1}^{(1)}, \quad H_N^{(2)}(\sigma) = -\sqrt{(1-a)N}\Xi_{\sigma_1\sigma_2}^{(2)}, \quad (1.2)$$

and $\Xi := \{\Xi_{\sigma_1}^{(1)}, \Xi_{\sigma}^{(2)}; \sigma \in \mathcal{V}_N\}$ is a family of i.i.d. standard Gaussian random variables. We call *random environment* and denote by $(\Omega, \mathcal{F}, \mathcal{P})$ the probability space on which the sequence of processes $(H_N(\sigma), \sigma \in \mathcal{V}_N)$, $N > 1$, is defined. As usual, we call $H_N(\sigma)$ the *Hamiltonian* or *energy* of σ . We refer to the minima of $H_N(\cdot)$ as *low energy* or *ground state* configurations. We will also refer to them, for being minima of $H_N(\cdot)$, as *top* configurations. Likewise for $H_N^{(i)}(\cdot)$, $i = 1, 2$. The associated Gibbs measure at inverse the temperature $\beta > 0$ is the (random) measure $G_{\beta,N}$ defined on \mathcal{V}_N through $G_{\beta,N}(\sigma) = e^{-\beta H_N(\sigma)} / Z_{\beta,N}$ where $Z_{\beta,N}$ is a normalization.

Let us briefly recall the key features of the statics of the 2-GREM (see [8] for a detailed account). As regard the Hamiltonian, two scenarios may be distinguished, related to the composition of the ground state energies in terms of their first and second level constituents: the *cascading phase*, when $a > p$, and where those energies are achieved by adding up the minimal energies of the two levels, so that to each first level ground state configuration, there corresponds many second level ground state configurations; in the complementary *non-cascading phase*, the composition of the ground state energies is different, and for each first level constituent there corresponds a single second level constituent.

Consider now the case where $a > p$. The free energy exhibits two discontinuities at the critical temperatures

$$\beta_1^{cr} \equiv \beta_* \sqrt{\frac{p}{a}} < \beta_2^{cr} \equiv \beta_* \sqrt{\frac{1-p}{1-a}}, \quad \beta_* = \sqrt{2 \ln 2}, \quad (1.3)$$

and the Gibbs measure behaves as follows. In the *high-temperature* region $\beta < \beta_1^{cr}$, no single configuration carries a positive mass in the limit $N \uparrow \infty$, \mathcal{P} -a.s.; here the measure resembles the high temperature Gibbs measure of the REM. On the contrary, in the *low temperature* region $\beta > \beta_2^{cr}$, the Gibbs measure becomes fully concentrated on the set of ground state configurations, yielding Ruelle’s two-level probability cascade. In between, when $\beta_1^{cr} < \beta < \beta_2^{cr}$, an intermediate situation occurs in which the first level Hamiltonian variables “freeze” close to their ground state values, but not the second level ones, so that, once again, no single configuration carries a positive mass in the limit $N \uparrow \infty$. To obtain a macroscopic mass, one must lump together an exponentially large number of second level configurations. In this paper we focus on the cascading phase ($a > p$) of the model at low temperature ($\beta > \beta_2^{cr}$). We will also treat the case $\beta_1^{cr} < \beta < \beta_2^{cr}$ in a sub-domain of the parameters where we can prove a scaling limit for the dynamics at the extreme time-scale. In the complementary sub-domain the process is in an aging phase and will thus be treated in the follow-up paper.

The dynamics we consider is a Markov jump process $(\sigma^N(t), t > 0)$ that evolves along the edges of \mathcal{V}_N with transition rates given by, for spin configurations $\sigma, \sigma' \in \mathcal{V}_N$,

$$Nw_N(\sigma, \sigma') = e^{\beta H_N(\sigma)} \mathbb{1}_{\sigma \overset{1}{\sim} \sigma'} + e^{\beta H_N^{(2)}(\sigma)} \mathbb{1}_{\sigma \overset{2}{\sim} \sigma'} \quad (1.4)$$

and $w_N(\sigma, \sigma') = 0$ else, where $\sigma \overset{i}{\sim} \sigma'$ iff $\sigma \sim \sigma'$ and $\sigma_i \sim \sigma'_i$. Let us briefly discuss this choice of rates. With the exception of [24], the only dynamics of mean-field spin glasses for which REM-like aging could be proved so far, and also the only one for which scaling limits at extreme time scales could be obtained is the Random Hopping dynamics [3, 4] which, when it jumps, chooses its next state uniformly from its n nearest neighbors, regardless of the random environment. This choice is also legitimate in the GREM, but it

could lead to an aging behavior different from that of a more “realistic” Glauber dynamic. Indeed, a moment’s thought reveals that the hierarchical structure of the Hamiltonian of the GREM is reflected in the dynamics, for typical choices of Glauber dynamics such as Metropolis dynamics, but that this property is lost in the above Random Hopping dynamics. The rates (1.4) retain this hierarchical property of Metropolis dynamics, but simplifies Metropolis rates within each given level by replacing them by the rates of the Random Hopping dynamics. We naturally call the resulting dynamics the *Hierarchical Random Hopping dynamics* (hereafter HRHD).

Following a standard notation $\sigma \sim \sigma'$ indicates that $d(\sigma, \sigma') = 1$, where $d(\cdot)$ stands for the usual Hamming distance in \mathcal{V}_N – we will below denote by d_i such distance in \mathcal{V}_{N_i} , $i = 1, 2$. In other words, $\sigma \sim \sigma'$ indicates that σ, σ' differ in exactly one coordinate; we say in this context that σ, σ' are (nearest) neighbors (in \mathcal{V}_N). We recognize the graph whose vertices are \mathcal{V}_N and whose edges are the neighboring pairs of configurations of \mathcal{V}_N , abusively denoted also by \mathcal{V}_N , as the N -dimensional hypercube. Clearly, σ^N is reversible w.r.t. $G_{\beta, N}$.

It now remains to specify the time-scale in which we observe this process. As mentioned earlier, we are interested in extreme time-scales, where the dynamics is close to equilibrium. What we mean here by the dynamics being close to equilibrium at a given extreme time-scale is that the dynamics with time rescaled by that time-scale converges in distribution to a nontrivial Markov process which is ergodic in the sense of having an irreducible (countable) state space and a unique equilibrium distribution. The limiting dynamics is thus close to equilibrium, since it converges to equilibrium as time diverges, and it is in this sense that we say that the original dynamics is close to equilibrium at the extreme time-scale – see Remark 2.10 for a more precise discussion. They also are the time-scales that determine the phase separation line between aging and stationarity in the sense that taking time to zero, the process enters an aging regime.

For future reference we call \mathbb{P} the law of σ^N conditional on the σ -algebra \mathcal{F} , i.e. for fixed realizations of the random environment, or \mathbb{P}_η , when the initial configuration η is specified. We will denote by $\mathcal{P} \otimes \mathbb{P}_\eta$ the probability measure obtained by integrating \mathbb{P}_η with respect to \mathcal{P} . Expectation with respect to \mathbb{P} , \mathcal{P} and $\mathcal{P} \otimes \mathbb{P}_\mu$ are denoted by \mathbb{E} , \mathcal{E} and $\mathcal{E} \otimes \mathbb{E}_\mu$, respectively, where μ is the uniform probability measure on \mathcal{V}_N .

1.2 Dynamical phase transitions

The distinct *static* phases of the cascading 2-GREM, determined by $\beta = \beta_i^{cr}$, $i = 1, 2$, are expected to exhibit different *dynamical* behaviors under the HRHD at extreme (and conceivably other) time scales. This will be seen when comparing the results of our analysis of the HRHD below the lowest critical temperature ($\beta > \beta_2^{cr}$) on the one hand, and those for intermediate temperatures ($\beta \in (\beta_1^{cr}, \beta_2^{cr})$), on the other hand. Another source of dynamical phase transition in the HRHD at extreme time scales is the *fine tuning* phenomenon discussed next.

1.2.1 Fine tuning; heuristics

There are two competing factors governing the behavior of the RHD at extreme time-scales. One is the number of jumps it takes for the dynamics to leave a first level ground state configuration σ_1 . This is a geometric random variable with mean $1 + \frac{N_2}{N_1} \exp\{\beta \sqrt{aN} \Xi_{\sigma_1}^{(1)}\} \approx \exp\{\beta \beta_* \sqrt{paN}\}$. The other factor is the number of jumps the process makes until it finds a second level low energy configuration. This is $\approx 2^{(1-p)N}$. The relative size of these numbers determines three temperature regimes.

At relatively high temperatures, the second number dominates, and so after leaving a ground state configuration σ , which it does at times of order

$$\exp\{\beta\sqrt{(1-a)N}\Xi_{\sigma}^{(2)}\} \approx \exp\{\beta\beta_*\sqrt{(1-p)(1-a)N}\},$$

σ^N will visit many first level ground state configurations before it finds a second level ground state configuration. When it first finds such a second level configuration, say σ'_2 , while in a first level ground state configuration σ'_1 (meaning that it first returned to an overall ground state configuration $\sigma'_1\sigma'_2$), σ'_1 will be effectively distributed proportionally to $\exp\{\beta\sqrt{aN}\Xi_{\sigma'_1}^{(1)}\}$; this can be explained by a size-bias mechanism that operates in the selection of σ'_1 . There is no such mechanism for the choice of σ'_2 , and it is distributed uniformly.

On the other hand, at low enough temperatures, the first factor dominates, and while staying at a first level low energy configuration, the process has time to reach equilibrium at the second level, so at the time scale where we see (uniform) transitions between first level low energy configurations, the second level is in equilibrium. This is a longer time scale, composed of the many jump times at second level till exiting first level.

In a narrow strip of borderline temperatures, we see nontrivial dynamics at both levels going on at the same time scale (corresponding to jump times out of second level ground state configurations, roughly of magnitude $\exp\{\beta\beta_*\sqrt{(1-p)(1-a)N}\}$, as at high temperatures).

In order for the above picture to represent the dynamics, we need the temperature to be below the static lowest phase transition temperature $1/\beta_2^{cr}$, so that the time spent off the ground state configurations is negligible. Moreover, this three-phase dynamical picture will take place if (and only if) the borderline temperatures alluded to above – and to be called *fine tuning* temperatures below – are (well) below the static lowest phase transition temperature; otherwise, we will see only one dynamical phase below that lowest critical temperature, namely the low temperature phase alluded to above.

1.2.2 Intermediate temperatures

For values of β between $\beta = \beta_1^{cr}$ and $\beta = \beta_2^{cr}$, we investigate the behavior of the dynamics at a time scale when we see transitions between the first level ground state configurations at times of order 1. In order that this time scale corresponds to an extreme time scale (as stipulated above – see the one but last paragraph of Subsection 1.1), we need a further a restriction in the temperature, to be seen below. The behavior of the dynamics of the first level configuration for intermediate temperatures in those conditions is similar to the one below the minimum between the lowest critical and the fine tuning temperature.

1.3 Aging results

Let us briefly anticipate our main aging results, holding in the case described at the last paragraph of Subsection 1.2.1, where fine tuning temperatures are below the lowest static critical temperature. In this case, as mentioned above, we have three phases for the dynamics at extreme time scales below the lowest static critical temperature. As already explained briefly above, our aging results in this paper are obtained by first taking the scaling limit of the dynamics at extreme time scales, thus obtaining ergodic processes, and next taking a small time limit in those processes, thus obtaining aging results. Let us suppose we have already taken the first, extreme time scale limit. We obtain three distinct dynamics in each of the temperature ranges: above fine tuning, at

fine tuning, and below fine tuning (see Theorems 2.4, 2.5, and 2.7 in Subsection 2.5). For $i = 1, 2$, let us consider the events

$$\mathfrak{N}_i = \mathfrak{N}_i(t_w, t) = \{Y_i \text{ does not jump between times } t_w \text{ and } t_w + t\}, \quad (1.5)$$

where Y_i represents the i -th level marginal of the process, and $t_w, t > 0$, and define

$$\Pi(t_w, t_w + t) = P(\mathfrak{N}_1 \cap \mathfrak{N}_2) + pP(\mathfrak{N}_1 \cap \mathfrak{N}_2^c) + (1 - p)P(\mathfrak{N}_1^c \cap \mathfrak{N}_2). \quad (1.6)$$

Π is an analogue of a (limiting) two-time overlap function of the HRHD. In the regime considered in this subsection, we have the following (vanishing time limit) aging result: writing FT as shorthand for fine tuning,

$$\lim_{\substack{t_w, t \rightarrow 0 \\ t/t_w \rightarrow \theta}} \Pi(t_w, t_w + t) = \begin{cases} \text{Asl}_{\alpha_2} \left(\frac{1}{1+\theta} \right), & \text{above FT,} \\ p\text{Asl}_{\alpha_1\alpha_2} \left(\frac{1}{1+\theta} \right) + (1 - p)\text{Asl}_{\alpha_2} \left(\frac{1}{1+\theta} \right), & \text{at FT,} \\ p\text{Asl}_{\alpha_1} \left(\frac{1}{1+\theta} \right), & \text{below FT,} \end{cases}$$

with proper choices of $\alpha_1, \alpha_2 \in (0, 1)$, where Asl . is the arcsine distribution function. See Section 10 and also the last paragraph of Section 11 for details and other regimes.

1.4 A 2-GREM-like trap model

The idea behind the construction of trap models for low temperature glassy dynamics is as follows: the traps represent the ground state configurations and, assuming that at low temperature the dynamics spends virtually all of the time on those configurations – here an extreme time-scale is assumed –, higher energy configurations are simply not represented, and all one needs to do is specify the times spent by the dynamics at each visit to a ground state configurations, and the transitions among those configurations, in such a way that the resulting process be Markovian.

The simplest such model to be proposed in the study of aging was put forth in [10], with $\{1, \dots, M\}$ as configuration space, mean waiting time at i given by X_i , with X_1, X_2, \dots iid random variables in the domain of attraction of of an α -stable law, $\alpha \in (0, 1)$, and uniform transitions among the configurations. This is the so called *REM-like trap model* or *trap model on the complete graph*. Models of a similar nature for the GREM were proposed in [10] and also in [29]. The scaling limit of the latter model for a *fine tuning* choice of level volumes was computed in [18] and its aging behavior away from fine tuning was studied in [23].

Out of our analysis of the HRHD in the cascading phase at low temperatures comes up the following GREM-like trap model on the ground state configurations of the GREM. The configuration space is represented by M_1 first level ground state configurations, labeled in decreasing order, and for each of those configurations, we have M_2 second level ground state configurations, labeled in decreasing order. The transition probabilities $p(x, y)$ between $x = (x_1, x_2)$ and $y = (y_1, y_2)$, $1 \leq x_i, y_i \leq M_i$, $i = 1, 2$, are given by

$$p(x, y) = \begin{cases} [(1 - \lambda^{y_1}) + \lambda^{y_1} \nu_1(y_1)] \frac{1}{M_2}, & \text{if } x_1 = y_1, \\ \nu_1(y_1) \lambda^{y_1} \frac{1}{M_2}, & \text{otherwise,} \end{cases} \quad (1.7)$$

where

$$\lambda^{y_1} = \frac{1}{1 + M_2 \psi \gamma_1(y_1)} \quad \text{and} \quad \nu_1(y_1) = \frac{\gamma_1(y_1) \lambda^{y_1}}{\sum_{z=1}^{M_1} \gamma_1(z) \lambda^z}. \quad (1.8)$$

The factor $\psi \in [0, \infty]$ in (1.8) interpolates between higher temperatures, above fine tuning ($\psi = 0$) and low temperatures, below fine tuning ($\psi = \infty$); $\psi \in (0, \infty)$ corresponds to borderline, fine tuning temperatures in the picture outlined above, and to be described

more precisely below. The factors $\gamma_1(\cdot)$ correspond to the scaled M_1 maxima first-level Boltzmann factors $\exp\{\beta\sqrt{aN}\Xi^{(1)}\}$.

The time spent at each visit to x in the appropriate time scale is an exponential random variable with mean $\gamma_2(x)$, where for each x_1 , $\gamma_2(x_1, \cdot)$ corresponds to the scaled M_2 maxima second-level Boltzmann factors $\exp\{\beta\sqrt{(1-a)N}\Xi_{\sigma_1}^{(2)}\}$, with σ_1 the first level configuration labeled x_1 , as explained above. It must be said that this time scale is of the order of magnitude of the time needed for the dynamics to jump out of ground state configurations, and it is indeed *extreme* in the above sense only for fine tuning temperatures and above. In these cases, (1.7) indeed represents the transitions among ground states (in the extreme time scale). At lower temperatures, as explained in the discussion on dynamical phase transition above, the extreme time scale is longer, with uniform transitions on first level, with exponential waiting times, and on second level the dynamics is a trivial product of equilibria at different times.

The results indicated above do not seem to be in or be predicted by the physics literature, which has focused on short time scales, where all levels age simultaneously, and thus no effect of the longer time dynamical phase transition is present. This matches our *short* extreme times aging results only at fine tuning, where that simultaneity takes place. Also, our GREM-like trap model differs from those considered in the literature (in [10, 29]).

1.5 Organization

In Section 2 we make precise the notions introduced in this introduction, and formulate our scaling limit results for σ^N on extreme time scales for $\beta > \beta_2^{cr}$. In Sections 3-6 we formulate and argue entrance law results leading in particular to the transition probabilities between ground state configurations described in 1.7. These results are key ingredients to the proofs of the above mentioned scaling limit results, which are undertaken in Sections 7-9. Section 10 is devoted to a brief discussion about aging results that we obtain for the limit processes, as already mentioned. In Section 11 we briefly discuss results for the intermediate temperature phase $(\beta_1^{cr}, \beta_2^{cr})$. An appendix contains definitions of the limit processes entering our scaling limit results, as well as auxiliary results.

2 Scaling limit of σ^N . Main Results

2.1 Choice of parameters

As mentioned above, we will study the cascading phase, which, we recall, corresponds to

$$a > p. \tag{2.1}$$

As regards temperatures, we want to take volume dependent ones (this is needed in order to capture the fine tuning phase transition). We also want low temperatures, which in the cascading phase corresponds to

$$\liminf_{N \rightarrow \infty} \beta > \beta_* \sqrt{\frac{1-p}{1-a}}, \tag{2.2}$$

where the dependence of β on N is implicit. In order to describe that dependence, let us start by setting $\beta_* = \sqrt{2 \ln 2}$, $\kappa = \frac{1}{2}(\ln \ln 2 + \ln 4\pi)$,

$$\alpha_1^N = \frac{\beta_*}{\beta} \sqrt{\frac{N_1}{Na}}, \quad c_1^N = \exp \left\{ -\frac{1}{\alpha_1^N} \left(\beta_*^2 N_1 - \frac{1}{2} \ln N_1 + \kappa \right) \right\}. \tag{2.3}$$

Given a sequence $\zeta_N < N_2\beta_*^2/2$ of real numbers, let $\beta(a, p, N, \zeta_N)$ be the solution in β of the equation

$$c_1^N 2^{N_2} = e^{\zeta_N + \kappa/\alpha_1^N}. \tag{2.4}$$

In explicit form

$$\beta(a, p, N, \zeta_N) = \frac{\beta_* N_2}{2 N_1} \sqrt{\frac{N_1}{Na}} \frac{1 - \frac{2\zeta_N}{N_2\beta_*^2}}{1 - \frac{\ln N_1}{2\beta_*^2 N_1}} = \beta^{FT} \left(1 - \frac{2\zeta_N}{N_2\beta_*^2}\right) (1 + o(1)) \tag{2.5}$$

where $\beta^{FT} := \frac{1-p}{2\sqrt{ap}}\beta_*$ is the inverse of the *fine tuning* temperature. Depending on the behavior of ζ_N we distinguish three types of temperature regimes. (Given two sequence s_N and \bar{s}_N we write $s_N \sim \bar{s}_N$ iff $\lim_{N \rightarrow \infty} s_N/\bar{s}_N = 1$. We also write $s_N = \mathcal{O}(1)$, resp. $s_N = o(1)$, iff $|s_N| \leq C' < \infty$, for some C' and all $N > 0$, resp. $\lim_{N \rightarrow \infty} s_N = 0$.)

Definition 2.1 (At/above/below fine tuning). *We say that a sequence $\beta^{-1} \equiv \beta_N^{-1} > 0$ of temperatures is in the fine tuning (FT) regime if there exists a finite real constant ζ and a convergent sequence $\zeta_N \sim \zeta$ such that $\beta = \beta(a, p, N, \zeta_N)$. We say that β^{-1} is below fine tuning if there exists a sequence ζ_N^- satisfying $\zeta_N^- \rightarrow -\infty$ as $N \rightarrow \infty$, and such that $\beta = \beta(a, p, N, \zeta_N^-)$. Finally, we say that β^{-1} is above fine tuning if there exists a sequence ζ_N^+ such that $\zeta_N^+ \rightarrow +\infty$ and $\limsup \frac{\zeta_N^+}{N_2} < \frac{\beta_*^2}{2}$ as $N \rightarrow \infty$, and $\beta = \beta(a, p, N, \zeta_N^+)$. (Note that for $\beta = \beta(a, p, N, \zeta_N)$ to be a convergent sequence, ζ_N/N_2 must be convergent.)*

In order to precisely describe our results, we start with some technical preliminaries. As described above, the way the ground state configurations are arranged in the cascading phase suggests the following relabeling of the state space \mathcal{V}_N .

2.2 Change of representation

Let $\mathcal{D}_i = \{1 \dots 2^{N_i}\}$, $i = 1, 2$. Call $\xi_1^{x_1}$, $x_1 \in \mathcal{D}_1$, the vertices of \mathcal{V}_{N_1} that carry the ranked variables

$$\Xi_{\xi_1^{x_1}}^{(1)} \geq \Xi_{\xi_2^{x_1}}^{(1)} \geq \dots \geq \Xi_{\xi_1^{x_1}}^{(1)} \geq \dots \tag{2.6}$$

and, similarly, for each $x_1 \in \mathcal{D}_1$ call $\xi_2^{x_1 x_2}$, $x_2 \in \mathcal{D}_2$, the vertices of \mathcal{V}_{N_2} such that

$$\Xi_{\xi_1^{x_1} \xi_2^{x_1 x_2}}^{(2)} \geq \Xi_{\xi_1^{x_1} \xi_2^{x_1 x_2}}^{(2)} \geq \dots \geq \Xi_{\xi_1^{x_1} \xi_2^{x_1 x_2}}^{(2)} \geq \dots \tag{2.7}$$

Let $\xi : \mathcal{D} \rightarrow \mathcal{V}_N$ be such that $\xi(x) = \xi^x := \xi_1^{x_1} \xi_2^{x_1 x_2}$. This is a one to one mapping for almost every realization of Ξ . Let now $X^N = X_1^N X_2^N$ be the mapping of σ^N on \mathcal{D} by the inverse of ξ . This is the process we will state scaling limit results for. This alternative representation suits our purpose of taking scaling limits, mainly due to the convenience of working with a state space which naturally extends to set of the natural numbers, which will be the state space of the limiting processes. The class to which these processes belong, namely K processes, is described in the appendix. In the build up for those scaling limit results, let us introduce next scaling factors, and then the scaling limit of the environment.

2.3 Scalings

Set

$$\alpha_2^N = \frac{\beta_*}{\beta} \sqrt{\frac{N_2}{N(1-a)}}, \quad c_2^N = \exp \left\{ -\frac{1}{\alpha_2^N} \left(\beta_*^2 N_2 - \frac{1}{2} \ln N_2 + \kappa \right) \right\}, \tag{2.8}$$

and define the scaled variables

$$\gamma_1^N(\sigma_1) \equiv c_1^N e^{\beta\sqrt{a_1 N} \Xi_{\sigma_1}^{(1)}} = e^{u_{N_1}^{-1}(\Xi_{\sigma_1}^{(1)})/\alpha_1^N} \tag{2.9}$$

$$\gamma_2^N(\sigma_1 \sigma_2) \equiv c_2^N e^{\beta\sqrt{a_2 N} \Xi_{\sigma_1 \sigma_2}^{(2)}} = e^{u_{N_2}^{-1}(\Xi_{\sigma_1 \sigma_2}^{(2)})/\alpha_2^N} \tag{2.10}$$

where for $i = 1, 2$, u_{N_i} is the scaling function for the maximum of 2^{N_i} i.i.d. standard Gaussians,

$$u_{N_i}(x) = \beta_* \sqrt{N_i} + \frac{1}{\beta_* \sqrt{N_i}} \{x - (\ln(N_i \ln 2) + \ln 4\pi)/2\}, \quad x \in \mathbb{R}. \quad (2.11)$$

For later use set

$$\psi_N^{-1} = \frac{N_1}{N_2} e^{\zeta_N + \kappa/\alpha_1^N}. \quad (2.12)$$

Clearly, in the fine tuning regime, $\beta \rightarrow \beta^{FT}, \alpha_1^N \rightarrow \alpha_1 \equiv 2^{\frac{p}{1-p}}$,

$$\alpha_2^N \rightarrow \alpha_2 \equiv 2\sqrt{\frac{a-p}{1-a} \frac{p}{1-p}} \text{ and } \psi_N^{-1} \rightarrow \psi^{-1} \equiv \frac{p}{1-p} e^{\zeta + \kappa \frac{1-p}{2p}}. \quad (2.13)$$

2.4 Scaling limit of the environment

For the remainder of this section, we assume that $\lim_{N \rightarrow \infty} \zeta_N/N_2$ exists. It follows that so does

$$\lim_{N \rightarrow \infty} \beta =: \bar{\beta}. \quad (2.14)$$

Let $\gamma_1^N = \{\gamma_1^N(x_1), x_1 \in \mathcal{D}_1\}$, and for $x_1 \in \mathcal{D}_1$, set $\gamma_2^{N,x_1} := \{\gamma_2^N(x_1 x_2), x_2 \in \mathcal{D}_2\}$, where $\gamma_1^N(x_1)$ and $\gamma_2^N(x_1 x_2)$ stand for $\gamma_1^N(\xi_1^{x_1})$ and $\gamma_2^N(\xi_2^{x_1 x_2})$, respectively. Then

$$\gamma_1^N \rightarrow \gamma_1, \quad \gamma_2^{N,x_1} \rightarrow \gamma_2^1, \quad (2.15)$$

$x_1 \in \mathbb{N}$, in distribution as $N \rightarrow \infty$, as point processes on \mathbb{R}^+ , and random measures on \mathbb{N} , where $\gamma_1 := \{\gamma_1(x_1), x_1 \in \mathbb{N}\}$, $\gamma_2^1 := \{\gamma_2(1 x_2), x_2 \in \mathbb{N}\}$ are independent Poisson point processes in \mathbb{R}^+ , enumerated in decreasing order, with respective intensity functions given by $\alpha_i/x^{1+\alpha_i}$, $i = 1, 2$, with

$$\alpha_1 = \frac{\beta_1^{cr}}{\bar{\beta}}, \quad \alpha_2 = \frac{\beta_2^{cr}}{\bar{\beta}}. \quad (2.16)$$

Notice that, as follows from our assumptions, $0 < \alpha_1 < \alpha_2 < 1$. We also have that

$$\gamma^N := \{\gamma_1^N(x_1)\gamma_2^N(x_1 x_2), x_1 x_2 \in \mathcal{D}\} \rightarrow \gamma := \{\gamma_1(x_1)\gamma_2(x_1 x_2), x_1 x_2 \in \mathbb{N}^2\} \quad (2.17)$$

in distribution as $N \rightarrow \infty$ as point processes on \mathbb{R}^+ , and (a.s. finite) random measures on \mathbb{N}^2 , where $\gamma_2^{x_1} := \{\gamma_2(x_1 x_2), x_2 \in \mathbb{N}\}$, $x_1 \geq 2$, are independent copies of γ_2^1 . We will sometimes below let γ_2 stand for the family $\{\gamma_2(x_1 x_2), x_1 x_2 \in \mathbb{N}^2\}$.

Remark 2.2. All of the convergence claims made in the above paragraph follow readily from convergence results of [14]. Indeed, we may apply Theorems 1.3 and 1.7 therein as follows. We preliminarily point out that in the cascading two-level system we are dealing with in the present paper, we have the following in terms of the notation of [14]: $n = m = 2$; $J_1 = 1, J_2 = 2$; $\bar{X}_{\sigma_1} = \Xi_{\sigma_1}^{(1)}, \bar{X}_{\sigma_2}^{\sigma_1} = \Xi_{\sigma_1 \sigma_2}^{(2)}$; $a_1 = \bar{a}_1 = a, a_2 = \bar{a}_2 = 1 - a$; $\alpha_1 = \bar{\alpha}_1 = 2^p, \alpha_2 = \bar{\alpha}_2 = 2^{1-p}$. We then have that for $\sigma_1 \sigma_2 \in \mathcal{V}_N$

$$\begin{aligned} \gamma_1^N(\sigma_1) &= \exp\{\beta u_{p \log 2, N}^{-1}(\sqrt{a} \Xi_{\sigma_1}^{(1)})\}, \\ \gamma_2^N(\sigma_1 \sigma_2) &= \exp\{\beta u_{(1-p) \log 2, N}^{-1}(\sqrt{1-a} \Xi_{\sigma_1 \sigma_2}^{(2)})\}, \end{aligned}$$

where $u_{\cdot, N}(\cdot)$ is defined in (1.7) of [14]. Theorem 1.3 of [14] now asserts the convergence of $\{u_{p \log 2, N}^{-1}(\sqrt{a} \Xi_{\sigma_1}^{(1)}), u_{(1-p) \log 2, N}^{-1}(\sqrt{1-a} \Xi_{\sigma_1 \sigma_2}^{(2)}); \sigma_1 \sigma_2 \in \mathcal{V}_N\}$ to the Poisson cascade introduced in [14]. (2.15) and (2.17) follow from that and Theorem 1.7 in the same reference after straightforward considerations – see also Proposition 1.8 of [14].

Remark 2.3. It follows from the above results that the Gibbs measure $G_{\beta, N}$ converges suitably to $G_{\bar{\beta}}$ – the normalized γ – as $N \rightarrow \infty$.

¹In this sentence, α_1 and α_2 are notations from [14], and should not be confused with the notation of the present paper introduced in (2.16).

2.5 Scaling limit of X^N

In order to have the three cases outlined in the heuristics discussion, namely, above, at and below fine tuning temperatures, we need that $\beta^{FT} > \beta_2^{cr}$, namely, that

$$\sqrt{\frac{1-p}{1-a}} < \frac{1-p}{2\sqrt{pa}}; \tag{2.18}$$

otherwise, all low temperatures according to (2.2) are below fine tuning. In each case we find a different scaling and different scaling limit for X^N .

To state the first theorem, we take ζ_N^+ as in Definition 2.1, with the extra assumption (2.14), i.e., we let the sequence of real numbers ζ_N^+ satisfy

$$\lim_{N \rightarrow \infty} \zeta_N^+ = \infty \quad \text{and} \quad \lim_{N \rightarrow \infty} \zeta_N^+/N_2 < \frac{\beta_*^2}{2} \left(1 - 2\sqrt{\frac{p}{1-p}} \sqrt{\frac{a}{1-a}} \right). \tag{2.19}$$

The latter condition is equivalent to (2.2) once we replace ‘lim inf’ by ‘lim’ there. Set

$$\tilde{X}^N(t) = X^N(t/c_2^N), \quad t \geq 0. \tag{2.20}$$

We recall that this is a process in the random environment γ^N . The limiting processes, which are K processes, described in the appendix, will naturally also be processes in random environment.

Consider $f_1, \mathfrak{w}_1 : \mathbb{N}^2 \rightarrow (0, \infty)$ such that $f_1(x_1x_2) = \tilde{\gamma}_2(x_1x_2) := \frac{1}{1-p}\gamma_2(x_1x_2)$, $\mathfrak{w}_1(x_1x_2) = \gamma_1(x_1)$ for all $x_1x_2 \in \mathbb{N}^2$. These functions will play the role of random environment for the limiting process in this case.

Theorem 2.4 (Above fine tuning temperatures). *As $N \rightarrow \infty$*

$$\tilde{X}^N \Rightarrow \mathcal{K}(f_1, \mathfrak{w}_1); \tag{2.21}$$

where \Rightarrow stands for convergence in $\mathcal{P} \otimes \mathbb{P}_\mu$ -distribution. The convergence takes place on the Skorohod space of trajectories of both processes, with the J_1 metric.

See the definition of $\mathcal{K}(\cdot, \cdot)$ in the first subsection of the appendix.

To state our second theorem, we assume $\lim_{N \rightarrow \infty} \zeta_N = \zeta$ for some real finite ζ .

Let $f_2 : \mathbb{N} \rightarrow (0, \infty)$, $f'_2 : \mathbb{N}^2 \rightarrow (0, \infty)$ such that $f_2(x_1) = \tilde{\gamma}_1(x_1) := \psi\gamma_1(x_1)$, $f'_2(x_1, x_2) = \tilde{\gamma}_2(x_1, x_2)$ for all $x_1x_2 \in \mathbb{N}^2$.

Theorem 2.5 (At fine tuning temperatures). *As $N \rightarrow \infty$*

$$\tilde{X}^N \Rightarrow \mathcal{K}_2(f_2, f'_2). \tag{2.22}$$

The convergence takes place in the Skorohod space of trajectories of both processes, with the J_1 metric.

See the definition of $\mathcal{K}_2(\cdot, \cdot)$ in the first subsection of the appendix.

Remark 2.6. In order for the above mentioned two-level K-process to be well defined, we need to make sure that f_2, f'_2 satisfy (almost surely) the summability conditions (A.1, A.4). This is a classic result for (A.1) – recall that $\mathfrak{w} \equiv 1$ in this case –, and follows by standard arguments for (A.4) from the fact that $\alpha_1 < \alpha_2 < 1$ (as noted above, below (2.16)).

For our last theorem, we take ζ_N^- as in Definition 2.1.

Let $\bar{c}^N = c_1^N 2^{N_2} c_2^N$ and make

$$\bar{X}^N(t) = X^N(t/\bar{c}^N), \quad t \geq 0. \tag{2.23}$$

Let also $f_3 : \mathbb{N} \rightarrow (0, \infty)$, with $f_3(x_1) = \gamma_1(x_1) \sum_{x_2 \in \mathbb{N}} \tilde{\gamma}_2(x_1x_2)$, $x_1 \in \mathbb{N}$.

Theorem 2.7 (Below fine tuning temperatures). As $N \rightarrow \infty$

$$\bar{X}^N \Rightarrow \bar{X}_1 \bar{X}_2, \tag{2.24}$$

where $\bar{X}_1 \sim \mathcal{K}(f_3, 1)$ and, given γ_2 and $\bar{X}_1 = x_1 \in \mathbb{N}$, \bar{X}_2 is an iid family of random variables on \mathbb{N} (indexed by time) each of which is distributed according to the weights given by $\gamma_2^{x_1}$. The marginal convergence of the first coordinate takes place in the Skorohod space of trajectories of both processes, with the J_1 metric, and the convergence of the second coordinate is in the sense of finite dimensional distributions only.

Remark 2.8. If Condition 2.18 is not satisfied (within the cascading, low temperature regime treated in this paper), then we are below fine tuning temperatures and Theorem 2.7 holds for all $\beta > \beta_2^{cr}$ temperatures, as can be readily checked from its proof. The other regimes are not present in this case.

Remark 2.9. It is either known or follows readily from known results that the limiting processes in the above theorems are ergodic Markov processes², having the infinite volume Gibbs measure $G_{\bar{\beta}}$ (see Remark 2.3 above) as their unique equilibrium distribution. See [21] for the case of the 2-level K-process, and [17] for the cases involving weighted/uniform K-processes.

Remark 2.10. As discussed earlier (see the one before last paragraph of Section 1.1) we may then say, after Remark 2.9, that the time scale $1/c_2^N$ is an extreme time scale at and above fine tuning, and the time scale $1/\bar{c}^N$ is an extreme time scale below fine tuning.

2.6 Strategy

The proofs of all the above theorems have the same structure³, reflecting the fact that at extreme time scales, the dynamics lives on ground state configurations. In each case, there are three ingredients:

1. Showing that the process spends virtually all of the time on ground state configurations;
2. Getting sharp estimates on the time spent on each visit to a ground state configurations;
3. Getting sharp estimates on the transition probabilities between ground state configurations.

Point 1 is quite simple in each case, resulting from straightforward expected value computations.

Point 2 is immediate for Theorems 2.4 and 2.5, since in those cases the time scale coincides with the order of magnitude of the waiting times at the ground state configurations. For Theorem 2.7, it is a longer time scale, corresponding to the time spent on the second level configurations before jumping from a given first level ground state configuration; the main point of this estimation involves deriving a law of large numbers for the number of visits to a given second level ground state configuration before the jump from the first level ground state configuration occurs – see (7.6).

As for Point 3, in Theorem 2.7 the transition between first level ground state configurations are already well known to be approximately uniform; and this time scale is well above the equilibration time scale for the second level, so the transitions we see are between independent configurations of equilibrium – we get that from known spectral gap estimates.

²I.e., Markov processes that have an irreducible state space and a unique invariant distribution.

³Except for the second level dynamics part in Theorem 2.7.

The bulk of our work comprise the estimation of the transition laws for Theorems 2.4 and 2.5. In the latter case, we heavily rely on a potential theoretic point of view, extending the *lumping* approach adopted in the analysis of the Random Hopping dynamics for the REM in [3], and further developed in [6]. See Proposition 3.1. There are many fine points to be considered here, making for a long analysis. The treatment of Theorem 2.4 demands a finer estimation, on the one hand – see Proposition 3.3 and the discussion around its statement; on the other hand, a more direct approach works in this case.

3 Entrance law. Main result

In this and the next three sections, we will mostly not be concerned with limits, so we find it convenient to revert to the original representation of σ^N as spin configuration. Given a subset $A \subset \mathcal{V}_N$, the hitting times τ_A of A by the continuous and discrete time processes σ^N and J_N^* are defined, respectively, as

$$\tau_A = \inf\{t > 0 \mid \sigma^N(t) \in A\} \quad \text{and} \quad \tau_A = \inf\{i \in \mathbb{N} \mid J_N^*(i) \in A\}. \tag{3.1}$$

3.1 The Top

Given two integers $M_i < 2^{N_i}$, $i = 1, 2$, set $\mathcal{M} = \mathcal{M}_1 \times \mathcal{M}_2$ where $\mathcal{M}_i = \{1, \dots, M_i\}$, $i = 1, 2$. We then let the Top be the set

$$T \equiv T((M_i)_{i \leq 2}, (N_i)_{i \leq 2}) \equiv \{\sigma \in \mathcal{V}_N \mid \sigma = \xi_1^{x_1} \xi_2^{x_1 x_2}, x_1 x_2 \in \mathcal{M}\} \tag{3.2}$$

Note that we may also write $T = \cup_{x_1 \in \mathcal{M}_1} T^{x_1}$ where, for each $x_1 \in \mathcal{M}_1$

$$T^{x_1} \equiv T^{x_1}(M_1, N_1) \equiv \{\sigma \in \mathcal{V}_N \mid \sigma = \xi_1^{x_1} \xi_2^{x_1 x_2}, x_2 \in \mathcal{M}_2\}. \tag{3.3}$$

Further introduce the canonical projection of T on \mathcal{V}_{N_1} ,

$$T_1 \equiv T_1(M_1, N_1) \equiv \{\sigma_1 \in \mathcal{V}_{N_1} \mid \sigma = \xi_1^{x_1}, x_1 \in \mathcal{M}_1\}. \tag{3.4}$$

To each $\xi_1^{x_1}$ in T_1 we associate the cylinder set

$$W^{x_1} \equiv W^{x_1}((N_i)_{i \leq 2}) \equiv \{\sigma \in \mathcal{V}_N \mid \sigma_1 = \xi_1^{x_1}\}. \tag{3.5}$$

Clearly, T^{x_1} is the restriction of T to this cylinder, $T^{x_1} = W^{x_1} \cap T$. Finally, set

$$\overline{W} \equiv \cup_{x_1 \in \mathcal{M}_1} W^{x_1}. \tag{3.6}$$

3.2 Main entrance law results

From now on we fix (ζ_N) , a sequence of real numbers such that $\beta = \beta(a, p, N, \zeta_N) > 0$ for all N , and let ψ_N be as in (2.12). For each $x_1 \in \mathcal{M}_1$ and $A \subseteq T^{x_1}$ set

$$\lambda_N^{x_1}(A) \equiv \lambda_N^{x_1}(|A|, N, \psi_N) = \frac{1}{1 + |A| \psi_N \gamma_1^N(\xi_1^{x_1})}. \tag{3.7}$$

We will see that this quantity can be interpreted as the probability that, starting in W^{x_1} , the process exits W^{x_1} before finding an element of A . Note that $\lambda_N^{x_1}(A)$ is a random variable. We use it to define the random probability measure ν_1 on \mathcal{M}_1 that assigns to x_1 the mass

$$\nu_1^N(x_1) = \frac{1 - \lambda_N^{x_1}(T^{x_1})}{\sum_{x'_1 \in \mathcal{M}_1} (1 - \lambda_N^{x'_1}(T^{x'_1}))} \tag{3.8}$$

Similarly, given $\bar{\eta} \in T$, we denote by $\bar{\nu}_1$ the random measure on \mathcal{M}_1 defined through

$$\bar{\nu}_1^N(x_1) = \frac{1 - \lambda_N^{x_1}(T^{x_1} \setminus \bar{\eta})}{\sum_{x'_1 \in \mathcal{M}_1} (1 - \lambda_N^{x'_1}(T^{x'_1} \setminus \bar{\eta}))} \tag{3.9}$$

(where clearly $T^{x_1} \setminus \bar{\eta} = T^{x_1}$ if $\bar{\eta} \notin T^{x_1}$).

Proposition 3.1. *There exists a subset $\tilde{\Omega} \subset \Omega$ with $\mathcal{P}(\tilde{\Omega}) = 1$ such that on $\tilde{\Omega}$, for all N large enough, the following holds in the temperature domain determined by $\beta = \beta(a, p, N, \zeta_N) > 0$ and $\zeta_N \ll \log N$. Let $\varepsilon_N = \mathcal{O}(N^{-1})$.*

i) Entrance law. Let $x_1 \in \mathcal{M}_1$ and $\eta \in T^{x_1}$. Then

i-1) for all $\sigma \in W^{x_1} \setminus T^{x_1}$

$$\mathbb{P}_\sigma(\tau_\eta < \tau_{T \setminus \eta}) = \frac{1}{M_2} [(1 - \lambda_N^{x_1}(T^{x_1})) + \nu_1^N(x_1)\lambda_N^{x_1}(T^{x_1})] (1 + \varepsilon_N), \quad (3.10)$$

i-2) for all $x'_1 \in \mathcal{M}_1 \setminus x_1$ and all $\sigma \in W^{x'_1} \setminus T^{x'_1}$

$$\mathbb{P}_\sigma(\tau_\eta < \tau_{T \setminus \eta}) = \frac{1}{M_2} \nu_1^N(x_1)\lambda_N^{x'_1}(T^{x'_1}) (1 + \varepsilon_N), \quad (3.11)$$

i-3) for all $\sigma \in \mathcal{V}_N \setminus \bar{W}$

$$\mathbb{P}_\sigma(\tau_\eta < \tau_{T \setminus \eta}) = \frac{\nu_1^N(x_1)}{M_2} (1 + \varepsilon_N). \quad (3.12)$$

i-4) Entrance in cylinder sets. For all $\sigma \in \mathcal{V}_N \setminus \bar{W}$ and all $x_1 \in \mathcal{M}_1$

$$\mathbb{P}_\sigma(\tau_{W^{x_1} \setminus T^{x_1}} < \tau_{\bar{W} \setminus (W^{x_1} \setminus T^{x_1})}) = \frac{1}{M_1} (1 + \varepsilon_N). \quad (3.13)$$

ii) Level 2 transitions. For all $\eta \in T^{x_1}$, $\bar{\eta} \in T^{\bar{x}_1}$, $x_1 = \bar{x}_1$, $\eta \neq \bar{\eta}$,

$$\mathbb{P}_{\bar{\eta}}(\tau_\eta < \tau_{T \setminus \{\eta, \bar{\eta}\}}) = \frac{1}{M_2 - 1} [(1 - \lambda_N^{x_1}(T^{x_1} \setminus \bar{\eta})) + \bar{\nu}_1^N(x_1)\lambda_N^{x_1}(T^{x_1} \setminus \bar{\eta})] (1 + \varepsilon_N). \quad (3.14)$$

iii) Level 1 transitions. For all $\eta \in T^{x_1}$, $\bar{\eta} \in T^{\bar{x}_1}$, $x_1 \neq \bar{x}_1$

$$\mathbb{P}_{\bar{\eta}}(\tau_\eta < \tau_{T \setminus \{\eta, \bar{\eta}\}}) = \frac{1}{M_2} \bar{\nu}_1^N(x_1)\lambda_N^{x_1}(T^{\bar{x}_1} \setminus \bar{\eta}) (1 + \varepsilon_N). \quad (3.15)$$

Remark 3.2. Taking limits of the above quantities we get, for ψ as in (2.13)

$$\lim_{N \rightarrow \infty} \lambda_N^{x_1}(T^{x_1}) = \begin{cases} 1, & \text{above FT,} \\ \frac{1}{1 + M_2 \psi \gamma_1(x_1)}, & \text{at FT,} \\ 0, & \text{below FT,} \end{cases} \quad (3.16)$$

which leads to

$$\lim_{N \rightarrow \infty} \nu_1^N(x_1) = \nu_1(x_1), \quad \nu_1(x_1) \equiv \begin{cases} \frac{\gamma_1(x_1)}{\sum_{z_1 \in \mathcal{M}_1} \gamma_1(z_1)}, & \text{above FT,} \\ \frac{h(\gamma_1(x_1))}{\sum_{z_1 \in \mathcal{M}_1} h(\gamma_1(z_1))}, & \text{at FT,} \\ \frac{1}{M_1}, & \text{below FT,} \end{cases} \quad (3.17)$$

where $h = h_{M_2, \psi} : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is such that $h(r) = r/(1 + M_2 \psi r)$. At FT both limits hold weakly with respect to the environment. Below FT, there is a window of values of ζ_N^- for which both limits hold almost surely, and above which both hold in probability. One may readily check that the following window has these properties: $\zeta_N^- \ll -\log \log N$; see Lemma 5.3, and its proof. Above FT, we have a mixed situation. For $\lambda_N^{x_1}(T^{x_1})$, there is a window above which the convergence is almost sure: $\zeta_N^+ \gg \log N$. And for $\nu_1^N(x_1)$, we need in addition the existence of $\lim_{N \rightarrow \infty} \zeta_N/N$, and the convergence is weak. The asymptotics of the probabilities follow readily.

Getting the estimates in Proposition 3.1 above fine tuning when we do not have that $\zeta_N \ll \log N$, requires an extra level of precision, related to the fact that, in that regime, $\nu_1^N(\cdot)$ is a quotient of vanishing terms. We state next a separate result where we deal with this case. Since it is a limit result, we require the existence of $\lim_{N \rightarrow \infty} \zeta_N/N$.

Proposition 3.3 (Above fine tuning temperatures). *Suppose (2.19) holds. Then for all $\sigma \notin T$ and $\eta = \xi^y \in T$*

$$\lim_{N \rightarrow \infty} \mathbb{P}_\sigma(\tau_\eta < \tau_{T \setminus \eta}) = \frac{1}{M_2} \nu_1(y), \tag{3.18}$$

where the limit holds in distribution in $(\Omega, \mathcal{F}, \mathcal{P})$.

The proof of Proposition 3.1 follows a strategy initiated in [3] for the study of aging in the REM and further developed in [6] with the main object of providing the tools to tackle more general spin glass models, such as the GREM. These tools are prepared in Section 4. They are used in Section 5 to prove basic probability estimates for the jump chain that, in turn, are the key ingredients of the proof of Proposition 3.1, concluded in Section 6, after which we prove Proposition 3.3.

4 Entrance law. Key tools

As announced above, the proof of Proposition 3.1 naturally draws upon [6] since that paper was designed precisely for that purpose. The strategy pursued in [6] (and initiated in [3] and [11]) consists, firstly, in reducing the probabilities of interest, say, the harmonic measure of a given set, A , to quantities which are functions only of the simple random walks $J_{N_i}^\circ$ and, secondly, in using lumping techniques (a coarse graining of the configuration space that is built from the set A) to express these quantities through a new chain, called *lumped chain*. This lumped chain, which is again Markovian, can be thought of as a (discrete analogue of a) diffusion in a convex potential (whose properties strongly depend on A) and is finally studied with precision using the potential theoretic approach developed in [11, 12]. Going back to the original state space, the resulting estimates on the harmonic measure of A come with conditions, which essentially relate to specific properties of the set A (e.g. its size and sparseness).

To see how the HRHD (1.4) links up to the simple random walk, first observe that it can be described alternatively through its jump chain, J_N^* , and jump rates, w_N , where

$$w_N(\sigma) = \sum_{\sigma' \sim \sigma} w_N(\sigma, \sigma') \tag{4.1}$$

and $(J_N^*(i), i \in \mathbb{N})$ is the discrete time Markov chain with one step transition probabilities

$$p_N^*(\sigma, \sigma') = w_N(\sigma, \sigma') w_N^{-1}(\sigma). \tag{4.2}$$

Then, introducing the parameters

$$q_N^*(\sigma_1) \equiv \frac{1}{1 + \frac{N_2}{N_1} e^{-\beta H_N^{(1)}(\sigma_1)}}, \tag{4.3}$$

one has that

$$p_N^*(\sigma, \sigma') = q_N^*(\sigma_1) p_{N_1}^\circ(\sigma_1, \sigma'_1) + (1 - q_N^*(\sigma_1)) p_{N_2}^\circ(\sigma_2, \sigma'_2) \tag{4.4}$$

where, for $i = 1, 2$, $p_{N_i}^\circ(\sigma_i, \sigma'_i) = N_i^{-1}$ if $\sigma_i \sim \sigma'_i$ and $p_{N_i}^\circ(\sigma_i, \sigma'_i) = 0$ else denote the one step transition probabilities of the simple random walk $(J_{N_i}^\circ(j), j \in \mathbb{N})$ on \mathcal{V}_{N_i} . Clearly, the J_N^* is reversible w.r.t. the measure $G_{\beta, N}^*$ defined through

$$C_{\beta,N}^*(\sigma) = w_N(\sigma)e^{-\beta H_N(\sigma)} (Z_{\beta,N}^*)^{-1} = (N_1/N) (q_N^*(\sigma_1)Z_{\beta,N}^*)^{-1} \tag{4.5}$$

where $Z_{\beta,N}^*$ is a normalization making this measure a probability.

In line with the strategy described above, this section has three parts. Subsection 4.1 gathers simple lemmata needed to link probabilities for the original chain σ^N (with rates (1.4)) to probabilities for the jump chain J_N^* (with transitions (4.4)), and to link the latter to quantities depending only on the simple random walks $J_{N_i}^\circ$. In Subsection 4.2 we introduce the notion of lumped chains. Finally, in Subsection 4.3, we state and prove the properties of various sets that are needed in Section 5 to make use of the known lumped chain estimates from [6]. This last section can be skipped at first reading.

For future reference, we call \mathbb{P}^* the law of the process J_N^* conditional on \mathcal{F} . We denote by $\mathbb{P}^{\circ,i}$ the law of $J_{N_i}^\circ$, $i = 1, 2$. If the initial state, say η , has to be specified we write \mathbb{P}_η , \mathbb{P}_η^* and $\mathbb{P}_{\eta_i}^{\circ,i}$. We will denote by $\mathcal{P} \otimes \mathbb{P}_\eta$ the probability measure obtained by integrating \mathbb{P}_η with respect to \mathcal{P} . Expectation with respect to \mathbb{P} , \mathbb{P}^* , $\mathbb{P}^{\circ,i}$, \mathcal{P} and $\mathcal{P} \otimes \mathbb{P}_\mu$ are denoted by \mathbb{E} , \mathbb{E}^* , $\mathbb{E}^{\circ,i}$, \mathcal{E} and $\mathcal{E} \otimes \mathbb{E}_\mu$, respectively, where μ is the uniform probability measure on \mathcal{V}_N .

4.1 Comparison lemmata

Our starting point is the observation that

Lemma 4.1. *For all $A, B \subseteq \mathcal{V}_N$ such that $A \cap B = \emptyset$ and for all $\sigma \in \mathcal{V}_N \setminus (A \cup B)$,*

$$\mathbb{P}_\sigma(\tau_A < \tau_B) = \mathbb{P}_\sigma^*(\tau_A < \tau_B). \tag{4.6}$$

We skip the proof of Lemma 4.1 which is elementary.

The next two lemmata deal with two classes of events that can be expressed through just one of the simple random walks $J_{N_i}^\circ$ on \mathcal{V}_{N_i} . The first are *REM-like events* that can be reduced to those of a REM (which is a 1-level GREM). Let π_i , $i = 1, 2$, denote the canonical projection of \mathcal{V}_N onto \mathcal{V}_{N_i} , that is

$$\pi_i \sigma = \sigma_i. \tag{4.7}$$

Lemma 4.2 (REM-like events). *Let $A, B \subseteq \mathcal{V}_{N_1}$ be such that $A \cap B = \emptyset$. Then, for all $\sigma_1 \in \mathcal{V}_{N_1} \setminus A$ and all $\sigma \in \pi_1^{-1} \sigma_1$,*

$$\mathbb{P}_\sigma^* \left(\tau_{\pi_1^{-1}A} < \tau_{\pi_1^{-1}B} \right) = (1 + \delta_{\sigma_1 \in B} (q_N^*(\sigma_1) - 1)) \mathbb{P}_{\sigma_1}^{\circ,1}(\tau_A < \tau_B). \tag{4.8}$$

Proof. Note that $\mathbb{P}_\sigma^*(\tau_{\pi_1^{-1}A} < \tau_{\pi_1^{-1}B}) = \mathbb{P}_{\sigma_1}^{\pi_1}(\tau_A < \tau_B)$ where $\mathbb{P}_{\sigma_1}^{\pi_1} = \mathbb{P}_\sigma^* \circ \pi_1^{-1}$ denotes the law of the projection $\pi_1 J_N^*$ of the jump chain on \mathcal{V}_{N_1} . By (4.4) this is a Markov chain with transition probabilities $p_N^\pi(\sigma_1, \sigma'_1) = q_N^*(\sigma_1) p_{N_1}^\circ(\sigma_1, \sigma'_1) \mathbb{1}_{\sigma_1 \sim \sigma'_1} + (1 - q_N^*(\sigma_1)) \mathbb{1}_{\sigma_1 = \sigma'_1}$. The lemma now easily follows. \square

Given $\sigma = \sigma_1 \sigma_2 \in \mathcal{V}_N$, define the cylinder sets

$$C(\sigma_i) \equiv \pi_i^{-1} \sigma_i = \{\sigma' \in \mathcal{V}_N \mid \sigma'_i = \sigma_i\}, \quad i = 1, 2. \tag{4.9}$$

The next lemma deals with so-called *level-2 events*, namely, events whose trajectories are confined to a given cylinder set $C(\sigma_1)$, and that can thus be expressed through just the simple random walk $J_{N_2}^\circ$ on \mathcal{V}_{N_2} . Define the outer boundary of a set $A \subset \mathcal{V}_N$ as

$$\partial A \equiv \{\sigma' \in (\mathcal{V}_N \setminus A) \mid \exists \sigma \in A \text{ s.t. } \sigma \sim \sigma'\}. \tag{4.10}$$

Lemma 4.3 (Level-2 events). *Given $\sigma_1 \in \mathcal{V}_{N_1}$, let $A, B \subseteq C \equiv C(\sigma_1)$ be such that $A \cap B = \emptyset$. Set $u(\sigma_1) \equiv \log(1 - q_N^*(\sigma_1))$. Then, for all $\sigma \in C$,*

$$\mathbb{P}_\sigma^*(\tau_A \leq \tau_{B \cup \partial C}) = \begin{cases} \mathbb{E}_{\sigma_2}^{\circ,2} \left(e^{u(\sigma_1) \tau_{\pi_2 A}} \mathbb{1}_{\{\tau_{\pi_2 A} < \tau_{\pi_2 B}\}} \right) & \text{if } B \neq \emptyset, \\ \mathbb{E}_{\sigma_2}^{\circ,2} \left(e^{u(\sigma_1) \tau_{\pi_2 A}} \right) & \text{if } B = \emptyset. \end{cases} \tag{4.11}$$

Proof. Write

$$\mathbb{P}_\sigma^*(\tau_A \leq \tau_{B \cup \partial C}) = \sum_{k=1}^\infty \mathbb{P}_\sigma^*(k = \tau_A \leq \tau_B \mid \tau_{\partial C} > k) \mathbb{P}_\sigma^*(\tau_{\partial C} > k) \tag{4.12}$$

and note that by (4.4), $\tau_{\partial C}$ is a geometric r.v. with success probability $q_N^*(\sigma_1)$. Thus, on the one hand, $\mathbb{P}_\sigma^*(\tau_{\partial C} > k) = (1 - q_1^*(\sigma_1))^k$ while on the other hand $\mathbb{P}_\sigma^*(k = \tau_A \leq \tau_B \mid \tau_{\partial C} > k) = \mathbb{P}_{\pi_2\sigma}^{\circ,2}(k = \tau_{\pi_2 A} \leq \tau_{\pi_2 B})$. \square

Estimates on the probabilities $\mathbb{P}_{\sigma_1}^{\circ,1}(\tau_A < \tau_B)$ appearing in Lemma 4.2 and the Laplace transform of Lemma 4.3 are derived in [6] using lumping techniques. For the one-dimensional case (when the set A reduces to a singleton), lumping reduces to the classical Ehrenfest chain. We recall below an expression for the probability generating function of hitting times of such chain, appearing in [27] (see (4.28,29) in that reference), and needed later. For $t \in [0, 1)$ and $\sigma_2, \sigma_2' \in \mathcal{V}_{N_2}$

$$\mathbb{E}_{\sigma_2}^{\circ,2}(t^{\tau_{\sigma_2'}}) = \frac{B_i(t')}{B_0(t')}, \tag{4.13}$$

where $i = d_2(\sigma_2, \sigma_2')$, $t' = \frac{N_2}{2} \frac{1-t}{t}$, and for $\alpha > 0$

$$B_i(\alpha) = \int_0^1 (1-u)^i (1+u)^{N_2-i} u^{\alpha-1} du = \sum_{j=0}^{N_2-i} \binom{N-i}{j} \frac{\Gamma(i+1)\Gamma(\alpha+j)}{\Gamma(\alpha+i+j+1)}. \tag{4.14}$$

4.2 Lumped chains and K -lumped chains

In this section we introduce certain functions of the simple random walks $J_{N_i}^\circ$ on \mathcal{V}_{N_i} , $i \in \{1, 2\}$, that play a key rôle in our proofs. Fix $i \in \{1, 2\}$. Given a partition Λ_i of $\{1, \dots, N_i\}$ into d classes, that is, non-empty disjoint subsets $\Lambda_i^1, \dots, \Lambda_i^d$, $1 \leq d \leq N_i$, satisfying $\Lambda_i^1 \cup \dots \cup \Lambda_i^d = \{1, \dots, N_i\}$, let m_i be the many-to-one function that maps the elements of \mathcal{V}_{N_i} onto d -dimensional vectors

$$m_i(\sigma_i) = (m_i^1(\sigma_i), \dots, m_i^k(\sigma_i), \dots, m_i^d(\sigma_i)), \quad \sigma_i \in \mathcal{V}_{N_i} \tag{4.15}$$

by setting, for all $k \in \{1, \dots, d\}$,

$$m_i^k(\sigma_i) \equiv \frac{1}{|\Lambda_i^k|} \sum_{j \in \Lambda_i^k} \sigma_{i,j} \tag{4.16}$$

where $\sigma_{i,j}$ denotes the j -th cartesian co-ordinate of σ_i . The image $\bar{I}_i \equiv m_i(J_{N_i}^\circ)$ of the simple random walk $J_{N_i}^\circ$, called *lumped chain*, also is a Markov chain that now takes value in a discrete grid $\Gamma_{N_i,d}$ that contains $\mathcal{V}_d = \{-1, 1\}^d$.

Different choices of the partition Λ_i yield different lumped chains. Given an integer n and a collection $K = \{\eta^1, \dots, \eta^x, \dots, \eta^n\}$ of elements of \mathcal{V}_{N_i} , the so-called *K -lumped chain* is induced by a partition $\Lambda_i(K)$ of $\{1, \dots, N_i\}$ into $d = 2^n$ classes, $\{1, \dots, N_i\} = \Lambda_i^1(K) \cup \dots \cup \Lambda_i^d(K)$, defined as follows. Let us identify the set K with the $n \times N_i$ matrix $(\eta_j^x)_{j=1, \dots, N_i}^{x=1, \dots, n}$ whose row vectors are the η^x 's, that is, $\eta^x \equiv (\eta_j^x)_{j=1, \dots, N_i} \in \mathcal{V}_{N_i}$, $x \in \{1, \dots, n\}$, and denote by η_j the column vectors $\eta_j \equiv (\eta_j^x)_{x=1, \dots, n} \in \mathcal{V}_n$, $j \in \{1, \dots, N_i\}$. Given an arbitrary labelling $\{e_1, \dots, e_k, \dots, e_d\}$ of the set of all $d = 2^n$ elements of \mathcal{V}_n , we then set

$$\Lambda_i^k(K) \equiv \{j \in \{1, \dots, N_i\} \mid \eta_j = e_k\}, \quad 1 \leq k \leq d. \tag{4.17}$$

We denote by $m_{i,K}$ the function (4.15)-(4.16) resulting from (4.17), by

$$\bar{I}_{i,K} \equiv m_{i,K}(J_{N_i}^\circ) \tag{4.18}$$

the associated K -lumped chain and by $\bar{\mathbb{P}}^{i,K}$ its law.

As mentioned earlier the K -lumped chain should be thought of as a discrete analogue of a diffusion in a convex potential (at least for values of d of order $o(\sqrt{N_i})$) with a drift induced by the convex potential $\psi_{i,K} \equiv -\frac{1}{N_i} \log |(m_{i,K})^{-1}(z)| + \log 2$, $z \in \Gamma_{N_i,d}$. This potential is very steep close to its global maxima, achieved at the 2^d vertices of \mathcal{V}_d , and has a unique global minimum, henceforth called the origin. Furthermore, there is a one-to-one correspondance between the set K and its image $\mathcal{K} \equiv m_{i,K}(K)$, and $\mathcal{K} \subseteq \mathcal{V}_d$. Intuitively, this chain exhibits the following scenario typical in the large deviation context of Freidlin and Wentzell: before reaching any vertex of \mathcal{V}_d it explores the bulk of the grid $\Gamma_{N_i,d}$ and remains there for an exponentially (in N_i) large time, revisiting the origin an exponential number of times before, at some random time, it quickly travels (in a polynomial time) from the origin to a vertex $z \in \mathcal{V}_d$, with almost uniform distribution. Analysing this behavior using the potential theoretic tools developed in [11, 12] for the analysis of metastability, Ref. [6] gives (among other results) precise sufficient conditions for the hitting time of subsets \mathcal{K} of \mathcal{V}_d to be asymptotically exponentially distributed, and for the hitting distribution to be uniform. These conditions essentially require that \mathcal{K} should be sparse enough (see Definition 1.2 in [6]) and that the partition $\Lambda_i(K)$ does not contain too many small boxes $\Lambda_i^k(K)$ (which would give flat directions to the potential). Because each vertex of \mathcal{V}_d has exactly one pre-image by $m_{i,K}$ these results and their conditions can be translated back for the original process in the original state space.

The conditions mentioned above as well as the precision of the estimates of [6] are expressed through the following key quantities: given a subset K of \mathcal{V}_{N_i} define

$$U_{N_i,d_i}(\sigma, K) \equiv \sum_{\eta \in A \setminus \sigma} F_{N_i,d_i}(\text{dist}(\sigma, \eta)), \quad \sigma \in \mathcal{V}_{N_i}, \quad (4.19)$$

$$\mathcal{U}_{N_i,d_i}(K) \equiv \max_{\sigma \in A} U_{N_i,d_i}(\sigma, K). \quad (4.20)$$

We further set $\mathcal{U}_{N_i,d_i}(K) = 0$ if $K = \emptyset$ and $U_{N_i,d_i}(\sigma, K) = 0$ if $K \setminus \sigma = \emptyset$. The function F_{N_i,d_i} in (4.19) is a function of the Hamming distance between the elements of A (depending on N_i and d_i), whose definition is explicit but pretty involved. We therefore refer to (3.5)-(3.8) of Section 3 of [6] for this definition. Its properties are analyzed in detail in Appendix A3 of [6].

4.3 Properties of the Top and other sets

The aim of this section is to prepare the ground for the use of the results of [6] by proving bounds for the quantities $\mathcal{U}_{N_i,d_i}(K)$, $U_{N_i,d_i}(\sigma, K)$ and F_{N_i,d_i} defined, respectively, in (4.19) and (4.20) above and in (3.5)-(3.8) of [6], for the three types of sets K that we will later encounter in our proofs: the Top, the Top plus a non random point, and large random subsets of \mathcal{V}_{N_1} .

In what follows $K = \{\eta^1, \dots, \eta^x, \dots, \eta^{m_i}\}$ denotes a collection of m_i elements of \mathcal{V}_{N_i} , and $\Lambda_i(K)$ is the partition of $\{1, \dots, N_i\}$ into $d_i = 2^{m_i}$ classes, $\Lambda_i^k(K)$, $1 \leq k \leq d_i$, induced by K through (4.17).

4.3.1 The Top

Consider the partitions $\Lambda_1(T_1)$ and $\Lambda_2(T^{x_1})$ induced respectively by T_1 and $\pi_2 T^{x_1}$, $x_1 \in \mathcal{M}_1$, through (4.17). Let K be any of the sets T_1 or $\pi_2 T^{x_1}$, $x_1 \in \mathcal{M}_1$ (thus $m_i = M_i$ and $i = 1$ if $K = T_1$ and $i = 2$ if $K = \pi_2 T^{x_1}$, $x_1 \in \mathcal{M}_1$). Introducing the sets

$$\bar{\Omega}_{N_i}(K) \equiv \left\{ \left| \frac{d_i}{N_i} |\Lambda_i^k(K)| - 1 \right| < \delta_i(N_i), \quad 1 \leq k \leq d_i \right\}, \quad \delta_i(N_i) \equiv 2\sqrt{\frac{d_i}{N_i}} \log N_i$$

and $\bar{\Omega}_i(K) \equiv \bigcup_{N'_i \geq i} \bigcap_{N_i \geq N'_i} \bar{\Omega}_{N_i}(K)$, define

$$\bar{\Omega} \equiv \bar{\Omega}_1(T_1) \cap \left(\bigcap_{x_1 \in \mathcal{M}_1} \bar{\Omega}_2(T^{x_1}) \right). \quad (4.21)$$

Lemma 4.4. $\mathcal{P}(\bar{\Omega}) = 1$.

Proof. The proof is an easy adaptation of that of Lemma 4.2 of [22]. \square

For $i = 1, 2$, $\eta \in \mathcal{V}_{N_i}$ and $\rho > 0$ set $\mathcal{B}_\rho(\eta) = \{\sigma \in \mathcal{V}_{N_i} \mid \text{dist}(\sigma, \eta) \leq \rho\}$.

Lemma 4.5. On $\bar{\Omega}$, for all large enough N the following holds: denoting by K any of the sets T_1 , $\pi_2 T^{x_1}$, $x_1 \in \mathcal{M}_1$, or $\cup_{x_1 \in \mathcal{M}_1} \pi_2 T^{x_1}$ we have, for all $\eta \in K$ and $\bar{\eta} \in K$, $\eta \neq \bar{\eta}$,

$$|\text{dist}(\eta, \bar{\eta}) - (N_i/2)| \leq (N_i/2)\delta_i(N_i) \tag{4.22}$$

and for all $0 \leq \epsilon < 1/2$

$$\mathcal{B}_{\epsilon N_i}(\eta) \cap \mathcal{B}_{\epsilon N_i}(\bar{\eta}) = \emptyset. \tag{4.23}$$

Proof. This is the analogue of Lemma 2.12 of [BBG1] and is proved in the same way. \square

Lemma 4.6. With the notation of Lemma 4.5, the following holds on $\bar{\Omega}$ for all large enough N : for all $\eta \in K$ and $\bar{\eta} \in K$, $\eta \neq \bar{\eta}$,

$$F_{N_i, d_i}(\text{dist}(\eta, \bar{\eta})) \leq 2^{-N_i/4}, \tag{4.24}$$

$$U_{N_i, d_i}(\eta, K) \leq \mathcal{U}_{N_i, d_i}(K) \leq |K| 2^{-N_i/4}. \tag{4.25}$$

Proof. Eq. (4.24) follows from (4.22) of Lemma 4.5 and (10.7) of Lemma 10.1 of [6] and implies the leftmost inequality of (4.25) which in turn implies the rightmost one. \square

For any $\sigma \in \mathcal{V}_{N_i}$ and any subset $A \subset \mathcal{V}_{N_i}$, set

$$j(\sigma, A) \equiv \begin{cases} 1 & \text{if } \text{dist}(\sigma, A) = 1, \\ 2 & \text{else.} \end{cases} \tag{4.26}$$

Lemma 4.7. With the notation of Lemma 4.5, the following holds on $\bar{\Omega}$ for all large enough N : for all $\eta \in K$ and all $\sigma \in \mathcal{V}_{N_i} \setminus K$,

$$F_{N_i, d_i}(\text{dist}(\sigma, \eta)) \leq \frac{j}{N_i^2}(1 + o(1)), \quad j = j(\eta, \sigma), \tag{4.27}$$

$$U_{N_i, d_i}(\sigma, K) \leq \frac{j}{N_i^2}(1 + o(1)), \quad j = j(\sigma, K). \tag{4.28}$$

Proof. Eq. (4.27) follows from the definition of the definition of F and case (a) and (b) of Lemma 10.1 of [6]. To prove (4.28) we distinguish two cases: (a) there exists $\eta \in K$ such that $\text{dist}(\sigma, \eta) \leq \epsilon N_i$ for some $0 \leq \epsilon < 1/2$ and (b) for all $\eta \in K$, $\text{dist}(\sigma, \eta) > \epsilon N_i$. In case (b) we have: $U_{N_i, d_i}(\sigma, K) \leq |K| 2^{-N_i/4} \leq o(N_i^{-2})$. This is proven just as (4.25). In case (a) we write

$$U_{N_i, d_i}(\sigma, K) = F_{N_i, d_i}(\text{dist}(\sigma, \eta)) + \sum_{\eta' \in K \setminus \eta} F_{N_i, d_i}(\text{dist}(\sigma, \eta')). \tag{4.29}$$

By (4.23) of Lemma 4.5 we may apply the bound just obtained in case (b) to bound the second term (namely the sum) in the right-hand side of (4.29) whereas the first term is bounded as in (4.27). \square

4.3.2 The Top and a non random point

We will frequently need to lump the simple random walk $J_{N_i}^\circ$ on \mathcal{V}_{N_i} with sets of the form $K \cup \sigma_i \subset \mathcal{V}_{N_i}$ where $\sigma_i \in \mathcal{V}_{N_i}$ is arbitrary and $K = T_1$ (then $i = 1$) or $K = \pi_2 T^{x_1}$ for some $x_1 \in \mathcal{M}_1$ (then $i = 2$). We are now interested in the partition $\Lambda_i(K \cup \sigma_i)$ of $\{1, \dots, N_i\}$ into $d'_i = 2^{M_i+1}$ classes, $\Lambda_i^k(K \cup \sigma_i)$, $1 \leq k \leq d'_i$, induced by $K \cup \sigma_i$ through (4.17). Lemma 4.6 and Lemma 4.7 can be extended to this setting as follows.

Lemma 4.8. *With the notation of Lemma 4.5, the following holds on $\bar{\Omega}$ for all large enough N : for all $\eta \in K$ and all $\sigma' \in \mathcal{V}_{N_i} \setminus (K \cup \sigma_i)$, $\sigma'' \in \mathcal{V}_{N_i}$,*

$$F_{N_i, d'_i}(\text{dist}(\sigma', \sigma'')) \leq (j/N_i^j)(1 + o(1)), \quad j = j(\sigma', \sigma''), \quad (4.30)$$

$$\mathcal{U}_{N_i, d'_i}(K \cup \sigma_i) \leq (j/N_i^j)(1 + o(1)), \quad j = j(\sigma_i, K). \quad (4.31)$$

Proof. This is a simple adaptation of the proofs of Lemmata 4.6 and 4.7. □

Remark 4.9. Upper bounds of the functions F_{N_i, d_i} , U_{N_i, d_i} and \mathcal{U}_{N_i, d_i} imply upper bounds on the quantities ϕ , V_{N_i, d_i} and \mathcal{V}_{N_i, d_i} defined (with obvious notation) in (4.1), (5.4) and (5.9) of [6] (see Lemma 4.2, Lemma 5.2, Lemma 5.4, and Lemma 5.7 of [6]). Furthermore, they imply upper bounds on the quantities U_{N_i, d_i}° , $\mathcal{U}_{N_i, d_i}^\circ$, V_{N_i, d_i}° and $\mathcal{V}_{N_i, d_i}^\circ$ for the associated lumped chain, defined in (5.3), (5.10), and (5.11) of [6] (see Lemma 5.3 and Lemma 5.5 of [6]). We do not repeat these arguments in the proofs of the statements of Section 5.1.

5 Basic estimates for the jump chains

This Section is concerned with the jump chain only. We state and prove a collection of probability estimates that will later be shown, in Section 6, to form the basic blocks of the proof of Proposition 3.1. This section heavily relies on Ref. [6] and the preparations of Section 4.3.

5.1 Main estimates

We recall that the set $\bar{\Omega}$ is defined in (4.21). From now on we drop the dependence on N in the notation.

Proposition 5.1 (REM-like estimates). *On $\bar{\Omega}$, for all large enough N , the following holds: for all $x_1 \in \mathcal{M}_1$ and all $\sigma \in \mathcal{V}_N \setminus \bar{W}$,*

$$\left| \mathbb{P}_\sigma^* \left(\tau_{W^{x_1}} < \tau_{\bar{W} \setminus W^{x_1}} \right) - \frac{1}{M_1} \right| = \mathcal{O} \left(N_1^{-i} \right) \quad (5.1)$$

where $i = 1$ if $\text{dist}(\sigma_1, \xi_1^{x_1}) = 1$ and $i = 2$ otherwise.

Proposition 5.2 (Level-2 motion). *On $\bar{\Omega}$, for all large enough N , the following holds: For $x_1 \in \mathcal{M}_1$ and $A \subset T^{x_1}$, set*

$$u(\xi_1^{x_1}) = \log(1 - q^*(\xi_1^{x_1})), \quad (5.2)$$

$$s^{x_1}(A) = |u(\xi_1^{x_1})| \frac{2^{N_2}}{|A|} (1 + N_2^{-1}), \quad (5.3)$$

$$\lambda^{x_1}(A) = \frac{s^{x_1}(A)}{1 + s^{x_1}(A)}. \quad (5.4)$$

Then, for all non empty subset $A \subseteq T^{x_1}$ and all $\sigma \in W^{x_1} \setminus A$, we have:

i) (Motion within the cylinder set W^{x_1} .) For all $\eta \in A$,

$$\mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{(A \setminus \eta) \cup \partial W^{x_1}} \right) = (1 - \lambda^{x_1}(A)) |A|^{-1} + b_A(\sigma, \eta) \quad (5.5)$$

where, setting $i = 1$ if $\text{dist}(\sigma_2, \eta_2) = 1$ and $i = 2$ otherwise, $b_A(\sigma, \eta) > 0$ and

$$0 \leq b_A(\sigma, \eta) \leq (1 - \lambda^{x_1}(A)) |A|^{-1} \mathcal{O}(N_2^{-i}) + \lambda^{x_1}(A) F_{N_2, d_2}(\text{dist}(\sigma_2, \eta_2)) + \frac{1}{1 + 2^{N_2} \psi_N \gamma_1^N(\xi_1^{x_1})} \mathcal{O} \left(N_2^{(d_2+1)/2} \log N_2 \right). \quad (5.6)$$

Here F_{N_2, d_2} , $d_2 = 2^{M_2}$, is the function introduced in (4.19) (see also the paragraph below (4.20)); in particular,

$$F_{N_2, d_2}(\text{dist}(\sigma_2, \eta_2)) \leq \frac{i}{N_2^i} (1 + o(1)). \tag{5.7}$$

ii) (Leaving the cylinder set W^{x_1} .) For all non empty subset $A \subseteq T^{x_1}$, and all $\sigma \in W^{x_1} \setminus A$,

$$\mathbb{P}_\sigma^*(\tau_{\partial W^{x_1}} < \tau_A) = \lambda^{x_1}(A) + c_A(\sigma) \tag{5.8}$$

where $c_A(\sigma) = \sum_{\eta \in A} b_A(\sigma, \eta)$.

The following almost sure (but rough) bounds on the ranked variables $\gamma_1^N(\xi_1^{x_1})$ will enable us to further express the quantity $\lambda^{x_1}(A)$. Define the set $\widehat{\Omega} \equiv \cap_{M>1} \widehat{\Omega}_M$ where $\widehat{\Omega}_M \equiv \bigcup_{N'_1 \geq 1} \bigcap_{N_1 \geq N'_1} \widehat{\Omega}_{M, N_1}$ and

$$\widehat{\Omega}_{M, N_1} \equiv \bigcap_{1 \leq x_1 \leq M} \left\{ \omega \in \Omega \mid N_1^{-2/\alpha_1^N} \leq (\gamma_1^N(\xi_1^{x_1}))^{-1} < (\ln N_1)^{2/\alpha_1^N} \right\}. \tag{5.9}$$

Lemma 5.3. $\mathcal{P}(\widehat{\Omega}) = 1$.

Proof of Lemma 5.3. By (2.9) and (2.6),

$$\exp \left\{ -u_{N_1}^{-1}(\Xi_{\xi_1^1}^{(1)})/\alpha_1^N \right\} \leq (\gamma_1^N(\xi_1^{x_1}))^{-1} \leq \exp \left\{ -u_{N_1}^{-1}(\Xi_{\xi_1^M}^{(1)})/\alpha_1^N \right\} \tag{5.10}$$

for each $1 \leq x_1 \leq M$. From the well known asymptotic distribution of $\Xi_{\xi_1^k}^{(1)}$ we get that

$$\mathcal{P}(u_{N_1}^{-1}(\Xi_{\xi_1^M}^{(1)}) \leq -2 \ln \ln N_1) \leq e^{-(\ln N_1)^2 \frac{(\ln N_1)^{2M}}{M!}} (1 + o(1)) < N_1^{-2} (1 + o(1))$$

and $\mathcal{P}(u_{N_1}^{-1}(\Xi_{\xi_1^1}^{(1)}) > 2 \ln N_1) \leq N_1^{-2} (1 + o(1))$. The lemma then follows from Borel-Cantelli Lemma. \square

Lemma 5.4. On $\widehat{\Omega}$, for all but a finite number of indices N_1 we have

$$\lambda^{x_1}(A) = \frac{1}{1 + |A| \psi_N \gamma_1^N(\xi_1^{x_1})} (1 + \mathcal{O}(N_2^{-1})), \tag{5.11}$$

$$1 - \lambda^{x_1}(A) = \frac{|A| \psi_N \gamma_1^N(\xi_1^{x_1})}{1 + |A| \psi_N \gamma_1^N(\xi_1^{x_1})} (1 + \mathcal{O}(N_2^{-1})). \tag{5.12}$$

Proof of Lemma 5.4. By (4.3), Lemma 5.3 and the fact that c_1^N decays exponentially fast to zero (see (2.8))

$$0 < q_N^*(\xi_1^{x_1}) = \frac{N_1}{N_2} \frac{c_1^N}{\gamma_1^N(\xi_1^{x_1})} \left(1 + \mathcal{O} \left(c_1^N (\ln N_1)^{2/\alpha_1^N} \right) \right) \ll 1. \tag{5.13}$$

Thus by (5.2)

$$u(\xi_1^{x_1}) = -q^*(\xi_1^{x_1}) \left(1 + \mathcal{O} \left(c_1^N (\ln N_1)^{2/\alpha_1^N} \right) \right). \tag{5.14}$$

The lemma now follows from (5.4), (2.4) and (2.12). For later use let us observe that $q_N^*(\xi_1^{x_1})$ decays exponentially fast. Indeed by (2.4) and (2.12) we have that $(N_1/N_2)c_1^N = 2^{-N_2} \psi_N^{-1}$, whereas, by assumption on ζ_N , there exists $\delta > 0$ such that $\zeta_N \leq (1 - \delta)N_2\beta_*^2/2$, so that

$$q_N^*(\xi_1^{x_1}) = \frac{1}{1 + 2^{N_2} \psi_N \gamma_1^N(\xi_1^{x_1})} \leq e^{-\delta N_2 \beta_*^2 / 2 + \kappa / \alpha_1^N} (\ln N_1)^{2/\alpha_1^N} (1 + o(1)). \tag{5.15}$$

\square

Remark 5.5. Note that if $1 - \lambda^{x_1}(A) \gg N^{-1}$ the term $b_A(\sigma, \eta)$ in (5.5) is sub-leading and if $\lambda^{x_1}(A) \gg N^{-1}$ the term $c_A(\sigma)$ in (5.8) is sub-leading. This will still be true when $1 - \lambda^{x_1}(A) < \text{const.}N^{-1}$, respectively, $\lambda^{x_1}(A) < \text{const.}N^{-1}$ provided only that σ_2 and η_2 are far enough from each other. Indeed the function $F_{N_2, d_2}(\text{dist}(\sigma_2, \eta_2))$ is decreasing and can be made exponentially small in N_2 by choosing $\text{dist}(\sigma_2, \eta_2)$ proportional to N_2 (see (10.7) of Lemma 10.1, Appendix A3 of [6]) while by (5.15) the last term in (5.6) always is exponentially small.

We now turn to “inter-level motions”.

Proposition 5.6 (Inter-level motion). *On $\bar{\Omega}$, for all large enough N , the following holds: for all $\eta \in T$, all $\sigma \in \mathcal{V}_N \setminus \bar{W}$, and some constant $c > 0$*

$$\mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{\bar{W} \setminus \eta} \right) \leq \frac{c}{N_1}. \tag{5.16}$$

It is not difficult to deduce from Proposition 5.6 that:

Corollary 5.7. *On $\bar{\Omega}$, for all large enough N , we have for all $\sigma \in \mathcal{V}_N \setminus \bar{W}$ and all $x_1 \in \mathcal{M}_1$,*

$$\mathbb{P}_\sigma^* \left(\tau_{T^{x_1}} < \tau_{\bar{W} \setminus T^{x_1}} \right) \leq \frac{c|T^{x_1}|}{N_1}, \tag{5.17}$$

$$\left| \mathbb{P}_\sigma^* \left(\tau_{W^{x_1} \setminus T^{x_1}} < \tau_{\bar{W} \setminus (W^{x_1} \setminus T^{x_1})} \right) - \frac{1}{M_1} \right| \leq \frac{c|T^{x_1}|+1}{N_1}. \tag{5.18}$$

5.2 Proofs of the statements of Section 5.1

Proof of Proposition 5.1 (REM-like estimates). Since $\sigma \notin \bar{W}$ then $\pi_1 \sigma \notin T_1$ so that using Lemma 4.2 with $A = \{\xi_1^{x_1}\}$ and $B = T_1 \setminus \xi_1^{x_1}$

$$\mathbb{P}_\sigma^* \left(\tau_{W^{x_1}} < \tau_{\bar{W} \setminus W^{x_1}} \right) = \mathbb{P}_{\sigma_1}^{0,1} \left(\tau_{\xi_1^{x_1}} < \tau_{T_1 \setminus \xi_1^{x_1}} \right). \tag{5.19}$$

Proposition 5.1 then follows from Theorem 1.4 of [6] using the partition $\Lambda_1(T_1)$ induced by T_1 through (4.17) together with (4.25) of Lemma 4.6 and (4.27) of Lemma 4.7. \square

Proof of Proposition 5.2. Throughout the proof we place ourselves on the set of full measure $\bar{\Omega} \cap \bar{\Omega} \subset \Omega$ (see (5.9) and (4.21)). We first prove Assertion (i). Set $d_2 = 2^{|\pi_2 T^{x_1}|} = 2^{M_2}$ and let $\Lambda_2(\pi_2 T^{x_1})$ be the partition into d_2 classes obtained by taking $i = 2$ and $K = \pi_2 T^{x_1}$ in (4.17). (The needed properties of this partition are established in Subsection 4.3.1). Recall that $m_{2, \pi_2 T^{x_1}}$ is the function defined in (4.15)-(4.16) using the partition $\Lambda_2(\pi_2 T^{x_1})$ and that $\bar{\mathbb{E}}^{2, \pi_2 T^{x_1}}$ denotes the expectation w.r.t. to the law $\bar{\mathbb{P}}^{2, \pi_2 T^{x_1}}$ of the $\pi_2 T^{x_1}$ -lumped chain $\bar{I}_{2, \pi_2 T^{x_1}}$ (4.18). By (4.11) of Lemma 4.3 (with $i = 1, j = 2, A = \{\eta\}, B = A \setminus \eta$ and $C(\sigma_1) = W^{x_1}$) and Lemma 2.5 of [6] we have, setting $y \equiv m_{2, \pi_2 T^{x_1}}(\sigma_2), x \equiv m_{2, \pi_2 T^{x_1}}(\eta_2), \mathcal{A} \equiv m_{2, \pi_2 T^{x_1}}(\pi_2 A)$ and $u(\xi_1^{x_1}) \equiv \log(1 - q^*(\xi_1^{x_1}))$

$$\begin{aligned} \mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{(A \setminus \eta) \cup \partial W^{x_1}} \right) &= \bar{\mathbb{E}}_y^{2, \pi_2 T^{x_1}} \left(e^{u(\xi_1^{x_1}) \tau_x} \mathbb{1}_{\{\tau_x < \tau_{\mathcal{A} \setminus x}\}} \right) \\ &\equiv G_{x, \mathcal{A}}^y(u(\xi_1^{x_1})). \end{aligned} \tag{5.20}$$

We now use Proposition 7.7 of [6] to express the Laplace transform $G_{x, \mathcal{A} \setminus x}^y(u), u < 0$. In view of remark 4.9 and Lemma 4.6, it is easy to check that condition (7.38) of Proposition 7.7 of [6] is satisfied: indeed, by (5.15) of Lemma 5.5 of [6], (5.16) of Lemma 5.7 of [6] and (4.25) of Lemma 4.6, for any $A \subseteq T^{x_1}$,

$$\mathcal{V}_{N_2, d_2}^0(\mathcal{A}) = \mathcal{V}_{N_2, d_2}(\pi_2 A) \leq \mathcal{U}_{N_2, d_2}(\pi_2 A) \leq |\pi_2 A| 2^{-N_2/4} \leq M_2 2^{-N_2/4}. \tag{5.21}$$

Next, the quantities $\underline{u}(d_2)$ and \bar{u} appearing in (7.39) and (7.40) of [6] are, here, given by

$$\bar{u}^{-1} \equiv \frac{2^{N_2}}{|\mathcal{A}|} \left(1 + \frac{1}{N_2} \right), \quad \underline{u}(d_2) \equiv C_{d_2} N_2^{d_2/2+1} (\log N_2)^2 (1 + o(1)) \tag{5.22}$$

where $C_{d_2} = C^2(2d_2/\pi)^{d_2/2}$ for some constant $0 < C < \infty$. Indeed, using that on $\bar{\Omega}$ the partition $\Lambda_2(\pi_2 T^{x_1})$ is very close to an equipartition, we have by Lemma 2.6 of [6] that $\mathbb{E}^\circ \tau_0^0 = (\pi/2d_2)^{d_2/2} N_2^{d_2/2} (1 + o(1))$ and by (6.7) of Theorem 6.3 of [6] that

$$\widehat{\Theta}(d_2) = CN_2^{(d_2+1)/2} \log N_2(1 + o(1)). \tag{5.23}$$

Finally, let us check that the point $u(\xi_1^{x_1})$ lies in the segment $(-\rho \underline{u}(d_2), \bar{u})$ for some $0 < \rho < 1$. Inserting (5.15) in (5.14) and using the bound on $\underline{u}(d_2)$ from (5.22), we see that

$$0 > u(\xi_1^{x_1}) = -\frac{1}{1 + 2^{N_2} \psi_N \gamma_1^N(\xi_1^{x_1})} (1 + o(1)) \gg -\underline{u}(d_2). \tag{5.24}$$

Thus $u(\xi_1^{x_1})$ lies in the segment $(-\rho \underline{u}(d_2), 0)$ for any $0 < \rho < 1$, and so, we can use assertion (i)-(a) of Proposition 7.7 of [6] to express the Laplace transform $G_{x,\mathcal{A}}^y(u(\xi_1^{x_1}))$ in (5.20). Using the notation (5.2)-(5.4) and writing $\bar{\mathbb{P}} \equiv \bar{\mathbb{P}}^{2,\pi_2 T^{x_1}}$, this yields

$$G_{x,\mathcal{A}\setminus x}^y(u(\xi_1^{x_1})) = \bar{\mathbb{P}}_y(\tau_x < \tau_{\mathcal{A}\setminus x}) (1 - \lambda^{x_1}(A)) + \bar{\mathbb{P}}_y(\tau_x < \tau_{(\mathcal{A}\setminus x)\cup 0}) \lambda^{x_1}(A) + \mathcal{R}_0 \tag{5.25}$$

where

$$\mathcal{R}_0 = \bar{\mathbb{P}}_0(\tau_x < \tau_{\mathcal{A}\setminus x}) (1 - \lambda^{x_1}(A)) [\mathcal{R}_1 + \bar{\mathbb{P}}_y(\tau_0 < \tau_{\mathcal{A}}) \mathcal{R}_2] + \mathcal{R}_3 \tag{5.26}$$

and where, by (5.21)-(5.24), the remainder terms are given by

$$\begin{aligned} \mathcal{R}_i &= \frac{1}{1 + 2^{N_2} \psi_N \gamma_1^N(\xi_1^{x_1})} \mathcal{O}\left(N_2^{(d_2+1)/2} \log N_2\right), \quad i = 1, 3, \\ \mathcal{R}_2 &= \mathcal{O}\left(\max\left\{\frac{1}{N_2^2} \lambda^{x_1}(A), \frac{1}{1 + 2^{N_2} \psi_N \gamma_1^N(\xi_1^{x_1})} N_2^{d_2/2+1} (\log N_2)^2\right\}\right). \end{aligned}$$

In particular, in view of (5.15) and the fact that $0 < \lambda^{x_1}(A) \leq 1$, $\mathcal{R}_2 = \mathcal{O}(N_2^{-2})$ and \mathcal{R}_i , $i = 1, 3$, decay exponentially fast. It now remains to estimate the four probabilities that enter the above expressions. We trivially bound $0 \leq \bar{\mathbb{P}}_y(\tau_0 < \tau_{\mathcal{A}}) \leq 1$. To deal with the prefactors of $(1 - \lambda^{x_1}(A))$ in (5.25) and (5.26), namely, the harmonic measures stating from 0 and from y we use, respectively, Lemma 4.4 of [6] and Theorem 4.5 of [6] combined with Lemma 4.2 of [6]. Together with Lemma 4.6 and Lemma 4.7, they yield,

$$\bar{\mathbb{P}}_0(\tau_x < \tau_{\mathcal{A}\setminus x}) = |A|^{-1} (1 + \mathcal{O}(N_2^{-2})), \tag{5.27}$$

$$0 \leq \bar{\mathbb{P}}_y(\tau_x < \tau_{\mathcal{A}\setminus x}) - |A|^{-1} (1 + \mathcal{O}(N_2^{-2})) \leq F_{N_2,d_2}(\text{dist}(\sigma_2, \eta_2)), \tag{5.28}$$

where F_{N_2,d_2} is the function introduced in Definition 3.3 of [6] and studied in Appendix 11 of [6]. It is a decreasing function that decays polynomially fast in N_2 for small distances and exponentially fast in N_2 for distances proportional to N_2 . In particular, setting $i = 1$ if $\text{dist}(\eta_2, \sigma_2) = 1$ and $i = 2$ otherwise, we have the rough bound

$$F_{N_2,d_2}(\text{dist}(\sigma_2, \eta_2)) \leq \frac{i}{N_2^i} (1 + o(1)). \tag{5.29}$$

It remains to deal with the pre-factors of $\lambda^{x_1}(A)$ in (5.25). For an upper bound, write

$$0 \leq \bar{\mathbb{P}}_y(\tau_x < \tau_{(\mathcal{A}\setminus x)\cup 0}) \leq \bar{\mathbb{P}}_y(\tau_x < \tau_0) \leq F_{N_2,d_2}(\text{dist}(\sigma_2, \eta_2)) \tag{5.30}$$

where the last inequality is Theorem 3.2 of [6]. Inserting our estimates in (5.25) and combining the result with (5.20), we arrive at

$$\mathbb{P}_\sigma^*(\tau_\eta < \tau_{(\mathcal{A}\setminus \eta)\cup \partial W^{x_1}}) = (1 - \lambda^{x_1}(A)) |A|^{-1} + b_A(\sigma_2, \eta_2) \tag{5.31}$$

where

$$0 \leq b_A(\sigma_2, \eta_2) \leq (1 - \lambda^{x_1}(A))|A|^{-1} \left(F_{N_2, d_2}(\text{dist}(\sigma_2, \eta_2)) + \mathcal{O}(N_2^{-2}) \right) + \lambda^{x_1}(A)F_{N_2, d_2}(\text{dist}(\sigma_2, \eta_2)) + \mathcal{R}_3 \tag{5.32}$$

which readily yields (5.5). The second assertion of Proposition 5.2 is a direct consequence of the first, observing that

$$\mathbb{P}_\sigma^* (\tau_{\partial W^{x_1}} < \tau_A) = 1 - \sum_{\eta \in A} \mathbb{P}_\sigma^* (\tau_\eta < \tau_{(A \setminus \eta) \cup \partial W^{x_1}}) = 1 - \sum_{x \in A} G_{x, \mathcal{A} \setminus x}^y(u(\xi_1^{x_1}))$$

where the last equality is (5.20), and use (5.31)-(5.32). The proof of Proposition 5.2 is now complete. \square

Proof of Proposition 5.6. Using (4.9), $\eta = \eta_1 \eta_2 \in T$ can be written as $\eta = C(\eta_2) \cap C(\eta_1)$ where $C(\eta_1) = W^{x_1}$ for some $x_1 \in \mathcal{M}_1$. Thus, for the event $\{\tau_\eta < \tau_{\overline{W} \setminus \eta}\}$ to take place, the chain must reach η “from within” the set $C(\eta_2)$, without of course having ever visited \overline{W} . Building on this observation we begin by establishing an a priori upper bound on the probability (5.16) that is valid for all starting points σ in $\mathcal{V}_N \setminus \overline{W}$.

Lemma 5.8. *The following holds on $\overline{\Omega}$ (see (4.21)) for all large enough N : for all $\eta \in T$ and all $\sigma \in \mathcal{V}_N \setminus \overline{W}$*

$$\mathbb{P}_\sigma^* (\tau_\eta < \tau_{\overline{W} \setminus \eta}) \leq \frac{q_N^*(\sigma_1)}{M_1} (1 + \mathcal{O}(M_1/N_1)). \tag{5.33}$$

Proof. Using the renewal identity (see e.g. Corollary 1.9 in [11])

$$\mathbb{P}_\sigma^* (\tau_\eta < \tau_{\overline{W} \setminus \eta}) \leq \frac{\mathbb{P}_\sigma^* (\tau_\eta < \tau_{(\overline{W} \setminus \eta) \cup \sigma})}{\mathbb{P}_\sigma^* (\tau_{\overline{W}} < \tau_\sigma)}. \tag{5.34}$$

To deal with the numerator we use reversibility and (4.5) to write

$$\mathbb{P}_\sigma^* (\tau_\eta < \tau_{(\overline{W} \setminus \eta) \cup \sigma}) = \frac{G_{\beta, N}^*(\eta)}{G_{\beta, N}^*(\sigma)} \mathbb{P}_\eta^* (\tau_\sigma < \tau_{\overline{W}}) = \frac{q_N^*(\sigma_1)}{q_N^*(\eta_1)} \mathbb{P}_\eta^* (\tau_\sigma < \tau_{\overline{W}}). \tag{5.35}$$

Now for the event $\{\tau_\sigma < \tau_{\overline{W}}\}$ to take place the chain, starting in η , must exit η through the set $C(\eta_2)$. Thus, setting $\partial_1 \eta \equiv C(\eta_2) \cap \partial \eta = \{\sigma \in \mathcal{V}_N \mid \sigma_1 \sim \eta_1, \sigma_2 = \eta_2\}$ we have

$$\begin{aligned} \mathbb{P}_\eta^* (\tau_\sigma < \tau_{\overline{W}}) &= \mathbb{P}_\eta^* (\tau_{\partial_1 \eta} < \tau_\sigma < \tau_{\overline{W}}) \\ &= \sum_{\sigma' \in \partial_1 \eta} \mathbb{P}_\eta^* (\tau_{\sigma'} < \tau_{(\partial_1 \eta \setminus \sigma') \cup \overline{W} \cup \sigma}) \mathbb{P}_{\sigma'}^* (\tau_\sigma < \tau_{\overline{W}}). \end{aligned} \tag{5.36}$$

To bound the last probability in the right-hand side of (5.36) we simply observe that on $\overline{\Omega}$, for all large enough N ,

$$\mathbb{P}_{\sigma'}^* (\tau_\sigma < \tau_{\overline{W}}) \leq \mathbb{P}_{\sigma'}^* (\tau_{C(\sigma_1)} < \tau_{\overline{W}}) = \frac{1}{M_1 + 1} (1 + \mathcal{O}(M_1/N_1)) \tag{5.37}$$

where the last equality is proved just as Proposition 5.1, namely, using first Lemma 4.2 with $A = \{\sigma_1\}$, $B = T_1$ and $\sigma'_1 \neq \sigma_1$ to write that $\mathbb{P}_{\sigma'}^* (\tau_{C(\sigma_1)} < \tau_{\overline{W}}) = \mathbb{P}_{\sigma'_1}^{\circ, 1} (\tau_{\sigma_1} < \tau_{T_1})$, and using next Theorem 1.4 of [6] with the partition $\Lambda_1(T_1 \cup \sigma_1)$ induced by $T_1 \cup \sigma_1$ (that is to say, the Top and a non random point) together with (4.30) and (4.31) of Lemma 4.8 and (4.27) of Lemma 4.7, under the assumptions therein (namely, on $\overline{\Omega}$, for all large enough N) which are assumed from now on to be verified. Plugging (5.37) in (5.36), we get

$$\mathbb{P}_\eta^* (\tau_\sigma < \tau_{\overline{W}}) \leq \frac{1}{M_1 + 1} (1 + \mathcal{O}(M_1/N_1)) \mathbb{P}_\eta^* (\tau_{\partial_1 \eta} < \tau_{\overline{W} \cup \sigma}) \tag{5.38}$$

where $\mathbb{P}_\eta^*(\tau_{\partial_1\eta} < \tau_{\overline{W}\cup\sigma}) \leq \mathbb{P}_\eta^*(\tau_{\partial_1\eta} < \tau_{W^{x_1}}) = \sum_{\sigma'' \in \partial_1\eta} p_2^*(\eta, \sigma'') = q_N^*(\eta_1)$. Combined with (5.35), this finally gives

$$\mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{(\overline{W}\setminus\eta)\cup\sigma} \right) \leq \frac{q_N^*(\sigma_1)}{M_1 + 1} (1 + \mathcal{O}(M_1/N_1)). \tag{5.39}$$

It now remains to bound the denominator of (5.34). For this we simply decompose on the first step of the jump chain,

$$\mathbb{P}_\sigma^* (\tau_{\overline{W}} < \tau_\sigma) = \sum_{\sigma' \sim \sigma: \sigma' \in \overline{W}} p_N^*(\sigma, \sigma') + \sum_{\sigma' \sim \sigma: \sigma' \notin \overline{W}} p_N^*(\sigma, \sigma') \mathbb{P}_{\sigma'}^* (\tau_{\overline{W}} < \tau_\sigma),$$

and observe that $\mathbb{P}_{\sigma'}^* (\tau_{\overline{W}} < \tau_\sigma) \geq \mathbb{P}_{\sigma'}^* (\tau_{\overline{W}} < \tau_{C(\sigma_1)}) = 1 - \frac{1}{M_1+1} (1 + \mathcal{O}(\frac{M_1}{N_1}))$ where the last equality follows from (5.37). Since $\sum_{\sigma' \sim \sigma} p_N^*(\sigma, \sigma') = 1$ we get

$$\mathbb{P}_\sigma^* (\tau_{\overline{W}} < \tau_\sigma) \geq 1 - \frac{1}{M_1 + 1} (1 + \mathcal{O}(M_1/N_1)). \tag{5.40}$$

Inserting (5.39) and (5.40) in (5.34) yields (5.33) and proves the lemma. □

Now choose $r(N)$ such that

$$2^{N_1/2} \mathbb{P} \left(\exp \left\{ \beta \sqrt{aN_1} \Xi_{\sigma_1}^{(1)} \right\} \geq r(N) \right) = 1, \tag{5.41}$$

and consider the set $\mathcal{V}_N^+ \equiv \mathcal{V}_{N_1}^+ \times \mathcal{V}_{N_2}$ where

$$\mathcal{V}_{N_1}^+ = \left\{ \sigma_1 \in \mathcal{V}_{N_1} \mid \Xi_{\sigma_1}^{(1)} \geq \frac{\log r(N)}{\beta \sqrt{aN_1}} \right\}. \tag{5.42}$$

Corollary 5.9. *Under the assumptions of Lemma 5.8, for all $\eta \in T$ and all $\sigma \in \mathcal{V}_N^+ \setminus \overline{W}$,*

$$\mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{\overline{W}\setminus\eta} \right) \leq \frac{1}{M_1} e^{-\sqrt{\frac{1}{2}}\beta\beta_*N_1(1+o(1))}. \tag{5.43}$$

Proof. By (4.3), (1.1) and (5.42), $q_N^*(\sigma_1) \leq [1 + (N_2/N_1)r(N)]^{-1} \leq e^{-\sqrt{\frac{1}{2}}\beta\beta_*N_1(1+o(1))}$ for all $\sigma \in \mathcal{V}_N^+$. Inserting this in Lemma 5.8 yields (5.43). □

This implies that the bound (5.16) holds true for all $\sigma \in \mathcal{V}_N^+ \setminus \overline{W}$. To extend this result to the entire set $\mathcal{V}_N \setminus \overline{W}$, observe that $\overline{W} \subset \mathcal{V}_N^+$ (see the definition (3.4)-(3.6) of \overline{W}) and decompose the probability in (5.16) according to whether the jump chain visits $\mathcal{V}_N^+ \setminus \overline{W}$ before η or not, namely, for $\sigma \in \mathcal{V}_N \setminus \mathcal{V}_N^+$ write

$$\mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{\overline{W}\setminus\eta} \right) = \mathbb{P}_\sigma^* \left(\tau_{\mathcal{V}_N^+ \setminus \overline{W}} < \tau_\eta < \tau_{\overline{W}\setminus\eta} \right) + \mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{(\overline{W}\setminus\eta)\cup(\mathcal{V}_N^+ \setminus \overline{W})} \right) \tag{5.44}$$

Call Q_1 and Q_2 , respectively, the first and second probabilities in the r.h.s. of (5.44). Then

$$Q_1 = \sum_{\sigma' \in \mathcal{V}_N^+ \setminus \overline{W}} \mathbb{P}_\sigma^* \left(\tau_{\sigma'} < \tau_{\mathcal{V}_N^+ \setminus \sigma'} \right) \mathbb{P}_{\sigma'}^* \left(\tau_\eta < \tau_{\overline{W}\setminus\eta} \right) \leq \frac{1}{M_1} e^{-\sqrt{\frac{1}{2}}\beta\beta_*N_1(1+o(1))} \tag{5.45}$$

where the inequality in (5.45) follows from Corollary 5.9. To bound Q_2 note that given any set $A_{N_1} \subset \mathcal{V}_{N_1}^+$, $A_N \equiv \cup_{\sigma_1 \in A_{N_1}} C(\sigma_1) \subset \cup_{\sigma_1 \in \mathcal{V}_{N_1}^+} C(\sigma_1)$, so that for all $\sigma \in \mathcal{V}_N \setminus \mathcal{V}_N^+$

$$Q_2 = \mathbb{P}_\sigma^* \left(\tau_\eta < \tau_{\mathcal{V}_N^+ \setminus \eta} \right) \leq \mathbb{P}_\sigma^* \left(\tau_{C(\eta_1)} < \tau_{A_N \setminus C(\eta_1)} \right) = \mathbb{P}_{\sigma_1}^{\circ,1} \left(\tau_{\eta_1} < \tau_{A_{N_1} \setminus \eta_1} \right) \tag{5.46}$$

where the last equality is Lemma 4.2 applied with $A = \{\eta_1\}$, $B = A_{N_1} \subset \mathcal{V}_{N_1}^+$ and $\sigma_1 \notin \mathcal{V}_{N_1}^+$.

Now let $\ell(N) = 2^{N_1} / |\mathcal{V}_{N_1}^+|$ be the inverse of the density of the random cloud $\mathcal{V}_{N_1}^+$. By (5.41), (5.42), a simple large deviation estimate and Borel-Cantelli lemma, there exists a set Ω_1^+ with $\mathcal{P}(\Omega_1^+) = 1$ and such that on Ω_1^+ , $\ell(N) = 2^{N_1/2}(1 + o(1))$ for all large enough N . Decomposing the event $\{\tau_{\eta_1} < \tau_{A_{N_1} \setminus \eta_1}\}$ according to whether $\tau_{A_{N_1} \setminus \eta_1} < \ell(N)/N_1$ or $\tau_{A_{N_1} \setminus \eta_1} \geq \ell(N)/N_1$ yields

$$\mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{\eta_1} < \tau_{A_{N_1} \setminus \eta_1}) \leq \mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{\eta_1} < \ell(N)/N_1) + \mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{A_{N_1} \setminus \eta_1} \geq \ell(N)/N_1). \quad (5.47)$$

By Theorem 1.1 of [16], carefully collecting all estimates in its proof, we get that

$$\max_{\sigma_1} \mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{A_{N_1} \setminus \eta_1} \geq \ell(N)/N_1) \leq \frac{c}{N_1} \quad (5.48)$$

for some constant $c > 0$. It remains to bound the first probability in the right-hand side of (5.47). Notice that the mean hitting time of a given vertex in $\mathcal{V}_{N_1}^+$ by the simple random walk is to leading order $2^{N_1} = \ell^2(N) \gg \ell(N)/N_1$. Assume first that $\text{dist}(\sigma_1, \eta_1) > 1$. Then, by Tchebychev exponential inequality, for all $u > 0$,

$$\mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{\eta_1} < \ell(N)/N_1) \leq \exp\left(\frac{u}{N_1 \ell(N)}\right) \mathbb{E}_{\sigma_1}^{\circ,1}\left(e^{-\frac{u}{\ell^2(N)} \tau_{A_{N_1} \setminus \eta_1}}\right) \quad (5.49)$$

$$\leq \exp\left(\frac{u}{N_1 \ell(N)}\right) \frac{1}{1 + u(1 + \frac{1}{N})} + \frac{c}{N_1^2} \quad (5.50)$$

where the last inequality follows from (7.7) of Theorem 7.2 of [6]. Choosing e.g. $u = N_1^2$ yields $\mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{\eta_1} < \ell(N)/N_1) \leq c/N_1^2$ for some $c > 0$. Proceeding in a similar way when $\text{dist}(\sigma_1, \eta_1) = 1$ but using (7.6) of Theorem 7.2 of [6] we get $\mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{\eta_1} < \ell(N)/N_1) \leq c/N_1$ for some $c > 0$. Hence, collecting our bounds,

$$\max_{\sigma_1} \mathbb{P}_{\sigma_1}^{\circ,1}(\tau_{A_{N_1} \setminus \eta_1} \geq \ell(N)/N_1) \leq \frac{c}{N_1}, \quad (5.51)$$

for some constant $c > 0$. Inserting (5.48) and (5.51) in (5.47) and plugging the result in (5.46), we finally get that for all $\sigma \in \mathcal{V}_N \setminus \mathcal{V}_N^+$ and some constant $c > 0$,

$$0 \leq Q_2 \leq c/N_1. \quad (5.52)$$

Plugging (5.45) and (5.52) in (5.44) we obtain that, on $\Omega_1^+ \cap \bar{\Omega}_1(T_1)$ and for all N_1 large enough, for all $\sigma \in \mathcal{V}_N \setminus \mathcal{V}_N^+$

$$\mathbb{P}_{\sigma}^*(\tau_{\eta} < \tau_{\bar{W} \setminus \eta}) \leq c/N_1. \quad (5.53)$$

Combining (5.53) and (5.43) yields the claim of Proposition 5.6. \square

Proof of Corollary 5.7. The bound (5.17) follows from Proposition 5.6 and the identity

$$\mathbb{P}_{\sigma}^*(\tau_{T^{x_1}} < \tau_{\bar{W} \setminus T^{x_1}}) = \sum_{\eta \in T^{x_1}} \mathbb{P}_{\sigma}^*(\tau_{\eta} < \tau_{\bar{W} \setminus \eta}). \quad (5.54)$$

To prove (5.18) note that

$$\mathbb{P}_{\sigma}^*(\tau_{W^{x_1}} < \tau_{\bar{W} \setminus W^{x_1}}) = \mathbb{P}_{\sigma}^*(\tau_{W^{x_1} \setminus T^{x_1}} < \tau_{\bar{W} \setminus (W^{x_1} \setminus T^{x_1})}) + \mathbb{P}_{\sigma}^*(\tau_{T^{x_1}} < \tau_{\bar{W} \setminus T^{x_1}}),$$

and use (5.1) of Proposition 5.1 to bound the probability in the left-hand side above and (5.17) to bound the second probability in the right-hand side. \square

6 Entrance law. Proofs

6.1 Proof of Proposition 3.1

The set $\tilde{\Omega}$ of Propositions 3.1 is chosen to be $\tilde{\Omega} = \bar{\Omega} \cap \Omega_1^+ \cap \hat{\Omega}$ where $\bar{\Omega}$, Ω_1^+ , and $\hat{\Omega}$ are defined, respectively, in (4.21), above (5.47) and in (5.9). Since each of these set has full measure (see Lemma 4.4, Lemma 5.3 and the paragraph above (5.47)) $\mathcal{P}(\tilde{\Omega}) = 1$. From now on we will assume that $\omega \in \tilde{\Omega}$. Lemma 4.1 will be frequently used without making explicit mention of it.

Proof of Assertion (i) of Proposition 3.1. We first work out general expressions, valid at all temperature, that relate the entrance probabilities $\mathbb{P}_\sigma(\tau_\eta < \tau_{T \setminus \eta})$, $\sigma \notin T$, $\eta \in T$, to the basic REM-like, level-2 and inter-level probabilities estimated in Propositions 5.1, 5.2 and 5.6. To shorten the notations we write, given $x_1 \in \mathcal{M}_1$ and $\eta \in T^{x_1}$

$$\mathbb{P}_\sigma(\tau_\eta < \tau_{T \setminus \eta}) = \begin{cases} P_\eta(\sigma) & \text{if } \sigma \in W^{x_1} \setminus T^{x_1}, \\ Q_\eta(\sigma) & \text{if } \sigma \in W^{x'_1} \setminus T^{x'_1} \text{ for some } x'_1 \in \mathcal{M}_1 \setminus x_1, \\ R_\eta(\sigma) & \text{if } \sigma \in \mathcal{S}_N \setminus \bar{W}. \end{cases} \quad (6.1)$$

Note that the probabilities (3.10), (3.11) and (3.12) are of the form, respectively, $P_\eta(\sigma)$, $Q_\eta(\sigma)$, and $R_\eta(\sigma)$. As the next lemma shows, both $P_\eta(\sigma)$ and $Q_\eta(\sigma)$ can be expressed as functions of $R_\eta(\sigma)$ whereas, using a renewal kind of argument, $R_\eta(\sigma)$ itself is the solution of a linear system of equations.

Lemma 6.1. Given $x_1 \in \mathcal{M}_1$ and $\eta \in T^{x_1}$ define, for $\sigma \in \mathcal{V}_N \setminus \bar{W}$,

$$b_\eta(\sigma) = \mathbb{P}_\sigma(\tau_\eta < \tau_{\bar{W} \setminus \eta}) + \sum_{\sigma' \in W^{x_1} \setminus T^{x_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'}) \mathbb{P}_{\sigma'}(\tau_\eta < \tau_{(T^{x_1} \setminus \eta) \cup \partial W^{x_1}}) \quad (6.2)$$

and, for $\sigma \in \mathcal{V}_N \setminus \bar{W}$ and $\sigma'' \in \partial \bar{W}$,

$$a(\sigma, \sigma'') = \sum_{x'_1 \in \mathcal{M}_1} \sum_{\sigma' \in W^{x'_1} \setminus T^{x'_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'}) \mathbb{P}_{\sigma'}(\tau_{\sigma''} < \tau_{(\partial W^{x'_1} \setminus \sigma'') \cup T^{x'_1}}). \quad (6.3)$$

If A denotes the square matrix $A = (a(\sigma, \sigma''))_{\sigma, \sigma'' \in \partial \bar{W}}$ and b_η the vector $b_\eta = (b_\eta(\sigma))_{\sigma \in \partial \bar{W}}$, then the vector $R_\eta \equiv (R_\eta(\sigma))_{\sigma \in \partial \bar{W}}$ obeys

$$R_\eta = b_\eta + AR_\eta. \quad (6.4)$$

Moreover, if $R_\eta^* = (R_\eta^*(\sigma))_{\sigma \in \partial \bar{W}}$ solves the linear system (6.4), then

(i) for all $\sigma \in \mathcal{S}_N \setminus \bar{W}$,

$$R_\eta(\sigma) = b_\eta(\sigma) + \sum_{\sigma'' \in \partial \bar{W}} a(\sigma, \sigma'') R_\eta^*(\sigma''), \quad (6.5)$$

(ii) for all $\sigma \in W^{x_1} \setminus T^{x_1}$,

$$P_\eta(\sigma) = \mathbb{P}_\sigma(\tau_\eta < \tau_{(T^{x_1} \setminus \eta) \cup \partial W^{x_1}}) + \sum_{\sigma' \in \partial W^{x_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{(\partial W^{x_1} \setminus \sigma') \cup T^{x_1}}) R_\eta^*(\sigma'), \quad (6.6)$$

(iii) for all $\sigma \in W^{x'_1} \setminus T^{x'_1}$ and all $x'_1 \in \mathcal{M}_1 \setminus x_1$,

$$Q_\eta(\sigma) = \sum_{\sigma' \in \partial W^{x'_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{(\partial W^{x'_1} \setminus \sigma') \cup T^{x'_1}}) R_\eta^*(\sigma'). \quad (6.7)$$

Proof. Let us first consider $P_\eta(\sigma)$. Decomposing the event $\{\tau_\eta < \tau_{T \setminus \eta}\}$ according to whether, starting in σ , the chain visits η before visiting the boundary ∂W^{x_1} or not, we get:

$$\begin{aligned} P_\eta(\sigma) &= \mathbb{P}_\sigma(\{\tau_\eta < \tau_{T \setminus \eta}\} \cap \{\tau_\eta < \tau_{\partial W^{x_1}}\}) + \mathbb{P}_\sigma(\{\tau_\eta < \tau_{T \setminus \eta}\} \cap \{\tau_{\partial W^{x_1}} < \tau_\eta\}) \\ &= \mathbb{P}_\sigma(\tau_\eta < \tau_{(T^{x_1} \setminus \eta) \cup \partial W^{x_1}}) + \mathbb{P}_\sigma(\tau_{\partial W^{x_1}} < \tau_\eta < \tau_{T \setminus \eta}) \\ &= \mathbb{P}_\sigma(\tau_\eta < \tau_{(T^{x_1} \setminus \eta) \cup \partial W^{x_1}}) + \sum_{\sigma' \in \partial W^{x_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{(\partial W^{x_1} \setminus \sigma') \cup T^{x_1}}) R_\eta(\sigma') \end{aligned} \quad (6.8)$$

where we used (6.1) in the last line. Proceeding in the same way with $Q_\eta(\sigma)$ yields

$$Q_\eta(\sigma) = \sum_{\sigma' \in \partial W^{x_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{(\partial W^{x_1} \setminus \sigma') \cup T^{x_1}}) R_\eta(\sigma'). \quad (6.9)$$

We now focus on $R_\eta(\sigma)$. Clearly $\mathbb{P}_\sigma(\tau_{\bar{W}} < \infty) = 1$ and by definition of \bar{W}

$$\{\tau_{\bar{W}} < \infty\} = \bigcup_{x'_1 \in \mathcal{M}_1} \{\tau_{W^{x'_1}} < \tau_{\bar{W} \setminus W^{x'_1}}\}. \quad (6.10)$$

Hence

$$R_\eta(\sigma) = \sum_{x'_1 \in \mathcal{M}_1} \mathbb{P}_\sigma(\{\tau_\eta < \tau_{T \setminus \eta}\} \cap \{\tau_{W^{x'_1}} < \tau_{\bar{W} \setminus W^{x'_1}}\}). \quad (6.11)$$

Setting $\mathcal{E}^{x'_1} \equiv \{\tau_\eta < \tau_{T \setminus \eta}\} \cap \{\tau_{W^{x'_1}} < \tau_{\bar{W} \setminus W^{x'_1}}\}$ and observing that for $x'_1 = x_1$

$$\mathcal{E}^{x_1} = \left(\bigcup_{\sigma' \in W^{x_1} \setminus T^{x_1}} \{\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'}\} \cap \{\tau_\eta < \tau_{T \setminus \eta}\} \right) \cup \{\tau_\eta < \tau_{\bar{W} \setminus \eta}\} \quad (6.12)$$

whereas for all $x'_1 \in \mathcal{M}_1 \setminus x_1$

$$\mathcal{E}^{x'_1} = \bigcup_{\sigma' \in W^{x'_1} \setminus T^{x'_1}} \{\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'}\} \cap \{\tau_\eta < \tau_{T \setminus \eta}\}, \quad (6.13)$$

(6.11) becomes

$$R_\eta(\sigma) = \mathbb{P}_\sigma(\tau_\eta < \tau_{\bar{W} \setminus \eta}) + \sum_{\sigma' \in W^{x_1} \setminus T^{x_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'}) P_\eta(\sigma') \quad (6.14)$$

$$+ \sum_{x'_1 \in \mathcal{M}_1 \setminus x_1} \sum_{\sigma' \in W^{x'_1} \setminus T^{x'_1}} \mathbb{P}_\sigma(\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'}) Q_\eta(\sigma'). \quad (6.15)$$

Plugging the expressions (6.8) and (6.9) of $P_\eta(\sigma)$ and $Q_\eta(\sigma)$ in (6.15) readily yield that for $b_\eta(\sigma)$ and $a(\sigma, \sigma'')$ as in (6.2) and (6.3), $R_\eta(\sigma)$ obeys, for all $\sigma \in \mathcal{S}_N \setminus \bar{W}$,

$$R_\eta(\sigma) = b_\eta(\sigma) + \sum_{\sigma'' \in \partial W^{x_1}} a(\sigma, \sigma'') R_\eta(\sigma''). \quad (6.16)$$

The restriction of this last relation to $\sigma \in \partial \bar{W}$ enables us to see the vector $(R_\eta(\sigma))_{\sigma \in \partial \bar{W}}$ as solution of the linear system of equations (6.4). This observation together with (6.16), (6.8) and (6.9) prove, respectively, (6.5), (6.6) and (6.7). Lemma 6.1 is proven. \square

Lemma 6.2. *Under the assumptions and with the notation of Proposition 3.1, the linear system (6.4) has a unique solution, $R_\eta^* = (R_\eta^*(\sigma))_{\sigma \in \partial \bar{W}}$, that obeys*

$$R_\eta^*(\sigma) = \frac{\nu_1(x_1)}{M_2} (1 + \varepsilon_N) \quad \forall \sigma \in \partial \bar{W}. \quad (6.17)$$

The proof of Lemma 6.2 makes use of the following two lemmata.

Lemma 6.3. *The matrix A has the following properties: for each $\sigma \in \partial\bar{W}$*

$$0 \leq \sum_{\sigma'' \in \partial\bar{W}} a(\sigma, \sigma'') = 1 - \sum_{\eta \in T} b_\eta(\sigma) \leq 1 \tag{6.18}$$

Proof. Summing both sides of (6.4) over $\eta \in T$ and using that by (6.1) and (6.2),

$$\sum_{\eta \in T} R_\eta(\sigma) = 1 \text{ for all } \sigma \in \mathcal{S}_N \setminus \bar{W} \tag{6.19}$$

yields the equality of (6.18). The first and final upper and lower bounds simply reflect the fact that A is a positive matrix and b_η a positive vector. \square

Lemma 6.4. *A necessary and sufficient condition for a solution to (6.4) to exist is that*

$$\min_{\sigma \in \partial\bar{W}} \sum_{\eta \in T} b_\eta(\sigma) > 0. \tag{6.20}$$

In this case the solution is unique, positive and given by $R_\eta = (I - A)^{-1}b_\eta$, where I denotes the identity matrix and $(I - A)^{-1} = \sum_{k=1}^{\infty} A^k$ exists.

Proof. Denote by $\rho(A)$ the Perron-Frobenius eigenvalue of A . By (6.18) and the standard min and max row-sum bounds on the Perron-Frobenius eigenvalue of primitive matrices (see [30], p. 8 Corollary 1)

$$0 \leq 1 - \max_{\sigma \in \partial\bar{W}} \sum_{\eta \in T} b_\eta(\sigma) \leq \rho(A) \leq 1 - \min_{\sigma \in \partial\bar{W}} \sum_{\eta \in T} b_\eta(\sigma) \leq 1. \tag{6.21}$$

The claim of the lemma now follows from (6.18), (6.21) and Theorem 2.1 p. 30 of [30] (see also Corollary 3 p. 31). \square

We are now ready to prove Lemma 6.2.

Proof of Lemma 6.2. Let us establish that if $1 - \lambda^{x_1}(T^{x_1}) \gg \mathcal{O}(N^{-1})$ then, there exists a constant $0 < c < \infty$ (that depends on a_1, a_2) such that, for all $x_1 \in \mathcal{M}_1$, all $\eta \in T^{x_1}$ and all $\sigma \in \mathcal{V}_N \setminus \bar{W}$

$$v(x_1) \leq b_\eta(\sigma) \leq v(x_1) + cN^{-1} \text{ where } v(x_1) \equiv \frac{1}{M_1 M_2} (1 - \lambda^{x_1}(T^{x_1})). \tag{6.22}$$

Inserting (5.5) of Proposition 5.2 in (6.2) and using (5.18) of Corollary 5.7 yields that for all $\sigma \in \mathcal{V}_N \setminus \bar{W}$ and all $\eta \in T^{x_1}$, $b_\eta(\sigma) = v(x_1)(1 + N_1^{-1}(1 + o(1))) + \delta b_\eta(\sigma)$ where $v(x_1)$ is as in (6.22) and

$$\delta b_\eta(\sigma) \equiv \mathbb{P}_\sigma \left(\tau_\eta < \tau_{\bar{W} \setminus \eta} \right) + \sum_{\sigma' \in W^{x_1} \setminus T^{x_1}} \mathbb{P}_\sigma \left(\tau_{\sigma'} < \tau_{\bar{W} \setminus \sigma'} \right) b_A(\sigma, \eta). \tag{6.23}$$

By (5.6), (5.7) and (5.15), $0 \leq b_{T^{x_1}}(\sigma, \eta) \leq N_2^{-1}(1 + o(1))$ for all for all $\sigma \in \mathcal{V}_N \setminus \bar{W}$ and all $\eta \in T^{x_1}$. Inserting this rough bound in (6.23) and using again (5.18) to bound the resulting sum, the first term in (6.22) being bounded in (5.16), we get

$$0 \leq \max_{\sigma \in \partial\bar{W}, \eta \in T^{x_1}} \delta b_\eta(\sigma) \leq \frac{1}{N_1}(1 + o(1)) + \frac{1}{M_1 N_2}(1 + (2M_1/N_1)(1 + o(1))). \tag{6.24}$$

This proves the claim (6.22). Eq. (6.22) in particular implies that the solution R_η^* of the linear system (6.4) obeys

$$v(x_1)(I - A)^{-1}\mathbf{1} \leq R_\eta^* \leq (v(x_1) + cN^{-1})(I - A)^{-1}\mathbf{1} \tag{6.25}$$

where $\mathbf{1}$ is the vector with all components equal to one and where the inequalities hold component wise, for each $R_\eta^*(\sigma)$, $\sigma \in \partial\bar{W}$. Now, by (6.18) of 6.3 and Lemma 6.4,

$$(I - A)^{-1}\mathbf{1} = \sum_{k=1}^{\infty} A^k \mathbf{1} = \left(\sum_{\eta \in T} b_\eta(\sigma) \right)^{-1} \mathbf{1} \tag{6.26}$$

where by (6.22), $\sum_{\eta \in T} v(x_1) \leq \sum_{\eta \in T} b_\eta(\sigma) \leq \sum_{\eta \in T} (v(x_1) + cN^{-1})$ and

$$\sum_{\eta \in T} v(x_1) = \sum_{x_1 \in \mathcal{M}_1} \sum_{\eta \in T^{x_1}} v(x_1) = \sum_{x_1 \in \mathcal{M}_1} \frac{1}{M_1} (1 - \lambda^{x_1}(T^{x_1})). \tag{6.27}$$

Thus if $1 - \lambda^{x_1}(T^{x_1}) \gg N^{-1}$, $v(x_1) \gg N^{-1}$. Now by (5.9) of Lemma 5.3, $\zeta_N \ll \log N$ implies that $\psi_N \gamma_1^N(\xi_1^{x_1}) \gg N^{-1}$ which in turn implies that $1 - \lambda^{x_1}(T^{x_1}) \gg N^{-1}$. The claim of the lemma now readily follows. \square

We are now ready to prove (3.10), (3.11) and (3.12). Clearly, Eq. (3.12) follows from (6.5), (6.17) of Lemma 6.2 and the bounds (6.22) which are valid for all $x_1 \in \mathcal{M}_1$, all $\eta \in T^{x_1}$ and all $\sigma \in \mathcal{V}_N \setminus \bar{W}$. Next, inserting (6.17) in (6.7) gives

$$Q_\eta(\sigma) = \mathbb{P}_\sigma \left(\tau_{\partial W^{x'_1}} < \tau_{T^{x'_1}} \right) \frac{\nu_1(x_1)}{M_2} (1 + \varepsilon_N). \tag{6.28}$$

Using (5.8) to express the probability in (6.28) and proceeding as in the proof of Lemma 6.2 to bound the term $c_{T^{x'_1}}(\sigma)$ (that is, the terms $b_{T^{x'_1}}(\sigma, \eta)$) appearing in that expression yields (3.11). Finally, inserting (6.17) in (6.6) gives

$$P_\eta(\sigma) = \mathbb{P}_\sigma \left(\tau_\eta < \tau_{(T^{x_1} \setminus \eta) \cup \partial W^{x_1}} \right) + \mathbb{P}_\sigma \left(\tau_{\partial W^{x_1}} < \tau_{T^{x_1}} \right) \frac{\nu_1(x_1)}{M_2} (1 + \varepsilon_N). \tag{6.29}$$

Using Lemma 6.2 to bound the two probabilities appearing above, reasoning again as in the proof of Lemma 6.2 to bound the terms $b_{T^{x_1}}(\sigma, \eta)$ and $c_{T^{x_1}}(\sigma)$, proves (3.10). As for (i-4), it follows from Corollary 5.7. The proof of Assertion (i) of Proposition 3.1 is complete. \square

Proof of Assertions (ii) and (iii) of Propositions 3.1. The proofs of these two assertions are similar to those of (3.10) and (3.11) and present no new difficulties. As before they center on “renewal systems” that closely resemble (6.16) and that we now describe.

Given $x_1, \bar{x}_1 \in \mathcal{M}_1$, let $\eta \in T^{x_1}$ and $\bar{\eta} \in T^{\bar{x}_1}$. Instead of the three quantities of (6.1) we now need to distinguish six quantities, denoted by $P_\eta^-(\sigma), Q_\eta^-(\sigma), R_\eta^-(\sigma)$ and $P_{\bar{\eta}}^\neq(\sigma), Q_{\bar{\eta}}^\neq(\sigma), R_{\bar{\eta}}^\neq(\sigma)$, and defined as follows: letting the symbol $*$ stand for $=$ if $x_1 = \bar{x}_1$ and \neq if $x_1 \neq \bar{x}_1$,

$$\mathbb{P}_\sigma \left(\tau_\eta < \tau_{T \setminus \{\eta, \bar{\eta}\}} \right) \equiv \begin{cases} P_\eta^*(\sigma) & \text{if } \sigma \in W^{x_1} \setminus (T^{x_1} \setminus \bar{\eta}), \\ Q_\eta^*(\sigma) & \text{if } \sigma \in W^{x'_1} \setminus (T^{x'_1} \setminus \bar{\eta}) \text{ for some } x'_1 \in \mathcal{M}_1 \setminus x_1, \\ R_\eta^*(\sigma) & \text{if } \sigma \in \mathcal{S}_N \setminus (\bar{W} \setminus \bar{\eta}) \end{cases} \tag{6.30}$$

where $T^{x'_1} \setminus \bar{\eta} = T^{x_1}$ if $\bar{\eta} \notin T^{x_1}$. Note that the probabilities (3.14) and (3.15) are of the form, respectively, $\bar{P}_\eta(\eta)$ and $\bar{Q}_\eta(\eta)$. As before they can be expressed as functions of, respectively, $R_\eta^-(\sigma)$ and $R_\eta^\neq(\sigma)$. Proceeding exactly as in the derivation of (6.8) and (6.9), we get

$$\begin{aligned} P_\eta^-(\bar{\eta}) &= \mathbb{P}_{\bar{\eta}} \left(\tau_\eta < \tau_{(T^{x_1} \setminus \{\eta, \bar{\eta}\}) \cup \partial W^{x_1}} \right) + \sum_{\sigma' \in \partial W^{x_1}} \mathbb{P}_{\bar{\eta}} \left(\tau_{\sigma'} < \tau_{(\partial W^{x_1} \setminus \sigma') \cup (T^{x_1} \setminus \bar{\eta})} \right) R_\eta^-(\sigma'), \\ Q_\eta^\neq(\bar{\eta}) &= \sum_{\sigma' \in \partial W^{\bar{x}_1}} \mathbb{P}_{\bar{\eta}} \left(\tau_{\sigma'} < \tau_{(\partial W^{\bar{x}_1} \setminus \sigma') \cup (T^{\bar{x}_1} \setminus \bar{\eta})} \right) R_\eta^\neq(\sigma'). \end{aligned} \tag{6.31}$$

A reasoning similar to that which leads to (6.16) yields, with the same notational convention as above

$$R_\eta^*(\sigma) = b_\eta^*(\sigma) + \sum_{\sigma'' \in \partial W^{x'_1}} a^*(\sigma, \sigma'') R_\eta^*(\sigma''), \tag{6.32}$$

where for all $\sigma \in \mathcal{S}_N \setminus (\overline{W} \setminus \bar{\eta})$

$$\begin{aligned} b_\eta^-(\sigma) &= (1 + P_\eta^-(\bar{\eta})) \mathbb{P}_\sigma \left(\tau_\eta < \tau_{\overline{W} \setminus \eta} \right) \\ &+ \sum_{\sigma' \in W^{x_1} \setminus T^{x_1}} \mathbb{P}_\sigma \left(\tau_{\sigma'} < \tau_{\overline{W} \setminus \sigma'} \right) \mathbb{P}_\sigma \left(\tau_\eta < \tau_{(T^{x_1} \setminus \{\eta, \bar{\eta}\}) \cup \partial W^{x_1}} \right), \end{aligned} \tag{6.33}$$

$$\begin{aligned} b_\eta^\neq(\sigma) &= (1 + Q_\eta^\neq(\bar{\eta})) \mathbb{P}_\sigma \left(\tau_\eta < \tau_{\overline{W} \setminus \eta} \right) \\ &+ \sum_{\sigma' \in W^{x_1} \setminus T^{x_1}} \mathbb{P}_\sigma \left(\tau_{\sigma'} < \tau_{\overline{W} \setminus \sigma'} \right) \mathbb{P}_\sigma \left(\tau_\eta < \tau_{(T^{x_1} \setminus \eta) \cup \partial W^{x_1}} \right) \end{aligned} \tag{6.34}$$

and, for $\sigma \in \mathcal{S}_N \setminus (\overline{W} \setminus \bar{\eta})$ and $\sigma'' \in \partial W^{x'_1}$, $a^-(\sigma, \sigma'') = a^\neq(\sigma, \sigma'') = \bar{a}(\sigma, \sigma'')$ where

$$\bar{a}(\sigma, \sigma'') \equiv \sum_{x'_1 \in \mathcal{M}_1} \sum_{\sigma' \in W^{x'_1} \setminus T^{x'_1}} \mathbb{P}_\sigma \left(\tau_{\sigma'} < \tau_{\overline{W} \setminus \sigma'} \right) \mathbb{P}_{\sigma'} \left(\tau_{\sigma''} < \tau_{(\partial W^{x'_1} \setminus \sigma'') \cup (T^{x'_1} \setminus \bar{\eta})} \right). \tag{6.35}$$

Note that by virtue of Proposition 5.6, the terms $P_\eta^-(\bar{\eta})$ and $\overline{Q}_\eta^\neq(\bar{\eta})$ in (6.33) and (6.34) are absorbed in the ε_N term of (3.14) and (3.15). The proof of Assertions (ii) and (iii) of Propositions 3.1 are now reruns of the proof of Assertion (i). We omit the details. \square

6.2 Proof of Proposition 3.3

There are two parts to the argument, a quite simple one: how the process leaves \mathcal{M} ; and the bulk of the argument: how it returns to \mathcal{M} . The latter part would be quite straightforward – apart from a result about the random walk in the hypercube and Ehrenfest chains; see Lemma A.5 –, if we could assume that both the returns to first level ground state configurations and to second level ground state configurations are uniform at each instance, irrespective of the starting point. Of course this is only approximately the case at each instance, and there are exponentially many instances to be controlled; so, in the end, indeed an extensive, careful analysis, is required, tailored to take advantage of existing bounds on the approximation of the law of the random walk in the hypercube to the uniform law.

6.2.1 Transition within \mathcal{M} : leaving \mathcal{M}

By the rules of our dynamics when leaving $x_1 x_2 \in \mathcal{M}$, the probability to jump to $x'_1 x_2$ for some $x'_1 \sim x_1$ equals $q_N^*(\xi^{x_1})$ (recall (4.3)), which vanishes as $N \rightarrow \infty$ for $x_1 \in \mathcal{M}_1$. Once \bar{X}^N leaves $x_1 x_2 \in \mathcal{M}$ and goes to some neighboring $x_1 x'_2$, while \bar{X}_1^N rests, the number of jumps \bar{X}_2^N would have to take before coming back to \mathcal{M}_2^4 is of the order of 2^{N_2} (by Corollary 1.8 of [6]), which in this temperature regime is much larger than $1/q_N^*(\xi^{x_1})$, the order of the number of jumps of \bar{X}^N before \bar{X}_1^N moves. The upshot is that with probability tending to 1 as $N \rightarrow \infty$, starting from \mathcal{M} , \bar{X}^N first leaves \mathcal{M} in such a way that \bar{X}_1^N leaves \mathcal{M}_1 before \bar{X}^N returns to \mathcal{M} .

⁴Let us recall from Lemma 4.5 that x'_2 may be assumed not in \mathcal{M}_2 .

6.2.2 Transition within \mathcal{M} : return to \mathcal{M}

In the presentation of the arguments in the remainder of this subsection, we find it convenient to go back to the representation \bar{X}^N of our process (introduced in Subsection 2.2). Let $\tau^1 = \inf\{t > 0 : \bar{X}_1^N \in \mathcal{M}_1\}$, $\tilde{\tau}^1 = \inf\{t > \tau^1 : \bar{X}_1^N \notin \mathcal{M}_1\}$ and, for $i \geq 2$, $\tau^i = \inf\{t > \tilde{\tau}^{i-1} : \sigma^N \in \mathcal{M}_1\}$, $\tilde{\tau}^i = \inf\{t > \tau^i : \bar{X}_1^N \notin \mathcal{M}_1\}$. We then have that τ^i , $i = 1, 2, \dots$, represent the successive hitting times of \mathcal{M}_1 by \bar{X}_1^N . Notice that $\tau^1 = \tau_{\bar{W}}$. Let also A_i , resp. A_i^y , $i = 1, 2, \dots$, denote the event that \bar{X}_2^N hits \mathcal{M}_2 , resp. $y \in \mathcal{M}_2$ during the i -th visit of \bar{X}_1^N to \mathcal{M}_1 . Let $\mathcal{I} = \min\{i \geq 1 : A_i \text{ occurs}\}$. Below we will compute the limit as $N \rightarrow \infty$ of

$$\mathbb{P}_\sigma(\bar{X}_1^N(\tau^{\mathcal{I}}) = x, A_{\mathcal{I}}^y), xy \in \mathcal{M}. \tag{6.36}$$

From the discussion on Subsubsection 6.2.1, we may take $\sigma \notin W$. The expression in (6.36) is not quite the probability in the left hand side of (3.18), but close enough in the sense that they turn out to have the same limit, as will also be argued below, in the conclusion of the our proof.

For $x \in \mathcal{M}_1$, set

$$\pi(x) = \mathbb{P}_{\mu_2}(A_1^y | \bar{X}_1^N(0) = x); \quad \hat{\pi}(x) = \mathbb{P}_{\mu_2}(A_1 | \bar{X}_1^N(0) = x), \tag{6.37}$$

where μ_2 is the uniform initial distribution of \bar{X}_2^N (on \mathcal{V}_{N_2}). Notice that $\pi(x)$ does not depend on y . We show in the appendix – see Lemma A.5 – that

$$\pi(x) \sim \frac{N_2}{N_1} \gamma_1(x) \frac{1}{c_1^N 2^{N_2}}; \quad \hat{\pi}(x) \sim M_2 \pi(x). \tag{6.38}$$

Let now D_i denote the event that \bar{X}_2^N gives at least N_2^3 steps between times τ^{i-1} and τ^i , $i \geq 2$. Now let $L_1 = \inf\{i \geq 2 : D_i \text{ occurs}\}$, and, for $k \geq 2$, $L_k = \inf\{i > L_{k-1} : D_i \text{ occurs}\}$, and define $\bar{\tau}^1 = \tau^1$ and $\bar{\tau}^k = \tau^{L_k}$, $k \geq 2$. Let also \bar{A}_i , $i = 1, 2, \dots$, denote the event that \bar{X}_2^N hits \mathcal{M}_2 while $\bar{X}_1^N \in \mathcal{M}_1$ between times $\bar{\tau}^i$ and $\bar{\tau}^{i+1}$, $i \geq 1$. Finally, let $\bar{\mathcal{I}} = \min\{i \geq 1 : \bar{A}_i \text{ occurs}\}$.

We will now couple \bar{X}^N to a process $\check{X}^N = \check{X}_1^N \check{X}_2^N$ which moves like \bar{X}^N , except that at the times $\bar{\tau}^i$, $i \geq 1$, it is uniformly distributed in the second coordinate. Lemma A.4 and Lemma 3.1 of [13] should in principle allow for a direct argument for such a coupling, but there is a periodicity issue for the application of the latter result, so we find it convenient to go through an intermediate step, in order to deal with that issue.

Let $\check{X}^N = \check{X}_1^N \check{X}_2^N$ be such that $\check{X}_1^N = \bar{X}_1^N$; for the definition of \check{X}_2^N , we partition the trajectory of \bar{X}_2^N into alternating portions, the first kind of which starts at a time $\bar{\tau}^i$ for some $i \geq 1$, when $\bar{X}_1^N = x_1$ for some $x_1 \in \mathcal{M}_1$, and ends when \bar{X}_1^N jumps (away from x_1); the second kind are the remaining portions of the trajectory of \bar{X}_2^N . Let us denote the number of steps in a given first kind of portion by \mathcal{G} (it is a geometric random variable).

\check{X}_2^N will be defined so that it gives the same number of steps as \bar{X}_2^N within each portion of their respective trajectories, in the same overall order, as follows. We let $\check{X}_2^N = \bar{X}_2^N$ for the second kind of portion. For any given first kind of portion, let Υ_i denote it, we will consider an extra step of \bar{X}_2^N after the \mathcal{G} -th step, independent of all else, except of course the position of \bar{X}_2^N after that step. Let $\check{\Upsilon}_i$ denote the trajectory obtained by putting that extra step at the end of the portion of Υ_i starting at the *second* step of \bar{X}_2^N within Υ_i . Notice that both portions of trajectory, Υ_i and $\check{\Upsilon}_i$, have the same number of steps, namely \mathcal{G} . We then toss an independent fair coin; if it comes up heads, then we make $\check{X}_2^N = \bar{X}_2^N$ for that portion of the trajectory as well; otherwise, if it comes up tails, then we make \check{X}_2^N coincide with $\check{\Upsilon}_i$, instead.

This extra coin-tossing randomness solves the periodicity issue, and we may now apply Lemma A.4 and Lemma 3.1 of [13] to couple \check{X}^N to a process $\check{\check{X}}^N = \check{\check{X}}_1^N \check{\check{X}}_2^N$ such

that $\check{X}_1^N = \dot{X}_1^N$, and, defining the random times in the above paragraph in the same way (and with the same notation) for \check{X}^N and \check{X}^N , we have that \check{X}_2^N is uniformly distributed on \mathcal{D}_2 at the times $\bar{\tau}^i$, $i \geq 2$, in such a way that, with probability tending to 1 as $N \rightarrow \infty$, we have that $\check{X}_2^N = \dot{X}_2^N$ for all times till $\bar{\tau}^{\bar{\mathcal{I}}}$. Notice that, since $\bar{\mathcal{I}} \leq \mathcal{I}$, Lemma A.4 holds also for $\bar{\mathcal{I}}$.

We now proceed to compute $\mathbb{P}_\sigma(\bar{X}_1^N(\tau^{\mathcal{I}}) = x, A_{\mathcal{I}}^y)$, $x \in \mathcal{M}_1$. The first step is to replace \bar{X}^N in that expression by \check{X}^N . The error in thus doing is bounded above by the probability that either there is a visit of \bar{X}_2^N to y on the first step of Υ_i or on the last (extra) step of $\check{\Upsilon}_i$ for some $i \leq \bar{\mathcal{I}}$. But this is readily checked to be an $o(1)$ by applying Lemma A.4 and Lemma 3.1 of [13].

For the next step, let $\mathcal{I}' = \min\{i \geq 0 : A_{L_{\bar{\mathcal{I}}+i}} \text{ occurs}\}$. Then for $x \in \mathcal{M}_1$, apart from an $o(1)$ error according to the above paragraphs, we have that

$$\mathbb{P}_\sigma(\bar{X}_1^N(\tau^{\mathcal{I}}) = x, A_{\mathcal{I}}^y) = \mathbb{P}_\sigma(\check{X}_1^N(\tau^{\mathcal{I}}) = x, A_{\mathcal{I}}^y) = \sum_{k=1}^{\infty} P_k(xy), \quad \text{where for } k \geq 2 \quad (6.39)$$

$$\begin{aligned} P_k(xy) &= \sum_{\ell=0}^{\infty} \sum_{z \in \mathcal{M}_1} \mathbb{P}_\sigma(\check{X}_1^N(\bar{\tau}^k) = z, \check{X}_1^N(\tau^{L_k+\ell}) = x, \bar{\mathcal{I}} = k, \mathcal{I}' = \ell, A_{L_k+\ell}^y) \\ &= \sum_{\ell=0}^{\infty} \sum_{x_1, \dots, x_{k+\ell-1} \in \mathcal{M}_1} \mathbb{P}_\sigma(\check{X}_1^N(\bar{\tau}^1) = x_1, \dots, \check{X}_1^N(\bar{\tau}^{k-1}) = x_{k-1}, \\ &\quad \check{X}_1^N(\tau^{L_k}) = x_k, \dots, \check{X}_1^N(\tau^{L_k+\ell-1}) = x_{k+\ell-1}, \check{X}_1^N(\tau^{L_k+\ell}) = x, \\ &\quad \bar{A}_1^c, \dots, \bar{A}_{k-1}^c, A_{L_k}^c, F_1, \dots, A_{L_k+\ell-1}^c, F_\ell, A_{L_k+\ell}^y), \end{aligned}$$

where F_i is the event that \check{X}_2^N gives less than N_2^3 jumps between τ^{L_k+i-1} and τ^{L_k+i} , $i = 1, 2, \dots$, and all the other quantities and events in the latter probability should be defined with \bar{X}^N replaced by \check{X}^N . Using the Markov property, the right-hand side above can be written as

$$\begin{aligned} &M_1^{-1} \sum_{x_0, \dots, x_{k-1} \in \mathcal{M}_1} \prod_{i=1}^{k-1} (1 - \bar{\pi}(x_i)) \mathbb{P}(\check{X}_1^N(\bar{\tau}^i) = x_i | \check{X}_1^N(\bar{\tau}^{i-1}) = x_{i-1}, \bar{A}_{i-1}^c) \\ &\times \left[\pi(x) \mathbb{P}(\check{X}_1^N(\tau^{L_k}) = x | \check{X}_1^N(\bar{\tau}^{k-1}) = x_{k-1}, \bar{A}_{k-1}^c) + \sum_{\ell=1}^{\infty} \sum_{x_k, \dots, x_{k+\ell-1} \in \mathcal{M}_1} \mathcal{Q}_{x_k, \dots, x_{k+\ell-1}}^\ell \right], \end{aligned} \quad (6.40)$$

where $\bar{\pi}(x_i) = \mathbb{P}_{\mu_2}(\bar{A}_i | \bar{X}_1^N(\bar{\tau}^i) = x_i)$ (it does not depend on $i \geq 1$ except through x_i) and $\mathcal{Q}_{x_k, \dots, x_{k+\ell-1}}^\ell$ equals

$$\begin{aligned} &\mathbb{P}(\check{X}_1^N(\tau^{L_k}) = x_k, A_{L_k}^c, F_1, \dots, \check{X}_1^N(\tau^{L_k+\ell-1}) = x_{k+\ell-1}, A_{L_k+\ell-1}^c, F_\ell, \\ &\quad \check{X}_1^N(\tau^{L_k+\ell}) = x, A_{L_k+\ell}^y | \check{X}_1^N(\bar{\tau}^{k-1}) = x_{k-1}, \bar{A}_{k-1}^c). \end{aligned} \quad (6.41)$$

Removing the A^c 's and using the Markov property, we find that the expression in (6.41) is bounded above by

$$\begin{aligned} &\mathbb{P}(\check{X}_1^N(\tau^{L_k}) = x_k | \check{X}_1^N(\bar{\tau}^{k-1}) = x_{k-1}, \bar{A}_{k-1}^c) \mathbb{P}(F_1 | \check{X}_1^N(\tau^{L_k}) = x_k) \times \\ &\quad \prod_{i=1}^{\ell-1} \{ \mathbb{P}(\check{X}_1^N(\tau^{L_k+i}) = x_{k+i} | \check{X}_1^N(\tau^{L_k+i-1}) = x_{k+i-1}, F_i) \mathbb{P}(F_{i+1} | \check{X}_1^N(\tau^{L_k+i}) = x_{k+i}) \} \times \\ &\quad \mathbb{P}(\check{X}_1^N(\tau^{L_k+\ell}) = x | \check{X}_1^N(\tau^{L_k+\ell-1}) = x_{k+\ell-1}, F_\ell) \mathbb{P}(A_{L_k+\ell}^y | \check{X}_1^N(\tau^{L_k+\ell}) = x), \end{aligned} \quad (6.42)$$

since the events F_1, F_2, \dots depend only on \check{X}_1^N , and starting with a uniform distribution on \mathcal{D}_2 , its invariant distribution, at time $L_k, \check{X}_2^N(L_k + \ell)$ still has that same distribution.

Now from Lemma A.2 and Remark A.3 above, each probability of the form $\mathbb{P}(F_i|\cdot)$ in (6.42) are bounded above by c/N for some constant c . We also notice that the latter probability in the same expression equals $\pi(x)$. Thus the second term within brackets on the bottom of (6.40) is bounded above by

$$\begin{aligned} \pi(x) \sum_{\ell=1}^{\infty} \left(\frac{c}{N}\right)^\ell \sum_{x_k, \dots, x_{k+\ell-1} \in \mathcal{M}_1} \mathbb{P}(\check{X}_1^N(\tau^{L_k}) = x_k | \check{X}_1^N(\tau^{k-1}) = x_{k-1}, \bar{A}_{k-1}^c) \\ \times \prod_{i=1}^{\ell-1} \mathbb{P}(\check{X}_1^N(\tau^{L_k+i}) = x_{k+i} | \check{X}_1^N(\tau^{L_k+i-1}) = x_{k+i-1}, F_i), \end{aligned}$$

and since the latter sum is over probabilities, it equals 1. It follows that the expression within brackets on the bottom of (6.40) equals

$$[\mathbb{P}(\check{X}_1^N(\tau^{L_k}) = x | \check{X}_1^N(\tau^{k-1}) = x_{k-1}, \bar{A}_{k-1}^c) + o_{x_{k-1}, x}] \pi(x), \tag{6.43}$$

where $o_{x', x}$ is an $o(1)$ for every $x, x' \in \mathcal{M}_1$.

Let us now consider $\bar{\pi}(z)$ for a given $z \in \mathcal{M}_1$; arguing in the same way as for the expression within brackets on the bottom of (6.40), we find that it equals $\hat{\pi}(z)$ plus an error bounded above by

$$\sum_{\ell=1}^{\infty} \left(\frac{c}{N}\right)^\ell \sum_{x'_1, \dots, x'_\ell \in \mathcal{M}_1} \prod_{i=1}^{\ell} \mathbb{P}(\check{X}_1^N(\tau^{L_{k-1}+i}) = x'_i | \check{X}_1^N(\tau^{L_{k-1}+i-1}) = x'_{i-1}, F'_i) \hat{\pi}(x'_\ell),$$

where F'_1, F'_2, \dots are defined in the obvious, parallel way to F_1, F_2, \dots above. From (6.38), we have that $\hat{\pi}(x), x \in \mathcal{M}_1$, are all of the same order of magnitude. It follows that

$$\bar{\pi}(x_i) = (1 + o_{x_i}) \hat{\pi}(x_i), \quad x_i \in \mathcal{M}_1, \tag{6.44}$$

where o_x is an $o(1)$ for every $x \in \mathcal{M}_1$.

In particular, we have that $\mathbb{P}_{\mu_2}(\bar{A}_i^c | \bar{X}_1^N(\bar{\tau}^i) = x_i) \sim 1$ for all $i = 1, 2, \dots$ and using also Corollary 1.5 of [6], we find that

$$\bar{P}_{x,y} := \mathbb{P}(\check{X}_1^N(\bar{\tau}^i) = y | \check{X}_1^N(\bar{\tau}^{i-1}) = x, \bar{A}_{i-1}^c) \sim \frac{1}{M_1}, \quad x, y \in \mathcal{M}_1. \tag{6.45}$$

We note that the latter conditional probability does not depend on $i = 1, 2, \dots$

The upshot of the above discussion is that the right-hand side of (6.39) may be written as

$$(1 + o(1)) \frac{1}{M_1} \sum_{x_0 \in \mathcal{M}_1} \sum_{k=1}^{\infty} \bar{R}_{x_0, x}^k \frac{\pi(x)}{1 - \bar{\pi}(x)}, \tag{6.46}$$

where \bar{R}^k is the k -th power of the matrix $\bar{R} = \bar{P}(I - \bar{\Pi})$, where I is the identity matrix in \mathcal{M}_1 and $\bar{\Pi}$ is the diagonal matrix in \mathcal{M}_1 with entries $\{\bar{\pi}(y), y \in \mathcal{M}_1\}$.

Now by (6.38), we have that $\bar{\Pi} \sim M_2 \epsilon \Gamma$, with $\epsilon = \frac{N_2}{N_1} \frac{1}{c_1^{N_2+1}}$ and $\Gamma = \text{diag} \{\gamma_1(y), y \in \mathcal{M}_1\}$ is the diagonal matrix in \mathcal{M}_1 with entries $\{\gamma_1(y), y \in \mathcal{M}_1\}$; thus, it is an $o(1)$ and from (6.45), we have that \bar{R} is a positive matrix for all large enough N . We may thus apply the Perron-Frobenius theory to write the internal sum in (6.46) as

$$\sum_{k=1}^{\infty} \rho^k [(\rho^{-1} \bar{R})^k - \bar{S}]_{x_0, x} \frac{\pi(x)}{1 - \bar{\pi}(x)} + \frac{1}{1 - \bar{\pi}(x)} \frac{\bar{S}_{x_0, x} \pi(x)}{1 - \rho}, \tag{6.47}$$

where ρ is the top eigenvalue of \bar{R} , and $\bar{S} = vw^T$, with v, w the right and left eigenvectors of \bar{R} associated to ρ such that $v^T w = 1$. See Theorem 8.2.11, its proof, and the preceding

and subsequent material of Section 2 of Chapter 8 of [26]. To check that $\rho < 1$ for all large enough N , we first note that \bar{R} is a perturbation of \bar{P} , which is stochastic, and thus has 1 as its top eigenvalue, and then resort to a standard perturbation result to the effect that $\rho = 1 - M_2 \epsilon' \frac{\bar{w}^T P \Gamma \bar{v}}{\bar{w}^T \bar{v}} + C\epsilon^2$, where \bar{v} and \bar{w} are right and left eigenvectors of \bar{P} associated to the eigenvalue 1, $\epsilon' \sim \epsilon$, and C is a constant. See Theorem IV.2.3 in [31]. Since the latter matrix is stochastic, we may take \bar{v} as the vector with all entries equal to 1. By (6.45) and again well known perturbation results, we may take $\bar{w} \sim \bar{v}$ (see Subsection V.2.3 of [31]), and from (6.45) we have that $\bar{P} \sim \frac{1}{M_1} I$; we thus get $\rho = 1 - M_2 \epsilon'' \bar{\gamma}_1 + C\epsilon^2$, where $\bar{\gamma}_1 = \frac{1}{M_1} \sum_{y \in \mathcal{M}_1} \gamma_1(y)$, and $\epsilon'' \sim \epsilon$. We then have that $\rho < 1$ for all large enough N .

Again the Perron-Frobenius theory tells us that the expression within brackets in (6.47) decays exponentially fast in k , uniformly in N . Since $\pi(x), \bar{\pi}(x)$ are $o(1)$ for all $x \in \mathcal{M}_1$, the infinite sum in (6.47) vanishes as $N \rightarrow \infty$.

Again resorting to well known perturbation theory results, since \bar{R} is also a perturbation of $\frac{1}{M_1} I$, we have that v and w may be taken as $v \sim \bar{v} = (1, \dots, 1)$, and $w \sim \frac{1}{M_1} v$; it follows that $\bar{S} \sim \frac{1}{M_1} I$; the upshot is that the second summand in (6.47) is asymptotic to

$$\frac{\frac{1}{M_1} \epsilon \gamma_1(x)}{M_2 \epsilon'' \bar{\gamma}} \sim \frac{\gamma_1(x)}{\sum_{y \in \mathcal{M}_1} \gamma_1(y)} \frac{1}{M_2}, \tag{6.48}$$

and thus so is the left hand side of (6.39).

6.2.3 Conclusion of the proof of Proposition 3.3

Let us show that $P(\bar{X}_1^N(\tau^{\mathcal{I}}) = x, A_{\mathcal{I}}^y)$ agrees with the entrance probability at left-hand side of (3.18) apart from an $o(1)$ error. As it stands, the former probability is actually the probability that \bar{X}_2^N visits y during the first visit of \bar{X}_1^N to \mathcal{M}_1 where \bar{X}_2^N hits \mathcal{M}_2 : the visit of \bar{X}_2^N to y may be not the first one to \mathcal{M}_2 . However, this probability is clearly an upper bound for the entrance probability, and if we subtract the following probability, we obtain a lower bound. For $i = 1, 2, \dots$, let B_i denote the event that \bar{X}_2^N hits \mathcal{M}_2 at least twice during the i -th visit of \bar{X}_1^N to \mathcal{M}_1 . We may then estimate $\mathbb{P}_{\mu_2}(\check{X}_1^N(\tau^{\mathcal{I}}) = x, B_{\mathcal{I}})$ in the same way as above (starting in (6.39)), by replacing $A_{\mathcal{I}}^y, A_{L_k+\ell}^y$ by $B_{\mathcal{I}}, B_{L_k+\ell}$, respectively, and $\pi(x)$ by $\mathbb{P}_{\mu_2}(B_1 | \bar{X}_1^N(\tau^1) = x)$. But this is bounded above by the right-hand side of (A.12), which was shown above to be an $o(\pi(x))$. We thus get that $\mathbb{P}_{\mu_2}(\check{X}_1^N(\tau^{\mathcal{I}}) = x, B_{\mathcal{I}})$ is an $o(1)$, and subtracting it from $\mathbb{P}_{\mu_2}(\bar{X}_1^N(\tau^{\mathcal{I}}) = x, A_{\mathcal{I}}^y)$ gives us a lower bound for the entrance law; the right-hand side of (6.48) as the limit for the latter quantity follows.

7 Proof of Theorem 2.7

In this and next two sections, we present the proofs of the scaling limit theorems for X^N , one section for each proof. We will use the results of Section 3 on entrance laws. There are two remaining things to establish in the case of Theorems 2.4 and 2.5: that the process spends virtually all of the time at the top, and what time is spent at each visit of a top configuration. The structure of the proof of Theorem 2.7 is not dissimilar: we control time spent off the top states as before, then evaluate the time spent on visits of X_1^N to top first level states, and finally resort to a spectral gap argument to get the behavior on second level.

Specifically in this section, we will concentrate on showing that

1. \bar{X}^N spends virtually all the time on \mathcal{M} ;
2. the time \bar{X}_1^N spends on each visit to each $x_1 \in \mathcal{M}_1$ is roughly exponential with mean $f_3(x_1)$;

3. given an interval of constancy $I = [a, b)$ of \bar{X}_1^N where $\bar{X}_1^N = x_1$ for some $x_1 \in \mathcal{M}_1$, and t_1, \dots, t_k such that $a < t_1 < \dots, t_k < b$ for some $k \geq 1$, we have that $\bar{X}_2^N(t_1), \dots, \bar{X}_2^N(t_k)$ are roughly independent random variables taking values on \mathbb{N} , each distributed roughly with probability weights given by $\gamma_2(x_1, \cdot)$ normalized.

As anticipated in Subsection 2.6, Point 1 is quite straightforward, and so is Point 3 as far as first level transitions are concerned; second level transitions are readily dealt with through spectral gap estimates. The main work concerns Point 2, where the gist of the argument is a law of large numbers for the number of visits to a given second level ground state configuration before the jump from a given first level ground state configuration – see (7.6).

After arguing these points in variable detail, we sketch an argument on how they fit together in a proof of Theorem 2.7. We start with the second point, after a few remarks.

Let us notice that the total time spent by $\sigma^N(\cdot/c_2^N)$ on any single visit to a given $\sigma_1 \in \mathcal{V}_{N_1}$ can be written as

$$\sum_{j=0}^{\mathcal{G}} \tilde{\gamma}_2^N(\sigma_1 J_{N_2}^{\circ}(j)) T_j, \tag{7.1}$$

where \mathcal{G} is the geometric random variable $\tau_{\partial C}$ with mean $\frac{N_2}{N_1} \exp\{\beta\sqrt{aN}\Xi_{\sigma_1}^{(1)}\}$, $T_i, i \geq 1$, are iid mean 1 exponential random variables, and

$$\tilde{\gamma}_2^N(\sigma) = \frac{\frac{N}{N_2} \gamma_2^N(\sigma)}{1 + \frac{N_1}{N_2} e^{\beta\sqrt{aN}\Xi_{\sigma_1}^{(1)}}}. \tag{7.2}$$

Remark 7.1. It should be quite clear that (2.15) and (2.17) remain valid when we replace γ_2^N and γ_2 by $\tilde{\gamma}_2^N$ and $\tilde{\gamma}_2$, respectively – see paragraph right above the statement of Theorem 2.4 –; one reason for this is that the denominator in (7.2) tends to 1 as $N \rightarrow \infty$ for every fixed x_1 .

Remark 7.2. We may indeed assume that all the convergences mentioned in the previous remark, which are weak for the original environment, may be taken strong by going to another, suitable probability space for the environment (resorting to Skorohod’s theorem). We will effectively, and for convenience, assume below that we are in the full measure event of such a probability space where those convergences take place, and omit further reference to it.

7.1 Time spent on top first level visits

Let $x_1 \in \mathcal{M}_1$. Set $\mathcal{X}_2 = (\xi_2^{x_1})^{-1}(J_{N_2}^{\circ})$, where $\xi_2^{x_1}$ is the function from \mathcal{D}_2 to \mathcal{V}_{N_2} mapping x_2 to $\xi_2^{x_1 x_2}$. The time spent in \mathcal{M} by \bar{X}_1^N on its i -th visit to $x_1 \in \mathcal{M}_1$ can be expressed as

$$\hat{\Upsilon}_i^N(x_1) := c_1^N 2^{N_2} \sum_{j=0}^{\mathcal{G}} 1_{\{\mathcal{X}_2(j) \in \mathcal{M}_2\}} \tilde{\gamma}_2^N(x_1 \mathcal{X}_2(j)) T_j. \tag{7.3}$$

Lemma 7.3. For each $x_1 \in \mathbb{N}$, $\hat{\Upsilon}_i^N(x_1)$ converges weakly as $N \rightarrow \infty$ to an exponential random variable of mean $\mathfrak{f}_3^M(x_1) = \gamma_1(x_1) \sum_{x_2 \in \mathcal{M}_2} \gamma_2(x_1 x_2)$.

Proof. We write $\hat{\Upsilon}_i^N(x_1) \stackrel{d}{=} c_1^N (1 + \mathcal{G}) \sum_{x_2 \in \mathcal{M}_2} \gamma_2^N(x_1 x_2) L_N(x_2)$, where

$$L_N(x_2) := \frac{2^{N_2}}{1 + \mathcal{G}} \sum_{j=0}^{\mathcal{G}} 1_{\{\mathcal{X}_2(j) = x_2\}} T_j(x_2), \tag{7.4}$$

with $\{T_j(x_2); j \geq 0, x_2 \geq 1\}$ an iid family of mean one exponential random variables independent from \mathcal{G} . One may readily check that

$$c_1^N(1 + \mathcal{G}) \rightarrow \gamma_1(x_1) \mathcal{E} \tag{7.5}$$

in distribution as $N \rightarrow \infty$, where \mathcal{E} is a mean one exponential random variable. We will next show that

$$L_N(x_2) \rightarrow 1 \tag{7.6}$$

in probability as $N \rightarrow \infty$ for every $x_2 \in \mathcal{M}_2$. This and (7.5) readily implies that

$$\hat{\Upsilon}_i^N(x_1) \rightarrow \mathfrak{f}_3^M(x_1) \mathcal{E} \tag{7.7}$$

in distribution as $N \rightarrow \infty$. From (7.5), since \mathcal{G} is independent from the family of exponential random variables entering $L_N(x_2)$, we may suppose that \mathcal{G} is roughly equal to $\hat{c}_1^N r$, with $r > 0$ a real number, where $\hat{c}_1^N = 1/c_1^N$. So rather than $L_N(x_2)$, we may consider instead

$$\hat{L}_N(x_2, r) := \frac{2^{N_2}}{1 + \hat{c}_1^N r} \sum_{j=0}^{\hat{c}_1^N r} 1_{\{x_2(j)=x_2\}} T_j(x_2), \tag{7.8}$$

and show that

$$\hat{L}_N(x_2, r) \rightarrow 1 \tag{7.9}$$

in probability as $N \rightarrow \infty$ for every $r > 0$.

We now note that the sum in (7.8) may be understood as the time spent on $\xi_2^{x_1 x_2}$ by a continuous time space homogeneous simple symmetric random walk in the \mathcal{V}_{N_2} during the first $\hat{c}_1^N r$ jumps.

Let us consider the delayed renewal process associated to that random walk consisting of successive return times of that random walk to $\xi_2^{x_1 x_2}$. Each such time (with the possible exception of the first one) can be decomposed as the sum of a mean one exponential random variable – the initial time spent on $\xi_2^{x_1 x_2}$ by the random walk prior to that particular return to $\xi_2^{x_1 x_2}$ – and the hitting time of $\xi_2^{x_1 x_2}$ by the random walk starting from the neighbor of $\xi_2^{x_1 x_2}$ it jumps to after that initial time. We then have a delayed renewal process with renewal times $\mathcal{E}_1 + \mathcal{R}_1, \mathcal{E}_2 + \mathcal{R}_2, \dots$, with $\mathcal{E}_1, \mathcal{R}_1, \mathcal{E}_2, \mathcal{R}_2, \dots$ independent, $\mathcal{E}_2, \mathcal{E}_3, \dots$ and $\mathcal{R}_2, \mathcal{R}_3, \dots$ identically distributed, \mathcal{E}_2 a mean one exponential random variable, and \mathcal{R}_2 is distributed as the hitting time of $\xi_2^{x_1 x_2}$ starting from a nearest neighbor of $\xi_2^{x_1 x_2}$. \mathcal{E}_1 may either be distributed as \mathcal{E}_2 or vanish, depending on whether the state of \bar{X}_2^N at the beginning of the visit of \bar{X}_1^N to x_1 was x_2 or not. Similarly, \mathcal{R}_1 may be either distributed as \mathcal{R}_2 or as the hitting time of $\xi_2^{x_1 x_2}$ by the random walk starting from another site of \mathcal{V}_{N_2} not $\xi_2^{x_1 x_2}$ or a neighbor of $\xi_2^{x_1 x_2}$. As will be clear below, neither (the distribution of) \mathcal{E}_1 or \mathcal{R}_1 will play a role in the result. Let $S_n = \sum_{i=1}^n (\mathcal{E}_i + \mathcal{R}_i)$ and $S'_n = \sum_{i=1}^n \mathcal{R}_i$, $n \geq 1$. Let N_t the counting process associated to S_n , namely $N_t = N(t) = \sup\{n \geq 0 : S_n \leq t\}$, $t \geq 0$, $S_0 = 0$. Notice that the sum in (7.8) is bounded below and above respectively by

$$\sum_{j=0}^{N(\hat{c}_1^N r)} T_j(x_2), \quad \sum_{j=0}^{N(\hat{c}_1^N r+1)} T_j(x_2). \tag{7.10}$$

We now claim that in order to establish (7.9), it is enough to show that

$$\frac{1}{K} S'_Q \rightarrow 1 \tag{7.11}$$

in probability as $N \rightarrow \infty$, where $K = K_N = \hat{c}_1^N r$, $Q = Q_N = \hat{c}_1^N r 2^{-N_2}$. Indeed, from (7.11) and the law of large numbers satisfied by iid mean one exponential random

variables, it readily follows that $\frac{1}{K}S_Q \rightarrow 1$ in probability as $N \rightarrow \infty$. This in turn readily implies that $\frac{1}{Q}N_K \rightarrow 1$ in probability as $N \rightarrow \infty$, and again the law of large numbers satisfied by iid mean one exponential random variables implies that either of the two expressions in (7.10), after division by Q , converges to 1 in probability as $N \rightarrow \infty$. The claim is established.

We may ignore \mathcal{R}_1 in the argument for (7.11), or take it identically distributed to \mathcal{R}_2 . We take the Laplace transform of the left hand side of (7.11) as follows. For $t > 0$

$$\bar{\mathbb{E}}\left(e^{-t\frac{1}{K}S'_Q}\right) = \left[\bar{\mathbb{E}}\left(e^{-t\frac{1}{K}\mathcal{R}_2}\right)\right]^Q \tag{7.12}$$

where $\bar{\mathbb{E}}$ denotes expectation with respect to the law of $\bar{J}_{N_2}^\circ$. It follows from Proposition 7.7.i.b of [6], after a simple adaptation to continuous time, that the expression within square brackets on the right of (7.12) can be written as

$$\frac{1 + \tilde{o}}{1 + \frac{t[1+o(1)]}{\bar{c}_1^N r 2^{-N_2}}}, \tag{7.13}$$

where $\tilde{o} = o(\hat{c}_1^N 2^{-N_2})$, and (7.11) follows. □

7.2 Equilibrium on the second level

Let us check the third point of the list outlined at the beginning of the section.

We initially remark that during a constancy interval of \bar{X}_1^N , where $\bar{X}_1^N = x_1$ for a given $x_1 \in \mathcal{D}_1$, \bar{X}_2^N is the mapping via ξ of a continuous time simple random walk on the hypercube \mathcal{V}_{N_2} with mean waiting time at $\xi_2^{x_1 x_2}$ given by $\gamma_2^N(x_1, x_2)$, starting from whichever second level configuration σ^N was at the beginning of the interval. We denote this random walk by $\bar{\sigma}_2^N$.

Let us now briefly argue the claim that the time to reach equilibrium for that random walk is of order smaller than that of the length of the constancy interval – which we just saw above to be of the order of the inverse of $\bar{c}^N = c_1^N 2^{N_2} c_2^N$.

After straightforward adjustments, we may check that the bound derived in [19] for the associated Metropolis dynamics for the REM applies for $\bar{\sigma}_2^N$, and we find that

$$\limsup_{N \rightarrow \infty} \frac{1}{N_2} \log \mathfrak{T}_2 \leq \sqrt{1 - a\beta\beta_*} \tag{7.14}$$

almost surely, where \mathfrak{T}_2 is the inverse of the spectral gap of $\bar{\sigma}_2^N$. It follows that

$$\liminf_{N \rightarrow \infty} \frac{1}{N} [\log(\bar{c}_N)^{-1} - \log \mathfrak{T}_2] > 0, \tag{7.15}$$

and

$$\max_{\sigma_2 \in \mathcal{V}_{N_2}} \left\| \mathbb{P}_{\sigma_2}(\bar{\sigma}_2^N(t) = \cdot) - G_{N_2}(\cdot) \right\| \leq \sqrt{Z_{N_2} \max_{\sigma_2 \in \mathcal{V}_{N_2}} \exp\{\beta\sqrt{(1-a)N}\Xi_{\xi_1^{x_1}\sigma_2}^{(2)}\}} e^{-t/\mathfrak{T}_2}, \tag{7.16}$$

where G_{N_2} is the equilibrium Gibbs measure for $\bar{\sigma}_2^N$, which is proportional to the weights $\exp\{-\beta\sqrt{(1-a)N}\Xi_{\xi_1^{x_1}}^{(2)}\}$, and Z_{N_2} is the partition function associated to G_{N_2} .

From well known results about the existence and exact expression for the limit of $\frac{1}{N} \log$ of both factors inside the square root above, we have that almost surely that square root is bounded from above by e^{cN} for some finite constant c for all large enough N . It immediately follows from (7.15) and the above that for times of the form $t = s(\bar{c}_N)^{-1}$, $s > 0$, we have that the left hand side of (7.16) is almost surely bounded above by $e^{cN} e^{-e^{dN}}$ for all large enough N , with $d > 0$ related to the left hand side of (7.15), and thus it almost surely vanishes as $N \rightarrow \infty$. This and (2.15) in turn readily imply the claim of the third point at the beginning of the section.

7.3 Time spent by \bar{X}^N outside \mathcal{M}

7.3.1 Preliminaries

We start with results about the number of visits of a given configuration σ' by a random walk on \mathcal{V}_N before reaching vertex $\sigma \neq \sigma'$. There are two initial situations: equilibrium and σ . Let $\tau_\sigma = \inf\{k \geq 1 : J_N^\circ(k) = \sigma\}$, where J_N° is the random walk on \mathcal{V}_N . We know from elementary theory of Markov chains that

$$\mathbb{E}_\sigma^\circ \left(\sum_{k=0}^{\tau_\sigma-1} 1\{J_N^\circ(k) = \sigma'\} \right) = 1 \tag{7.17}$$

(see e.g. Theorems 1.7.5 and 1.7.6 in [28]; to get (7.17) we use also the fact that the uniform measure on the vertices of the hypercube is invariant for \mathcal{X} , which is moreover irreducible). Let μ denote the uniform invariant measure for J° .

Lemma 7.4. *Suppose $\sigma \neq \sigma'$. For all large enough n*

$$\mathbb{E}_\mu^\circ \left(\sum_{k=0}^{\tau_\sigma-1} 1\{J_N^\circ(k) = \sigma'\} \right) \leq 2. \tag{7.18}$$

Proof. We write the left hand side of (7.18) as follows

$$\begin{aligned} \sum_{k=0}^{\infty} \mathbb{P}_\mu^\circ(J_N^\circ(k) = \sigma', \tau_\sigma \geq k+1) &= \frac{1}{2^n} + \frac{1}{2^n} \sum_{k=1}^{\infty} \sum_{\sigma'' \neq \sigma} \mathbb{P}_{\sigma''}^\circ(J_N^\circ(k) = \sigma', \tau_\sigma \geq k+1) \\ &\stackrel{*}{=} \frac{1}{2^n} + \frac{1}{2^n} \sum_{k=1}^{\infty} \sum_{z \neq y} \mathbb{P}_{\sigma'}^\circ(\tau_\sigma \geq k, J_N^\circ(k) = \sigma'') \\ &= \frac{1}{2^n} + \frac{1}{2^n} \sum_{k=1}^{\infty} \mathbb{P}_{\sigma'}^\circ(\tau_\sigma \geq k) - \frac{1}{2^n} \sum_{k=1}^{\infty} \mathbb{P}_{\sigma'}^\circ(\tau_\sigma \geq k, J_N^\circ(k) = \sigma) \\ &= \frac{1}{2^n} + \frac{1}{2^n} \sum_{k=1}^{\infty} \mathbb{P}_{\sigma'}^\circ(\tau_\sigma \geq k) - \frac{1}{2^n} \sum_{k=1}^{\infty} \mathbb{P}_{\sigma'}^\circ(\tau_\sigma = k) = \frac{1}{2^n} \mathbb{E}_{\sigma'}^\circ(\tau_\sigma) \leq 2, \end{aligned}$$

for all large enough n , where $\stackrel{*}{=}$ is due to the reversibility of \mathcal{X} , and the inequality at the end follows from Theorem 1.6 of [6]. □

7.3.2 Time outside T_1

Let us estimate the time spent by \bar{X}_1^N outside T_1 till the first visit to a vertex $\sigma_1 \in T_1$, and between two visits to σ_1 . Let us fix $x_1 \in \mathcal{M}_1$ and take $\sigma_1 = \xi_1^{x_1}$.

Let \mathcal{U} denote the first such time, which can be written as follows.

$$\mathcal{U} = c_1^N 2^{N_2} \sum_{k=0}^{\tau_{\sigma_1}-1} 1\{J_{N_1}^\circ(k) \notin T_1\} \sum_{j=0}^{\mathcal{G}_{k+1}(J_{N_1}^\circ(k))} \tilde{\gamma}_2^N(J_{N_1}^\circ(k) J_{N_2}^\circ(\mathfrak{G}_k + j)) T_j^k, \tag{7.19}$$

where $\mathfrak{G} := \{\mathcal{G}_1(\sigma'_1), \mathcal{G}_2(\sigma'_1), \dots; \sigma'_1 \in \mathcal{V}_{N_1}\}$ is a family of independent geometric random variables with mean $\frac{N_2}{N_1} e^{-\beta H_1(\sigma'_1)}$, independent of $J_{N_1}^\circ$ and $J_{N_2}^\circ$; $\mathfrak{G}_k = \sum_{i=1}^k \mathcal{G}_i(J_{N_1}^\circ(i))$, $k \geq 1$; T_j^k , $j, k \geq 0$ are iid mean 1 exponential random variables, independent from all the other random variables.

We note that $J_{N_1}^\circ$ and $J_{N_2}^\circ$ are independent discrete time random walks on the hypercubes \mathcal{V}_{N_1} and \mathcal{V}_{N_2} , respectively, each starting from its respective equilibrium distribution. Thus

$$\mathbb{E}^\circ(\mathcal{U} | J_{N_1}^\circ, \mathfrak{G}) = c_1^N 2^{N_2} \sum_{k=0}^{\tau_{\sigma_1}-1} 1\{J_{N_1}^\circ(k) \notin T_1\} \sum_{j=0}^{\mathcal{G}_{k+1}(J_{N_1}^\circ(k))} E(\tilde{\gamma}_2^N(J_{N_1}^\circ(k) J_{N_2}^\circ(\mathfrak{G}_k + j)) | J_{N_1}^\circ, \mathfrak{G}).$$

The conditional expectation on the right-hand side may be written as

$$\sum_{\sigma_2 \in \mathcal{V}_{N_2}} \gamma_2^N(J_{N_1}^\circ(k)\sigma_2)P(J_{N_2}^\circ(\mathfrak{G}_k + j) = \sigma_2) = \frac{1}{2^{N_2}} \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \tilde{\gamma}_2^N(J_{N_1}^\circ(k)\sigma_2).$$

Thus

$$\begin{aligned} \mathbb{E}^\circ(\mathcal{U}|J_{N_1}^\circ) &= \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \sum_{k=0}^{\tau_{\sigma_1}-1} 1\{J_{N_1}^\circ(k) \notin T_1\} \tilde{\gamma}_2^N(J_{N_1}^\circ(k)\sigma_2) c_1^N \mathbb{E}^\circ(\mathcal{G}_{k+1}(J_{N_1}^\circ(k))|J_{N_1}^\circ) \\ &= \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \sum_{k=0}^{\tau_{\sigma_1}-1} 1\{J_{N_1}^\circ(k) \notin T_1\} \gamma_1^N(J_{N_1}^\circ(k)) \tilde{\gamma}_2^N(J_{N_1}^\circ(k)\sigma_2), \quad \text{and} \\ \mathbb{E}^\circ(\mathcal{U}) &= \sum_{\sigma'_1 \notin T_1} \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \gamma_1^N(\sigma'_1) \tilde{\gamma}_2^N(\sigma'_1\sigma_2) E\left(\sum_{k=0}^{\tau_{\sigma_1}-1} 1\{J_{N_1}^\circ(k) = \sigma'_1\}\right) \\ &\leq 2 \sum_{\sigma'_1 \notin T_1} \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \gamma_1^N(\sigma'_1) \tilde{\gamma}_2^N(\sigma'_1\sigma_2), \end{aligned} \tag{7.20}$$

where the inequality holds for all large enough N , according to Lemma 7.4.

It follows from (2.15) and Remark 7.1 that

$$\limsup_{N \rightarrow \infty} \mathbb{E}^\circ(\mathcal{U}) \leq 2 \sum_{\sigma'_1 \notin T_1} \sum_{\sigma_2 \in \mathbb{N}} \gamma_1(\sigma'_1) \gamma_2(\sigma'_1\sigma_2), \tag{7.21}$$

and thus $\lim_{M_1 \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}^\circ(\mathcal{U}) = 0$.

Let now \mathcal{W}_i denote the time spent by \bar{X}_1^N outside T_1 between the i -th and $i+1$ -st visit to $\sigma_1 \in T_1$, $i \geq 1$. A similar reasoning as above yields

$$\begin{aligned} \mathbb{E}^\circ(\mathcal{W}_i) &= \sum_{\sigma'_1 \notin T_1} \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \gamma_1^N(\sigma'_1) \tilde{\gamma}_2^N(\sigma'_1\sigma_2) E_{\sigma_1}\left(\sum_{k=0}^{\tau_{\sigma_1}-1} 1\{J_{N_1}^\circ(k) = \sigma'_1\}\right) \\ &= \sum_{\sigma'_1 \notin T_1} \sum_{\sigma_2 \in \mathcal{V}_{N_2}} \gamma_1^N(\sigma'_1) \tilde{\gamma}_2^N(\sigma'_1\sigma_2), \end{aligned} \tag{7.22}$$

where we used (7.17) in the last passage; again

$$\lim_{M_1 \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}^\circ(\mathcal{W}_i) = 0, \quad i \geq 1. \tag{7.23}$$

7.3.3 Time inside T_1 and outside \mathcal{M}_2

Let now

$$\tilde{\Upsilon}_i^N(\sigma_1) = c_1^N 2^{N_2} \sum_{j=0}^{\mathcal{G}} 1\{J_{N_2}^\circ(j) \notin \mathcal{M}_2\} \tilde{\gamma}_2^N(\sigma_1 J_{N_2}^\circ(j)) T_j \tag{7.24}$$

be the time spent outside \mathcal{M} by \bar{X}_1^N on its i -th visit to $\sigma_1 \in T_1$ - recall the notation on the paragraph of (7.3). A similar reasoning to that leading to (7.20) and (7.22) yields

$$\mathbb{E}^\circ(\tilde{\Upsilon}_i^N(\sigma_1)) = \gamma_1^N(\sigma'_1) \sum_{\eta \in T^{x'_1}} \tilde{\gamma}_2^N(\eta), \quad i \geq 1 \tag{7.25}$$

and again

$$\lim_{M_2 \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}^\circ(\tilde{\Upsilon}_i^N(\sigma_1)) = 0, \quad i \geq 1. \tag{7.26}$$

As a corollary to (7.26) and (7.7), we have that, recalling that $x_1 = (\xi_1)^{-1}(\sigma_1)$,

$$\Upsilon_i^N(\sigma_1) \rightarrow \mathfrak{f}_3(x_1) \mathcal{E} \tag{7.27}$$

in distribution as $N \rightarrow \infty$, where $\Upsilon_i^N(\sigma_1) = \hat{\Upsilon}_i^N(\sigma_1) + \tilde{\Upsilon}_i^N(\sigma_1)$ is the time spent by \bar{X}_1^N on its i -th visit to $\sigma_1 \in T_1$, and \mathcal{E} is a mean one exponential random variable.

7.4 Conclusion of the proof of Theorem 2.7

Let us now fit together the above points in an argument for Theorem 2.7.

From the first claim at the beginning of the section (argued in Subsection 7.3), it is enough to show that \bar{X}_1^N restricted to \mathcal{M}_1 converges to \bar{X}_1 restricted to \mathcal{M}_1 . We already know from (7.27) that the sojourn times of \bar{X}_1^N on the various vertices of \mathcal{M}_1 converge in distribution to the respective sojourn times of \bar{X}_1 . We only have to argue that the jump probabilities of \bar{X}_1^N restricted to \mathcal{M}_1 converge to the uniform jump probabilities of \bar{X}_1 restricted to \mathcal{M}_1 . But that is established in (3.13). We have then that \bar{X}_1^N converges in distribution to \bar{X}_1 in Skorohod space, and the full statement readily follows from the third point claimed at the beginning of the section (and argued in Subsection 7.2).

This concludes the proof of Theorem 2.7.

8 Proof of Theorem 2.4

We start by showing that \bar{X}^N spends virtually all the time on \mathcal{M} .

8.1 Time spent by \bar{X}^N outside \mathcal{M}

We will show that the expected time spent by \bar{X}^N outside \mathcal{M} until the first visit to \mathcal{M} or between consecutive visits to \mathcal{M} is small. Indeed, we will argue that the first such time (the others can be similarly treated) vanishes in probability as $M_1, M_2 \rightarrow \infty$ uniformly in N . We will be more precise next.

The first such time is bounded above by $\bar{U} = \sum_{i=1}^{\mathcal{I}} (\hat{U}_i + \tilde{U}_i)$, where

$$\hat{U}_i = \sum_{k=\tilde{\tau}^{i-1}+1}^{\tilde{\tau}^i-1} \sum_{j=0}^{\mathcal{G}_{k+1}(J_{N_1}^\circ(k))} \tilde{\gamma}_2^N(J_{N_1}^\circ(k)J_{N_2}^\circ(\mathfrak{S}_k + j)) T_j^k, \tag{8.1}$$

and, writing $\mathcal{X}_1 = (\xi_1)^{-1}(J_{N_1}^\circ)$,

$$\tilde{U}_i = \sum_{j=0}^{\mathcal{G}_{\tilde{\tau}^i+1}(J_{N_1}^\circ(\tilde{\tau}^i))} 1_{\{J_{N_2}^\circ(\mathfrak{S}_{\tilde{\tau}^i+j}) \notin T^{\mathcal{X}_1}\}} \tilde{\gamma}_2^N(J_{N_1}^\circ(\tilde{\tau}^i)J_{N_2}^\circ(\mathfrak{S}_{\tilde{\tau}^i} + j)) T_j^{\tilde{\tau}^i}, \tag{8.2}$$

where $\tilde{\tau}^i, i = 1, 2, \dots$ denote the successive hitting times of T_1 by $J_{N_1}^\circ$ (the jump chain of \bar{X}_1^N) – when $i = 0$, then in (8.1) $\tilde{\tau}_0$ is either 0 or -1 depending on whether $\bar{X}_1^N(0) \in T_1$ or not, respectively.

We will show that $\lim_{M_1 \rightarrow \infty} \lim_{M_2 \rightarrow \infty} \limsup_{N \rightarrow \infty} \bar{U} = 0$ in probability. Given the tightness result for $\mathcal{I}/[c_1^N 2^{N_2}]$ given in Lemma A.4, it is enough to show that for all R

$$\lim_{M_1 \rightarrow \infty} \lim_{M_2 \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}^\circ \left(\sum_{i=1}^{Rc_1^N 2^{N_2}} \hat{U}_i \right) = 0, \tag{8.3}$$

$$\lim_{M_1 \rightarrow \infty} \lim_{M_2 \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{E}^\circ \left(\sum_{i=1}^{Rc_1^N 2^{N_2}} \tilde{U}_i \right) = 0. \tag{8.4}$$

Let us first point out that for every $i \geq 1$, \hat{U}_i is bounded above stochastically by $\mathcal{U}/[c_1^N 2^{N_2}]$ (see (7.19) above). (8.3) then follows from the above and (7.23).

Now

$$\mathbb{E}^\circ(\tilde{U}_i) \leq \max_{x_1 \in \mathcal{M}_1} \mathbb{E}^\circ \left(\sum_{j=0}^{\mathcal{G}(x_1)} 1_{\{J_{N_2}^\circ(j) \notin T^{x_1}\}} \tilde{\gamma}_2^N(x_1 J_{N_2}^\circ(j)) \right), \tag{8.5}$$

where $\mathcal{G}(x_1)$ is a geometric random variable with mean $\frac{N_2}{N_1} \exp\{\beta \sqrt{aN} \Xi_{\xi_1^{x_1}}^{(1)}\}$. We find that the expectation on the right-hand side equals $\gamma_1^N(x_1) \sum_{x_2 > M_2} \tilde{\gamma}_2^N(x_1 x_2) / [c_1^N 2^{N_2}]$, and (8.4) follows upon substitutions into (8.5) and the left hand side of (8.4).

8.2 Conclusion of the proof of Theorem 2.4

Given the result in Subsection 8.1 and the usual *constancy interval matching* argument that can be used to show convergence in Skorohod spaces, it is enough to show convergence of the transition probabilities among sites in \mathcal{M} (in the process restricted to \mathcal{M} , which is a Markov jump process) to the respective ones of the respective limit process (the one restricted to \mathcal{M} , which is also a Markov jump process), and the convergence of the respective sojourn times. The latter convergence is quite clear, and the former follows immediately from Proposition 3.3.

9 Proof of Theorem 2.5

We start by observing that we can check that \bar{X}^N spends virtually all the time on \mathcal{M} by virtually the same argument as for below fine tuning. Indeed, the corresponding expressions in the present regime of (7.19), (7.22) and (7.24) are the same, except for the factor of $c_1^N 2^{N_2}$, which is absent in the present regime. But notice that that factor is bounded as $N \rightarrow \infty$, and so the arguments of Subsections 7.3.2 and 7.3.3 carry through.

We can then repeat the argument for the conclusion of the proof of Theorem 2.4 on Subsection 8.2, once we have the convergence of the transition probabilities of \bar{X}^N restricted to \mathcal{M} to those of the limiting 2-level K process restricted to \mathcal{M} .

For $x, y \in \mathbb{N}$, let $\mathfrak{P}_N(x, y)$ denote the transition probability of $\bar{X}^N|_{\mathcal{M}}$. Then, using the remark in Subsubsection 6.2.1 and Proposition 3.1.i-1 and i-2, we have that

$$\lim_{N \rightarrow \infty} \mathfrak{P}_N(x, y) = \mathfrak{P}(x, y) := \begin{cases} [(1 - \bar{\lambda}^{x_1}) + \bar{\lambda}^{x_1} \bar{\nu}_1(x_1)] \frac{1}{M_2}, & \text{if } x_1 = y_1, \\ \bar{\nu}_1(y_1) \bar{\lambda}^{y_1} \frac{1}{M_2}, & \text{otherwise,} \end{cases} \tag{9.1}$$

where $\bar{\lambda}^{y_1} = \frac{1}{1 + M_2 \bar{\gamma}_1(y_1)}$ and $\bar{\nu}_1(y_1) = \frac{1 - \bar{\lambda}^{y_1}}{\sum_{y'_1 \in \mathcal{M}_1} (1 - \bar{\lambda}^{y'_1})}$.

It is then enough to argue that $\mathfrak{P}(x, y)$ is the transition probability from x to y for the 2-level K process restricted to \mathcal{M} . We do that next.

Let $X|_{\mathcal{M}}$ denote X restricted to \mathcal{M} . We can construct $X|_{\mathcal{M}}$ as follows. Let \hat{X} denote the 1-level K process used in the construction of X as at the end of Subsection A.1. Let us now construct a 2-level process \hat{X} in the same way as X , except that we use $\hat{X}|_{\mathcal{M}_1}$ instead of \hat{X} . One readily checks that

1. $\hat{X}|_{\mathcal{M}_1}$ is a Markov jump process on \mathcal{M}_1 with uniform initial state, uniform transitions on \mathcal{M}_1 (we should allow loops), and jump rate at $x_1 \in \mathcal{M}_1$ given by $1/\bar{\gamma}_1(x_1)$;
2. $X|_{\mathcal{M}} = \hat{X}|_{\mathcal{M}}$;
3. letting $\hat{\mathcal{X}}_1$ denote the jump chain of $\hat{X}|_{\mathcal{M}_1}$, and, for $z \in \mathcal{D}_2$, defining the events

$$A_n = \{\text{during the } n + 1\text{-st sojourn period of } \hat{X}_1, \hat{X}_2 \text{ visits } \mathcal{M}_2\},$$

$$A_n^z = \{\text{during the } n + 1\text{-st sojourn period of } \hat{X}_1, \hat{X}_2 \text{ visits } z \text{ before visiting } \mathcal{M}_2 \setminus \{z\}\},$$

we have that, given $\hat{\mathcal{X}}_1$, the events A_n^{*n} , $n \geq 0$, with $*n = y$ or blank for each n , are independent, having respective (conditional) probabilities given by

- (a) in the case of $*n = \text{blank}$: $P(N_T > 0)$, where N is a Poisson counting process with intensity M_2 , and T is exponential with mean $\bar{\gamma}_1(\hat{\mathcal{X}}_1(n))$, N and T independent; we then have that N_T has a geometric distribution with success parameter $\frac{1}{1 + M_2 \bar{\gamma}_1(\hat{\mathcal{X}}_1(n))} = \bar{\lambda}^{\hat{\mathcal{X}}_1(n)}$, and thus $P(N_T > 0) = 1 - \bar{\lambda}^{\hat{\mathcal{X}}_1(n)}$;
- (b) similarly, the probability of A_n^y given $\hat{\mathcal{X}}_1$ equals $\frac{1 - \bar{\lambda}^{\hat{\mathcal{X}}_1(n)}}{M_2}$.

Let $\mathfrak{P}'(x, y)$ denote the transition probability of $X|_{\mathcal{M}}$ from x to y . From the above, we conclude that if $x_1 = y_1$, $\mathfrak{P}'(x, y) = P(A_0^{y_2}) +$

$$\sum_{n \geq 0} \sum_{w_1, \dots, w_{n-1} \in \mathcal{M}_1} P(A_0^c, \dots, A_{n-1}^c, A_n^{y_2}, \hat{\mathcal{X}}_1(1) = w_1, \dots, \hat{\mathcal{X}}_1(n-1) = w_{n-1}, \hat{\mathcal{X}}_1(n) = x_1) \\ = \frac{1 - \bar{\lambda}^{x_1}}{M_2} + \bar{\lambda}^{x_1} \frac{1}{M_1} \frac{1 - \bar{\lambda}^{x_1}}{M_2} \sum_{n \geq 0} \left(\frac{1}{M_1} \sum_{w_1 \in \mathcal{M}_1} \bar{\lambda}^{x_1} \right)^n,$$

which is readily checked to equal $\mathfrak{P}(x, y)$ given in (9.1), in this case. When $x_1 \neq y_1$, we have that the same expression as in (9.2) holds for $\mathfrak{P}'(x, y)$, except for the first term in the sum, which is absent, and thus it agrees with $\mathfrak{P}(x, y)$ again in this case.

10 Aging in the K processes

As anticipated in the introduction, we will derive aging results for σ^N in a two-stage scaling limit process. We first take the limit in the extreme time scale, where there is no aging, since σ^N is close to equilibrium in that time scale: we have already done that in our scaling limit theorems. In the second stage, we take a small time limit of the limiting K processes. We will be concerned with correlation functions which involve only clock processes of the limiting processes, so we will take the second limit only of the relevant clocks. We will keep the presentation brief, in particular at and below fine tuning, since the issues involved are quite clear in those regimes, and technicalities are quite well known and fairly straightforward.

Let $Y = Y_1 Y_2$ be the K process representing the scaling limit of either \tilde{X}^N or \bar{X}^N in Theorems 2.4, 2.5 and 2.7. We assume $Y(0) = \infty$ in the first case, $\infty\infty$ in the second case, and (∞, \bar{X}_2) in the latter case, where \bar{X}_2 is as in Theorem 2.7. Given $\theta > 0$, we are interested in taking the following limit

$$\lim_{\substack{t_w, t \rightarrow 0 \\ t/t_w \rightarrow \theta}} \Pi(t_w, t_w + t), \tag{10.1}$$

where $\Pi(\cdot, \cdot)$ was defined in 1.6. Recall also \mathfrak{N}_i defined in 1.5.

10.1 Below fine tuning

This is the simplest case, since on the one hand, Y_2 jumps at every time interval, so there is no point in considering \mathfrak{N}_2 (which has probability 0). On the other hand, Y_1 is a uniform K process with waiting function given by a Poisson process with intensity measure $\frac{c_1}{x^{\alpha_1+1}} dx$ for some constant c_1 . It is well known (see e.g., [7]) that the limit in (10.1) is given by the arcsine law $\text{Asl}_{\alpha_1}(1/(1+\theta))$, where

$$\text{Asl}_{\alpha}(u) = \frac{\sin \pi \alpha}{\pi} \int_0^u x^{-(1-\alpha)} (1-x)^{-\alpha} dx, \quad u, \alpha \in (0, 1). \tag{10.2}$$

This follows readily from the scaling limit of the clock process of Y_1 at small times. This issue will come up again in the other temperature regimes, so we let it rest for this regime.

10.2 At fine tuning

We first notice that, for $i = 1, 2$, $\mathfrak{N}_i = \{\mathcal{R}_i \cap (t_w, t_w + t) = \emptyset\}$, where \mathcal{R}_i , $i = 1, 2$, is the range of the clock processes Γ_1 and Γ' , respectively (see end of Subsection A.1 below). It is a simple matter to check that in this case $\mathfrak{N}_2 \subset \mathfrak{N}_1$, so indeed $\Pi(t_w, t_w + t) = pP(\mathfrak{N}_1) + (1-p)P(\mathfrak{N}_2)$. Let us now point out that, as is well known, (10.1) follows from

a small time scaling limit for the respective clocks in an appropriate topology. Let us first examine Γ' – recall the definition at the end of Subsection A.1, and the statement of Theorem 2.5. As pointed out at the end of Subsection A.1, within intervals of constancy of \dot{X} , the increments of Γ' are those of a uniform K process with waiting function $f(x_1, \cdot)$, where x_1 is the constant value of \dot{X} within the interval. We know that the clock process of a uniform K process with waiting function given by a Poisson process with intensity measure $\frac{c}{x^{\alpha+1}}dx$, $\alpha \in (0, 1)$, c any constant in $(0, \infty)$, converges in the J_1 Skorohod metric to an α -stable subordinator in the small time limit for almost every realization of the Poisson process (see e.g., [7] with $a = 0$). This and the fact that $f'_2(x_1, \cdot)$ are given by iid in x_1 Poisson processes with intensity measure $\frac{c_2}{x^{\alpha_2+1}}dx$, for some constant c_2 , yields $\varepsilon^{-1}\Gamma'(\varepsilon^{\alpha_2} \times \cdot) \rightarrow S_2(\cdot)$ in distribution on the J_1 Skorohod space as $\varepsilon \rightarrow 0$ for a.e. realization of f_2, f'_2 , where S_2 is an α_2 -stable subordinator. It readily follows that almost surely

$$\lim_{\substack{t_w, t \rightarrow 0 \\ t/t_w \rightarrow \theta}} P(\mathcal{R}_2 \cap (t_w, t_w + t) = \emptyset) = \text{Asl}_{\alpha_2}(1/(1 + \theta)). \tag{10.3}$$

It can be also readily checked that $\varepsilon^{-1}\Gamma_1(\varepsilon^{\alpha_1\alpha_2} \times \cdot) \rightarrow S'_1 := S_1 \circ S_2(\cdot)$ in distribution on the J_1 Skorohod space as $\varepsilon \rightarrow 0$ for a.e. realization of f_2, f'_2 , where S_1 is an α_1 -stable subordinator. S'_1 is thus an $\alpha_1\alpha_2$ -stable subordinator, and it follows that almost surely

$$\lim_{\substack{t_w, t \rightarrow 0 \\ t/t_w \rightarrow \theta}} P(\mathcal{R}_1 \cap (t_w, t_w + t) = \emptyset) = \text{Asl}_{\alpha_1\alpha_2}(1/(1 + \theta)). \tag{10.4}$$

10.3 Above fine tuning

We first point out that right after the weighted K process Y jumps out of any state in \mathbb{N}^2 , it gives infinitely many jumps within any nonempty open time interval. This tells us that $\mathfrak{N}_1 = \mathfrak{N}_2 = \{\mathcal{R} \cap (t_w, t_w + t) = \emptyset\}$, where \mathcal{R} is the range of Γ , the clock process of Y (see Subsection A.1 below).

Let us then derive the small time limit of Γ . Recall from [17] that we may write

$$\Gamma(r) = \sum_{x \in \mathbb{N}^2} \sum_{i=1}^{N_x(r)} \gamma_2(x) T_i^x, \tag{10.5}$$

where $N_x, x \in \mathbb{N}^2$, are independent Poisson counting processes with rate $\gamma_1(x_1)$, respectively, and T_i^x are iid mean one exponential random variables. We will now argue that

$$\varepsilon^{-1}\Gamma(\varepsilon^{\alpha_2} \times \cdot) \rightarrow S_2(\cdot) \tag{10.6}$$

in distribution on the J_1 Skorohod space as $\varepsilon \rightarrow 0$ for a.e. realization of γ_1, γ_2 , where S_2 is an α_2 -stable subordinator. It is enough to establish this convergence for

$$\tilde{\Gamma}(r) := \sum_{x \in \mathbb{N}^2} \gamma_2(x) N_x(r) \tag{10.7}$$

(see Lemma 2.1 in [7]). Since this is a subordinator, in order to get the small time convergence, it is enough to consider the Laplace exponent of the small time scaled $\tilde{\Gamma}$, given by

$$\tilde{\varphi}_\varepsilon(\lambda) = r \sum_{x_1 \in \mathbb{N}} \gamma_1(x_1) \varepsilon^{\alpha_2} \sum_{x_2 \in \mathbb{N}} (1 - e^{\lambda \varepsilon^{-1} \gamma_2(x)}). \tag{10.8}$$

It is convenient at this point to write the scaled inner sum as $\varepsilon^{\alpha_2} \sum_{u \in [0,1]} (1 - e^{\lambda \varepsilon^{-1} \gamma_2^{x_1}(u)})$, where $\gamma_2^{x_1}(\cdot), x_1 \in \mathbb{N}$, are iid sets of increments of an α_2 -stable subordinator in $[0, 1]$. By the scale invariance of that subordinator, we have that it equals in distribution

$$\varepsilon^{\alpha_2} \sum_{i=1}^{\lfloor \lambda^{\alpha_2} \varepsilon^{-\alpha_2} \rfloor} \sum_{u \in [0,1]} (1 - e^{\lambda \varepsilon^{-1} \gamma_2^i(u)}) \tag{10.9}$$

plus an independent random variable which is stochastically dominated by

$$\varepsilon^{\alpha_2} \sum_{u \in [0,1]} (1 - e^{\lambda \varepsilon^{-1} \gamma_2^1(u)}).$$

Now standard large deviation estimates and an application of Campbell's Theorem, showing that $Z := \sum_{u \in [0,1]} (1 - e^{\lambda \varepsilon^{-1} \gamma_2^1(u)})$ has an exponential moment, imply that $\tilde{\varphi}_\varepsilon(\lambda) \rightarrow \text{const} \times r \lambda^{\alpha_2}$, where $\text{const} = E(Z) \sum_{x_1 \in \mathbb{N}} \gamma_1(x_1)$. Given γ_1 , this is the Laplace exponent of an α_2 -stable subordinator, and (10.6) follows. Thus

$$\lim_{\substack{t_w, t \rightarrow 0 \\ t/t_w \rightarrow \theta}} \Pi(t_w, t_w + t) = \text{Asl}_{\alpha_2}(1/(1 + \theta)). \tag{10.10}$$

10.4 Final remark about aging at $\beta > \beta_2^{cr}$

We end this section on aging by pointing out, as anticipated in the introduction, that the aging results in [29] are consistent with ours only in the fine tuning regime. As also earlier anticipated, this is explained by the shorter time scale considered in that reference. In those time scales, all levels are supposed to be aging simultaneously. That indeed also happens in the fine tuning regime at the *short* extreme scale considered in this section. Recall the discussion on Subsubsection 1.2.2.

We may understand the aging results in the other regimes treated in detail so far as follows. Below fine tuning, we already explained that the second level is well within equilibrium, so it does not age in the short extreme time scale (also in not much shorter time scales). The aging behavior in that phase comes from the first level, with its characteristic α_1 exponent.

Above fine tuning, we have the opposite behavior: the first level by itself would be in equilibrium, thus not aging, and aging comes from the second level, with its characteristic α_2 exponent.

11 Scaling limit at intermediate temperatures

In this section, we briefly state and discuss our scaling limit and aging results for $\beta \in (\beta_1^{cr}, \beta_2^{cr})$. We will be rather sketchy, trusting the experienced reader to be able to readily fill in the gaps with standard arguments. Recall from the discussion around (7.1) that the total time spent by $\sigma^N(\cdot)$ on a single visit to a given $\sigma_1 \in \mathcal{V}_{N_1}$ can be written as

$$\frac{N}{N_2 + N_1 e^{\beta \sqrt{aN} \Xi_{\sigma_1}^{(1)}}} \sum_{j=0}^{\mathcal{G}} \exp\{\beta \sqrt{(1-a)N} \Xi_{\sigma_1 J_{N_2}^{\circ}(j)}^{(2)}\} T_j, \tag{11.1}$$

where again \mathcal{G} is a geometric random variable with mean $\frac{N_2}{N_1} \exp\{\beta \sqrt{aN} \Xi_{\sigma_1}^{(1)}\}$, and $T_i, i \geq 0$, are iid standard exponential random variables.

For top first level configurations $\sigma_1 = \xi_1^{x_1}$, the factor in front of the sum above contributes a constant $(1/(1-p))$ almost surely in the limit as $N \rightarrow \infty$, and thus we are left to properly scale the sum itself. At this point we may replace \mathcal{G} by $\frac{1-p}{p} \frac{1}{c_1^N} \gamma_1(x_1) T$, with T a standard exponential, independent of all the other remaining random variables, and then resort to Proposition 1.9.ii of [25] which gives the proper scaling of the sum, as well as conditions under which the scaled sum satisfies a law of large numbers. Since the result in [25] applies for the REM (which is indeed the model appearing in the above sum), we need to do some translation in terms of our parameters. Upon doing that, we find that the proper scaling is given by $\tilde{c}_N = c_1^N \exp\{-\frac{\beta^2 N}{2}(1-a)\}$, and, provided

$\beta < \frac{1-p}{2\sqrt{ap}}\beta_* = \beta^{FT}$, the following law of large numbers holds:

$$\tilde{c}_N \sum_{j=0}^{t/\tilde{c}_1^N} \exp\{\beta\sqrt{(1-a)N} \Xi_{\sigma_1 J_{N_2}^{(2)}}^{(2)}\} T_j \rightarrow t \tag{11.2}$$

as $N \rightarrow \infty$ in probability for each $t > 0$. It follows that the sum in (11.1) scaled by \tilde{c}_N converges in distribution as $N \rightarrow \infty$ to $\hat{\gamma}_1(x_1)T$, where $\hat{\gamma}_1(\cdot) := \frac{1}{p}\gamma_1(\cdot)$. Given also that the transition among top first level configurations is asymptotically uniform, we find that the asymptotic motion among the top first level configurations is consistent with a (uniform) K process. We can state the following result.

Theorem 11.1 (Intermediate temperatures). *If $\beta \in (\beta_1^{cr}, \beta_2^{cr}) \cap (0, \beta^{FT})$, then*

$$X_1^N(\cdot/\tilde{c}^N) \Rightarrow \tilde{X}_1(\cdot) \tag{11.3}$$

as $N \rightarrow \infty$, where $\tilde{X}_1 \sim \mathcal{K}(\hat{f}, 1)$ starting at ∞ , with $\hat{f} : \mathbb{N} \rightarrow (0, \infty)$, $\hat{f}(x_1) = \hat{\gamma}_1(x_1)$.

We have indeed given all of the main ingredients of the proof above, except for an estimate establishing that $X_1^N(\cdot/\tilde{c}^N)$ spends virtually all of its time on the first level top configurations. This can be done as in the proofs of the Theorems 2.4, 2.5 and 2.7, by a first moment estimate. For that to work, we resort to a further known result, namely that the first level marginal of the Gibbs measure converges to the normalization of \hat{f} . This is given by Theorem 9.2 of [8]; indeed that result is stated for intermediate temperatures for the case where $p = 1/2$, but one may readily check that it holds in general.

In order that the conditions on β above are not empty, we need of course $\beta^{FT} > \beta_1^{cr}$, which is equivalent to $p < 1/3$; in this case, it may or may not happen that $\beta^{FT} < \beta_2^{cr}$; in the former case, (11.3) holds only in a (nonempty) subinterval of $(\beta_1^{cr}, \beta_2^{cr})$, namely $(\beta_1^{cr}, \beta^{FT})$; otherwise, it holds in the full intermediate interval.

If $\beta^{FT} < \beta < \beta_2^{cr}$, then Theorem 1.10 in [25] tells us that the sum on the left hand side of (11.2), when properly rescaled, converges weakly to a stable subordinator instead. This signals aging, and thus the time scale is not extreme. This clarifies a point raised in Subsubsection 1.2.2.

Under the conditions of the above theorem, the following aging result readily follows in the same way as in the previous section

$$\lim_{\substack{t_w, t \rightarrow 0 \\ t/t_w \rightarrow \theta}} \tilde{\Pi}_1(t_w, t_w + t) = \text{Asl}_{\alpha_1}(1/(1 + \theta)), \tag{11.4}$$

where $\tilde{\Pi}_1(t_w, t_w + t)$ is the probability that \tilde{X}_1 gives no jump within $(t_w, t_w + t)$.

A Appendix

A.1 K-processes

Let \mathcal{D} be a countably infinite set, and let ∞ denote a point not in \mathcal{D} , and make $\bar{\mathcal{D}} = \mathcal{D} \cup \{\infty\}$. Let $f, \mathfrak{w} : \mathcal{D} \rightarrow (0, \infty)$ be such that

$$\sum_{x \in \mathcal{D}} \mathfrak{w}(x) = \infty, \quad \sum_{x \in \mathcal{D}} \mathfrak{w}(x)f(x) < \infty. \tag{A.1}$$

Consider $\{\mathcal{C}_x, x \in \mathcal{D}\}$, an independent family of Poisson counting processes such that \mathcal{C}_x has intensity \mathfrak{w}_x for each $x \in \mathcal{D}$, with associated point processes $\mathcal{S} = \{(\theta_x(i), i \geq 1), x \in \mathcal{D}\}$ (the *event times* of the respective counting processes). Let $\omega : \mathbb{R}^+ \rightarrow \bar{\mathcal{D}}$ be such that $\omega(\theta_x(i)) = x$ for $x \in \mathcal{D}, i \geq 1$, and $\omega(s) = \infty$ if $s \notin \mathcal{S}$. We note that ω is well

defined almost surely. Let $\{T_s, s \in \mathbb{R}^+\}$ be an iid family of mean 1 exponential random variables. Let now ν be an atomic measure on \mathbb{R}^+ concentrated on \mathcal{S} as follows

$$\nu(\{s\}) = f(\omega(s)) T_s, s \in \mathcal{S}, \tag{A.2}$$

and let Γ be its distribution function, namely, $\Gamma : \mathbb{R}^+ \rightarrow \mathbb{R}^+$ is such that $\Gamma(r) = \nu([0, r])$, and let φ be the right continuous inverse of Γ . Then for $t \geq 0$ let

$$X(t) := \omega(\varphi(t)). \tag{A.3}$$

We call X thus defined a K-process on $\bar{\mathcal{D}}$ with *waiting time* function f , and *weight* function ω , starting at ∞ . Notation: $X \sim \mathcal{K}(f, \omega)$. In the particular case where $\omega \equiv 1$, we call X a *uniform* K-process on $\bar{\mathcal{D}}$ with waiting time function f , and use the notation $X \sim \mathcal{K}(f, 1)$. Also, we call Γ the clock process of X .

We next define 2-level K-processes, as follows. Let \hat{X} be a *uniform* K-process on $\bar{\mathcal{D}}$ with $(\omega \equiv 1)$ and f as above such that (A.1) is satisfied. Let \mathcal{D}' be a countably infinite set and as before make $\bar{\mathcal{D}}' = \mathcal{D}' \cup \{\infty\}$. Let $f' : \mathcal{D} \times \mathcal{D}' \rightarrow (0, \infty)$ be such that

$$\sum_{xy \in \mathcal{D} \times \mathcal{D}'} f(x)f'(xy) < \infty. \tag{A.4}$$

Let $\{C'_x, x \in \mathcal{D}'\}$ be an iid family of intensity 1 Poisson counting processes, independent of $\{C_x, x \in \mathcal{D}\}$, with associated point processes $S' = \{(\theta'_x(i), i \geq 1), x \in \mathcal{D}'\}$.

Let $\omega' : \mathbb{R}^+ \rightarrow \bar{\mathcal{D}}'$ be such that $\omega'(\theta'_x(i)) = x$ for $x \in \mathcal{D}'$, $i \geq 1$, and $\omega'(s) = \infty$ if $s \notin S'$. We note that ω' is well defined almost surely.

Let ν' be an atomic measure on \mathbb{R}^+ concentrated on S' as follows.

$$\nu'(\{s\}) = f'(\hat{X}(s)\omega'(s)) T_s, s \in S', \tag{A.5}$$

and let Γ' be its distribution function, and let φ' be the right continuous inverse of Γ' . Then for $t \geq 0$ let

$$X(t) = X_1(t)X_2(t) := \hat{X}(\varphi'(t))\omega'(\varphi'(t)). \tag{A.6}$$

We call $X = (X(t), t \geq 0)$ a 2-level K-process (starting at (∞, ∞)) on $\bar{\mathcal{D}} \times \bar{\mathcal{D}}'$ with waiting time functions f, f' . Notation: $X \sim \mathcal{K}_2(f, f')$. This process was introduced in [18] (with $\mathcal{D} = \mathcal{D}' = \mathbb{N} = \{1, 2, \dots\}$), where some of its properties and those of a finite volume version were studied (and where illustrations of their construction can be found). It may be understood as two 1-level uniform K-processes arranged in hierarchies as follows.

Given the realization of the (1-level) K-process \hat{X} , let \mathcal{I} be the set of maximal intervals of constancy of \hat{X} (maximal time intervals where \hat{X} is constant), the second step of the above description amounts to constructing within each such interval, say $[a, b)$, a 1-level uniform K-process with waiting time function $f'(x, \cdot)$, where x is the constant value of \hat{X} within that interval. This results in what can be seen as an excursion of a 2-level K-process $X = X_1X_2$ with $X_1 \equiv x$. This excursion takes place within the time interval $[a', b')$, with $a' = \Gamma'(a)$, $b' = \Gamma'(b)$. Outside the union of all such intervals, $X \equiv (\infty, \infty)$.

We call Γ' the clock processes of X . We also call $\Gamma_1 := \hat{\Gamma} \circ \Gamma'$ the clock process of X_1 , where $\hat{\Gamma}$ is the clock process of \hat{X} .

A.2 Auxiliary results for the proof of Proposition 3.3

Let $r = p/(1 - p)$. Let C_N be the event that for all $x_1 \in \mathcal{M}_1$ and $x'_1 \in \mathcal{V}_{N_1}$ such that $d_1(\xi_1^{x'_1}, \xi_1^{x_1}) \leq \epsilon_0 N_1$ we have that

$$\#\{x''_1 \in \mathcal{D}_1 : \xi_1^{x''_1} \sim \xi_1^{x'_1} \text{ and } q_N^*(\xi_1^{x''_1}) \leq re^{-\sqrt{N}}\} \geq \epsilon_1 N_1, \tag{A.7}$$

Lemma A.1. *There exists $\epsilon_0, \epsilon_1 > 0$ such that $\mathcal{P}(C_N) \rightarrow 1$ as $N \rightarrow \infty$.*

Proof. Since $H_1(y_1) = \sqrt{aN}\Xi_{y_1}$, and $\{\Xi_{y_1}, y_1 \in \mathcal{D}_1\}$ are iid standard Gaussian random variables, we have that the left hand side of (A.7) above dominates a binomial random variable with N_1 trials and probability of success $\Phi(-1/(\beta\sqrt{a}))$ in each trial, where Φ is the standard Gaussian distribution function. Therefore, by standard large deviation bounds, there exists $\epsilon_1 > 0$ such that the probability of the complement of (A.7) may be bounded above by $c_1 2^{-\epsilon_1 N_1}$ for some constant c_1 , and $\epsilon'_1 > 0$. Now the probability of C_N^c may be bounded above by

$$c_1 M_1(\epsilon_0 N_1 + 1) \sqrt{N_1} \binom{N_1}{\epsilon_0 N_1} 2^{-\epsilon'_1 N_1} \leq c_0 N_1 2^{-(\epsilon'_1 - \epsilon'_0) N_1},$$

for some constant c_0 , and $\epsilon'_0 = \epsilon'_0(\epsilon_0) > 0$ such that $\epsilon'_0 \rightarrow 0$ as $\epsilon_0 \rightarrow 0$; the result follows by choosing $\epsilon_0, \epsilon_1 > 0$ such that $0 < \epsilon'_0 < \epsilon'_1$. \square

Let us fix $\epsilon, \epsilon_0, \epsilon_1 > 0$ satisfying the conditions of Lemma 4.5 and Lemma A.1, and such that $\epsilon_0 < \epsilon$. Let B_N be the event that for all $\eta, \bar{\eta} \in T_1, \eta \neq \bar{\eta}$, (4.22) holds.

Lemma A.2. *Let \mathfrak{J}_2 be the number of jumps of \bar{X}_2^N till \bar{X}_1^N reaches $\xi_1^{-1}(T_1)$. Given $\sigma_1 \in \mathcal{V}_{N_1}$ such that $d_1(\sigma_1, T_1) > \epsilon_0 N_1$, then provided B_N and C_N occur, and \bar{X}_1^N starts from $\xi_1^{-1}(\sigma_1)$, we have that $\mathbb{P}_{\sigma_1}(\mathfrak{J}_2 \leq N^3)$ vanishes exponentially fast as $N \rightarrow \infty$, i.e., there exists $R > 1$ such that*

$$\mathbb{P}_{\sigma_1}(\mathfrak{J}_2 < N^3) \leq R^{-N}. \tag{A.8}$$

Proof. Let \mathfrak{J}_1 be the number of jumps of \bar{X}_1^N till \bar{X}_1^N reaches $\xi_1^{-1}(T_1)$. Then

$$\mathfrak{J}_2 \stackrel{d}{=} \sum_{j=0}^{\mathfrak{J}_1-1} \mathcal{G}_j(J_{N_1}^\circ(j)) \geq \sum_{j=0}^{\mathfrak{J}_1-1} \mathcal{G}_j(J_{N_1}^\circ(j)) \mathbf{1}_{\{d_1(J_{N_1}^\circ(j), T_1) \leq \epsilon_0 N_1\}}, \tag{A.9}$$

where $J_{N_1}^\circ$ is the jump chain of \bar{X}_1^N , namely, a simple symmetric discrete time random walk on \mathcal{V}_{N_1} , and $\mathcal{G}_j(\sigma'_1), j \geq 0, \sigma'_1 \in \mathcal{V}_{N_1}$, are independent geometric random variables with success parameter $q'_N := re^{-\sqrt{N}} \wedge 1$. The right-hand side of (A.9) may be bounded stochastically from below as follows. Notice that since B_N occurs, the chain $d_1(J_{N_1}^\circ(\cdot), T_1)$ observed only when $J_{N_1}^\circ$ is at a distance at most $\epsilon_0 N_1$ from T_1 may be identified (in distribution) to a Markov chain \mathcal{Z} in $\{0, 1, \dots, \epsilon_0 N\}$ which has the same transition probabilities as $\mathcal{Y} := d'(J_N^\circ(\cdot), O)$, where J_N° is a simple symmetric random walk on a hypercube of dimension N , and O is a given site of such a hypercube, except that \mathcal{Z} is lazy at $\epsilon_0 N$ – the jumps of \mathcal{Y} to the right of $\epsilon_0 N$ are replaced by self jumps at $\epsilon_0 N$ of \mathcal{Z} . Also, in this case, \mathcal{Z} starts from $\epsilon_0 N$. Since C_N occurs, every time $J_{N_1}^\circ$ gives a jump within distance at most $\epsilon_0 N$ from T_1 , independent of all else, there is an at least ϵ_1 probability it will land on a site $\sigma'_1 \in \mathcal{V}_{N_1}$ such that $q_N^*(\sigma'_1) \leq q'_N$. The above justifies bounding stochastically the right-hand side of (A.9) from below by

$$\mathfrak{J}'_2 := \sum_{j=0}^{\mathfrak{J}'_1-1} \mathcal{G}'_j \geq N^3 \sum_{j=0}^{\mathfrak{J}'_1-1} \mathbf{1}_{\{\mathcal{G}'_j \geq N^3\}} \geq N^3 \sum_{j=0}^{\epsilon_0 N-1} \mathbf{1}_{\{\mathcal{G}'_j \geq N^3\}}, \tag{A.10}$$

where \mathfrak{J}'_1 is the number of jumps \mathcal{Y} takes to reach 0 starting from $\epsilon_0 N$, and $\mathcal{G}'_j, j = 0, 1, \dots$, are iid random variables, independent of \mathfrak{J}'_1 ; \mathcal{G}'_0 is a mixture of two random variables, the first of which, with weight $1 - \epsilon_1$ is 0, and the other, with weight ϵ_1 , is geometric with success parameter q'_N . Now, $P(\mathcal{G}'_0 \geq N^3) = \epsilon_0(1 - q'_N)^{N^3} \geq \epsilon''_0 > 0$ for all large enough N , and (A.8) follows with $R = (1 - \epsilon''_0)^{-\epsilon_0}$. \square

Remark A.3. From Theorem 3.1 of [6] starting from any $\sigma_i \in \mathcal{V}_{N_i}$, the probability that \bar{X}_i^N does not go a distance $\frac{1}{2}N_i$ from σ_i before it returns to σ_i is bounded above by $1/N + 4/N^2$.

Lemma A.4. For \mathcal{I} as in (6.36) we have that $\lim_{R \rightarrow \infty} \limsup_{N \rightarrow \infty} \mathbb{P}(\mathcal{I} \geq Rc_1^N 2^{N_2}) \rightarrow 0$.

Proof. We may stochastically bound \mathcal{I} from below by a geometric random variable with success parameter $\max_{x_1 \in \mathcal{M}_1, \sigma \in W^{x_1}} \mathbb{P}_\sigma(\mathcal{G}^* > \Theta^*)$, where \mathcal{G}^* is a geometric random variable with success parameter $q_N^*(\xi_1^{x_1})$, independent of Θ^* , and $\Theta^* = \tau_{\{\sigma'_2 \in \pi_2 T^{x_1} : d_2(\sigma'_2, \sigma_2) = N_2\}}$. Notice that the distribution of \mathcal{G}^* only depends on σ through x_1 , and that the distribution of Θ^* is independent of σ . Therefore we can drop the maximum over $\sigma \in W^{x_1}$ in the above formula. Furthermore the maximum over x_1 is achieved at $x_1 = M_1$. We then have

$$\max_{x_1 \in \mathcal{M}_1} \mathbb{P}(\mathcal{G}^* > \Theta^*) = \mathbb{E}^{\circ,2}[(1 - q_N^*(\xi_1^{M_1}))^{\Theta^*}] =: \rho_N \geq \frac{\psi_N \gamma_1^N(\xi_1^{M_1})}{1 + \psi_N \gamma_1^N(\xi_1^{M_1})} (1 + o(1)), \quad (\text{A.11})$$

where the inequality follows from (5.5) and (5.12) (with $A = \{\xi_1^{x_1}\}$ and where in (5.6), $d_2 = 1$ and F_{N_2, d_2} is absent). The probability in the statement of Lemma A.4 is bounded above by $(1 - \rho_N)^{Rc_1^N 2^{N_2}}$. Since $\rho_N \rightarrow 0$ as $N \rightarrow \infty$ and $\rho_N c_1^N 2^{N_2} \gtrsim \text{const} \frac{N_2}{N_1} \gamma_1(M_1)$, the result follows. \square

Lemma A.5. (6.38) holds.

Proof of Lemma A.5. We start by computing $\pi(x)$. Let \mathcal{G} denote the number of jumps of \bar{X}_2^N before \bar{X}_1^N leaves x . \mathcal{G} is a geometric random variable with success parameter $q_N^*(\xi_1^x)$, independent of τ_{σ_2} . Then, setting $\lambda'_N = \frac{N_1}{2} \frac{c_1^N}{\gamma_1^N(x)}$, and applying (4.13, 4.14), we get

$$\begin{aligned} \pi(x) &= \frac{1}{2^{N_2}} \sum_{\sigma'_2 \in \mathcal{V}_{N_2}} \mathbb{P}_{\sigma'_2}(\mathcal{G} \geq \tau_{\sigma_2}) = \frac{1}{2^{N_2}} \sum_{\sigma'_2 \in \mathcal{V}_{N_2}} \mathbb{E}_{\sigma'_2}^{\circ,2}[(1 - q_N^*(\xi_1^x))^{\tau_{\sigma_2}}] \\ &= \frac{1}{B_0(\lambda'_N)} \frac{1}{2^{N_2}} \sum_{i=0}^{N_2} \sum_{\sigma'_2 \in \mathcal{V}_{N_2} : d(\sigma_2, \sigma'_2) = i} B_i(\lambda'_N) = \frac{1}{B_0(\lambda'_N)} \frac{1}{2^{N_2}} \sum_{i=0}^{N_2} \binom{N_2}{i} B_i(\lambda'_N) \\ &= \frac{1}{B_0(\lambda'_N)} \int_0^1 u^{\lambda'_N - 1} du \frac{1}{2^{N_2}} \sum_{i=0}^{N_2} \binom{N_2}{i} (1 - u)^i (1 + u)^{N_2 - i} \\ &= \frac{1}{B_0(\lambda'_N)} \int_0^1 u^{\lambda'_N - 1} du = \frac{1}{\lambda'_N B_0(\lambda'_N)} = \frac{1}{1 + \lambda'_N \sum_{i=1}^{N_2} \binom{N_2}{i} \frac{1}{i + \lambda'_N}}, \end{aligned}$$

and the first claim of (6.38) follows upon noticing that the sum in the denominator on the right-hand side above is $\sim 2^{N_2+1}/N_2$.

As for the second claim, we write $\hat{\pi}(x_1) = \mathbb{P}(\cup_{x_2 \in \mathcal{M}_2} H_{x_2} | \bar{X}_1^N(0) = x_1)$, where H_{x_2} is the event that \bar{X}_2^N hits x_2 before the first jump of \bar{X}_1^N . By the Bonferroni inequalities, we have that

$$0 \leq \sum_{x_2 \in \mathcal{M}_2} \mathbb{P}(H_{x_2} | \bar{X}_1^N(0) = x_1) - \hat{\pi}(x_1) \leq \sum_{x_2 \neq x'_2 \in \mathcal{M}_2} \mathbb{P}(H_{x_2} \cap H_{x'_2} | \bar{X}_1^N(0) = x_1). \quad (\text{A.12})$$

Since the summands on the central expression above are identically equal to $\pi(x_1)$, and the expression on the right-hand side equals $\pi(x_1) \sum_{x_2 \neq x'_2 \in \mathcal{M}_2} \mathbb{P}(H_{x_2} | \bar{X}_1^N(0) = x_1, H_{x_2})$, it

is enough to argue that each summand in the expression above is an $o(1)$. But, given Lemma 4.5 above, each such summand is, apart from an $o(1)$ error, the probability that, starting from the origin, an Ehrenfest chain on $\{0, \dots, N_2\}$ passes by bN_2 , with $b = 1/3$,

before an independent time which is geometrically distributed with success probability $q_N^*(x_1)$. Writing that probability as a moment generating function as above (see e.g. the first equality in (A.11)), and applying (4.13, 4.14), we have that that equals

$$\frac{1}{B_0(\lambda'_N)} \sum_{j=0}^{\bar{b}N_2} \binom{\bar{b}N_2}{j} \frac{\Gamma(bN_2 + 1)\Gamma(j + \lambda'_N)}{\Gamma(bN_2 + 1 + j + \lambda'_N)}, \quad (\text{A.13})$$

where $\bar{b} = 1 - b$. The quotient inside the latter sum is bounded above by 1, and thus (A.13) is bounded above by

$$\frac{1}{\lambda'_N B_0(\lambda'_N)} + \frac{1}{B_0(\lambda'_N)} \sum_{j=1}^{\bar{b}N_2} \binom{\bar{b}N_2}{j}. \quad (\text{A.14})$$

As we saw above the first term of this sum is $\sim \gamma_1(x_1) \frac{N_2}{N_1} \frac{1}{c_1^{N_2/2}}$, which is an $o(1)$. The second term is readily checked to also be an $o(1)$, and the claim is established. \square

References

- [1] M. Baity-Jesi, G. Biroli, and C. Cammarota, *Activated aging dynamics and effective trap model description in the random energy model*, J. Stat. Mech. Theory Exp. (2018), no. 1, 013301, 25. MR-3761614
- [2] Gérard Ben Arous, Anton Bovier, and Jiří Černý, *Universality of the REM for dynamics of mean-field spin glasses*, Comm. Math. Phys. **282** (2008), no. 3, 663–695. MR-2426140
- [3] Gérard Ben Arous, Anton Bovier, and Véronique Gayraud, *Glauber dynamics of the random energy model. I. Metastable motion on the extreme states*, Commun. Math. Phys. **235** (2003), no. 3, 379–425. MR-1974509
- [4] Gérard Ben Arous, Anton Bovier, and Véronique Gayraud, *Glauber dynamics of the random energy model. II. Aging below the critical temperature*, Commun. Math. Phys. **236** (2003), no. 1, 1–54. MR-1977880
- [5] Gérard Ben Arous and Jiří Černý, *The arcsine law as a universal aging scheme for trap models*, Comm. Pure Appl. Math. **61** (2008), no. 3, 289–329. MR-2376843
- [6] Gérard Ben Arous and Véronique Gayraud, *Elementary potential theory on the hypercube*, Electron. J. Probab. **13** (2008), no. 59, 1726–1807. MR-2448129
- [7] S. C. Bezerra, L. R. G. Fontes, R. J. Gava, V. Gayraud, and P. Mathieu, *Scaling limits and aging for asymmetric trap models on the complete graph and K processes*, ALEA Lat. Am. J. Probab. Math. Stat. **9** (2012), no. 2, 303–321. MR-3069367
- [8] Erwin Bolthausen and Alain-Sol Sznitman, *Ten lectures on random media*, DMV Seminar, vol. 32, Birkhäuser Verlag, Basel, 2002. MR-1890289
- [9] J.-P. Bouchaud, L. Cugliandolo, J. Kurchan, and M. Mézard, *Out of equilibrium dynamics in spin-glasses and other glassy systems*, Spin glasses and random fields (A. P. Young, ed.), World Scientific, Singapore, 1998. MR-1488891
- [10] J.-P. Bouchaud and D. S. Dean, *Aging on Parisi’s tree*, J. Phys I (France) **5** (1995), 265.
- [11] Anton Bovier, Michael Eckhoff, Véronique Gayraud, and Markus Klein, *Metastability in stochastic dynamics of disordered mean-field models*, Probab. Theory Related Fields **119** (2001), no. 1, 99–161. MR-1813041
- [12] Anton Bovier, Michael Eckhoff, Véronique Gayraud, and Markus Klein, *Metastability and low lying spectra in reversible Markov chains*, Comm. Math. Phys. **228** (2002), no. 2, 219–255. MR-1911735
- [13] Anton Bovier and Véronique Gayraud, *Convergence of clock processes in random environments and ageing in the p -spin SK model*, Ann. Probab. **41** (2013), no. 2, 817–847. MR-3077527
- [14] Anton Bovier and Irina Kurkova, *Derrida’s generalised random energy models. I. Models with finitely many hierarchies*, Ann. Inst. H. Poincaré Probab. Statist. **40** (2004), no. 4, 439–480. MR-2070334

- [15] D. Capocaccia, M. Cassandro, and P. Picco, *On the existence of thermodynamics for the generalized random energy model*, J. Statist. Phys. **46** (1987), no. 3-4, 493–505. MR-0883541
- [16] Jiří Černý and Véronique Gayrard, *Hitting time of large subsets of the hypercube*, Random Structures Algorithms **33** (2008), no. 2, 252–267. MR-2436848
- [17] L. R. Fontes and G. R. C. Peixoto, *Elementary results on K processes with weights*, Markov Process. Related Fields **19** (2013), no. 2, 343–370. MR-3113947
- [18] L. R. G. Fontes, R. J. Gava, and V. Gayrard, *The K -process on a tree as a scaling limit of the GREM-like trap model*, Ann. Appl. Probab. **24** (2014), no. 2, 857–897. MR-3178499
- [19] L. R. G. Fontes, M. Isopi, Y. Kohayakawa, and P. Picco, *The spectral gap of the REM under Metropolis dynamics*, Ann. Appl. Probab. **8** (1998), no. 3, 917–943. MR-1627811
- [20] L. R. G. Fontes and P. H. S. Lima, *Convergence of symmetric trap models in the hypercube*, New Trends in Mathematical Physics (Dordrecht) (Vladas Sidoravičius, ed.), Springer Netherlands, 2009, pp. 285–297.
- [21] R. J. Gava, *Scaling limit of the trap model on a tree*, Ph.D. thesis, University of São Paulo (2011 [in portuguese]). Url: teses.usp.br/teses/disponiveis/45/45133/tde-24092014-224507/publico/TeseGava.pdf
- [22] V. Gayrard, *Thermodynamic limit of the q -state Potts-Hopfield model with infinitely many patterns*, J. Statist. Phys. **68** (1992), no. 5-6, 977–1011. MR-1180270
- [23] V. Gayrard and O. Gün, *Aging in the GREM-like trap model*, Markov Process. Related Fields **22** (2016), no. 1, 165–202. MR-3523983
- [24] Véronique Gayrard, *Aging in Metropolis dynamics of the REM: a proof*, Probab. Theory Related Fields **174** (2019), no. 1-2, 501–551. MR-3947330
- [25] Véronique Gayrard and Lisa Hartung, *Dynamic phase diagram of the REM*, Statistical Mechanics of Classical and Disordered Systems (Gayrard V., L.-P. Arguin, N. Kistler, and I. Kurkova, eds.), Springer Proceedings in Mathematics & Statistics, Springer International Publishing, 2019, pp. 111–170, arXiv:1008.3849. MR-3947330
- [26] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1985. MR-0832183
- [27] J. H. B. Kemperman, *The passage problem for a stationary Markov chain*, Statistical Research Monographs, Vol. I, The University of Chicago Press, Chicago, Ill., 1961. MR-0119226
- [28] J. R. Norris, *Markov chains*, Cambridge Series in Statistical and Probabilistic Mathematics, vol. 2, Cambridge University Press, Cambridge, 1998, Reprint of 1997 original. MR-1600720
- [29] M. Sasaki and K. Nemoto, *Analysis on aging in the generalized random energy model*, J. Phys. Soc. Jpn. **69** (2000), 3045–3050.
- [30] E. Seneta, *Nonnegative matrices and Markov chains*, second ed., Springer Series in Statistics, Springer-Verlag, New York, 1981. MR-0719544
- [31] G. W. Stewart and Ji Guang Sun, *Matrix perturbation theory*, Computer Science and Scientific Computing, Academic Press, Inc., Boston, MA, 1990. MR-1061154