



**HAL**  
open science

# The disordered lattice free field pinning model approaching criticality

Giambattista Giacomin, Hubert Lacoin

► **To cite this version:**

Giambattista Giacomin, Hubert Lacoin. The disordered lattice free field pinning model approaching criticality. *Annals of Probability*, 2022, 50 (4), pp.1478-1537. 10.1214/22-aop1566 . hal-03911973

**HAL Id: hal-03911973**

**<https://hal.science/hal-03911973v1>**

Submitted on 23 Dec 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THE DISORDERED LATTICE FREE FIELD PINNING MODEL APPROACHING CRITICALITY

BY GIAMBATTISTA GIACOMIN AND HUBERT LACOIN

We continue the study, initiated in [25], of the localization transition of a lattice free field  $\phi = (\phi(x))_{x \in \mathbb{Z}^d}$ ,  $d \geq 3$ , in presence of a quenched disordered substrate. The presence of the substrate affects the interface at the spatial sites in which the interface height is close to zero. This corresponds to the Hamiltonian

$$\sum_{x \in \mathbb{Z}^d} (\beta \omega_x + h) \delta_x,$$

where  $\delta_x = \mathbf{1}_{[-1,1]}(\phi(x))$ , and  $(\omega_x)_{x \in \mathbb{Z}^d}$  is an IID centered field. A transition takes place when the average pinning potential  $h$  goes past a threshold  $h_c(\beta)$ : from a delocalized phase  $h < h_c(\beta)$ , where the field is macroscopically repelled by the substrate, to a localized one  $h > h_c(\beta)$  where the field sticks to the substrate. In [25] the critical value of  $h$  is identified and it coincides, up to the sign, with the log-Laplace transform of  $\omega = \omega_x$ , that is  $-h_c(\beta) = \lambda(\beta) := \log \mathbb{E}[e^{\beta \omega}]$ . Here we obtain the sharp critical behavior of the free energy approaching criticality:

$$\lim_{u \searrow 0} \frac{\mathbb{F}(\beta, h_c(\beta) + u)}{u^2} = \frac{1}{2 \operatorname{Var}(e^{\beta \omega - \lambda(\beta)})}.$$

Moreover, we give a precise description of the trajectories of the field in the same regime: to leading order as  $h \searrow h_c(\beta)$  the absolute value of the field is  $\sqrt{2\sigma_d^2 |\log(h - h_c(\beta))|}$  except on a vanishing fraction of sites ( $\sigma_d^2$  is the single site variance of the free field).

## 1. Introduction.

1.1. *Disorder and critical phenomena: an overview, till the free field pinning case.* A fundamental issue in statistical mechanics is the effect of *disorder*, synonymous of *random environment* and (with more old fashioned language) of *impurities*, on phase transitions. The issue is very general and applies to any statistical model that exhibits a phase transition: in mathematical terms, a phase transition happens at a given value, called critical, of a parameter (the temperature, an external field,...) when one or more observables on the system have a singular, i.e. non-analytic, behavior in the

---

*MSC 2010 subject classifications:* 60K35, 60K37, 82B27, 82B44

*Keywords and phrases:* Lattice Free Field, Disordered Pinning Model, Localization Transition, Critical Behavior, Disorder Relevance, Multiscale Analysis

parameter we are considering, at the critical value. The behavior of systems approaching criticality is particularly interesting because of the appearance of *universal behaviors* that are, to a certain extent, characterized in terms of *critical exponents* (e.g. [20, 21]). Consider now a *spatially homogeneous* model, i. e. a model in which the interactions are translation invariant, that has a phase transition. If we modify the interactions by perturbing them in a spatially random way, we obtain, for every realization of the randomness, a different non homogeneous model (that we call *disordered*): does the phase transition survive to the introduction of this randomness? And, if it does, is the nature of the transition affected? That is, are the critical exponents the same as in the homogeneous case?

In spite of the general nature of the problem, phase transitions and critical phenomena are under control only for particular homogeneous models, or classes of homogeneous models. The most famous one, and first (nontrivial) one to be solved (in 1944, by Lars Onsager) is the two dimensional Ising model (on square lattice, with ferromagnetic interactions and in absence of external field): this model has been at the heart of the activity of a large community of researchers since. A part of this community, mostly on the physical side, focused on the issue that interests us, that is whether Onsager's results withstand the introduction of disorder, for example a small amount of disorder. And it is precisely in the Ising model context that A. B. Harris [28] took an approach to this question that turned out to be very successful in the physical community. Harris' approach is based on the renormalization group and can be summed up (in a vague but hopefully evocative fashion) by saying that one has to consider what is the effect of the renormalization transformation on the disorder when the system is close to criticality. If the renormalization tends to suppress the disorder then one expects that on large scale the disorder will be *irrelevant*, and the critical phenomenon will not be affected by the disorder. On the other hand, if disorder is enhanced by the renormalization group transformation, one generically expects that the critical behavior is affected by the disorder, that is therefore dubbed *relevant*. The success of Harris' arguments is in part due to the fact that he was able to make them boil down to a very simple criterion, called *Harris criterion*.

In spite of the fact that these ideas are around since at least 45 years and that they are commonly applied in physics, from the mathematical viewpoint the understanding of the Harris criterion is very limited, notably for the original example of the two dimensional Ising model (see [24, Ch. 5] for a review). Only more recently (in a sequence of works including [1, 2, 4, 15, 11, 27, 29], we refer to [24] for a more thorough bibliography) the

Harris criterion prediction has been proven in full for a class of statistical mechanics models: the pinning models.

Keeping at a very informal level, pinning models can be visualized as *interface* pinning models. An *effective*  $d+1$  dimensional interface is modeled by considering a random function from  $\mathbb{Z}^d$  to  $\mathbb{R}$  (or to  $\mathbb{Z}$ ): examples include the Lattice Free Field (LFF) or other *gradient* fields like the massless fields or the Solid On Solid (SOS) models (see e.g. [21, Ch.8] for an introduction to the LFF or [22, 36, 38] for more advanced material). In  $d = 1$  these interface models just reduce to random walk models. The pinning potential is a reward that is introduced via an energy term (we are taking a Gibbsian viewpoint of the probability law of the model) that rewards or penalizes the visit to level zero (if the interface takes values in  $\mathbb{Z}$ ) or a neighborhood of level 0 (if the interface takes values in  $\mathbb{R}$ ): we call these visits *contacts*. The intensity of the reward is parametrized by a variable  $h$ , and it can become a penalization if we change the sign of  $h$ . In the disordered case we simply make  $h$  depend on  $x$ : the parametrization we choose is  $h + \beta\omega_x$ , where  $(\omega_x)_{x \in \mathbb{Z}^d}$  are IID centered random variables (with suitable integrability properties), and  $\beta \geq 0$ .

As already understood in [19], the  $d = 1$  model has an intrinsic independence structure that allows in particular a generalization of the model that turns out to be very important in the Harris criterion perspective: in mathematical terms the model can be rewritten (for every value of  $d$ ) just in terms of the point process represented by the location of the contacts and if  $d = 1$  the contact set is just a renewal process (if the interface takes values in  $\mathbb{Z}$ , otherwise it is a Markov renewal process)[23, 24]. This is not only precious in solving the model – notably, the  $d = 1$  homogeneous case is exactly solvable – but it offers an immediate natural generalization to the large class of renewal pinning models. So for  $d = 1$ , in the generalized context we just hinted to, one can obtain models for which Harris criterion predicts disorder irrelevance and other ones for which it predicts disorder relevance. We refer to [23, 24] for the large literature on  $1 + 1$  dimensional pinning and renewal pinning. But we want to stress that if the irrelevant disorder results are very satisfactory (and they are proven exactly when Harris criterion predicts irrelevance), relevant disorder results are much weaker. This is not surprising: Harris criterion does not bear information about what the critical behavior is, if disorder is relevant. Nevertheless, it has been shown that, when the Harris criterion predicts disorder relevance, the critical behavior is not the same as the one of the homogeneous model. What the disordered critical behavior really is remains mathematically a fully open issue. Substantial progress on this problem has been recently achieved, but not for the pinning

models itself: the critical behavior of a relevant disorder case for one class of *copolymer* pinning models and of a simplified version of the hierarchical pinning model have been identified respectively in [3] and in [12].

The  $d > 1$  case is a priori more difficult to handle and, above all, the contact set does not enjoy the independence (renewal) structure of the  $d = 1$  case: it is replaced by a more geometric *spatial* Markov property. The problem has been attacked in [14, 25, 30] by using the LFF as interface model: the homogeneous model turns out to be solvable (or, at least, has a certain degree of solvability) and, as far as the questions we raise, in a rather elementary way. In particular, it displays a delocalization to localization transition at a critical value  $h_c$ , as  $h$  grows. It is also rather straightforward to see that this transition survives when disorder is introduced, that is for  $\beta > 0$ . A peculiar feature of this transition is that  $h_c$ , or  $h_c(\beta)$  when  $\beta > 0$ , separates the regime  $h < h_c(\beta)$  in which the contact fraction is zero (delocalized regime), and the regime  $h > h_c(\beta)$  in which the contact fraction is positive (localized regime).

What is instead much less obvious [25, 30] is the identification of the critical value  $h_c(\beta)$ , along with estimates on the contact fraction of the system that show that disorder is relevant in all dimensions  $d \geq 2$ . In particular, for  $d \geq 3$ , the case on which we focus here, we have proven in [25] that the contact fraction approaching criticality, i.e.  $h \searrow h_c(\beta)$ , is bounded above and below by  $h - h_c(\beta)$  times a positive constant (different for lower and upper bound). This result has been established only for Gaussian disorder, while for more general disorder a lower bound of  $(h - h_c(\beta))^c$ ,  $c > 1$  is a constant that depends on  $d$ . Therefore the contact fraction is (Lipschitz) continuous for  $\beta > 0$  and this is sufficient to infer that disorder is relevant. In fact for  $\beta = 0$  the contact fraction is discontinuous at the critical value.

The content of the work we present now is:

1. showing that, when  $h \searrow h_c(\beta)$ , the contact fraction behaves like  $h - h_c(\beta)$  times a constant that depends on  $\beta$  and on the law of the disorder, on which we make only integrability assumptions;
2. providing precise path estimates in the same limit. That is, describing the trajectories on which the system concentrates near criticality.

Precise contact fraction estimates like the ones in point (1) typically demand at least a heuristic understanding of the path behavior of point (2). Therefore in our context they demand a good understanding of the *localization mechanism* for the disordered system near criticality. This is one of the main achievements of our analysis.

On the other hand, the step from (1) to (2) is by no mean evident and, as

a matter of fact, it is technically the most demanding part of our analysis, involving in particular a full multiscale analysis.

1.2. *The model's building blocks: Lattice Free Field and disorder.* We set  $\Lambda_N := \llbracket 0, N \rrbracket^d = \{0, 1, \dots, N\}^N$ ,  $d \geq 3$ ,  $N \in \mathbb{N}$ , and consider the centered free field on this set. That is, we consider a Gaussian family of centered random variables  $(\phi(x))_{x \in \mathbb{Z}^d}$ , whose law is denoted by  $\mathbf{P}_N$ , with  $\mathbf{E}_N[\phi(x)\phi(y)] := G_N(x, y)$  where  $G_N(\cdot, \cdot)$  is the Green function associated with the simple symmetric random walk on  $\mathbb{Z}^d$  killed upon exiting  $\mathring{\Lambda}_N := \llbracket 1, N-1 \rrbracket^d$ . More explicitly, if  $P_x$  denotes the distribution law of a simple symmetric continuous time random walk  $(S_t)_{t \geq 0}$  with jump rate one in each direction and initial condition  $S_0 = x$ , then

$$(1.1) \quad G_N(x, y) = E_x \left[ \int_0^{\tau_N} \mathbf{1}_{\{S_t=y\}} dt \right], \quad \text{with} \quad \tau_N := \inf \left\{ t > 0 : S_t \notin \mathring{\Lambda}_N \right\}.$$

It is well known that the Green function  $G(\cdot, \cdot)$  of the simple random walk without killing (obtained by replacing  $\tau_N$  in (1.1) by  $\infty$ ) exists for  $d \geq 3$ . We let  $\mathbf{P}$  denote the law of the Gaussian field on  $\mathbb{Z}^d$  with covariance  $G(\cdot, \cdot)$ . We also set  $\sigma_d^2 := G(0, 0)$ .

Let us recall from now some well known random walk estimates in transient dimensions that we will repeatedly use. First of all  $\lim_{x \rightarrow \infty} |x|^{d-2} G(0, x)$  exists and it is positive, so in particular we can find  $C_d > 1$  such that for every  $x \neq 0$

$$(1.2) \quad C_d^{-1} |x|^{2-d} \leq G(0, x) \leq C_d |x|^{2-d}.$$

Moreover if  $\liminf_{N \rightarrow \infty} \text{dist}(z_N, \Lambda_N^c)/N > 0$  (e.g.  $z_N := \lceil N/2 \rceil (1, 1, \dots, 1)$ )

$$(1.3) \quad 0 \leq G(0, 0) - G_N(z_N, z_N) = O\left(1/N^{d-2}\right),$$

and, always aiming at comparing the Green function and its killed version, we have if  $\lim_{N \rightarrow \infty} \text{dist}(z_N, \Lambda_N^c) = \infty$  for every  $x, y \in \mathbb{Z}^d$

$$(1.4) \quad G(x, y) = G(0, y - x) = \lim_{N \rightarrow \infty} G_N(x + z_N, y + z_N).$$

Of course, for  $x \in \mathbb{Z}^d$  and  $A \subset \mathbb{Z}^d$ ,  $\text{dist}(x, A) := \min_{y \in A} |x - y|$  and we make the choice (irrelevant in most of the cases, but of some importance for some geometric constructions) that  $|\cdot|$  denotes the  $\ell_1$  distance in  $\mathbb{Z}^d$ , that is  $|x| = \sum_{j=1}^d |x_j|$  (but for  $x \in \mathbb{R}^d$  we use  $|x|$  for the Euclidean norm).

The disorder, or random environment,  $(\omega_x)_{x \in \mathbb{Z}^d}$  is a family of IID random variables,  $\mathbb{P}$  is its law. Free field and disorder are independent. We assume that

$$(1.5) \quad \lambda(s) := \log \mathbb{E}[\exp(s\omega)] < \infty \quad \text{for every } s \in \mathbb{R},$$

and that  $\lambda'(\cdot)$  is not a constant, i. e.  $\omega$  is not a constant. In (1.5) we have dropped the index  $x$  for obvious reasons. Without loss of generality we assume  $\mathbb{E}[\omega] = 0$ : this is largely irrelevant because  $\omega$  appears in the model in the form  $\beta\omega - \lambda(\beta)$  which is invariant under the transformation  $\omega \mapsto \omega + \text{constant}$ . However,  $\mathbb{E}[\omega] = 0$  assures that  $\lambda(\cdot)$  is increasing on the positive semi-axis and decreasing in the negative one, and this is practical.

The generalization of the results to the case in which we assume (1.5) only, say, for  $|s|$  smaller than a constant is not straightforward. The full hypothesis (1.5) is used for a cut-off estimate, see Section A.2, that is probably not necessary but it does not appear to be easy to circumvent. On the other hand, a part of the main results (notably, the probability upper bound) can be obtained under very mild hypothesis on the lower (negative) tail of  $\omega$ , in particular for this results (1.5) is exploited only for  $s > 0$ . For sake of readability we will make precise this aspect only in the technical part of the work (see Remark 2.1).

1.3. *The disordered lattice free field pinning model.* The model we consider is the disordered pinning model based on the LFF with law  $\mathbf{P}_N$ . For  $\beta \geq 0$  and  $h \in \mathbb{R}$  we set

$$(1.6) \quad \frac{d\mathbf{P}_{N,\omega,\beta,h}}{d\mathbf{P}_N}(\phi) := \frac{1}{Z_{N,\omega,\beta,h}} e^{\sum_{x \in \Lambda_N} (\beta\omega_x - \lambda(\beta) + h)\delta_x}, \quad \text{with } \delta_x := \mathbf{1}_{[-1,1]}(\phi(x)),$$

where of course  $Z_{N,\omega,\beta,h}$  is the normalization constant (or *partition function*)

$$(1.7) \quad Z_{N,\omega,\beta,h} := \mathbf{E}_N \left[ \exp \left( \sum_{x \in \Lambda_N} (\beta\omega_x - \lambda(\beta) + h)\delta_x \right) \right].$$

In the proofs, i.e. starting from Section 2, we will use as far as possible the concise notation  $\omega_x(\beta, h) := \beta\omega_x - \lambda(\beta) + h$ . We will often use also

$$(1.8) \quad \mathcal{F}_A^\phi := \sigma(\phi(x) : x \in A).$$

Note that  $\mathbf{P}_{N,\omega,\beta,h}$  does not change if we replace summing over  $x \in \Lambda_N$  with  $x \in \mathring{\Lambda}_N$  or any other set  $\Lambda$ ,  $\mathring{\Lambda}_N \subset \Lambda \subset \Lambda_N$ .  $Z_{N,\omega,\beta,h}$  is affected by such a

change, but only in a trivial way and, in particular, the *free energy density* (that we will simply call *free energy*)

$$(1.9) \quad \mathbb{F}(\beta, h) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \mathbb{E} \log Z_{N, \omega, \beta, h},$$

is clearly not affected either. The proof of the existence of the limit in (1.9) can be found in [14]: the argument is based on the almost super-additive behavior of the sequence  $(\mathbb{E} \log Z_{N, \omega, \beta, h})_{N \in \mathbb{N}}$ . We will come back to this issue (see Proposition 3.2) because a sharp super-additive behavior for a modified partition function, that gives rise to the same free energy, is going to be important, notably for the lower bound on the free energy in Section 3. Here are some basic, but crucial, properties of the free energy (see the introduction of [25] for full details):

- The map  $(\beta, h) \mapsto \mathbb{F}(\beta, h)$  is convex, moreover it is non decreasing in  $h \in \mathbb{R}$  for  $\beta$  fixed and in  $\beta \geq 0$  for  $h$  fixed;
- The inequality  $\mathbb{F}(\beta, h) \geq 0$  holds because of the rough *entropic repulsion* estimate  $\log \mathbf{P}_N(\phi(x) > 1 \text{ for every } x \in \Lambda_N) = o(|\Lambda_N|)$  which is easily derived by exploiting the continuum symmetry of the LFF [35];
- By Jensen's inequality, we have  $\mathbb{F}(\beta, h) \leq \mathbb{F}(0, h)$  (*annealed bound*) .

The convexity and monotonicity properties in  $h$  lead to identifying

$$(1.10) \quad h_c(\beta) := \inf\{h : \mathbb{F}(\beta, h) > 0\} = \inf\{h : \partial_h \mathbb{F}(\beta, h) > 0\},$$

as a critical point, provided that  $h_c(\beta) \neq \pm\infty$ . Elementary estimates lead to excluding  $h_c(\beta) \neq \pm\infty$ , but much more than that is true: in [25] is shown that  $h_c(\beta) = 0$  for every  $\beta \geq 0$ . Again we refer to the introduction of [25] for full details, but we stress that establishing  $h_c(\beta) \geq 0$  is a rather straightforward consequence of comparison with the model without disorder: of course  $\mathbb{F}(0, h) = 0$  for  $h \leq 0$  and a very moderate amount of work leads to

$$(1.11) \quad \mathbb{F}(0, h) \stackrel{h \searrow 0}{\sim} c_d h, \quad \text{with } c_d := \mathbf{P}(\phi(0) \in [-1, 1]) = P(\sigma_d \mathcal{N} \in [-1, 1]),$$

and  $\mathcal{N}$  is our notation for a standard Gaussian variable. In particular (1.11) yields  $h_c(0) = 0$  and, in turn, from the annealed bound we obtain  $h_c(\beta) \geq h_c(0)$ . Remark also that from the annealed bound we extract  $\mathbb{F}(\beta, h) \leq ch$ , for any  $c > c_d$  and  $h > 0$  small. However this result is poor precisely because disorder is relevant for this model: the main result in [25] is that if  $\omega_x$ 's are standard Gaussian variables then for  $h \in (0, 1)$

$$(1.12) \quad C_-(\beta)h^2 \leq \mathbb{F}(\beta, h) \leq C_+(\beta)h^2,$$



with  $C_{\pm}(\beta) > 0$  satisfying  $\lim_{\beta \searrow 0} \beta^2 C_{\pm}(\beta) = c_{\pm} > 0$ , and  $c_- < c_+$ . We remark that, exploiting the convexity of  $F(\beta, \cdot)$  and the fact that  $\partial_h \log Z_{N,\omega,\beta,h} = \mathbf{E}_{N,\omega,\beta,h}[\sum_{x \in \Lambda_N} \delta_x]$ , one readily obtains that the infinite volume contact density  $\rho(\beta, h) := \lim_N \rho_N(\beta, h)$ , with  $\rho_N(\beta, h) := \mathbf{E}_{N,\omega,\beta,h} \left[ \sum_{x \in \Lambda_N} \delta_x / |\Lambda_N| \right]$ , exists except possibly for countably many values of  $h$ : the non decreasing function  $\rho(\beta, \cdot)$  may have jumps and, when it does, its value is not well defined at the jump. In order to avoid this nuisance we extend the definition of  $\rho(\beta, h)$  choosing the right continuous version of  $\rho(\beta, \cdot)$  (the results that follow are exactly the same for the left continuous version). From (1.12) one easily obtains

$$(1.13) \quad C_-(\beta)h \leq \rho(\beta, h) \leq 4C_+(\beta)h,$$

for every  $h \in (0, 1)$  for the lower bound and for every  $h \in (0, 1/2)$  for the upper bound, see Remark 1.3.

Also results for general disorder are given in [25], but they are rougher than (1.12), in the sense that, in the lower bound in (1.12),  $h^2$  is replaced by  $h$  to a large power. Nevertheless, the results we just cited show that disorder is relevant for the model we consider: the critical behavior changes when disorder is switched on.

1.4. *The main results.* The first main result of this paper is a sharp version of (1.12), valid for general disorder distribution. We prove that  $F(\beta, h)$  is asymptotically proportional to  $h^2$  when  $h \searrow 0$  and identify the value of the constant in front of  $h^2$ .

THEOREM 1.1. *For every  $\beta > 0$  we have that*

$$(1.14) \quad F(\beta, h) \stackrel{h \searrow 0}{\sim} \chi(\beta)h^2,$$

with

$$(1.15) \quad \chi(\beta) := \frac{1}{2\text{Var}(e^{\beta\omega_x - \lambda(\beta)})} = \frac{1}{2(e^{\lambda(2\beta) - 2\lambda(\beta)} - 1)}.$$

Note that  $\chi(\beta) \sim 1/(2\beta^2)$  for  $\beta \searrow 0$ , showing thus that the smoothing effect weakens for small disorder. Theorem 1.1 sums up quantitative upper and lower bounds (Propositions 2.2 and 3.1 respectively) that bear more information on the rate of convergence of (1.14). Moreover one easily extracts

from (1.14) the following asymptotic equivalence on the contact fraction (see Remark 1.3)

$$(1.16) \quad \rho(\beta, h) \stackrel{h \searrow 0}{\sim} 2\chi(\beta)h.$$

Of course (1.16) gives already a precise information on the behavior of the trajectories with law  $\mathbf{P}_{N,\omega,\beta,h}$  in the infinite volume limit and near criticality. Our second main result, Theorem 1.2 below, goes much farther in this direction.

Recall, cf. the end of Section 1.2, that  $\sigma_d$  denotes the standard deviation of the one site marginal of the infinite volume LFF. The following result shows that asymptotically most of the points in the field are located around height  $\sigma_d\sqrt{2\log(1/h)}$ .

**THEOREM 1.2.** *Given  $\varepsilon > 0$ , there exists  $h_0(\varepsilon) > 0$  such that for every  $h \in (0, h_0(\varepsilon))$  we can find  $c := c(\varepsilon, h) > 0$  and  $N_0 := N_0(\omega, \varepsilon, h)$ , with  $\mathbb{P}(N_0 < \infty) = 1$ , such that  $\mathbb{P}(d\omega)$ -a.s. we have for  $N > N_0$*

$$(1.17) \quad \mathbf{P}_{N,\omega,\beta,h} \left( \left| \left\{ x \in \Lambda_N : \left| |\phi(x)|(\log(1/h))^{-1/2} - \sqrt{2}\sigma_d \right| > \varepsilon \right\} \right| \geq \varepsilon N^d \right) \leq e^{-cN^d}.$$

This result considerably refines previous estimates obtained on the trajectories. In [25] it was only proved that typically the field is on most of the sites above height  $c_d\sqrt{\log(1/h)}$  (in absolute value) for an explicit non-optimal constant  $c_d$ .

**REMARK 1.3.** *The arguments to go from (1.12) to (1.13) and from (1.14) to (1.16) are standard, but we sketch them here. If  $f : [0, \infty) \rightarrow [0, \infty)$  is a convex increasing function such that  $C_1h^2 \leq f(h) \leq C_2h^2$  for  $h \in [0, h_0]$ , with  $C_1 > 0$ , then  $\partial f(h) \geq f(h)/h \geq C_1h$ . Here  $\partial f(\cdot)$  is either the upper or the lower differential of  $f$ . Moreover  $\partial f(h) \leq (f(2h) - f(h))/h \leq 4C_2h$ , if  $h \in [0, h_0/2]$ . If we have instead  $f(h) \sim Ch^2$  for  $h \searrow 0$  and  $C > 0$ , then  $\partial f(h) \sim 2Ch$ . In fact, for every  $\delta \in (0, 1)$  we have  $C_1h^2 \leq f(h) \leq C_2h^2$  with  $C_1/C_2 = 1 - \delta^2$ , for  $h$  sufficiently small: of course  $C \in [C_1, C_2]$ . Therefore  $\partial f(h) \leq (f((1 + \delta)h) - f(h))/(\delta h) \leq 2C_2h(1 + \delta) \leq 2Ch(1 + \delta)/(1 - \delta^2)$ . For the opposite bound we use the incremental ratio between  $(1 - \delta)h$  and  $h$  and we obtain  $\partial f(h) \geq 2Ch(1 - \delta)$ .*

1.5. *Discussion of the results, relevant literature and organization of the paper.*

1.5.1. *Localization strategy and sketch of proofs.* A key point and one of the main novelty of our work is that we identify the *localization mechanism*. And this mechanism is crucially suggested by the argument for the upper bound on the free energy. The upper bound we give is a universal bound: it holds for an arbitrary random contact set. More explicitly: the partition function (1.7) depends only on the random field  $(\delta_x)_{x \in \mathring{\Lambda}_N}$  or, equivalently, on the random set  $\{x \in \mathring{\Lambda}_N : \delta_x = 1\}$ , and we give an upper bound on the free energy density not only for  $\delta_x = \mathbf{1}_{[-1,1]}(\phi(x))$  with  $\phi$  the LFF on  $\Lambda_N$  with zero boundary conditions, but for an arbitrary law of  $(\delta_x)_{x \in \mathring{\Lambda}_N}$ . Moreover this bound is saturated by choosing  $(\delta_x)_{x \in \mathring{\Lambda}_N}$  IID Bernoulli variables  $\text{Ber}(p)$  of a parameter  $p$  chosen to be the value  $p(\beta, h)$  that maximizes the function (1.18)

$$p \mapsto \mathbb{E} \log \mathbf{E} \left[ e^{(\beta\omega - \lambda(\beta) + h)\text{Ber}(p)} \right] = \mathbb{E} \log (1 + p (\exp(\beta\omega - \lambda(\beta) + h) - 1)) ,$$

and we stress that this expression is just the free energy density with the  $\delta$  contact random variables replaced by IID  $\text{Ber}(p)$  variables. In the limit  $h \searrow 0$  we obtain that  $p(\beta, h) \sim 2\chi(\beta)h$ : this follows from a rather elementary computation detailed in the proof of Proposition 2.2. Why should the contact set that we obtain from the LFF be close to saturating this bound too? At a heuristic level the reason is a combination of two well known facts on the LFF:

1. the continuum symmetry of the LFF that makes *rigid* vertical translations of the interface little expensive (this is of course very much in the logic of the *entropic repulsion phenomena* [6, 7, 16, 17, 35])
2. large excursions of the LFF have very mild correlations (an issue underlying [7, 17] and developed in detail in [13]).

This suggests the following behavior for the field when  $h > 0$  is small: the field shifts away from level 0, precisely it shifts to a level  $u(\beta, h)$ , or  $-u(\beta, h)$ , so that the probability that  $\phi(x) \sim \mathcal{N}(\sigma_d^2, \pm u(\beta, h))$  belongs to  $[-1, 1]$  is equal to  $p(\beta, h)$ . As the reader might expect in view of Theorem 1.2, it turns out that  $u(\beta, h)$  is asymptotically equivalent to the square root of  $2\sigma_d^2 \log(1/h)$ . The details on the Gaussian computation leading to this shift are a bit subtle (see Lemma 3.4). This is because what we want is that  $\mathbf{P}(\mathcal{N}(u(\beta, h), \delta_d^2) \in [-1, 1]) \sim 2\chi(\beta)h$ , but  $u(\beta, h) \sim 2\sigma_d^2 \log(1/h)$  yields the poor result  $\mathbf{P}(\mathcal{N}(u(\beta, h), \delta_d^2) \in [-1, 1]) = h^{2+o(1)}$ . So a much sharper choice of  $u(\beta, h)$ , actually called  $u(2\chi(\beta), h)$  in Section 3, is required. See also Remark 3.5 for more on this.

To substantiate this localization strategy we need to provide a lower bound. This is achieved by considering the field with boundary conditions  $u(\beta, h)$  – the value of the free energy does not depend on on this choice – and via a two step decomposition of the LFF that is in the spirit of several earlier works: in three or more dimension the LFF can be written as the superposition of a field with small variance, and spatially power law decaying covariance, and a field that accounts for almost all the variance of the original field, but with exponentially decaying covariance: it is for example the case of the decomposition of the field theory literature [9, 20] in which  $G = G_\epsilon + (G - G_\epsilon)$  where  $G_\epsilon$  is the Green function of a walk with a rate of death  $\epsilon > 0$ , see [17, Sec. 4] for a probabilistic presentation. We propose instead a decomposition that is much more geometrically structured: we write  $\phi$  as a power law correlated field (with small variance) plus independent fields that are compactly supported over boxes (i.e. hypercubes). The boxes have edge length proportional to  $h^{-c}$ ,  $c > 0$  a small constant, and they overlap only near the boundary. Recalling that the boundary of the LFF is set to height  $u(\beta, h)$ , hence the mean of the field is  $u(\beta, h)$ , in each one of these boxes we typically expect no contact, because the contact density is proportional to  $h$  and the volume of each box is  $h^{-cd}$  (and  $c$  is small, in particular  $c < 1/d$ ). On this scale we are able to perform an accurate analysis that shows that the leading contribution to the free energy in each of these boxes is given by configurations with a number of contact that is bounded by a constant  $\kappa$  that does not depend on  $h$ . This accurate analysis is an implementation of the constrained second moment method which exploits the fact that the leading contribution comes from essentially independent variables (see Sec. 3.3.5). The errors introduced by the power law correlated field (with small variance) and by the overlap regions turn out to be higher order corrections.

The next step is proving that the trajectories of the field behave like what is suggested by the asymptotic of the free energy and its proof. This is a matter of proving upper and lower bounds on the height of the field with law  $\mathbf{P}_{N,\omega,\beta,h}$ , that is

$$(1.19)$$

$$(A) \quad \mathbf{P}_{N,\omega,\beta,h} \left( \left| \left\{ x \in \Lambda_N : \frac{|\phi(x)|}{\sqrt{\log(1/h)}} < \sqrt{2}\sigma_d - \varepsilon \right\} \right| \geq \varepsilon N^d \right) \leq e^{-cN^d},$$

$$(B) \quad \mathbf{P}_{N,\omega,\beta,h} \left( \left| \left\{ x \in \Lambda_N : \frac{|\phi(x)|}{\sqrt{\log(1/h)}} > \sqrt{2}\sigma_d + \varepsilon \right\} \right| \geq \varepsilon N^d \right) \leq e^{-cN^d}.$$

Showing (A), that is that the field shifts, except possibly for  $\varepsilon|\Lambda_N|$  sites,

to a height of at least  $(1 - \varepsilon)(2\sigma_d^2 \log(1/h))^{1/2}$  is not too difficult. In fact, again by a two scale argument (but rather standard, in the spirit of the arguments in [7, 17] and already exploited in [25]) one establishes that if one partitions  $\Lambda_N$  in boxes of edge length  $L$  (large but fixed), in most of these boxes there is a point in which the field is close to the correct height  $(2\sigma_d^2 \log(1/h))^{1/2}$ . This is incompatible with having a density of sites on which the field is below  $(1 - \varepsilon)(2\sigma_d^2 \log(1/h))^{1/2}$ , because it forces a density of sites  $x$  to have a neighbor  $y$  with  $|\phi(x) - \phi(y)| \geq c(\log(1/h))^{1/2}$  ( $c > 0$  small but not depending on  $h$ ). And this is highly penalized by the LFF Hamiltonian.

The lower bound on the trajectory we just outlined is relatively short and it is just a refinement of the argument in [25]. This is because being too close to the pinning region, i.e. level zero, is penalized in an obvious way (since on the average the potential is  $-\lambda(\beta) + h$ , and  $h$  tends to zero). On the other hand, showing (B) in (1.19), that is that the field shifts, except possibly for  $\varepsilon|\Lambda_N|$  sites, to a height of at most  $(1 + \varepsilon)(2\sigma_d^2 \log(1/h))^{1/2}$  is substantially harder and requires novel arguments. Indeed the fact that the field is too far from the pinning region in a small (but positive) density of sites says that it cannot collect the expected amount of rewards on those sites, but this does not exclude, at least not in an obvious way, that these rewards are collected elsewhere. After all, only rare spikes hit the pinning potential region with the strategy we have outlined. This estimate therefore has to exploit a more collective behavior of the field and the keyword at this stage is certainly *rigidity of the interface*. But, in practice, implementing a proof along the reasoning that we just sketched is not straightforward and the control from above of the trajectories comes via two non trivial and technically demanding estimates:

1. a control of the contact fraction on *mesoscopic* scales, notably down to the boxes of volumes that are just a bit larger than  $h^{-2}$ : we stress that these boxes become large when  $h \searrow 0$ , but they are of constant size with respect to  $N$ ;
2. an estimate of the rigidity of the field that demands a full multiscale analysis.

1.5.2. *Open problems: sign of  $\phi$  and disordered induced symmetry breaking.* An obvious question at this stage is: what about the sign of the field? It is natural to conjecture that for small values of  $h$  most sites are located on the same side of the interface. On the other hand, we believe that  $N^{-d} \sum_{x \in \Lambda_N} \mathbf{1}_{\{\phi_x > 0\}}$  converges to  $1/2$  for sufficiently large values of  $h$ , since, in that regime, most sites are favorable to contact. This corresponds to the

following convergence in law (conjectural) statement on  $\mathbf{P}_{N,\omega,\beta,h}$  for  $h > 0$ :

$$(1.20) \quad \frac{1}{N^d} \sum_{x \in \Lambda_N} \mathbf{1}_{\{\phi_x > 0\}} \xrightarrow{N \rightarrow \infty} \rho_\beta(h) \text{Ber}(1/2) + (1 - \rho_\beta(h))(1 - \text{Ber}(1/2)),$$

where  $\rho_\beta(h) \in [1/2, 1)$  and approaches 1 as  $h \searrow 0$ , while  $\rho_\beta(h) = 1/2$  when  $h$  is above a threshold. According to this conjecture, the interface lies above level zero in a majority of sites if  $\text{Ber}(1/2) = 1$  and  $\rho_\beta(h) > 1/2$ :  $\rho_\beta(h)$  is precisely the density of these sites. In this case there is therefore a density  $1 - \rho_\beta(h)$  of sites below level zero and, by symmetry, if  $\text{Ber}(1/2) = 0$ , i.e. the interface lies below level zero in a majority of sites, the density of sites above zero is  $1 - \rho_\beta(h)$ . This phenomenon disappears when, for  $h$  above a threshold,  $\rho_\beta(h)$  becomes  $1/2$ , and the right-hand side of (1.20) becomes equal to  $1/2$  too.

As we already pointed out, the value  $1/2$  for the parameter of the limiting Bernoulli variable comes from symmetry. We believe that this probability for the field to be *mostly positive* is very sensitive to boundary condition: if we replace the centered LFF  $\phi$  that defines the model with  $\phi + c$ , any  $c > 0$ , we expect (1.20) to become

$$(1.21) \quad \lim_{N \rightarrow \infty} \frac{1}{N^d} \sum_{x \in \Lambda_N} \mathbf{1}_{\{\phi_x > 0\}} = \rho_\beta(h),$$

in  $\mathbf{P}_{N,\omega,\beta,h}$ -probability.

Obtaining a proof of (1.20) and/or (1.21) appears to be very challenging. Nevertheless, sidetracking farther, we observe that they suggest the following consideration concerning Gibbs states for the disordered model. It seems reasonable to expect that for positive values of  $h$  there is a unique translation invariant Gibbs state associated with the homogenous model (we warn the reader that already this step is speculative and it represents in itself a challenging conjecture). On the other hand, (1.20)-(1.21) indicate that for the disordered model there are at least two Gibbs states: one corresponding the limit obtained with positive boundary condition (i.e.,  $c > 0$ ) and another one corresponding to the limit with negative boundary condition.

1.5.3. *Comparison with another interface repulsion phenomenon.* The disorder-induced repulsion phenomenon highlighted in the present paper bears some analogy with the entropic repulsion phenomenon observed in the low temperature SOS model constrained to remain positive and recently studied in detail by one of the authors [31, 32]. The introduction of disorder has in fact effects that are very similar to those induced by imposing a

positivity constraint to the SOS model: the phase transition is smoothed, it vanishes like of  $(h - h_c)^\nu$  with  $\nu \geq 2$  approaching the critical point (as opposed to linearly for the model without constraint), and the interface is repelled to a distance from level zero that diverges in this limit.

While the specific mechanisms that triggers these phenomena are different for the two models, two common ingredients can be identified. Firstly, in both models, the contact set is well approximated (at least at a heuristic level) by an IID Bernoulli field. Secondly, the optimal value for the contact fraction  $p$  is obtained by optimizing the balance between a reward which is proportional to  $(h - h_c)p$  and a penalty term which takes the form  $p^\mu$  for some  $\mu > 1$ . Then one can easily conclude that the optimal balance between penalties occurs for a contact fraction of order  $(h - h_c)^{\frac{1}{\mu-1}}$ , which therefore yields a critical exponent for the free energy equal to  $\frac{\mu}{\mu-1}$ . We have  $\mu = 2$  for the disordered pinning and  $\mu \in (1, 2)$  for SOS (the specific value depends on the lattice, it is equal to  $3/2$  on  $\mathbb{Z}^2$ ).

One important difference between the two models lies in the origin of the penalty term. In the disordered model we study here, this penalty term is produced by the second order term in the Taylor expansion of (1.18). We have

$$(1.22) \quad \mathbb{E} [\log (1 + p (\exp (\beta \omega - \lambda(\beta) + h) - 1))] = \\ ph - \frac{p^2}{2} e^{2h} \text{Var} (\exp (\beta \omega - \lambda(\beta))) + O(p^3).$$

This quadratic terms in  $p$  indicates by how much Jensen's inequality fails to be an equality, in other words it quantifies by how much the disorder can fail to self average for a fixed contact fraction: recall (observation right after (1.18)) that  $p$  becomes asymptotically proportional to  $h$  when  $h \searrow 0$ , so the first two terms in the right-hand side of (1.22) are competing with each other. For SOS, the penalty term comes from a rewriting of the model, which transforms the wall constraint into a shift of  $h$  and an additional penalty for pairs of neighboring contact points, and a similar *first versus second order* competition arises.

REMARK 1.4. *Note that the LFF pinning model in presence of a hard wall (studied in [8, 26]) presents a different phenomenology, since the critical exponent changes from 1 to  $\infty$  when the hard wall constraint is introduced. Informally the reason why this happens is that in that case the penalty term induced by the hard wall constraint is of the form  $p\sqrt{|\log p|}$  instead of  $p^\mu$ , which results in a much smaller optimal value for  $p$ .*

REMARK 1.5. *After this work was complete one of the authors investigated the problem of SOS pinning, in presence of disorder and without wall [33], focusing on the free energy behavior. The results present similarities with Theorem 1.1, especially in terms of the critical exponent associated with the phase transition: the introduction of disorder changes the phase transition behavior from first order to quadratic. However discretization effects of SOS make the free energy curve for disordered SOS less regular (at least in asymptotic terms) than the one for the LFF pinning model: it displays angular points corresponding to the discrete transition of the localization height. On the technical side, it may be worth noting that while the harder part in Theorem 1.1 is the proof of the lower bound, the opposite is true for [33]. The proof of the lower bound in [33, Theorem 2.5] is made easier by the application of contour arguments which are specific to discrete models, while the upper bound relies on an intricate coarse-graining procedure.*

1.5.4. *Entropic repulsion and critical disordered pinning* . Entropic repulsion models (like [7, 16, 17, 35] for LFF and [10] for SOS) have already entered the discussion and there is of course more than a flavor of a connection between our results and entropic repulsion phenomena. There is however the substantial difference that the repulsion phenomenon we observe is to a height that is finite, and diverges only approaching the critical point. We believe that a direct connection between the wall repulsion studied in [7, 16, 17, 35] can be made with the critical disordered pinning model: we quickly develop this next, just keeping at a heuristic level.

The main question is: what is the typical value  $u(N)$  for  $|\phi(x)|$ , for  $x$  in the bulk of the box, at the critical point, that is, under the measure  $\mathbf{P}_{N,\omega,\beta,0}$ . Clearly Theorem 1.2 indicates that  $\lim_n u(N) = \infty$  and, at an intuitive level, the pinning *strength*  $\beta\omega_x - \lambda(\beta)$  is negative, i.e. repulsive, on average, but of course there is a density of sites that are attracting the interface. Let us recall the mechanism which induces entropic repulsion for the the measure  $\mathbf{P}_N(\cdot | \phi(x) \geq 0 \text{ for every } x \in \Lambda_N)$ . We let  $u_{\text{wall}}(N)$  denote the typical bulk height under this measure (note that a softer potential would lead to a similar heuristics). The height  $u_{\text{wall}}(N)$  can be understood as the one that balances the two penalties  $N^{d-2}u$  that accounts for shifting the field away from level zero in the bulk and  $N^d \exp(-u_N^2/(2\sigma_d^2))$  that comes from the penalization coming from hitting the penalized (or forbidden) region: if we want to minimize the sum of these two quantities, we find that  $u_{\text{wall}}(N)$  has to be asymptotically equivalent to the square root of  $4\sigma_d^2 \log N$  (see [7, 16, 17]). In the critical disordered model the penalization coming from the potential is weaker: it is proportional to  $N^d(\exp(-u_N^2/(2\sigma_d^2)))^2$ , as it is



strongly suggested by the leading  $p^2$  in the Taylor expansion of (1.18) at  $h = 0$  (see (1.22)). This leads to a (conjectured) repulsion  $u(N)^2 \sim 2\sigma_d^2 \log N$  for the critical disordered pinning model.

On the other hand, we believe that the square root of  $4\sigma_d^2 \log N$  corresponds to the typical height for negative values of  $h$  both in the homogeneous and disordered case. We mention in relation to this problem the disordered entropic repulsion model studied in [5]: for the repulsion mechanism of [5], quenched and annealed models have, to leading order, the same behavior.

**1.5.5. Organization of the paper.** In Section 2 and 3, we prove quantitative upper bound and lower bounds for the free energy respectively. The proof of Theorem 1.2 is also split into two sections: Section 4 for (A) in (1.19) and Section 5 for (B) in (1.19).

The four sections – upper and lower bounds on the free energy, lower and upper bound on the height of the field – are almost completely independent from the technical viewpoint.

**Acknowledgements:** This work has been performed in part while G.G. was visiting IMPA with the support of the Franco-Brazilian network in mathematics. G.G. also acknowledges support from grant ANR-15-CE40-0020. H.L. acknowledges support from a productivity grant from CNPq and a Jovem Cientista do Nosso Estado grant from FAPERJ.

**2. Proof of Theorem 1.1: Upper bound on the free energy.** Let us first present the quantitative bound proved in this section. As anticipated in the introduction, for the results in this section we can sensibly weaken the assumptions on  $\omega$ .

**REMARK 2.1.** *The main result of this section, that is Proposition 2.2, holds (and it is proven) assuming less than (1.5). More precisely for Proposition 2.2 we only require (1.5) for  $s > 0$  on the upper tail of the disorder. For the lower tail we assume that  $\mathbb{E}[\omega_-] < \infty$ , with the notation  $\omega_- = -\omega \mathbf{1}_{\{\omega < 0\}}$  for the negative part. An analogous standard notation for the positive part is used below. Like before, we keep the convention that  $\mathbb{E}[\omega] = 0$ .*

**PROPOSITION 2.2.** *Recall that  $\chi(\beta)$  is given (1.15). For every  $\beta > 0$  there exists a constant  $C_\beta$  such that for every  $h \in [0, 1]$*

$$(2.1) \quad \mathbb{F}(\beta, h) \leq \chi(\beta)h^2 + C_\beta h^3.$$

To achieve this bound, we do not rely at all on the fact that  $(\delta_x)_{x \in \Lambda_N}$  is the contact set of the LFF. We instead prove a general statement which says that the averaged log partition function associated to any point process  $(\delta_x)_{x \in \Lambda_N}$  – or any family of Bernoulli random variables (with arbitrary parameters and correlations) – is always smaller than the one obtained when the  $(\delta_x)_{x \in \Lambda_N}$  are IID Bernoulli variables with an optimal density.

**PROPOSITION 2.3.** *Consider  $\Lambda$  a finite non empty set,  $(\xi_x)_{x \in \Lambda}$  a field of IID random variables satisfying  $\mathbb{P}(\xi \geq -1) = 1$  and  $\mathbb{E}[(\log(1 + \xi))_+] < \infty$ . Moreover we assume that  $\mathbf{P}_\Lambda$  is the probability distribution of an arbitrary random vector  $(\delta_x)_{x \in \Lambda}$  on  $\{0, 1\}^\Lambda$ . Then we have*

$$(2.2) \quad \mathbb{E} \log \mathbf{E}_\Lambda \left[ \prod_{\{x : \delta_x=1\}} (1 + \xi_x) \right] \leq |\Lambda| \max_{p \in [0,1]} \mathbb{E} [\log(1 + p\xi)] ,$$

with the convention  $\prod_{\{x \in \emptyset\}} (1 + \xi_x) = 1$ .

Proposition 2.2 then follows from Proposition 2.3 by solving the corresponding optimization problem. This is what we do first.

**PROOF OF PROPOSITION 2.2.** Recall that we use  $\omega(\beta, h)$  for  $\beta\omega - \lambda(\beta)$ . As a consequence of (2.2) we have that for  $\xi_x := e^{\omega_x(\beta, h)} - 1$

$$(2.3) \quad \mathbb{E} \log Z_{N, \omega, \beta, h} \leq |\Lambda_N| \max_{p \in [0,1]} \mathbb{E} [\log(1 + p\xi)] ,$$

and thus for every  $\beta > 0$  and every  $h$  we have

$$(2.4) \quad \mathbb{F}(\beta, h) \leq \max_{p \in [0,1]} \mathbb{E} \left[ \log \left( 1 + p \left( e^{\omega(\beta, h)} - 1 \right) \right) \right] .$$

We are now going to expand the right-hand side for  $h \searrow 0$ , and everything we are going to require is that  $\lambda(3\beta) < \infty$ . We use the fact that the maximum is achieved for some  $p$  that we call  $p_h$  (which is unique as shown in the proof of Proposition 2.3 although it is not needed in the argument here). First of all remark that  $\lim_{h \searrow 0} p_h = 0$ . Indeed if we set  $p_0 = \limsup_{h \searrow 0} p_h$ , using dominated convergence (here  $\mathbb{E}[\omega_-] < \infty$  is exploited) and the definition of  $p_h$ , we have, for some positive sequence  $h_n$  tending to 0

$$(2.5) \quad \lim_{n \rightarrow \infty} \max_{p \in [0,1]} \mathbb{E} \left[ \log \left( 1 + p \left( e^{\omega(\beta, h_n)} - 1 \right) \right) \right] = \mathbb{E} [\log(1 + p_0(e^{\beta\omega - \lambda(\beta)} - 1))] .$$

The right-hand side is non-negative (set  $p = 0$  in the right-hand side) while by the (strict) Jensen's inequality the left-hand side is strictly negative if  $p_0 > 0$  since  $\beta > 0$  and the distribution of  $\omega$  is non degenerate. Therefore  $p_0 = 0$ .

To conclude, we analyse the asymptotic behavior  $\mathbb{E}[\log(1 + p(e^{\omega(\beta, h)} - 1))]$  in the limit when  $p, h \searrow 0$  via Taylor expansion. This will allow to determine the asymptotics for  $p_h$  and subsequently also that of r.h.s of (2.4). We use the elementary bound

$$(2.6) \quad -(x_-)^3 \leq \log(1 + x) - x + \frac{1}{2}x^2 \leq \frac{1}{3}x^3,$$

where the upper bounds holds for every  $x > -1$ , while the lower bound holds for  $x > -0.8156\dots$ . Since we have assumed that  $\lambda(3\beta) < \infty$  we obtain that

$$(2.7) \quad \mathbb{E} \left[ \log \left( 1 + p \left( e^{\omega(\beta, h)} - 1 \right) \right) \right] - p(e^h - 1) + \frac{p^2}{2} \left( e^{\lambda(2\beta) - 2\lambda(\beta) + 2h} - 2e^h + 1 \right),$$

is  $O(p^3)$  and this readily entails ( $\chi(\beta)$  is given in (1.15))

$$(2.8) \quad \mathbb{E} \left[ \log \left( 1 + p \left( e^{\omega(\beta, h)} - 1 \right) \right) \right] = ph - \frac{p^2}{4\chi(\beta)} + O(p^3) + O(p^2h) + O(ph^2).$$

We stress that an expression (that depends on  $p, h$  and  $\beta$ ) is  $O(p)$  (or  $O(h)$ , etc. . .) means that there exists a constant  $C_\beta$  such that its absolute value is bounded by  $C_\beta p$  (or by  $C_\beta h$ , etc. . .) for all  $h$  and  $p$  sufficiently small. It follows then by simple computations (using the fact that  $p_h$  is the maximizer and that it tends to 0) that  $p_h \sim 2h\chi(\beta)$ . So for  $p = p_h$  the error term in the right-hand side of (2.8) is  $O(h^3)$  and this in turn implies that

$$(2.9) \quad p_h = 2h\chi(\beta) + O(h^2).$$

Therefore (2.1) follows and the proof of Proposition 2.2 is complete.  $\square$

REMARK 2.4. *For future use let us remark that what we have just proven implies that, with  $\xi_x := e^{\omega_x(\beta, h)} - 1$ , we have that for every  $\beta > 0$  there exists  $C_\beta > 0$  such that*

$$(2.10) \quad \max_{p \in [0, 1]} \mathbb{E}[\log(1 + p\xi)] \leq \chi(\beta)h^2 + C_\beta h^3,$$

for every  $h \in [0, 1]$ .

PROOF OF PROPOSITION 2.3. Let us first observe that we can restrict to the case when  $\mathbb{E}[\xi] > 0$ . When this is not the case, by Jensen's inequality the left hand side of (2.2) is bounded above by 0 and thus, considering the case  $p = 0$ , the inequality holds.

By analyzing the function  $p \mapsto \log(1 + p\xi)$ , which is in particular strictly concave and smooth for  $p \in (0, 1)$ , one readily establishes also that

$$(2.11) \quad [0, 1] \ni p \mapsto \mathbb{E}[\log(1 + p\xi)] \in \mathbb{R} \cup \{-\infty\},$$

is a strictly concave smooth function which is continuous up to the boundary points. More precisely, keeping in mind that  $(\log(1 + p\xi))_+ = \log(1 + p\xi_+)$  and  $(\log(1 + p\xi))_- = -\log(1 - p\xi_-)$ , we have that the function in (2.11)

- converges to zero when  $p \searrow 0$ . This is a consequence of the Dominated Convergence Theorem:  $(\log(1 + p\xi))_+ \leq (\log(1 + \xi))_+$  (recall that  $\mathbb{E}[(\log(1 + \xi))_+] < \infty$ ) and  $(\log(1 + p\xi))_-$  is bounded for  $p$  away from one;
- converges to its boundary values when  $p \nearrow 1$ : the boundary value is finite if  $\mathbb{E}[(\log(1 + \xi))_-] < \infty$  and it takes values  $-\infty$  otherwise. This follows because  $(\log(1 + p\xi))_+$  is non decreasing in  $p$  to the limit value  $(\log(1 + \xi))_+$  that has bounded expectation and because also  $(\log(1 + p\xi))_-$  is non decreasing in  $p$ .

Let  $p_\star$  denote the (unique) value of  $p$  for which the maximum in (2.4) is attained.

Let us argue first that  $p_\star > 0$ . In fact,  $p_\star = 0$  is not possible because the function (2.11) takes value zero for  $p = 0$ , but its derivative is equal to  $\mathbb{E}[\xi/(1 + p\xi)]$  which approaches  $\mathbb{E}\xi > 0$  (see the beginning of the proof) for  $p \searrow 0$ : this follows by separating once again the case of  $\xi$  positive, for which we apply the Monotone Convergence Theorem, and  $\xi$  negative, for which the integrand is bounded.

Suppose now that  $p_\star \in (0, 1)$  (the case  $p_\star = 1$  is treated at the end). In this case we exploit the fact that the first derivative of the map (2.11) is zero at  $p_\star$  and we obtain

$$(2.12) \quad \mathbb{E} \left[ \frac{\xi}{1 + p_\star \xi} \right] = 0 \quad \text{which implies} \quad \mathbb{E} \left[ \frac{1}{1 + p_\star \xi} \right] = 1.$$

Reintroducing the dependence on  $x$  we set

$$(2.13) \quad Y_\Lambda := \prod_{x \in \Lambda} (1 + p_\star \xi_x) \quad \text{and} \quad Z_{\Lambda, \xi} := \mathbf{E}_\Lambda \left[ \prod_{\{x : \delta_x=1\}} (1 + \xi_x) \right].$$

With these notations we have

$$(2.14) \quad \mathbb{E} \log Z_{\Lambda, \xi} = \mathbb{E} \log \left( \frac{Z_{\Lambda, \xi}}{Y_{\Lambda}} \right) + \mathbb{E} \log Y_{\Lambda} \leq \log \mathbb{E} \left[ \frac{Z_{\Lambda, \xi}}{Y_{\Lambda}} \right] + |\Lambda| \mathbb{E} \log(1 + p_{\star} \xi).$$

Hence to conclude it suffices to show that  $\mathbb{E} [Z_{\Lambda, \xi}/Y_{\Lambda}] \leq 1$ . To establish this we introduce a new law  $\tilde{\mathbb{P}}$  for the disorder

$$(2.15) \quad \frac{d\tilde{\mathbb{P}}}{d\mathbb{P}}(\omega) := \frac{1}{Y_{\Lambda}(\omega)},$$

where  $\mathbb{E}[Y_{\Lambda}^{-1}] = 1$  because of the second equality in (2.12). Under this new probability, the variables  $\xi_x$  are still IID and we have from (2.12)

$$(2.16) \quad \tilde{\mathbb{E}}[1 + \xi_x] = 1.$$

Hence for this reason we have (recall the convention that a product over an empty set is equal to one)

$$(2.17) \quad \mathbb{E} \left[ \frac{Z_{\Lambda, \xi}}{Y_{\Lambda}} \right] = \tilde{\mathbb{E}} Z_{\Lambda, \xi} = \mathbf{E}_{\Lambda} \tilde{\mathbb{E}} \left[ \prod_{x \in \Lambda: \delta_x=1} (1 + \xi_x) \right] = 1.$$

We are left with the case  $p_{\star} = 1$ . In this case the derivative of the map (2.11) must be positive for every  $p \in (0, 1)$ , that is  $\mathbb{E} [(1 + p\xi)^{-1}] < 1$  for every  $p \in (0, 1)$ , so  $\mathbb{E} [(1 + \xi)^{-1}] \leq 1$ : the continuity for  $p \nearrow 1$  is established by splitting the expectations according to  $\xi \geq 0$ , in this case the integrand is bounded, and  $\xi < 0$  for which we can apply the Monotone Convergence Theorem. We use again (2.14), even if in this case  $1/Y_{\Lambda}$  is not a probability density. But we can argue directly that

$$(2.18) \quad \begin{aligned} \mathbb{E} \left[ \frac{Z_{\Lambda, \xi}}{Y_{\Lambda}} \right] &= \mathbf{E}_{\Lambda} \mathbb{E} \left[ \prod_{x \in \Lambda} \frac{1}{(1 + \xi_x)} \prod_{x \in \Lambda: \delta_x=1} (1 + \xi_x) \right] \\ &= \mathbf{E}_{\Lambda} \prod_{x \in \Lambda: \delta_x=0} \mathbb{E} \left[ \frac{1}{(1 + \xi_x)} \right] \leq 1. \end{aligned}$$

This completes the proof of Proposition 2.3.  $\square$

An alternative proof of Proposition 2.3 can be obtained by adapting the proof of Lemma 5.6.

**3. Proof of Theorem 1.1: Lower bound on the free energy.** The lower bound we obtain on the free energy is in a sense less precise than the upper bound since the correction we obtain is  $h^{2+\varepsilon}$  instead of  $h^3$ .

PROPOSITION 3.1. *Choose  $\beta > 0$ . There exist  $\varepsilon = \varepsilon_d$  and  $C_\beta = C_{\beta,d} > 0$  such that for every  $h > 0$*

$$(3.1) \quad \mathbb{F}(\beta, h) \geq \chi(\beta)h^2 - C_\beta h^{2+\varepsilon}.$$

Note that we can assume that  $h$  is sufficiently small whenever needed. Indeed if (3.1) holds for  $h < h_0 = h_0(d, \beta)$  and a constant  $C_\beta$  then it necessary holds for all  $h > 0$  with a modified constant  $C'_\beta = \max(C_\beta, \chi(\beta)h_0^{-\varepsilon})$  (this choice makes the right-hand side in (3.1) negative for  $h > h_0$ ).

The proof of Proposition 3.1 is essentially self-contained except for a result of super-additivity connected to the existence of the free energy that we cite from [25]. For this result we introduce

$$(3.2) \quad \tilde{\Lambda}_N := \llbracket 1, N \rrbracket^d,$$

and we let  $\partial\Lambda_N := \Lambda_N \setminus \mathring{\Lambda}_N$  denote the internal boundary of  $\Lambda_N$  ( $\Lambda_N$  and  $\mathring{\Lambda}_N$  are defined at the beginning of Section 1.2) and set

$$(3.3) \quad \delta_x^u := \mathbf{1}_{[u-1, u+1]}(\phi(x)).$$

PROPOSITION 3.2 ([25, Prop. 4.2]). *For every  $\beta > 0$ , every  $h \in \mathbb{R}$  and every  $u \in \mathbb{R}$  we have that*

$$(3.4) \quad \mathbb{F}(\beta, h) = \lim_{N \rightarrow \infty} \frac{1}{N^d} \mathbb{E} \mathbb{E} \left[ \log \mathbf{E} \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \middle| \mathcal{F}_{\partial\Lambda_N}^\phi \right] \right].$$

Moreover for every  $N$

$$(3.5) \quad \mathbb{F}(\beta, h) \geq \frac{1}{N^d} \mathbb{E} \mathbb{E} \left[ \log \mathbf{E} \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \middle| \mathcal{F}_{\partial\Lambda_N}^\phi \right] \right].$$

Proposition 3.2 deserves some discussion. The point is that it introduces a different partition function: let us discuss first the case  $u = 0$ . The main difference in this case is that we are not considering 0 boundary conditions, but boundary conditions that are random and that they are sampled from

a LFF, so they are zero only in some averaged sense (there is also the milder difference that the contacts are only the ones in the box  $\tilde{\Lambda}_N$  which is slightly smaller than  $\Lambda_N$ ). When we introduce  $u \neq 0$  we can think of this new partition function as the partition function of the model in which the boundary conditions are not sampled from a LFF with mean zero, but with mean  $-u$ : we have then written the partition function by exploiting the continuum symmetry of the LFF and we have translated the region in which the contact potential acts up by  $u$  and reset the boundary mean to zero. Proposition 3.4 states two facts:

- (A) The free energy associated to this new model coincides with the free energy of the original model, and this regardless of the value of  $u$ ;
- (B) The free energy dominates its finite  $N$  approximation, if we choose the modified partition function we have just introduced for the finite  $N$  approximation: this is proven in [25] as a direct consequence of the fact that the logarithm of the modified partition function forms a super-additive sequence.

The proof of Proposition 3.1, which involves several steps, is given in Section 3.3. Before going through it we provide (in Section 3.1 below) a quick exposition of the main underlying ideas, and introduce in Section 3.2 a decomposition of the field which serves as an important technical tool for the proof.

3.1. *Sketch of proof for Proposition 3.1.* The intuition behind our proof of Proposition 3.1 comes from the inequality (2.2), which implies that for every choice of  $u$  and  $N$

$$(3.6) \quad \log \mathbf{E} \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \middle| \mathcal{F}_{\partial \Lambda_N}^\phi \right] \leq |\tilde{\Lambda}_N| \max_{p \in [0,1]} \mathbb{E} \left[ \log \left( 1 + p[e^{\beta\omega - \lambda(\beta) + h} - 1] \right) \right].$$

We observe that this inequality is an equality if  $(\delta_x^u)_{x \in \mathbb{Z}^d}$  is replaced by a field of Bernoulli variable with parameter  $p(\beta, h)$  which is the maximizer of the right-hand side of (3.6). Hence our strategy relies on fixing  $N$  large so that the influence of the boundary condition vanishes, and the value of  $u$  in such a way that  $(\delta_x^u)_{x \in \mathbb{Z}^d}$  resembles an IID Bernoulli field with optimal density. This is achieved by fixing  $u = u(\beta, h)$  in such a way so that  $\mathbb{E}[\delta_x^u] = 2h\chi(\beta)$  (recall that from (2.9) this ensures that the density is close to optimal). Moreover, at a heuristical level, when  $h \searrow 0$ , the dependence between the

variables  $(\delta_x^u)_{x \in \tilde{\Lambda}_N}$  vanishes: indeed as our fixed density vanishes, we have  $u(\beta, h) \rightarrow \infty$ , and high peaks of the LFF are known to display some asymptotic independence (see [13] for an illustration). Most of the challenge is then to transform this intuition of asymptotic independence into a quantitative statement.

The strategy of proof is the following: we split  $\tilde{\Lambda}_N$  into smaller boxes of edge length  $M$  with  $1 \ll M \ll N$ , and we wish to consider the contribution of each box separately. To do so we write  $\phi$  as a sum of “local” fields whose compact supports corresponds roughly to a box, plus a negligible rest (see Proposition 3.3). It is not possible for the support of the local field to match exactly with boxes and they must display some overlap, but we play with an extra parameter  $1 \ll L \ll M$  to make the total area of overlapping region negligible.

Once this decomposition is made, we need to show the following two estimates:

- (i) The contribution per site to the free energy inside the region where there is no overlap is given to leading order in  $h \searrow 0$  by  $\chi(\beta)h^2$ . This is the content of Lemma 3.9.
- (ii) The contribution per site to the free energy in regions where the support of different local fields intersect is larger than  $-h^{2-\varepsilon}$ . While the second estimate might seem very rough, it turns out to be sufficient for our purpose since the overlap of the supports of local fields only accounts for a small portion of the box  $\tilde{\Lambda}_N$ .

*3.2. A finite range decomposition of the free field.* Let us explain in this section our decomposition of  $\phi$  into a sum of random field supported cubic boxes plus a random field of much smaller amplitude (which *contains* all the long range correlations).

Throughout the text we use cube for hyper-cube. When  $d \geq 3$ , given  $L \geq 1$  ( $L$  will later be chosen as a function of  $h$  that diverges in the limit  $h \searrow 0$ , so we can think of  $L$  as a large integer), we choose the support of the local fields to be cubes of edge length  $M = L^2 + L$  while the length  $L$  corresponds to the width of the overlap region between the support of two neighboring local fields.

**PROPOSITION 3.3.** *If  $\phi = (\phi(x))_{x \in \mathbb{Z}^d}$  is a LFF on  $\mathbb{Z}^d$ , then one can construct a collection of independent random fields  $\{\phi_0, (\phi^{(z)})_{z \in \mathbb{Z}^d}\}$  (with non negative covariance entries) which satisfy the following properties*



(i) We have

$$(3.7) \quad \phi \stackrel{\text{law}}{=} \phi_0 + \sum_{z \in \mathbb{Z}^d} \phi^{(z)},$$

(ii) The field  $\phi_0$  satisfies

$$(3.8) \quad \sup_x \text{Var}(\phi_0(x)) = \begin{cases} O(L^{-1}(\log L)) & \text{if } d = 3, \\ O(L^{-1}) & \text{if } d \geq 4. \end{cases}$$

(ii) The fields  $\phi^{(z)}$  are identically distributed up to a lattice translation and they are supported in a box of diameter  $M + L$ . More precisely we have  $\phi^{(z)}(\cdot) \stackrel{\text{law}}{=} \phi^{(0)}(\cdot - Mz)$  for every  $z$  and almost surely

$$(3.9) \quad \text{Supp}(\phi^{(z)}) = Mz + \llbracket 1, M + L - 1 \rrbracket^d.$$

where  $\text{Supp}(\phi^{(z)}) := \{x \in \mathbb{Z}^d : \phi^{(z)}(x) \neq 0\}$ .

One way to picture the decomposition is thinking that the support of the  $z$ -local field are the integer points in  $(0, M + L)^d$  translated by  $Mz$ . The supports of the fields  $\phi^{(z)}$  and  $\phi^{(z')}$ ,  $z \neq z'$ , do not overlap if  $|z - z'| \geq 2$  (recall that  $|\cdot|$  denotes the  $l_1$  norm). If  $|z - z'| = 1$  they do overlap, but at most on  $(M + L - 1)^{d-1}(L - 1)$  sites. There are regions (of size  $(L - 1)^d$  sites: the *corners*) where  $2^d$  local fields overlap and this is the maximal number of overlapping fields.

PROOF. We first decompose  $\phi(x)$  into  $(L + 1)^d$  independent fields

$$(3.10) \quad \phi(x) = (L + 1)^{-d/2} \sum_{y \in \llbracket 0, L \rrbracket^d} \varphi^{(y)}(x),$$

where  $(\varphi^{(y)})_{y \in \llbracket 0, L \rrbracket^d}$  are IID infinite volume LFF. Then we introduce the grid  $\mathbb{H}_M$

$$(3.11) \quad \mathbb{H}_M := \{x \in \mathbb{Z}^d : \text{ther exists } i \in \llbracket 1, d \rrbracket \text{ such that } x_i \in M\mathbb{Z}\},$$

which splits the lattice  $\mathbb{Z}^d$  into cubic boxes of edge length  $M - 1$ . Let us also introduce the translations of  $\mathbb{H}_M$ :

$$(3.12) \quad \mathbb{H}_M^{(y)} := y + \mathbb{H}_M,$$

which will be used for  $y \in \llbracket 0, L \rrbracket^d$  and the boxes (with boundaries) delimited by  $\mathbb{H}_M^{(y)}$  are:

$$(3.13) \quad B_M^{(y,z)} := \prod_{i=1}^d \llbracket y_i + Mz_i, y_i + M(z_i+1) \rrbracket \quad \text{for } y \in \llbracket 0, L \rrbracket^d \text{ and } z \in \mathbb{Z}^d.$$

We let  $H^{(y)}$  be the harmonic extension of the restriction of  $\varphi^{(y)}$  to  $\mathbb{H}_M^{(y)}$ , which is the solution of the system

$$(3.14) \quad \begin{cases} (\Delta H^{(y)})(x) = 0 & \text{for } x \in \mathbb{Z}^d \setminus \mathbb{H}_M^{(y)}, \\ H^{(y)}(x) = \varphi^{(y)}(x) & \text{for } x \in \mathbb{H}_M^{(y)}. \end{cases}$$

Recall that  $H^{(y)}(\cdot)$  is the conditional expectation of  $\varphi^{(y)}$  knowing its value on  $\mathbb{H}_M^{(y)}$ , that is

$$(3.15) \quad \mathbb{E} \left[ \varphi^{(y)}(x) \middle| \mathcal{F}_{\mathbb{H}_M^{(y)}}^{\varphi^{(y)}} \right] = H^{(y)}(x).$$

Now we define the fields

$$(3.16) \quad \psi^{(y,z)}(x) := \left( \varphi^{(y)}(x) - H^{(y)}(x) \right) \mathbf{1}_{B_M^{(y,z)}}(x),$$

and it follows from the spatial Markov property that, for every  $y$ , the random fields  $\{\psi^{(y,z)}\}_{z \in \mathbb{Z}^d}$  are independent, and  $\psi^{(y,z)}$  is a free field on  $B_M^{(y,z)}$  with zero boundary conditions. We are now ready to make the fields  $\phi_0$  and  $\phi^{(z)}$  explicit:

$$(3.17) \quad \begin{aligned} \phi_0(x) &:= (L+1)^{-d/2} \sum_{y \in \llbracket 0, L \rrbracket^d} H^{(y)}(x), \\ \phi^{(z)}(x) &:= (L+1)^{-d/2} \sum_{y \in \llbracket 0, L \rrbracket^d} \psi^{(y,z)}(x). \end{aligned}$$

The support property of  $\phi^{(z)}$  is evident, so it remains to show that (3.8) holds. This is a consequence of the bound (Lemma A.4 for a proof)

$$(3.18) \quad \text{Var} \left( H^{(y)}(x) \right) \leq C_d \left( \left( \text{dist} \left( x, \mathbb{H}^{(y)} \right) + 1 \right)^{2-d} \right).$$

In fact since  $(\varphi^{(y)})_{y \in \llbracket 0, L \rrbracket^d}$  is a family of independent fields, from (3.18) we have

$$(3.19) \quad \begin{aligned} \text{Var}(\phi_0(x)) &= \frac{1}{(L+1)^d} \sum_{y \in \llbracket 0, L \rrbracket^d} \text{Var} \left( H^{(y)}(x) \right) \\ &\leq \frac{C_d}{(L+1)^d} \sum_{y \in \llbracket 0, L \rrbracket^d} \frac{1}{\left( \text{dist} \left( x, \mathbb{H}^{(y)} \right) + 1 \right)^{d-2}}, \end{aligned}$$

and simple symmetry arguments, assuming  $L > 1$ , show that the last expression is bounded by the case in which  $-x_j$  is equal to the (upper or lower) integer part of  $L/2$  for every  $j$ : this corresponds to summing on  $y$  over a cube of edge length  $L + 1$  centered, up to parity issues, at  $x$ . We now assume  $L$  even to simplify the notations: the last expression in (3.19) is therefore bounded for every  $x$  by

$$(3.20) \quad \frac{2^d C_d}{(L+1)^d} \sum_{y \in \llbracket 0, L/2 \rrbracket^d} \frac{1}{(\text{dist}(0, \mathbb{H}(y)) + 1)^{d-2}} \leq$$

$$\frac{C_d 2^d d!}{(L+1)^d} \sum_{\substack{y \in \llbracket 0, L/2 \rrbracket^d \\ y_1 \leq y_2 \leq \dots \leq y_d}} \frac{1}{(y_1 + 1)^{d-2}} \leq \frac{C_d 2^d d! (\frac{L}{2} + 1)^{d-1}}{(L+1)^d} \sum_{y=0}^{L/2} \frac{1}{(y+1)^{d-2}},$$

which, separating the cases  $d = 3$  and  $d > 3$ , directly yields the desired estimate. This completes the proof of Proposition 3.3.  $\square$

### 3.3. Proof of Proposition 3.1.

3.3.1. *Step 1: Choice of the finite size parameters.* Recall that by Proposition 3.2 it suffices to show that there exists  $h_0 > 0$  such that for every  $h \in (0, h_0)$  there exist  $N$  and  $u$  such that

$$(3.21) \quad \mathbb{E} \mathbf{E} \left[ \log \mathbf{E} \left[ e^{\sum_{x \in \bar{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \middle| \mathcal{F}_{\partial \Lambda_N}^\phi \right] \right] \geq N^d (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}).$$

So  $N$  may (and will) depend on  $h$  as well as  $u$ . In view of exploiting the finite range decomposition of Proposition 3.3 we introduce also an  $h$  dependent quantity  $L$  and recall that  $M = L^2 + L$ :

$$(3.22) \quad L := h^{-\frac{1}{\kappa_d}}, \quad k := h^{-10} \quad \text{and}$$

$$N := kM + L = kL^2 + (k+1)L \stackrel{h \searrow 0}{\sim} h^{-10 - \frac{2}{\kappa_d}},$$

where  $\kappa_d \geq 4d$  is a positive integer that depends only on  $d$  (the choice is made just after (3.62) below). We drop integer parts in the notation for the sake of readability. Finally we fix  $u_h$  (we will often omit the subscript for better readability) to be the unique positive solution  $u$  to the equation

$$(3.23) \quad \mathbf{P}(\sigma_d \mathcal{N} \in [u-1, u+1]) = 2h\chi(\beta).$$

Note that the left-hand side of (3.23) decreases from  $\mathbf{P}(\sigma_d \mathcal{N} \in [-1, +1])$  to zero when  $u$  goes from 0 to  $\infty$ . Hence a unique solution  $u_h$  to (3.23) exists

if (and only if)  $h \in [0, \mathbf{P}(\sigma_d \mathcal{N} \in [-1, +1]) / (2\chi(\beta))]$  (which we assume). Lemma 3.4 below provides a sharp asymptotic expression for  $u_h$  along with a useful technical estimate.

We need a preliminary notation: for  $a > 0$  and  $h > 0$  small we set

$$(3.24) \quad u(a, h) := \sigma_d \sqrt{2 \log(1/h)} + 1 - \frac{\sigma_d \log \log(1/h)}{2 \sqrt{2 \log(1/h)}} - \sigma_d \frac{\log(2a\sqrt{\pi})}{\sqrt{2 \log(1/h)}}.$$

LEMMA 3.4. *For  $h \searrow 0$*

$$(3.25) \quad u_h = u(2\chi(\beta), h) + o\left(1/\sqrt{\log(1/h)}\right),$$

and if  $0 \leq r(h) = o(1/\sqrt{\log(1/h)})$  then both for  $I_h = [0, r(h)]$  and  $I_h = [-r(h), 0]$  we have

$$(3.26) \quad \mathbf{P}(\sigma_d \mathcal{N} - u_h + 1 \in I_h) = \frac{2\chi(\beta)}{\sigma_d^2} u_h h r(h) (1 + o(1)).$$

On the other hand, for every choice of two positive constants  $c$  and  $C$  we have for  $h$  sufficiently small

$$(3.27) \quad \mathbf{P}(\sigma_d \mathcal{N} \geq u_h - 1 + c) \leq h (\log(1/h))^{-C}.$$

REMARK 3.5. *Our analysis depends on the specific form taken by the potential. If we consider an arbitrary bounded compactly supported non-negative function  $\vartheta$ , the analysis of the model in which  $(\beta\omega_x + h - \lambda(\beta))$  times  $\mathbf{1}_{[-1,1]}(\phi(x))$  is replaced by  $(\beta\omega_x + h)\vartheta(\phi(x))$  appears to be challenging even in its annealed version! On the other the arguments we use generalizes to other cases in which the annealed model coincides with the  $\beta = 0$  case. More precisely our analysis does cover the case in which  $(\beta\omega_x + h - \lambda(\beta))\mathbf{1}_{[-1,1]}(\phi(x))$  is replaced by  $(\beta\omega_x + h - \lambda(\beta\vartheta(\phi(x))))$  times  $\vartheta(\phi(x))$ . Such a  $\phi(x)$  dependent shift is not very natural, but notice that if  $\vartheta(\cdot)$  is the indicator function of a bounded set of positive Lebesgue measure, we are back to a constant shift of the value of  $h$ . The solution of this generalized model requires a different choice of the global shift  $u(2\chi(\beta), h)$ : this new quantity differs from the one we are using only for a term proportional to  $1/\sqrt{|\log h|}$ , i.e. only the lowest order term in (3.24) is modified. This is technically relevant for the proof, but bears no consequence on the final result about the height at which the field is pushed near criticality. This height is in fact determined only with a precision  $o(\sqrt{|\log h|})$  (see Theorem 1.2).*

*Proof of Lemma 3.4.* Everything is based on the well known asymptotic ( $x \nearrow \infty$ ) estimate

$$(3.28) \quad \mathbf{P}(\mathcal{N} > x) = \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \left(1 + O\left(\frac{1}{x^2}\right)\right),$$

which in particular implies (via a relatively cumbersome computation) that

$$(3.29) \quad \mathbf{P}(-u(a, h) + \sigma_d \mathcal{N} \in [-1, 1]) \stackrel{h \searrow 0}{\sim} ah,$$

and we point out that the result is the same if  $[u_h - 1, u_h + 1]$  is replaced by  $[u_h - 1, u_h - 1 + c]$ , any  $c > 0$ : that is, the contribution to the asymptotic behavior is all near  $u_h - 1$ . Using (3.29) together with (3.23) we readily extract (3.25).

At this point it is rather straightforward to realize that (3.27) holds (this is just a quantitative version of the observation that we just made that the contribution to the asymptotic behavior is all near  $u_h - 1$ ): the leading effect generated by a shift by  $c$  in the Gaussian term is  $\exp(-(2c/\sigma_d)\sqrt{2\log(1/h)})$ , that is a factor that vanishes faster than any power of  $1/\log(1/h)$  and this is the content of (3.27).

The estimate for (3.26) requires more care. For  $I_h = [-r(h), 0]$  (the argument for  $I_h = [0, r(h)]$  is essentially the same) we have

$$(3.30) \quad \mathbf{P}(\sigma_d \mathcal{N} - u_h + 1 \in [-r(h), 0]) = \int_{u_h - 1 - r(h)}^{u_h - 1} g_{\sigma_d}(z) dz,$$

with  $g_{\sigma}(\cdot)$  the density of  $\sigma \mathcal{N}$ . For  $r(h) = o(1/\sqrt{\log(1/h)})$  one directly verifies that

$$(3.31) \quad \lim_{h \searrow 0} \sup_{z: |z - u_h + 1| \leq r(h)} \left| \frac{g_{\sigma_d}(z)}{g_{\sigma_d}(u_h - 1)} - 1 \right| = 0,$$

so that

$$(3.32) \quad \int_{u_h - 1 - r(h)}^{u_h - 1} g_{\sigma_d}(z) dz = r(h) g_{\sigma_d}(u_h - 1) (1 + o(1)).$$

Using the asymptotic equivalence (3.28) and (3.23) we have, when  $h \searrow 0$

$$(3.33) \quad g_{\sigma_d}(u_h - 1) = \frac{u_h}{\sigma_d^2} \mathbf{P}(\sigma_d \mathcal{N} \in [u_h - 1, u_h + 1]) (1 + o(1)) = \frac{2\chi(\beta) h u_h}{\sigma_d^2} (1 + o(1)),$$

which yields (3.26). The proof of Lemma 3.4 is therefore complete.  $\square$

3.3.2. *Step 2: field decomposition and boundary control estimate.* Let us use the decomposition (3.7) of Proposition 3.3 for the LFF. Using the information we have concerning the support of  $\phi^{(z)}$  we see that  $\phi^{(z)}(x) = 0$  for  $z \in \llbracket 0, k-1 \rrbracket^d$  and  $x \in \partial\Lambda_N$ , with  $\partial\Lambda_N$  defined right after (3.2). Therefore the value of  $\phi(x)$  for  $x \in \partial\Lambda_N$  is not affected by the realization of  $(\phi^{(z)})_{z \in \llbracket 0, k-1 \rrbracket^d}$ . Hence letting

$$(3.34) \quad \mathbf{P}_N^0, \mathbf{P}_N^1 \text{ and } \mathbf{P}_N^2,$$

denote, respectively, the distribution of

$$(3.35) \quad \phi_0, \left( \phi^{(z)} \right)_{z \in \llbracket 0, k-1 \rrbracket^d} \text{ and } \left( \phi^{(z)} \right)_{z \in \mathbb{Z}^d \setminus \llbracket 0, k-1 \rrbracket^d},$$

we obtain from Jensen's inequality that

$$(3.36) \quad \mathbf{E} \left[ \log \mathbf{E} \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \middle| \mathcal{F}_{\partial\Lambda_N}^\phi \right] \right] = \\ \mathbf{E} \left[ \log \mathbf{E}_N^0 \otimes \mathbf{E}_N^2 \left[ \mathbf{E}_N^1 \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \middle| \mathcal{F}_{\partial\Lambda_N}^\phi \right] \right] \right] \geq \\ \mathbf{E}_N^0 \otimes \mathbf{E}_N^2 \left[ \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \right] \right].$$

Now let us show that one can replace  $\tilde{\Lambda}_N$  (see (3.2)), in (3.36) by a smaller set in order to avoid boundary effects. Set

$$(3.37) \quad \Lambda'_N = \llbracket L+1, kM \rrbracket^d \quad \text{and} \quad \Gamma_N := \tilde{\Lambda}_N \setminus \Lambda'_N.$$

We are going to prove:

LEMMA 3.6. *For  $h \geq 0$  we have  $\mathbf{P}_N^0 \otimes \mathbf{P}_N^2$  a.s.*

$$(3.38) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \right] \geq \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] - \lambda(\beta) |\Gamma_N|.$$

Since the cardinality of  $\Gamma_N$  is  $O(N^{d-1}L)$ , our choice of parameter makes the last term negligible with respect to  $N^d h^{2+\varepsilon}$ , and thus combining (3.36) and (3.38), the inequality (3.21) follows if  $\varepsilon < 1/\kappa_d$  (which we therefore assume;  $\kappa_d$  is introduced right after (3.22)) and if we show that

$$(3.39) \quad \mathbb{E} \mathbf{E}_N^0 \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \geq N^d (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}).$$

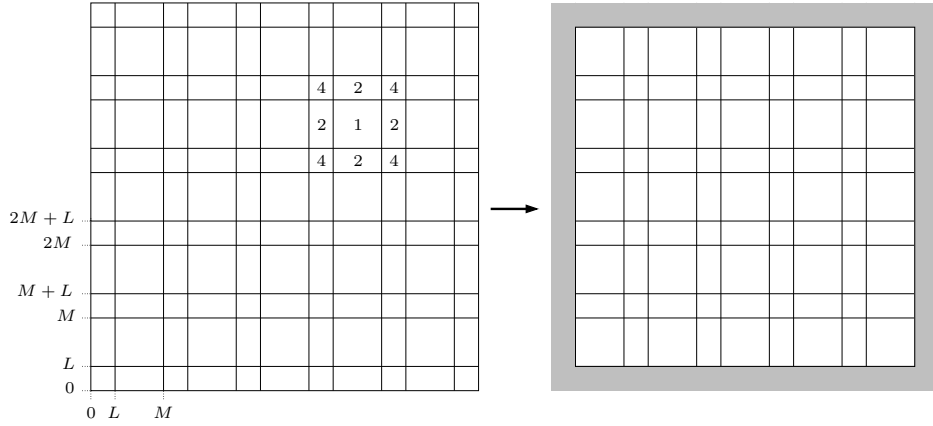


FIG 1. On the left there is the set  $\Lambda_N$ , covered by partially superposed boxes of edge length  $M + L$  (the picture is drawn for  $d = 2$  just for the sake of visualization purpose:  $d \geq 2$  is not a relevant dimension for our problem). One of this boxes is highlighted by the numbers 1, 2 and 4: one (and only one) of the fields  $\phi^{(z)}$  is supported, for a well chosen  $z = z_* \in \mathbb{Z}^d$ , exactly on this box. On the region marked by 1, of edge length  $M - L$ , only one field  $\phi^{(z)}$  is non zero, that is  $\phi^{(z)}(x) = 0$  for every  $z \neq z_*$  and  $x$  in the region marked by 1. In the regions, belonging to the frame of region 1, marked by 2 and 4, there are respectively 2 and 4 values of  $z$  (one is  $z_*$ ) for which the  $\phi^{(z)}$  fields are non zero. Of course this frame is much smaller than the box ( $M \sim L^2$ ). moreover in dimension  $d$  there are regions in which up to  $2^d$  fields superpose. Step 2 in the proof consists in showing that we can erase the contacts in the shadowed frame of the large box  $\Lambda_N$  at little price: we will see that in the end this fully deals with the boundary effects because the long range correlations of the field are all carried by the field  $\phi_0$  and we get rid of this field in Step 3.

Note that the expectation w.r.t. to  $\mathbf{P}_N^2$  is not displayed because  $\phi$  restricted to  $\Lambda'_N$  is completely determined by  $\phi_0$  and  $(\phi^{(z)})_{z \in \llbracket 0, k-1 \rrbracket^d}$ .

*Proof of Lemma 3.6.* We let  $\mu_N^A$  be the probability measure defined by

$$(3.40) \quad \frac{d\mu_N^A}{d\mathbf{P}_N^1} \left( (\phi^{(z)})_{z \in \llbracket 0, k-1 \rrbracket} \right) = \frac{e^{\sum_{x \in A} \omega_x(\beta, h) \delta_x^u}}{\mathbf{E}_N^1 \left[ e^{\sum_{x \in A} \omega_x(\beta, h) \delta_x^u} \right]},$$

that is the distribution of  $(\phi^{(z)})_{z \in \llbracket 0, k-1 \rrbracket}$  when interactions with sites in  $A$  are taken into account (this measure depends on the realization of  $\phi_0$  and

possibly also of that of  $\phi^{(z)}$  for  $z \notin \llbracket 0, k-1 \rrbracket$ . We have

$$(3.41) \quad \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \tilde{\Lambda}_N} \omega_x(\beta, h) \delta_x^u} \right] \\ = \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] + \log \mu_N^{\Lambda'_N} \left[ e^{\sum_{x \in \Gamma_N} \omega_x(\beta, h) \delta_x^u} \right],$$

And thus (3.38) follows by Jensen's inequality as follows:

$$(3.42) \quad \mathbb{E} \log \mu_N^{\Lambda'_N} \left[ e^{\sum_{x \in \Gamma_N} \omega_x(\beta, h) \delta_x^u} \right] \geq \\ \mathbb{E} \mu_N^{\Lambda'_N} \left[ \sum_{x \in \Gamma_N} \omega_x(\beta, h) \delta_x^u \right] \geq -\lambda(\beta) |\Gamma_N|,$$

and in the last step we have used the fact that  $\mu_N^{\Lambda'_N}$  does not depend on  $\{\omega_x\}_{x \in \Gamma_N}$  and then we have just used that  $\sum_{x \in \Gamma_N} \delta_x^u \leq |\Gamma_N|$ . The proof of Lemma 3.6 is complete.  $\square$

**3.3.3. Step 3: getting rid of the base field  $\phi_0$ .** Our next step is to get rid of the dependence in  $\phi_0$  in (3.39). We can do this combining two facts. Firstly from (3.8) and our choice of the parameters we know that with an overwhelming large probability  $\phi_0$  is small everywhere. Secondly, from (3.26), we know that the expectation  $\mathbf{E}_N^1[\delta_x^u]$  for  $x \in \delta_x^u$  is not much affected by small variations of  $\phi_0$ . Let us define the event  $\mathcal{C}_N$  by

$$(3.43) \quad \mathcal{C}_N := \left\{ |\phi_0(x)| \leq L^{-1/3} \text{ for every } x \in \Lambda'_N \right\}.$$

A simple union bound using the estimate on the variance (3.8) – recall the choice of parameters (3.22) – yields immediately for  $h$  small

$$(3.44) \quad \mathbf{P}_N^0 \left( \mathcal{C}_N^c \right) \leq \exp \left( -L^{1/4} \right).$$

By applying Jensen's inequality we have also that for every realization of  $\phi_0$

$$(3.45) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \geq \mathbf{E}_N^1 \mathbb{E} \left[ \sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u \right] \geq -|\Lambda'_N| \lambda(\beta),$$

and therefore

$$(3.46) \quad \mathbb{E} \mathbf{E}_N^0 \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \\ \geq \mathbb{E} \mathbf{E}_N^0 \left[ \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \mathbf{1}_{\mathcal{C}_N} \right] - \lambda(\beta) N^d \mathbf{P}_N^0 \left( \mathcal{C}_N^c \right).$$



Now recalling (3.44) and (3.22), this implies that (3.39) follows if one proves that there exists  $h_0 > 0$  such that for every  $h \in (0, h_0)$  and every  $\phi_0 \in \mathcal{C}_N$

$$(3.47) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \geq N^d (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}) .$$

Note that expressions like for example the leftmost sides of (3.45) and (3.47) are now random variables: they are measurable with respect to the  $\sigma$ -field generated by  $\phi_0$ . We are going to see that not averaging over  $\phi_0$  is typically not a problem, at least as long as  $\phi_0 \in \mathcal{C}_N$ . The key result in this direction says that the contact density is not very much affected by conditioning with respect to  $\phi_0$ , as long as  $\phi_0 \in \mathcal{C}_N$ . It will be repeatedly used in the remainder. Here it is:

**LEMMA 3.7.** *For every  $h$  sufficiently small, every  $\phi_0 \in \mathcal{C}_N$  and every  $x \in \Lambda'_N$  we have*

$$(3.48) \quad |\mathbf{E}_N^1[\delta_x^u] - 2\chi(\beta)h| \leq hL^{-1/4} .$$

*Proof of Lemma 3.7.* We introduce the practical notation

$$(3.49) \quad \phi_1(x) := \sum_{z \in \llbracket 0, k-1 \rrbracket^d} \phi^{(z)}(x)$$

(note that the summation defining  $\phi_1(x)$  contains between one and  $2^d$  non-zero terms), and for  $x \in \Lambda'_N$  we have that

$$(3.50) \quad \begin{aligned} \mathbf{E}_N^1[\delta_x^u] &= \mathbf{P}_N^1(\phi_1(x) \in [u-1-\phi_0(x), u+1-\phi_0(x)]) \\ &= \mathbf{P} \left( \sqrt{\sigma_d^2 - \text{Var}(\phi_0(x))} \mathcal{N} \in [u-1-\phi_0(x), u+1-\phi_0(x)] \right) , \end{aligned}$$

where  $\mathbf{P}$  in the last line is just the law of  $\mathcal{N} \sim \mathcal{N}(0, 1)$ , which is of course the only random variable in the expression. Now, for small  $h$  and assuming that  $\mathcal{C}_N$  holds, by monotonicity the right-hand side in (3.50) is maximized when  $\phi_0(x) = L^{-1/3}$  and minimized when  $\phi_0(x) = -L^{-1/3}$ . If we set  $\tilde{\sigma} := \sqrt{\sigma_d^2 - \text{Var}(\phi_0(x))}$ , (3.8) (and the fact that  $u$  is asymptotically proportional to  $\log L$ ) implies that for  $h$  sufficiently small we have

$$\begin{aligned} \frac{\sigma_d}{\tilde{\sigma}} [u-1+L^{-1/3}, u+1+L^{-1/3}] &\supset [u-1+2L^{-1/3}, u+1], \\ \frac{\sigma_d}{\tilde{\sigma}} [u-1-L^{-1/3}, u+1-L^{-1/3}] &\subset [u-1-L^{-1/3}, u+1], \end{aligned}$$

so that

$$(3.51) \quad \mathbf{P} \left( \sigma_d \mathcal{N} \in [u - 1 + 2L^{-1/3}, u + 1] \right) \leq \mathbf{E}_N^1 [\delta_x^u] \leq \mathbf{P} \left( \sigma_d \mathcal{N} \in [u - 1 - L^{-1/3}, u + 1] \right).$$

Now, using (3.26) we obtain that

$$(3.52) \quad \left| \mathbf{E}_N^1 [\delta_x^u] - 2\chi(\beta)h \right| \leq h \mathbf{P} \left( \sigma_d \mathcal{N} \in [u - 1 - L^{-1/3}, u - 1 + 2L^{-1/3}] \right) \leq Ch u_h L^{-1/3},$$

with  $C = C(\beta, d)$  which can be easily read out from (3.26). The proof of Lemma 3.7 is therefore complete.  $\square$

3.3.4. *Step 4: reducing to estimates on  $M$ -boxes.* We are now going to state two technical results. The first (Lemma 3.9) is (3.47), but with  $\Lambda'_N$  replaced by the set of vertices  $x$  which are in the support of a unique  $\phi^{(z)}$

$$(3.53) \quad \Lambda''_N := \left\{ x \in \Lambda'_N : \exists z, \phi_1(x) = \phi^{(z)}(x), \mathbf{P}_N^1 - \text{a.s.} \right\}.$$

Note that our condition  $L \ll M$  ensures that  $|\Lambda''_N| \sim |\Lambda'_N|$ . The reason why the quantity with  $\Lambda''_N$  is easier to handle is that by independence of the  $\phi^{(z)}$ , the partition function can be factorized (and its log becomes simply a sum). For technical reason, we prove in fact (3.47) with  $\Lambda''_N$  but also with an additional constraint of the fields  $\phi^{(z)}$ .

The second result (Lemma 3.10) states that the portion of the domain that we leave out, i.e.  $\Lambda'_N \setminus \Lambda''_N$ , gives a negligible contribution, this is where the constraint imposed in the statement of the first result are used.

Both results are needed for our proof, but while the second is of a technical nature, the first contains the key second moment argument on which the proof relies. Let us observe that  $\Lambda''_N$  is a disjoint union of cubes of diameter  $M - L$

$$(3.54) \quad \Lambda''_N = \bigcup_{z \in \llbracket 0, k-1 \rrbracket^d} \mathcal{B}_L^{(z)},$$

where

$$(3.55) \quad \mathcal{B}_L^{(z)} := \llbracket L + 1, M \rrbracket^d + zM.$$

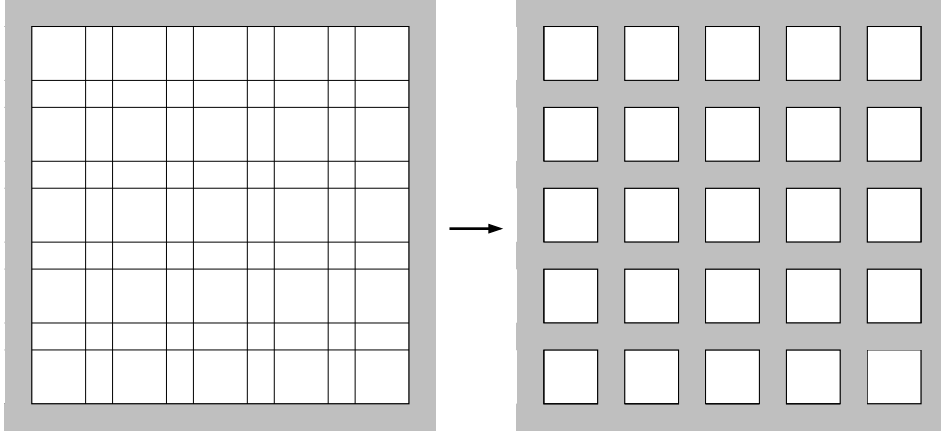


FIG 2. In Step 4 to Step 6 we show that it is sufficient to consider the system in which the contacts in the shadowed grid-like region are erased. The independence properties of the  $\phi^{(z)}$  fields reduce the problem to estimate the contributions to the free energy in each of the non shadowed boxes of edge length  $M - L$ . One important ingredient is also the fact that, for every such a box, we introduce a constraint on the  $\phi^{(z)}$  fields whose support intersects the box or its frame (the box with the frame is a box of edgelength  $M + L$ ). This constraint limits the number of contacts that can take place in this box and its frame. A second moment argument turns out to be very efficient thanks to this control on the number of contacts.

The space between cubes being much smaller than the cube diameter,  $\Lambda_N''$  covers most of the original box. Furthermore let us also introduce

$$(3.56) \quad \mathcal{D}_N := \bigcap_{z \in \llbracket 0, k-1 \rrbracket^d} \mathcal{D}_N^{(z)},$$

with

$$(3.57) \quad \mathcal{D}_N^{(z)} := \left\{ \phi^{(z)} : \left| \left\{ x : \phi^{(z)}(x) \geq 2^{-d}(u-2) \right\} \right| \leq \kappa_d \right\}.$$

where  $\kappa_d$  is fixed in such a way that

$$(3.58) \quad 1 - \mathbf{P}_N^1 \left( \mathcal{D}_N^{(z)} \right) = \mathbf{P}_N^1 \left( \left| \left\{ x : \phi^{(z)}(x) \geq 2^{-d}(u-2) \right\} \right| > \kappa_d \right) \leq h^2.$$

Note that  $\mathcal{D}_N$  limits the number of contact points. Let us show that we can find  $\kappa_d$  such that (3.58) holds.

LEMMA 3.8. *For any  $c > 0$  we can find  $\kappa = \kappa(d, c) \in \mathbb{N}$  such that for  $h$  small*

$$(3.59) \quad \mathbf{P}_N^1 \left( \left| \left\{ x : \phi^{(z)}(x) \geq c\sqrt{2 \log(1/h)} \right\} \right| > \kappa \right) \leq h^2.$$

Of course (3.58) holds by Lemma 3.8 (keep in mind (3.24) and (3.25)) by choosing a value of  $c \in (0, 2^{-d}\sigma_d)$  and then fixing  $\kappa_d = \max(\kappa(d, c), 4d)$  (the requirement  $\kappa_d \geq 4d$  is due to (3.92) below).

PROOF. By a union bound the probability on the left-hand side of (3.59) is not larger than

$$(3.60) \quad (M + L)^{d\kappa} \max_{x_1, \dots, x_\kappa} \mathbf{P}_N^1 \left( \phi^{(z)}(x_j) \geq c\sqrt{2\log(1/h)} \text{ for } j = 1, \dots, \kappa \right) \\ \leq (2h^{-2/\kappa})^{d\kappa} \max_{x_1, \dots, x_\kappa} \mathbf{P}_N^1 \left( \sum_{j=1}^{\kappa} \phi^{(z)}(x_j) \geq c\kappa\sqrt{2\log(1/h)} \right),$$

where of course the  $x_j$ 's are in the support of  $\phi^{(z)}$  and we have used  $M = L^2 + L \leq 2h^{-2/\kappa}$ . We are therefore reduced to estimating the variance of  $\sum_{j=1}^{\kappa} \phi^{(z)}(x_j)$ . The first observation is that we can replace  $\mathbf{P}_N^1$  with  $\mathbf{P}$ , that is we can work with the free field in  $\mathbb{Z}^d$  (this is just because  $G(x, y)$  is larger than the covariance of  $\phi^{(z)}$  for every choice of  $x$  and  $y$ ). The next step is realizing that this variance is maximal when the  $x_j$  are *closely packed*: more precisely that the maximum of the variance is bounded above by a constant that depends only on the dimension time the variance of the case in which the set  $\{x_1, \dots, x_\kappa\}$  is replaced by the cube of edge length equal to  $\lceil \kappa^{1/d} \rceil$  (of course this cube contains more than  $\kappa$  points, unless  $\kappa^{1/d}$  is an integer number: the full argument is left to the reader and can be found for example in [25, Lemma 6.11] where one can refer also for an explicit constant (that is however largely overestimated). The computation for the cube is straightforward and yields a variance which behaves for  $\kappa$  large like  $C_d \kappa^{(d+2)/d}$ . Therefore

$$(3.61) \quad \mathbf{P}_N^1 \left( \sum_{j=1}^{\kappa} \phi^{(z)}(x_j) \geq c\kappa\sqrt{2\log(1/h)} \right) \leq \\ \mathbf{P} \left( \sigma_d \mathcal{N} \geq c \frac{\kappa^{1-(d+2)/(2d)}}{C_d^{1/2}} \sqrt{2\log(1/h)} \right) \leq h^{c^2 \kappa^{(d-2)/d} / (C_d \sigma_d^2)} = h^{c'_d \kappa^{(d-2)/d}}.$$

Therefore

$$(3.62) \quad \mathbf{P}_N^1 \left( \left| \left\{ x : \phi^{(z)}(x) \geq c\sqrt{2\log(1/h)} \right\} \right| > \kappa \right) \leq 2^{d\kappa} h^{-2d + c'_d \kappa^{(d-2)/d}} \leq h^2,$$

with the last step that holds if  $\kappa$  is chosen so that  $c'_d \kappa^{(d-2)/d} > 2d + 2$  and if  $h$  is sufficiently small. Therefore  $\kappa(d, c)$  is identified and the proof of Lemma 3.8 is complete.  $\square$

Here is the first of the two results that we announced:

LEMMA 3.9. *For every  $h$  sufficiently small and  $\phi_0 \in \mathcal{C}_N$  we have (note that  $|\mathcal{B}_L^{(z)}| = (M - L)^d = L^{2d}$ )*

$$(3.63) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}_L^{(z)}} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(z)}} \right] \geq L^{2d} (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}).$$

As a consequence for  $\phi_0 \in \mathcal{C}_N$  we have

$$(3.64) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda_N''} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N} \right] = \sum_{z \in \llbracket 0, k-1 \rrbracket^d} \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}_L^{(z)}} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(z)}} \right] \geq N^d (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}).$$

Of course this is not yet sufficient to conclude because there is no direct way to show that adding the sites of  $\Lambda_N' \setminus \Lambda_N''$  has a positive contribution on the free energy.

We will in fact content ourselves with showing that this contribution is not too negative. This is the object of our second result. It requires the introduction of some further notation. We define for  $A \subset \Lambda_N'$ ,  $\tilde{Z}_{A, \omega}$  as a partition function restricted to  $\mathcal{D}_N$  for which the interaction is present only for sites in  $A$

$$(3.65) \quad \tilde{Z}_{A, \omega} := \mathbf{E}_N^1 \left[ e^{\sum_{x \in A} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N} \right].$$

Moreover we define

$$(3.66) \quad K_h := \inf \{ K \geq \sqrt{\log 1/h} : \mathbb{E}[|\omega_x|; |\omega_x| > K] \leq h^3 \}.$$

The condition  $K \geq \sqrt{\log 1/h}$  is artificial, but it is convenient for us to have  $\lim_{h \searrow 0} K_h = \infty$ . Since we have assumed  $\mathbb{E}[\exp(t\omega_x)] < \infty$  for every  $t \in \mathbb{R}$  we have that  $K_h = o(\log(1/h))$ . We set

$$(3.67) \quad \bar{\omega}_x := \omega_x \mathbf{1}_{\{|\omega_x| \leq K_h\}}.$$

Here is the second result.

LEMMA 3.10. *For every  $\gamma > 0$  there exists  $h_0$  such that for all  $h \in [0, h_0]$  we have*

$$(3.68) \quad \mathbb{E} \log \tilde{Z}_{B, \bar{\omega}} - \mathbb{E} \log \tilde{Z}_{A, \bar{\omega}} \geq -h^{2-\gamma} |B \setminus A|,$$

for every  $A$  and  $B$  with  $A \subset B \subset \Lambda_N'$

Let us show that combining Lemma 3.9 and Lemma 3.10 we obtain (3.47) (and the proof of Proposition 3.1 is therefore reduced to proving the two lemmas). First of all just by restricting the expectation to the event  $\mathcal{D}_N$  we have

$$(3.69) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \geq \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N} \right].$$

The next step is to remark that we can replace  $\omega_x$  by  $\bar{\omega}_x$  without changing much the expectation. In fact, by applying Lemma A.3 with  $K = K_h$  we see that for any  $A \subset \Lambda'_N$  we have

$$(3.70) \quad \left| \mathbb{E} \left[ \log \tilde{Z}_{A, \omega} \right] - \mathbb{E} \left[ \log \tilde{Z}_{A, \bar{\omega}} \right] \right| \leq \beta |A| h^3.$$

In particular, using (3.70) for  $A = \Lambda'_N$  we see that it is sufficient to prove (3.47) with  $\mathbb{E} \log \tilde{Z}_{\Lambda'_N, \omega}$  replaced by  $\mathbb{E} \log \tilde{Z}_{\Lambda'_N, \bar{\omega}}$ . We write (recall (3.53))

$$(3.71) \quad \mathbb{E} \left[ \log \tilde{Z}_{\Lambda'_N, \bar{\omega}} \right] = \mathbb{E} \left[ \log \tilde{Z}_{\Lambda''_N, \bar{\omega}} \right] + \mathbb{E} \left[ \log \tilde{Z}_{\Lambda'_N, \bar{\omega}} - \log \tilde{Z}_{\Lambda''_N, \bar{\omega}} \right],$$

and we use Lemma 3.9 and (3.70) to bound the first term and Lemma 3.10 for the second one. So we obtain that for  $\gamma > 0$  and  $h$  sufficiently small (allowed to depend on  $\gamma$ )

$$(3.72) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda'_N} \omega_x(\beta, h) \delta_x^u} \right] \geq N^d (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}) - |\Lambda'_N \setminus \Lambda''_N| h^{2-\gamma}.$$

Now it suffices to recall the choice of the parameters and to choose  $\gamma = 1/2\kappa_d$  and (3.22) allows to conclude that (3.47) holds for some  $\varepsilon > 0$ . More precisely we have

$$(3.73) \quad |\Lambda'_N \setminus \Lambda''_N| \leq d(k+1)LN^{d-1} \leq 2dN^d L^{-1} = 2dN^d h^{\frac{1}{\kappa_d}}.$$

As announced, this means that we are *just* left with proving Lemma 3.9 and Lemma 3.10. This is the content of the next two steps of the proof.

*3.3.5. Step 5: the second moment estimate (proof of Lemma 3.9).* To see that (3.64) follows from (3.63), one simply observes that once  $\phi_0$  is fixed,  $(\delta_x^u)_{x \in \mathcal{B}_L^{(z)}}$  is determined by  $\phi^{(z)}$ . Hence, since  $\mathbf{P}_N^1$  is a product measure (recall the support properties of  $\phi^{(z)}$  and the definition (3.55) of  $\mathcal{B}_L^{(z)}$ ), the expectation can be factorized. Applying (3.63) to each term of the sum thus obtained we have

$$(3.74) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \Lambda''_N} \omega_x(\beta, h) \delta_x^u} \right] \geq |\Lambda''_N| (\chi(\beta) h^2 - C_\beta h^{2+\varepsilon}),$$

and we conclude (modifying the value of  $C_\beta$  if necessary) by using the fact that our choice of parameters implies  $|\Lambda_N''| \geq N^d(1 - h^\varepsilon)$ .

Let us now turn to the important part which is the proof of (3.63). Since the result does not depend on  $z$  let us assume that  $z = 0$  and write  $\mathcal{B}$  for  $\mathcal{B}_L^{(z)}$ . With this notational simplification we remark the splitting:

$$\begin{aligned}
(3.75) \quad \mathbb{E} \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right] &= \\
\mathbb{E} \log \frac{\mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right]}{\mathbf{E} \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right]} &+ \log \mathbf{E} \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta, h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right] \\
&= \mathbb{E} \log \mathbf{E}_N^{1, h} \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta, 0) \delta_x^u} \middle| \mathcal{D}_N^{(0)} \right] + \log \mathbf{E}_N^1 \left[ e^{h \sum_{x \in \mathcal{B}} \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right],
\end{aligned}$$

and in the last step we have introduced the *non disordered pinning* Gibbs measure  $\mathbf{P}_N^{1, h}$  which is the probability with partition function  $\mathbf{E}_N^1[e^{h \sum_{x \in \mathcal{B}} \delta_x^u}]$ . Let us first estimate the last addendum in (3.75). We have

$$\begin{aligned}
(3.76) \quad \log \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}} h \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right] &\geq h \mathbf{E}_N^1 \left[ \sum_{x \in \mathcal{B}} \delta_x^u \middle| \mathcal{D}_N^{(0)} \right] + \log \mathbf{P}_N^1(\mathcal{D}_N^{(0)}) \\
&\geq h \mathbf{E}_N^1 \left[ \left( \sum_{x \in \mathcal{B}} \delta_x^u \right) \mathbf{1}_{\mathcal{D}_N^{(0)}} \right] + \log \mathbf{P}_N^1(\mathcal{D}_N^{(0)}) \\
&= h \mathbf{E}_N^1 \left[ \sum_{x \in \mathcal{B}} \delta_x^u \right] - h \mathbf{E}_N^1 \left[ \left( \sum_{x \in \mathcal{B}} \delta_x^u \right) \mathbf{1}_{(\mathcal{D}_N^{(0)})^c} \right] + \log \mathbf{P}_N^1(\mathcal{D}_N^{(0)}) \\
&\geq L^{2d} \left( 2\chi(\beta)h^2 + O\left(h^2 L^{-1/4}\right) \right) - L^{2d}h^3 - 2h^2 \\
&\geq L^{2d} \left( 2\chi(\beta)h^2 - C_\beta h^{2+\varepsilon} \right),
\end{aligned}$$

where in the step before the last one the three terms correspond to the three terms in the previous line and we have used, in order, (3.48), then (3.58) together with  $\sum_{x \in \mathcal{B}} \delta_x^u \leq L^{2d}$  and finally again (3.58).

We are left with estimating first addendum in (3.75). Note that since  $\mathbf{P}_N^{1, h}(\cdot | \mathcal{D}_N^{(0)})$  is a small modification of  $\mathbf{P}_N^1$ , the probability of making one

contact in  $\mathcal{B}$  is small also under the former measure. In fact

$$\begin{aligned}
(3.77) \quad \mathbf{P}_N^{1,h} \left( \sum_{x \in \mathcal{B}} \delta_x^u \geq 1 \mid \mathcal{D}_N^{(0)} \right) &= \frac{\mathbf{E}_N^1 \left[ e^{h \sum_{x \in \mathcal{B}} \delta_x^u} \mathbf{1}_{\{\sum_{x \in \mathcal{B}} \delta_x^u \geq 1\}} \cap \mathcal{D}_N^{(0)} \right]}{\mathbf{E}_N^1 \left[ e^{h \sum_{x \in \mathcal{B}} \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(0)}} \right]} \\
&\leq \frac{e^{h\kappa_d} \mathbf{P}_N^1 \left( \{\sum_{x \in \mathcal{B}} \delta_x^u \geq 1\} \cap \mathcal{D}_N^{(0)} \right)}{\mathbf{P}_N^1 \left( \mathcal{D}_N^{(0)} \right)} \\
&\leq (1 + 2\kappa_d h) \mathbf{P}_N^1 \left( \left\{ \sum_{x \in \mathcal{B}} \delta_x^u \geq 1 \right\} \cap \mathcal{D}_N^{(0)} \right) \\
&\leq (1 + 2\kappa_d h) \mathbf{P}_N^1 \left( \sum_{x \in \mathcal{B}} \delta_x^u \geq 1 \right) \\
&\leq (1 + 2\kappa_d h) \sum_{x \in \mathcal{B}} \mathbf{E}_N^1 [\delta_x^u] \leq C_1(\beta) L^{2d} h,
\end{aligned}$$

where in the second inequality we used (3.58) and in the last one we have used (3.48) and  $C_1(\beta) := 4\chi(\beta)$ . As a consequence we have for every  $\omega$

$$(3.78) \quad \mathbf{E}_N^{1,h} \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta,0) \delta_x^u} \mid \mathcal{D}_N^{(0)} \right] \geq 1 - C_1(\beta) L^{2d} h.$$

Hence using the formula  $\log(1+y) \geq y - \frac{1}{2(1-\eta)} y^2$ , valid for all  $y \geq -\eta$  and  $\eta \in (0, 1)$ , with  $\eta = C_1(\beta) L^{2d} h$  we obtain that

$$\begin{aligned}
(3.79) \quad \mathbb{E} \log \mathbf{E}_N^{1,h} \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta,0) \delta_x^u} \mid \mathcal{D}_N^{(0)} \right] \\
\geq -\frac{1}{2(1 - C_1(\beta) L^{2d} h)} \left( \mathbb{E} \left( \mathbf{E}_N^1 \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta,0) \delta_x^u} \mid \mathcal{D}_N^{(0)} \right]^2 \right) - 1 \right) \\
\geq -\frac{1}{2} \left( 1 + 2C_1(\beta) L^{2d} h \right) \left( \left( \tilde{\mathbf{E}}_N^1 \right)^{\otimes 2} \left[ e^{\sum_{x \in \mathcal{B}} (\lambda(2\beta) - 2\lambda(\beta)) \delta_x^{(1)} \delta_x^{(2)}} \right] - 1 \right),
\end{aligned}$$

where the last inequality it is simply the fact that  $1 + 2x \geq 1/(1-x)$  for  $x \in [0, 1/2]$ , so we are assuming  $C_1(\beta) L^{2d} h \leq 1/2$ , and we have introduced the notation

$$(3.80) \quad \tilde{\mathbf{P}}_N^1(\cdot) := \mathbf{P}_N^1(\cdot \mid \mathcal{D}_N^{(0)}).$$

The notation  $\delta_x^{(i)}$ ,  $i = 1, 2$  denote the set of contact point for the two marginals  $\phi^{(i)}$ ,  $i = 1, 2$  of the product measure  $\left( \tilde{\mathbf{P}}_N^1 \right)^{\otimes 2}$ . Now taking advan-



tage of the fact that, because of the restriction to  $\mathcal{D}_N^{(0)}$ , under  $(\tilde{\mathbf{P}}_N^1)^{\otimes 2}$  we have  $\sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \leq \kappa_d$ , we deduce that

$$\begin{aligned}
(3.81) \quad & \left(\tilde{\mathbf{E}}_N^1\right)^{\otimes 2} \left[ e^{\sum_{x \in \mathcal{B}} (\lambda(2\beta) - 2\lambda(\beta)) \delta_x^{(1)} \delta_x^{(2)}} \right] \leq 1 - \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 1 \right) + \\
& e^{\lambda(2\beta) - 2\lambda(\beta)} \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} = 1 \right) + e^{\kappa_d(\lambda(2\beta) - 2\lambda(\beta))} \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 2 \right) \\
& = 1 + \left( e^{\lambda(2\beta) - 2\lambda(\beta)} - 1 \right) \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} = 1 \right) \\
& \quad + \left( e^{\kappa_d(\lambda(2\beta) - 2\lambda(\beta))} - 1 \right) \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 2 \right).
\end{aligned}$$

In the following lemma (whose proof we postpone), we compute a sharp upper bound for the second term in the sum, and show that the third one is negligible.

LEMMA 3.11. *Set  $\varepsilon = 1/(4\kappa_d)$ . There exists  $h_0 = h_0(\beta, d)$  such that for  $h \in (0, h_0(\beta, d))$  and  $\phi_0 \in \mathcal{C}_N$  we have*

$$(3.82) \quad \left(\tilde{\mathbf{E}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \right) \leq ((2\chi(\beta)h)^2 + h^{2+\varepsilon}) L^{2d},$$

and

$$(3.83) \quad \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 2 \right) \leq h^{2+\varepsilon} L^{2d}.$$

The application of Lemma 3.11 to (3.81) yields (for adequate choice of constants)

$$\begin{aligned}
(3.84) \quad & \left(\tilde{\mathbf{E}}_N^1\right)^{\otimes 2} \left[ e^{\sum_{x \in \mathcal{B}} (\lambda(2\beta) - 2\lambda(\beta)) \delta_x^{(1)} \delta_x^{(2)}} \right] - 1 \leq \\
& \frac{1}{2\chi(\beta)} \left( (2\chi(\beta)h)^2 + Ch^{2+\varepsilon} \right) L^{2d} - Ch^{2+\varepsilon} \left( e^{\kappa_d(\lambda(2\beta) - 2\lambda(\beta))} - 1 \right) L^{2d} \\
& = (2\chi(\beta)h^2 - C'h^{2+\varepsilon}) L^{2d},
\end{aligned}$$

and going back to (3.79) we get (for a different constant  $C$ )

$$(3.85) \quad \mathbb{E} \log \mathbf{E}_N^{1,h} \left[ e^{\sum_{x \in \mathcal{B}} \omega_x(\beta,0) \delta_x^u} \middle| \mathcal{D}_N^{(0)} \right] \geq -(\chi(\beta)h^2 - Ch^{2+\varepsilon})L^{2d},$$

and we are done with the proof of Lemma 3.9.  $\square$

*Proof of Lemma 3.11.* The proof relies essentially on controlling the first and second moment of the sum  $\sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)}$ . For what concerns (3.82) we observe that

$$(3.86) \quad \left( \tilde{\mathbf{E}}_N^1 \right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \right) = \sum_{x \in \mathcal{B}} \left( \tilde{\mathbf{E}}_N^1 [\delta_x^u] \right)^2,$$

and that

$$(3.87) \quad \begin{aligned} \tilde{\mathbf{E}}_N^1 [\delta_x^u] &= \frac{\mathbf{E}_N^1 \left[ \delta_x^u \mathbf{1}_{\mathcal{D}_N^{(0)}} \right]}{\mathbf{P}_N^1 \left( \mathcal{D}_N^{(0)} \right)} \leq (1 + 2h^2) \mathbf{E}_N^1 [\delta_x^u] \leq (1 + 2h^2) \left( 2\chi(\beta)h + hL^{-1/4} \right) \\ &\leq 2\chi(\beta)h + h^{1+b}, \end{aligned}$$

and in the last step  $b$  is any positive number smaller than  $1/(4\kappa_d)$  and the step holds for  $h$  smaller than a constant that depends on  $\beta$ ,  $d$  and on the choice of  $b$ . Therefore the square of this last expression is bounded by  $(2\chi(\beta)h)^2 + h^{2+b}$ , for example  $b = 1/(5\kappa_d)$  for  $h$  smaller than a constant that depends only on  $\beta$  and  $d$ . The proof of (3.82) is therefore complete.

For what concerns (3.83) we start like for (3.82), that is we get rid of the conditioning with respect to  $\mathcal{D}_N^{(0)}$  and then we proceed with a union bound:

$$(3.88) \quad \begin{aligned} \left( \tilde{\mathbf{P}}_N^1 \right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 2 \right) &\leq (1 + 2h^2)^2 \left( \mathbf{P}_N^1 \right)^{\otimes 2} \left( \sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 2 \right) \\ &\leq (1 + 2h^2)^2 \sum_{x,y \in \mathcal{B}: x \neq y} \left( \mathbf{P}_N^1 \right)^{\otimes 2} \left( \delta_x^{(1)} \delta_x^{(2)} \delta_y^{(1)} \delta_y^{(2)} = 1 \right) \\ &= (1 + 2h^2)^2 \sum_{x,y \in \mathcal{B}: x \neq y} \mathbf{P}_N^1 \left( \delta_x^u \delta_y^u = 1 \right)^2 \\ &\leq (1 + 2h^2)^2 L^{4d} \max_{x \neq y} \mathbf{P}_N^1 \left( \delta_x^u \delta_y^u = 1 \right)^2. \end{aligned}$$

We proceed by observing that (recall (3.43) and (3.49))

$$(3.89) \quad \begin{aligned} \mathbf{P}_N^1(\delta_x^u \delta_y^u = 1) &\leq \mathbf{P}_N^1(\phi_1(x) \geq u - 1 - \phi_0(x), \phi_1(y) \geq u - 1 - \phi_0(y)) \\ &\leq \mathbf{P}_N^1(\phi_1(x) + \phi_1(y) \geq 2u - 3). \end{aligned}$$

This is just a Gaussian tail estimate for a centered Gaussian variable with variance equal to  $\text{Var}(\phi(x) + \phi(y) - \phi_0(x) - \phi_0(y)) = \text{Var}(\phi(x) + \phi(y)) - \text{Var}(\phi_0(x) + \phi_0(y))$  and, for  $x \neq 0$ , it is therefore smaller than  $\text{Var}(\phi(x) + \phi(y)) = 2(1 + p(d))\sigma_d^2$ , where  $p(d) \in (0, 1)$  is the probability that the simple random walk, starting from the origin, revisits the origin. The constant  $p(d)$  has an expression in terms of an integral involving a Bessel function (see for example [18, Section 5.9]), in particular  $p(d)$  decreases to zero as  $d$  becomes large: here we will just use  $p(d) \leq p(3) = 0.3405\dots < 7/20$ . Hence we are left with estimating

$$(3.90) \quad \mathbf{P}_N^1(\delta_x^u \delta_y^u = 1) \leq \mathbf{P}\left(\sqrt{2(1 + p(d))}\sigma_d \mathcal{N} \geq 2u - 3\right),$$

and (3.28), along with the fact that  $u_h$  is asymptotically equivalent to  $\sigma_d \sqrt{2 \log(1/h)}$ , readily yields that

$$(3.91) \quad \mathbf{P}_N^1(\delta_x^u \delta_y^u = 1) \leq h^{1+b},$$

for every  $b \in (0, (1 - p(d))/(1 + p(d)))$ , and  $h$  sufficiently small. Going back to (3.88) we conclude that

$$(3.92) \quad \left(\tilde{\mathbf{P}}_N^1\right)^{\otimes 2} \left(\sum_{x \in \mathcal{B}} \delta_x^{(1)} \delta_x^{(2)} \geq 2\right) \leq L^{4d} h^{2+2b} = L^{2d} h^{2+2b-2d/\kappa_d},$$

for every  $b \in (0, (1 - p(d))/(1 + p(d)))$  and  $h$  sufficiently small. Since  $(1 - x)/(1 + x) > (1 - 2x)$  for every  $x > 0$  we can choose  $b = 1 - 2p(d) > 1 - 7/10 = 3/10$ . Since  $2b - 2d/\kappa_d > 3/5 - 2d/\kappa_d \geq 1/10 > 1/(4\kappa_d)$  (for the last inequality and the one before the last one, recall that  $\kappa_d$  is chosen to be at least  $4d$ ) and this completes the proof of Lemma 3.11.  $\square$

**3.3.6. Step 6: the proof of Lemma 3.10.** Of course it is sufficient to prove the result for  $B = A \cup \{x\}$ , because then the general results can be obtained by adding vertices one by one. We just have to prove then that

$$(3.93) \quad \mathbb{E} \log \tilde{\nu}^{A, \bar{\omega}} \left( e^{(\beta \bar{\omega}_x - \lambda(\beta) + h)\delta_x^u} \right) \geq -Ch^{2-\gamma}.$$

where  $\tilde{\nu}^{A, \bar{\omega}}$  denotes to the distribution of  $(\phi^{(z)})_{z \in \mathbb{Z}^d}$  associated with the partition function  $\tilde{Z}_{A, \bar{\omega}}$ , see (3.65): note that it depends only on  $(\bar{\omega}_y)_{y \in A}$ .

We are going to show a stronger statement than (3.93): in fact we are going to show that (3.93) holds also if we average only with respect to  $\bar{\omega}_x$  (we use the notation  $\mathbb{E}_x$ ) and we freeze the realization of  $(\bar{\omega}_y)_{y \neq x}$ . Setting  $p_{A, \bar{\omega}} = p_{A, \bar{\omega}, \beta, h} := \tilde{\nu}^{A, \bar{\omega}}(\delta_x^u = 1)$  – it is a random variable measurable with respect to  $(\bar{\omega}_y)_{y \in A}$  – we have

$$(3.94) \quad \begin{aligned} \mathbb{E}_x \log \tilde{\nu}^{A, \bar{\omega}} \left[ e^{(\beta \bar{\omega}_x - \lambda(\beta) + h) \delta_x^u} \right] &= \mathbb{E}_x \log \left( 1 + p_{A, \bar{\omega}} \left( e^{\beta \bar{\omega}_x - \lambda(\beta) + h} - 1 \right) \right) \\ &\geq -p_{A, \bar{\omega}}^2 \left( e^{\lambda(2\beta) - 2\lambda(\beta)} - 1 \right), \end{aligned}$$

where in the last step holds if  $p_{A, \bar{\omega}} \leq 1/2$ , so we can use the inequality  $\log(1 + y) \geq y - y^2$  that holds for  $y \geq -1/2$ : in fact  $p_{A, \bar{\omega}} \leq 1/2$  yields  $p_{A, \bar{\omega}}(e^{\beta \bar{\omega}_x - \lambda(\beta)} - 1) \geq -1/2$  for every value of  $\bar{\omega}$  and  $\beta$ .

Hence we are reduced to proving that for every  $\beta, d$  and every  $\gamma > 0$  there exists  $h_0$  such that for every finite subset  $A$  of  $\mathbb{Z}^d$ , every  $h \in (0, h_0)$ , every  $\phi_0 \in \mathcal{C}_N$  and every realization of  $\bar{\omega}$

$$(3.95) \quad p_{A, \bar{\omega}} \leq h^{1-\gamma/2}.$$

Recalling (3.48), the strategy is now to show that  $\tilde{\nu}^{A, \bar{\omega}}$  is not too different from  $\mathbf{P}_N^1$ . And, since  $\phi_0$  is fixed,  $\phi(x)$  is determined by the realization  $\phi^{(z)}$  for at most  $2^d$  values of  $z$ . We will prove that conditioned to all the rest, the marginal distribution of  $\phi^{(z)}$  has a bounded Radon-Nikodym derivative. Namely (recall that  $|\bar{\omega}_x| \leq K_h$ , with  $K_h$  given in (3.66)):

LEMMA 3.12. *For every  $z, A$  and every measurable subset  $B$  of  $\mathbb{R}^{\mathbb{Z}^d}$  we have that when  $\phi^{(y)} \in \mathcal{D}_N^{(y)}$  for  $y \neq z$ ,*

$$(3.96) \quad \tilde{\nu}^{A, \bar{\omega}} \left( \phi^{(z)} \in B \mid \phi^{(y)}, y \neq z \right) \leq 2e^{\beta K_h 3^{d+1} \kappa_d} \mathbf{P}_N^1 \left( \phi^{(z)} \in B \right).$$

We want to apply Lemma 3.12 to estimate from above  $\tilde{\nu}^{A, \bar{\omega}}(\delta_x^u = 1)$  and for this we observe that the event  $\delta_x^u = 1$  relies only on the realization of finitely many  $\phi^{(z)}$ . More precisely

$$(3.97) \quad \left\{ (\phi^{(z)})_{z \in \mathbb{Z}^d} : \delta_x^u = 1 \right\} = \left\{ (\phi^{(z)})_{z \in \mathbb{Z}^d} : \sum_{z \in I_x} \phi^{(z)}(x) \in [u-1, u+1] - \phi_0(x) \right\},$$

where the set  $I_x := \{z : x \in \text{Supp}(\phi^{(z)})\}$ , contains at least one point and at most  $2^d$ . An immediate consequence of Lemma 3.12 is that the Radon-Nykodym derivative with respect to  $\mathbf{P}_N^1$  of the distribution of  $(\phi^{(z)})_{z \in I_x}$  under  $\tilde{\nu}^{A, \bar{\omega}}$  is bounded above by

$$(3.98) \quad (2e^{\beta K_h 3^{d+1} \kappa_d})^{|I_x|} \leq (2e^{\beta K_h 3^{d+1} \kappa_d})^{2^d}$$

As the event in (3.97) depends only on  $(\phi^{(z)})_{z \in I_x}$  we have in particular (recall that  $K_h = o(\log(1/h))$ )

$$(3.99) \quad \begin{aligned} \tilde{\nu}^{A, \bar{\omega}}(\delta_x^u = 1) &\leq (2e^{\beta K_h 3^{d+1} \kappa_d})^{2^d} \mathbf{P}_N^1(\delta_x^u = 1) \\ &\leq (2e^{\beta K_h 3^{d+1} \kappa_d})^{2^d} 3\chi(\beta)h \leq h^{-\gamma/2}h, \end{aligned}$$

where we have applied (3.48) in the inequality before the last one.  $\square$

PROOF OF LEMMA 3.12. We recall that

$$(3.100) \quad \tilde{\mathcal{B}}_z = (L^2 + L)z + \llbracket 1, L^2 + 2L - 1 \rrbracket^d = Mz + \llbracket 1, M + L - 1 \rrbracket^d.$$

is the support of  $\phi^{(z)}$ . Now the conditional Radon-Nikodym derivative given  $\phi^{(y)}$ ,  $y \neq z$  is equal to

$$(3.101) \quad \frac{e^{\sum_{x \in \tilde{\mathcal{B}}_z \cap A} (\beta \bar{\omega}_x - \lambda(\beta) + h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(z)}}}{\mathbf{E}_N^{1, (z)} \left[ e^{\sum_{x \in \tilde{\mathcal{B}}_z \cap A} (\beta \bar{\omega}_x - \lambda(\beta) + h) \delta_x^u} \mathbf{1}_{\mathcal{D}_N^{(z)}} \right]}$$

where the superscript  $(z)$  in the expectation underlines that the average is taken only w.r.t.  $\phi^{(z)}$ . We consider the sum over  $\tilde{\mathcal{B}}_z \cap A$  and not  $A$  because terms coming from  $x \in (\tilde{\mathcal{B}}_z)^c$  are completely determined by  $\phi^{(y)}$ ,  $y \neq z$  and therefore cancel out in the numerator and denominator.

Because of the restriction to  $\mathcal{D}_N$  there are at most  $3^d \kappa_d$  contacts in  $\tilde{\mathcal{B}}_z$ . Indeed if  $\delta_x^u = 1$  for some  $x \in \tilde{\mathcal{B}}_z$  then there must exist  $z'$  with  $|z - z'| \leq 1$  for which  $\phi^{(z')}(x) \geq 2^{-d}(u - 2)$  (because otherwise the sum  $\sum_{y \in \mathbb{Z}^d} \phi^{(y)}$  which contains at most  $2^d$  non-zero terms is smaller than  $u - 2$ ). We conclude using the constraint  $\mathcal{D}_N^{(z')}$  and the fact that there are  $3^d$  choices for  $z'$ . Thus using our uniform bound on  $\bar{\omega}$  we have

$$(3.102) \quad \left| \sum_{x \in \tilde{\mathcal{B}}_z \cap A} (\beta \bar{\omega}_x - \lambda(\beta) + h) \delta_x^u \right| \leq 3^d \kappa_d (\beta K_h + |h| + \lambda(\beta)) \leq \frac{1}{2} 3^{d+1} \kappa_d \beta K_h,$$

where the last inequality is valid for  $h$  small enough (recall that  $K_h$  diverges when  $h \rightarrow 0$ ). Using this bound on the numerator and denominator we obtain that the Radon Nykodym is bounded above by

$$(3.103) \quad \frac{1}{\mathbf{P}_N^1(\mathcal{D}_N^{(z)})} e^{3^{d+1} \kappa_d \beta K_h},$$

which yields the desired estimates because  $\mathbf{P}_N^1(\mathcal{D}_N^{(z)}) \geq 1/2$ .  $\square$

**4. Proof of Theorem 1.2: Lower bound on the height.** The main object of this Section is to prove inequality (A) in (1.19) holds. It can be reformulated as follows:

PROPOSITION 4.1. *Given  $\varepsilon > 0$ , for all  $h \leq h_0(\varepsilon)$ , there exists  $c(\varepsilon, h) > 0$  such that almost surely for  $N$  sufficiently large (depending on  $\omega$ ) we have*

$$(4.1) \quad \mathbf{P}_{N,\omega,\beta,h} \left( \sum_{x \in \hat{\Lambda}_N} \mathbf{1}_{\{|\phi(x)| \leq (1-\varepsilon)\sigma_d \sqrt{2 \log(1/h)}\}} \geq \varepsilon N^d \right) \leq e^{-cN^d}.$$

For simplicity we redefine from now till the end, that is the entire Sections 4 and 5, the value of  $u_h$  by keeping only the leading behavior, that is we set

$$(4.2) \quad u_h := \sigma_d \sqrt{2 \log(1/h)}.$$

Recalling (3.23)-(3.25), we see that the newly defined  $u_h$  coincides to first order with the one used in Section 3.

The proof of Proposition 4.1 is divided into three steps. To each step is devoted one of the subsections that follow.

4.1. *Step 1: Upper bound on the contact fraction.* The first important ingredient to prove Proposition 4.1 is a quantitative control on the contact density under  $\mathbf{P}_{N,\omega,\beta,h}$ . We know the contact density is close to the optimal density  $p_h$  when  $h \searrow 0$ , and in this limit  $p_h \sim 2\chi(\beta)h$  (see (2.9)). But we can extract from Theorem 1.1 also upper and lower Large Deviations estimates: for our arguments we just need a control from above, and this is what we are going to develop next. Given  $\eta > 0$  we set

$$(4.3) \quad B_{N,\eta} := \left\{ \phi \in \mathbb{R}^{\hat{\Lambda}_N} : \frac{1}{N^d} \sum_{x \in \hat{\Lambda}_N} \delta_x \leq (2\chi(\beta) + \eta) h \right\}.$$

Then we have the following:

LEMMA 4.2. *For every  $\eta > 0$ , there exists constant  $c(\beta, \eta) > 0$  and  $h_0$  such that for all  $h \in (0, h_0)$ , we have almost-surely for  $N$  sufficiently large*

$$(4.4) \quad \mathbf{P}_{N,\omega,\beta,h} \left( B_{N,\eta}^{\mathbb{C}} \right) \leq e^{-ch^2 N^d}.$$

REMARK 4.3. *Actually we only need to show the result for one positive value of  $\eta$ , no need to choose it arbitrarily small. In fact, we are going to apply Lemma 4.2 with  $\eta = \chi(\beta)$ . Nevertheless, we feel that a precise result gives more intuition about the proof.*

PROOF. Given an event  $A$ , we use the notation  $Z_{N,\omega,\beta,h}(A)$  for the partition function restricted to the set  $A$  that is

$$(4.5) \quad Z_{N,\omega,\beta,h}(A) = \mathbf{P}_{N,\omega,\beta,h}(A) Z_{N,\omega,\beta,h}.$$

For every  $v > 0$ , using the convexity of  $v \mapsto \log Z_{N,\omega,\beta,h+v} \left( B_{N,\eta}^{\mathbb{C}} \right)$  and the fact that its derivative at the origin is

$$(4.6) \quad \mathbf{E}_{N,\omega,\beta,h} \left[ \sum_{x \in \dot{\Lambda}_N} \delta_x \mid B_{N,\eta}^{\mathbb{C}} \right] \geq (2\chi(\beta) + \eta) h N^d,$$

we have

$$(4.7) \quad \log \frac{Z_{N,\omega,\beta,h+v}}{Z_{N,\omega,\beta,h}} \geq \log \frac{Z_{N,\omega,\beta,h+v} \left( B_{N,\eta}^{\mathbb{C}} \right)}{Z_{N,\omega,\beta,h}} \\ \geq (2\chi(\beta) + \eta) h v N^d + \log \mathbf{P}_{N,\omega,\beta,h} \left( B_{N,\eta}^{\mathbb{C}} \right),$$

where we have used that

$$(4.8) \quad \mathbf{P}_{N,\omega,\beta,h} \left( B_{N,\eta}^{\mathbb{C}} \right) = \frac{Z_{N,\omega,\beta,h} \left( B_{N,\eta}^{\mathbb{C}} \right)}{Z_{N,\omega,\beta,h}}.$$

Dividing by  $N^d$  and taking the limit in (4.7) we obtain that  $\mathbb{P}(d\omega)$ -a.s.

$$(4.9) \quad \limsup_{N \rightarrow \infty} \frac{1}{N^d} \log \mathbf{P}_{N,\omega,\beta,h} \left( B_{N,\eta}^{\mathbb{C}} \right) \leq \mathbb{F}(\beta, h+v) - \mathbb{F}(\beta, h) - (2\chi(\beta) + \eta) h v.$$

Choose now  $v = bh$  with  $b := \eta/(2\chi(\beta))$ . By applying the precise asymptotic results of Theorem 1.1 we obtain, in the small  $h$  limit

$$(4.10) \quad \begin{aligned} & \mathbb{F}(\beta, h + v) - \mathbb{F}(\beta, h) - (2\chi(\beta) + \eta)hv \\ &= h^2 (\chi(\beta)b^2 - \eta b) + o(h^2) = -h^2 \frac{\eta^2}{4\chi(\beta)} (1 + o(1)), \end{aligned}$$

which completes the proof of Lemma 4.2.  $\square$

*4.2. Step 2: lower bound on harmonic averages.* We introduce a length  $L \in \mathbb{N}$  sufficiently large (how large is specified below: we stress that this length  $L$  does not depend on  $h$  and it is unrelated to its homonym of Section 3, see (3.22)) and divide  $\Lambda_N$  into disjoint cubes of edge length  $L$ . We introduce the disjoint cubes  $C_L^z := \llbracket 0, L \rrbracket^d + zL$  and their *centers*  $x_L(z) := zL + (1, \dots, 1)\lfloor L/2 \rfloor$ . We choose  $L$  so that the variance of a zero-boundary free field on  $C_L^z$  at the center of the cube is close to variance of the infinite volume field (recall (1.3))

$$(4.11) \quad \sigma_{d,L}^2 := G_L(x_L(0), x_L(0)) \geq \sigma_d^2(1 - \varepsilon/2).$$

We consider only  $z \in \llbracket 0, \lfloor N/L \rfloor - 1 \rrbracket^d =: I_{N,L}$  meaning that we consider cubes for which  $C_L^z \subset \Lambda_N$ . We let  $H_L(z)$  denote the harmonic average, at the center  $x_L(z)$  of the cube  $C_L^z$ , of the field on the boundary of  $C_L^z$

$$(4.12) \quad H_L(z) := \sum_{x \in \partial C_L^z} p_{L,z}(x) \phi(x).$$

where  $\partial C_L^z := C_L^z \setminus (\llbracket 1, L - 1 \rrbracket^d + zL)$  is the internal boundary of  $C_L^z$  and  $p_{L,z}(x)$  is the probability that a simple symmetric random walk issued from the center  $x_L(z)$  hits  $\partial C_L^z$  at  $x$ .

We are going to show that for most  $z$ 's,  $H_L(z)$  lies above height  $(1 - \varepsilon/2)u_h$ . We introduce the event

$$(4.13) \quad F_{N,\varepsilon} := \left\{ \phi \in \mathbb{R}^{\Lambda_N} : |\Xi_{N,L}(\phi)| \leq \frac{\varepsilon}{2L^d} N^d \right\},$$

where

$$(4.14) \quad \Xi_{N,L}(\phi) := \{z \in I_{N,L} : |H_L(z)| \leq (1 - \varepsilon/2)u_h\},$$

is a random subset of  $I_{N,L}$ .



LEMMA 4.4. *Given  $\varepsilon > 0$ , there exists  $h_0(\varepsilon)$  such for all  $h \in (0, h_0)$ , there exists  $c(\varepsilon, h) > 0$  for which for all  $N$  sufficiently large we have*

$$(4.15) \quad \mathbf{P}_{N,\omega,\beta,h} \left( F_{N,\varepsilon}^{\mathbb{C}} \right) \leq e^{-cN^d}.$$

PROOF. Using Lemma 4.2 we can in fact look only at the probability of  $F_{N,\varepsilon}^{\mathbb{C}} \cap B_{N,\eta}$  for some arbitrary value of  $\eta$ . We choose  $\eta = \chi(\beta)$  and simply denote the corresponding event by  $B_N$ . Now we have for any event  $A$  and  $h > 0$

$$(4.16) \quad \begin{aligned} \mathbb{E} [\mathbf{P}_{N,\omega,\beta,h} (A \cap B_N)] &= \mathbb{E} \left[ \frac{Z_{N,\omega,\beta,h} (A \cap B_N)}{Z_{N,\omega,\beta,h}} \right] \\ &\leq \mathbb{E} [Z_{N,\omega,\beta,h} (A \cap B_N)] / \mathbf{P}_N \left( \forall x \in \mathring{\Lambda}_N, \phi(x) > 1 \right) \\ &= \mathbf{E}_N \left[ e^{h \sum_{x \in \mathring{\Lambda}_N} \delta_x} \mathbf{1}_{A \cap B_N} \right] / \mathbf{P}_N \left( \forall x \in \mathring{\Lambda}_N, \phi(x) > 1 \right), \end{aligned}$$

where we have simply bounded the denominator by the contribution of trajectories with no contacts. Given that the probability in the denominator behaves sub-exponentially in the volume  $N^d$  (recall the entropic repulsion estimates discussed below (1.9)) and that the number of contact is bounded above by  $3\chi(\beta)hN^d$ , we have for  $N$  sufficiently large

$$(4.17) \quad \mathbb{E} [\mathbf{P}_{N,\omega,\beta,h} (A \cap B_N)] \leq e^{4\chi(\beta)h^2N^d} \mathbf{P}_N (A \cap B_N).$$

Hence to prove (4.15) it is sufficient to show that

$$(4.18) \quad \mathbf{P}_N \left( F_{N,\eta,\varepsilon}^{\mathbb{C}} \cap B_N \right) \leq e^{-5\chi(\beta)h^2N^d},$$

and then apply the Borel-Cantelli via a Markov inequality bound. We are in fact going to show that for any realization of  $H_L$  for which  $|\Xi_{N,L}(\phi)| > \frac{\varepsilon}{2L^2}N^d$ ,

$$(4.19) \quad \mathbf{P}_N (B_N \mid H_L(z), z \in I_{N,L}) \leq e^{-5\chi(\beta)h^2N^d}.$$

The Markov property for the LFF states that under  $\mathbf{E}_N$  the random variables  $\phi(x_L(z)) - H_L(z)$  are IID Gaussian variables with variance  $\sigma_{d,L}^2$ , which are independent of  $(H_L(z))_{z \in I_{N,L}}$ . In particular, conditioned to  $H_L$ ,  $(\delta_{x_L(z)})_{z \in [0, \lfloor N/L \rfloor - 1]^d}$  are independent Bernoulli variables with respective parameters

$$(4.20) \quad q_z(H_L) := \frac{1}{\sqrt{2\pi}} \int_{-1}^1 \exp \left( -\frac{(t - H_L(z))^2}{2\sigma_{d,L}^2} \right) dt.$$

Note in particular that the above expression is decreasing in  $|H(z)|$  and using also (4.11) we see that for  $z \in \Xi_{N,L}$

$$(4.21) \quad \begin{aligned} q_z(H_L) &\geq \frac{1}{\sqrt{2\pi}} \int_{-1}^1 \exp\left(-\frac{(t - (1 - \varepsilon/2)u_h)^2}{2\sigma_{d,L}^2}\right) dt \\ &\geq P(\sigma_d \mathcal{N} \in [a_\varepsilon u_h - 1/a_\varepsilon, a_\varepsilon u_h + 1/a_\varepsilon]) =: \bar{q}(h, d, L, \varepsilon), \end{aligned}$$

where in the intermediate step we used  $a_\varepsilon := (1 - \varepsilon/2)^{1/2}$ . Replacing  $u_h$  by its value (4.2) and using (3.28) in a rough way, we obtain that  $\bar{q} \geq h^{1-\varepsilon/2}$ , at least for  $h$  sufficiently small.

Hence in particular, considering within  $\mathring{\Lambda}_N$  only the points of the form  $x_L(z)$  with  $z \in \Xi_{N,L}$  we obtain that, conditioned to  $H$ , when  $|\Xi_{N,L}| > \frac{\varepsilon}{2L^2} N^d$ , the quantity  $\sum_{x \in \mathring{\Lambda}_N} \delta_x$  stochastically dominates a binomial random variable of parameters  $\bar{q}$  and  $\lceil \varepsilon/(2L^2) \rceil N^d$  (we then omit the integer part for notational convenience). We have thus

$$(4.22) \quad \mathbf{P}_N(B_N \mid H_L(z), z \in I_{N,L}) \leq P\left(\text{Bin}(\varepsilon N^d/(2L^2), h^{1-\varepsilon/2}) > 3\chi(\beta)hN^d\right),$$

where  $\text{Bin}(n, p)$  denotes a binomial random variable of parameters  $n$  and  $p$ . By the first inequality in Lemma A.1 applied with  $n = \varepsilon N^d/(2L^2)$  and  $\Delta = 6\chi(\beta)h^{\varepsilon/2}L^d/\varepsilon$ , it is just a matter of choosing  $h$  suitably small to get to

$$(4.23) \quad \mathbf{P}_N(B_N \mid \bar{\phi}(z), z \in I_{N,L}) \leq \exp\left(-\frac{\varepsilon h^{1-\varepsilon/2}}{4L^d} N^d\right),$$

which largely proves (4.19).  $\square$

*4.3. Step 3: positive density of low sites is incompatible with harmonic average lower bound.* Now let us consider the event whose probability we wish to bound in Proposition 4.1 which is

$$(4.24) \quad A_{N,\varepsilon} := \left\{ \phi \in \mathbb{R}^{\mathring{\Lambda}_N} : \sum_{x \in \mathring{\Lambda}_N} \mathbf{1}_{\{|\phi(x)| \leq (1-\varepsilon)u_h\}} \geq \varepsilon N^d \right\}.$$

Using Lemma 4.2 and Lemma 4.4, it is sufficient to prove that the probability of  $A_{N,\varepsilon} \cap B_N \cap F_{N,\varepsilon}$  decays exponentially with the volume. From (4.17) (with  $A = A_{N,\varepsilon} \cap F_{N,\varepsilon}$ ) we deduce that

$$(4.25) \quad \mathbb{E}[\mathbf{P}_{N,\omega,\beta,h}(A_{N,\varepsilon} \cap B_N \cap F_{N,\varepsilon})] \leq e^{4\chi(\beta)h^2 N^d} \mathbf{P}_N(A_{N,\varepsilon} \cap B_N \cap F_{N,\varepsilon}).$$

Hence we are left with showing that

$$(4.26) \quad \mathbf{P}_N(A_{N,\varepsilon} \cap F_{N,\varepsilon}) \leq e^{-5\chi(\beta)h^2N^d}.$$

Now to conclude we need observe that on the event  $A_{N,\varepsilon} \cap F_{N,\varepsilon}$  the Hamiltonian  $H_N(\phi)$  is anomalously large. Indeed, on  $A_{N,\varepsilon}$ , we have necessarily

$$(4.27) \quad \left| \left\{ z \in \llbracket 0, (N/L) - 1 \rrbracket : \exists x \in zL + \llbracket 1, L \rrbracket^d, |\phi(x)| \leq (1 - \varepsilon)u_h \right\} \right| \geq \varepsilon \frac{3N^d}{4L^d},$$

and hence on  $A_{N,\varepsilon} \cap F_{N,\varepsilon}$  we have

$$(4.28) \quad \left| \left\{ z : |H(z)| \geq (1 - \varepsilon/2)u_h \text{ and } \exists x \in zL + \llbracket 1, L \rrbracket^d, |\phi(x)| \leq (1 - \varepsilon)u_h \right\} \right| \geq \varepsilon \frac{N^d}{4L^d}.$$

Note that for each  $z$  which satisfies the property in the right-hand side of (4.28) we can find  $x_1 \in zL + \llbracket 1, L \rrbracket^d$  such that  $|\phi(x_1)| \leq (1 - \varepsilon)u_h$ , and since  $H_L(z)$  is a weighted average of the values of  $\phi$  on the boundary of  $zL + \llbracket 0, L \rrbracket^d$ , there exists  $x_2$  in the boundary of  $zL + \llbracket 0, L \rrbracket^d$  such that  $|\phi(x_2)| \geq (1 - \varepsilon/2)u_h$ .

Considering a path of minimal length (which in this case has to be smaller than  $dL$ ) between  $x_1$  and  $x_2$ , we obtain that there exists a pair of neighbors  $y_1 \in zL + \llbracket 0, L \rrbracket^d$   $y_2 \in zL + \llbracket 1, L \rrbracket^d$ , such that

$$(4.29) \quad |\phi(x_1) - \phi(x_2)| \geq \frac{\varepsilon}{2dL}u_h.$$

Given that these edges are necessarily distinct for different values of  $z$ , we obtain that

$$(4.30) \quad A_{N,\varepsilon} \cap F_{N,\varepsilon} \subset \left\{ \phi \in \mathbb{R}^{\Lambda_N} : \sum_{\substack{\{x,y\} \in \Lambda_N \\ x \sim y}} (\phi(x) - \phi(y))^2 \geq \frac{\varepsilon^3 u_h^2}{16d^2 L^{d+2}} N^d \right\}$$

To conclude, we observe that  $\varepsilon^3 u_h^2 / (16d^2 L^{d+2})$  can be made arbitrarily large by choosing  $h$  small, and use that by [25, (B.8)] one can find  $C > 0$  such that

$$(4.31) \quad \mathbf{P}_N \left( \sum_{\substack{\{x,y\} \in \Lambda_N \\ x \sim y}} (\phi(x) - \phi(y))^2 \geq CN^d \right) \leq e^{-N^d}.$$

Therefore (4.26) holds and the proof of Proposition 4.1 is therefore complete.  $\square$

**5. Proof of Theorem 1.2: Upper bound on the height.** In this Section we prove inequality (B) in (1.19). We keep the definition in the previous section (4.2) for the value of  $u_h$ .

**PROPOSITION 5.1.** *Given  $\varepsilon > 0$ , for all  $h \leq h_0(\varepsilon)$ , there exists  $c = c(\varepsilon, h) > 0$  and  $N_0(\varepsilon, h, \omega)$ , with  $N_0 = N_0(\varepsilon, h, \omega) < \infty$   $\mathbb{P}(d\omega)$ -a.s., such that  $\mathbb{P}(d\omega)$ -a.s.  $N \geq N_0$  we have*

$$(5.1) \quad \mathbf{P}_{N, \omega, \beta, h}(\mathcal{D}(N, \varepsilon)) \leq e^{-cN^d},$$

with

$$(5.2) \quad \mathcal{D}(N, \varepsilon) := \left\{ \sum_{x \in \Lambda_N} \mathbf{1}_{\{|\phi(x)| \geq (1+\varepsilon)u_h\}} \geq \varepsilon N^d \right\}.$$

The proof will be achieved through various steps of which we give here a quick sketch that could be useful as a guideline:

1. Section 5.1: *construction of a hierarchy of (almost) coverings of  $\Lambda_N$ .* To prove Proposition 5.1 we need to exploit the fact that the contact fraction is close to  $p_h := 2h\chi(\beta)$ . However, a statement about the global density like Lemma 4.2 is not sufficient. We need and will show that if we divide  $\Lambda_N$  into boxes (hyper-cubes) – we will call them *level-0 boxes* – of volume roughly  $h^{-2}$ , the empirical contact density in *most of these boxes* is close to  $p_h$ . Such a statement is only about level-0 boxes, but later on in the proof we will need a full hierarchy of boxes. In such hierarchy, level-0 is the lowest level, the highest being the one of  $\Lambda_N$  itself. It is more practical to introduce from the start the full hierarchy even if up to Subsection 5.5 only level-0 is used. For a part of the argument the level-0 boxes will be further split into  $6^d$  sub-boxes that will be called *elementary boxes*.
2. Section 5.2: *control of contact density on elementary boxes.* We will introduce an event  $\mathcal{C}(N, \delta)$ , with  $\delta > 0$  a parameter that is simply going to be chosen proportional to  $\varepsilon$  in the end, on which the field has approximately the correct contact fraction in most of the elementary boxes (that are just a further partition of each of the level-0 boxes into

a finite number, precisely  $6^d$ , boxes). We will show that the  $\mathbf{P}_{N,\omega,\beta,h}$  probability of the complement of  $\mathcal{C}(N, \delta)$  is negligible, in the sense that it is  $O(\exp(-cN^d))$  for some  $c > 0$ .

3. Section 5.3: *the main body of the argument*. We write  $\phi = \psi + \bar{\psi}$ , with  $\psi$  and  $\bar{\psi}$  independent. This decomposition is similar to the one made in Section 3.2:  $\bar{\psi}$  has small variance and contains the long range correlations, while the field  $\psi$  has no correlations when we consider sites that belong to different level-0 boxes. We introduce at this stage one more event, called  $\mathcal{B}(N, \delta)$  that contains the requirements we demand on the  $\bar{\psi}$  and such that the  $\mathbf{P}_{N,\omega,\beta,h}$  probability of the complement of  $\mathcal{B}(N, \delta)$  is  $O(\exp(-cN^d))$ . In a nutshell, what we require on  $\bar{\psi}$  is that it is close to being affine on most of the level-0 boxes and we do this by passing through second order discrete derivatives of  $\phi$ : arguments would have been much more straightforward if we were able to show (with the proper exponential probability estimate) that the field  $\bar{\psi}$  is small or that  $\bar{\psi}$  is almost flat, i.e. its gradient is small: the point is that we would like to get rid of  $\bar{\psi}$  and exploit the independence properties of  $\psi$ . We will however explain why we cannot prove this. Nevertheless, working with *locally affine*  $\bar{\psi}$  will turn out to be sufficient to bound in a satisfactory way the  $\mathbf{P}_{N,\omega,\beta,h}$  probability of  $\mathcal{B}(N, \delta) \cap \mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon)$ . For readability we split this section into two and we devote a separate section to the probability estimates on the  $\psi$  field inside the level-0 boxes.
4. Section 5.4: *Level-0 estimates*. Here we exploit the fact that we are in  $\mathcal{C}(N, \delta)$ , hence with a control on the contact density in most of the level-0 boxes, and in  $\mathcal{B}(N, \delta)$ , hence with a strong control on the long range correlated field  $\bar{\psi}$  (in most of the level-0 boxes). We pick any of the good level-0 boxes and develop geometric arguments, coupled with probability bounds, that show that it is improbable that the absolute value of the field goes above level  $(1 + \varepsilon)u_h$ .
5. Section 5.5: the multiscale bound. This section is devoted to bounding the probability of the complement of  $\mathcal{B}(N, \delta)$  and this is the step in which the multiscale construction introduced in Section 5.1 is exploited.

5.1. *Construction of nested (almost) coverings of  $\Lambda_N$* . Before stating the main result of this subsection we must introduce some notations. As announced, we want to *cover*  $\Lambda_N$  with cubic boxes with volume  $c(\log(1/h)/h)^2$ ,  $c$  a positive constant: we will call them *level-0 boxes*. On top of this we want to construct a hierarchy of boxes: for each  $i \geq 1$  we want to construct boxes

of level  $i$  which are obtained by grouping  $2^d$  disjoint boxes at level  $i - 1$  (meaning that boxes at level  $i$  of the hierarchy will have volume asymptotically equivalent to  $2^{di}c(\log(1/h)/h)^2$ ). Finally, on top of this we require that at each level of the hierarchy, some amount of free space is left between the boxes. We stop the procedure once *we reach*  $\Lambda_N$ . Since we want to cover most of  $\Lambda_N$  with level-0 boxes, in the sense that we want that the fraction of uncovered sites can be made arbitrarily small, it turns out to be more practical for the construction to start from the top level box, that is  $\Lambda_N$ , and work down to when we get to level-0.

This structure will be of fundamental importance for the multiscale analysis introduced in Section 5.5. We introduce it beforehand because the statement about the local contact density presented in Section 5.2 needs to be formulated in terms of level-0 boxes in our hierarchy, but we stress that up to Section 5.5 we are going to need only the level-0 of this construction.

Set  $\tilde{N}_0 := N$  and define  $\tilde{N}_j$  recursively for  $j \geq 1$ . Given  $\varkappa \in (0, 1)$  we set also

$$(5.3) \quad \tilde{N}_j := \left\lfloor \frac{\tilde{N}_{j-1} - 4 \lfloor \tilde{N}_{j-1}^{1-\varkappa} \rfloor}{2} \right\rfloor,$$

and

$$(5.4) \quad J(h, N) := \inf \left\{ j \geq 1 : \tilde{N}_{j+1} \leq 7(\log(1/h)/h)^{2/d} \right\}.$$

The parameter  $\varkappa$  can be chosen arbitrarily in  $(0, 1)$ , for example  $\varkappa = 1/2$ , but this time readability is helped if we do not make the constant explicit. Note also that, if  $h$  is sufficiently small and  $N > 7(\log(1/h)/h)^{2/d}$ , we have

$$(5.5) \quad 7(\log(1/h)/h)^{2/d} \leq \tilde{N}_J \leq 15(\log(1/h)/h)^{2/d}.$$

Then we construct recursively a sequence  $(\tilde{\mathcal{C}}_j)_{0 \leq j \leq J}$  of collections of  $2^{dj}$  disjoint boxes of edge length  $\tilde{N}_j$  within  $\Lambda_N$ .

- We let  $\tilde{\mathcal{C}}_j = \{\tilde{B}_{j,k}, k \in \llbracket 1, 2^{dj} \rrbracket\}$  denote the collection of boxes at step  $j$ . We initiate with  $\tilde{B}_{0,1} := \Lambda_N$  and  $\tilde{\mathcal{C}}_0 := \{\Lambda_N\}$ .
- Now for  $j \geq 1$ , given  $\tilde{B}_{j-1,k_0}$  a generic box in  $\tilde{\mathcal{C}}_{j-1}$  we introduce the  $2^d$  disjoint hypercubes  $\{\tilde{B}_{j,k} : k \in \llbracket 2^d k_0 + 1, 2^d(k_0 + 1) \rrbracket\}$  of edge length  $\tilde{N}_j$  satisfying  $\tilde{B}_{j,k} \subset \tilde{B}_{j-1,k_0}$ . The cubes  $\tilde{B}_{j,k}$  are placed inside  $\tilde{B}_{j-1,k_0}$  as explained in Fig. 3. If we consider the  $d$  hyperplanes that bisect  $\tilde{B}_{j-1,k_0}$  and are orthogonal respectively to  $e_1, e_2, \dots, e_d$  we split

$\tilde{B}_{j-1,k_0}$  into  $2^d$  chambers and the  $2^d$  disjoint hypercubes we have just introduced are simply placed at the center of each chamber.

- We also introduce the *conditioning grid*  $\tilde{G}_j$ : for every  $\tilde{B}_{j-1,k_0}$  we build a portion of the grid by considering the union of the external boundary of  $\tilde{B}_{j-1,k_0}$  and of the portion in  $\tilde{B}_{j-1,k_0}$  of the  $d$  bisecting hyperplanes we have introduced at the previous point (they are the boundaries of the  $2^d$  chambers). We repeat the procedure for each one of the  $2^{d(j-1)}$  hypercubes  $\tilde{B}_{j-1,k_0}$  and, by considering the union of the sets we have constructed we obtain  $\tilde{G}_j$  (that has therefore  $2^{d(j-1)}$  connected components). We add to this collection  $\tilde{G}_0$  which is  $\mathbb{Z} \setminus \dot{\Lambda}$ : we could have added just the external boundary, but this is notationally convenient. In practice, it is more compact to work with the cumulative grid, we define the cumulative grid for  $j = 0, 1, \dots$  by

$$(5.6) \quad \tilde{\mathbb{G}}_j := \bigcup_{j'=0}^j G_{j'}.$$

REMARK 5.2. *In the above construction, bisecting hyperplanes' and at the center of each chamber have to be considered after integer rounding if necessary (so that the hyperplanes and the boxes  $\tilde{B}_j$  are subsets of  $\mathbb{Z}^d$ ).*

Now we introduce for  $j = 0, 1, \dots, J$  the decreasing sequence of sets (that are all unions of hypercubes) with  $\tilde{D}_0 := \Lambda_N$  and for  $j = 1, \dots, J$

$$(5.7) \quad \tilde{D}_j := \bigcup_{\tilde{B} \in \tilde{\mathcal{C}}_j} \tilde{B}.$$

Now we reverse the order by introducing for  $j = 0, \dots, J$

$$(5.8) \quad N_j := \tilde{N}_{J-j}, \quad B_{j,k} := \tilde{B}_{J-j,k}, \quad D_j := \tilde{D}_{J-j}, \quad G_j := \tilde{G}_{J-j}, \quad \mathbb{G}_j := \tilde{\mathbb{G}}_{J-j}.$$

Note that with this order reversing,  $N_j$  is close to  $2^j N_0$  and  $N_0$  does not depend much on  $N$  (see (5.5)), namely  $h^{2/d} N_0 / (\log(1/h))^{2/d} \in [7, 15]$ . Furthermore, with our construction, the fraction of  $\Lambda_N$  which is not covered by level zero boxes  $B_{0,k}$  is small. The content of this paragraph is made more precise and quantitative by:

LEMMA 5.3. *With the notations specified above, for every  $\varkappa > 0$  there exists  $C_\varkappa > 0$  and  $h_0 > 0$  such that for every  $j \geq 1$  and every  $h \in (0, h_0]$  we have*

$$(5.9) \quad 2^j N_0 \leq N_j \leq (1 + C_\varkappa N_0^{-\varkappa}) 2^j N_0.$$

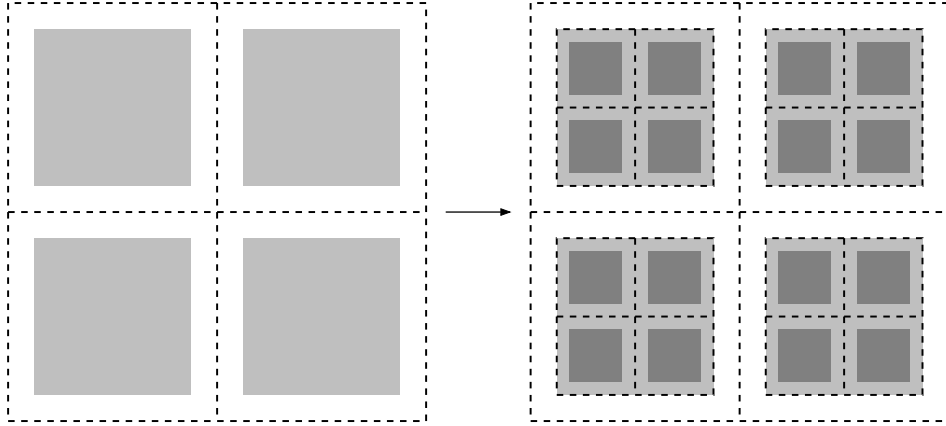


FIG 3. We draw two bisecting steps of the construction of the nested almost coverings of  $\Lambda_N$ , going say from level  $j - 1$  to  $j$  and from level  $j$  to  $j + 1$  in the preliminary version of the construction (that is before inversion: but of course we can view it the other way around). The dashed lines in the two figures mark a portion of the conditioning grid, but they are also the boundary of boxes: notably on the left the largest cube (square in the figure) delimited by dashed line is one of the  $\tilde{B}$  boxes at level  $j - 1$ , say  $\tilde{B}_{j-1, k_0}$ . This box contains  $2^d$  boxes of level  $j$ : they are denoted by  $\tilde{B}_{j, k}$ ,  $k \in \llbracket 2^d k_0 + 1, 2^d k_0 + 2^d \rrbracket$ . This operation is then repeated in the drawing in the right and if we choose one of this boxes, say  $\tilde{B}_{j, k_1}$ , the next level boxes inside  $\tilde{B}_{j, k_1}$  are  $\tilde{B}_{j+1, k}$  with  $k \in \llbracket 2^d k_1 + 1, 2^d k_1 + 2^d \rrbracket$ .

In particular we have for  $C'_\varkappa = 2dC_\varkappa$

$$(5.10) \quad \frac{\left| \bigcup_{k=1}^{2^{dJ}} B_{0, k} \right|}{N^d} \geq (1 - C'_\varkappa N_0^{-\varkappa}).$$

PROOF. The lower bound in (5.9) follows because (5.3) implies  $N_j \geq 2N_{j-1}$ . For what concerns the upper bound in (5.9), by definition  $N_0 = \tilde{N}_j \geq 7((1/h) \log(1/h))^{2/d}$  (recall (5.5)), (5.3) implies that, if  $h$  is sufficiently small, for every  $j \geq 1$  we have

$$(5.11) \quad N_j \leq 2N_{j-1} + 10(N_{j-1})^{1-\varkappa} = 2N_{j-1} \left( 1 + 5N_{j-1}^{-\varkappa} \right).$$

Hence iterating and using the lower bound we obtain

$$(5.12) \quad \frac{N_j}{N_0} \leq 2^j \prod_{i=1}^j (1 + 5N_{i-1}^{-\varkappa}) \leq 2^j \prod_{i=0}^{\infty} (1 + 5(2^i N_0)^{-\varkappa}) \leq 2^j (1 + C_\varkappa N_0^{-\varkappa}),$$



with  $C_\varkappa = 6/(1 - 2^{-\varkappa})$ . Therefore (5.9) is proven. The inequality (5.10) comes from the fact that from (5.9) we have

$$(5.13) \quad \frac{\left| \bigcup_{k=1}^{2^{dJ}} B_{0,k} \right|}{N^d} = \left( \frac{2^J N_0}{N_J} \right)^d \geq (1 + C_\varkappa N_0^{-\varkappa})^{-d},$$

from which the result follows.  $\square$

Finally, we divide the level-0 boxes, whose edge length  $N_0$  satisfies (because of (5.5))

$$(5.14) \quad 7(\log(1/h)/h)^{2/d} \leq N_0 \leq 15(\log(1/h)/h)^{2/d},$$

into  $6^d$  cubes of edge length  $\lfloor N_0/6 \rfloor$ . More precisely  $\bar{B}_{6^d k+i}$ ,  $i = 1, \dots, 6^d$  are obtained by dividing  $B_{0,k}$ , see Fig. 4. We set  $\bar{N}_h = \lfloor N_0/6 \rfloor$  and set  $\bar{\mathcal{C}}_0 := \{\bar{B}_l, l \in \llbracket 1, 6^d 2^{dJ} \rrbracket\}$ . Note that (5.10) is also valid for  $\bigcup_l \bar{B}_l$ , possibly increasing the value of  $C'_\varkappa$ . Moreover  $\bar{N}_h$  depends on  $N$  only mildly: in fact from (5.5) we have

$$(5.15) \quad (\log(1/h)/h)^{2/d} \leq \bar{N}_h \leq 3(\log(1/h)/h)^{2/d}.$$

We refer to  $\bar{B}_l$  as an *elementary box* and to  $B_{0,k}$  as a level-0 box.

We let

$$(5.16) \quad K := 2^{dJ} \quad \text{and} \quad \bar{K} := 6^d 2^{dJ},$$

denote the number of as level-0 boxes and elementary boxes respectively.

5.2. *Control of the contact density in elementary boxes.* For  $k \in \llbracket 1, \bar{K} \rrbracket$  let us define  $\zeta(k)$  to be the contact fraction inside  $\bar{B}_k$ .

$$(5.17) \quad \zeta(k) := \frac{1}{\bar{N}_h^d} \sum_{x \in \bar{B}_k} \mathbf{1}_{[-1,1]}(\phi(x)),$$

and let  $\mathcal{C}(N, \delta)$  be the event that most boxes have a contact fraction *reasonably* close to the optimal value  $p_h = 2\chi(\beta)h$

$$(5.18) \quad \mathcal{C}(N, \delta) := \left\{ \# \left\{ k \in \llbracket 1, \bar{K} \rrbracket : \frac{\zeta(k)}{\chi(\beta)h} \notin [1, 3] \right\} \leq \delta \bar{K} \right\}.$$

In the end  $\delta$  will be chosen proportional to  $\varepsilon$ . Our first step is to prove that  $\mathcal{C}(N, \delta)$  has probability close to one.

LEMMA 5.4. *Choose an arbitrary value of  $\delta > 0$ . Then there exists  $h_0 > 0$  such that for  $h \in (0, h_0]$  there exists  $c = c(h) > 0$  and  $N_h > 0$  such that for  $h \in (0, h_0]$  and  $N \geq N_h$  we have*

$$(5.19) \quad \mathbb{E} \left[ \mathbf{P}_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right) \right] = e^{-c(h)\delta N^d}.$$

REMARK 5.5. *The result would be valid also replacing [1, 3] by an arbitrarily small interval centered at 2, but this is useless for the rest of the proof and, with our choice, we avoid introducing one more parameter.*

PROOF. We are going to prove an upper bound on  $Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right)$  which shows that it is typically much smaller than  $Z_{N,\omega,\beta,h}$ . More precisely, by Markov inequality it is sufficient to show that for every  $h > 0$  small there exists  $\theta \in (0, 1)$  such that

$$(5.20) \quad \limsup_{N \rightarrow \infty} \frac{1}{\theta N^d} \log \mathbb{E} \left[ Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right)^\theta \right] < F(\beta, h).$$

We are going to prove a more quantitative statement that implies the one above (recall Theorem 1.1), that is that for  $N$  sufficiently large

$$(5.21) \quad \frac{1}{\theta N^d} \log \mathbb{E} \left[ Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right)^\theta \right] \leq \left( 1 - \frac{\delta}{8} \right) \chi(\beta) h^2.$$

We decompose  $Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right)$  according to the position of atypical density blocks. Given  $I \subset \llbracket 1, \bar{K} \rrbracket$  we set

$$(5.22) \quad \mathcal{A}_I := \{ \{k : |\zeta(k) - 2\chi(\beta)h| > \chi(\beta)h\} = I \},$$

so  $\mathcal{C}(N, \delta)^{\mathbb{C}}$  coincides with the union of the events  $\mathcal{A}_I$  with  $|I| \geq \delta \bar{K}$ . We then observe that for  $\theta \in (0, 1]$

$$(5.23) \quad Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right)^\theta \leq \sum_{\{I : |I| \geq \delta \bar{K}\}} Z_{N,\omega,\beta,h} (\mathcal{A}_I)^\theta,$$

where we have used the elementary inequality  $(\sum_j a_j)^\theta \leq \sum_j a_j^\theta$  that holds for  $a_j > 0$ . Therefore

$$(5.24) \quad \mathbb{E} \left[ Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^{\mathbb{C}} \right)^\theta \right] \leq 2^{\bar{K}} \max_{|I| \geq \delta \bar{K}} \mathbb{E} \left[ Z_{N,\omega,\beta,h} (\mathcal{A}_I)^\theta \right].$$

Now we can use the fact that  $\bar{K} \leq (N/\bar{N}_h)^d$  (simply because the boxes are disjoint), and we obtain that

$$(5.25) \quad \frac{1}{\theta N^d} \log \mathbb{E} \left[ Z_{N,\omega,\beta,h} \left( \mathcal{C}(N, \delta)^c \right)^\theta \right] \leq \frac{\log 2}{\theta \bar{N}_h^d} + \frac{1}{\theta N^d} \max_{|I| \geq \delta \bar{K}} \log \mathbb{E} \left[ Z_{N,\omega,\beta,h} (\mathcal{A}_I)^\theta \right].$$

If one sets  $\theta = \theta_h := (\log 1/h)^{-1}$ , the first term on the right-hand side is of order  $h^{-2}(\log 1/h)^{-1}$  and hence smaller than  $h^2 \chi(\beta)/24$ . Hence we can conclude that (5.21) if we show that that for  $h$  sufficiently small

$$(5.26) \quad \frac{1}{\theta_h N^d} \max_{|I| \geq \delta \bar{K}} \log \mathbb{E} \left[ Z_{N,\omega,\beta,h} (\mathcal{A}_I)^{\theta_h} \right] \leq \chi(\beta) \left(1 - \frac{\delta}{6}\right) h^2.$$

For this we use the following technical lemma, whose proof is postponed till the end of the proof we are developing.

LEMMA 5.6. *Recall that  $\xi_x = e^{\omega_x(\beta,h)} - 1$ . Let  $\Lambda$  be an arbitrary subset of  $\mathcal{Z}^d$ . For any measure on  $\mathbf{P}_\Lambda$  on  $\{0,1\}^\Lambda$  and every  $\theta \in (0,1)$  we have for every  $h \geq 0$  that*

$$(5.27) \quad \mathbb{E} \left[ \mathbf{E}_\Lambda \left[ \prod_{\{x : \delta_x=1\}} (1 + \xi_x) \right]^\theta \right] \leq \left( \max_{p \in (0,1)} \mathbb{E} \left[ (1 + p\xi)^\theta \right] \right)^{|\Lambda|}.$$

In the limit where  $\theta$  and  $h$  tend to zero we have

$$(5.28) \quad \frac{1}{\theta} \max_{p \in (0,1)} \log \mathbb{E} \left[ (1 + p\xi)^\theta \right] = \chi(\beta) h^2 + O(h^2 \theta) + O(h^3).$$

If furthermore  $\mathbf{P}_\Lambda$  is such that for  $\eta > 0$  and  $h > 0$

$$(5.29) \quad \mathbf{P}_\Lambda \left( \left| \frac{1}{|\Lambda|} \sum_{x \in \Lambda} \delta_x - 2\chi(\beta)h \right| \geq \eta h \right) = 1,$$

then there exist  $C = C_\beta > 0$ ,  $h_0 = h_{0,\beta,\eta}$  and  $\theta_0 = \theta_{0,\beta}$  such that for all  $h \in (0, h_0]$  and  $\theta \in (0, \theta_0]$  and for all  $\Lambda$

$$(5.30) \quad \frac{1}{\theta |\Lambda|} \log \mathbb{E} \left[ \mathbf{E}_\Lambda \left[ \prod_{\{x : \delta_x=1\}} (1 + \xi_x) \right]^\theta \right] \leq \chi(\beta) \left( 1 - \frac{\eta^2}{4\chi(\beta)^2} \right) h^2 + \frac{\log 2}{\theta |\Lambda|} + C(h^2 \theta + h^3).$$

REMARK 5.7. *The inequality (5.30) is valid for all size of boxes. However it provides a better bound than (5.28) only if  $\theta|\Lambda|$  is much larger than  $h^{-2}$ .*

Consider  $I \subset \llbracket 1, \bar{K} \rrbracket$  with  $|I| \geq \delta \bar{K}$ . Let  $k_1 < \dots < k_{|I|}$  denote the elements of  $I$ . We can prove by induction that for every  $j \leq |I|$

$$(5.31) \quad T_1(I, j) := \frac{1}{\theta_h} \log \mathbb{E} \left[ \left( \mathbf{E}_N \left[ e^{\sum_{x \in \cup_{1 \leq i \leq j} \bar{B}_{k_i}} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right] \right)^{\theta_h} \right] \leq \frac{4j}{5} \chi(\beta) h^2 \bar{N}_h^d.$$

The result for  $j = 1$  is a direct consequence of (5.30): we work with  $\eta = \chi(\beta)$  and then we use that  $1/(\bar{N}_h^d \theta_h) = O(h^2 / \log(1/h)) \ll h^2$ , so that the term  $(\log 2)/(\theta_h |\Lambda|)$ , as well as the term  $C(h^2 \theta_h + h^3)$ , can be absorbed into the leading order, for  $h$  sufficiently small (hence the factor  $4/5$  instead of  $3/4$ ). Now for  $j > 1$ , we just need to apply (5.30) to the measure  $\mu$  defined by

$$(5.32) \quad \mu(A) := \frac{\mathbf{E}_N \left[ e^{\sum_{x \in (\cup_{1 \leq i \leq j-1} \bar{B}_{k_i})} \omega_x(\beta, h) \delta_x} \mathbf{1}_A \mid \mathcal{A}_I \right]}{\mathbf{E}_N \left[ e^{\sum_{x \in (\cup_{1 \leq i \leq j-1} \bar{B}_{k_i})} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right]}.$$

We obtain that

$$(5.33) \quad \log \mathbb{E} \left[ \left( \frac{\mathbf{E}_N \left[ e^{\sum_{x \in (\cup_{1 \leq i \leq j} \bar{B}_{k_i})} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right]}{\mathbf{E}_N \left[ e^{\sum_{x \in (\cup_{1 \leq i \leq j-1} \bar{B}_{k_i})} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right]} \right)^{\theta_h} \right] = \log \mathbb{E} \left[ \mu \left( e^{\sum_{x \in \bar{B}_{k_j}} \omega_x(\beta, h) \delta_x} \right)^{\theta_h} \right] \leq \frac{4}{5} \theta_h \bar{N}_h^d \chi(\beta) h^2$$

where the inequality is obtained by applying (5.30), in the same way as for the case  $j = 1$ , when averaging with respect to  $(\omega_x)_{x \in \bar{B}_{k_j}}$ . This completes the induction argument and (5.31) is proven.

Using the same trick and (5.27)-(5.28) we obtain that

$$(5.34) \quad T_2(I) := \frac{1}{\theta_h} \log \mathbb{E} \left[ \left( \frac{\mathbf{E}_N \left[ e^{\sum_{x \in \hat{\Lambda}_N} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right]}{\mathbf{E}_N \left[ e^{\sum_{x \in \cup_{i \in I} \bar{B}_{k_i}} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right]} \right)^{\theta_h} \right] \\ \leq \left[ (N-1)^d - |I| \bar{N}_h^d \right] (\chi(\beta) h^2 + C \theta_h h^2).$$

Finally, combining (5.31) and (5.34), by apply the Cauchy-Schwarz inequality we obtain

$$(5.35) \quad \frac{2}{\theta_h} \log \mathbb{E} \left[ Z_{N, \omega, \beta, h}(\mathcal{A}_I)^{(\theta_h/2)} \right] \\ \leq \frac{2}{\theta_h} \log \mathbb{E} \left[ \left( \mathbf{E}_N \left[ e^{\sum_{x \in \hat{\Lambda}_N} \omega_x(\beta, h) \delta_x} \mid \mathcal{A}_I \right] \right)^{\theta_h/2} \right] \\ \leq T_1(I, |I|) + T_2(I) \leq \chi(\beta) h^2 \left( N^d - \frac{1}{5} |I| \bar{N}_h^d \right) + C \theta_h h^2 N^d \\ \leq \left( 1 - \frac{\delta}{6} \right) \chi(\beta) h^2 N^d,$$

which is (5.26), except for the (clearly irrelevant) fact that  $\theta_h$  is replaced by  $\theta_h/2$ .  $\square$

PROOF OF LEMMA 5.6. The inequality (5.27) can be proven using the argument of proof of Proposition 2.3, but we give here a different proof. We proceed by induction on  $N = |\Lambda|$ . Note that we can assume without loss of generality that  $\Lambda := \llbracket 0, N \rrbracket$  and write  $\mathbf{E}_N$  for  $\mathbf{E}_\Lambda$ . The result is obvious for  $N = 0$ . Now given  $N \geq 1$  set

$$(5.36) \quad p_N(\xi_1, \dots, \xi_{N-1}) := \frac{\mathbf{E}_N \left[ \prod_{\{x \in \llbracket 1, N-1 \rrbracket : \delta_x=1\}} (1 + \xi_x) \delta_N \right]}{\mathbf{E}_N \left[ \prod_{\{x \in \llbracket 1, N-1 \rrbracket : \delta_x=1\}} (1 + \xi_x) \right]},$$

together with

$$(5.37) \quad Z_N(\xi) := \mathbf{E}_N \left[ \prod_{\{x \in \llbracket 1, N \rrbracket : \delta_x=1\}} (1 + \xi_x) \right],$$

and

$$(5.38) \quad \tilde{Z}_N(\xi) := \mathbf{E}_N \left[ \prod_{\{x \in \llbracket 1, N-1 \rrbracket : \delta_x = 1\}} (1 + \xi_x) \right].$$

Note that

$$(5.39) \quad Z_N(\xi) = \tilde{Z}_N(\xi)(1 + p_N \xi_N),$$

and by raising both sides to the power  $\theta$  and taking the average with respect to  $\xi_N$  we obtain

$$(5.40) \quad \begin{aligned} \mathbb{E} \left[ Z_N^\theta(\xi) \mid (\xi_x)_{1 \leq x \leq N-1} \right] &= \tilde{Z}_N^\theta(\xi) \mathbb{E} \left[ (1 + p_N \xi_N)^\theta \right] \\ &\leq \tilde{Z}_N^\theta(\xi) \max_{p \in (0,1]} \mathbb{E} \left[ (1 + p\xi)^\theta \right], \end{aligned}$$

and by taking the average with respect to all other variables on both sides we obtain that

$$(5.41) \quad \mathbb{E} \left[ Z_N^\theta(\xi) \right] \leq \mathbb{E} \left[ \tilde{Z}_N^\theta(\xi) \right] \max_{p \in (0,1]} \mathbb{E} \left[ (1 + p\xi)^\theta \right].$$

We can now take the supremum over  $\mathbf{P}_N$  and we can conclude the induction step because

$$(5.42) \quad \sup_{\mathbf{P}_N} \mathbb{E} \left[ \tilde{Z}_N^\theta(\xi) \right] = \sup_{\mathbf{P}_{N-1}} \mathbb{E} \left[ Z_{N-1}^\theta(\xi) \right].$$

The bound (5.27) is therefore established.

As for the optimizing problem (5.28) we can proceed as for Proposition 2.2 using Taylor expansion and showing first that  $p_{h,\theta}$  tends to 0, then that  $p_{h,\theta} \sim 2\chi(\beta)h$ , then  $p_{h,\theta} = 2\chi(\beta)h + O(\theta h) + O(h^2)$  and then the final claim. We do not detail these steps.

For what concerns (5.30) (recall that we are therefore assuming (5.29)), at the cost of losing a factor two in the estimate on  $\mathbb{E}[Z_N^\theta]$ , setting  $I_N := \sum_{x \in \llbracket 1, N \rrbracket} \delta_x$  we can assume that one of the following holds

$$(5.43) \quad \mathbf{P}_N(I_N \geq (2\chi(\beta) + \eta)hN) = 1 \text{ or } \mathbf{P}_N(I_N \leq (2\chi(\beta) - \eta)hN) = 1.$$

Using Hölder inequality we have that for every positive random variable  $g$

$$(5.44) \quad \mathbb{E} \left[ Z_N^\theta \right] \leq \mathbb{E} \left[ Z_N g^{\theta-1} \right]^\theta \mathbb{E} \left[ g^\theta \right]^{1-\theta}.$$

Now we set  $g = \prod_{x \in \llbracket 1, N \rrbracket} (1 + q\xi')^{\frac{1}{1-\theta}}$  with  $\xi' = e^{\beta\omega - \lambda(\beta)} - 1$  (using  $\xi$  instead of  $\xi'$  would give an analogous result but computations are easier with  $\xi'$ ). The first term in the r.h.s. of (5.44) is

$$(5.45) \quad \mathbf{E}_N \left[ \left( e^h \mathbb{E} \left[ (1 + \xi')(1 + q\xi')^{-1} \right] \right)^{I_N} \mathbb{E} \left[ (1 + q_-\xi')^{-1} \right]^{N-I_N} \right]$$

We are going to set  $q = q_{\pm} := (2\chi(\beta) \pm \eta)h$  depending on which assumption we have on the contact fraction.

The following asymptotic statements hold in the limit where both  $q$  and  $\theta$  go to zero

$$(5.46) \quad \begin{aligned} \mathbb{E} \left[ (1 + q\xi')^{-1} \right] &= 1 + q^2 \text{Var}(\xi') + O(q^3), \\ \mathbb{E} \left[ (1 + \xi')(1 + q\xi')^{-1} \right] &= 1 - q \text{Var}(\xi') + O(q^2), \\ \mathbb{E} \left[ (1 + q\xi')^{\frac{\theta}{1-\theta}} \right]^{\frac{1-\theta}{\theta}} &= 1 - \frac{q^2}{2} \text{Var}(\xi') + O(q^3) + O(\theta q^2). \end{aligned}$$

Note in particular performing a second order expansion in  $h$  that for fixed  $\eta$  with our choice of  $q$  we have for  $h$  sufficiently small (depending on  $\eta$  and on  $\beta$ )

$$(5.47) \quad \begin{aligned} e^h \mathbb{E} \left[ (1 + \xi')(1 + q_-\xi')^{-1} \right] &\geq \mathbb{E} \left[ (1 + q_-\xi')^{-1} \right], \\ e^h \mathbb{E} \left[ (1 + \xi')(1 + q_+\xi')^{-1} \right] &\leq \mathbb{E} \left[ (1 + q_+\xi')^{-1} \right]. \end{aligned}$$

Hence if  $I_N \leq Nq_-$ ,  $\mathbf{P}_N$ -a.s. using (5.47) we can replace  $I_N$  by its upper bound  $q_-|\Lambda|$  in (5.45) and altogether we obtain

$$(5.48) \quad \begin{aligned} \mathbb{E} \left[ Z_N^\theta \right] &\leq \left( e^h \mathbb{E} \left[ (1 + \xi')(1 + q_-\xi')^{-1} \right] \right)^{q_- N \theta} \\ &\quad \times \left( \mathbb{E} \left[ (1 + q_-\xi')^{-1} \right] \right)^{(1-q_-)N\theta} \left( \mathbb{E} \left[ (1 + q_-\xi')^{\frac{\theta}{1-\theta}} \right] \right)^{N(1-\theta)}. \end{aligned}$$

Hence applying (5.46) we obtain

$$(5.49) \quad \begin{aligned} \frac{1}{N\theta} \log \mathbb{E} \left[ Z_N^\theta \right] &\leq q_- h - \frac{q_-^2}{2} \text{Var}(\xi') + O(\theta q_-^2) + O(q_-^3) \\ &= \left( \chi(\beta) - \frac{\eta^2}{4\chi(\beta)} \right) h^2 + O(h^2\theta) + O(h^3). \end{aligned}$$

In the same manner when  $I_N \geq Nq_+$ ,  $\mathbf{P}_N$ -a.s. we have by (5.47)

$$(5.50) \quad \begin{aligned} \mathbb{E} \left[ Z_N^\theta \right] &\leq \left( e^h \mathbb{E} \left[ (1 + \xi')(1 + q_+\xi')^{-1} \right] \right)^{q_+ N \theta} \\ &\quad \times \left( \mathbb{E} \left[ (1 + q_+\xi')^{-1} \right] \right)^{(1-q_+)N\theta} \left( \mathbb{E} \left[ (1 + q_+\xi')^{\frac{\theta}{1-\theta}} \right] \right)^{N(1-\theta)}. \end{aligned}$$

and we obtain

$$(5.51) \quad \begin{aligned} \frac{1}{N\theta} \log \mathbb{E} \left[ Z_N^\theta \right] &\leq q_+ h - \frac{q_+^2}{2} \text{Var}(\xi') + O(\theta q^2) + O(q^3) \\ &= \left( \chi(\beta) - \frac{\eta^2}{4\chi(\beta)} \right) h^2 + O(h^2\theta) + O(h^3). \end{aligned}$$

This completes the proof of Lemma 5.6.  $\square$

5.3. *The main body of the argument (Proof of Proposition 5.1).* We start by an important preliminary result. By using the control we have on the local contact density and on the free energy, we are going to show that if the probability of a sequence of events decays exponentially in the volume size under  $\mathbf{P}_N$  with a rate suitably controlled from below, then it also decays exponentially under  $\mathbf{P}_{N,\omega,\beta,h}$ . Said differently, the result we are going to state allows to neglect the environment and to reduce the estimates to Gaussian field estimates.

LEMMA 5.8. *Recall  $\varepsilon_d > 0$  from Proposition 3.1. Consider for  $h \leq h_0$ ,  $N$  larger than a suitable  $N_0(h)$  and a sequence of events  $A_N$  such that  $\mathbf{P}_N(A_N) \leq \exp(-h^{2+\eta}N^d)$ , with  $\eta \in (0, \varepsilon_d)$ . Then there exists  $h'_0 > 0$  and  $N'_0(h, \omega)$  such that for every  $h \leq h'_0$  and  $N \geq N'_0(h)$*

$$(5.52) \quad \mathbf{P}_{N,\omega,\beta,h}(A_N) \leq \exp(-N^d h^{2+\eta}/4).$$

PROOF. Because  $N^{-d} \log Z_{N,\omega,\beta,h}$  converges, it is sufficient to show that (recall (4.5))

$$(5.53) \quad \frac{1}{N^d} \log Z_{N,\omega,\beta,h}(A_N) \leq \mathbb{F}(\beta, h) - h^{2+\eta}/3.$$

We are going to show first that

$$(5.54) \quad \frac{1}{N^d} \mathbb{E} [\log Z_{N,\omega,\beta,h}(A_N)] \leq \mathbb{F}(\beta, h) - h^{2+\eta}/2.$$

To see this, we note that we have (using Proposition 2.3 for  $\mathbf{P}_N(\cdot | A_N)$  and keeping into account Remark 2.4)

$$(5.55) \quad \begin{aligned} \mathbb{E} [\log Z_{N,\omega}(A_N)] &= \log \mathbf{P}_N(A_N) + \mathbb{E} \log \mathbf{E}_N \left[ e^{\sum_{x \in \Lambda_N} \omega_x(\beta, h) \delta_x} \mid A_N \right] \\ &\leq -N^d h^{2+\eta} + |\mathring{\Lambda}_N| (h^2 \chi(\beta) + C_\beta h^3). \end{aligned}$$



We have used  $Z_N$  as a shortcut for  $Z_{N,\omega,\beta,h}$  and we recall that  $\mathring{\Lambda}_N = \llbracket 1, N-1 \rrbracket^d$ , hence  $|\mathring{\Lambda}_N| \leq N^d$ . Now in view of Proposition 3.1, (5.55) implies that (5.54) holds for  $h$  sufficiently small.

Now to conclude we only need to show that  $\log Z_{N,\omega}(A_N)$  is concentrated around its mean. This follows from a standard concentration argument, for which we introduce a cut-off  $M_N := N^{d/6}$  on the disorder variables  $\bar{\omega}_x := \omega_x \mathbf{1}_{\{|\omega_x| \leq M_N\}}$ . A union bound and the finiteness of the exponential moment of all orders for  $\omega$  implies that for  $N$  sufficiently large we have

$$(5.56) \quad \mathbb{P}(\exists x \in \Lambda_N \text{ such that } |\omega_x| \geq M_N) \leq e^{-N^{d/6}}.$$

We now apply Azuma's inequality to the martingale with bounded increments  $(\mathbb{E}[\log Z_{N,\bar{\omega}}(A_N) \mid \mathcal{G}_n])_{n=0}^{|\mathring{\Lambda}_N|}$ , where  $\mathcal{G}_n := \sigma(\omega_{x_i}, i \leq n)$  and  $x_1, \dots, x_{|\mathring{\Lambda}_N|}$  is an arbitrary enumeration of  $\mathring{\Lambda}_N$ . Note that if  $\omega$  and  $\omega'$  are such that  $\omega_n = \omega'_n$  for every  $n \neq j$  we have

$$|\log Z_{N,\omega}(A_N) - \log Z_{N,\omega'}(A_N)| \leq \beta |\omega_j - \omega'_j| \leq 2\beta M_N,$$

so the increment of the martingale are bounded by  $2\beta M_N$ . Therefore we obtain that

$$(5.57) \quad \mathbb{P}(|\log Z_{N,\bar{\omega}}(A_N) - \mathbb{E}[\log Z_{N,\bar{\omega}}(A_N)]| \geq u) \leq e^{-\frac{cu^2}{N^d M_N^2}},$$

with  $c = 1/(8\beta^2)$ . Applying this for  $u = N^{3d/4}$  and using (5.56) to bound the probability of  $\{Z_{N,\omega}(A_N) \neq Z_{N,\bar{\omega}}(A_N)\}$  we obtain (for  $c' = c \wedge 1$ )

$$(5.58) \quad \mathbb{P}\left(|\log Z_{N,\omega}(A_N) - \mathbb{E}[\log Z_{N,\bar{\omega}}(A_N)]| > N^{3d/4}\right) \leq 2e^{-c'N^{d/6}}$$

At this point we can conclude by observing that the difference between  $\mathbb{E}[\log Z_{N,\bar{\omega}}(A_N)]$  and  $\mathbb{E}[\log Z_{N,\omega}(A_N)]$  is small: this follows by using Lemma A.3 with  $K = M_N$  that yields

$$(5.59) \quad \mathbb{E}[|\log Z_{N,\omega}(A_N) - \log Z_{N,\bar{\omega}}(A_N)|] \leq \beta N^d \mathbb{E}[|\omega_x|; |\omega_x| > M_N] \leq e^{-N^{d/6}},$$

for  $N$  sufficiently large, again because all exponential moments of  $\omega$  are finite. We can then obtain (5.53) combining the concentration estimate above with (5.54) by Borel-Cantelli. This completes the proof of Lemma 5.8.  $\square$

*Proof of Proposition 5.1.* We aim at applying Lemma 5.8 so to reduce the proof of Proposition 5.1 to proving a statement about  $\mathbf{P}_N$ . However this

fails if applied directly to  $\mathcal{D}(N, \varepsilon)$  because  $\mathbf{P}_N(\mathcal{D}(N, \varepsilon))$  does not decrease exponentially in the volume (due to the massless character of the LFF, see e.g. [7]). However, we will show that the event  $\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon)$  does not have this drawback and Lemma 5.4 assures that we can limit ourselves to studying this event. More formally, thanks to Lemma 5.4 and Lemma 5.8 it suffices to show that given  $\varepsilon > 0$  there exist  $\delta > 0$  and  $\eta < \varepsilon_d$  such that for  $h$  sufficiently small for all  $N$  sufficiently large, we have

$$(5.60) \quad \frac{1}{N^d} \log \mathbf{P}_N(\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon)) \leq -h^{2+\eta}.$$

To prove (5.60) we are going to use a decomposition of  $\phi$  into a field  $\psi$  which is independent in each level-0 boxes, and a field  $\bar{\psi}$  which displays long range correlation but has a small amplitude.

Recall that  $\mathbb{G}_0$  denotes a grid which isolates each of the boxes at level zero of the hierarchy. By the Markov property of the LFF, we can write

$$(5.61) \quad \phi = \phi_{\Lambda_N} \stackrel{(\text{law})}{=} \psi + \bar{\psi},$$

where  $\bar{\psi}$  is the harmonic continuation of the restriction of  $\phi$  to  $\mathbb{G}_0$ , that is the solution of

$$(5.62) \quad \begin{cases} \Delta \bar{\psi}(x) = 0, & \text{for all } x \in \mathring{\Lambda}_N \setminus \mathbb{G}_0, \\ \bar{\psi}(x) = \phi, & \text{for } x \in \mathbb{G}_0, \end{cases}$$

and  $\psi$  a free field on  $\mathring{\Lambda}_N \setminus \mathbb{G}_0$  with 0 boundary condition: therefore  $\psi$  is a collection of independent free fields on each of the level-0 boxes, and  $\bar{\psi}$  carries all the long range correlations of  $\phi$ . We are going to reduce the proof of (5.60) to that of the two following facts about  $\psi$  and  $\bar{\psi}$  (rigorously stated as Proposition 5.10 and Proposition 5.12 below):

- (A) With very large probability on most level-0 boxes  $B_{0,k}$ , the field  $\bar{\psi}$  is almost flat (in the sense that it is very close to an affine function). We mean by this that the probability of the complement is smaller than  $e^{-N^d h^b}$  for a value of  $b < 2$ : this is largely sufficient because also a value of  $b$  slight larger than 2 would have been sufficient (see Lemma 5.8).
- (B) When  $\bar{\psi}$  is flat on the box  $B_{0,k}$ , the probability of having both the right number of contact (that is, about  $2\bar{N}_0^d \chi(\beta)h$ ) in each of the  $6^d$  elementary boxes inside  $B_{0,k}$ , and a density of high points is small.

To conclude from these two statements, we only need to use the fact that conditioned to  $\bar{\psi}$ , the various level 0 boxes are independent.

REMARK 5.9. *The fact that the 0 level boxes are separated from the grid of a distance equal (up to integer rounding) to  $N_0^{1-\varkappa}$  and recalling that  $N_0 \approx (\log(1/h)/h)^{2/d}$ , more precisely (5.14), readily yields (use (1.3)) that the variance of  $\psi_x$  can differ from  $\sigma_d^2$  at most of a term  $O(1/N_0^{(d-2)(1-\varkappa)})$ , that is by  $O(h^c)$  with  $c$  a positive constant. In view of the estimates we aim at, these variations of the variance turn out to be irrelevant: details will be given in due times but the reason is simply that for every  $b > 0$*

$$(5.63) \quad \frac{P((1 + \epsilon(h))\mathcal{N} \geq b \log(1/h))}{P(\mathcal{N} \geq b \log(1/h))} - 1 = O(h^{c'}) ,$$

if  $\epsilon(h) = O(h^c)$ . Moreover the distance from the conditioning grid is not only used at level-0, but at all levels. However for the higher levels it is used to assure that harmonic extension has a small variance uniformly in the box: the tool we use in that case is Lemma A.4.

Let us give a more quantitative version of (A) and (B). First let us define for  $e$  and  $g$  two unit vectors (not necessarily distinct) of the canonical base of  $\mathbb{R}^d$   $\nabla_{eg}\bar{\psi}$  to be the *bi-gradient* of  $\bar{\psi}$  in directions  $e$  and  $g$

$$(5.64) \quad \nabla_{eg}\bar{\psi}(x) := \bar{\psi}(x + e + g) - \bar{\psi}(x + e) - \bar{\psi}(x + g) + \bar{\psi}(x) .$$

Now for  $k \in \llbracket 1, K \rrbracket$  (recall  $K = 2^{dJ}$  is the number of level-0 boxes) we set

$$(5.65) \quad \|\nabla_{eg}\bar{\psi}\|_{\infty, k} := \max_{x \in B_{0, k}} |\nabla_{eg}\bar{\psi}(x)| .$$

Finally we let  $\mathcal{B}(N, \delta)$  the event that all bi-gradients are small on most boxes, that is

$$(5.66) \quad \mathcal{B}(N, \delta) := \{ |\{k \in \llbracket 1, K \rrbracket : \|\nabla_{eg}\bar{\psi}\|_{\infty, k} \geq N_0^{-2}\}| \leq \delta K \} .$$

Here is the result we need on the even  $\mathcal{B}(N, \delta)$ : it is proven in Section 5.5.

PROPOSITION 5.10. *For any  $\delta > 0$  and for  $h$  sufficiently small*

$$(5.67) \quad \mathbf{P}_N \left( \mathcal{B}(N, \delta)^c \right) \leq e^{-N^d h^{11/6}} .$$

By putting together Lemma 5.8 and Lemma 5.10 we have that for every  $\eta > 0$ ,  $h$  sufficiently small and  $N \geq N_0(\omega, \delta, \eta, h)$

$$(5.68) \quad \mathbf{P}_{N, \omega, \beta, h} \left( \mathcal{B}(N, \delta)^c \right) \leq e^{-h^{2+\eta} N^d} .$$

But since  $11/6 < 2$  we can avoid Lemma 5.8 and obtain in a more direct way from Lemma 5.10 that the right-hand side in (5.68) can be improved to  $e^{-h^{2-\eta}N^d}$ , any  $\eta < 1/6$ , but this is irrelevant for what follows.

REMARK 5.11. *The reason for us to consider bi-gradient of  $\bar{\psi}$  instead of  $\bar{\psi}$  or its gradient is that the spatial correlations of the bi-gradient are summable (which is not the case for neither  $\bar{\psi}$  nor for its gradient). The corresponding statement for  $\psi$  or for the gradient of  $\psi$  would not hold for this reason.*

Given  $k \in \llbracket 1, K \rrbracket$ , and recalling that  $\bar{B}_{6^d k+i}$ ,  $i = 1, \dots, 6^d$  are disjoint boxes of edge length  $\bar{N}_h$  located in  $B_{0,k}$  we define  $\mathcal{E}^{(1)}(k)$  to be the event that all elementary boxes in  $B_{0,k}$  have a typical contact density. Recalling (5.17) we set

$$(5.69) \quad \mathcal{E}^{(1)}(k) := \bigcap_{i=1}^{6^d} \left\{ \frac{\zeta(6^d k + i)}{\chi(\beta)h} \in [1, 3] \right\},$$

We let  $\mathcal{E}^{(2)}(k) = \mathcal{E}^{(2)}(k, \delta, \varepsilon)$  be the event that the box  $B_{0,k}$  displays a density  $\delta$  of high points (recall  $u_h = \sigma_d \sqrt{2 \log(1/h)}$ )

$$(5.70) \quad \mathcal{E}^{(2)}(k) := \left\{ \sum_{x \in B_{0,k}} \mathbf{1}_{\{|\phi(x)| \geq (1+\varepsilon)u_h\}} \geq \delta N_0^d \right\}$$

Our second result is that  $\mathcal{E}^{(1)}(k) \cap \mathcal{E}^{(2)}(k)$  is unlikely in boxes where the bi-gradient of  $\bar{\psi}$  are small: to the proof is devoted Section 5.4.

PROPOSITION 5.12. *Given  $\eta > 0$  there exists  $h_0(\eta, \delta, \varepsilon) > 0$  such that if  $h \leq h_0(\eta, \delta, \varepsilon)$  and  $\max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty, k} \leq N_0^{-2}$  we have*

$$(5.71) \quad \mathbf{P}_N \left( \mathcal{E}^{(1)}(k) \cap \mathcal{E}^{(2)}(k) \mid \bar{\psi} \right) \leq \eta.$$

By Proposition 5.10, in order to get to (5.60) it is sufficient to show that

$$(5.72) \quad \mathbf{P}_N (\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon) \cap \mathcal{B}(N, \delta)) \leq \exp \left( -h^{2+\eta} N^d \right),$$

for some small  $\delta > 0$  (which is allowed to depend on  $\varepsilon$ ). We are going to show that whenever  $\bar{\psi} \in \mathcal{B}(N, \delta)$  we have

$$(5.73) \quad \mathbf{P}_N (\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon) \mid \bar{\psi}) \leq \exp \left( -h^{2+\eta} N^d \right).$$

Now we record as a lemma the observation that when  $\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon)$  holds then  $\mathcal{E}^{(1)}(k) \cap \mathcal{E}^{(2)}(k)$  holds in a lot of boxes: also the proof of this result is delayed till the end of the main argument.

LEMMA 5.13. *Given  $\varepsilon > 0$  sufficiently small and  $\delta \leq \varepsilon/(6^d + 4)$  then we have for all  $N$  sufficiently large*

$$(5.74) \quad \mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon) \cap \mathcal{B}(N, \delta) \subset \left\{ \sum_{k=1}^K \mathbf{1}_{\mathcal{E}^{(1)}(k) \cap \mathcal{E}^{(2)}(k)} \mathbf{1}_{\{\max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty, k} < N_0^{-2}\}} \geq \delta K \right\}.$$

According to Proposition 5.12, conditioning to  $\bar{\psi}$  and assuming  $\bar{\psi} \in \mathcal{B}(N, \delta)$  the sum

$$(5.75) \quad \sum_{\{k : \max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty, k} \leq N_0^{-2}\}} \mathbf{1}_{\mathcal{E}^{(1)}(k) \cap \mathcal{E}^{(2)}(k)}$$

is stochastically dominated by a binomial of parameters  $\eta$  and  $K$ . We choose  $\eta := \delta/3$  and from the second bound in Lemma A.1 applied for  $\Delta = 3$ , we obtain that when  $\bar{\psi} \in \mathcal{B}(N, \delta)$

$$(5.76) \quad \mathbf{P}_N \left( \sum_{\{k : \max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty, k} \leq N_0^{-2}\}} \mathbf{1}_{\mathcal{E}^{(1)}(k) \cap \mathcal{E}^{(2)}(k)} \geq \delta K \mid \bar{\psi} \right) \leq P(\text{Bin}(K, \delta/3) \geq \delta K) \leq e^{-(\delta/6)K}.$$

But  $K = 2^{dJ} = (N/N_0)^d \geq cN^d h^2 (\log h)^{-2}$  for some  $c > 0$ , this is sufficient to conclude that (5.73) holds. This completes the proof of (5.60) and therefore Proposition 5.1 is established.  $\square$

PROOF OF LEMMA 5.13. We need to show that on the event  $\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon) \cap \mathcal{B}(N, \delta)$  we have

$$(5.77) \quad \sum_{k=1}^K \mathbf{1}_{\mathcal{E}^{(1)}(k)^c} + \sum_{k=1}^K \mathbf{1}_{\mathcal{E}^{(2)}(k)^c} + \sum_{k=1}^K \mathbf{1}_{\{\max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty, k} \geq N_0^{-2}\}} \leq (1 - \delta)K.$$

First of all on  $\mathcal{C}(N, \delta)$ , recall (5.18), we have

$$(5.78) \quad \sum_{k=1}^K \mathbf{1}_{\mathcal{E}^{(1)}(k)\mathfrak{E}} \leq \delta \bar{K} = 6^d \delta K.$$

Moreover on  $\mathcal{D}(N, \varepsilon)$

$$(5.79) \quad \varepsilon N^d \leq \sum_{x \in \tilde{\Lambda}_N} \mathbf{1}_{\{|\phi(x)| \geq (1+\varepsilon)u_h\}} \leq N_0^d \sum_{k=1}^K \mathbf{1}_{\mathcal{E}^{(2)}(k)} + \delta N_0^d K + (N^d - KN_0^d),$$

where the first term follows estimating from above with  $N_0^d$  the number of points that are too high in a box  $B_{0,k}$  where there are at least  $\delta N_0^d$  points that are too high, the second accounts for the fact that in all other boxes (overestimated by *all boxes tout court*) there at most  $\delta N_0^d$  high points, and the third accounts for the points that are not in the level 0 boxes. Recalling that  $(N/N_0)^d \geq K$ , if  $h$  is sufficiently small (so that the term  $(N^d - KN_0^d)/N^d$  is small), we have

$$(5.80) \quad \sum_{k=1}^K \mathbf{1}_{\mathcal{E}^{(2)}(k)\mathfrak{E}} \leq (1 + 2\delta - \varepsilon)K.$$

Finally the third sum is smaller than  $\delta K$  by the definition of  $\mathcal{B}(N, \delta)$ . Collecting all the estimates we see that on  $\mathcal{C}(N, \delta) \cap \mathcal{D}(N, \varepsilon) \cap \mathcal{B}(N, \delta)$  the left-hand side of (5.77) is bounded above by  $(1 + 6^d \delta + 3\delta - \varepsilon)K$ . Since  $\delta \leq \varepsilon/(6^d + 4)$ , (5.77) holds and the proof of Lemma 5.13 is complete.  $\square$

5.4. *Level-0 estimates: the proof of Proposition 5.12.* A first observation is the following

LEMMA 5.14. *If  $\max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty,k} \leq N_0^{-2}$ , then there exists  $a \in \mathbb{R}$  and  $b \in \mathbb{R}^d$  such that we have,*

$$(5.81) \quad \forall x \in B_{0,k} \quad |\bar{\psi}(x) - a + b \cdot x| \leq d^2.$$

PROOF. We can replace without loss of generality  $B_{0,k}$  by  $\llbracket -N_0/2, N_0/2 \rrbracket^d$  with appropriate rounding. We set

$$(5.82) \quad a := \bar{\psi}(0) \quad \text{and} \quad b_i := \bar{\psi}(e_i) - \bar{\psi}(0).$$

We have for every  $x \in \llbracket -N_0/2, N_0/2 \rrbracket^d$  and  $i \in \llbracket 1, d \rrbracket$

$$(5.83) \quad |\bar{\psi}(x+e) - \bar{\psi}(x) - b_i| \leq dN_0^{-1},$$

simply because the term in the absolute value can be written as the sum of at most  $N_0d$  terms of the form  $\nabla_{eg}\bar{\psi}(z)$ . Using this we see that ( $\cdot$  is the scalar product)

$$(5.84) \quad |\bar{\psi}(y) - \bar{\psi}(0) - b \cdot y| \leq d^2,$$

since the term in the absolute value can be written as the sum of at most  $N_0d$  terms of the form  $\bar{\psi}(x+e) - \bar{\psi}(x) - b_i$ .  $\square$

We recall that if we work on  $\bar{\psi} \in \mathcal{B}(N, \delta)$ , most level-0 boxes are good in the sense that they satisfy the bi-gradient requirement of Lemma 5.14. We assume that the (good) level-0 box  $B_{0,k}$  we are considering is centered at 0, that  $\bar{\psi}(0) \geq 0$  and that  $b_i \geq 0$  for all  $i$ : by symmetry this yields no loss of generality, as we explain in the caption of Figure 4. We set

$$(5.85) \quad H_0(x) := a + b \cdot x,$$

so Lemma 5.14 implies that  $\max_{x \in B_{0,k}} |\bar{\psi}(x) - H_0(x)| \leq d^2$ . We set  $y_1 := (\bar{N}_h, \bar{N}_h, \dots, \bar{N}_h)$ ,  $y_2 := 2y_1$ ,  $y_3 := 3y_1$  and we let  $\bar{B}_i$  denote the cube of edge length  $\bar{N}_h$  whose maximal corner (for the lexicographic order) is  $y_i$ . For  $i = 1, 2, 3$ ,  $\bar{B}_i \subset B_{0,k}$ , see Figure 4 and its caption. To prove Proposition 5.12, we are going to distinguish three cases.

- (i)  $H_0(y_1) \leq (1 - \delta)u_h$ . In this case  $H_0$  is small in the whole  $\bar{B}_1$  and this yields, with large probability, too many contacts with respect to what  $\mathcal{E}^{(1)}$  allows.
- (ii)  $H_0(y_2) \geq (1 + \delta)u_h$ . In this case  $H_0$  is large in  $\bar{B}_3$ , which implies that, with large probability, there are too few contacts in this box with respect to what is required by  $\mathcal{E}^{(1)}$ .
- (iii)  $H_0(y_1) \geq (1 - \delta)u_h$  and  $H_0(y_2) \leq (1 + \delta)u_h$ . In this case

$$H_0(y_3) = 2H_0(y_2) - H_0(y_1) \leq (1 + 3\delta)u_h,$$

meaning in particular that  $|H_0(x)| \leq (1 + 3\delta)u_h$  on the full level 0 box (because of our assumption  $a$  and  $b_i$  positive). This makes  $\mathcal{E}^{(2)}$  unlikely if one chooses  $\delta = \varepsilon/6$  and  $h$  sufficiently small.

We will treat cases (ii) and (iii) using first moment estimates and Markov inequality. Case (i) requires a less straightforward second moment computation.

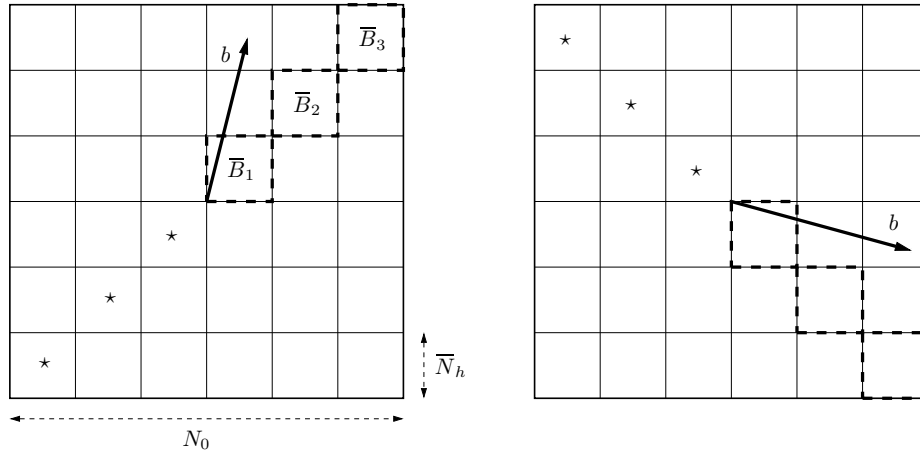


FIG 4. A level-0 box  $B_{0,k}$ , of edge length  $N_0$ , is partitioned into  $6^d$  boxes (hypercubes)  $\bar{B}_1, \dots, \bar{B}_{6^d}$  of edge length  $\bar{N}_0$ , up to integer rounding (so it may be that only a subset of  $B_{0,k}$  is really covered (but what is left out is just a negligible fraction,  $O(1/N_0)$ , of the sites in  $B_{0,k}$ ). Note that we have relabeled the elementary boxes contained in  $B_{0,k}$  both because they were originally labeled with indexes going from  $6^d k + 1$  to  $6^d(k+1)$  and because we have made a specific choice of the first three elementary boxes that minimizes notations. With reference to (5.82), the drawing considers the case  $a \geq 0$  and, in the case on the left,  $b = (b_1, \dots, b_d)$  has non negative entries. In this case the three steps analysis – see points (i)–(iii) – just focuses on three elementary boxes  $\bar{B}_1$  to  $\bar{B}_3$ . Whenever one or more entries of  $b$  are non positive it suffices to change of quadrant, like in the case on the right. And if  $a \leq 0$  it suffices to change the orientation and choose the elementary boxes marked by  $\star$ . So the  $2^{d+1}$  cases that we need to analyse are all equivalent.

5.4.1. *The case (i).* On  $\bar{B}_1$ , since  $a$  and all the  $b_i$  are positive, we have ( $0 \leq \bar{H}_0 \leq 1 - \delta u_h$ ) and thus from Lemma 5.14  $-d^2 \leq \bar{\psi} \leq (1 - \delta)u_h + d^2$ . Hence we have for  $h$  sufficiently small

$$(5.86) \quad \mathbf{E}_N \left[ \sum_{x \in \bar{B}_1} \mathbf{1}_{[-1,1]}(\phi(x)) \mid \bar{\psi} \right] \geq \bar{N}_h^d P(\sigma_d \mathcal{N} \in (1 - \delta)u_h + d^2 + [0, 2]) \geq \bar{N}_h^d h^{1-\delta},$$

where we have taken the worst case scenario for  $\bar{\psi}$  and we have added a shift of one to compensate for the boundary effects on the variance, see Remark 5.9: the effect of the boundary on the variance is  $O(h^c)$  so a shift of a unit is largely sufficient because  $u_h = O(\sqrt{\log(1/h)})$ . The probability in the intermediate term turns out to be bounded below by  $h^b$  for every



$b > (1 - \delta)^2$  by a direct Gaussian estimate. So the choice  $b = 1 - \delta$  yields the lower bound of  $h^{1-\delta}$  and (5.86) holds. We now proceed to a second moment estimate to get a concentration result on the number of contacts. The most technical estimate of this step is in the following lemma:

LEMMA 5.15. *When  $|x - y| \geq |\log h|^2$  and  $|v_1|, |v_2| \leq u_h - 1$  we have*

$$(5.87) \quad \mathbf{E}_N [\mathbf{1}_{[-1,1]}(\psi(x) + v_1)\mathbf{1}_{[-1,1]}(\psi(y) + v_2)] \\ \leq (1 + C|\log h|^{-1})\mathbf{E}_N [\mathbf{1}_{[-1,1]}(\psi(x) + v_1)] \mathbf{E}_N [\mathbf{1}_{[-1,1]}(\psi(y) + v_2)] ,$$

with  $C$  a constant that depends only on  $d$ .

PROOF. We need to bound  $\mathbf{E}_N[\mathbf{1}_{[-1,1]}(\psi(y) + v_2) \mid \psi(x)]$  when  $\psi(x) \in [-1, 1] - v_1$ . Let  $\sigma_y^2$  denote the variance of  $\psi(y)$  and  $(\sigma_y^{(x)})^2$  is the variance of  $\psi(y)$  conditional to  $\psi(x)$ : let us remark from now that these variances are bounded below by the variance of the field on one site conditioned to its nearest neighbors, that is  $1/(2d)$ . The covariance of  $\psi(x)$  and  $\psi(y)$ , that is  $\mathbf{E}_N[\psi(x)\psi(y)]$ , is denoted, as usual, by  $G_N(x, y)$ . Recall that, by standard computations on bi-variate Gaussian vectors,  $\mathbf{E}_N[\psi(y) \mid \psi(x)] = c_N(x, y)\psi(x)$  with

$$(5.88) \quad c_N(x, y) := \frac{G_N(x, y)}{\sigma_x^2} \leq 2dG(0, y - x) \leq C|\log h|^{2(2-d)} ,$$

with  $C$  a constant that depends only on  $d$ : in this proof we re-use  $C$  precisely in this sense, possibly updating its value. Moreover for the conditional variance we have the formula

$$(5.89) \quad (\sigma_y^{(x)})^2 = \left(1 - \left(\frac{G_N(x, y)}{\sigma_x \sigma_y}\right)^2\right) \sigma_y^2 = \sigma_y^2 - (c_N(x, y)\sigma_x)^2 .$$

Therefore (5.88) and (5.89) yield

$$(5.90) \quad (\Delta\sigma)^2 := \sigma_y^2 - (\sigma_y^{(x)})^2 = (c_N(x, y)\sigma_x)^2 \leq G(0, 0)C^2 |\log h|^{4(2-d)} .$$

Now we remark that we have

$$(5.91) \quad \mathbf{E}_N[\mathbf{1}_{[-1,1]}(\psi(y) + v_2)] = \frac{1}{\sqrt{2\pi}\sigma_y} \int_{[-1,1]} e^{-\frac{(u-v_2)^2}{2\sigma_y^2}} du , \\ \mathbf{E}_N[\mathbf{1}_{[-1,1]}(\psi(y) + v_2) \mid \psi(x) = z] = \frac{1}{\sqrt{2\pi}\sigma_y^{(x)}} \int_{[-1,1]} e^{-\frac{(u-v_2-c_N(x,y)z)^2}{2(\sigma_y^{(x)})^2}} du ,$$

with some small abuse of notation in the second identity. We aim at controlling the ratio between these two quantities. For this it is sufficient to observe that for every  $|z|, |v_2| \leq u_h$  and  $|u| \leq 1$

$$\begin{aligned}
(5.92) \quad & \frac{\sigma_y}{\sigma_y^{(x)}} e^{\frac{(u-v_2)^2}{2\sigma_y^2} - \frac{(u-v_2-c_N(x,y)z)^2}{2(\sigma_y^{(x)})^2}} \\
&= \sqrt{1 + \left(\frac{\Delta\sigma}{\sigma_y^{(x)}}\right)^2} e^{-\frac{(u-v_2)^2(\Delta\sigma)^2}{2\sigma_y^2(\sigma_y^{(x)})^2} + \frac{(u-v_2)^2 - (u-v_2-c_N(x,y)z)^2}{2(\sigma_y^{(x)})^2}} \\
&\leq \sqrt{1 + \left(\frac{\Delta\sigma}{\sigma_y^{(x)}}\right)^2} e^{\frac{|u-v_2|c_N(x,y)|z|}{(\sigma_y^{(x)})^2}} \leq 1 + \frac{C}{|\log h|},
\end{aligned}$$

where in the last inequality we have used (5.88) and (5.89), obtaining thus an upper bound of  $1 + C \max(|\log h|^{4(2-d)}, u_h^2 |\log h|^{2(2-d)})$  and the worst case estimate is for  $d = 3$ . We have therefore completed the proof of Lemma 5.15.  $\square$

Using (5.87) when  $|x - y| \geq |\log h|^2$  and  $\mathbf{1}_{[-1,1]}(\phi(x))\mathbf{1}_{[-1,1]}(\phi(y)) \leq \mathbf{1}_{[-1,1]}(\phi(x))$  when  $|x - y| < (\log h)^2$  we obtain that

$$\begin{aligned}
(5.93) \quad & \mathbf{E}_N \left[ \sum_{x,y \in \bar{B}_1} \mathbf{1}_{[-1,1]}(\phi(x))\mathbf{1}_{[-1,1]}(\phi(y)) \mid \bar{\psi} \right] \\
&\leq (1 + C|\log h|^{-1}) \left( \mathbf{E}_N \left[ \sum_{x \in \bar{B}_1} \mathbf{1}_{[-1,1]}(\phi(x)) \mid \bar{\psi} \right] \right)^2 \\
&\quad + (\log h)^{2d} \mathbf{E}_N \left[ \sum_{x \in \bar{B}_1} \mathbf{1}_{[-1,1]}(\phi(x)) \mid \bar{\psi} \right].
\end{aligned}$$

This inequality can be written in compact form if we call  $V_N$  the conditional variance of  $\sum_{x \in \bar{B}_1} \mathbf{1}_{[-1,1]}(\phi(x))$  and  $m_N$  the conditional mean, which, by (5.86), is bounded below by  $\bar{N}_h^d h^{1-\delta} \geq h^{-1-\delta}$  (by (5.15)):

$$(5.94) \quad \frac{V_N}{m_N^2} \leq \frac{C}{|\log h|} + \frac{|\log h|^{2d}}{m_N} \leq \frac{2C}{|\log h|}.$$

Therefore, using again (5.86), we see that for every  $\eta > 0$

$$(5.95) \quad \mathbf{P}_N \left( \sum_{x \in \bar{B}_1} \mathbf{1}_{[-1,1]}(\phi(x)) \leq \bar{N}_h^d h^{1-\delta/2} \mid \bar{\psi} \right) \leq \eta,$$

by choosing  $h$  small, and, by recalling (5.69), we see that in the case (i), inequality (5.71) holds.

5.4.2. *The case (ii).* In this case it suffices to observe that since  $H_0 \geq (1 + \delta)u_h$  on  $\bar{B}_3$ , we have  $\bar{\psi} \geq (1 + \delta)u_h - d^2$  and thus

$$(5.96) \quad \mathbf{E}_N \left[ \sum_{x \in \bar{B}_3} \mathbf{1}_{[-1,1]}(\phi(x)) \mid \bar{\psi} \right] \leq \bar{N}_h^d P(\sigma_d \mathcal{N} \geq (1 + \delta)u_h - d^2) \\ \leq \bar{N}_h^d \exp \left( -\frac{((1 + \delta)u_h - d^2)^2}{2\sigma_d^2} \right) \leq \bar{N}_h^d h^{1+\delta},$$

where, in the last step, we have used (4.2) and from this, applying Markov's inequality we have (recall (5.69))

$$(5.97) \quad \mathbf{P}(\mathcal{E}^{(1)} \mid \bar{\psi}) \leq \mathbf{P}_N \left( \sum_{x \in \bar{B}_3} \mathbf{1}_{[-1,1]}(\phi(x)) \geq h\chi(\beta)\bar{N}_h^d \mid \bar{\psi} \right) \leq (\chi(\beta))^{-1} h^{-\delta}.$$

So, also in the case (ii) we have that (5.71) holds.

5.4.3. *The case (iii).* Here again a first moment estimate suffices. As we pointed out

$$\max_{x_0 \in B_{0,k}} |H_0(x)| = H_0(y_3) \leq (1 + 3\delta)u_h$$

for every  $x$  in the level-0 box  $B_{0,k}$  we are considering (the fact that  $|H_0(x)|$  is maximized at  $y_3$  comes from the fact that  $a$  and the  $b_i$ s are positive). Hence  $|\bar{\psi}(x)| \leq (1 + 3\delta)u_h + d^2$  for  $x \in B_{0,k}$  and if we choose  $\delta \leq \varepsilon/6$  we have that for  $h$  sufficiently small

$$(5.98) \quad \mathbf{E}_N \left[ \sum_{x \in B_{0,k}} \mathbf{1}_{(-(1+\varepsilon)u_h, (1+\varepsilon)u_h)}(\phi(x)) \mid \bar{\psi} \right] \leq 2N_0^d P(\sigma_d \mathcal{N} \geq \varepsilon u_h/2) \leq 2N_0^d h^{\varepsilon^2/5},$$

and the Markov inequality immediately yields that the conditional probability of the event  $\mathcal{E}^{(2)}$ , defined in (5.70), can be made arbitrarily small, in particular smaller than  $\eta$ , by choosing  $h$  smaller than a suitable constant (that depends on  $\varepsilon$ ). Therefore also in the case (iii) we obtain (5.71).

This completes the proof of Proposition 5.12.  $\square$

5.5. *The multiscale bound: proof of Proposition 5.10.* Recall the two scale decomposition (5.61) and that we need to control  $\bar{\psi}$ . We start by a notational remark: we have  $\bar{\psi} = \mathbf{E}[\phi_{\Lambda_N} | \mathbb{G}_0]$  and  $\psi = \phi_{\Lambda_N} - \mathbf{E}[\phi_{\Lambda_N} | \mathbb{G}_0]$ , where we set for conciseness  $\mathbf{E}[\cdot | A] := \mathbf{E}[\cdot | \mathcal{F}_A]$ . We recall that we use the notation of  $\phi_A$  for the free field with 0 boundary conditions outside  $A \subset \mathbb{Z}^d$ . In fact, in this section we avoid using  $\mathbf{P}_N$ : the law of  $\phi$  under  $\mathbf{P}_N$  just coincides with the law of  $\phi_{\Lambda_N}$  (and the probability is just denoted by  $\mathbf{P}$ ). By noticing that  $\mathbf{E}[\phi_{\Lambda_N} | \mathbb{G}_J] = 0$  simply because by definition  $\mathbb{G}_J = \tilde{\mathbb{G}}_0 = \Lambda_N^c$  we see that

$$(5.99) \quad \bar{\psi} = \sum_{j=1}^J \psi_j,$$

where

$$(5.100) \quad \psi_j := \mathbf{E}[\phi_{\Lambda_N} | \mathbb{G}_{j-1}] - \mathbf{E}[\phi_{\Lambda_N} | \mathbb{G}_j].$$

By construction and by using the Markov and the Gaussian properties of the field,  $\psi_j(x)$  and  $\psi_j(y)$  are independent if  $x$  and  $y$  belong to different level- $j$  boxes. And of course  $(\psi_j)_{1, \dots, J}$  is a family of independent fields. This independence is going to play a crucial role.

Let us recall now the definitions (5.64)-(5.66). From (5.99) of course we have

$$(5.101) \quad |\nabla_{eg} \bar{\psi}(x)| \leq \sum_{j=1}^J |\nabla_{eg} \psi_j(x)|.$$

*Proof of proof of Proposition 5.10.* With reference to (5.101): we need to estimate the left-hand side and we are going to do so by estimating every level- $j$  separately. We start by estimating the variance of  $\nabla_{eg} \psi_j(x)$  and for this we use

$$(5.102) \quad \begin{aligned} \text{Var}(\nabla_{eg} \psi_j(x)) &= \mathbf{E}[\text{Var}_{\mathbb{G}_j}(\mathbf{E}[\nabla_{eg} \phi_{\Lambda_N}(x) | \mathbb{G}_{j-1}])] \\ &\leq \text{Var}(\mathbf{E}[\nabla_{eg} \phi_{\Lambda_N}(x) | \mathbb{G}_{j-1}]), \end{aligned}$$

where  $\text{Var}_{\mathbb{G}_j}(\cdot)$  is the variance with respect to  $\mathbf{P}(\cdot | \mathbb{G}_j)$ . But we can go even farther: in fact

$$(5.103) \quad \text{Var}(\nabla_{eg}\psi_j(x)) \leq \text{Var}(\mathbf{E}[\nabla_{eg}\phi(x) | \mathbb{G}_{j-1}]),$$

where  $\phi$  is now the infinite volume field (note the difference between the right(most)-hand sides in (5.102) and (5.103)). This is simply because, using the same trick we have repeatedly used up to now,  $\phi$  can be written as sum of two independent fields: the first one is the trace of the infinite volume LFF over the complement of  $\Lambda_N$  harmonically continued in  $\Lambda_N$  and the second one is a free field with Dirichlet boundary conditions outside  $\Lambda_N$ . Hence the variance in right-hand side of (5.103) can be written as the sum of two variances and one of them coincides with the right-hand side of (5.102) and (5.103) holds.

At this point we can apply Lemma A.5, together with Lemma 5.3, to the right-hand side of (5.103) and we see that for every  $j \in \llbracket 1, 2^J \rrbracket$

$$(5.104) \quad \max_{e,g} \max_{k \in \llbracket 1, 2^{d(J-j+1)} \rrbracket} \max_{x \in B_{j-1,k}} \text{Var}(\nabla_{eg}\psi_j(x)) \leq \frac{C}{N_j^{d+3/2}},$$

where the exponent  $3/2$  has been chosen arbitrarily in (1, 2) (we could even choose 2 if we introduce a logarithmic correction, see Lemma A.5), but any number larger than 1 suffices for our purposes.  $C$  is just a  $d$  dependent constant and it has been chosen also to compensate the fact that for readability we replaced  $N_{j-1}$  with  $N_j$ . From (5.104) we directly obtain the Gaussian tail estimate

$$(5.105) \quad \mathbf{P}\left(|\nabla_{eg}\psi_j(x)| \geq \frac{h^{4/d}}{j^2(\log(1/h))^2}\right) \leq 2 \exp\left(-\frac{h^{8/d}N_j^{d+3/2}}{2Cj^4(\log(1/h))^4}\right),$$

uniformly in  $e, g$  and in  $x$  in all the level- $(j-1)$  boxes. Therefore, by a union bound, we have that if  $\check{B}_j$  is the union of the  $2^d$  level- $(j-1)$  boxes contained in the level- $j$  box  $B_j$

$$(5.106) \quad \mathbf{P}\left(\max_{e,g} \max_{x \in \check{B}_j} |\nabla_{eg}\psi_j(x)| \geq \frac{h^{4/d}}{j^2(\log(1/h))^2}\right) \leq 2d^2 N_j^d \exp\left(-\frac{h^{8/d}N_j^{d+3/2}}{2Cj^4(\log(1/h))^4}\right) =: p_j(h).$$

Now we remark that (5.14) implies

$$(5.107) \quad h^{8/d}N_0^{d+3/2} \geq h^{(8/d)-(2/d)(d+3/2)} \geq h^{-1/3},$$

and for  $h$  sufficiently small we have for every  $j \in \llbracket 1, J \rrbracket$

$$(5.108) \quad \log p_j(h) \leq -\frac{h^{-1/3}2^{j(d+3/2)}}{2Cj^4(\log(1/h))^4} + d \log N_j + \log(2d^2) \leq -\frac{h^{-1/4}2^{jd} \exp(j)}{j^4},$$

because  $d \log N_j \leq jd \log 2 + \log(1/h)$  for  $h$  small and in the last step we have used  $2^{3/2} > e$ . We record explicitly the bound that we will use

$$(5.109) \quad p_j(h) \leq \exp\left(-\frac{h^{-1/4}2^{jd} \exp(j)}{j^4}\right),$$

and that we choose  $h$  sufficiently small to guarantee that  $\sup_j p_j(h) < 1$ . But we are going to choose  $h$  small to satisfy also the stronger requirement that for given  $\eta > 0$  we have  $p_j(h) < (1/2)^{2j^2/\eta}$  for every  $j$ : we are going to choose  $\eta := 6\delta/\pi^2$ , with  $\delta$  the constant entering the definition (5.66) of  $\mathcal{B}(N, \delta)$ , but this is going to be irrelevant till the very last steps of the proof. We can then apply the binomial bound in Lemma A.2 and obtain

$$(5.110) \quad \mathbf{P}\left(\# \left\{ k \in \llbracket 1, 2^{d(J-j+1)} \rrbracket : \max_{e,g} \max_{x \in \tilde{B}_{j,k}} |\nabla_{eg} \psi_j(x)| \geq \frac{h^{4/d}}{j^2(\log(1/h))^2} \right\} \geq \frac{\eta}{j^2} 2^{d(J-j+1)}\right) \\ \leq (p_j(h))^{\eta 2^{d(J-j+1)}/(2j^2)} \leq \exp\left(-\eta \frac{h^{-1/4}2^{d(J+1)} \exp(j)}{2j^6}\right).$$

Using the independence of  $(\psi_j)_{j=1, \dots, J}$  we can control all levels at the same time:

$$\mathbf{P}\left(\exists j \in \llbracket 1, J \rrbracket \text{ s.t. } \# \left\{ k : \max_{e,g} \max_{x \in \tilde{B}_{j,k}} |\nabla_{eg} \psi_j(x)| \geq \frac{h^{4/d}}{j^2(\log(1/h))^2} \right\} \geq \frac{\eta}{j^2} 2^{d(J-j+1)}\right)$$

$$\begin{aligned}
&\leq 1 - \prod_{j=1}^J \left( 1 - \exp \left( -\eta \frac{h^{-1/4} 2^{d(J+1)} \exp(j)}{2j^6} \right) \right) \\
&\leq 1 - \prod_{j=1}^J \left( 1 - \exp \left( -2\eta h^{-1/5} 2^{d(J+1)} j \right) \right) \\
(5.111) \quad &\leq 1 - \exp \left( \frac{1}{2} \sum_{j=1}^{\infty} \exp \left( -2\eta h^{-1/5} 2^{d(J+1)} j \right) \right) \\
&\leq \exp \left( -\eta h^{-1/5} 2^{d(J+1)} \right) \leq \exp \left( -h^{2-\frac{1}{6}} N^d \right),
\end{aligned}$$

where in the last line we have used that  $2^{dJ} = N^d h^{2-o(1)}$  (cf. (5.3)-(5.4)) Now we recall (5.101) and the definition (5.66) of the event  $\mathcal{B}(N, \delta)$  and we see that on the complementary of the event whose probability is estimated in (5.111) we have that

$$(5.112) \quad \max_{e,g} \|\nabla_{eg} \bar{\psi}\|_{\infty,k} \leq \frac{h^{4/d}}{(\log(1/h))^2} \sum_{j=1}^{\infty} \frac{1}{j^2} \leq \frac{1}{N_0^2},$$

except for at most a fraction  $\delta = \eta \sum_{j \in \mathbb{N}} j^{-2}$  of the  $K = 2^{dJ}$  level-0 boxes. This completes the proof of Proposition 5.10.  $\square$

## APPENDIX A: TECHNICAL ESTIMATES

**A.1. Standard binomial bounds.**  $\text{Bin}(n, p)$  denotes a binomial random variable of parameters  $n$  and  $p$ .

LEMMA A.1. *For  $\Delta \in [0, 1/6]$  we have*

$$(A.1) \quad P(\text{Bin}(n, p) \leq p\Delta n) \leq \exp(-np/2),$$

and for  $\Delta \geq 3$

$$(A.2) \quad P(\text{Bin}(n, p) \geq p\Delta n) \leq \exp(-np/2).$$

*Proof.* For  $\Delta \in [0, 1/p]$  set with

$$(A.3) \quad f(p, \Delta) := p\Delta \log \Delta + (1 - p\Delta) \log((1 - p\Delta)/(1 - p)).$$

By the exponential Markov inequality we have that for every  $\Delta \in [0, 1]$

$$(A.4) \quad P(\text{Bin}(n, p) \leq p\Delta n) \leq \exp(-nf(p, \Delta)),$$

and for every  $\Delta \in [1, 1/p]$

$$(A.5) \quad P(\text{Bin}(n, p) \geq p\Delta n) \leq \exp(-nf(p, \Delta)) .$$

The convexity of  $p \mapsto f(p, \Delta)$  yields  $f(p, \Delta) \geq p\partial_p f(0, \Delta) = p(1 - \Delta(1 + \log(1/\Delta)))$ , so the result follows because  $1 - \Delta(1 + \log(1/\Delta)) \geq 1/2$  for  $\Delta \leq 1/6$  and  $\Delta \geq 3$ .  $\square$

LEMMA A.2. *Given  $\eta > 0$  such that  $p \leq (1/2)^{2/\eta}$ , we have  $P(\text{Bin}(n, p) \geq \eta n) \leq p^{\eta n/2}$ .*

PROOF. It follows by remarking that

$$(A.6) \quad P(\text{Bin}(n, p) \geq \eta n) = \sum_{k \geq \eta n} \binom{n}{k} p^k (1-p)^{n-k} \leq p^{\eta n} \sum_{k \geq \eta n} \binom{n}{k} \leq p^{\eta n} 2^n .$$

$\square$

**A.2. A disorder cut-off estimate.** Let us consider a finite subset  $A$  of  $\mathbb{Z}^d$  and an arbitrary measure  $\mathbf{P}$  on  $\{0, 1\}^{\mathbb{Z}^d}$  and the corresponding product  $\sigma$ -algebra, with  $\mathcal{B}$  a measurable event in this  $\sigma$ -algebra. We set  $Z_{A, \omega}(\mathcal{B}) := \mathbf{E}[\exp(\sum_{x \in A} \omega_x(\beta, h)\delta_x) \mathbf{1}_{\mathcal{B}}]$  where  $(\delta_x)_{x \in \mathbb{Z}^d}$  denotes the coordinate process on  $\{0, 1\}^{\mathbb{Z}^d}$ . For  $K > 0$  we introduce also  $\bar{\omega}_x := \omega_x \mathbf{1}_{|\omega_x| \leq K}$  and

$$(A.7) \quad L_K := \mathbb{E}[|\omega_x|; |\omega_x| > K] .$$

LEMMA A.3. *We have*

$$(A.8) \quad |\mathbb{E}[\log Z_{A, \omega}(\mathcal{B})] - \mathbb{E}[\log Z_{A, \bar{\omega}}(\mathcal{B})]| \leq \beta |A| L_K .$$

PROOF. We can assume  $\mathcal{P}(\mathcal{B}) > 0$ . We have

$$(A.9) \quad \mathbb{E}[\log Z_{A, \omega}] - \mathbb{E}[\log Z_{A, \bar{\omega}}] = \mathbb{E} \log \mathbf{E}_{A, \bar{\omega}} \left[ \exp \left( \beta \sum_{x \in A} \omega_x \mathbf{1}_{\{|\omega_x| > K\}} \delta_x \right) \right] ,$$



where we have introduced the measure  $\mathbf{P}_{A,\bar{\omega}}$  associated to the partition function  $Z_{A,\bar{\omega}}(\mathcal{B})$  neglecting  $\mathcal{B}$  in the notation for conciseness. As we trivially have (for every realization of  $\phi$ )

$$(A.10) \quad \sum_{x \in A} \omega_x \mathbf{1}_{\{\omega_x < -K\}} \leq \sum_{x \in A} \omega_x \mathbf{1}_{\{|\omega_x| > K\}} \delta_x \leq \sum_{x \in A} \omega_x \mathbf{1}_{\{\omega_x > K\}},$$

the right-hand side of (A.9) is smaller, respectively larger, than

$$(A.11) \quad \beta \mathbb{E} \left[ \sum_{x \in A} \omega_x \mathbf{1}_{\{\omega_x > K\}} \right], \quad \text{respectively } \beta \mathbb{E} \left[ \sum_{x \in A} \omega_x \mathbf{1}_{\{\omega_x < -K\}} \right],$$

both of which in absolute value are smaller than

$$(A.12) \quad \beta |A| \mathbb{E} [\omega_x \mathbf{1}_{\{|\omega_x| > K\}}] = \beta |A| L_K.$$

□

**A.3. Harmonic extension estimates.** For the next result  $B \neq \emptyset$  is an arbitrary finite connected subset of  $\mathbb{Z}^d$  and  $H = H_B$  is the harmonic extension of the trace on  $\mathbb{Z}^d \setminus B$  of a LFF  $\phi$ , that is  $\Delta H(x) = 0$  for every  $x \in B$  and  $H(x) = \phi(x)$  for  $x \in \mathbb{Z}^d \setminus B$ . We recall for this section that  $G(\cdot, \cdot)$  is the Green function of the walk in  $\mathbb{Z}^d$ , see Section 1.2.

LEMMA A.4. *For  $d \geq 3$  for every  $x \in B$  we have*

$$(A.13) \quad \text{Var}(H(x)) \leq C_d \left( \left( \text{dist}(x, \mathbb{Z}^d \setminus B) + 1 \right)^{2-d} \right),$$

with  $C_d > 0$  the constant appearing in (1.2).

PROOF. We use that for  $x \in B$  we have

$$(A.14) \quad \text{Var}(H(x)) = \text{Var}(\phi(x)) - \text{Var}(\phi_x^B),$$

with  $\phi^B$  a free field with zero boundary conditions in  $\mathbb{Z}^d \setminus B$ . A proof of (A.14) is for example in [6, p. 1676]. If we reinterpret this formula in random walk terms we have that the variance under analysis is the expected number of visits to  $x$  (by the walk that starts from  $x$ ) after hitting the external boundary of  $B$ . If we call  $p_x(z)$ ,  $z \in \mathbb{Z}^d \setminus B$ , the hitting probability of the

walk issued from  $x$  – of course  $p_x(z) > 0$  only if  $\text{dist}(z, B) = 1$  – then, by using (1.2), we get to

$$(A.15) \quad \begin{aligned} \text{Var}(H(x)) &= \sum_z p_x(z) G(z, x) \\ &\leq C_d \sum_z \frac{p_x(z)}{(|z-x|+1)^{d-2}} \leq \frac{C_d}{(\text{dist}(x, \mathbb{Z}^d \setminus B) + 1)^{d-2}}. \end{aligned}$$

□

Next we consider the case of  $B = \mathring{\Lambda}_M = \llbracket 1, M-1 \rrbracket^d$  of (the trace of) and infinite volume LFF on  $\mathbb{Z}^d \setminus \mathring{\Lambda}_M$ . So now  $H = H_{\mathring{\Lambda}_M}$ . The covariance of  $H$  can be written for  $x, y \in \Lambda_M$  as

$$(A.16) \quad \bar{G}(x, y) := \mathbf{E}_x \left[ \sum_{n=\tau_M}^{\infty} \mathbf{1}_{\{X_n=y\}} \right],$$

with  $\tau_M$  the hitting time of the internal boundary of  $\Lambda_M$  as in (1.1).

LEMMA A.5. *For every  $d \geq 3$  and every  $\kappa > 1$  there is  $C > 0$  such that for every  $r \geq 2$*

$$(A.17) \quad \sup_{x: \text{dist}(x, \Lambda_M^c) \geq r} \text{Var}(H(x) - H(x+e) - H(x+g) + H(x+e+g)) \leq C \frac{(\log(r))^\kappa}{r^{d+2}}.$$

*Proof.* Setting  $p_M(x, y) := P_x(X_{\tau_M} = y)$  we have

$$(A.18) \quad \begin{aligned} \text{Var}(H(x) - H(x+e) - H(x+g) + H(x+e+g)) &= \\ &\sum_{y \in \partial \Lambda_M} [p_M(x, y) - p_M(x+e, y) - p_M(x+g, y) + p_M(x+e+g, y)] \\ &\quad \times (G(y, x) - G(y, x+e) - G(y, x+g) + G(y, x+e+g)). \end{aligned}$$

The proof therefore follows if we show that for every  $i, j \in \{1, \dots, d\}$  and every  $x$  and  $y$

$$(A.19) \quad |G(y, x) - G(y, x+e) - G(y, x+g) + G(y, x+e+g)| \leq C|x-y|^d,$$

with  $C$  a positive constant that depends only on  $d$ , and that for every  $x \in \Lambda_M$  at distance at least  $r$  from  $\Lambda_M^c$  we have

$$(A.20) \quad \sum_{y \in \partial \Lambda_M} |p_M(x, y) - p_M(x + e, y) - p_M(x + g, y) + p_M(x + e + g, y)| \leq Cr^{-2}(\log r)^\kappa,$$

with  $C$  a positive constant that depends only on  $d$  and  $\kappa$ .

For what concerns (A.19), the bound follows directly from the asymptotic Green function estimate in the limit when  $|x| \rightarrow \infty$  [34, Th. 4.3.1],

$$(A.21) \quad G(0, x) - \frac{c_d}{|x|^{d-2}} = O\left(\frac{1}{|x|^d}\right),$$

where  $c_d$  is a positive constant.

For what concerns (A.20) we observe the left-hand side of (A.20) coincides with four times the total variation distance between the distribution of  $X_{\tau_M}$  for the simple random walk starting with respective initial condition  $\frac{1}{2}(\delta_x + \delta_{x+e+g})$  and  $\frac{1}{2}(\delta_{x+e} + \delta_{x+g})$ . The factor four comes from the usual factor two that relates the  $L^1$  norm and the total distance and the fact two that comes from normalizing the initial measures. We let  $P_1$  and  $P_2$  denote the respective law of these two walks. Set  $t := r^2/(\log r)^\kappa$ . We have

$$(A.22) \quad \|P_1(X_{\tau_M} \in \cdot) - P_2(X_{\tau_M} \in \cdot)\|_{TV} \leq \|P_1(X_t \in \cdot) - P_2(X_t \in \cdot)\|_{TV} + P_1(\tau_M < t) + P_2(\tau_M < t).$$

The validity of (A.22) follows by observing that

$$(A.23) \quad \sum_{y \in \partial \Lambda_M} |P_1(X_{\tau_M} = y, \tau_M \geq t) - P_2(X_{\tau_M} = y, \tau_M \geq t)| = \sum_{y \in \partial \Lambda_M} \left| \sum_{x \in \Lambda_M} (P_1(X_{\tau_M} = y, X_t = x, \tau_M \geq t) - P_2(X_{\tau_M} = y, X_t = x, \tau_M \geq t)) \right| \\ = \sum_{y \in \partial \Lambda_M} \left| \sum_{x \in \Lambda_M} P_x(X_{\tau_M} = y) (P_1(X_t = x, \tau_M \geq t) - P_2(X_t = x, \tau_M \geq t)) \right| \\ \leq \sum_{x \in \Lambda_M} \left( \sum_{y \in \partial \Lambda_M} P_x(X_{\tau_M} = y) \right) |P_1(X_t = x, \tau_M \geq t) - P_2(X_t = x, \tau_M \geq t)| \\ \leq 2 \|P_1(X_t \in \cdot) - P_2(X_t \in \cdot)\|_{TV} + P_1(\tau_M < t) + P_2(\tau_M < t),$$

where  $P_x$  is used in the obvious sense. From (A.23) one easily obtains (A.22)

Now we estimate the right-hand side of (A.22): for the last two terms we observe that, with  $P = P_1$  or  $P = P_2$ , we can apply directly [34, Prop. 2.1.2, part 2] and we have that there exists a constant  $C$  that depends only on  $d$  such that

$$(A.24) \quad P(\tau_M < t) \leq C \exp(-r^2/(Ct)) = C \exp(-(\log(t))^\kappa)/C,$$

which is much smaller than any power than  $1/t$  because  $\kappa > 1$ .

We can therefore focus on the the right-hand side of (A.22). We have (translating every coordinate by  $x$  and using the symmetries),

$$(A.25) \quad \|P_1(X_n \in \cdot) - P_2(X_n \in \cdot)\|_{TV} = \frac{1}{4} \sum_{z \in \mathbb{Z}^d} |p_t(z) - p_t(z+e) - p_t(z+g) + p_t(z+e+g)|,$$

where  $p_t(z) = P_0(X_t = z)$ . Using the Local Central Limit Theorem [34, Theorem 2.1.1 and Theorem 2.1.3] we have for some  $d$  dependent constant  $C$

$$(A.26) \quad \left| p_t(z) - \frac{1}{(\sqrt{2\pi t})^{d/2}} e^{-\frac{|z|^2}{2t}} \right| \leq C \left( t^{-\frac{d+2}{2}} \left( \frac{|z|^4}{t^2} + 1 \right) e^{-\frac{|z|^2}{2t}} + t^{-\frac{d+3}{2}} \right).$$

On the other hand, it is elementary to see that  $p_t(z) \leq \exp(-z^2/t)$  so we choose  $b \in (0, 1/(2d))$  and we have

$$(A.27) \quad \sum_{z \in \mathbb{Z}^d} |p_t(z) - p_t(z+e) - p_t(z+g) + p_t(z+e+g)| \leq C t^{-d/2} \sum_{|z| \leq t^{b+1/2}} \left| e^{-\frac{|z|^2}{2t}} + e^{-\frac{|z+e+g|^2}{2t}} - e^{-\frac{|z+e|^2}{2t}} - e^{-\frac{|z+g|^2}{2t}} \right| + C t^{-d/2} \sum_{|z| \leq t^{b+1/2}} \left( \frac{1}{t} \left( \frac{|z|^4}{t^2} + 1 \right) e^{-\frac{|z|^2}{2t}} + \frac{1}{t^{3/2}} \right) + 2 \sum_{|z| > t^{b+1/2}} \exp(-z^2/t).$$

It is straightforward to see that the last line is bounded by  $C/t$ . We are therefore left with controlling the second line of (A.27): we use Taylor formula and we obtain that there exists  $C > 0$  such that for every choice of  $i$  and  $j$  and for every  $|z| \leq t^{b+1/2}$  we have

$$(A.28) \quad \left| e^{-\frac{|z|^2}{2t}} + e^{-\frac{|z+e+g|^2}{2t}} - e^{-\frac{|z+e|^2}{2t}} - e^{-\frac{|z+g|^2}{2t}} \right| \leq \frac{C}{2t} \left( 1 + \frac{z_i^2 + z_j^2}{t} \right) e^{-\frac{|z|^2}{2t}} \leq \frac{C}{t},$$

where the last inequality holds for  $t$  sufficiently large. Altogether we obtain that

$$(A.29) \quad \|P_1[X_t \in \cdot] - P_2[X_t \in \cdot]\|_{TV} \leq \frac{C}{t},$$

for a  $C$  depending only on  $d$  and  $\kappa$  and the proof of (A.20) is complete.

The proof of Lemma A.5 is thus completed.  $\square$

## REFERENCES

- [1] K. S. Alexander, *The effect of disorder on polymer depinning transitions*, Commun. Math. Phys. **279** (2008), 117-146.
- [2] K.S. Alexander and N. Zygouras, *Quenched and annealed critical points in polymer pinning models*, Commun. Math. Phys. **291** (2009) 659-689.
- [3] Q. Berger, G. Giacomin and H. Lacoïn, *Disorder and critical phenomena: the  $\alpha = 0$  copolymer model*, Probab. Theory Rel. Fields, **174** (2019), 787–819.
- [4] Q. Berger and H. Lacoïn, *Pinning on a defect line: characterization of marginal disorder relevance and sharp asymptotics for the critical point shift*, J. Inst. Math. Jussieu **17** (2018), 305-346.
- [5] D. Bertacchi and G. Giacomin, *Enhanced interface repulsion from quenched hard-wall randomness*, Probab. Theory Rel. Fields, **124** (2002), 487-516.
- [6] E. Bolthausen, J.-D. Deuschel and G. Giacomin, *Entropic repulsion and the maximum of the two-dimensional harmonic crystal*, Ann. Probab. **29** (2001), 1670-1692.
- [7] E. Bolthausen, J.-D. Deuschel and O. Zeitouni, *Entropic repulsion of the lattice free field*, Commun. Math. Phys. **170** (1995), 417-443.
- [8] E. Bolthausen, J.-D. Deuschel and O. Zeitouni, *Absence of a wetting transition for a pinned harmonic crystal in dimensions three and larger*, J. Math. Phys. **41** (2000), 1211-1223.
- [9] D. C Brydges, G. Guadagni and P. K. Mitter, *Finite range decomposition of Gaussian processes*, J. Statist. Phys. **115** (2004), 415-449.
- [10] P. Caputo, E. Lubetzky, F. Martinelli, A. Sly, and F.L. Toninelli *Dynamics of (2+1)-dimensional SOS surfaces above a wall: Slow mixing induced by entropic repulsion* Ann. Probab. **42** (2014) 1516-1589.
- [11] F. Caravenna and F. den Hollander, *A general smoothing inequality for disordered polymers*, Elec. Comm. Probab. **18** (2013), 1-15.
- [12] X. Chen, V. Dagar, B. Derrida, Y. Hu, M. Lifshits and Z. Shi, *The Derrida–Retaux conjecture on recursive models*, Ann. Probab. **49** (2021), 637-670.
- [13] A. Chiarini, A. Cipriani and R. S. Hazra, *A note on the extremal process of the supercritical Gaussian Free Field*, Electron. Comm. Probab. **20** (2015) 74.
- [14] L. Coquille and P. Miloš, *A note on the discrete Gaussian free field with disordered pinning on  $\mathbb{Z}^d$ ,  $d \geq 2$* , Stoch. Proc. and Appl. **123** (2013) 3542-3559.
- [15] B. Derrida, G. Giacomin, H. Lacoïn and F. L. Toninelli, *Fractional moment bounds and disorder relevance for pinning models*, Commun. Math. Phys. **287** (2009) 867–887.
- [16] J.-D. Deuschel, *Entropic repulsion of the lattice free field. II. The 0-boundary case*, Commun. Math. Phys. **181** (1996), 647-665.
- [17] J.-D. Deuschel and G. Giacomin, *Entropic repulsion for the free field: pathwise characterization in  $d \geq 3$* , Commun. Math. Phys. **206** (1999), 447-462.

- [18] S. Finch, *Mathematical Constants*, Encyclopedia of Mathematics and its Applications, Cambridge University Press, 2003.
- [19] M. E. Fisher, *Walks, walls, wetting, and melting*, J. Statist. Phys. **34** (1984), 667-729.
- [20] R. Fernández, J. Fröhlich, and A. D. Sokal, *Random walks, critical phenomena, and triviality in quantum field theory*, Texts and Monographs in Physics, Springer-Verlag, Berlin, 1992.
- [21] S. Friedli and Y. Velenik, *Statistical mechanics of lattice systems. A concrete mathematical introduction*, Cambridge University Press, Cambridge, 2018.
- [22] T. Funaki, *Stochastic interface models*, École d'été de probabilités de Saint-Flour XL-2010, Lecture Notes in Mathematics **1869**, Springer, 2005 pp 103-274.
- [23] G. Giacomin, *Random polymer models*, Imperial College Press, World Scientific (2007).
- [24] G. Giacomin, *Disorder and critical phenomena through basic probability models*, École d'été de probabilités de Saint-Flour XL-2010, Lecture Notes in Mathematics **2025**, Springer, 2011.
- [25] G. Giacomin and H. Lacoin, *Pinning and disorder relevance for the lattice Gaussian free field*, J. Eur. Math. Soc. (JEMS) **20** (2018), 199-257.
- [26] G. Giacomin and H. Lacoin, *Disorder and wetting transition: The pinned harmonic crystal in dimension three or larger*, Ann. Appl. Probab. **28** (2018), 577-606.
- [27] G. Giacomin and F. L. Toninelli, *Smoothing effect of quenched disorder on polymer depinning transitions*, Commun. Math. Phys. **266** (2006), 1-16.
- [28] A. B. Harris, *Effect of random defects on the critical behaviour of Ising models* J. Phys. C **7** (1974), 1671-1692.
- [29] H. Lacoin, *The martingale approach to disorder irrelevance for pinning models*, Elec. Comm. Probab. **15** (2010) 418-427.
- [30] H. Lacoin, *Pinning and disorder for the Gaussian free field II: the two dimensional case*, Ann. Sci. Éc. Norm. Sup **52** (2019), 1331-1401.
- [31] H. Lacoin, *Wetting and layering for Solid-on-Solid I: Identification of the wetting point and critical behavior*, Commun. Math. Phys. **362** (2018) 1007-1048.
- [32] H. Lacoin, *Wetting and layering for Solid-on-Solid II: Layering transitions, Gibbs states, and regularity of the free energy*, Journal de l'École polytechnique **7** (2020), 1-62.
- [33] H. Lacoin, *Solid-On-Solid interfaces with disordered pinning*, Commun. Math. Phys. **383** (2021), 489-536.
- [34] G. F. Lawler and V. Limic, *Random Walk: A Modern Introduction*, Cambridge Studies in Advanced Mathematics **123**, 2010.
- [35] J. L. Lebowitz and C. Maes, *The effect of an external field on an interface, entropic repulsion*, J. Statist. Phys. **46** (1987), 39-49.
- [36] S. Sheffield, *Random surfaces*, Astérisque **304** (2005) vi+175 pp.
- [37] F. L. Toninelli, *A replica-coupling approach to disordered pinning models*, Commun. Math. Phys. **280** (2008) 389-401.
- [38] Y. Velenik, *Localization and delocalization of random interfaces*, Probab. Surv. **3** (2006), 112-169.

UNIVERSITÉ DE PARIS,  
 LABORATOIRE DE PROBABILITÉS,  
 STATISTIQUES ET MODÉLISATION,  
 UMR 8001,  
 F-75205 PARIS, FRANCE,  
 E-MAIL: giambattista.giacomin@u-paris.fr

IMPA  
 INSTITUTO NACIONAL DE MATEMÁTICA PURA E APLICADA,  
 ESTRADA DONA CASTORINA 110,  
 RIO DE JANEIRO, BRASIL 22460-320,  
 E-MAIL: lacoin@impa.br