On the discrete Gaussian Free Field with disordered pinning on Z^d, $d \geq 2$

Coquille, Loren; Miłoś, Piotr

On the discrete Gaussian Free Field with disordered pinning on $\mathbb{Z}^d$, $d \geq 2$

Loren Coquille, Piotr Miłoś

July 26, 2012

Abstract

We study the discrete massless Gaussian Free Field on $\mathbb{Z}^d$, $d \geq 2$, in presence of two types of random environments : (1) δ-pinning at height 0 of inhomogenous i.i.d. Bernoulli strengths; (2) square-well potential supported on a finite strip with i.i.d. Bernoulli reward/penalty coefficients $\mathbf{e}$. We prove that the quenched free energy associated to these models exists in $\mathbb{R}^+$, is self-averaging, and strictly smaller than the annealed free energy (whenever the latter is strictly positive). Moreover, for model(2), we prove that in the plane $(\mathbb{V}\mathrm{ar}(\mathbf{e}), \mathbb{E}(\mathbf{e}))$, the quenched critical line (separating the phases of positive and zero free energy) lies strictly below the line $\mathbb{E}(\mathbf{e}) = 0$, showing in particular that there exists a non trivial region where the field is localized though repulsed on average by the environment.

Keywords : Random interfaces, random surfaces, pinning, disordered systems, Gaussian free field.
MSC2010 : 60K35, 82B44, 82B41.

Contents

1 Introduction

1.1 The models

We study two models belonging to the so-called effective interface class. For Λ being a finite subset of $\mathbb{Z}^d$ (denoted by $\Lambda \Subset \mathbb{Z}^d$), let $\varphi = (\varphi_x)_{x\in\Lambda}$ represent the heights of the interface above or below sites in Λ. The interface is thus living in $\mathbb{Z}^{d+1}$.

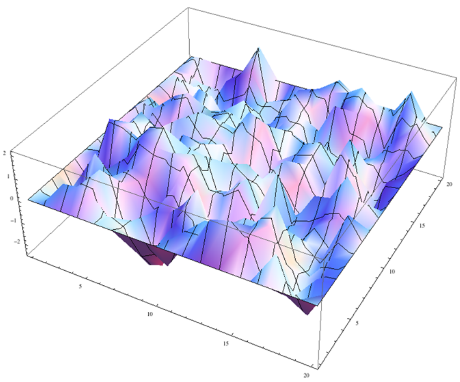

Figure 1: A 2-dimensional interface.

The variables φ_x can also be seen as continuous unbounded (spin) variables, we will refer to φ as “the interface” or “the field”. The models are defined in terms of a Gaussian pair potential and a random potential term at interface height zero (or close to zero).
In the first model, we consider only attractive potential. The so-called δ-pinning measure is given by :

$$\mu_\Lambda^{\mathbf{e}}(d\varphi) = \frac{1}{Z_\Lambda^{\mathbf{e}}} \exp\left(-\frac{\beta}{4d}\sum_{x\sim y}(\varphi_x - \varphi_y)^2\right) \prod_{x\in\Lambda}\left(d\varphi_x + \sqrt{\beta} e_x \delta_0(d\varphi_x)\right) \prod_{y\in\Lambda^c} \delta_0(d\varphi_y), \tag{1}$$

where $\beta > 0$ is the inverse temperature, $\Lambda \Subset \mathbb{Z}^d$, $x \sim y$ denotes nearest neighbour relation in $\Lambda \cup \partial\Lambda$ and δ_0 is the Dirac mass at height 0. $\mathbf{e} := (e_x)_{x\in\Lambda}$ is given by positive i.i.d. Bernoulli random variables on $\mathbb{Z}^d$ standing for the quenched disorder, which thus represents a random attraction strength at height 0. $Z_\Lambda^{\mathbf{e}}$ is the partition function, i.e. it normalizes $\mu_\Lambda^{\mathbf{e}}$ so it is a probability measure.

In the second model, we consider attractive/repulsive potential, and define the following measure :

$$\mu_{\Lambda,a}^{\mathbf{e}}(d\varphi) = \frac{1}{Z_{\Lambda,a}^{\mathbf{e}}} \exp\left(-\frac{\beta}{4d} \sum_{x\sim y} (\varphi_x - \varphi_y)^2 + \beta \sum_{x\in\Lambda} e_x \mathbb{1}_{[\varphi_x \in [-a,a]]} \right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y), \tag{2}$$

where $\mathbb{1}_{[.]}$ denotes the indicator function, $a > 0$ and $\mathbf{e}$ is again a Bernoulli product random field on $\mathbb{Z}^d$, which is now allowed to take negative values, thus repulsing the interface.
Note that both models contain two levels of randomness. The first one is $\mathbf{e}$ which we will refer to as "the environment". The second one is the actual interface model, depending on $\mathbf{e}$. By performing a standard scaling we can eliminate the inverse temperature β, thus henceforth we assume $\beta = 1$; cf. Remark 2.1.
The dimensions 1 and 2 are physically relevant as interface models. In this paper we focus on $d \geq 2$ since 1-dimensional models have been well-studied in the last decade (see Section (1.3.1) below).
The questions we are addressing in this framework are the usual ones concerning statistical mechanics models in random environment : Is the quenched free energy non-random ? Does it differ from the annealed one ? Can we give a physical meaning to the strict positivity (resp. vanishing) of the free energy ? What can be said concerning the quenched and annealed critical lines (surfaces) in the space of the relevant parameters of the system ?
For models defined by (1) and (2), we prove that the quenched free energy exists and is non-random. It is always strictly positive in case of model (1), and strictly smaller than the annealed free energy for both models.
For model (2), we also investigate the corresponding phase diagram : in the plane $(\mathbb{V}\mathrm{ar}_{\mathbf{e}}(\mathbf{e}), \mathbb{E}(\mathbf{e}))$ (i.e. the variance and expectation of $\mathbf{e}$), we prove that the quenched critical line (separating the phases of positive and zero free energy) lies strictly below the line $\mathbb{E}(\mathbf{e}) = 0$. Thus there exists a non trivial region where the field is localized though repulsed on average by the environment.
We do not treat here the questions concerning critical exponents of the system, nor the order of the phase transition, in presence (absence) of disorder. This has been done for a certain class of 1-dimensional systems (see below), and is a much more difficult problem in dimension 2 and above.

1.2 Known results about homogenous models

Our work relies on [5],[9]. They focus on the case $e_x = const$ which we will refer to as the homogenous pinning model. A review article about localization and delocalization of random interfaces (in a non-random environment) can be found in [23]. We briefly describe some important results here.
Let us now consider the model (1) [1] with $e_x = \varepsilon$ with $\varepsilon \geq 0$ for all $x \in \mathbb{Z}^d$, and denote the associated measure by $\mu_\Lambda^{\varepsilon,0}$, the superscript 0 standing for 0 boundary condition. The so-called free measure is

$$\mu_\Lambda^{0,0} = (Z_\Lambda^{0,0})^{-1} \exp\left(-\frac{\beta}{4d} \sum_{x\sim y} (\varphi_x - \varphi_y)^2 \right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y),$$

is centered Gaussian, and thus characterized by its covariance matrix. There exists a useful representation of the latter in terms of the Green function of the simple symmetric random walk on $\mathbb{Z}^d$

$$\mu_\Lambda^{0,0}(\varphi_x\varphi_y) = E_x\left[\sum_{n=0}^{\tau_\Lambda - 1} \mathbb{1}_{[X_n = y]} \right], \tag{3}$$

[1] Similar results hold for the homogenous version of model (2) with $e_x = \varepsilon > 0$ for all $x \in \mathbb{Z}^d$, see [5].

where, $((X_n)_n, P_x)$ is a simple symmetric random walk on $\mathbb{Z}^d$, starting at x, and τ_Λ is its exit time from Λ. As the simple random walk is recurrent in $d = 2$ and transient in $d \geq 3$, the variances diverge in $d = 2$, and stay bounded for $d \geq 3$ as $\Lambda \uparrow \mathbb{Z}^d$. We refer to [21] for random walks estimates.
To analyze the properties of the model with pinning ($\varepsilon > 0$), it is convenient to map it onto a model of a random walk in an annealed random environment of killing obstacles. To achieve that, we can expand the product $\prod_{x\in\Lambda} d\varphi_x + \varepsilon\delta_0(d\varphi_x)$, as in a high temperature expansion. Let $\mathcal{A} = \{x \in \Lambda : \varphi_x = 0\}$ be the random variable describing the set of pinned sites. By simple calculations (see (16) for more general case), for any measurable function $f : \mathbb{R}^\Lambda \mapsto \mathbb{R}$, we have :

$$\mu_\Lambda^{\varepsilon,0}(f) = \sum_{A\subset\Lambda} \mu_\Lambda^{\varepsilon,0}(f \mid \mathcal{A} = A) \cdot \mu_\Lambda^{\varepsilon,0}(\mathcal{A} = A) = \sum_{A\subset\Lambda} \mu_{\Lambda\setminus A}^{0,0}(f) \cdot \mu_\Lambda^{\varepsilon,0}(\mathcal{A} = A). \tag{4}$$

For simplicity we denote $\nu_\Lambda^\varepsilon(A) := \mu_\Lambda^{\varepsilon,0}(\mathcal{A} = A)$. Conditionally on $\mathcal{A}$, $\mu_\Lambda^{\varepsilon,0}$ is a Gaussian measure with covariances described by the random walk representation above.
Important results concerning the pinned sites distribution were obtained in [5]. First of all, the measure ν_Λ^ε is strong FKG in the sense of [10]. Moreover, it can be compared with i.i.d. Bernoulli distributions. Let $\mathbb{B}_\Lambda^\alpha$ be the Bernoulli product measure (or site percolation) of parameter $\alpha \in [0, 1]$ in the box Λ, namely the measure on subsets of Λ given by :

$$\mathbb{B}_\Lambda^\alpha(\mathcal{A} = A) = \alpha^{|A|}(1 - \alpha)^{|\Lambda|-|A|}$$

By [5, Theorem 2.4] there exist constants $0 < c_-(d) < c_+(d) < \infty$ such that for any Λ, any $B \subset \Lambda$, and ε sufficiently small in $d = 2$, we have

$$\mathbb{B}_\Lambda^{c_-(d)g(\varepsilon)}(\mathcal{A} \cap B = \varnothing) \leq \nu_\Lambda^\varepsilon(\mathcal{A} \cap B = \varnothing) \leq \mathbb{B}_\Lambda^{c_+(d)g(\varepsilon)}(\mathcal{A} \cap B = \varnothing), \tag{5}$$

where

$$g(\varepsilon) = \begin{cases} \varepsilon|\log\varepsilon|^{-1/2} & d = 2, \\ \varepsilon & d \geq 3. \end{cases}$$

For $d \geq 3$, an even stronger statement is true : ν_Λ^ε is strongly stochastically dominated by $\mathbb{B}_\Lambda^{c_+(d)\varepsilon}$ and strongly stochastically dominates $\mathbb{B}_\Lambda^{c_-(d)\varepsilon}$, i.e. for any $C \subset \Lambda$:

$$c_-(d)\varepsilon \leq \nu_\Lambda^\varepsilon(x \in \mathcal{A} \mid \mathcal{A}\setminus\{x\} = C) \leq c_+(d)\varepsilon. \tag{6}$$

Concerning the behavior of the interface, it is known that an arbitrarily weak pinning ε is sufficient to localize the interface. Indeed, in [9], Deuschel and Velenik proved, for a class of models including ours, that the infinite volume Gibbs measure $\mu^{\varepsilon,0}$ exists in all $d \geq 1$ and that for any ε small enough and all K large enough [2],

$$-\log\mu^{\varepsilon,0}(\varphi_0 > K) \asymp_d \begin{cases} K & d = 1, \\ K^2/\log K & d = 2, \\ K^2 & d \geq 3. \end{cases} \tag{7}$$

The so-called mass, or rate of exponential decay of the two-point function, associated to the infinite volume Gibbs measure $\mu^{\varepsilon,0}$ is defined, for any $x \in \mathbb{S}^{d-1}$, by

$$m^\varepsilon(x) := -\lim_{k\to\infty} \frac{1}{k} \log \mathbb{C}\mathrm{ov}_\varepsilon(\varphi_0\varphi_{[kx]}).$$

[2] $a \asymp_d b$ means that there exist two constants $0 < c_1 \leq c_2 < \infty$, depending only on d, such that $c_1 b \leq a \leq c_2 b$.

where $[x]$ is the vector of integer parts of x's coordinates. In [17] Ioffe and Velenik showed that for any $d \geq 1$,

$$\inf_{x \in \mathbb{S}^{d-1}} m^{\varepsilon}(x) > 0. \tag{8}$$

The localization of the interface becomes weaker as $\varepsilon \to 0$, we can quantify this by studying the behavior of the variance and the mass of the field in this limit. The most precise results were proved by Bolthausen and Velenik in [5], and can be stated as follows:
For $d = 2$ and ε small enough,

$$\mu^{\varepsilon,0}(\varphi_0^2) = \frac{1}{\pi}|\log \varepsilon| + O(\log|\log \varepsilon|) \tag{9}$$

For ε small enough,

$$m^{\varepsilon} \asymp_d \begin{cases} \sqrt{\varepsilon}|\log \varepsilon|^{-3/4} & d = 2 \\ \sqrt{\varepsilon} & d \geq 3 \end{cases} \tag{10}$$

1.3 Known results about disordered models

1.3.1 Models on $\mathbb{Z}$

In [2], Alexander and Sidoravicius studied the 1-dimensional (attractive/ repulsive) model. They consider a polymer, with monomer locations modeled by the trajectory of a Markov chain $(X_i)_{i \in \mathbb{Z}}$, in the presence of a potential (usually called a "defect line") that interacts with the polymer when it visits 0. Formally, let V_i be an i.i.d. sequence of 0-mean random variables, the model is given by weighting the realisation of the chain with the Boltzmann term

$$\exp\left(\beta \sum_{i=1}^{n} (u + V_i) \mathbb{1}_{[X_i = 0]}\right).$$

Their purpose was to study the localization transition in this model. If a positive fraction of monomers is at 0, we say that the polymer is pinned. In the plane β vs. u, critical lines are defined : for β fixed, let $u_c^{\mathbf{q}}(\beta)$ (resp. $u_c^{\mathbf{a}}(\beta)$) the quenched (resp. annealed) critical value of u above which the polymer is pinned with probability 1 (for the quenched (resp. annealed) measure). They show that the quenched free energy and critical point are nonrandom, calculated the critical point for a deterministic interaction (i.e. $V_i \equiv 0$) and proved that the critical point in the quenched case is strictly smaller.
Note that when the underlying chain is a symmetric simple random walk on $\mathbb{Z}$, the deterministic critical point is known to be 0, so having the quenched critical point ($u_c(\beta)$) strictly negative means that, even when the disorder is repulsive on average, the chain is pinned. This result was obtained by Galluccio and Graber in [11] for a periodic potential, which is frequently used in the physics literature as a "toy model" for random environment.
In [16], Giacomin and Toninelli investigated the order of the localization transition in general models of directed polymers pinned on a defect line.
They prove that for quite a general class of models, as soon as disorder is present, the transition is at least of the second order, i.e. the free energy is differentiable at the critical line and the order parameter (which is the density of pinned sites) vanishes continuously at the transition.
This is particularly interesting as there are examples of non-disordered systems with first order transition (cf. for example [13], Proposition 1.6, for $(1 + d)$-dimensional directed polymers and $d \geq 5$). The result thus implies that the introduction of a disorder may have a smoothening effect on the transition.

For 1-dimensional models, the renewal structure of the return times to 0 plays important role, in particular it simplifies a lot of calculations. In [1], Alexander emphasized this fact by assuming that the tails of the excursion length between consecutive returns of X to 0 are as $n^{-c}\phi(n)$ (for some $1 < c < 2$ and slowly varying ϕ). He analysed the quenched and annealed critical curves in the plane (u, β) for different values of c, showing that for $c > 3/2$ at high temperature the quenched and annealed curves differ significantly only in a very small neighborhood of the critical point, whereas for $c < 3/2$ the quenched and annealed critical points are equal. This was a prediction made by theoretical physicists on the basis of the so-called Harris criterion (see [13], Section 5.5, for more informations). The relevant case in the framework of this paper is the case of the Markov chain given by a simple symetric random walk on $\mathbb{Z}$, which corresponds to $c = 3/2$ and $\phi(n) \sim K$ for $K > 0$, which is borderline. The author is unable to say whether the critical behavior is altered by the disorder in this case, but progress on this question has been made recently by Giacomin, Lacoin and Toninelli in [14] and [15]. They prove that in the borderline case $c = 3/2$, the disorder is relevant in the sense that the quenched critical point in shifted with respect to the annealed one. They consider i.i.d. Gaussian disorder in the first paper and extend the result to more general i.i.d. laws, as well as refine the lower bound on the shift, in the second paper. Note that this case includes pinning of a directed polymer in dimension (1+1) as already mentionned[3], but also the classical models of two-dimensional wetting of a rough substrate, pinning of directed polymers on a defect line in dimension $(3 + 1)$ and pinning of an heteropolymer by a point potential in three-dimensional space.

1.3.2 Models on $\mathbb{Z}^d$, $d \geq 2$

The only result about random pinning models we are aware of is [18]. In this paper, Janvresse, De La Rue and Velenik considered the model (1) in dimension 1 and 2 where $\mathbf{e} \in \{0, \eta\}^{\mathbb{Z}^d}$, which models a interface interacting with an attractive diluted potential. They show that the interface is localized in a sufficiently large but finite box (in the sense that there is a density of pinned sites) if and only if the sites at which the pinning potential is non-zero have positive density. Note that in this paper they characterize the set of realizations of the environment for which pinning holds (the disorder is fixed, not sampled from some given distribution), which is stronger than an almost sure result.

We also mention a series of papers by Külske *et al.* ([19], [22], [20] which study a model with disordered magnetic field (instead of disordered pinning potential). For example, Külske and Orlandi studied the following model in dimension 2

$$\mu_\Lambda^{\varepsilon,0,(\eta)}(d\varphi) = \frac{1}{Z_\Lambda^{\varepsilon,0,(\eta)}} \exp\left(-\frac{1}{4d}\sum_{x\sim y} V(\varphi_x - \varphi_y) + \sum_{x\in\Lambda} \eta_x\varphi_x\right) \prod_{x\in\Lambda} (d\varphi_x + \varepsilon\delta_0(d\varphi_x)) \prod_{y\in\Lambda^c} \delta_0(d\varphi_y), \quad (11)$$

where $(\eta_x)_{x\in\Lambda}$ is an arbitrary fixed configuration of external fields and V is not growing too slowly at infinity. Without disorder ($\eta \equiv 0$), the interface is localized for any $\varepsilon > 0$ [5]. One could expect that in presence of disorder and at least for very large ε the interface is pinned. However, the authors show that this is not the case : the interface diverges regardless of the pinning strength. This implies that an infinite-volume Gibbs measure for this model does not exist. One could hope for the existence of the so-called gradient Gibbs measure (Gibbs distributions of the increments of the interface). In [22] Van Enter and Külske proved that such (infinite volume) measures do not exist in the random field model in dimension 2. Note that gradient Gibbs measures may exist, even when the corresponding Gibbs measure does not. This happens when the interface is locally smooth, although at large scales its fluctuations diverge. This is the case for the two-dimensional Gaussian free field.

[3]We speak about "polymer in dimension $(d + 1)$" when the state space of the Markov chain X is of dimension d.

1.4 Open problems

The model studied in this paper lends itself to number of extensions. We list here a selection, with brief comments.

- **Path-wise description of the interface.** In the positive free energy region, for both of our models, one expects localization, i.e. the finite variance of φ_x and exponential decay of correlations. A much more difficult question concerns the behavior of the interface near the critical line. Does it behave the same as in the homogenous case (i.e. second order transition with the density of pinned sites decreasing linearly for $d \geq 3$, and with a logarithmic correction for $d = 2$) ? Or does the presence of disorder have a smoothening effect on the transition (as it was proven for certain 1-dimensional models) ?
  In the zero free energy region (in the attractive/repulsive setting), we expect the behavior similar to the entropic repulsion for the GFF : in a box of size n the interface should be repelled at height $\pm \log n$ in $d = 2$ and $\pm\sqrt{\log n}$ in $d \geq 3$ (see [8] and references therein for details). The $\pm$ stems from the fact that our model is symmetric with respect to reflection at zero height, hence with probability 1/2 it either goes upwards or downwards.

- **Description for non Gaussian case.** A natural conjecture is that the behavior of the model is the same if we change the Gaussian term $(\varphi_x - \varphi_y)^2$ to any other uniformly convex potential $V(\varphi_x - \varphi_y)$. Let us note that Griffiths inequality is not proven even in the homogenous case, which can make the problem difficult to handle.

- **Non-local interactions.** We restricted our work to the case of the nearest neighbors interactions. We suspect that the results holds true for fast decaying interactions, at least with condition like in [5, (2.1)] (which ensures a control of the random walk's behavior in the random walk representations). As the behavior of the homogenous pinning model beyond this regime is not known, we are unable to pose any further conjectures.

- **Non i.i.d. pinning laws.** Going beyond the i.i.d. case is a very interesting direction. Two natural cases would be the stationary Bernoulli field or the chessboard like configuration. These questions may be closely connected to convexity/concavity properties of the free energy function in the homogenous case. The understanding of this case is still limited. It would be interesting to know if finite range environment laws change the picture, as it is sometimes the case in models with bulk disorder but it seems difficult to answer to this question rigorously.

- **Geometry of pinned sites.** The geometry of the pinned sites is still not fully understood in the homogenous case. For the $d \geq 3$ the law of pinned sites resemble a Bernoulli point process. It is conjectured that once the pinning tends to zero, under suitable re-scaling, this field converges to Poisson point process. For $d = 2$ the situation is not clear at all, though it is expected that the dependency between the points will be preserved in the limit (implying the limit being non-Poissonian).
  Not only these questions propagate to the non-homogenous case but also new ones arise. E.g. for the attracting/repulsive model it would be interesting to study the joint geometry of attractive and repulsive sites.

- **Models with wetting transition.** The effects of introducing a disorder in other models with pinning might be interesting, for example in models exibiting a wetting phenomenon. In the case of the massless Gaussian model in $d = 2$, it is known [6] that the wetting transition takes place at a non-trivial point. A natural question to ask is, if adding disorder shifts the transitions point.

2 Attractive environment

2.1 Introduction

2.1.1 Description of the results

We denote by $\Omega := \mathbb{R}^{\mathbb{Z}^d}$ the set of configurations of the model. Let Λ be a finite subset of $\mathbb{Z}^d$. The finite volume Gibbs measure in Λ for the discrete Gaussian Free Field with random δ-pinning $\mathbf{e}$, with 0 boundary condition, is the probability measure on Ω defined by :

$$\mu_\Lambda^{\mathbf{e},0}(d\varphi) = \frac{1}{Z_\Lambda^{\mathbf{e},0}} e^{-\beta H_\Lambda(\varphi)} \prod_{x\in\Lambda} \left(d\varphi_x + \sqrt{\beta} e_x \delta_0(d\varphi_x) \right) \prod_{y\in\Lambda^c} \delta_0(d\varphi_y), \tag{12}$$

where $\Lambda^c := \mathbb{Z}^d \backslash \Lambda$ and δ_0 is the point mass at 0. The normalization constant $Z_\Lambda^{\mathbf{e}}$ is the so-called partition function of the model. The Hamiltonian in Λ is given by

$$H_\Lambda(\varphi) = \frac{1}{4d} \sum_{\substack{\{x,y\}\cap\Lambda\neq\varnothing \\ x\sim y}} (\varphi_x - \varphi_y)^2, \tag{13}$$

where $x \sim y$ means that x and y are neighbouring vertices in $\mathbb{Z}^d$. The non-homogenous pinning environment is denoted by $\mathbf{e} := (e_x)_{x\in\Lambda}$. We consider the law of $\mathbf{e}$ given by a Bernoulli product measure. Namely, independently on each site $x \in \Lambda$:

$$e_x = \begin{cases} \underline{e} & \text{with probability } p \\ \overline{e} & \text{with probability } (1-p), \end{cases} \tag{14}$$

for some fixed triplet $(\underline{e}, \overline{e}, p)$ such that $0 < \underline{e} < \overline{e}$ and $p \in (0,1)$. In this section we consider only positive values of the environment, meaning purely attractive environment. Let

$$e^\star := p\underline{e} + (1-p)\overline{e}, \quad \text{and} \quad \sigma^2 := p\underline{e}^2 + (1-p)\overline{e}^2 - (e^\star)^2. \tag{15}$$

be the average and variance of e_x, respectively.

Remark 2.1. *1. The measure (12) is the weak limit of the measures $\mu_{\Lambda,a^n}^{\mathbf{e}^n,0}$, given by (2), once we choose $(a^n)_{n\geq 1}, (\mathbf{e}^n)_{n\geq 1}$ such that $a^n \to 0$ and $2a^n(e^{e_x^n} - 1) = e_x$.*

2. The temperature parameter enters only in a trivial way. If we replace the field $(\varphi_x)_{x\in\Lambda}$ by $(\sqrt{\beta}\phi_x)_{x\in\Lambda}$, and $(e_x)_{x\in\Lambda}$ by $(\sqrt{\beta}e_x)_{x\in\Lambda}$ we have transformed the model to temperature parameter $\beta = 1$. Thus in the sequel, we will assume $\beta = 1$.

By definition the annealed model is obtained by averaging the Gibbs weights over the disorder; the annealed model corresponding to environment $\mathbf{e}$ is thus the same as the homogenous model with pinning strength $e^\star$.
We define the quenched (resp. annealed) free energy per site in $\Lambda_n := \{0, ..., n-1\}^d$ by $f_{\Lambda_n}^{\mathbf{q}}(\mathbf{e})$ (resp. $f_{\Lambda_n}^{\mathbf{a}}(\mathbf{e})$) :

$$f_{\Lambda_n}^{\mathbf{q}}(\mathbf{e}) := n^{-d} \log \left(\frac{Z_{\Lambda_n}^{\mathbf{e},0}}{Z_{\Lambda_n}^{0,0}} \right), \quad f_{\Lambda_n}^{\mathbf{a}}(\mathbf{e}) := n^{-d} \log \left(\frac{\mathbb{E}_{\mathbf{e}} Z_{\Lambda_n}^{\mathbf{e},0}}{Z_{\Lambda_n}^{0,0}} \right).$$

where $Z_\Lambda^{0,0}$ denotes the partition function of the model with no pinning (i.e. $\underline{e} = \overline{e} = 0$). This normalization is chosen such that the free energy of the free model is zero.

In the next section we show that $\left(f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e})\right)_n$ converges to a non-random quantity: the infinite volume quenched free energy, denoted by $f^{\mathbf{q}}(\mathbf{e})$, i.e.

$$f^{\mathbf{q}}_{\Lambda}(\mathbf{e}) \to f^{\mathbf{q}}(\mathbf{e}) = \mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}(\mathbf{e})) \qquad \text{a.s. and in } L^2 \text{ as } \Lambda \uparrow \mathbb{Z}^d \text{ provided } |\partial\Lambda|/|\Lambda| \to 0.$$

The same proof can be used to show that $f^{\mathbf{a}}_{\Lambda}(\mathbf{e}) \to f^{\mathbf{a}}(\mathbf{e})$ as $\Lambda \uparrow \mathbb{Z}^d$ provided $|\partial\Lambda|/|\Lambda| \to 0$ (see [16] for the proof in $d = 1$). Comparing the free energy for the quenched and annealed models is useful for studying the effect of the environment on the system. The case when they are different indicates that the disorder of the environment has some macroscopic effect. By the Jensen inequality (applied to log) and L^1 convergence of $f^{\mathbf{q}}_{\Lambda}(\mathbf{e})$, it is easy to present that

$$f^{\mathbf{q}}(\mathbf{e}) \le f^{\mathbf{a}}(\mathbf{e}).$$

We recall that in this section we consider purely attracting environment. Thus we have $f^{\mathbf{q}}(\mathbf{e}) > 0$ for all $(\underline{e}, \overline{e}, p)$ such that $0 < \underline{e} < \overline{e}$ and $p \in (0,1)$, as we prove in Fact 2.5. Moreover, the density of pinned sites in the box Λ_n, is strictly positive as n tends to infinity, as we prove in Fact 2.6. In Theorem 2.8, we show the main result of this section, namely that $f^{\mathbf{q}}(\mathbf{e}) < f^{\mathbf{a}}(\mathbf{e})$ for all $(\underline{e}, \overline{e}, p)$ such that $0 < \underline{e} < \overline{e}$ and $p \in (0,1)$. We provide also some estimates on the difference $f^{\mathbf{a}}(\mathbf{e}) - f^{\mathbf{q}}(\mathbf{e})$ in terms of $\sigma^2 = \mathbb{V}\mathrm{ar}(\mathbf{e})$.

2.1.2 Extension of pinned sites representation to inhomogenous pinning

In this section we extend the decomposition (4) to our inhomogenous model. We write

$$\begin{aligned}
\mu^{\mathbf{e},0}_{\Lambda}(d\varphi) &= \frac{1}{Z^{\mathbf{e},0}_{\Lambda}} e^{-H_\Lambda(\varphi)} \prod_{x\in\Lambda} (d\varphi_x + e_x\delta_0(d\varphi_x)) \prod_{y\in\Lambda^c} \delta_0(d\varphi_y) \\
&= \frac{1}{Z^{\mathbf{e},0}_{\Lambda}} e^{-H_\Lambda(\varphi)} \sum_{A\subset\Lambda} \left(\prod_{x\in A} e_x\delta_0(d\varphi_x) \prod_{y\in\Lambda\setminus A} d\varphi_y \right) \prod_{z\in\Lambda^c} \delta_0(d\varphi_z) \\
&= \sum_{A\subset\Lambda} \underbrace{\left(\prod_{x\in A} e_x\right) \frac{Z^{0,0}_{\Lambda\setminus A}}{Z^{\mathbf{e},0}_{\Lambda}}}_{=:\nu^{\mathbf{e}}_{\Lambda}(A)} \mu^{0,0}_{\Lambda\setminus A}(d\varphi).
\end{aligned} \tag{16}$$

In other words $\nu^{\mathbf{e}}_{\Lambda}$ is the marginal of the measure $\mu^{\mathbf{e},0}_{\Lambda}(d\varphi)$ giving the distribution of pinned points. We can obtain a formula for the ratio of partition functions appearing in the free energy :

$$\frac{Z^{\mathbf{e},0}_{\Lambda}}{Z^{0,0}_{\Lambda}} = \sum_{A\subset\Lambda} \left(\prod_{x\in A} e_x\right) \frac{Z^{0,0}_{\Lambda\setminus A}}{Z^{0,0}_{\Lambda}}. \tag{17}$$

2.2 Existence of quenched free energy

Theorem 2.2. *For $d \ge 2$, let $f^{\mathbf{q}}_{\Lambda}(\mathbf{e}) := |\Lambda|^{-1} \log(Z^{\mathbf{e},0}_{\Lambda}/Z^{0,0}_{\Lambda})$ be the quenched free energy per site in $\Lambda \Subset \mathbb{Z}^d$. Then, the limit*

$$f^{\mathbf{q}}(\mathbf{e}) := \lim_{\Lambda\uparrow\mathbb{Z}^d} f^{\mathbf{q}}_{\Lambda}(\mathbf{e}),$$

exists almost surely and in L^2, and does not depend on the sequence $\Lambda \uparrow \mathbb{Z}^d$ provided it satisfies $|\partial\Lambda|/|\Lambda| \to 0$. Moreover, $f^{\mathbf{q}}(\mathbf{e})$ is non-random, i.e.

$$f^{\mathbf{q}}(\mathbf{e}) = \mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}(\mathbf{e})) \quad a.s.$$

Proof. We prove the existence of the limit along the sequence of boxes $B_n := \Lambda_{2^n-1}$. The generalization to all sequences $\Lambda \uparrow \mathbb{Z}^d$ such that $|\partial\Lambda|/|\Lambda| \to 0$ is rather standard (cf. for example [24]).
Let us write $\alpha_{\mathbf{e},0}(d\varphi) := \prod_{x\in B_n}(d\varphi_x + e_x\delta_0(d\varphi_x))\prod_{y\notin B_n}\delta_0(d\varphi_y)$. We recall also notation (13) for the Hamiltonian. We will cut B_n in 2^d sub-boxes denoted by $B^{(i)}_{n-1}$. Let $X := (\bigcup_{i=1}^{2^d}\partial B^{(i)}_{n-1})\backslash\partial B_n$ be the interface between the sub-boxes. In order to prove the existence of the limit along $(B_n)_n$, we first derive a "decoupling property". Namely there exist $c_n \geq 0$ such that $\sum_n c_n < \infty$ and $|Z^{\mathbf{e},0}_{B_n} - \prod_{i=1}^{2^d} Z^{\mathbf{e}^{(i)},0}_{B_{n-1}}| \leq c_n$ for any realisation of $\mathbf{e}$, where $\mathbf{e}^{(i)}$ is the restriction of $\mathbf{e}$ to the box $B^{(i)}_{n-1}$. This allows us to prove that expectation of $f^{\mathbf{q}}_{B_n}(\mathbf{e})$ converges, and its variance tends to zero, i.e. $f^{\mathbf{q}}_{B_n}(\mathbf{e}) \to c \in \mathbb{R}$ almost surely, in L^1 and in L^2.

Lower bound on $Z^{\mathbf{e},0}_{B_n}$
We have,

$$Z^{\mathbf{e},0}_{B_n} = \int_{\varphi\in\mathbb{R}^{B_n}} e^{-H(\varphi)}\alpha_{\mathbf{e},0}(d\varphi) \geq \mathbf{e}^X \int_{\varphi|_X\equiv 0} e^{-H(\varphi)}\alpha_{\mathbf{e},0}(d\varphi) \geq \underline{e}^{|X|}\prod_{i=1}^{2^d} Z^{\mathbf{e}^{(i)},0}_{B_{n-1}} = \underline{e}^{d2^{n(d-1)}}\prod_{i=1}^{2^d} Z^{\mathbf{e}^{(i)},0}_{B_{n-1}},$$

where $\mathbf{e}^{(i)}$ denotes $\mathbf{e}$ restricted to $B^{(i)}_{n-1}$. Hence,

$$f^{\mathbf{q}}_{B_n}(\mathbf{e}) = 2^{-nd}\log\left(\frac{Z^{\mathbf{e},0}_{B_n}}{Z^{0,0}_{B_n}}\right) \geq \frac{1}{2^d}\sum_{i=1}^{2^d} f^{\mathbf{e}^{(i)}}_{B_{n-1}} + C_{\min}2^{-n}.$$

for some constant $C_{\min} > 0$.

Upper bound on $Z^{\mathbf{e},0}_{B_n}$
We will first prove that with high probability $|\varphi_i| \leq 2^{n\delta}$ for all $i \in X$ and some small $\delta > 0$. This will allow us to “force” φ to be zero on X for small energetic cost.

Lemma 2.3. *There exists $C_1, C_2 > 0$ such that for all $i \in \Lambda_n$ and for all n sufficiently large,*

$$\mu^{\mathbf{e},0}_n(|\varphi_i| > T) \leq C_1 e^{-C_2T^2/\log n}.$$

Proof. Using (16),

$$\mu^{\mathbf{e},0}_n(|\varphi_i| > T) \;=\; \sum_{A\subset\Lambda_n}\nu^{\mathbf{e}}_n(A)\cdot\mu^{0,0}_{\Lambda_n\backslash A}(|\varphi_i| > T)$$

Now, $\mu^{0,0}_{\Lambda_n\backslash A}$ is Gaussian, therefore, $\varphi_i \sim \mathcal{N}(0,\sigma_A)$ with

$$\sigma_A := \mathrm{Var}^{0,0}_{\Lambda_n\backslash A}(\varphi_i) \leq \mathrm{Var}^{0,0}_{\Lambda_n}(\varphi_0) \leq \tilde{C}\log n$$

for some $\tilde{C} > 0$ and n large (cf. [7]). We can use the Gaussian tail estimate :

$$\mu^{0,0}_{\Lambda_n\backslash A}(|\varphi_i| > T) \leq C_1 e^{-C_2T^2/\log n}.$$ ☐

In the sequel, the notation C, C', C'' will be used for positive constants that may change from line to line. Let us define $\tilde{X} := X \cup (\partial X \cap B_n)$, which is a thickening of X consisting of 3 “layers”. Then, Lemma 2.3 allow us to control the height of the field on $\tilde{X} \subset B_n$:

$$\mu^{\mathbf{e},0}_{B_n}(\exists i \in \tilde{X} : |\varphi_i| > 2^{\delta n}) \leq \sum_{i\in\tilde{X}}\mu^{\mathbf{e},0}_{B_n}(|\varphi_i| > 2^{\delta n}) \leq C2^{n(d-1)}e^{-C_2 2^{\delta n}} \leq e^{-C2^{\delta n}} \tag{18}$$

Hence,

$$\begin{aligned}
Z^{\mathbf{e},0}_{B_n} &= (1+C'e^{-C2^{\delta n}})\int_{\varphi|_{\tilde{X}}\in[-2^{\delta n},2^{\delta n}]} e^{-H(\varphi)}\alpha_{\mathbf{e},0}(d\varphi) \\
&\le (1+C'e^{-C2^{\delta n}})C''e^{2^{2\delta n}2^{n(d-1)}}\int_{\varphi|_X\equiv 0}\int_{\varphi|_{\partial X\cap B_n}\in[-2^{\delta n},2^{\delta n}]} e^{-H(\varphi)}\alpha_{\mathbf{e},0}(d\varphi) \\
&\le (1+C'e^{-C2^{\delta n}})C''e^{2^{2\delta n}2^{n(d-1)}}2^{\delta n 2^{n(d-1)}}\int_{\varphi|_{B_n\setminus X}} e^{-H(\varphi|_{B_n\setminus X})}\alpha_{\mathbf{e},0}(d\varphi|_{B_n\setminus X}) \\
&= (1+C'e^{-C2^{\delta n}})C''e^{2^{n(2\delta+d-1)}}2^{\delta n 2^{n(d-1)}}\prod_{i=1}^{2^d} Z^{\mathbf{e}^{(i)},0}_{B_{n-1}}.
\end{aligned}$$

This leads to

$$f^{\mathbf{q}}_{B_n}(\mathbf{e}) \le \frac{1}{2^d}\sum_{i=1}^{2^d} f^{\mathbf{e}^{(i)}}_{B_{n-1}} + C2^{n(\delta-d)} + C'2^{n(2\delta-1)} + C''2^{-n}.$$

Combining our two bounds, we obtain :

$$C_{\min}2^{-n} \le f^{\mathbf{q}}_{B_n}(\mathbf{e}) - \frac{1}{2^d}\sum_{i=1}^{2^d} f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}^{(i)}) \le C_{\max}2^{n(2\delta-1)} \tag{19}$$

for some constants $C_{\min}, C_{\max} > 0$.

Expectation of $f^{\mathbf{q}}_{B_n}(\mathbf{e})$ converges and its variance tends to zero

By (19), it is easy to see that :

$$|\mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}_{B_n}(\mathbf{e})) - \mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}))| \le |\mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}_{B_n}(\mathbf{e})) - \mathbb{E}_{\mathbf{e}}(\frac{1}{2^d}\sum_{i=1}^{2^d} f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}^{(i)}))| \le C2^{n(2\delta-1)} \tag{20}$$

The right hand side is summable for $\delta < 1/2$ hence $\mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}_{B_n}(\mathbf{e}))$ converges as $n \to \infty$. Using independence of environment among the boxes $B^{(i)}_{n-1}$, we can deal with the variance. Let us write

$$f^{\mathbf{q}}_{B_n}(\mathbf{e}) = 2^{-d}\sum_{i=1}^{2^d} f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}^{(i)}) + \mathcal{E}_n$$

where $\mathcal{E}_n$ is the above error term, $|\mathcal{E}_n| \le C'2^{n(2\delta-1)}$. Then,

$$\begin{aligned}
\mathbb{V}\mathrm{ar}_{\mathbf{e}}(f^{\mathbf{q}}_{B_n}(\mathbf{e})) &= \mathbb{V}\mathrm{ar}_{\mathbf{e}}(2^{-d}\sum_{i=1}^{2^d} f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}^{(i)})) + \mathbb{V}\mathrm{ar}_{\mathbf{e}}(\mathcal{E}_n) + 2\mathbb{C}\mathrm{ov}_{\mathbf{e}}(2^{-d}\sum_{i=1}^{2^d} f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}^{(i)}), \mathcal{E}_n) \\
&= 2^{-d}\mathbb{V}\mathrm{ar}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e})) + \mathbb{V}\mathrm{ar}_{\mathbf{e}}(\mathcal{E}_n) + 2\mathbb{C}\mathrm{ov}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}), \mathcal{E}_n) \\
&\le 2^{-d}\mathbb{V}\mathrm{ar}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e})) + \mathbb{V}\mathrm{ar}_{\mathbf{e}}(\mathcal{E}_n) + C'2^{n(2\delta-1)}\mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e}))
\end{aligned}$$

Now, since $(\mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e})))_n$ converges, we get the following upper bound on the variance:

$$\mathbb{V}\mathrm{ar}_{\mathbf{e}}(f^{\mathbf{q}}_{B_n}(\mathbf{e})) \le 2^{-d}\mathbb{V}\mathrm{ar}_{\mathbf{e}}(f^{\mathbf{q}}_{B_{n-1}}(\mathbf{e})) + C2^{2n(2\delta-1)} + C'2^{n(2\delta-1)} \tag{21}$$

We deduce $\mathbb{V}\mathrm{ar}_{\mathbf{e}}(f^{\mathbf{q}}_{B_n}(\mathbf{e})) \to 0$, and conclude $f^{\mathbf{q}}_{B_n}(\mathbf{e}) \to c \in \mathbb{R}$ in L^2. The sequence converges also almost surely, since all error terms are summable. □

Exactly the same proof until (19) yields

Theorem 2.4. *For $d \geq 2$, let $f^{\mathbf{a}}_\Lambda(\mathbf{e}) := |\Lambda|^{-1}\log(\mathbb{E}_{\mathbf{e}} Z^{\mathbf{e},0}_\Lambda / Z^{0,0}_\Lambda)$ be the annealed free energy per site in $\Lambda \Subset \mathbb{Z}^d$. Then, the limit*

$$f^{\mathbf{a}}(\mathbf{e}) := \lim_{\Lambda \uparrow \mathbb{Z}^d} f^{\mathbf{a}}_\Lambda(\mathbf{e})$$

exists and does not depend on the sequence $\Lambda \uparrow \mathbb{Z}^d$ provided it satisfies $|\partial\Lambda|/|\Lambda| \to 0$.

2.3 Bounds on the quenched free energy

2.3.1 Strict positivity of the quenched free energy

In this section we prove that the free energy of the model is strictly positive for all environments $(\underline{e}, \overline{e}, p)$ such that $0 < \underline{e} < \overline{e}$ and $p \in (0,1)$. That implies the localization of the field, namely the strict positivity of the limiting density of pinned sites. Our results are valid for all $d \geq 2$.

Fact 2.5. *For all $(\underline{e}, \overline{e}, p)$ such that $0 < \underline{e} < \overline{e}$ and $p \in (0,1)$, we have $f^{\mathbf{q}}(\mathbf{e}) > 0$.*

Proof. Using (17), and the fact that $0 < \underline{e} < \overline{e}$, we can write

$$f^{\mathbf{q}}(\mathbf{e}) = \lim_{n\to\infty} n^{-d} \log\left(\sum_{A\subset\Lambda_n}\prod_{x\in A} e_x \frac{Z^{0,0}_{\Lambda_n\setminus A}}{Z^{0,0}_{\Lambda_n}}\right) \geq \lim_{n\to\infty} n^{-d}\log\left(\sum_{A\subset\Lambda_n} \underline{e}^{|A|}\frac{Z^{0,0}_{\Lambda_n\setminus A}}{Z^{0,0}_{\Lambda_n}}\right) = f^{\mathbf{a}}(\underline{e})$$

Moreover, it follows from [5] Theorem 2.4, that the asymptotic density of pinned points under $\mu^{\underline{e},0}_{\Lambda_n}$ is stictly positive uniformly in Λ_n for n large enough. This immediately implies by definition of the free energy that $f^{\mathbf{a}}(\underline{e}) > 0$. ☐

Fact 2.6. *Let $0 < \underline{e} < \overline{e}$ and $p \in (0,1)$ then there exists $c > 0$ such that for almost every realization $\mathbf{e}$ we have*

$$\lim_{n\to\infty} \mu^{\mathbf{e},0}_{\Lambda_n}\left(n^{-d}\sum_{x\in\Lambda_n} \mathbb{1}_{[\varphi_x=0]} > c\right) = 1.$$

Remark 2.7. *More is known for $d = 1$. By [2] there exists $\rho = \rho(e^\star, \sigma)$ such that*

$$\lim_{\epsilon\to0}\lim_{n\to\infty} \mu^{\mathbf{e},0}_{\Lambda_n}\left(n^{-d}\sum_{x\in\Lambda_n}\mathbb{1}_{[\varphi_x=0]} \in (\rho-\epsilon, \rho+\epsilon)\right) = 1.$$

Proof. Let $\delta > 0$, and $A_\delta := \{\sum_{x\in\Lambda_n}\mathbb{1}_{[\varphi_x=0]} \leq \delta n^d\}$. Let us fix some $\epsilon > 0$. By definition we have

$$\mu^{\mathbf{e},0}_{\Lambda_n}(A_\delta) = \frac{Z^{\epsilon,0}_{\Lambda_n}}{Z^{0,0}_{\Lambda_n}}\frac{Z^{0,0}_{\Lambda_n}}{Z^{\mathbf{e},0}_{\Lambda_n}}(Z^{\epsilon,0}_{\Lambda_n})^{-1}\int \mathbb{1}_{[A_\delta]} e^{-H(\varphi)}\prod_{x\in\Lambda}(d\varphi + e_x\delta_0(d\varphi_x)).$$

Using the pinned sites decomposition (16) we get :

$$\frac{Z^{\mathbf{e},0}_{\Lambda_n}}{Z^{\epsilon,0}_{\Lambda_n}} = \sum_{A\subset\Lambda_n}\mathbf{e}^A\frac{Z^{0,0}_{\Lambda_n\setminus A}}{Z^{\epsilon,0}_{\Lambda_n}} = \sum_{A\subset\Lambda_n}\frac{\mathbf{e}^A}{\epsilon^{|A|}}\epsilon^{|A|}\frac{Z^{0,0}_{\Lambda_n\setminus A}}{Z^{e^\star,0}_{\Lambda_n}} = \nu^\epsilon_n\left(\boldsymbol{\gamma}^{\mathcal{A}}\right), \tag{22}$$

where $\mathbf{e}^A := \prod_{x\in A} e_x$, $\boldsymbol{\gamma} = (\gamma_x)_{x\in\Lambda_n}$ and $\gamma_x := e_x/\epsilon$. Similarly,

$$(Z^{\epsilon,0}_{\Lambda_n})^{-1}\int \mathbb{1}_{[A_\delta]}e^{-H(\varphi)}\prod_{x\in\Lambda}(d\varphi + e_x\delta_0(d\varphi_x)) = \nu^\epsilon_n\left(\boldsymbol{\gamma}^{\mathcal{A}}; |\mathcal{A}| \leq n^d\delta\right) \leq (\overline{e}/\epsilon)^{n^d\delta}.$$

We thus have

$$n^{-d}\log\mu^{\mathbf{e},0}_{\Lambda_n}(A_\delta)\le\delta\log(\overline{e}/\epsilon)+n^{-d}\log\frac{Z^{\epsilon,0}_{\Lambda_n}}{Z^{0,0}_{\Lambda_n}}-n^{-d}\log\frac{Z^{\mathbf{e},0}_{\Lambda_n}}{Z^{0,0}_{\Lambda_n}}.$$

By Fact 2.5 the last term converges to a strictly positive quantity. Moreover, it is not difficult to show that $Z^{\epsilon,0}_{\Lambda_n}\le Z^{0,0}_{\Lambda_n}(1+C\epsilon)^{n^d}$ for some $C>0$ uniformly in n, hence the second term tends to 0 uniformly in n as $\epsilon\to 0$. Choosing $\epsilon>0$ and $\delta>0$ small enough we get

$$\limsup_{n\to+\infty} n^{-d}\log\mu^{\mathbf{e},0}_{\Lambda_n}(A_\delta)<0.$$

and the result follows. $\square$

2.3.2 Strict inequality between quenched and annealed free energies

Theorem 2.8. *Take $(p,\underline{e},\overline{e})$ such that $p\in(0,1)$ and $0<\underline{e}<\overline{e}$. Let $\mathbf{e}$ be an i.i.d. Bernoulli environment of parameters $(p,\underline{e},\overline{e})$ as described in (14). We recall the notations $e^\star=p\underline{e}+(1-p)\overline{e}$ and $\sigma^2=\underline{e}^2p+\overline{e}^2(1-p)-e^{\star 2}$.*

(a) Let $d\ge 3$. Then,

$$f^{\mathbf{q}}(\mathbf{e})<f^{\mathbf{a}}(\mathbf{e}).$$

Moreover, there exists $c_3(d,p)>0$ such that,

$$f^{\mathbf{a}}(\mathbf{e})-f^{\mathbf{q}}(\mathbf{e})\ge c_3(d,p)\sigma+O(\sigma^2)\quad \text{as}\quad \sigma\to 0.$$

(b) Let $d=2$. There exists $\epsilon>0$ uniform in $p,\underline{e},\overline{e}$ such that if $e^\star<\epsilon$, then

$$f^{\mathbf{q}}(\mathbf{e})<f^{\mathbf{a}}(\mathbf{e}).$$

Moreover, under the same hypothesis on $e^\star$, there exist $c_2(p,e^\star)>0$ such that,

$$f^{\mathbf{a}}(\mathbf{e})-f^{\mathbf{q}}(\mathbf{e})\ge c_2(p,e^\star)\sigma+O(\sigma^2)\quad \text{as}\quad \sigma\to 0.$$

Proof. By Theorem 2.2 we have

$$f^{\mathbf{q}}(\mathbf{e})-f^{\mathbf{a}}(\mathbf{e})=\lim_{n\to\infty}n^{-d}\mathbb{E}_{\mathbf{e}}\log\left(\frac{Z^{\mathbf{e},0}_{\Lambda_n}}{Z^{e^\star}_{\Lambda_n}}\right)$$

By (22) we conclude that our goal is to prove

$$\limsup_{n\to\infty}n^{-d}\mathbb{E}_{\mathbf{e}}\log\left(\nu^{e^\star}_n\left(\gamma^{\mathcal{A}}\right)\right)<0,$$

where $\gamma_x:=e_x/e^\star$.

We will now outline the proof strategy. We notice that if $\nu^{e^\star}_n$ was a Bernoulli product measure with some intensity $\lambda\in(0,1)$ then the theorem would follow easily. Indeed, in such a case, we can calculate directly the expectation under log and the expression in question is equal to

$$n^{-d}\mathbb{E}_{\mathbf{e}}\log\left(\prod_{x\in\Lambda_n}(p\gamma_x+1-p)\right)=\mathbb{E}_{\mathbf{e}}\log(p\gamma_0+1-p)<0.$$

The last inequality follows by the strict concavity of the logarithm and the Jensen inequality. Our aim is to use (6) in order to make this argument rigorous.
Let us now introduce an additional randomization. We will use the uniformly distributed permutations of Λ_n, writing the corresponding expectation by $\tilde{\mathbb{E}}$. It is easy to check that for any i.i.d. pinning law $\mathbb{E}_{\mathbf{e}}\tilde{\mathbb{E}}(\cdot) = \mathbb{E}_{\mathbf{e}}(\cdot)$. Let π be a permutation of the n^d vertices of Λ_n, chosen uniformly at random. By the Jensen inequality, we have

$$\begin{aligned} n^{-d}\mathbb{E}_{\mathbf{e}} \log\left(\nu_n^{e^\star}\left(\gamma^{\mathcal{A}}\right)\right) &= n^{-d}\mathbb{E}_{\mathbf{e}}\tilde{\mathbb{E}} \log \nu_n^{e^\star}\left(\prod_{i\in\mathcal{A}}\gamma_{\pi(i)}\right) \\ &\leq n^{-d}\mathbb{E}_{\mathbf{e}} \log \tilde{\mathbb{E}}\nu_n^{e^\star}\left(\prod_{i\in\mathcal{A}}\gamma_{\pi(i)}\right) = n^{-d}\mathbb{E}_{\mathbf{e}} \log \tilde{\mathbb{E}}\nu_n^{e^\star}\left(\gamma^{\pi(\mathcal{A})}\right), \end{aligned}$$

where $\pi(A) := \{\pi(i) : i \in A\}$. Intuitively, $\tilde{\mathbb{E}}\nu_n^{e^\star}$ is the expectation of the distribution of pinned sites "scattered" by a random permutation. Thanks to this we can work with a uniformly distributed set of pinned points, provided we know its cardinality. More precisely,

$$\tilde{\mathbb{E}}\nu_n^{e^\star}\left(\gamma^{\pi(\mathcal{A})}\right) = \sum_{k=0}^{n^d}\binom{n^d}{k}^{-1}\left(\sum_{A\subset\Lambda_n:|A|=k}\gamma^A\right)\nu_n^{e^\star}\left(|\mathcal{A}|=k\right).$$

Our aim is now to use stochastic majoration/minoration by Bernoulli measures to obtain an upper-bound for $\nu_n^{e^\star}(|\mathcal{A}|=k)$. Difficulties will follow from the fact that $\{|\mathcal{A}|=k\}$ is not an increasing event.

Case $d \geq 3$. We recall the notation and the stochastic dominance results introduced in (6). In $d \geq 3$, we know that $\nu_n^{e^\star}$ stochastically dominates a Bernoulli product measure $\mathbb{B}_n^\lambda$ with $\lambda := c_-(d)e^\star > 0$. For any set $A \subset \Lambda$, we have

$$\mathbb{B}_n^\lambda(\mathcal{A}=A) = \lambda^{|A|}(1-\lambda)^{n^d-|A|}$$

Let $b_{n,k}(\lambda) := \binom{n^d}{k}\lambda^k(1-\lambda)^{n^d-k}$. As the event $\{|\mathcal{A}|\leq k\}$ is decreasing, we have the following upper-bound :

$$\begin{aligned} \nu_n^{e^\star}(|\mathcal{A}|=k) &\leq \nu_n^{e^\star}(|\mathcal{A}|\leq k) \leq \mathbb{B}_n^\lambda(|\mathcal{A}|\leq k) \\ &= \sum_{j=0}^{k} b_{n,\lambda}(j) = b_{n,\lambda}(k)\underbrace{\left(1+\sum_{j=0}^{k-1}\frac{b_{n,\lambda}(j)}{b_{n,\lambda}(k)}\right)}_{(*)} = b_{n,\lambda}(k)\left(1+\sum_{j=0}^{k-1}\prod_{i=j}^{k-1}\frac{b_{n,\lambda}(i)}{b_{n,\lambda}(i+1)}\right) \end{aligned} \tag{23}$$

Now, for $i \leq \lfloor \lambda n^d \rfloor$,

$$\frac{b_{n,\lambda}(i)}{b_{n,\lambda}(i+1)} = \frac{i+1}{n^d-i}\frac{1-\lambda}{\lambda} \leq 1.$$

This gives an upper bound for $k \leq \lfloor \lambda n^d \rfloor$, namely

$$\nu_n^{e^\star}(|\mathcal{A}|=k) \leq c_1 n^d \cdot \mathbb{B}_n^\lambda(|\mathcal{A}|=k).$$

For $k \in [\lceil \lambda n^d \rceil, n^d]$, define $\lambda_k := kn^{-d}$. As by the Stirling formula $b_{n,\lambda_k}(k) \geq c_2 n^{-d/2} > 0$ uniformly in k, we can use the bound:

$$\nu_n^{e^\star}(|\mathcal{A}|=k) \leq c_2 n^{d/2} \cdot \mathbb{B}_n^{\lambda_k}(|\mathcal{A}|=k).$$

This leads to :

$$
\begin{aligned}
\tilde{\mathbb{E}}\nu_n^{e^\star}\left(\gamma^{\pi(\mathcal{A})}\right) &\leq \sum_{k=0}^{n^d}\binom{n^d}{k}^{-1}\left(\sum_{A\subset\Lambda_n:|A|=k}\gamma^A\right)\left[c_1 n^d\cdot\mathbb{B}_n^{\lambda}\left(|\mathcal{A}|=k\right)+\left(\sum_{j=\lceil\lambda n^d\rceil}^{n^d} c_2 n^{d/2}\cdot\mathbb{B}_n^{\lambda_j}\left(|\mathcal{A}|=k\right)\right)\right]\\
&\leq C' n^d\sum_{k=0}^{n^d}\binom{n^d}{k}^{-1}\left(\sum_{A\subset\Lambda_n:|A|=k}\gamma^A\right)\left[\mathbb{B}_n^{\lambda}\left(|\mathcal{A}|=k\right)+\sum_{j=\lceil\lambda n^d\rceil}^{n^d}\mathbb{B}_n^{\lambda_j}\left(|\mathcal{A}|=k\right)\right]\\
&\leq C n^{2d}\max_{\alpha\in[\lambda,1]}\sum_{k=0}^{n^d}\binom{n^d}{k}^{-1}\left(\sum_{A\subset\Lambda_n:|A|=k}\gamma^A\right)\cdot\mathbb{B}_n^{\alpha}(|\mathcal{A}|=k)\\
&\leq C n^{2d}\max_{\alpha\in[\lambda,1]}\mathbb{B}_n^{\alpha}\left(\gamma^{\mathcal{A}}\right)\\
&= C n^{2d}\max_{\alpha\in[\lambda,1]}\prod_{x\in\Lambda_n}\mathbb{B}_n^{\alpha}\left(\gamma_x^{\mathbb{1}[x\in\mathcal{A}]}\right).
\end{aligned}
$$

Hence,

$$
n^{-d}\log\tilde{\mathbb{E}}\nu_n^{e^\star}\left(\gamma^{\pi(\mathcal{A})}\right)\leq\left(n^{-d}\max_{\alpha\in[\lambda,1]}\sum_{x\in\Lambda_n}\log(\alpha\gamma_x+(1-\alpha))\right)+o(1)\qquad\text{as } n\to\infty.
$$

Let us call $\alpha_n^{\max}\in[\lambda,1]\subset(0,1]$ the (random) parameter which maximizes the expression above. Then, taking the expectation with respect to the environment, and the limit $n\to\infty$ we get

$$
\lim_{n\to\infty} n^{-d}\mathbb{E}_{\mathbf{e}}\log\tilde{\mathbb{E}}\nu_n^{e^\star}\left(\gamma^{\pi(\mathcal{A})}\right)\quad\leq\quad\lim_{n\to\infty} n^{-d}\mathbb{E}_{\mathbf{e}}\left[\sum_{x\in\Lambda_n}\log\left(\alpha_n^{\max}(\gamma_x-1)+1\right)\right]\tag{24}
$$

For Bernoulli pinning law, we have to study the function :

$$
g_n(\alpha):=\left(\frac{\underline{N}_n}{n^d}\right)\log(\alpha(\underline{e}/e^\star-1)+1))+\left(\frac{\overline{N}_n}{n^d}\right)\log(\alpha(\overline{e}/e^\star-1)+1),
$$

where $\underline{N}_n=\#\{x\in\Lambda_n: e_x=\underline{e}\}$ and $\overline{N}_n=\#\{x\in\Lambda_n: e_x=\overline{e}\}$. It is easy to check that

$$
\frac{\partial^2 g_n}{\partial\alpha^2}(\alpha)<0\qquad\text{and}\qquad g_n(0)=0.
$$

Moreover, the maximum is attained for $\tilde{\alpha}=\frac{e^\star}{(\overline{e}-\underline{e})^2 p(1-p)}\left(\left(\frac{\overline{N}_n}{n^d}\right)\overline{e}+\left(\frac{\underline{N}_n}{n^d}\right)\underline{e}-e^\star\right)\to 0$ almost surely as $n\to\infty$ by the law of large numbers. Hence there exists almost surely n_0 sufficiently large such that for $n\geq n_0$, $\alpha_n^{\max}=\lambda$. Noticing that for $\alpha\in\mathbb{R}^+$, $g_n(\alpha)\leq\log(\lambda(\overline{e}/e^\star-1)+1)<\infty$ almost surely, uniformly in $n\geq n_0$, we deduce that

$$
\begin{aligned}
g_n(\alpha) &\overset{n\to\infty}{\longrightarrow} p\log(1+\lambda(\underline{e}/e^\star-1)))+(1-p)\log(1+\lambda(\overline{e}/e^\star-1)))\\
&= \mathbb{E}_{\mathbf{e}}\log(\lambda(e_x/e^\star-1)+1)\qquad\text{a.s. and in } L^1.
\end{aligned}
$$

Observe that the r.h.s. is independent of n, so that we can pass to the limit $n\to\infty$ and obtain

$$
\begin{aligned}
(24) &\leq \mathbb{E}_{\mathbf{e}}\log(\lambda(e_x/e^\star-1)+1)\\
&< \log\mathbb{E}_{\mathbf{e}}(\lambda(e_x/e^\star-1)+1)\\
&= \log(1)=0.
\end{aligned}
$$

Let us now do a Taylor expansion of the upperbound (24) around $\sigma = 0$, recalling that $\sigma^2 = \mathbb{V}\mathrm{ar}_{\mathbf{e}}(e_x)$, and that $\lambda = c_-(d)e^\star$. We obtain:

$$\begin{aligned} f^{\mathbf{q}}(\mathbf{e}) - f^{\mathbf{a}}(\mathbf{e}) &\leq p\log(1+\lambda(\underline{e}/e^\star - 1))) + (1-p)\log(1+\lambda(\overline{e}/e^\star - 1))) \\ &= -c_-(d)K(p)\sigma + O(\sigma)^2 \quad \text{ as } \sigma \to 0 \end{aligned}$$

where $K(p) = \sqrt{\frac{p}{1-p}} + p\left(\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}}\right) > 0$ for $p \in [0,1]$. This concludes the proof for $d \geq 3$.

$\square$

Case $d = 2$. The case $d = 2$ requires some slight modification as $\nu_n^{e^\star}$ can be stochastically estimated by Bernoulli product measures only in a weaker sense. Consequently, we cannot use the same argument as in (23).

By [5, Theorem 2.4, (2.13)] there exists $\tilde{c} > 0$ such that for all $e^\star$ sufficiently small, and any set $B \subset \Lambda_n$ we have $\nu_n^{e^\star}(\mathcal{A} \cap B = \varnothing) \leq (1-\lambda)^{|B|}$ where $\lambda = \lambda(e^\star) := \tilde{c}e^\star/\sqrt{|\log(e^\star)|} \in (0,1]$.

Then,

$$\begin{aligned} \nu_n^{e^\star}(|\mathcal{A}| = k) &\leq \nu_n^{e^\star}(|\mathcal{A}| \leq k) = \nu_n^{e^\star}(\exists B \subset \Lambda_n : |B| = n^d - k \text{ and } \mathcal{A} \cap B = \varnothing) \\ &\leq \sum_{B \subset \Lambda_n : |B| = n^d - k} \nu_n^{e^\star}(\mathcal{A} \cap B = \varnothing) \leq \binom{n^d}{k}(1-\lambda)^{n^d-k} \end{aligned}$$

We will show that there exists $\tilde{\lambda} = \tilde{\lambda}(\lambda) > 0$ such that,

$$\binom{n^d}{k}(1-\lambda)^{n^d-k} \leq \mathbb{B}_n^{\tilde{\lambda}}(|\mathcal{A}| = k), \qquad \text{for } k \leq \lfloor \tilde{\lambda} n^d \rfloor$$

Indeed, let $k = \alpha n^d$, we must find $\alpha_0 > 0$ and $\tilde{\lambda}(\alpha)$ such that :

$$1 - \lambda \leq \tilde{\lambda}^{\frac{\alpha}{1-\alpha}}(1-\tilde{\lambda}) =: f_\alpha(\tilde{\lambda}) \tag{25}$$

for all $\alpha \leq \alpha_0$. It is easy to check that $f_\alpha(\tilde{\lambda})$ has a unique maximum in $[0,1]$ at $\tilde{\lambda} = \alpha$. Now, observing that $f_\alpha(\alpha) = \alpha^{\frac{\alpha}{1-\alpha}}(1-\alpha) \to 1$ as $\alpha \to 0$, we deduce that for all $\lambda > 0$ there exists $\alpha_0(\lambda) \in (0,1]$ such that $f_\alpha(\alpha) = \max_{\tilde{\lambda}\in[0,1]} f_\alpha(\tilde{\lambda}) > 1 - \lambda$, so that (25) is satisfied. Note that this is sufficient to prove $f^{\mathbf{q}}(\mathbf{e}) < f^{\mathbf{a}}(\mathbf{e})$ for any environment $\mathbf{e}$ such that $e^\star > 0$: we have $\nu_n^{e^\star}(|\mathcal{A}| = k) \leq \mathbb{B}_n^{\tilde{\lambda}}(|\mathcal{A}| = k)$ with $\tilde{\lambda} > 0$, for $k \leq \lfloor \tilde{\lambda} n^d \rfloor$ and we can proceed analogously as for $d \geq 3$.

But we need to control how α_0 varies as a function of λ to obtain an estimate on the gap $f^{\mathbf{q}}(\mathbf{e}) - f^{\mathbf{a}}(\mathbf{e})$. As the value $\lambda = \tilde{c}e^\star/\sqrt{|\log e^\star|}$ is known only for small values of $e^\star$, we will only study the case $\lambda < 1$, and find an approximation of the solution to the equation $f_\alpha(\alpha) = \alpha^{\alpha/1-\alpha}(1-\alpha) = 1 - \lambda$ around $\alpha = 0$. Developping the left hand side, we get:

$$-\alpha + \alpha\log\alpha + O(\alpha^2) = -\lambda$$

One can check that the behavior of the solution of $-\alpha + \alpha\log\alpha = -\lambda$ close to $\alpha = 0$ is the following :

$$\alpha_0(\lambda) = \frac{\lambda}{|\log\lambda|}\left(1 + O\left(\frac{\log|\log\lambda|}{|\log\lambda|}\right)\right) \qquad \text{as } \lambda \to 0$$

A natural guess for the solution of (25) is then $\alpha^\star = -\frac{C\lambda}{\log\lambda}$. Indeed, $f(\alpha^\star(\lambda))$ is a function of λ which is concave on $[0,1]$ (i.e. $\frac{d^2}{d\lambda^2}f(\alpha^\star(\lambda)) < 0$) and such that $\frac{d}{d\lambda}f(\alpha^\star(\lambda))|_{\lambda=0} = -C$. Moreover, for $C < 1$

the solution to the equation $f(\alpha^\star(\lambda)) = 1 - \lambda$ tends to 1 as C tends to 0. We deduce the following result :
For all $\lambda_0 \in (0,1)$ there exists some $\delta > 0$ sufficiently small such that $f_\alpha(\alpha)$ with $\alpha(\lambda) = -\frac{\delta\lambda}{\log\lambda}$ satisfies (25) for all $\lambda \in [0, \lambda_0]$.
We thus obtain the following upper bound, with $\tilde\lambda = \tilde\lambda(\lambda,\delta) = -\frac{\delta\lambda}{\log\lambda}$ and some fixed $e^\star > 0$ sufficiently small such that $\lambda = \lambda(e^\star) = \tilde c e^\star/\sqrt{|\log(e^\star)|} < 1$:

$$\begin{aligned} f^{\mathbf{q}}(\mathbf{e}) - f^{\mathbf{a}}(\mathbf{e}) &\leq p\log(1+\tilde\lambda(e^\star,\delta,\tilde c)(\underline{e}/e^\star - 1)) + (1-p)\log(1+\tilde\lambda(e^\star,\delta,\tilde c)(\overline{e}/e^\star - 1)) \\ &= \frac{\tilde c \delta K(p)\sigma}{\sqrt{|\log e^\star|}\log\left(\tilde c e^\star/\sqrt{|\log e^\star|}\right)} + O(\sigma^2) \quad \text{as } \sigma \to 0, \end{aligned}$$

with $K(p) = \sqrt{\frac{p}{1-p}} + p\left(\sqrt{\frac{1-p}{p}} - \sqrt{\frac{p}{1-p}}\right)$. Note that the term $\log\left(\tilde c e^\star/\sqrt{|\log e^\star|}\right)$ is negative. This finishes the proof for $d = 2$. □

3 Attractive/Repulsive environment

3.1 Introduction

3.1.1 Description of the results

In this section, we study a model in which the values of the potential can be negative, so the field is penalized if it enters a (small) strip of width $2a$ around 0. As before, we denote by $\Omega := \mathbb{R}^{\mathbb{Z}^d}$ the set of configurations of the model. Let $\Lambda \Subset \mathbb{Z}^d$. The finite volume Gibbs measure in Λ for the discrete Gaussian Free Field with random "strip"-potential, and 0 boundary condition, is the probability measure on Ω defined by :

$$\mu^{\mathbf{e},0}_{\Lambda,a}(d\varphi) = \frac{1}{Z^{\mathbf{e},0}_{\Lambda,a}} \exp\left(-\beta H_\Lambda(\varphi) + \beta\sum_{x\in\Lambda} e_x \mathbb{1}_{[\varphi_x \in [-a,a]]}\right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y). \tag{26}$$

where $a, \beta \in \mathbb{R}^+$, and $H_\Lambda(\varphi)$ is given by (13).The non-homogenous environment is written $\mathbf{e} := (e_x)_{x\in\Lambda}$. We consider again the law of $\mathbf{e}$ given by a Bernoulli product measure. Namely, independently on each site $x \in \Lambda$:

$$e_x = \begin{cases} \underline{e} & \text{with probability } p \\ \overline{e} & \text{with probability } (1-p), \end{cases} \tag{27}$$

for some fixed triplet $(\underline{e}, \overline{e}, p)$. This time $\underline{e}$ and $\overline{e}$ can be negative, i.e. $\underline{e} < \overline{e} \in \mathbb{R}$ and $p \in (0,1)$. We write again $e^\star$ and σ^2 the average and variance of e_x, respectively (c.f. (15)).
The temperature parameter enters again only in a trivial way. If we replace the field $(\varphi_x)_{x\in\Lambda}$ by $(\sqrt{\beta}\phi_x)_{x\in\Lambda}$, a by $\sqrt{\beta}a$, and $(e_x)_{x\in\Lambda}$ by $(\beta e_x)_{x\in\Lambda}$ we have transformed the model to temperature parameter $\beta = 1$. In the sequel, we will therefore work with $\beta = 1$.
We define the quenched (resp. annealed) free energy per site in $\Lambda_n := \{0, ..., n-1\}^d$ by $f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e})$ (resp. $f^{\mathbf{a}}_{\Lambda_n}(\mathbf{e})$) :

$$f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) := n^{-d}\log\left(\frac{Z^{\mathbf{e},0}_{\Lambda_n,a}}{Z^{0,0}_{\Lambda_n}}\right), \quad f^{\mathbf{a}}_{\Lambda_n}(\mathbf{e}) := n^{-d}\log\left(\frac{\mathbb{E}_{\mathbf{e}} Z^{\mathbf{e},0}_{\Lambda_n,a}}{Z^{0,0}_{\Lambda_n}}\right).$$

We first prove in Theorem 3.1 that $(f^{\mathbf{q}}_\Lambda(\mathbf{e}))_\Lambda$ converges to a non-random quantity, denoted by $f^{\mathbf{q}}(\mathbf{e})$, as $\Lambda \uparrow \mathbb{Z}^d$ provided $|\partial\Lambda|/|\Lambda| \to 0$. Again, the same proof shows that $(f^{\mathbf{a}}_\Lambda(\mathbf{e}))_\Lambda$ converges in $\mathbb{R}$ (towards

$f^{\mathbf{a}}(\mathbf{e})$) as $\Lambda \uparrow \mathbb{Z}^d$ provided $|\partial\Lambda|/|\Lambda| \to 0$.
The annealed model is the homogenous one with pinning $\ell(\mathbf{e})$ given by

$$\ell(\mathbf{e}) = \log(p\exp(\underline{e}) + (1-p)\exp(\overline{e})))$$

(which is different from $e^\star$ in the δ-pinning case). In Lemma 3.4 we prove that regardless of $\mathbf{e}$ both the quenched and annealed free energies are non-negative. This motivates the following notions. We introduce the quenched (resp. annealed) critical lines, which are delimiting the region where $f^{\mathbf{q}}(\mathbf{e}) = 0$ (resp. $f^{\mathbf{a}}(\mathbf{e}) = 0$) from the region $f^{\mathbf{q}}(\mathbf{e}) > 0$ (resp. $f^{\mathbf{a}}(\mathbf{e}) > 0$). We are interested in describing the behaviour of these quantities in the phase diagram described by the plane $(\sigma, e^\star) = (\sqrt{\mathbb{V}\mathrm{ar}(\mathbf{e})}, \mathbb{E}(\mathbf{e}))$. We hence introduce the so-called quenched and annealed critical lines:

$$e_c^\star(\sigma) := \sup\{e^\star \in \mathbb{R} : f^{\mathbf{q}}(\mathbf{e}) = f^{\mathbf{q}}(\sigma, e^\star) = 0\} \quad \text{and} \quad e^{\star\mathbf{a}}_c(\sigma) := \sup\{e^\star \in \mathbb{R} : f^{\mathbf{a}}(\mathbf{e}) = f^{\mathbf{a}}(\sigma, e^\star) = 0\}$$

Knowing the behavior of the homogenous model for positive pinning [5], we easily deduce that the annealed critical line is given by the equation

$$\ell(\mathbf{e}) := \log(p\exp(\underline{e}) + (1-p)\exp(\overline{e}))) = 0 \quad \Leftrightarrow \quad e^{\star\mathbf{a}}_c(\sigma) = -\log\left[(1-p)e^{\sigma\sqrt{\frac{p}{1-p}}} + pe^{-\sigma\sqrt{\frac{1-p}{p}}}\right]$$

Note that $f^{\mathbf{q}}(\mathbf{e}) \le f^{\mathbf{a}}(\mathbf{e})$ implies that $e_c^\star(\sigma) \ge e^{\star\mathbf{a}}_c(\sigma)$. We present three results concerning the study of the quenched free energy. First of all, in Theorem 3.8, we prove that $f^{\mathbf{q}}(\mathbf{e}) < f^{\mathbf{a}}(\mathbf{e})$ whenever $\ell(\mathbf{e}) > 0$. Secondly in Theorem 3.6 we prove that if the environment is inside $\{(\underline{e}, \overline{e}, p) : f^{\mathbf{q}}(\mathbf{e}) > 0\}$, then the density of pinned sites in a box of size n is lower bounded by a positive number as $n \to \infty$, while in the interior of $\{(\underline{e}, \overline{e}, p) : f^{\mathbf{q}}(\mathbf{e}) = 0\}$, this density tends to 0. Finally in Theorem 3.9 we show that the quenched critical line lies strictly below the axis $e^\star = 0$ in the neighborhood of $\sigma = 0$ for all $d \ge 2$. Our result shows in particular that there exists a non trivial region where $e^\star < 0$, $\sigma > 0$ and $f^{\mathbf{q}}(\mathbf{e}) > 0$, i.e. where the field is localized though it is repulsed on average by the environment. Unfortunately, we don't have any estimate on the behaviour of $e_c^\star(\sigma) - e^{\star\mathbf{a}}_c(\sigma)$. Moreover, our theorems do not say anything about the existence or absence of smoothing of the transition due to the presence of disorder.

3.1.2 Extension of pinned sites representation for inhomogenous "strip"-potential

Following the same lines as Section 2.1.2 for the attractive model, we can perform a high temperature expansion of the pinning term :

$$\begin{aligned}
\mu^{\mathbf{e},0}_{\Lambda,a}(d\varphi) &= \frac{1}{Z^{\mathbf{e},0}_{\Lambda,a}} \exp\left(-\beta H_\Lambda(\varphi) + \beta\sum_{x\in\Lambda} e_x \mathbb{1}_{[\varphi_x\in[-a,a]]}\right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y) \qquad (28)\\
&= \frac{1}{Z^{\mathbf{e},0}_{\Lambda,a}} \exp\left(-\beta H_\Lambda(\varphi)\right) \prod_{x\in\Lambda}\left((e^{\beta e_x}-1)\mathbb{1}_{[\varphi_x\in[-a,a]]}+1\right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y)\\
&= \sum_{A\subset\Lambda} \underbrace{\left(\prod_{x\in A}\left(e^{\beta e_x}-1\right) \frac{Z^{0,0}_\Lambda(\varphi_x\in[-a,a]\,,\forall x\in A)}{Z^{\mathbf{e},0}_{\Lambda,a}}\right)}_{=:\nu^{\mathbf{e}}_{\Lambda,a}(A)} \mu^{0,0}_\Lambda(d\varphi \mid \varphi_x\in[-a,a]\,,\forall x\in A)
\end{aligned}$$

Observe that the weights can be negative if $e_x < 0$. For this reason we will only use the measure $\nu^\varepsilon_{\Lambda,a}$ corresponding to a homogenous pinning potential of strength $\varepsilon > 0$.
For the homogenous case with $\varepsilon > 0$, it is known [5] that $\nu^\varepsilon_{\Lambda,a}$ is strong FKG in the sense of [10]. Moreover, it can be compared to Bernoulli product measures in the same fashion as in the δ-pinning

case, namely (5) is still true in this setting. These will be very useful futher. However, let us stress that unlike δ-pinning case this measure is not a marginal of $\mu^{\mathbf{e},0}_{\Lambda,a}$ giving the "pinned sites". We will define it in a next step.
For some reasons that will be explained later, we will need another representation

$$\begin{aligned}
\mu^{\mathbf{e},0}_{\Lambda,a}(d\varphi) &= \frac{1}{Z^{\mathbf{e},0}_{\Lambda,a}} \exp\left(-\beta H_\Lambda(\varphi) + \beta \sum_{x\in\Lambda} e_x \mathbb{1}_{[\varphi_x\in[-a,a]]}\right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y) \qquad (29)\\
&= \frac{1}{Z^{\mathbf{e},0}_{\Lambda,a}} \exp\left(-\beta H_\Lambda(\varphi)\right) \prod_{x\in\Lambda} \left((e^{\beta e_x} \mathbb{1}_{[\varphi_x\in[-a,a]]})^A (\mathbb{1}_{[\varphi_x\notin[-a,a]]})^{A^c}\right) \prod_{x\in\Lambda} d\varphi_x \prod_{y\in\Lambda^c} \delta_0(d\varphi_y)\\
&= \sum_{A\subset\Lambda} \underbrace{\left(\left(\prod_{x\in A} e^{\beta e_x}\right) \frac{Z^{0,0}_\Lambda(A=\{x:|\varphi_x|\le a\})}{Z^{\mathbf{e},0}_{\Lambda,a}}\right)}_{=:\tilde\nu^{\mathbf{e}}_{\Lambda,a}(A)} \mu^{0,0}_\Lambda(d\varphi \mid A=\{x:|\varphi_x|\le a\})
\end{aligned}$$

In particular, we obtain a formula for the following ratio of the partition functions :

$$\frac{Z^{\mathbf{e},0}_{\Lambda,a}}{Z^{\ell(\mathbf{e}),0}_{\Lambda,a}} = \sum_{A\subset\Lambda} \prod_{x\in A} e^{\beta(e_x-\ell(\mathbf{e}))} \tilde\nu^{\ell(\mathbf{e})}_{\Lambda,a}(A) = \tilde\nu^{\ell(\mathbf{e})}_{\Lambda,a}\left(\tilde\gamma^{\mathcal{A}}\right).$$

where $\ell(\mathbf{e})$ stands for a homogenous environment for which $e_x=\ell(\mathbf{e})$ for all $x\in\Lambda$, and $\tilde\gamma=(\tilde\gamma_x)_{x\in\Lambda}$, $\tilde\gamma_x := e^{\beta(e_x-\ell(\mathbf{e}))}$. Observe that, contrary to (28), we do not obtain a mixture of Gaussian measures in (29).

3.2 Existence of the quenched free energy

Theorem 3.1. *For $d\ge 2$, let $f^{\mathbf{q}}_\Lambda(\mathbf{e}) := |\Lambda|^{-1}\log(Z^{\mathbf{e},0}_{\Lambda,a}/Z^{0,0}_\Lambda)$ be the quenched free energy per site in $\Lambda \Subset \mathbb{Z}^d$. Then, the limit*

$$f^{\mathbf{q}}(\mathbf{e}) := \lim_{\Lambda\uparrow\mathbb{Z}^d} f^{\mathbf{q}}_\Lambda(\mathbf{e})$$

exists almost surely and in L^2, and does not depend on the sequence $\Lambda\uparrow\mathbb{Z}^d$ provided $|\partial\Lambda|/|\Lambda|\to 0$. Moreover, $f^{\mathbf{q}}(\mathbf{e})$ is non-random, i.e.

$$f^{\mathbf{q}}(\mathbf{e}) = \mathbb{E}_{\mathbf{e}}(f^{\mathbf{q}}(\mathbf{e})) \quad a.s.$$

Proof. The idea of the proof is the same as for Theorem 2.2, so we only write the part that must be adapted, namely the upper-bound on the height of the field under $\mu^{\mathbf{e},0}_{n,a}$. We keep the same notations as previously.

Lemma 3.2. *There exist $C_1, C_2, C_3 > 0$ such that for all $x\in\Lambda_n$ and for all n sufficiently large,*

$$\mu^{\mathbf{e},0}_{\Lambda_n,a}(|\varphi_x|>T) \le C_1 e^{-C_2(T-C_3\log n)^2/\log n}.$$

Proof. We first use (29) to get

$$\mu^{\mathbf{e},0}_{\Lambda_n,a}(|\varphi_x|>T) = \sum_{A\subset\Lambda} \tilde\nu^{\mathbf{e}}_{n,a}(A)\cdot \mu^{0,0}_{\Lambda_n}(|\varphi_x|>T \mid A=\{x:|\varphi_x|\le a\}).$$

As it was mentioned before $\mu^{0,0}_{\Lambda_n}(d\varphi \mid A=\{x:|\varphi_x|\le a\})$ is not Gaussian making the proof considerably harder. We will use the results listed in survey [12]. We are going to show that

$$\mu^{0,0}_{\Lambda_n}(d\varphi \mid A=\{x:|\varphi_x|\le a\}) \prec \mu^{0,0}_{\Lambda_n}(d\varphi \,|\, \forall x\in\Lambda_n,\ \varphi_x\ge a),$$

where $\prec$ denotes the stochastic domination. To this end we define, for $k \geq 0$:

$$\begin{aligned} U_1^k(\varphi_x) &:= \begin{cases} k(\varphi_x + a)^4 \mathbb{1}_{[\varphi_x \in [-a,0]]} + k(\varphi_x - a)^4 \mathbb{1}_{[\varphi_x \in [0,a]]} & \text{for } x \in A, \\ k(\varphi_x - a)^4 \mathbb{1}_{[\varphi_x > a]} + k(\varphi_x + a)^4 \mathbb{1}_{[\varphi_x < -a]} & \text{for } x \in A^c. \end{cases} \\ U_2^k(\varphi_x) &:= k\varphi_x^4 \mathbb{1}_{[\varphi_x \leq a]}, \end{aligned}$$

Further we define also measures $\mu_i^k(\varphi) \propto \exp(-H_n(\varphi) - \sum_{x\in\Lambda_n} U_i^k(\varphi_x))$ for $i \in \{1,2\}$. One checks that condition [12, (B.3)] is fulfilled hence $\mu_1^k \prec \mu_2^k$. Passing with $k \to +\infty$ we get the aforementioned domination and as a consequence

$$\mu_{\Lambda_n}^{0,0}(\varphi_x \geq T |\, A = \{y : |\varphi_y| \leq a\}) \leq \mu_{\Lambda_n}^{0,0}(\varphi_x \geq T \,|\, \forall y \in \Lambda_n,\ \varphi_y \geq a).$$

For $d \geq 3$ we use [12, Theorem 3.1], which in our notation implies

$$A_n := \mu_{\Lambda_n}^{0,0}(\varphi_x \,|\, \forall y \in \Lambda_n,\ \varphi_y \geq a) \leq C\sqrt{\log n},$$

for some $C > 0$. By the Brascamp–Lieb inequality [12, Section B.2] we obtain

$$\begin{aligned} \mu_{\Lambda_n}^{0,0}(\varphi_x \geq T \,|\, \forall y \in \Lambda_n,\ \varphi_y \geq a) &\leq \mu_{\Lambda_n}^{0,0}(\varphi_x - A_n \geq T - A_n \,|\, \forall y \in \Lambda_n,\ \varphi_y \geq a) \\ &\leq \exp(-C_1(T - A_n)^2/\log n) \end{aligned}$$

for another constant $C_1 > 0$.
The case $d = 2$ is slightly more involved. Our first aim is to prove that $A_n \leq C \log n$, for some constant $C > 0$. Let us assume on contrary that $A_n > C \log n$. By [12, (B.14)] we have $\mu_{\Lambda_n}^{0,0}(\varphi_i^2) \leq \log n + A_n^2$. Using the Paley-Zygmund inequality we get

$$\mu_{\Lambda_n}^{0,0}(\varphi_i \geq A_n/2) \geq \frac{1}{4} \cdot \frac{A_n^2}{\log n + A_n^2} \geq 1/8,$$

for n large enough. Now choosing C large enough we get a contradiction of [3, Theorem 4]. The rest of the reasoning follows now the same line as before.

□

Again, exactly the same proof suffices to show the following

Theorem 3.3. *For $d \geq 2$, let $f_\Lambda^{\mathbf{a}}(\mathbf{e}) := |\Lambda|^{-1} \log(\mathbb{E}_{\mathbf{e}} Z_{\Lambda,a}^{\mathbf{e},0}/Z_\Lambda^{0,0})$ be the annealed free energy per site in $\Lambda \Subset \mathbb{Z}^d$. Then, the limit*

$$f^{\mathbf{a}}(\mathbf{e}) := \lim_{\Lambda \uparrow \mathbb{Z}^d} f_\Lambda^{\mathbf{a}}(\mathbf{e})$$

exists in $\mathbb{R}$ and does not depend on the sequence $\Lambda \uparrow \mathbb{Z}^d$ provided $|\partial\Lambda|/|\Lambda| \to 0$.

□

3.3 Bounds on the quenched free energy

3.3.1 Positivity of the quenched free energy

Fact 3.4. *For $d \geq 2$,*

$$f^{\mathbf{q}}(\mathbf{e}) := \lim_{n\to\infty} n^{-d} \log\left(\frac{Z_{\Lambda_n,a}^{\mathbf{e},0}}{Z_n^{0,0}}\right) \geq 0$$

Remark 3.5. *This proves also that $f^{\mathbf{a}}(\mathbf{e}) \geq 0$ since by the Jensen inequality $f^{\mathbf{q}}(\mathbf{e}) \leq f^{\mathbf{a}}(\mathbf{e})$.*

Proof. For $0 \leq \underline{e} < \overline{e}$ (and any $d \geq 1$) the proof is easy, since by the Jensen inequality

$$f^{\mathbf{a}}(\mathbf{e}) = \lim_{n\to\infty} n^{-d} \log \mu_{\Lambda_n}^{0,0}\left(\exp\left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[|\varphi_x|\leq a]}\right)\right) \geq \lim_{n\to\infty} n^{-d} \mu_{\Lambda_n}^{0,0}\left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[|\varphi_x|\leq a]}\right) \geq 0$$

The non-trivial part of the claim concerns $\underline{e} < 0$ (no restriction on $\overline{e}$ which is by convention bigger than $\underline{e}$). We will prove that $f^{\mathbf{a}}(\mathbf{e}) \geq \underline{e}\eta$ for all $\eta > 0$, with different strategies for $d = 2$ and $d \geq 3$.

Proof for $d = 2$. In this case,

$$\mu_{\Lambda_n}^{0,0}\left(\sharp\{x : |\varphi_x| \leq a\} > \eta n^d\right) \leq \frac{\mu_{\Lambda_n}^{0,0}\left(\sharp\{x : |\varphi_x| \leq a\}\right)}{\eta n^d} = \frac{1}{\eta n^d}\sum_{x\in\Lambda_n} \mu_{\Lambda_n}^{0,0}(|\varphi_x| \leq a) \leq \frac{2a}{\eta n^d \sqrt{2\pi\sigma_n^2}}.$$

The last inequality comes from the fact that φ_x is a Gaussian variable under $\mu_{\Lambda_n}^{0,0}$ with variance $\sigma_n^2 = \mu_{\Lambda_n}^{0,0}(\varphi_x^2) \geq C \log n \to \infty$ as $n \to \infty$. The right hand side is then less than $1/2$ for n sufficiently large, hence

$$\mu_{\Lambda_n}^{0,0}\left(\exp\left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[|\varphi_x|\leq a]}\right)\right) \geq \mu_{\Lambda_n}^{0,0}\left(\exp\left(\underline{e}\sum_{x\in\Lambda_n} \mathbb{1}_{[|\varphi_x|\leq a]}\right)\right) \geq e^{\underline{e}\eta n^d}/2,$$

from which we deduce $f^{\mathbf{a}}(\mathbf{e}) \geq \underline{e}\eta$. Finally we set $\eta \to 0$.

Proof for $d \geq 3$. In dimension 3 and above, the variance of the free field is bounded, so the previous argument does not work. But for a cost of surface order we can shift the boundary condition to a constant $M > 0$, and from this height the probability of having a density of sites where $|\varphi_x| \leq a$ which is bigger than η is less than a half for M sufficiently large. Indeed, recalling that we write $\mu_{\Lambda_n}^{0,M}$ for the measure with boundary condition M, for some $C > 0$ we have

$$\mu_{\Lambda_n}^{0,0}\left(\exp\left(\underline{e}\sum_{x\in\Lambda_n} \mathbb{1}_{[|\varphi_x|\leq a]}\right)\right) \geq Ce^{M^2 n^{d-1}} \mu_{\Lambda_{n-1}}^{0,M}\left(\exp\left(\underline{e}\sum_{x\in\Lambda_n} \mathbb{1}_{[|\varphi_x|\leq a]}\right)\right).$$

Now,

$$\begin{aligned}
\mu_{\Lambda_{n-1}}^{0,M}\left(\sharp\{x : |\varphi_x| \leq a\} > \eta(n-1)^d\right) &\leq \frac{\mu_{\Lambda_{n-1}}^{0,M}\left(\sharp\{x : |\varphi_x| \leq a\}\right)}{\eta(n-1)^d} = \frac{1}{\eta(n-1)^d}\sum_{x\in\Lambda_n} \mu_{\Lambda_{n-1}}^{0,M}(|\varphi_x| \leq a) \\
&= \frac{1}{\eta(n-1)^d}\sum_{x\in\Lambda_{\Lambda_{n-1}}} \mu_{\Lambda_{n-1}}^{0,0}(\varphi_x > M - a) \\
&\leq \frac{1}{(M-a)^2\eta(n-1)^d}\sum_{x\in\Lambda_{n-1}} \mu_{\Lambda_{n-1}}^{0,0}(\varphi_x^2) \\
&\leq \frac{C(d)}{\eta(M-a)^2} < 1/2 \quad \text{for } M \text{ sufficiently large.}
\end{aligned}$$

Hence,

$$\mu_{\Lambda_n}^{0,0}\left(\exp\left(\underline{e}\sum_{x\in\Lambda_n} \mathbb{1}_{[|\varphi_x|\leq a]}\right)\right) \geq Ce^{M^2 n^{d-1}} e^{\underline{e}\eta n^d}/2,$$

which implies as before that $f^{\mathbf{a}}(\mathbf{e}) \geq \underline{e}\eta$ for all $\eta > 0$. □

3.3.2 Density of pinned points

Theorem 3.6. *Let* $\mathbf{e}$ *be the environment as described in (27), the following holds :*

1. *If, additionally,* $\mathbf{e}$ *is such that* $f^{\mathbf{q}}(\mathbf{e}) > 0$*. Then there exists some* $c > 0$ *such that for almost every realisation of* $\mathbf{e}$ *we have*

$$\lim_{n\to\infty} \mu^{\mathbf{e},0}_{\Lambda_n,a}\left(n^{-d}\sum_{i\in\Lambda_n}\mathbb{1}_{[\varphi_i\in[-a,a]]} > c\right) = 1.$$

2. *Let* $\mathcal{Z} := \{\mathbf{e}\in\{\underline{e},\overline{e}\}^{\mathbb{Z}^d} : f^{\mathbf{q}}(\mathbf{e}) = 0\}$*. If* $\mathbf{e}\in\overset{\circ}{\mathcal{Z}}$*, then for any* $c > 0$ *and for almost every realisation of* $\mathbf{e}$ *we have*

$$\lim_{n\to\infty} \mu^{\mathbf{e},0}_{\Lambda_n,a}\left(n^{-d}\sum_{i\in\Lambda_n}\mathbb{1}_{[\varphi_i\in[-a,a]]} > c\right) = 0. \tag{30}$$

where $\overset{\circ}{\mathcal{Z}}$ *denotes the interior of the set* $\mathcal{Z}$*.*

Proof. The proof of the first claim goes similar lines of the proof of Fact 2.6 and hence is skipped. For the second claim note that $f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e})$ is a function of $\underline{e}$ and $\overline{e}$. Let us write

$$\bar{\rho}_n := n^{-d}\sum_{x\in\Lambda_n}\mathbb{1}_{[e_x=\overline{e}]}\mathbb{1}_{[\varphi_x\in[-a,a]]}.$$

It is easy to prove that

$$\frac{\partial f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e})}{\partial\overline{e}} = \mu^{\mathbf{e},0}_{\Lambda_n,a}(\bar{\rho}_n), \quad \text{and} \quad \frac{\partial^2 f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e})}{\partial\overline{e}^2} = \mathbb{V}\mathrm{ar}^{\mathbf{e},0}_{\Lambda_n,a}(\bar{\rho}_n) \geq 0.$$

Hence $(f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}))_n$ is a sequence of convex functions in the variable $\overline{e}$. We know by Theorem 3.1 that it converges to $f^{\mathbf{q}}(\mathbf{e})$. Therefore, for $\mathbf{e}\in\overset{\circ}{\mathcal{Z}}$ we have

$$\lim_{n\to\infty}\mu^{\mathbf{e},0}_{\Lambda_n,a}(\bar{\rho}_n) = \lim_{n\to\infty}\frac{\partial f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e})}{\partial\overline{e}} = \frac{\partial f^{\mathbf{q}}(\mathbf{e})}{\partial\overline{e}} = 0.$$

The same reasoning holds for $\underline{\rho} := n^{-d}\sum_{x\in\Lambda_n}\mathbb{1}_{[e_x=\underline{e}]}\mathbb{1}_{[\varphi_x\in[-a,a]]}$. The second claim is now a straightforward consequence of the Chebychev inequality. □

Remark 3.7. *The question of the behaviour on the boundary of the* 0*-phase is intimately connected with the question of the degree of the phase transition which stays unresolved for the moment.*

3.3.3 Strict inequality between quenched and annealed free energies

Theorem 3.8. *Take* $(p,\underline{e},\overline{e})$ *such that* $p\in(0,1)$ *and* $\underline{e}<\overline{e}$ *(no restriction on the sign of* $\underline{e}$*). Let* $\mathbf{e}$ *be an i.i.d. Bernoulli environment of parameters* $(p,\underline{e},\overline{e})$ *as described in (27). We recall the notations* $e^\star = p\underline{e}+(1-p)\overline{e}$ *and* $\sigma^2 = \underline{e}^2p+\overline{e}^2(1-p)-e^{\star 2}$*.*

(a) *Let* $d\geq 3$*. Then,*

$$f^{\mathbf{q}}(\mathbf{e}) < f^{\mathbf{a}}(\mathbf{e}).$$

Moreover, there exists $c_3(d,a,e^\star) > 0$ *such that,*

$$f^{\mathbf{a}}(\mathbf{e}) - f^{\mathbf{q}}(\mathbf{e}) \geq c_3(d,a,e^\star)\sigma^2 + O(\sigma^3) \quad \text{as} \quad \sigma\to 0.$$

(b) Let $d = 2$. There exists $\epsilon > 0$ uniform in $p, \underline{e}, \overline{e}$ such that if $e^{\ell(\mathbf{e})} - 1 < \epsilon$, then

$$f^{\mathbf{q}}(\mathbf{e}) < f^{\mathbf{a}}(\mathbf{e}).$$

Moreover, under the same hypothesis on $\ell(\mathbf{e})$, there exist $c_2(a, e^\star) > 0$ such that,

$$f^{\mathbf{a}}(\mathbf{e}) - f^{\mathbf{q}}(\mathbf{e}) \geq c_2(a, e^\star)\sigma^2 + O(\sigma^3) \quad \textit{as} \quad \sigma \to 0.$$

Proof. The same idea as the proof for the attractive δ-pinning case can be As in Section 2.3.2, we have

$$f^{\mathbf{q}}(\mathbf{e}) - f^{\mathbf{a}}(\mathbf{e}) = \lim_{n\to\infty} n^{-d}\mathbb{E}_{\mathbf{e}} \log\left(\frac{Z^{\mathbf{e},0}_{\Lambda_n,a}}{Z^{\ell(\mathbf{e}),0}_{\Lambda_n,a}}\right) = \lim_{n\to\infty} n^{-d}\mathbb{E}_{\mathbf{e}} \log(\tilde{\nu}^{\ell(\mathbf{e})}_{n,a}(\tilde{\gamma}^{\mathcal{A}}))$$

with $\ell(\mathbf{e}) = \log(p\exp(\underline{e}) + (1-p)\exp(\overline{e})))$ is the annealed environment corresponding to $\mathbf{e}$, and $\tilde{\gamma}_x = e^{e_x - \ell(\mathbf{e})}$.
The following simple estimation is crucial for our proof. We are able to upperbound $\tilde{\nu}^{\ell(\mathbf{e})}_n(|\mathcal{A}| = k)$ using $\nu^{\ell(\mathbf{e})}_n(|\mathcal{A}| = k)$ and consequently, using results of [5], by a Bernoulli measure

$$\begin{aligned}
\tilde{\nu}^{\ell(\mathbf{e})}_{n,a}(|\mathcal{A}| \leq k) &= \mu^{\ell(\mathbf{e}),0}_{\Lambda_n,a}(|\mathcal{A}| \leq k) = \sum_{A\subset\Lambda_n} \nu^{\ell(\mathbf{e})}_{n,a}(A)\mu^{0,0}_{\Lambda_n}(|\mathcal{A}| \leq k||\phi_x| \leq a, \forall x \in A) \\
&= \sum_{A\subset\Lambda_n, |A|\leq k} \nu^{\ell(\mathbf{e})}_{n,a}(A)\mu^{0,0}_{\Lambda_n}(|\mathcal{A}| \leq k||\phi_x| \leq a, \forall x \in A) \leq \sum_{A\subset\Lambda_n, |A|\leq k} \nu^{\ell(\mathbf{e})}_{n,a}(A) \\
&= \nu^{\ell(\mathbf{e})}_{n,a}(|\mathcal{A}| \leq k) \leq \mathbb{B}^{\lambda}_n(|\mathcal{A}| \leq k)
\end{aligned}$$

with $\lambda := c_-(a)(e^{\ell(\mathbf{e})} - 1)$. Let us notice that the above much resembles (23). The proof follows further unchanged up to (24) where we have to analyse the function

$$g_n(\alpha) := \left(\frac{\underline{N}_n}{n^d}\right)\log\left(\alpha\cdot e^{\underline{e}-\ell(\mathbf{e})} + (1-\alpha)\right) + \left(\frac{\overline{N}_n}{n^d}\right)\log\left(\alpha\cdot e^{\overline{e}-\ell(\mathbf{e})} + (1-\alpha)\right).$$

We can easily check that $\partial^2_\alpha g_n(\alpha) < 0$ and $g_n(0) = 0$. Moreover, the maximum is attained for $\tilde{\alpha} = \frac{(e^{\overline{e}}(1-p)+e^{\underline{e}}p)}{(e^{\overline{e}}-e^{\underline{e}})(1-p)p}(-\frac{\underline{N}_n}{n^d}(1-p) + \frac{\overline{N}_n}{n^d}p) \to 0$ almost surely as $n \to \infty$ by the law of large numbers. The same reasoning holds then, and we end up with

$$g_n(\alpha) \overset{n\to\infty}{\to} \mathbb{E}_{\mathbf{e}}\log(\lambda(e^{e_x-\ell(\mathbf{e})} - 1) + 1) \quad \text{a.s. and in } L^1.$$

leading to the upper bound on the difference of free energies

$$\begin{aligned}
f^{\mathbf{q}}(\mathbf{e}) - f^{\mathbf{a}}(\mathbf{e}) &\leq \mathbb{E}_{\mathbf{e}}\log(\lambda(e^{e_x-\ell(\mathbf{e})} - 1) + 1) \\
&< \log\mathbb{E}_{\mathbf{e}}(\lambda(e^{e_x-\ell(\mathbf{e})} - 1) + 1) \\
&= \log(1) = 0
\end{aligned} \tag{31}$$

using the strict convexity of the logarithm and the definition of $\ell(\mathbf{e})$. The Taylor expansion of the upperbound (31) around $\sigma = 0$ gives the following estimate. (We recall that $\sigma^2 = \mathbb{V}\mathrm{ar}_{\mathbf{e}}(e_x)$, and that $\lambda = c_-(a)(e^{\ell}(\mathbf{e}) - 1)$.)

$$f^{\mathbf{a}}(\mathbf{e}) - f^{\mathbf{q}}(\mathbf{e}) \geq \left(\frac{1}{2}\left(e^{e^\star} - 1\right)^2 c_-(a)^2\right)\sigma^2 + O(\sigma^3) \quad \text{as } \sigma \to 0$$

This concludes the proof for $d \geq 3$.
The generalization for $d = 2$ follows the same lines, with $\lambda = \tilde{c}_-(a)(e^{\ell(\mathbf{e})} - 1)/\sqrt{|\log(e^{\ell(\mathbf{e})} - 1)|}$. We obtain :

$$f^{\mathbf{a}}(\mathbf{e}) - f^{\mathbf{q}}(\mathbf{e}) \quad \geq \quad \left(\frac{\tilde{c}_-(a)^2 \left(e^{e^\star} - 1\right)^2}{2|\log\left(e^{e^\star} - 1\right)|} \right) \sigma^2 + O(\sigma^3) \qquad \text{as } \sigma \to 0$$

This concludes the proof for $d = 2$. $\square$

3.4 Attraction by a repulsive on average environment

Theorem 3.9. *Take $(p, \underline{e}, \overline{e})$ such that $p \in (0, 1)$ and $\underline{e} < 0 < \overline{e}$. Let $\mathbf{e}$ be an i.i.d. Bernoulli environment of parameters $(p, \underline{e}, \overline{e})$ as described in (27). We recall the notations $e^\star = p\underline{e} + (1 - p)\overline{e}$ and $\sigma^2 = \underline{e}^2 p + \overline{e}^2(1 - p) - e^{\star 2}$.*

For $d \geq 2$, the quenched critical line is located in the quadrant $\{(\sigma, e^\star) : \sigma \geq 0, e^\star < 0\}$.

More precisely, for $d \geq 3$, there exists some $C, C' > 0$ depending on d, a, p only and $\epsilon \in (0, 1)$ such that for any environment $\mathbf{e}$ which fulfills $\overline{e} > 0$, $-\epsilon < \underline{e} < 0$ and

$$\begin{cases} e^\star > \frac{C' \underline{e}^2}{\log(-\underline{e})} & \text{for } d = 2 \\ e^\star > -C \cdot \underline{e}^2 & \text{for } d \geq 3, \end{cases}$$

we have $f^{\mathbf{q}}(\mathbf{e}) > 0$.

Remark 3.10.

1. *Both of this bounds can be seen on Figures (2) in the plane $(\sigma, e^\star)$. Moreover, the bound for $d \geq 3$ can be rewritten as $e^\star > -C''(d, a, p) \cdot \sigma^2$.*
2. *Jensen's inequality gives us an upper bound on C, C'. Indeed, as $f^{\mathbf{a}}(\mathbf{e}) \geq f^{\mathbf{q}}(\mathbf{e})$, if $f^{\mathbf{a}}(\mathbf{e}) = 0$ then $f^{\mathbf{q}}(\mathbf{e}) = 0$. In particular, we must have $-C \leq \frac{\partial^2}{\partial \underline{e}^2} e^\star_c|_{\underline{e}=0} < 0$. Our result gives thus an upper-bound on the behavior of the quenched critical line near $\sigma = 0$.*

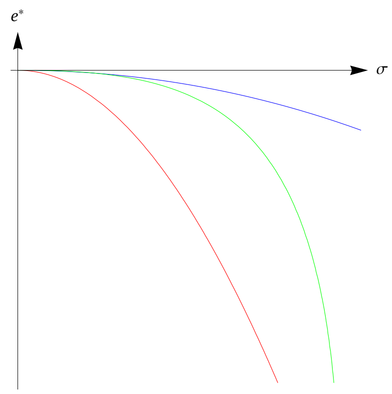

Figure 2: Phase diagram of the model. The red curve is the annealed critical line; the two other curves are the bounds we get on the position of the quenched critical line for $d \geq 3$ (blue curve) and $d = 2$ (green curve). They are both valid for small σ. (Note that the green curve could be above the blue one, depending on the constants C and C', which we do not control.)

3.4.1 Proof for $d \geq 3$

Proof. The idea is to tilt the measure such that the field φ is shifted up of an amount h on the sites x for which $e_x < 0$. In this way the shift of the field follows the environment. For some technical reasons, we need to work with the measure with boundary condition a, so perform two changes of measure (first one changing boundary condition and the second one to following the environment). Let $h > 0$ (to be fixed later).

$$f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) = n^{-d} \log \mu^{0,a}_{\Lambda_n,\mathbf{e},h}\left(\frac{\mathrm{d}\mu^{0,a}_{\Lambda_n}}{\mathrm{d}\mu^{0,a}_{\Lambda_n,\mathbf{e},h}} \frac{\mathrm{d}\mu^{0,0}_{\Lambda_n}}{\mathrm{d}\mu^{0,a}_{\Lambda_n}} \exp\left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[\varphi_x \in [-a,a]]} \right)\right),$$

where $(\varphi_x)_{x\in\Lambda_n}$ under $\mu^{0,a}_{\Lambda_n,\mathbf{e},h}$ is distributed as $\left(\varphi_x + h\mathbb{1}_{[e_x<0]}\right)_{x\in\Lambda_n}$ under $\mu^{0,a}_{\Lambda_n}$. More formally, introducing $T_{\mathbf{e},h} : ((\varphi_x)_{x\in\Lambda_n}) \mapsto \left(\varphi_x + h\mathbb{1}_{[e_x<0]}\right)_{x\in\Lambda_n}$, we define $\mu^{0,a}_{\Lambda_n,\mathbf{e},h}$ as $\mu^{0,a}_{\Lambda_n} \circ T^{-1}_{\mathbf{e},h}$.
Using Jensen's inequality, we get

$$\begin{aligned} f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) &= n^{-d} \log \left[\mu^{0,a}_{\Lambda_n,\mathbf{e},h} \exp\left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[\varphi_x\in[-a,a]]} + \log \frac{\mathrm{d}\mu^{0,a}_{\Lambda_n}}{\mathrm{d}\mu^{0,a}_{\Lambda_n,\mathbf{e},h}} + \log \frac{\mathrm{d}\mu^{0,0}_{\Lambda_n}}{\mathrm{d}\mu^{0,a}_{\Lambda_n}} \right)\right] \\ &\geq n^{-d} \mu^{0,a}_{\Lambda_n,\mathbf{e},h} \left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[\varphi_x\in[-a,a]]} + \underbrace{\log \frac{\mathrm{d}\mu^{0,a}_{\Lambda_n}}{\mathrm{d}\mu^{0,a}_{\Lambda_n,\mathbf{e},h}}}_{(1)} + \underbrace{\log \frac{\mathrm{d}\mu^{0,0}_{\Lambda_n}}{\mathrm{d}\mu^{0,a}_{\Lambda_n}}}_{(2)} \right) \end{aligned}$$

As $Z^{0,0}_{\Lambda_n,\mathbf{e},h} = Z^{0,0}_{\Lambda_n}$ (which follows by change of variables in the Gaussian integral), the first term can be written as

$$(1) = -\frac{1}{4d} \sum_{\substack{\{x,y\}\cap\Lambda_n \neq \varnothing \\ x\sim y}} (\varphi_x - \varphi_y)^2 - (\hat{\varphi}_x - \hat{\varphi}_y)^2$$

where $\hat{\varphi}_x := \varphi_x + h\mathbb{1}_{[e_x<0]}$. Hence, using the definition of $\mu^{0,a}_{\Lambda_n,\mathbf{e},h}$,

$$\begin{aligned} n^{-d}\mu^{0,a}_{\Lambda_n,\mathbf{e},h}((1)) &= -\frac{n^{-d}}{4d} \mu^{0,a}_{\Lambda_n} \left(\sum_{\substack{\{x,y\}\cap\Lambda_n \neq \varnothing \\ x\sim y}} (\hat{\varphi}_x - \hat{\varphi}_y)^2 - (\varphi_x - \varphi_y)^2 \right) \\ &= -\frac{h^2 n^{-d}}{4d} \sum_{\substack{\{x,y\}\cap\Lambda_n \neq \varnothing \\ x\sim y}} \left(\mathbb{1}_{[e_x<0]} - \mathbb{1}_{[e_y<0]}\right)^2 \end{aligned}$$

The second term contains only boundary contribution of order n^{d-1}. Indeed,

$$(2) = \left(2a \sum_{x\in\partial\Lambda_n} \varphi_x - a^2|\partial\Lambda_n| \right) + \log\left(\frac{Z^{0,a}_{\Lambda_n}}{Z^{0,0}_{\Lambda_n}} \right) \geq \left(2a \sum_{x\in\partial\Lambda_n} \varphi_x - a^2|\partial\Lambda_n| \right) - Cn^{d-1}$$

where we used the same methods as in Section 2.2 to estimate the ratio $Z^{0,a}_{\Lambda_n}/Z^{0,0}_{\Lambda_n}$. Hence,

$$n^{-d}\mu^{0,a}_{\Lambda_n,\mathbf{e},h}((2)) \geq 2a \cdot n^{-d} \sum_{x\in\partial\Lambda_n} \mu^{0,a}_{\Lambda_n}(\hat{\varphi}_x) - Cn^{-1} \geq h \sum_{x\in\partial\Lambda_n} \mathbb{1}_{[e_x<0]} - Cn^{-1} \geq -Cn^{-1}$$

We get

$$f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) \quad \geq \quad n^{-d} \sum_{x \in \Lambda_n} e_x \mu^{0,a}_{\Lambda_n}(\hat{\varphi}_x \in [-a,a]) - \frac{h^2 n^{-d}}{4d} \sum_{\substack{\{x,y\} \cap \Lambda_n \neq \varnothing \\ x \sim y}} (\mathbb{1}_{[e_x<0]} - \mathbb{1}_{[e_y<0]})^2 - Cn^{-1}$$

Now we use the fact that the marginal laws of all φ_x, $x \in \Lambda_n$ under $\mu^{0,a}_{\Lambda_n}$ are Gaussian variables centered at a, i.e. $\varphi_x \sim \mathcal{N}(a, \sigma^x_n)$ where $\sigma^x_n = \mathrm{Var}^{0,a}_{\Lambda_n}(\varphi_x) \leq \mathrm{Var}^{0,a}_{\infty}(\varphi_x) \leq c(d) < \infty$ for $d \geq 3$. Therefore,

$$\begin{aligned} \mu^{0,a}_{\Lambda_n}(\varphi_x \in [-a,a]) - \mu^{0,a}_{\Lambda_n}(\varphi_x + h \in [-a,a]) &= \mu^{0,0}_{\Lambda_n}(\varphi_x \in [-2a,0]) - \mu^{0,0}_{\Lambda_n}(\varphi_x \in [-2a-h,-h]) \\ &= C \left(\int_{-h}^{0} - \int_{-2a-h}^{-2a} \right) e^{-y^2/2\sigma^{x\,2}_n} dy \\ &\asymp h \quad \text{as } n \to \infty, \end{aligned} \tag{32}$$

for $c(d) >> h$. In particular we will use that :

$$\mu^{0,a}_{\Lambda_n}(\varphi_x \in [-a,a]) - \mu^{0,a}_{\Lambda_n}(\varphi_x + h \in [-a,a]) \quad \geq \quad C_1(d,a) \cdot h,$$

for some $C_1(d,a) > 0$.

$$\begin{aligned} f^{\mathbf{q}}_n(\mathbf{e}) \geq n^{-d} \sum_{x \in \Lambda_n} e_x & \left(\mu^{0,a}_{\Lambda_n}(\varphi_x \in [-a,a]) - C_1(d,a) h \mathbb{1}_{[e_x<0]} \right) \\ & - \frac{h^2 n^{-d}}{4d} \sum_{\substack{\{x,y\} \cap \Lambda_n \neq \varnothing \\ x \sim y}} (\mathbb{1}_{[e_x<0]} - \mathbb{1}_{[e_y<0]})^2 - Cn^{-1} \end{aligned}$$

Observe that $\mu^{0,a}_{\Lambda_n}(\varphi_x \in [-a,a]) = \mu^{0,0}_{\Lambda_n}(\varphi_x \in [-2a,0]) \geq \mu^{0,0}_{\infty}(\varphi_x \in [-2a,0]) \geq C_2(d,a)$ for some $C_2(d,a) > 0$. By taking the expectation with respect to the environment, using the bounded convergence theorem and Theorem 3.1 we get :

$$f^{\mathbf{q}}(\mathbf{e}) = \lim_{n \to \infty} \mathbb{E}_{\mathbf{e}} f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) \quad \geq \quad e^\star C_2(d,a) - hC_1(d,a)p\underline{e} - h^2 p(1-p)/4 \tag{33}$$

We may optimize over h as the left hand side does not depend on it. Doing this one checks that $f^{\mathbf{q}}(\mathbf{e}) > 0$ as soon as

$$e^\star > -\frac{C_1(d,a)}{C_2(d,a)} \cdot \frac{p}{1-p} \cdot \underline{e}^2 =: -K(d,a,p) \cdot \underline{e}^2$$

This gives the implicit equation in terms of the variance σ^2 of e_x :

$$e^\star > -\frac{1}{2K} + \sigma \sqrt{\frac{1-p}{p}} + \frac{1}{2} \sqrt{\frac{1}{K^2} - \frac{4\sigma}{Kp^2} \sqrt{p(1-p)}} = -\frac{K(1-p)\sigma^2}{p} + O(\sigma^3)$$

The annealed critical curve as well as this bound are drawn on Figure 2. We recall that (32) is valid under assumption that h is small. The maximum of (33) is realized at $h_{\max} = -\frac{2C_1}{1-p} \cdot \underline{e}$, thus it is enough to assume that $\underline{e}$ is small.

□

3.4.2 Proof for $d = 2$

In the case $d = 2$, the variance of the Gaussian free field diverges with the size of the box, so we cannot use the previous estimates. To circumvent this problem we introduce the so-called massive free field. Let $m > 0$,

$$\mu^{0,s}_{\Lambda_n,m}(d\varphi) = \frac{1}{Z^{0,s}_{\Lambda_n,m}} \exp\left(-\frac{1}{8d} \sum_{\substack{\{x,y\}\cap\Lambda_n\neq\varnothing \\ x\sim y}} (\varphi_x - \varphi_y)^2 - m^2 \sum_{x\in\Lambda_n} (\varphi_x - s)^2 \right) \prod_{x\in\Lambda_n} d\varphi_x \prod_{x\in\partial\Lambda_n} \delta_0(d\varphi_x), \tag{34}$$

Known facts about this model can be found in [7, Section 3.3]. In particular, the random walk representation introducted for the massless GFF is still true, but for a random walk Y_t that is killed with rate $\xi(m) = \frac{m^2}{1+m^2}$, namely at each time ℓ, if the walker has not already been killed, the process is killed with probability $\xi(m)$, where the killing is independent of the walk.

Lemma 3.11. *Let $d = 2$. Then,*

1. *There exists some $C_1 > 0$ such that for n large enough, $m > 0$ small enough and all $x \in \Lambda_n$,*

$$\mu^{0,0}_{\Lambda_n,m}(\varphi_x^2) \leq C_1 |\log(m)|.$$

2. *There exists some $C_2 > 0$ such that for n large enough and $m > 0$ small enough, we have*

$$n^{-2} \log \frac{Z^{0,0}_{\Lambda_n,m}}{Z^{0,0}_{\Lambda_n,0}} \geq -C_2 m^2 |\log(m)|.$$

Proof. These bounds are rather standard. We give here the main steps of the proofs with some references. For the first claim, we use the random walk representation [7] to write

$$\mu^{0,0}_{\Lambda_n,m}(\varphi_x^2) = \sum_{\ell=0}^{\infty} P_x(Y_\ell = x\,,\, \tau_{\Lambda_n} \wedge \aleph > \ell) = \sum_{\ell=0}^{\infty} (1 - \xi(m))^\ell P_x(Y_\ell = x\,,\, \tau_{\Lambda_n} > \ell)$$

where τ_{Λ_n} is the first exit time of Λ_n and $\aleph$ is the killing time of the random walk Y_t. Hence,

$$\mu^{0,0}_{\Lambda_n,m}(\varphi_x^2) \leq \mu^{0,0}_{\Lambda_n,m}(\varphi_0^2) \leq \sum_{\ell=1}^{\infty} (1 - \xi(m))^\ell P_0(X_\ell = 0) \tag{35}$$

where X_ℓ is a simple random walk (without killing). The projections of X_ℓ onto the two coordinate axis are two independant 1−dimensional random walks X^1_ℓ and X^2_ℓ, then by Stirling formula,

$$P_0(X_{2\ell} = 0) = (P_0(X^1_{2\ell} = 0))^2 = \left(\binom{2\ell}{\ell} 2^{-2\ell} \right)^2 = \frac{1}{\pi\ell}(1 + o(1)) \quad \text{as } \ell \to \infty$$

The asymptotics of (35) for small m gives the desired upperbound.
To prove the second claim we use the representation of the partition function described in [4, p.542] (it applies to the massive GFF with an obvious modification). We denote by $\tilde{P}$ the coupling of a random

walk X_n and a killed random walk Y_n such that $Y_n = X_n$ up to its killing time $\aleph$,

$$
\begin{aligned}
|\Lambda_n|^{-1} \log \frac{Z^{0,0}_{\Lambda_n,0}}{Z^{0,0}_{\Lambda_n,m}} &= |\Lambda_n|^{-1} \left(\frac{1}{2} \sum_{x\in\Lambda_n} \sum_{\ell=1}^{\infty} \frac{1}{2\ell} \left(\tilde{P}_x(X_{2\ell} = x, \tau_{\Lambda_n} > 2\ell) - \tilde{P}_x(Y_{2\ell} = x, \tau_{\Lambda_n} \wedge \aleph > 2\ell) \right) \right) \\
&\leq \frac{1}{2} \sum_{\ell=1}^{\infty} \frac{1}{2\ell} \left(\tilde{P}_0(X_{2\ell} = 0, \tau_{\Lambda_n} > 2\ell) - \tilde{P}_0(Y_{2\ell} = 0, \tau_{\Lambda_n} \wedge \aleph > 2\ell) \right) \\
&= \frac{1}{2} \sum_{\ell=1}^{\infty} \frac{1}{2\ell} \tilde{P}_0(X_{2\ell} = 0, \tau_{\Lambda_n} > 2\ell, \aleph \leq 2\ell) \\
&\leq \frac{1}{2} \sum_{\ell=1}^{\infty} \frac{1}{2\ell} \tilde{P}_0(X_{2\ell} = 0) \left(1 - (1-\xi(m))^{2\ell} \right) \qquad (36)
\end{aligned}
$$

Using the same estimate as in (35), the asymptotics of (36) for small m gives the desired upperbound. □

The idea is to tilt the measure, as in the proof for $d \geq 3$, first to work with the massive measure, and second to follow the environment such that the field φ is shifted up of an amount h on the sites x for which $e_x < 0$. For some technical reason, we need to work with the measure with boundary condition a, so we perform three changes of measure (first one for changing boundary condition, a second one for adding mass, and a third one for following the environment).
Let $h > 0$ and $m > 0$ to be fixed later.

$$
\begin{aligned}
f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) &= n^{-2} \log \mu^{0,0}_{\Lambda_n} \left(\exp \sum_{x\in\Lambda_n} e_x \mathbb{1}_{[\phi_x \in [-a,a]]} \right) \\
&= n^{-2} \log \mu^{0,s}_{\Lambda_n,m,\mathbf{e},h} \left(\exp \left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[\phi_x\in[-a,a]]} + \log \frac{\mathrm{d}\mu^{0,0}_{\Lambda_n}}{\mathrm{d}\mu^{0,s}_{\Lambda_n}} + \log \frac{\mathrm{d}\mu^{0,s}_{\Lambda_n}}{\mathrm{d}\mu^{0,s}_{\Lambda_n,m}} + \log \frac{\mathrm{d}\mu^{0,s}_{\Lambda_n,m}}{\mathrm{d}\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}} \right) \right)
\end{aligned}
$$

where $(\varphi_x)_{x\in\Lambda_n}$ under $\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}$ is distributed as $\left(\varphi_x + h\mathbb{1}_{[e_x<0]}\right)_{x\in\Lambda_n}$ under $\mu^{0,s}_{\Lambda_n,m}$; more formally, introducing $T_{\mathbf{e},h} : ((\varphi_x)_{x\in\Lambda_n}) \mapsto \left(\varphi_x + h\mathbb{1}_{[e_x<0]}\right)_{x\in\Lambda_n}$, we define $\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}$ as $\mu^{0,s}_{\Lambda_n,m} \circ T^{-1}_{\mathbf{e},h}$. Using Jensen's inequality we get

$$
f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) \geq n^{-2} \mu^{0,s}_{\Lambda_n,m,\mathbf{e},h} \left(\sum_{x\in\Lambda_n} e_x \mathbb{1}_{[\phi_x\in[-a,a]]} + \underbrace{\log \frac{\mathrm{d}\mu^{0,0}_{\Lambda_n}}{\mathrm{d}\mu^{0,s}_{\Lambda_n}}}_{(1)} + \underbrace{\log \frac{\mathrm{d}\mu^{0,s}_{\Lambda_n}}{\mathrm{d}\mu^{0,s}_{\Lambda_n,m}}}_{(2)} + \underbrace{\log \frac{\mathrm{d}\mu^{0,s}_{\Lambda_n,m}}{\mathrm{d}\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}}}_{(3)} \right)
$$

As in the proof for $d \geq 3$, we have

$$
n^{-2} \mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}((1)) \geq -Cn^{-1}.
$$

By Lemma 3.11, we have $\frac{Z^{0,s}_{\Lambda_n,m}}{Z^{0,s}_{\Lambda_n,0}} = \frac{Z^{0,s}_{\Lambda_n,m}}{Z^{0,0}_{\Lambda_n,m}} \frac{Z^{0,0}_{\Lambda_n,m}}{Z^{0,0}_{\Lambda_n,0}} \frac{Z^{0,0}_{\Lambda_n,0}}{Z^{0,s}_{\Lambda_n,0}} \geq -Cn - C_2 n^2 m^2 |\log m|$, and then

$$
(2) = \log \left(\frac{Z^{0,s}_{\Lambda_n,m}}{Z^{0,s}_{\Lambda_n}} \right) + m^2 \sum_{x\in\Lambda_n} \varphi_x^2 \geq -Cn - C_2 n^2 m^2 |\log m| + m^2 \sum_{x\in\Lambda_n} \varphi_x^2
$$

hence,

$$n^{-2}\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}((2)) \geq -Cn^{-1} - C_2 m^2|\log m|.$$

Finally, noticing that $Z^{0,s}_{\Lambda_n,m,\mathbf{e},h} = Z^{0,s}_{\Lambda_n,m}$ (just perform a change of variables in the Gaussian integral), we can compute the third term.

$$(3) = -\frac{1}{8}\sum_{\substack{\{x,y\}\cap\Lambda_n\neq\varnothing\\ x\sim y}} (\varphi_x-\varphi_y)^2 - (\hat\varphi_x-\hat\varphi_y)^2 - m^2\sum_{x\in\Lambda_n}(\varphi_x - s)^2 - (\hat\varphi_x - s)^2$$

where $\hat\varphi_x := \varphi_x + h\mathbb{1}_{[e_x<0]}$. Now we will use the fact that the marginal laws of all φ_x, $x\in\Lambda_n$ under $\mu^{0,s}_{\Lambda_n,m}$ are Gaussian variables, i.e. $\varphi_x \sim \mathcal{N}(\mu^x_n, \sigma^{x2}_n)$ where $\mu^x_n \approx s$ except for x close to the boundary of the box. Indeed, by the random walk representation of the mean, there is $C>0$ such that $|\mu^{0,s}_{\Lambda_n,m}(\varphi_x) - s| \leq C(1+m^2)^{-\|x-\partial\Lambda_n\|}$. Moreover, $(\sigma^x_n)^2 = \mathrm{Var}^{0,s}_{\Lambda_n,m}(\varphi_x) \leq C_1|\log m|$. Using the definition of $\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}$, and computing the terms as in the proof for $d\geq 3$,

$$n^{-2}\mu^{0,s}_{\Lambda_n,m,\mathbf{e},h}((3)) \quad\geq\quad -\frac{h^2n^{-2}}{8}\sum_{\substack{\{x,y\}\cap\Lambda_n\neq\varnothing\\ x\sim y}}(\mathbb{1}_{[e_x<0]}-\mathbb{1}_{[e_y<0]})^2 - n^{-2}m^2h^2\sum_{x\in\Lambda_n}\mathbb{1}_{[e_x<0]} + Cn^{-1}.$$

We get, for n large enough and m small enough

$$\begin{aligned} f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) \quad\geq\quad & n^{-d}\sum_{x\in\Lambda_n} e_x\mu^{0,s}_{\Lambda_n,m}(\hat\varphi_x\in[-a,a]) - \frac{h^2n^{-2}}{8}\sum_{\substack{\{x,y\}\cap\Lambda_n\neq\varnothing\\ x\sim y}}(\mathbb{1}_{[e_x<0]}-\mathbb{1}_{[e_y<0]})^2\\ & -n^{-2}m^2h^2\sum_{x\in\Lambda_n}\mathbb{1}_{[e_x<0]} - C'm^2|\log m| - Cn^{-1}, \end{aligned}$$

for some C and $C'>0$. Note that for $O(n^2)$ sites x, we have

$$\mu^{0,s}_{\Lambda_n,m}(\varphi_x\in[-a,a]) - \mu^{0,s}_{\Lambda_n,m}(\varphi_x + h\in[-a,a]) \quad\asymp\quad \Phi'_{s,\sigma^2}(a)\cdot h \qquad \text{as } n\to\infty, \tag{37}$$

for $h << a \leq s = C_1|\log m|$, and $\sigma = C_1|\log m|$. Above $\Phi_{s,\sigma}$ stands for the p.d.f. of the above Gaussian distribution with mean s and variance σ^2. In particular, for a positive fraction of x (close to 1) and m sufficiently small, we have the upper bound :

$$\mu^{0,s}_{\Lambda_n,m}(\varphi_x\in[-a,a]) - \mu^{0,s}_{\Lambda_n,m}(\varphi_x + h\in[-a,a]) \quad\geq\quad \frac{C_1(a)}{|\log m|}\cdot h,$$

for some $C_1(a)>0$. Now we can compute :

$$\begin{aligned} f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) \quad\geq\quad & n^{-2}\sum_{x\in\Lambda_n} e_x(\mu^{0,s}_{\Lambda_n,m}(\varphi_x\in[-a,a]) - \frac{C_1(a)}{|\log m|}h\mathbb{1}_{[e_x<0]})\\ & -\frac{h^2n^{-2}}{8}\sum_{\substack{\{x,y\}\cap\Lambda_n\neq\varnothing\\ x\sim y}}(\mathbb{1}_{[e_x<0]}-\mathbb{1}_{[e_y<0]})^2 - n^{-2}m^2h^2\sum_{x\in\Lambda_n}\mathbb{1}_{[e_x<0]} - C'm^2|\log m| - Cn^{-1} \end{aligned}$$

Observe that $\mu^{0,s}_{\Lambda_n,m}(\varphi_x\in[-a,a]) \geq 2a\cdot\Phi'_{s,\sigma^2}(-a) = \frac{\tilde C_1(a)}{|\log m|}$ uniformly in $x\in\Lambda_n$. Let us take the expectation with respect to the environment, and use the bounded convergence theorem and Theorem 3.1, we get :

$$f^{\mathbf{q}}(\mathbf{e}) = \lim_{n\to\infty}\mathbb{E}_{\mathbf{e}} f^{\mathbf{q}}_{\Lambda_n}(\mathbf{e}) \quad\geq\quad e^\star\frac{\tilde C_1(a)}{|\log m|} - h\cdot\frac{C_1(a)p\underline{e}}{|\log m|} - h^2m^2p - h^2\cdot\frac{p(1-p)}{4} - C'm^2|\log m|. \tag{38}$$

Our aim now is to show that the right hand side can be positive even when $e^\star$ is negative. In the above expression h, m are free parameters which we may vary. However, we have to remember that both h and m need to be small enough, which makes standard optimisation analysis cumbersome. We are going to show that there exists $C > 0$ and $\underline{e}_0 < 0$ such that for any $\underline{e} \in (\underline{e}_0, 0)$ and

$$e^\star := C\frac{\underline{e}^2}{\log(-\underline{e})},$$

there exist small h and m such that the rhs of (38) is positive. Notice that the result will imply that for any $e^\star \geq C\frac{\underline{e}^2}{\log(-\underline{e})}$ the free energy is positive. Let us choose the value of h which maximizes (38) for fixed m, i.e.

$$h = -\frac{C_1(a)\underline{e}}{(2m^2p + (1-p)/2)|\log(m)|}. \tag{39}$$

and for m let us take

$$m^2 = -k/(\log k)^3, \text{ where } k := -e^\star \tilde{C}_1(a)/C'. \tag{40}$$

One can verify that with the above choice of parameters both h and m are as small as we want. Let us first put (39) into the rhs of (38) and obtain

$$f^{\mathbf{q}}(\mathbf{e}) \geq e^\star \frac{\tilde{C}_1(a)}{|\log m|} + \frac{C_1(a)^2 p\underline{e}^2}{(4m^2p + (1-p))(\log m)^2} - C'm^2|\log m|.$$

For k and consequently m small enough we have

$$f^{\mathbf{q}}(\mathbf{e}) \geq e^\star \frac{\tilde{C}_1(a)}{|\log m|} + \frac{C_1(a)^2 p\underline{e}^2}{2(1-p)(\log m)^2} - C'm^2|\log m|.$$

Further let us multiply both sides by $(\log m)^2$ and insert (40).

$$\begin{aligned} f^{\mathbf{q}}(\mathbf{e})(\log m)^2 &\geq -e^\star\tilde{C}_1(a)\log m + \frac{C_1(a)^2p\underline{e}^2}{2(1-p)} + C'm^2(\log m)^3 \\ &= -\frac{e^\star\tilde{C}_1(a)}{2}(\log k - 3\log(|\log k|)) + \frac{C_1(a)^2p\underline{e}^2}{2(1-p)} - C'\frac{k}{8(\log k)^3}(\log k - 3\log(|\log k|))^3 \\ &= -e^\star\frac{\tilde{C}_1(a)}{2}\log(-e^\star\tilde{C}_1(a)/C') + \frac{C_1(a)^2p\underline{e}^2}{2(1-p)} + e^\star\frac{\tilde{C}_1(a)}{8} + o(e^\star) \quad \text{as} \quad |e^\star| \to 0 \end{aligned}$$

From the last claim it is straightforward to conclude existence of C (sufficiently small) and $\underline{e}_0$ with the properties described above. □

Acknowledgements. We would like to warmly acknowledge Yvan Velenik for introducing us to the topic, and for many usefull discussions. L.C. thanks Fabio Toninelli for his advice.
L.C. was partially supported by the Swiss National Foundation. P.M. was supported by a Sciex Fellowship grant no. 10.044.

References

[1] K. S. Alexander. The effect of disorder on polymer depinning transitions. *Comm. Math. Phys.*, 279(1):117–146, 2008.

[2] K. S. Alexander and V. Sidoravicius. Pinning of polymers and interfaces by random potentials. *Ann. Appl. Probab.*, 16(2):636–669, 2006.

[3] E. Bolthausen, J.-D. Deuschel, and G. Giacomin. Entropic repulsion and the maximum of the two-dimensional harmonic crystal. *Ann. Probab.*, 29(4):1670–1692, 2001.

[4] E. Bolthausen and D. Ioffe. Harmonic crystal on the wall: a microscopic approach. *Comm. Math. Phys.*, 187(3):523–566, 1997.

[5] E. Bolthausen and Y. Velenik. Critical behavior of the massless free field at the depinning transition. *Comm. Math. Phys.*, 223(1):161–203, 2001.

[6] P. Caputo and Y. Velenik. A note on wetting transition for gradient fields. *Stochastic Process. Appl.*, 87(1):107–113, 2000.

[7] A. Dembo and T. Funaki. *Lectures on probability theory and statistics*, volume 1869 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2005. Lectures from the 33rd Probability Summer School held in Saint-Flour, July 6–23, 2003, Edited by Jean Picard.

[8] J.-D. Deuschel and G. Giacomin. Entropic repulsion for the free field: pathwise characterization in $d \geq 3$. *Comm. Math. Phys.*, 206(2):447–462, 1999.

[9] J.-D. Deuschel and Y. Velenik. Non-Gaussian surface pinned by a weak potential. *Probab. Theory Related Fields*, 116(3):359–377, 2000.

[10] C. M. Fortuin, P. W. Kasteleyn, and J. Ginibre. Correlation inequalities on some partially ordered sets. *Comm. Math. Phys.*, 22:89–103, 1971.

[11] S. Galluccio and R. Graber. Depinning transition of a directed polymer by a periodic potential: A d-dimensional solution. *Phys. Rev. E*, 53:R5584–R5587, Jun 1996.

[12] G. Giacomin. Aspects of statistical mechanics of random surfaces. Notes of the lectures given at IHP, 2001.

[13] G. Giacomin. *Random polymer models.* Imperial College Press, London, 2007.

[14] G. Giacomin, H. Lacoin, and F. Toninelli. Marginal relevance of disorder for pinning models. *Comm. Pure Appl. Math.*, 63(2):233–265, 2010.

[15] G. Giacomin, H. Lacoin, and F. L. Toninelli. Disorder relevance at marginality and critical point shift. *Ann. Inst. Henri Poincaré Probab. Stat.*, 47(1):148–175, 2011.

[16] G. Giacomin and F. L. Toninelli. Smoothing effect of quenched disorder on polymer depinning transitions. *Comm. Math. Phys.*, 266(1):1–16, 2006.

[17] D. Ioffe and Y. Velenik. A note on the decay of correlations under δ-pinning. *Probab. Theory Related Fields*, 116(3):379–389, 2000.

[18] É. Janvresse, T. de la Rue, and Y. Velenik. Pinning by a sparse potential. *Stochastic Process. Appl.*, 115(8):1323–1331, 2005.

[19] C. Külske and E. Orlandi. A simple fluctuation lower bound for a disordered massless random continuous spin model in $D = 2$. *Electron. Comm. Probab.*, 11:200–205 (electronic), 2006.

[20] C. Külske and E. Orlandi. Continuous interfaces with disorder: even strong pinning is too weak in two dimensions. *Stochastic Process. Appl.*, 118(11):1973–1981, 2008.

[21] F. Spitzer. *Principles of random walks.* Springer-Verlag, New York, second edition, 1976. Graduate Texts in Mathematics, Vol. 34.

[22] A. C. D. van Enter and C. Külske. Nonexistence of random gradient Gibbs measures in continuous interface models in $d = 2$. *Ann. Appl. Probab.*, 18(1):109–119, 2008.

[23] Y. Velenik. Localization and delocalization of random interfaces. *Probab. Surv.*, 3:112–169 (electronic), 2006.

[24] Y. Velenik. Le modèle d'Ising. Lecture notes, 2009.