



Localization of a Gaussian membrane model with weak pinning potentials

Hironobu Sakagawa

Department of Mathematics, Faculty of Science and Technology, Keio University,
3-14-1 Hiyoshi, Kouhoku-ku, Yokohama 223-8522, JAPAN.
E-mail address: sakagawa@math.keio.ac.jp

Abstract. We consider a class of effective models on \mathbb{Z}^d called Gaussian membrane models with square-well pinning or δ -pinning. It is known that when $d = 1$ this model exhibits a localization/delocalization transition that depends on the strength of the pinning. In this paper, we show that when $d \geq 2$, once we impose weak pinning potentials the field is always localized in the sense that the corresponding free energy is always positive. We also discuss the case that both square-well potentials and repulsive potentials are acting in high dimensions.

1. Introduction

The study of the many equilibrium/non-equilibrium phenomena taking place at interfaces and membranes is an important and fascinating field in statistical mechanics. In particular, the probabilistic study of the effective interface model has been an active area of research for several decades. Yet, in spite of its physical and mathematical importance, the study of the effective membrane model has been hindered by the lack of analytic tools. The main purpose of this paper is to study the localization of the Gaussian membrane model under the effect of weak pinning potentials. First, we introduce our model. Let $d \geq 1$. For a configuration $\phi = \{\phi_x\}_{x \in \mathbb{Z}^d} \in \mathbb{R}^{\mathbb{Z}^d}$, consider the following formal Hamiltonian:

$$H(\phi) = \frac{1}{2} \sum_{x \in \mathbb{Z}^d} (\Delta \phi_x)^2, \quad (1.1)$$

Received by the editors September 13th, 2017; accepted August 23th, 2018.

2010 Mathematics Subject Classification. 60K35, 82B24, 82B41.

Key words and phrases. Random membrane, Localization, Pinning, Free energy.

Research partially supported by JSPS KAKENHI Grant Number 26400147.

where $\Delta = \{\Delta(x, y)\}_{x, y \in \mathbb{Z}^d}$ is a discrete Laplacian on \mathbb{Z}^d , namely,

$$\Delta(x, y) = \begin{cases} \frac{1}{2d} & \text{if } |x - y| = 1, \\ -1 & \text{if } x = y, \\ 0 & \text{otherwise,} \end{cases}$$

and $\Delta f(x) = \sum_{y \in \mathbb{Z}^d} \Delta(x, y)f(y)$ for $f : \mathbb{Z}^d \rightarrow \mathbb{R}$. The configuration ϕ is considered an effective model of a (discretized) random membrane embedded in $d + 1$ -dimensional space, and the spin ϕ_x denotes the height at the position $x \in \mathbb{Z}^d$. This model is also called a $\Delta\phi$ model and is a model of a (tension-less) semiflexible membrane. More precisely, in the model of a membrane such as a lipid bilayer, the energy of the surface separating the water phase and the lipid phase is given by

$$H(\phi) = \sum_{x \in \mathbb{Z}^d} \{\kappa_1(\nabla\phi_x)^2 + \kappa_2(\Delta\phi_x)^2\}, \tag{1.2}$$

where κ_1 and κ_2 are the lateral tension and the bending rigidity, respectively (see [Leiber, 2004](#); [Lipowsky, 1995](#), etc.). When $\kappa_1 > 0$ and $\kappa_2 = 0$, the resulting model is called the $\nabla\phi$ model, and it describes the phase-separating interface. Roughly speaking, the energy of the interface in the $\nabla\phi$ model is determined from the surface area of the microscopic interface ϕ . The $\Delta\phi$ model captures the situation where the surface area of the membrane is preserved, and the energy is determined from the curvature of ϕ . For every $\Lambda \Subset \mathbb{Z}^d$, the finite-volume Gibbs measure on \mathbb{R}^Λ corresponding to the Hamiltonian (1.1) with 0-boundary conditions is defined by

$$P_\Lambda(d\phi) = \frac{1}{Z_\Lambda} \exp\{-H_\Lambda^\Delta(\phi)\} \prod_{x \in \Lambda} d\phi_x,$$

where

$$H_\Lambda^\Delta(\phi) = \frac{1}{2} \sum_{x \in \bar{\Lambda}} (\Delta\phi_x)^2 \Big|_{\phi \equiv 0 \text{ on } \partial_2^+ \Lambda}, \tag{1.3}$$

and

$$Z_\Lambda = \int_{\mathbb{R}^\Lambda} \exp\{-H_\Lambda^\Delta(\phi)\} \prod_{x \in \Lambda} d\phi_x.$$

Here, $d\phi_x$ denotes the Lebesgue measure on \mathbb{R} . We denote the outer and inner boundaries of $\Lambda \subset \mathbb{Z}^d$ with width $k \in \mathbb{N}$ by

$$\partial_k^+ \Lambda := \{x \notin \Lambda; \|y - x\|_1 \leq k \text{ for some } y \in \Lambda\},$$

$$\partial_k^- \Lambda := \{x \in \Lambda; \|y - x\|_1 \leq k \text{ for some } y \notin \Lambda\},$$

respectively, where $\|\cdot\|_1$ denotes the l^1 -norm on \mathbb{Z}^d . For ease of notation, we write $\partial^+ \Lambda$ for $\partial_1^+ \Lambda$, $\partial^- \Lambda$ for $\partial_1^- \Lambda$ and define $\bar{\Lambda} = \Lambda \cup \partial^+ \Lambda$.

We recall some basic properties of the model. First, we have

$$\begin{aligned} \frac{1}{2} \sum_{x \in \mathbb{Z}^d} (\Delta\phi_x)^2 \Big|_{\phi \equiv 0 \text{ on } \Lambda^c} &= H_\Lambda^\Delta(\phi) = \frac{1}{2} \sum_{x \in \bar{\Lambda}} \left(\sum_{y \in \Lambda} \Delta(x, y)\phi_y \right) \left(\sum_{z \in \Lambda} \Delta(x, z)\phi_z \right) \\ &= \frac{1}{2} \sum_{y \in \Lambda} \sum_{z \in \Lambda} \Delta^2(y, z)\phi_y\phi_z \\ &= \frac{1}{2} \langle \phi, \Delta_\Lambda^2 \phi \rangle_\Lambda, \end{aligned} \tag{1.4}$$

where $\langle \cdot, \cdot \rangle_\Lambda$ denotes the $l^2(\Lambda)$ -inner product. We define $\Delta^2 = \Delta\Delta$ as a matrix product and use Δ_Λ^2 to represent its restriction to Λ , namely, $\Delta_\Lambda^2 = \{\Delta_\Lambda^2(x, y)\}_{x, y \in \mathbb{Z}^d}$ with

$$\Delta_\Lambda^2(x, y) = \begin{cases} \Delta^2(x, y) & \text{if } x, y \in \Lambda, \\ 0 & \text{otherwise.} \end{cases}$$

Similarly, Δ_Λ represents the restriction of Δ to Λ . (1.4) shows that the measure P_Λ coincides with the law of a centered Gaussian field on \mathbb{R}^Λ with the covariance matrix $(\Delta_\Lambda^2)^{-1}$. Note that $\Delta_\Lambda^2 \neq (\Delta_\Lambda)^2$. The model is not a ferromagnetic spin system because $\Delta^2(x, y)$ can be both positive and negative for x , and y with x different from y . This means that the basic analytic tools for spin systems or Gaussian fields such as useful correlation inequalities (FKG, Griffith, etc.) and the random walk representation of the covariance, do not hold (see Georgii, 1988; Giacomin, 2001, and the references therein). Therefore, from the mathematical point of view, the membrane model is much less tractable than the $\nabla\phi$ model. However, it is well-known that $(\Delta_\Lambda)^{-1}$ can be represented by the Green function of a simple random walk on \mathbb{Z}^d , and by comparing $(\Delta_{\Lambda_N}^2)^{-1}$ with $(\Delta_{\Lambda_N})^{-2}$, the following estimate on the variance of the field is obtained (see Kurt, 2012, Section 4 and the references therein):

$$\text{Var}_{P_{\Lambda_N}}(\phi_0) = (\Delta_{\Lambda_N}^2)^{-1}(0, 0) = \begin{cases} O(N^{4-d}) & \text{if } d \leq 3, \\ O(\log N) & \text{if } d = 4, \\ O(1) & \text{if } d \geq 5, \end{cases} \quad (1.5)$$

as $N \rightarrow \infty$ where $\Lambda_N := [-N, N]^d \cap \mathbb{Z}^d$. The field is said to be delocalized if $d \leq 4$ because, in this case, the variance diverges as $N \rightarrow \infty$. It is said to be localized if $d \geq 5$ because the variance remains finite. In particular, the corresponding infinite-volume Gaussian measure $Q_\infty \sim \mathcal{N}(0, (-\Delta)^{-2})$ exists only for $d \geq 5$. One of the important properties of is that the field has long-range correlations: $Q_\infty[\phi_x \phi_y] \sim C|x - y|^{-d+4}$ as $|x - y| \rightarrow \infty$ for some constant $C > 0$ (see Sakagawa, 2003, Section 2). We denote expectation with respect to a measure P by $P[\cdot]$. When $d \geq 3$, the asymptotics of the variance (1.5) correspond to those for the $\nabla\phi$ model in dimension $d - 2$. When $d \leq 2$, the model exhibits huge fluctuations known as membrane undulations.

We are now in a position to state the problems and the results of this paper. One problem related to interfaces and membranes is the study of the behavior of the field under the effect of various external potentials. See review articles Funaki (2005); Velenik (2006) for a background on the development of the study for the $\nabla\phi$ model. For a self-potential $U : \mathbb{R} \rightarrow \mathbb{R}$, the corresponding model is generally defined as follows:

$$P_\Lambda^U(d\phi) = \frac{1}{Z_\Lambda^U} \exp\{-H_\Lambda^\Delta(\phi) - \sum_{x \in \Lambda} U(\phi_x)\} \prod_{x \in \Lambda} d\phi_x,$$

where

$$Z_\Lambda^U = \int_{\mathbb{R}^\Lambda} \exp\{-H_\Lambda^\Delta(\phi) - \sum_{x \in \Lambda} U(\phi_x)\} \prod_{x \in \Lambda} d\phi_x. \quad (1.6)$$

To begin with, we consider the effect of weak pinning potentials called square-well pinning. In this case, the self-potential is given by $U_1(r) = -bI(|r| \leq a)$, $a \geq 0$, $b \geq 0$. We denote the corresponding Gibbs measure $P_\Lambda^{U_1}$ and partition function

$Z_\Lambda^{U_1}$ as $P_\Lambda^{a,b}$ and $Z_\Lambda^{a,b}$, respectively. We also define the model with δ -pinning by the following:

$$\tilde{P}_\Lambda^\varepsilon(d\phi) = \frac{1}{\tilde{Z}_\Lambda^\varepsilon} \exp\{-H_\Lambda^\Delta(\phi)\} \prod_{x \in \Lambda} (\varepsilon \delta_0(d\phi_x) + d\phi_x),$$

where

$$\tilde{Z}_\Lambda^\varepsilon = \int_{\mathbb{R}^\Lambda} \exp\{-H_\Lambda^\Delta(\phi)\} \prod_{x \in \Lambda} (\varepsilon \delta_0(d\phi_x) + d\phi_x). \tag{1.7}$$

Denoted by δ_0 a Dirac mass at zero, and let $\varepsilon > 0$ represent the strength of the pinning. The measure $\tilde{P}_\Lambda^\varepsilon$ is obtained as a weak limit of $P_\Lambda^{a,b}$ as $a \rightarrow 0$ and $b \rightarrow \infty$ while maintaining the relation $2a(e^b - 1) = \varepsilon$. For the corresponding $\nabla\phi$ model, detailed properties of the field have been studied and it is known that the field is always localized by the pinning effect for every $d \geq 1$ (see [Funaki, 2005](#), Section 7, [Velenik, 2006](#), Section 5, and the references therein).

The main result of this paper is Theorem 1.1 for the membrane model described therein. Throughout this paper, in the case of $\Lambda = \Lambda_N := [-N, N]^d \cap \mathbb{Z}^d$ we always use subscript N instead of Λ_N and denote as P_N, Z_N, Δ_N etc..

Theorem 1.1. *Let $d \geq 1$. For every $a \geq 0, b \geq 0$ and $\varepsilon \geq 0$ the free energies*

$$F(a, b) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_N^{a,b}}{Z_N}, \quad \tilde{F}(\varepsilon) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\tilde{Z}_N^\varepsilon}{Z_N},$$

exist, and the following hold.

- (1) *When $d = 1$, for every $a > 0$ there exists $b_c = b_c(a) > 0$ such that $F(a, b) > 0$ for every $b > b_c$ and $F(a, b) = 0$ for every $0 \leq b \leq b_c$.*
- (2) *When $d \geq 2$, if $a > 0$ and $b > 0$ then $F(a, b) > 0$. Also, if $\varepsilon > 0$ then $\tilde{F}(\varepsilon) > 0$.*

We note that the model with δ -pinning for $d = 1$ has been studied in [Caravenna and Deuschel \(2008\)](#), and a phase transition occurs that depends on the strength of the pinning. Namely, when $d = 1$ there exists $\varepsilon_c > 0$ such that $\tilde{F}(\varepsilon) > 0$ holds for every $\varepsilon > \varepsilon_c$ and $\tilde{F}(\varepsilon) = 0$ holds whenever $0 \leq \varepsilon \leq \varepsilon_c$ is satisfied. In [Sakagawa \(2012\)](#), the model with δ -pinning was studied, and a result analogous to Theorem 1.1 was obtained for $d \geq 4$. Combining Theorem 1.1 with the results in [Caravenna and Deuschel \(2008\)](#), we can characterize the existence of the localization/delocalization transition by weak pinning potentials in terms of the free energy for all dimensions. For both the model with δ -pinning and the one with square-well pinning, a phase transition occurs only when $d = 1$. When $d \geq 2$, the field is always localized once we put pinning potentials.

When $d \geq 3$, the asymptotics of the variance given in (1.5) for the membrane model correspond to those for the $\nabla\phi$ model in dimension $d - 2$. Therefore, with the result for the $\nabla\phi$ model, we can guess the localization of the membrane model. The most interesting part of Theorem 1.1 is the case $d = 2$. In spite of the large fluctuation, the field is localized by pinning potentials however small the positive parameters a, b , and ε are chosen to be.

Next, let $\rho_N(a, b) := \frac{1}{|\Lambda_N|} P_N^{a,b} [\{|x \in \Lambda_N; |\phi_x| \leq a\}]$ and $\tilde{\rho}_N(\varepsilon) := \frac{1}{|\Lambda_N|} \tilde{P}_N^\varepsilon [\{|x \in \Lambda_N; \phi_x = 0\}]$ be densities of the pinned sites. Then, Theorem 1.1 yields the localization of the field by pinning potentials in the following sense.

Corollary 1.2. (1) When $d = 1$, for every $a > 0$ there exists $b_c = b_c(a) > 0$ such that $\liminf_{N \rightarrow \infty} \rho_N(a, b) > 0$ holds for every $b > b_c$ and $\lim_{N \rightarrow \infty} \rho_N(a, b) = 0$ holds for every $0 \leq b < b_c$.
 (2) When $d \geq 2$, for every $a > 0$, $b > 0$ and $\varepsilon > 0$, both $\liminf_{N \rightarrow \infty} \rho_N(a, b) > 0$ and $\liminf_{N \rightarrow \infty} \tilde{\rho}_N(\varepsilon) > 0$ hold.

Remark 1.3. One of the next steps is to give a pathwise description of the localization of the field by pinning potentials such as uniform bounds on variances and exponential decay of correlations. In the one-dimensional case, the membrane model corresponds to the law of an integrated random walk and a renewal-type argument works well. In particular, its scaling limits have been studied in detail (see Adams et al., 2016; Caravenna and Deuschel, 2009).

However, when $d \geq 2$, there is no known suitable graphical representation of the correlations for the membrane model, and without this crucial tool, the state of knowledge is quite limited. To the best of our knowledge, the only recent relevant result Bolthausen et al. (2017) is that when $d \geq 5$, the membrane model with δ -pinning exhibits exponentially decaying correlations.

The next theorem of the paper pertains to the case when the self-potential is given by $U_2(r) = -bI(0 \leq r \leq a) + lI(r < 0)$, $a \geq 0$, $b \geq 0$, $l \geq 0$. We denote the corresponding Gibbs measure and partition function $P_\Lambda^{U_2}$ and $Z_\Lambda^{U_2}$ by $P_\Lambda^{a,b,l}$ and $Z_\Lambda^{a,b,l}$, respectively. This describes the system that both of weak pinning potentials to height 0 and repulsive potentials from height 0 are imposed. Then, we have the following theorem.

Theorem 1.4. For every $a \geq 0$, $b \geq 0$ and $l \geq 0$ the free energy

$$F(a, b, l) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_N^{a,b,l}}{Z_N},$$

exists, and when $d \geq 5$, if $a > 0$ and $b > 0$ then $F(a, b, l) > 0$ for every $l \geq 0$.

Therefore, when $d \geq 5$, the wetting transition does not occur under the competition between square-well pinning and the soft wall condition of bounded repulsive potentials from $(-\infty, 0)$. The field is always localized by weak pinning.

Remark 1.5. For the $\nabla\phi$ model, it has been shown in Giacomini and Lacoïn (2018) that the conclusion of Theorem 1.4 yields positivity of the free energy for the case with square-well pinning together with the following hard wall condition: $\{\phi_x \geq 0 \text{ for every } x \in \Lambda_N\}$. For technical reasons, including lack of monotonicity, we have not been able to prove the corresponding result for the membrane model.

As a strategy for the proof of Theorem 1.1, we modify the original model to construct the “massive” membrane model by inserting proper potentials that localize the field. Variances of the massive model can be compared with those for a centered Gaussian field on \mathbb{R}^{Λ_N} with the covariance matrix $(m^2 I_N - \Delta_N)^{-2}$, which has a nice random walk representation. Then, the lower bound on the free energy for the massive model is given by Jensen’s inequality and Gaussian computations. We can also show that when $d \geq 2$, the cost of the above modification of the measure is small enough. Hence, we obtain the positivity of the free energy for the original model when $d \geq 2$. For the proof of Theorem 1.4, we modify the original model to a Gaussian field on \mathbb{R}^{Λ_N} with mean $h \in \mathbb{R}$ and covariance matrix $(-\Delta_N)^{-2}$.

Then, an argument similar to the proof of Theorem 1.1 and optimization about the mean height yield the result. We note that these types of arguments have recently been used to study the discrete Gaussian free field with pinning (see [Giacomin and Lacoïn, 2018](#)). One advantage of these arguments is that we do not need to expand the pinning potential parts of (1.6) and (1.7), and therefore we do not need a random walk representation of the covariance under P_N . The key ingredients of our proof are the Gaussian property and the spatial Markov property of the field. In the case of $\Delta\phi$ model, the field has the Markov property with distance two: that is, if $A, B \subset \Lambda$ satisfies $\text{dist}(A, B) > 2$, then $\{\phi_x\}_{x \in A}$ and $\{\phi_x\}_{x \in B}$ are independent under the conditional measure $P_\Lambda(\cdot | \sigma(\{\phi_x; x \notin A \cup B\}))$.

The proofs of Theorems 1.1 and 1.4 apply to the model with both gradient interactions and Laplacian interactions, which correspond to the Hamiltonian (1.2). Let $\kappa_1 > 0$ and $\kappa_2 > 0$ be fixed and consider a Hamiltonian

$$\mathcal{H}_\Lambda(\phi) = \kappa_1 H_\Lambda^\nabla(\phi) + \kappa_2 H_\Lambda^\Delta(\phi), \tag{1.8}$$

for $\Lambda \Subset \mathbb{Z}^d$ where $H_\Lambda^\Delta(\phi)$ is given by (1.3). We define

$$H_\Lambda^\nabla(\phi) = \frac{1}{8d} \sum_{\substack{\{x,y\} \cap \Lambda \neq \emptyset \\ |x-y|=1}} (\phi_x - \phi_y)^2 \Big|_{\phi \equiv 0 \text{ on } \partial^+ \Lambda}.$$

For $U : \mathbb{R} \rightarrow \mathbb{R}$ and $\varepsilon \geq 0$, we define the partition function of the corresponding Gibbs measure with self-potentials and δ -pinning, respectively, by

$$\begin{aligned} \mathcal{Z}_N^U &= \int_{\mathbb{R}^{\Lambda_N}} \exp\left\{-\mathcal{H}_N(\phi) - \sum_{x \in \Lambda_N} U(\phi_x)\right\} \prod_{x \in \Lambda_N} d\phi_x, \\ \tilde{\mathcal{Z}}_N^\varepsilon &= \int_{\mathbb{R}^{\Lambda_N}} \exp\left\{-\mathcal{H}_N(\phi)\right\} \prod_{x \in \Lambda_N} (\varepsilon \delta_0(d\phi_x) + d\phi_x). \end{aligned}$$

Denote the partition function of the finite-volume Gibbs measure on \mathbb{R}^{Λ_N} without self-potentials by $\mathcal{Z}_N := \mathcal{Z}_N^0 = \tilde{\mathcal{Z}}_N^0$. Then, we have the following theorem.

Theorem 1.6. *Let $U_1(r) = -bI(|r| \leq a)$ and $U_2(r) = -bI(0 \leq r \leq a) + lI(r < 0)$, $a \geq 0, b \geq 0, l \geq 0, \varepsilon \geq 0$. The free energies*

$$\begin{aligned} \mathcal{F}(a, b) &:= \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\mathcal{Z}_N^{U_1}}{\mathcal{Z}_N}, \quad \tilde{\mathcal{F}}(\varepsilon) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\tilde{\mathcal{Z}}_N^\varepsilon}{\mathcal{Z}_N}, \\ \mathcal{F}(a, b, l) &:= \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\mathcal{Z}_N^{U_2}}{\mathcal{Z}_N}, \end{aligned}$$

exist and the following hold.

- (1) When $d \geq 1$, $\mathcal{F}(a, b) > 0$ holds for every $a > 0, b > 0$ and $\tilde{\mathcal{F}}(\varepsilon) > 0$ holds for every $\varepsilon > 0$.
- (2) When $d \geq 3$, if $a > 0$ and $b > 0$ then $\mathcal{F}(a, b, l) > 0$ for every $l \geq 0$.

We note that for the case $d = 1$ with δ -pinning, the above result was presented in [Borecki and Caravenna \(2010\)](#). Intuitively, if the energy determined from gradient interactions is small then the energy determined from Laplacian interactions automatically turns out to be small. Hence, the gradient interactions term should dominate over the Laplacian interactions term under the Gibbs measure. For the $\nabla\phi$ model, we know that if we impose only pinning potentials, then localization occurs for every $d \geq 1$ and the wetting transition does not occur when $d \geq 3$.

Therefore, Theorem 1.6 does not come as a surprise. Of course, there are similar technical difficulties in the proof as in the case of $\kappa_1 = 0$.

The rest of the paper is organized as follows. We provide the proofs of Theorems 1.1 and 1.4 in Sections 2 and 3 and the proof of Theorem 1.6 in Section 4. Throughout the paper, C represents a positive constant that does not depend on the size of the system N , but may depend on other parameters. The particular positive constant denoted by C in various estimates may change from place to place in the paper.

2. Proofs of Theorems 1.1 and 1.4

To provide an understanding of the large picture, we outline the proofs of Theorems 1.1 and 1.4 in this section while postponing the proofs of all the technical lemmas to the next section. The first step of the proof is to show the existence of free energy.

Lemma 2.1. *Let $d \geq 1$. Assume that the function $U : \mathbb{R} \rightarrow \mathbb{R}$ is bounded from above. Then, $\lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log Z_N^U$ exists. Also, $\lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \tilde{Z}_N^\varepsilon$ exists for every $\varepsilon \geq 0$.*

Proof of Theorem 1.1: We introduce the finite-volume “massive” membrane model. For $m \geq 0$, consider the following Hamiltonian on Λ_N with 0-boundary conditions:

$$H_{N,m}(\phi) := H_N^\Delta(\phi) + 2m^2 H_N^\nabla(\phi) + \frac{1}{2} m^4 \sum_{x \in \Lambda_N} (\phi_x)^2.$$

We also define the corresponding finite-volume Gibbs measure on \mathbb{R}^{Λ_N} with self-potential $U : \mathbb{R} \rightarrow \mathbb{R}$ and 0-boundary conditions by

$$P_{N,m}^U(d\phi) = \frac{1}{Z_{N,m}^U} \exp\{-H_{N,m}(\phi) - \sum_{x \in \Lambda_N} U(\phi_x)\} \prod_{x \in \Lambda_N} d\phi_x,$$

where

$$Z_{N,m}^U = \int_{\mathbb{R}^{\Lambda_N}} \exp\{-H_{N,m}(\phi) - \sum_{x \in \Lambda_N} U(\phi_x)\} \prod_{x \in \Lambda_N} d\phi_x.$$

The cases without self-potentials are denoted by $P_{N,m}$ and $Z_{N,m}$. Note that both the ϕ -field and the $\nabla\phi$ -field are penalized under this measure, and $Z_{N,m}^U \leq Z_N^U$ holds for every $m \geq 0$.

Now, we take U to be $U_1(r) = -bI(|r| \leq a), a \geq 0, b \geq 0$. We have the following inequalities for every $m \geq 0$:

$$\begin{aligned} F(a,b) &:= \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_N^{U_1}}{Z_N} \\ &\geq \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}^{U_1}}{Z_N} \\ &\geq \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}^{U_1}}{Z_{N,m}} + \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}}{Z_N}. \end{aligned} \tag{2.1}$$

The first term of (2.1) is the free energy of the massive model with square-well pinning. By Jensen’s inequality, the following holds:

$$\log \frac{Z_{N,m}^{U_1}}{Z_{N,m}} = \log P_{N,m} \left[\exp \left\{ - \sum_{x \in \Lambda_N} U_1(\phi_x) \right\} \right] \geq b \sum_{x \in \Lambda_N} P_{N,m}(|\phi_x| \leq a). \tag{2.2}$$

For the estimate on the right hand side, we use the following simple but useful equalities:

$$\begin{aligned} H_N^\Delta(\phi) &= \frac{1}{2} \sum_{x \in \bar{\Lambda}_N} (\Delta\phi_x)^2 \Big|_{\phi \equiv 0 \text{ on } \partial_2^+ \Lambda_N} \\ &= \frac{1}{2} \sum_{x \in \Lambda_N} \left(\sum_{y \in \Lambda_N} \Delta(x, y)\phi_y \right) \left(\sum_{z \in \Lambda_N} \Delta(x, z)\phi_z \right) \\ &\quad + \frac{1}{2} \sum_{x \in \partial^+ \Lambda_N} \left(\sum_{y \in \Lambda_N} \Delta(x, y)\phi_y \right) \left(\sum_{z \in \Lambda_N} \Delta(x, z)\phi_z \right) \\ &= \frac{1}{2} \langle \phi, (-\Delta_N)^2 \phi \rangle_N + B_N(\phi), \end{aligned} \tag{2.3}$$

where $\langle \cdot, \cdot \rangle_N$ denotes the $l^2(\Lambda_N)$ -inner product,

$$B_N(\phi) = \frac{1}{2} \sum_{x \in \partial^+ \Lambda_N} \left(\frac{1}{2d} \sum_{\substack{y \in \Lambda_N \\ |x-y|=1}} \phi_y \right)^2 = \sum_{x \in \partial^- \Lambda_N} \frac{r_N(x)}{8d^2} (\phi_x)^2,$$

and $r_N(x) = |\{y \in \partial^+ \Lambda_N; |y - x| = 1\}|$. Also, by summation by parts, we have the identity

$$H_N^\nabla(\phi) = \frac{1}{8d} \sum_{\substack{\{x,y\} \cap \Lambda_N \neq \emptyset \\ |x-y|=1}} (\phi_x - \phi_y)^2 \Big|_{\phi \equiv 0 \text{ on } \partial^+ \Lambda_N} = \frac{1}{2} \langle \phi, (-\Delta_N)\phi \rangle_N.$$

Then, the equalities

$$\begin{aligned} H_{N,m}(\phi) &= \frac{1}{2} \langle \phi, (-\Delta_N)^2 \phi \rangle_N + B_N(\phi) + m^2 \langle \phi, (-\Delta_N)\phi \rangle_N + \frac{1}{2} m^4 \langle \phi, \phi \rangle_N \\ &= \frac{1}{2} \langle \phi, (m^2 I_N - \Delta_N)^2 \phi \rangle_N + B_N(\phi), \end{aligned} \tag{2.4}$$

holds, where I_N denotes the $\Lambda_N \times \Lambda_N$ identity matrix. (2.4) means that $P_{N,m}$ is obtained by adding convex self-potentials at the inner boundary $\partial^- \Lambda_N$ to a centered Gaussian field on \mathbb{R}^{Λ_N} with the covariance $(m^2 I_N - \Delta_N)^{-2}$. We define the corresponding Gaussian measure as follows:

$$Q_{N,m}(d\phi) = \frac{1}{\Xi_{N,m}} \exp \left\{ -\frac{1}{2} \langle \phi, (m^2 I_N - \Delta_N)^2 \phi \rangle_N \right\} \prod_{x \in \Lambda_N} d\phi_x,$$

where

$$\Xi_{N,m} = \int_{\mathbb{R}^{\Lambda_N}} \exp \left\{ -\frac{1}{2} \langle \phi, (m^2 I_N - \Delta_N)^2 \phi \rangle_N \right\} \prod_{x \in \Lambda_N} d\phi_x.$$

Also, $Q_{\infty,m}$ denotes the law of a centered Gaussian field on $\mathbb{R}^{\mathbb{Z}^d}$ with the covariance $(m^2 I - \Delta)^{-2}$ where I denotes $\mathbb{Z}^d \times \mathbb{Z}^d$ identity matrix. Note that when $m > 0$, infinite-volume measure $Q_{\infty,m}$ exists for every $d \geq 1$. On the other hand, when

$m = 0$, $Q_{\infty,0}$ exists only for $d \geq 5$. Then, by the Brascamp-Lieb inequality (see [Brascamp and Lieb, 1976](#), Theorem 5.1) and Lemma 2.2 (2) below, we have that

$$\text{Var}_{P_{N,m}}(\phi_x) \leq \text{Var}_{Q_{N,m}}(\phi_x) \leq \text{Var}_{Q_{\infty,m}}(\phi_0),$$

holds for every $x \in \Lambda_N$. Combining this variance estimate with (2.2), we obtain the estimate

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}^{U_1}}{Z_{N,m}} \geq b Q_{\infty,m}(|\phi_0| \leq a).$$

The second term of (2.1) is non-positive, which corresponds to the cost for adding mass terms. For this estimate we compare $P_{N,m}$ with $Q_{N,m}$ as follows:

$$\begin{aligned} & \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}}{Z_N} \\ & \geq \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,m}}{\Xi_{N,0}} + \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,0}}{Z_{N,0}} + \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}}{\Xi_{N,m}}. \end{aligned} \quad (2.5)$$

It is well-known that $(m^2 I_N - \Delta_N)^{-1}$ is equal to the Green function of a simple random walk on \mathbb{Z}^d with killing rate $\frac{m^2}{m^2+1}$ and Dirichlet boundary conditions outside Λ_N . Therefore, $Q_{N,m}$ has a nice random walk representation of the covariance and we can show the following asymptotics.

Lemma 2.2. *There exist constants $C_d, \tilde{C}_d > 0$ depending only on the dimension d such that the following hold as $m \downarrow 0$.*

(1)

$$\limsup_{N \rightarrow \infty} \left\{ -\frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,m}}{\Xi_{N,0}} \right\} \leq J_d(m) := \begin{cases} C_d m(1 + o(1)) & \text{if } d = 1, \\ C_d m^2 |\log m|(1 + o(1)) & \text{if } d = 2, \\ C_d m^2(1 + o(1)) & \text{if } d \geq 3. \end{cases}$$

(2)

$$\begin{aligned} \max_{x \in \Lambda_N} \text{Var}_{Q_{N,m}}(\phi_x) & \leq \text{Var}_{Q_{\infty,m}}(\phi_0) \\ & = \sigma_d^2(m) := \begin{cases} \tilde{C}_d m^{-4+d}(1 + o(1)) & \text{if } d = 1, 2, 3, \\ \tilde{C}_d |\log m|(1 + o(1)) & \text{if } d = 4, \\ \tilde{C}_d(1 + o(1)) & \text{if } d \geq 5. \end{cases} \end{aligned}$$

By (2.4) and the fact that $B_N(\phi) \geq 0$, we have that

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,0}}{Z_{N,0}} = \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{1}{Q_{N,m}[e^{-B_N(\phi)}]} \geq 0,$$

trivially. Also, by Jensen's inequality, the definition of $B_N(\phi)$, and uniform bounds on variances under $Q_{N,m}$ we have that

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}}{\Xi_{N,m}} \geq -\limsup_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} Q_{N,m}[B_N(\phi)] = 0,$$

holds for every $m > 0$. Hence, (2.5) and Lemma 2.2 yield that

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_{N,m}}{Z_N} \geq -J_d(m),$$

holds for every $m > 0$.

As a result, we obtain the following estimate on the free energy:

$$\begin{aligned}
 F(a, b) &:= \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_N^{U_1}}{Z_N} \geq b Q_{\infty, m}^0(|\phi_0| \leq a) - J_d(m) \\
 &\geq \frac{2ab}{\sqrt{2\pi\sigma_d^2(m)}} e^{-\frac{a^2}{2\sigma_d^2(m)}} - J_d(m). \tag{2.6}
 \end{aligned}$$

By Lemma 2.2, we know that $\frac{1}{\sqrt{\sigma_d^2(m)}} \gg J_d(m)$ as $m \downarrow 0$ when $d \geq 2$. Hence, by taking $m > 0$ small enough, we have $F(a, b) > 0$ for every $a > 0, b > 0$ when $d \geq 2$. Also, when $d = 1$ if we fix $m > 0$ then we can see that for every $a > 0$ there exists $b_c = b_c(a) > 0$ such that $F(a, b) > 0$ holds for every $b > b_c$.

For the proof of the remaining cases of Theorem 1.1, we use the following comparison estimate. Recall that $F(a, b) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_N^{U_1}}{Z_N}$, $U_1(r) = -bI(|r| \leq a)$ denotes the free energy for the square-well pinning case and $\tilde{F}(\varepsilon) := \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{\tilde{Z}_N^\varepsilon}{Z_N}$ denotes the free energy for the δ -pinning case.

Lemma 2.3. *Let $d \geq 1$. We have $F(a, b) \leq \tilde{F}(\varepsilon(a, b))$ for every $a \geq 0, b \geq 0$ where $\varepsilon(a, b) := 2a(e^b - 1)$.*

Once we have this lemma, Theorem 1.1 for the δ -pinning case when $d \geq 2$ directly follows from the result for the square-well potential case. Also, by the results in Caravenna and Deuschel (2008), we know that there exists ε_c such that $\tilde{F}(\varepsilon) = 0$ for every $0 \leq \varepsilon \leq \varepsilon_c$ when $d = 1$. Therefore, the above comparison estimate shows that for every $a > 0$ there exists $b'_c = b'_c(a) > 0$ such that $F(a, b) = 0$ holds for every $b < b'_c$. Finally, it is easy to see that $F(a, \cdot)$ is convex, and hence continuous, and this completes the proof of the square-well potential case when $d = 1$. \square

Proof of Corollary 1.2: For the case of square-well pinning, it is easy to compute that

$$\frac{\partial}{\partial b} \log Z_N^{a, b} = |\Lambda_N| \rho_N(a, b),$$

and

$$\frac{\partial^2}{\partial b^2} \log Z_N^{a, b} = \text{Var}_{P_N^{a, b}} \left(\sum_{x \in \Lambda_N} I(|\phi_x| \leq a) \right) \geq 0.$$

Hence, we have

$$\frac{1}{|\Lambda_N|} \log \frac{Z_N^{a, b}}{Z_N} = \frac{1}{|\Lambda_N|} \int_0^b \frac{\partial}{\partial b'} \log Z_N^{a, b'} db' = \int_0^b \rho_N(a, b') db',$$

and $\rho_N(a, \cdot)$ is non-decreasing for every $a > 0$. Corollary 1.2 follows from these facts and Theorem 1.1. For the case of δ -pinning, by expanding the δ -pinning part of (1.7) we have

$$\tilde{Z}_N^\varepsilon = \sum_{A \subset \Lambda_N} \varepsilon^{|A|} \int_{\mathbb{R}^{\Lambda_N}} e^{-H_N^\Delta(\phi)} \prod_{x \in \Lambda_N \setminus A} d\phi_x \prod_{x \in A} \delta_0(d\phi_x) = \sum_{A \subset \Lambda_N} \varepsilon^{|A|} Z_{\Lambda_N \setminus A}. \tag{2.7}$$

Then, we can compute that

$$\frac{d}{dt} \log \tilde{Z}_N^{e^t} = |\Lambda_N| \tilde{\rho}_N(e^t),$$

and

$$\frac{d^2}{dt^2} \log \tilde{Z}_N^{\varepsilon^t} = \text{Var}_{\tilde{P}_N^{\varepsilon^t}} \left(\sum_{x \in \Lambda_N} I(\phi_x = 0) \right) \geq 0.$$

Hence, we have

$$\frac{1}{|\Lambda_N|} \log \frac{\tilde{Z}_N^\varepsilon}{Z_N} = \frac{1}{|\Lambda_N|} \int_0^\varepsilon \frac{d}{d\varepsilon'} \log \tilde{Z}_N^{\varepsilon'} d\varepsilon' = \int_0^\varepsilon \frac{1}{\varepsilon'} \tilde{\rho}_N(\varepsilon') d\varepsilon',$$

and $\tilde{\rho}_N(\cdot)$ is non-decreasing. Then, an application of Theorem 1.1 yields Corollary 1.2. \square

Proof of Theorem 1.4: Let $d \geq 5$ and Q_N^h denote the law of a Gaussian field on \mathbb{R}^{Λ_N} with mean $h \in \mathbb{R}$ and covariance $(-\Delta_N)^{-2}$. We compare P_N with Q_N^h as follows:

$$\begin{aligned} & \frac{1}{2} \langle (\phi - h), (-\Delta_N)^2 (\phi - h) \rangle_N \\ &= \frac{1}{2} \langle \phi, (-\Delta_N)^2 \phi \rangle_N + \langle \Delta_N h, \Delta_N \phi \rangle_N + \frac{1}{2} \langle \Delta_N h, \Delta_N h \rangle_N \\ &= \frac{1}{2} \langle \phi, (-\Delta_N)^2 \phi \rangle_N + h T_N(\phi) + \frac{1}{2} h^2 \gamma_N, \end{aligned} \tag{2.8}$$

where $T_N(\phi) := \langle \Delta_N \mathbf{1}, \Delta_N \phi \rangle_N$ and $\gamma_N := \langle \Delta_N \mathbf{1}, \Delta_N \mathbf{1} \rangle_N$. $\mathbf{1}$ denotes a function that is always equal to 1. Because $(\Delta_N \mathbf{1})_x = 0$ holds for every $x \in \Lambda_N \setminus \partial^- \Lambda_N$, T_N is a Gaussian random variable with mean $O(N^{d-1})$ under the law Q_N^h and we have $\gamma_N = O(N^{d-1})$. Then, by (2.3) and (2.8), we have the following:

$$Z_N^U = \int_{\mathbb{R}^{\Lambda_N}} e^{-H_N^{\Delta}(\phi) - U_N(\phi)} \prod_{x \in \Lambda_N} d\phi_x = e^{\frac{1}{2} h^2 \gamma_N} \Xi_N Q_N^h [e^{h T_N(\phi) - B_N(\phi) - U_N(\phi)}],$$

where we set $U_N(\phi) := \sum_{x \in \Lambda_N} U(\phi_x)$ and $\Xi_N := \Xi_{N,0}$.

Now, we take U to be $U_2(r) = -bI(0 \leq r \leq a) + lI(r < 0)$, $a \geq 0$, $b \geq 0$, $l \geq 0$. Jensen's inequality yields that

$$\begin{aligned} F(a, b, l) &:= \lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \frac{Z_N^{U_2}}{Z_N} \\ &\geq \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \left\{ \frac{1}{2} h^2 \gamma_N + \log \frac{\Xi_N}{Z_N} + Q_N^h [h T_N(\phi) - B_N(\phi) - U_N(\phi)] \right\} \\ &\geq \liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \sum_{x \in \Lambda_N} \left\{ b Q_N^h(0 \leq \phi_0 \leq a) - l Q_N^h(\phi_0 < 0) \right\}, \end{aligned}$$

for every $h \in \mathbb{R}$. Note that when $d \geq 5$, we have uniform bounds on variances under Q_N^h and $\lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} Q_N^h [B_N(\phi)] = 0$. Also, Q_N^h weakly converges to Q_∞^h as $N \rightarrow \infty$ where Q_∞^h denotes the law of a Gaussian field on $\mathbb{R}^{\mathbb{Z}^d}$ with mean $h \in \mathbb{R}$ and covariance $(-\Delta)^{-2}$. Therefore,

$$\begin{aligned} F(a, b, l) &\geq b Q_\infty^h(0 \leq \phi_0 \leq a) - l Q_\infty^h(\phi_0 < 0) \\ &\geq b \frac{C_1 \sqrt{\sigma_d^2}}{\sqrt{2\pi h}} e^{-\frac{(h-a)^2}{2\sigma_d^2}} - l \frac{C_2 \sqrt{\sigma_d^2}}{\sqrt{2\pi h}} e^{-\frac{h^2}{2\sigma_d^2}}, \end{aligned}$$

for every $h > 0$ large enough, where the last inequality follows from the Gaussian tail estimate $P(\mathcal{N} > t) \sim \frac{1}{\sqrt{2\pi r}} e^{-\frac{1}{2}r^2}$ as $r \rightarrow \infty$. \mathcal{N} is a standard normal random

variable. By this estimate, if $a > 0$ and $b > 0$ then for every $l > 0$ by choosing $h > 0$ large enough we have $F(a, b, l) > 0$. \square

3. Proofs of the lemmas

In this section we prove all the lemmas presented in Section 2.

Proof of Lemma 2.1: Assume that $U : \mathbb{R} \rightarrow \mathbb{R}$ satisfies the condition $\sup_{r \in \mathbb{R}} U(r) \leq L$ for some $L \in \mathbb{R}$. For the existence of the limit $\lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log Z_N^U$, we first show that there exists $C_0 = C_0(d) > 0$ such that for every $x_0 \in \Lambda \Subset \mathbb{Z}^d$, it holds that

$$\int_{\mathbb{R}} e^{-H_\Lambda(\phi) - \sum_{x \in \Lambda} U(\phi_x)} d\phi_{x_0} \geq C_0 e^{-H_{\Lambda \setminus \{x_0\}}(\phi) - \sum_{x \in \Lambda \setminus \{x_0\}} U(\phi_x)}. \tag{3.1}$$

Define $F(t) := H_\Lambda(\phi) \Big|_{\phi_{x_0}=t}$, $t \in \mathbb{R}$. F is a sum of quadratic functions of t and $|F''(t)|$ is bounded above by some constant $C = C(d) > 0$. Then, for fixed $t_0 \in \mathbb{R}$ we have that

$$F(t) - F(t_0) \leq F'(t_0)(t - t_0) + C(t - t_0)^2,$$

holds for every $t \in \mathbb{R}$. This yields the inequality

$$\int_{\mathbb{R}} e^{-H_\Lambda(\phi) - \sum_{x \in \Lambda} U(\phi_x)} d\phi_{x_0} \geq e^{-L} e^{-F(t_0) - \sum_{x \in \Lambda \setminus \{x_0\}} U(\phi_x)} \int_{\mathbb{R}} e^{-F'(t_0)(t-t_0) - C(t-t_0)^2} dt.$$

By restricting the integral of the right-hand side to $(-\infty, t_0]$ or $[t_0, \infty)$ depending on whether $F'(t_0)$ is positive or negative, the integral is bounded below by $\int_0^\infty e^{-Ct^2} dt < \infty$. Hence we obtain the inequality

$$\int_{\mathbb{R}} e^{-H_\Lambda(\phi) - \sum_{x \in \Lambda} U(\phi_x)} d\phi_{x_0} \geq C' e^{-F(t_0) - \sum_{x \in \Lambda \setminus \{x_0\}} U(\phi_x)},$$

for some $C' > 0$. Because $H_\Lambda(\phi) \Big|_{\phi_{x_0}=0}$ is equal to $H_{\Lambda \setminus \{x_0\}}(\phi)$, if we choose t_0 to be zero then we obtain the inequality (3.1).

Now, we use a super-additivity argument. Let $n \in \mathbb{N}$ be fixed and divide Λ_N into disjoint small boxes with side-length $2n+1$ by imposing 0-boundary conditions on the layer with width two. Explicitly, given $\Gamma_N := \Lambda_N \setminus \left(\bigcup_{x \in (2n+3)\mathbb{Z}^d} (x + \Lambda_n) \right)$,

then $\Lambda_N \setminus \Gamma_N$ consists of $K := \lceil \frac{2N+1}{2n+3} \rceil^d$ disjoint small boxes with the same size as Λ_n , where the distance between each small box is more than 2. Repeated use of the inequality (3.1) yields the inequality $Z_\Lambda^U \geq C_0^{|\Gamma|} Z_{\Lambda \setminus \Gamma}^U$ for every $\Gamma \subset \Lambda \Subset \mathbb{Z}^d$ and using this inequality for $\Lambda = \Lambda_N$ and $\Gamma = \Gamma_N$, we have that $Z_N^U \geq C_0^{(CKn^{d-1} + C'nN^{d-1})} (Z_n^U)^K$ holds for some $C, C' > 0$ by the Markov property of the field. Therefore, we have

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log Z_N^U \geq \frac{C}{n} + \left(\frac{2n+1}{2n+3} \right)^d \frac{1}{|\Lambda_n|} \log Z_n^U,$$

and by taking \limsup we obtain that

$$\liminf_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log Z_N^U \geq \limsup_{n \rightarrow \infty} \frac{1}{|\Lambda_n|} \log Z_n^U.$$

This means that the limit $\lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log Z_N^U$ exists.

Next, for the proof of the existence of the limit $\lim_{N \rightarrow \infty} \frac{1}{|\Lambda_N|} \log \tilde{Z}_N^\varepsilon$, we restrict the sum in (2.7) to sets that include Γ_N . Then, by the Markov property of the field we have that

$$\tilde{Z}_N^\varepsilon = \sum_{A \subset \Lambda_N} \varepsilon^{|A|} Z_{\Lambda_N \setminus A} \geq \sum_{A \subset \Lambda_N \setminus \Gamma_N} \varepsilon^{|A| + |\Gamma_N|} Z_{\Lambda_N \setminus (A \cup \Gamma_N)} = \varepsilon^{|\Gamma_N|} (\tilde{Z}_N^\varepsilon)^K,$$

holds. The rest of the proof is the same as in the case with self-potentials. □

Proof of Lemma 2.2: (1) Since

$$\Xi_{N,m} = (2\pi)^{\frac{|\Lambda_N|}{2}} \sqrt{\det((m^2 I_N - \Delta_N)^{-2})} = (2\pi)^{\frac{|\Lambda_N|}{2}} (\det(m^2 I_N - \Delta_N))^{-1},$$

we can use the following random walk representation (see Bolthausen and Ioffe, 1997, Section 4.1):

$$\begin{aligned} \log \Xi_{N,m} &= \frac{1}{2} |\Lambda_N| \log(2\pi) - |\Lambda_N| \log(m^2 + 1) \\ &\quad + \sum_{x \in \Lambda_N} \sum_{n=1}^{\infty} \frac{1}{n} \left(\frac{1}{m^2 + 1}\right)^n \mathbb{P}_x(S_n = x, n < \tau_{\Lambda_N}), \end{aligned}$$

where $\{S_n\}_{n \geq 0}$ is a simple random walk on \mathbb{Z}^d , and \mathbb{P}_x denotes its law starting at $x \in \mathbb{Z}^d$. $\tau_A := \inf\{n \geq 0; S_n \notin A\}$ denotes the first exit time from $A \subset \mathbb{Z}^d$. Therefore,

$$\begin{aligned} \limsup_{N \rightarrow \infty} \left\{ -\frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,m}}{\Xi_{N,0}} \right\} &\leq m^2 + \sum_{n=1}^{\infty} \frac{1}{n} \left\{ 1 - \left(\frac{1}{m^2 + 1}\right)^n \right\} \mathbb{P}_0(S_n = 0) \\ &\leq m^2 + C \sum_{n=1}^{\infty} n^{-1 - \frac{d}{2}} \left\{ 1 - \left(\frac{1}{m^2 + 1}\right)^n \right\} \\ &\leq m^2 + C \sum_{n=1}^{[m^{-2}]} m^2 n^{-\frac{d}{2}} + C \sum_{n=[m^{-2}]+1}^{\infty} n^{-1 - \frac{d}{2}}, \end{aligned}$$

holds for some constant $C = C(d) > 0$, where we used the local central limit theorem $\mathbb{P}_0(S_n = 0) \sim Cn^{-\frac{d}{2}}$ to obtain for the second inequality. The last inequality follows from an estimate $(1+a)^{-n} \geq 1-na$ for $n \geq 1$ and $a > 0$. Hence, we obtain the estimate

$$\limsup_{N \rightarrow \infty} \left\{ -\frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,m}}{\Xi_{N,0}} \right\} \leq J_d(m) := \begin{cases} C_d m(1 + o(1)) & \text{if } d = 1, \\ C_d m^2 |\log m| (1 + o(1)) & \text{if } d = 2, \\ C_d m^2 (1 + o(1)) & \text{if } d \geq 3, \end{cases}$$

as $m \downarrow 0$. □

(2) By the random walk representation of $(m^2 I_N - \Delta_N)^{-1}$, we have that

$$\begin{aligned} \text{Var}_{Q_{N,m}}(\phi_x) &= (m^2 I_N - \Delta_N)^{-2}(x, x) \\ &= \sum_{y \in \Lambda_N} (m^2 I_N - \Delta_N)^{-1}(x, y)(m^2 I_N - \Delta_N)^{-1}(y, x) \\ &= \sum_{y \in \Lambda_N} \left\{ \frac{1}{m^2 + 1} \sum_{k=0}^{\infty} \left(\frac{1}{m^2 + 1}\right)^k \mathbb{P}_x(S_k = y, k < \tau_{\Lambda_N}) \right\} \\ &\quad \times \left\{ \frac{1}{m^2 + 1} \sum_{l=0}^{\infty} \left(\frac{1}{m^2 + 1}\right)^l \mathbb{P}_y(S_l = x, l < \tau_{\Lambda_N}) \right\} \\ &= \frac{1}{(m^2 + 1)^2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left(\frac{1}{m^2 + 1}\right)^{k+l} \mathbb{P}_x(S_{k+l} = x, k+l < \tau_{\Lambda_N}) \\ &\leq \frac{1}{(m^2 + 1)^2} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left(\frac{1}{m^2 + 1}\right)^{k+l} \mathbb{P}_x(S_{k+l} = x) \\ &= \frac{1}{(m^2 + 1)^2} \sum_{n=0}^{\infty} (n+1) \left(\frac{1}{m^2 + 1}\right)^n \mathbb{P}_x(S_n = x) = \text{Var}_{Q_{\infty,m}}(\phi_0). \end{aligned}$$

For the asymptotics of $\text{Var}_{Q_{\infty,m}}(\phi_0)$, we use the local central limit theorem and the asymptotics of the polylogarithm: $\sum_{k=1}^{\infty} \frac{(e^\mu)^k}{k^s} \sim \Gamma(1-s)(-\mu)^{s-1}$ as $\mu \uparrow 0$ if $s < 1$ when $d \leq 3$ and the expansion $\sum_{k=1}^{\infty} \frac{x^k}{k} = -\log(1-x)$, $0 < x < 1$ when $d = 4$. Then, we can obtain

$$\text{Var}_{Q_{\infty,m}}(\phi_0) = \sigma_d^2(m) := \begin{cases} \tilde{C}_d m^{-4+d}(1 + o(1)) & \text{if } d = 1, 2, 3, \\ \tilde{C}_d |\log m|(1 + o(1)) & \text{if } d = 4, \\ \tilde{C}_d(1 + o(1)) & \text{if } d \geq 5, \end{cases}$$

as $m \downarrow 0$. □

Proof of Lemma 2.3: By expanding the pinning part, we have

$$\begin{aligned} \frac{Z_N^{a,b}}{Z_N} &= P_N \left[\exp \left\{ \sum_{x \in \Lambda_N} bI(|\phi_x| \leq a) \right\} \right] = \sum_{A \subset \Lambda_N} (e^b - 1) P_N(|\phi_x| \leq a \text{ for every } x \in A), \\ \frac{\tilde{Z}_N^\varepsilon}{Z_N} &= \sum_{A \subset \Lambda_N} \varepsilon^{|A|} \frac{Z_N(A)}{Z_N}, \end{aligned}$$

where

$$Z_N(A) = \int_{\mathbb{R}^{\Lambda_N}} e^{-H_N^\Delta(\phi)} \prod_{x \in \Lambda_N \setminus A} d\phi_x \prod_{x \in A} \delta_0(d\phi_x).$$

If we can show that

$$P_N(|\phi_x| \leq a \text{ for every } x \in A) \leq (2a)^{|A|} \frac{Z_N(A)}{Z_N}, \tag{3.2}$$

for every $A \subset \Lambda_N$, then we have

$$\frac{Z_N^{a,b}}{Z_N} \leq \sum_{A \subset \Lambda_N} (e^b - 1)^{|A|} (2a)^{|A|} \frac{Z_N(A)}{Z_N} = \frac{\tilde{Z}_N^{\varepsilon(a,b)}}{Z_N},$$

with $\varepsilon(a, b) = 2a(e^b - 1)$, and this yields $F(a, b) \leq \tilde{F}(\varepsilon(a, b))$. Now, we prove the inequality (3.2). For a given $\{\phi_x\}_{x \in A}$, let ϕ^* be a solution of the following Dirichlet boundary problem:

$$\begin{cases} \Delta^2 \phi^*(x) = 0 & \text{for every } x \in \Lambda_N \setminus A, \\ \phi^*(x) = \phi(x) & \text{for every } x \in A, \\ \phi^*(x) = 0 & \text{for every } x \in \Lambda_N^c. \end{cases}$$

Then, for every configuration $\tilde{\phi} \in \mathbb{R}^{\Lambda_N}$ that coincides with ϕ on A , we have

$$H_N^\Delta(\tilde{\phi}) = H_N^\Delta(\tilde{\phi} - \phi^*) + H_N^\Delta(\phi^*) \geq H_N^\Delta(\tilde{\phi} - \phi^*),$$

where the equality follows from the fact that $\sum_{x \in \overline{\Lambda_N}} (\Delta \psi_x) \eta_x = \sum_{x \in \overline{\Lambda_N}} \psi_x (\Delta \eta_x)$ for every $\eta \in \mathbb{R}^{\Lambda_N \cup \partial_2^+ \Lambda_N}$ and $\psi \in \mathbb{R}^{\Lambda_N \cup \partial_2^+ \Lambda_N}$ which satisfies $\psi_x = 0$ for every $x \in \partial_2^+ \Lambda_N$ combining with $\Delta^2 \phi^* = 0$ on $\Lambda_N \setminus A$ and $\tilde{\phi} - \phi^* = 0$ on A . By this estimate,

$$\begin{aligned} & Z_N P_N(|\phi_x| \leq a \text{ for every } x \in A) \\ &= \int_{\mathbb{R}^A} I(|\phi_x| \leq a \text{ for every } x \in A) \left(\int_{\mathbb{R}^{\Lambda_N \setminus A}} e^{-H_N^\Delta(\phi)} \prod_{x \in \Lambda_N \setminus A} d\phi_x \right) \prod_{x \in A} d\phi_x \\ &\leq \int_{\mathbb{R}^A} I(|\phi_x| \leq a \text{ for every } x \in A) \left(\int_{\mathbb{R}^{\Lambda_N \setminus A}} e^{-H_N^\Delta(\tilde{\phi} - \phi^*)} \prod_{x \in \Lambda_N \setminus A} d\tilde{\phi}_x \right) \prod_{x \in A} d\phi_x \\ &= (2a)^{|A|} Z_N(A), \end{aligned}$$

and we obtain the inequality (3.2), as desired. □

4. Proof of Theorem 1.6

In this section we consider the model with both gradient interactions and Laplacian interactions. By a simple change of variables, we may assume that $\kappa_1 = \kappa > 0$ and $\kappa_2 = 1$ in Hamiltonian (1.8), and we may further define $\mathcal{H}_{\Lambda, \kappa}(\phi) := \kappa H_\Lambda^\nabla(\phi) + H_\Lambda^\Delta(\phi)$ for $\Lambda \in \mathbb{Z}^d$. We insert mass terms for this model as follows:

$$\mathcal{H}_{N, \kappa, m}(\phi) := \mathcal{H}_{N, \kappa}(\phi) + m^2 H_N^\nabla(\phi) + \frac{1}{2} m^2 \kappa \sum_{x \in \Lambda_N} (\phi_x)^2.$$

The corresponding Gibbs measure is defined by

$$\mathcal{P}_{N, \kappa, m}(d\phi) = \frac{1}{\mathcal{Z}_{N, \kappa, m}} \exp\{-\mathcal{H}_{N, \kappa, m}(\phi)\} \prod_{x \in \Lambda_N} d\phi_x,$$

where $\mathcal{Z}_{N, \kappa, m}$ denotes the normalization factor.

Now, we have

$$\begin{aligned} \mathcal{H}_{N, \kappa, m}(\phi) &= \frac{1}{2} \kappa \langle \phi, (-\Delta_N) \phi \rangle_N + \frac{1}{2} \langle \phi, (-\Delta_N)^2 \phi \rangle_N + B_N(\phi) \\ &\quad + \frac{1}{2} m^2 \langle \phi, (-\Delta_N) \phi \rangle_N + \frac{1}{2} m^2 \kappa \langle \phi, \phi \rangle_N \\ &= \frac{1}{2} \langle \phi, (m^2 I_N - \Delta_N)(\kappa I_N - \Delta_N) \phi \rangle_N + B_N(\phi). \end{aligned}$$

Then, we define a Gaussian measure $Q_{N,\kappa,m}$ as follows:

$$Q_{N,\kappa,m}(d\phi) = \frac{1}{\Xi_{N,\kappa,m}} \exp\left\{-\frac{1}{2}\langle\phi, (m^2I_N - \Delta_N)(\kappa I_N - \Delta_N)\phi\rangle_N\right\} \prod_{x \in \Lambda_N} d\phi_x,$$

where

$$\Xi_{N,\kappa,m} = \int_{\mathbb{R}^{\Lambda_N}} \exp\left\{-\frac{1}{2}\langle\phi, (m^2I_N - \Delta_N)(\kappa I_N - \Delta_N)\phi\rangle_N\right\} \prod_{x \in \Lambda_N} d\phi_x.$$

The law of the corresponding centered Gaussian field on $\mathbb{R}^{\mathbb{Z}^d}$ with the covariance $(m^2I - \Delta)^{-1}(\kappa I - \Delta)^{-1}$ is denoted by $Q_{\infty,\kappa,m}$. As for $P_{N,m}$, the finite-volume Gibbs measure $\mathcal{P}_{N,\kappa,m}$ has the Markov property with distance two and an argument similar to that in the proofs of Theorem 1.1 and 1.4 works well. Differences are only the way of adding the mass term and asymptotic behaviors as $m \downarrow 0$. These are given by Lemma 4.1 below.

Lemma 4.1. *There exist constants $C'_d, \tilde{C}'_d > 0$ such that the following hold:*

(1)

$$\limsup_{N \rightarrow \infty} \left\{ -\frac{1}{|\Lambda_N|} \log \frac{\Xi_{N,\kappa,m}}{\Xi_{N,\kappa,0}} \right\} \leq J'_d(m) := \begin{cases} C'_d m(1 + o(1)) & \text{if } d = 1, \\ C'_d m^2 |\log m|(1 + o(1)) & \text{if } d = 2, \\ C'_d m^2(1 + o(1)) & \text{if } d \geq 3. \end{cases}$$

(2)

$$\begin{aligned} \max_{x \in \Lambda_N} \text{Var}_{Q_{N,\kappa,m}}(\phi_x) &\leq \text{Var}_{Q_{\infty,\kappa,m}}(\phi_0) \\ &\leq \sigma'^2_d(m) := \begin{cases} \tilde{C}'_d m^{-1}(1 + o(1)) & \text{if } d = 1, \\ \tilde{C}'_d |\log m|(1 + o(1)) & \text{if } d = 2, \\ \tilde{C}'_d(1 + o(1)) & \text{if } d \geq 3. \end{cases} \end{aligned}$$

Proof: (1) Since $\Xi_{N,\kappa,m} = (2\pi)^{\frac{|\Lambda_N|}{2}} \sqrt{\det((m^2I_N - \Delta_N)^{-1}(\kappa I_N - \Delta_N)^{-1})}$, we have $\log \frac{\Xi_{N,\kappa,m}}{\Xi_{N,\kappa,0}} = \frac{1}{2} \log \frac{\Xi_{N,m}}{\Xi_{N,0}}$. Hence the same estimate as given for Lemma 2.2 (1) holds.

(2) Similar to the proof of Lemma 2.2 we can compute that

$$\begin{aligned} \max_{x \in \Lambda_N} \text{Var}_{Q_{N,\kappa,m}}(\phi_x) &\leq \text{Var}_{Q_{\infty,\kappa,m}}(\phi_0) \\ &= ((m^2I - \Delta)^{-1}(\kappa I - \Delta)^{-1})(0, 0) \\ &= \sum_{x \in \mathbb{Z}^d} (m^2I - \Delta)^{-1}(0, x)(\kappa I - \Delta)^{-1}(x, 0) \\ &= \frac{1}{(m^2 + 1)(\kappa + 1)} \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} \left(\frac{1}{m^2 + 1}\right)^k \left(\frac{1}{\kappa + 1}\right)^l \mathbb{P}_0(S_{k+l} = 0). \end{aligned}$$

By the local central limit theorem we have that the estimate

$$\mathbb{P}_0(S_{k+l} = 0) \leq C(k+l)^{-\frac{d}{2}} \leq Ck^{-\frac{d}{2}},$$

holds for every $k \geq 1$ and $l \geq 1$. These estimates yield the bound

$$\text{Var}_{Q_{\infty,\kappa,m}}(\phi_0) \leq \frac{C}{m^2 + 1} \sum_{k=0}^{\infty} \frac{1}{k^{\frac{d}{2}}} \left(\frac{1}{m^2 + 1}\right)^k,$$

for some $C = C(\kappa, d) > 0$. Hence we obtain the variance

$$\mathrm{Var}_{Q_{\infty, \kappa, m}}(\phi_0) \leq \sigma_d^{\prime 2}(m) := \begin{cases} \tilde{C}'_d m^{-1}(1 + o(1)) & \text{if } d = 1, \\ \tilde{C}'_d |\log m|(1 + o(1)) & \text{if } d = 2, \\ \tilde{C}'_d(1 + o(1)) & \text{if } d \geq 3, \end{cases}$$

as $m \downarrow 0$. □

For the case of square-well potentials, we have $\frac{1}{\sqrt{\sigma_d^{\prime 2}(m)}} \gg J'_d(m)$ as $m \downarrow 0$ for every $d \geq 1$ in the argument of the inequality (2.6) by Lemma 4.1 and we can see that the field is always localized for every $d \geq 1$. Also, if $d \geq 3$ then the law $Q_{\infty, \kappa, 0}$ exists, and the argument of the proof of Theorem 1.4 works for the case of square-well potentials and repulsive potentials. □

References

- S. Adams, A. Kister and H. Weber. Sample path large deviations for Laplacian models in $(1 + 1)$ -dimensions. *Electron. J. Probab.* **21**, Paper No. 62, 36 (2016). [MR3563890](#).
- E. Bolthausen, A. Cipriani and N. Kurt. Exponential decay of covariances for the supercritical membrane model. *Comm. Math. Phys.* **353** (3), 1217–1240 (2017). [MR3652489](#).
- E. Bolthausen and . Ioffe. Harmonic crystal on the wall: a microscopic approach. *Comm. Math. Phys.* **187** (3), 523–566 (1997). [MR1468313](#).
- M. Borecki and F. Caravenna. Localization for $(1 + 1)$ -dimensional pinning models with $(\nabla + \Delta)$ -interaction. *Electron. Commun. Probab.* **15**, 534–548 (2010). [MR2737711](#).
- H. J. Brascamp and E. H. Lieb. On extensions of the Brunn-Minkowski and Prékopa-Leindler theorems, including inequalities for log concave functions, and with an application to the diffusion equation. *J. Functional Analysis* **22** (4), 366–389 (1976). [MR0450480](#).
- F. Caravenna and J.-D. Deuschel. Pinning and wetting transition for $(1 + 1)$ -dimensional fields with Laplacian interaction. *Ann. Probab.* **36** (6), 2388–2433 (2008). [MR2478687](#).
- F. Caravenna and J.-D. Deuschel. Scaling limits of $(1 + 1)$ -dimensional pinning models with Laplacian interaction. *Ann. Probab.* **37** (3), 903–945 (2009). [MR2537545](#).
- T. Funaki. Stochastic interface models. In *Lectures on probability theory and statistics*, volume 1869 of *Lecture Notes in Math.*, pages 103–274. Springer, Berlin (2005). [MR2228384](#).
- H.-O. Georgii. *Gibbs measures and phase transitions*, volume 9 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin (1988). ISBN 0-89925-462-4. [MR956646](#).
- G. Giacomin. Aspects of statistical mechanics of random surfaces (2001). Notes of the lectures given at IHP in the fall 2001 (available at the web page of the author).
- G. Giacomin and H. Lacoin. Disorder and wetting transition: the pinned harmonic crystal in dimension three or larger. *Ann. Appl. Probab.* **28** (1), 577–606 (2018). [MR3770884](#).

-
- N. Kurt. Laplacian interface models with strictly convex potential. *Markov Process. Related Fields* **18** (1), 9–30 (2012). [MR2952017](#).
- S. Leibler. Equilibrium statistical mechanics of fluctuating films and membranes. *Statistical Mechanics of Membranes and Surfaces, 2nd ed.* pages 49–101 (2004). [DOI: 10.1142/9789812565518_0003](#).
- R. Lipowsky. Generic interaction of flexible membranes. *Handbook of Biological Physics* **1**, 521–602 (1995).
- H. Sakagawa. Entropic repulsion for a Gaussian lattice field with certain finite range interaction. *J. Math. Phys.* **44** (7), 2939–2951 (2003). [MR1982781](#).
- H. Sakagawa. On the free energy of a Gaussian membrane model with external potentials. *J. Stat. Phys.* **147** (1), 18–34 (2012). [MR2922757](#).
- Y. Velenik. Localization and delocalization of random interfaces. *Probab. Surv.* **3**, 112–169 (2006). [MR2216964](#).