

THE ENDPOINT DISTRIBUTION OF DIRECTED POLYMERS

BY ERIK BATES¹ AND SOURAV CHATTERJEE²

¹*Department of Mathematics, Stanford University, ewbates@stanford.edu*

²*Department of Statistics, Stanford University, souravc@stanford.edu*

Probabilistic models of directed polymers in random environment have received considerable attention in recent years. Much of this attention has focused on integrable models. In this paper, we introduce some new computational tools that do not require integrability. We begin by defining a new kind of abstract limit object, called “partitioned subprobability measure,” to describe the limits of endpoint distributions of directed polymers. Inspired by a recent work of Mukherjee and Varadhan on large deviations of the occupation measure of Brownian motion, we define a suitable topology on the space of partitioned subprobability measures and prove that this topology is compact. Then using a variant of the cavity method from the theory of spin glasses, we show that any limit law of a sequence of endpoint distributions must satisfy a fixed point equation on this abstract space, and that the limiting free energy of the model can be expressed as the solution of a variational problem over the set of fixed points. As a first application of the theory, we prove that in an environment with finite exponential moment, the endpoint distribution is asymptotically purely atomic if and only if the system is in the low temperature phase. The analogous result for a heavy-tailed environment was proved by Vargas in 2007. As a second application, we prove a subsequential version of the longstanding conjecture that in the low temperature phase, the endpoint distribution is asymptotically localized in a region of stochastically bounded diameter. All our results hold in arbitrary dimensions, and make no use of integrability.

CONTENTS

1. Introduction	818
1.1. The model of directed polymers in random environment	819
1.2. An overview of known results about general polymer models	820
1.2.1. High and low temperature phases	820
1.2.2. Weak and strong disorder regimes	821
1.2.3. Further versions of localization	823
1.2.4. A remark about the definition of endpoint distribution	824
1.3. Related results in integrable models	824
1.4. An overview of results proved in this paper	825
1.4.1. Compactification of measures on \mathbb{Z}^d	826
1.4.2. The update map	827
1.4.3. The energy functional	828
1.5. Main applications	829
The single-copy condition	830
2. Partitioned subprobability measures	830
2.1. Definition and properties	830
2.2. Compactness	836
3. The update map	837
3.1. Definition of the update map	837
3.2. Lifting the update map	840

Received June 2018; revised February 2019.

MSC2010 subject classifications. Primary 60K37; secondary 82B26, 82B44, 82D60.

Key words and phrases. Directed polymer, free energy, disordered system, phase transition.

4. The empirical measure of the endpoint distribution	840
4.1. Definition and properties	840
4.2. Convergence to fixed points of the update map	843
4.3. Variational formula for the free energy	844
5. Limits of empirical measures	847
5.1. Characterizing high and low temperature phases	848
5.2. An illustrative application	849
6. Asymptotic pure atomicity	851
6.1. Definitions of relevant functionals	851
6.2. Proof of asymptotic pure atomicity at low temperature	852
7. Geometric localization	854
7.1. Definitions of relevant functionals	855
7.2. Proof of geometric localization with positive density	857
7.3. Localization in a favorite region	858
8. Proof of Proposition 3.2	860
9. Open problems	866
Acknowledgments	867
Supplementary Material	867
References	867

1. Introduction. The model of directed polymers in random environment was introduced in the physics literature by Huse and Henley [44] to represent the phase boundary of the Ising model in the presence of random impurities. It was later mathematically reformulated as a model of random walk in random potential by Imbrie and Spencer [45]. Over the last thirty years, the directed polymer model has played an important role as a source of many fascinating problems in the probability literature, culminating in the amazing recent developments in integrable polymer models. However, in spite of the wealth of information now available for integrable models, our knowledge about the general case is fairly limited, especially in spatial dimension greater than one. The goal of this paper is to introduce an abstract theory that allows computations for polymer models that are not integrable. Before we discuss our approach in detail, let us very briefly communicate the main consequences:

- The probabilistic model of $(d + 1)$ -dimensional directed polymers of length n assigns a random probability measure to the set of random walk paths of length n in \mathbb{Z}^d that start at the origin. The precise mathematical model is defined in Section 1.1 below. The “endpoint distribution” is the probability distribution of the final vertex reached by the random walk. Note that this is a random probability measure on \mathbb{Z}^d , which we will denote f_n . Understanding the behavior of f_n is the main goal of this article. Although a number of results about f_n were known prior to this work (reviewed in Section 1.2), this is the first paper that puts forward a comprehensive theoretical framework for analyzing the asymptotic properties of f_n .
- Like many statistical mechanical models, directed polymers have a low temperature phase and a high temperature phase. One of the most striking features of directed polymers is that in the low temperature phase, f_n has “atoms” whose weights do not decay to zero as $n \rightarrow \infty$. This is in stark contrast with the endpoint distribution of simple random walk, where the most likely site has mass of order $n^{-d/2}$. The precise statement of this well-known localization phenomenon will be discussed in Section 1.2. One of the main applications of the abstract machinery developed in this paper is to show that in the low temperature regime, the atoms account for *all* of the mass—that is, there is no part of the mass that diffuses out. Such a result was proved earlier for directed polymers in heavy-tailed environment, and also for a particular $(1 + 1)$ -dimensional integrable model. We prove it under finite exponential moments, to which previously known techniques do not apply.

- Our second main application is to show that in the low temperature regime, there is almost surely a subsequence of positive density along which the endpoint distribution concentrates mass $> 1 - \delta$ on a set of diameter $\leq K(\delta)$, where δ is arbitrary and $K(\delta)$ is a deterministic constant that depends only on δ and some features of the model. In other words, not only does the mass localize on atoms, but the atoms themselves localize in a set of bounded diameter. This proves a subsequential version of a longstanding conjecture about the endpoint distribution. Prior to this work, the only case where a similar statement could be proved was for an integrable model.
- Crucial to each of these results is a new variational formula for the limiting free energy, given as the infimum of a continuous functional over a certain closed subset \mathcal{K} of a compact space. The high and low temperature phases are then characterized respectively by whether \mathcal{K} is just a single trivial object or instead contains nontrivial elements.

We will now begin a more detailed presentation by defining the model below. This is followed by a discussion of the known results about polymer models, and then a general overview of the results proved in this paper and the ideas involved in the proofs.

REMARK. For the sake of presentation and length, some of the more technical proofs needed for this manuscript are contained in supplementary material, henceforth referred to as Appendix A and Appendix B [11]. When a proof has been omitted from the present text, a reference to its location in the appendices will be provided. Alternatively, a complete version of the manuscript can be found at <https://arxiv.org/abs/1612.03443>.

1.1. *The model of directed polymers in random environment.* Take any integer $d \geq 1$. The probabilistic model of $(d + 1)$ -dimensional *directed polymers in random environment* is defined as follows. We begin with a simple random walk $\omega = (\omega_i)_{i \geq 0}$ on \mathbb{Z}^d , letting P denote the law of the walk when started at the origin. Let E denote expectation according to P .

Next, let $\mathbb{N} := \{1, 2, \dots\}$ and introduce a collection of i.i.d. random variables $(X_u)_{u \in \mathbb{N} \times \mathbb{Z}^d}$ called the *random environment*, defined on some probability space $(\Omega_e, \mathcal{F}, \mathbf{P})$. We will write \mathbf{E} for expectation according to \mathbf{P} . In what follows, it will not be problematic to define all random variables on the abstract probability space $(\Omega_e, \mathcal{F}, \mathbf{P})$. Unless stated otherwise, “almost sure” statements are made with respect to \mathbf{P} .

Let $\beta > 0$ be a parameter, called the *inverse temperature*. Let \mathcal{L} denote the common law of the X_u , often called the *disorder distribution*. We will assume the logarithmic moment generating function for \mathcal{L} satisfies

$$(1.1) \quad \lambda(\alpha) := \log \mathbf{E}(e^{\alpha X_u}) < \infty \quad \text{for all } \alpha \in [-2\beta, 2\beta].$$

We now have the notation to define the model of $(d + 1)$ -dimensional directed polymers in random environment. The *quenched polymer measure* of length $n \geq 0$, denoted ρ_n , is the Gibbs measure for P with Hamiltonian

$$H_n(\omega) := - \sum_{i=1}^n X_{i, \omega_i}.$$

That is,

$$(1.2) \quad \rho_n(d\omega) = \frac{1}{Z_n} e^{-\beta H_n(\omega)} P(d\omega),$$

where the normalization constant $Z_n := E(e^{-\beta H_n(\omega)})$ is called the *quenched partition function*. Explicitly,

$$Z_n = \frac{1}{(2d)^n} \sum_{\gamma} \exp\left(\beta \sum_{i=1}^n X_{i, \gamma(i)}\right),$$

where the sum is over the $(2d)^n$ nearest-neighbor paths $\gamma : \{0, 1, \dots, n\} \rightarrow \mathbb{Z}^d$ of length $|\gamma| = n$, starting at the origin ($\gamma(0) = 0$). We can equivalently write

$$Z_n = \frac{1}{(2d)^n} \sum_{x \in \mathbb{Z}^d} Z_n(x),$$

where

$$Z_n(x) := \sum_{\gamma: \gamma(n)=x} \exp\left(\beta \sum_{i=1}^n X_{i, \gamma(i)}\right)$$

is often referred to as a *point-to-point partition function*. When $d = 1$, convention often replaces Z_n by $(2d)^n Z_n$, which is correspondingly called the *point-to-line partition function*. So that the model does not reduce to a simple random walk (which is the case when $\beta = 0$), we assume the X_u are nondegenerate random variables. That is, \mathfrak{L} is not supported on a single point.

1.2. *An overview of known results about general polymer models.* In this section, we will review the results that are known in arbitrary dimensions, with usually mild assumptions on the distribution of the environment. Indeed, our paper works in this general setting, and so these results will be most relevant to the present study. One can find a more detailed review in the recent notes by Comets [25].

1.2.1. *High and low temperature phases.* The qualitative behavior of directed polymers depends on the disorder distribution \mathfrak{L} , the inverse temperature β , and the traversal dimension d . Much of this dependence can be observed through the system’s *quenched free energy*, given by

$$F_n := \frac{\log Z_n}{n},$$

where $F_0 := 0$. Like $(\rho_n)_{n \geq 0}$ and $(Z_n)_{n \geq 0}$, the sequence $(F_n)_{n \geq 0}$ is a random process with respect to the filtration $(\mathcal{F}_n)_{n \geq 0}$, where

$$(1.3) \quad \mathcal{F}_n := \sigma(X_{i,x} : 1 \leq i \leq n, x \in \mathbb{Z}^d).$$

When the randomness of the environment is averaged out, one obtains the *averaged quenched free energy*,

$$\mathbf{E}(F_n) = \frac{1}{n} \mathbf{E} \log Z_n.$$

In general, this quantity is distinct from the *annealed free energy*, which is

$$\frac{1}{n} \log \mathbf{E}(Z_n) = \frac{1}{n} \log e^{n\lambda(\beta)} = \lambda(\beta).$$

Indeed, Jensen’s inequality gives the comparison

$$(1.4) \quad \mathbf{E}(F_n) < \lambda(\beta),$$

where the inequality is strict because F_n is not an almost sure constant, and $x \mapsto \log x$ is not linear. A superadditivity argument (e.g., see [23], proof of Proposition 1.4) shows that $\mathbf{E}(F_n)$ converges to $\sup_{n \geq 0} \mathbf{E}(F_n)$ as $n \rightarrow \infty$. It has also been shown in [85] that under hypotheses much weaker than (1.1), $F_n - \mathbf{E}(F_n)$ tends to 0 almost surely. Therefore, there is a deterministic limit

$$(1.5) \quad p(\beta) := \lim_{n \rightarrow \infty} \mathbf{E}(F_n) = \lim_{n \rightarrow \infty} F_n \quad \text{a.s.}$$

Using the FKG inequality, Comets and Yoshida [33] identified a phase transition:

THEOREM A ([33], Theorem 3.2). *There exists a critical inverse temperature $\beta_c = \beta_c(\mathfrak{L}, d) \in [0, \infty]$ such that*

$$\begin{aligned} 0 \leq \beta \leq \beta_c &\Rightarrow p(\beta) = \lambda(\beta), \\ \beta > \beta_c &\Rightarrow p(\beta) < \lambda(\beta). \end{aligned}$$

We refer to the region $0 \leq \beta < \beta_c$ as the “high temperature phase,” while $\beta > \beta_c$ defines the “low temperature phase.” Roughly speaking, high temperatures reduce the influence of the random environment, and so polymer growth resembles a simple random walk, while at low temperatures the random impurities force a much different behavior. This distinction has been most frequently made in terms of the *endpoint distribution* $\rho_n(\omega_n \in \cdot)$, which is a random probability measure on \mathbb{Z}^d . For instance, one striking result proved by Carmona and Hu [23] (for a Gaussian environment) and then by Comets, Shiga, and Yoshida [28] (in the general case) is the following: If $\beta > \beta_c$, then the polymer endpoint observes so-called *strong localization*:

$$(SL) : \quad \exists c > 0, \quad \limsup_{n \rightarrow \infty} \max_{x \in \mathbb{Z}^d} \rho_n(\omega_n = x) \geq c \quad \text{a.s.}$$

That is, infinitely often the polymer has “favorite sites” at which its endpoint distribution concentrates.

1.2.2. *Weak and strong disorder regimes.* While examining the high and low temperature regimes is very natural, the mathematical development of directed polymers in random environment has often followed an ostensibly different route. Since the work of Bolthausen [14], analysis of the directed polymer model has frequently focused on the *normalized partition function*,

$$\tilde{Z}_n := Z_n e^{-n\lambda(\beta)},$$

with $\tilde{Z}_0 = Z_0 = 1$ and $\lambda(\beta)$ defined as in (1.1). It is not difficult to check that \tilde{Z}_n is a positive martingale adapted to the filtration $(\mathcal{F}_n)_{n \geq 0}$ defined in (1.3). In particular, $\mathbf{E}(\tilde{Z}_n) = 1$ for all n , and the martingale convergence theorem implies that there is an \mathcal{F} -measurable random variable \tilde{Z}_∞ such that

$$(1.6) \quad \lim_{n \rightarrow \infty} \tilde{Z}_n = \tilde{Z}_\infty \quad \text{a.s.}$$

Furthermore, we necessarily have $\tilde{Z}_\infty \geq 0$ almost surely, and the event of positivity $\{\tilde{Z}_\infty > 0\}$ is measurable with respect to the tail σ -algebra,

$$\bigcap_{n=1}^{\infty} \sigma(X_{i,x} : i \geq n, x \in \mathbb{Z}^d).$$

By Kolmogorov’s zero-one law we have either *weak* or *strong disorder*,

$$(WD) : \quad \tilde{Z}_\infty > 0 \quad \text{a.s.}, \quad (SD) : \quad \tilde{Z}_\infty = 0 \quad \text{a.s.}$$

Carmona and Hu [23] and Comets, Shiga, and Yoshida [28] gave the following characterization of the two phases. It says that there is strong disorder exactly when the overlap of two independent polymers has infinite expectation.

THEOREM B ([23], Proposition 5.1 and [28], Theorem 2.1). *Strong disorder (SD) is equivalent to weak localization,*

$$(WL) : \quad \sum_{n=0}^{\infty} \rho_n^{\otimes 2}(\omega_n = \omega'_n) = \infty \quad \text{a.s.},$$

where ω and ω' are independent samples from ρ_n .

Despite the unfortunate clash of terminology in the above theorem, one should not mistake weak localization to be associated with weak disorder. The use of “weak” in the first case is only to distinguish this notion of localization from the other (SL) notion defined before. Indeed, it is apparent from the inequality

$$\rho_n^{\otimes 2}(\omega_n = \omega'_n) \geq \left(\max_{x \in \mathbb{Z}^d} \rho_n(\omega_n = x) \right)^2$$

that (SL) \Rightarrow (WL). In fact, Carmona and Hu [24] proved in a continuous-time model with an environment of i.i.d. Brownian motions—the parabolic Anderson model—that the converse is also true: (SL) \Leftarrow (WL). It is believed that the two notions are equivalent in general.

As with the high and low temperature regimes, there is a phase transition between weak and strong disorder.

THEOREM C ([33], Theorem 12.1). *There exists a critical inverse temperature $\tilde{\beta}_c = \tilde{\beta}_c(\mathcal{L}, d) \in [0, \infty]$ such that*

$$\begin{aligned} 0 \leq \beta < \tilde{\beta}_c &\Rightarrow \text{(WD)}, \\ \beta > \tilde{\beta}_c &\Rightarrow \text{(SD)}. \end{aligned}$$

Determining the behavior of \tilde{Z}_∞ at $\tilde{\beta}_c$ is an open problem. For analogous models on b -ary trees, it is known from work of Kahane and Peyrière [48] that strong disorder occurs at $\tilde{\beta}_c$. Interestingly, in dimensions $d = 1$ and $d = 2$, there is strong disorder at all finite temperatures (i.e., $\tilde{\beta}_c = 0$), while in higher dimensions weak disorder occurs at sufficiently high temperatures ($\tilde{\beta}_c > 0$). Precise conditions on $\lambda(\beta)$ guaranteeing either behavior can be found in Theorem 2.3.2 of the review [29]. These conditions are in fact a culmination of results from [4, 14, 23, 28, 45, 50, 77, 79].

Whereas Theorem B characterizes the disorder regimes in terms of endpoint localization, one can also attempt to give a characterization based on endpoint diffusion. For instance, Comets and Yoshida [33] showed, as part of a more general Brownian central limit theorem, that

$$(1.7) \quad \text{(WD)} \Rightarrow \lim_{n \rightarrow \infty} \int \frac{|\omega_n|}{n^{1/2}} \rho_n(d\omega) = 1 \quad \text{a.s.}$$

Notice that the scaling $|\omega_n| \sim n^{1/2}$ matches that of simple random walk on \mathbb{Z}^d . On the other hand, it is believed (at least in low dimensions, see [68]) that the polymer endpoint is superdiffusive in strong disorder. Specifically, when $d = 1$ (for which any $\beta > 0$ yields strong disorder), it is conjectured that

$$(1.8) \quad \lim_{n \rightarrow \infty} \int \frac{|\omega_n|}{n^{2/3}} \rho_n(d\omega) = 1 \quad \text{a.s.}$$

Apart from the integrable models discussed in Section 1.3, the only result in this direction is an upper bound due to Piza [68] and is conditional on a curvature assumption. In some related models on \mathbb{R} , both upper and lower bounds are known: $n^{3/5} \lesssim |\omega_n| \lesssim n^{3/4}$ [12, 31, 32, 58, 67, 87].

A major challenge is to reconcile the notion of high and low temperature with the notion of weak and strong disorder. The first is more natural from the perspective of statistical physics, because phase transitions, in the standard sense, are defined in terms of nonanalyticities of limiting free energies. Since the annealed free energy is typically analytic everywhere, the transition of the difference $\lambda(\beta) - p(\beta)$ from zero to nonzero denotes the first point of non-analyticity of the limiting free energy $p(\beta)$. On the other hand, the more abstract conditions

of (WD) and (SD) have lead to a wealth of mathematical results. It is conjectured (see [24, 33]) that $\beta_c = \tilde{\beta}_c$, which would mean

$$0 \leq \beta < \beta_c \Rightarrow \text{(WD)} \quad \text{and} \quad \beta > \beta_c \Rightarrow \text{(SD)}.$$

Evidence for this belief includes the result by Comets and Vargas [30] that $\beta_c = 0$ universally in $d = 1$, and the subsequent proof by Lacoïn [52] that $\beta_c = 0$ in $d = 2$. For $d \geq 3$, only the trivial second implication above is known (i.e., $\beta_c \geq \tilde{\beta}_c$). Indeed, it is clear from (1.5) and (1.6) that $p(\beta) < \lambda(\beta) \Rightarrow \text{(SD)}$. Therefore, it has become common to say that in the low temperature phase, there is *very strong disorder*,

$$\text{(VSD)} : \quad p(\beta) < \lambda(\beta).$$

In recent work, Rassoul-Agha, Seppäläinen and Yilmaz [71, 72] expressed $p(\beta)$ in terms of several variational formulas. A method for finding minimizers to these formulas was proposed in [40], and it was noted in [72] that they may not admit minimizers when $\beta > \beta_c$. These results suggest the possibility of a more general correspondence between the disorder regimes and the existence of variational minimizers (see [72], Conjecture 2.13).

When $d = 1$, a so-called “intermediate” disorder regime was discovered in [3] by scaling the inverse temperature to 0 as $n \rightarrow \infty$, in such a way that features of both weak and strong disorder are observed in the limit. More specifically, when $\beta_n = \beta n^{-1/4}$, the polymer measure exhibits diffusivity (1.7), and the fluctuations of $\log Z_n$ are order 1 as in weak disorder. The actual fluctuations, however, are not Gaussian, but rather depend on β and the random environment, interpolating between Gaussian at $\beta = 0$ and Tracy–Widom at $\beta = \infty$.

1.2.3. *Further versions of localization.* Another central task in the theory of directed polymers is determining if and when weak localization results imply stronger ones. The following analog to Theorem B shows that (VSD) \Rightarrow (SL), which means the strongest type of localization occurs throughout the low temperature regime. Therefore, if $\beta_c = \tilde{\beta}_c$, then all notions of disorder and localization are equivalent, except possibly at the critical temperature.

THEOREM D ([28], Corollary 2.2 and Theorem 2.3(a)). *There is very strong disorder (VSD) if and only if there exists $c > 0$ such that*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i^{\otimes 2}(\omega_i = \omega'_i) \geq c \quad a.s.,$$

or, equivalently, there exists $c > 0$ such that

$$(1.9) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \max_{x \in \mathbb{Z}^d} \rho_i(\omega_i = x) \geq c \quad a.s.$$

Strong localization (SL) or the stronger property (1.9) captures the tendency of the endpoint distribution to localize mass when $\beta > \beta_c$. In this scenario, it is natural to ask whether the entire mass localizes, or if some positive proportion of the mass remains delocalized. In [85], Vargas proposed the following definition for complete localization, or as Vargas called it, *asymptotic pure atomicity*. We consider the set of “ ε -atoms,”

$$\mathcal{A}_i^\varepsilon := \{x \in \mathbb{Z}^d : \rho_i(\omega_i = x) > \varepsilon\}, \quad i \geq 0, \varepsilon > 0,$$

and say that the sequence $(\rho_i(\omega_i \in \cdot))_{i \geq 0}$ is *asymptotically purely atomic* if for every sequence $(\varepsilon_i)_{i \geq 0}$ tending to 0 as $i \rightarrow \infty$, we have

$$\frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^{\varepsilon_i}) \rightarrow 1 \quad \text{in probability, as } n \rightarrow \infty.$$

The following result was obtained in [85].

THEOREM E ([85], Theorem 3.2 and Corollary 3.3). *If $\lambda(\beta) = \infty$, then $(\rho_i(\omega_i \in \cdot))_{i \geq 0}$ is asymptotically purely atomic.*

Actually, Vargas showed that $(\rho_{i-1}(\omega_i \in \cdot))_{i \geq 1}$ is asymptotically purely atomic, with the set $\mathcal{A}_i^\varepsilon$ replaced by $\{x \in \mathbb{Z}^d : \rho_{i-1}(\omega_i = x) > \varepsilon\}$. From observation (1.10), it is not difficult to check that the two notions are equivalent (a proof can be found in Appendix A.7). One of the main results of this paper, Theorem 6.3, asserts that the conclusion of Theorem E continues to hold even if $\lambda(\beta)$ is finite, as long as $\beta > \beta_c$ and (1.1) holds. This is discussed in greater detail in Section 1.4 below.

1.2.4. *A remark about the definition of endpoint distribution.* All the results of Section 1.2 are normally stated in the literature using the measure $\rho_{n-1}(\omega_n \in \cdot)$, as opposed to $\rho_n(\omega_n \in \cdot)$. The reason is that the former arises naturally out of the martingale analysis for \tilde{Z}_n (for instance, see the proof of Theorem 3.3.1 in [29]), but in all cases, the statements are equivalent when the latter is used instead. This equivalence can be seen by writing

$$(1.10) \quad \rho_{n-1}(\omega_n = x) = \frac{1}{2d} \sum_{\|x-y\|_1=1} \rho_{n-1}(\omega_{n-1} = y),$$

(where $\|\cdot\|_1$ denotes the ℓ^1 norm on \mathbb{Z}^d), which implies

$$\frac{1}{2d} \max_{y \in \mathbb{Z}^d} \rho_{n-1}(\omega_{n-1} = y) \leq \max_{x \in \mathbb{Z}^d} \rho_{n-1}(\omega_n = x) \leq \max_{y \in \mathbb{Z}^d} \rho_{n-1}(\omega_{n-1} = y).$$

Similarly, we have

$$\frac{1}{2d} \rho_{n-1}^{\otimes 2}(\omega_{n-1} = \omega'_{n-1}) \leq \rho_{n-1}^{\otimes 2}(\omega_n = \omega'_n) \leq \rho_{n-1}^{\otimes 2}(\omega_{n-1} = \omega'_{n-1}),$$

where the first inequality follows from the observation that if $\omega_{n-1} = \omega'_{n-1}$, then $\omega_n = \omega'_n$ with probability $(2d)^{-1}$ under ρ_{n-1} . The second inequality is due to Cauchy–Schwarz applied to (1.10):

$$\begin{aligned} \rho_{n-1}^{\otimes 2}(\omega_n = \omega'_n) &= \sum_{x \in \mathbb{Z}^d} \rho_{n-1}(\omega_n = x)^2 \\ &\leq \sum_{x \in \mathbb{Z}^d} \frac{1}{2d} \sum_{\|x-y\|_1=1} \rho_{n-1}(\omega_{n-1} = y)^2 \\ &= \sum_{y \in \mathbb{Z}^d} \rho_{n-1}(\omega_{n-1} = y)^2 = \rho_{n-1}^{\otimes 2}(\omega_{n-1} = \omega'_{n-1}). \end{aligned}$$

As $\rho_n(\omega_n \in \cdot)$ will be the more natural object for our purposes, we reserve the term “endpoint distribution” for this measure.

1.3. *Related results in integrable models.* Until recently, no directed polymer model offered the possibility of exact asymptotic calculations. This was in contrast to the integrable last passage percolation (LPP) models in $1 + 1$ dimensions, for which specific choices of the passage time distribution (namely geometric or exponential) allow one to use representation theory to derive exact formulas for the distribution of passage times, which are the LPP analogs of partition functions. Such was the approach advanced in the seminal work of Johansson [46], which showed that asymptotic fluctuations of passage times follow the Tracy–Widom distributions at scale $n^{1/3}$ —GUE for point-to-point passage times, and GOE for the point-to-line passage time. The connection to Tracy–Widom laws can be seen more

explicitly in a model known as Brownian LPP [63], which also exhibits the $n^{1/3}$ scaling. These results place the LPP models within the KPZ universality class (see [19, 34]). For a review of related works, including results verifying spatial fluctuations at scale $n^{2/3}$, we refer the reader to [69] and references therein.

In the zero-temperature limit $\beta \rightarrow \infty$, the polymer measure ρ_n concentrates on the path that is most likely given the random environment. When $\beta = \infty$, the directed polymer model is equivalent to LPP. The natural conjecture, therefore, is that directed polymers in strong disorder obey the same KPZ scaling relations. In particular, models in $1 + 1$ dimensions should have energy fluctuations of order $n^{1/3}$, and endpoint fluctuations of order $n^{2/3}$ (see (1.8)). The first case permitting exact calculations was the integrable log-gamma model introduced by Seppäläinen [75], whose breakthrough work proved the conjectured exponents. Subsequent studies showed that free energy fluctuations were again of Tracy–Widom type [18, 35], gave a large deviation rate function [42], and computed the limiting value $p(\beta)$ [41].

Regarding spatial fluctuations, Comets and Nguyen [27] found an explicit limiting endpoint distribution in the equilibrium case of the log-gamma model. More precisely, they showed that the endpoint distribution, when recentered around the most likely site, converges in law to a certain random distribution on \mathbb{Z} . While the boundary conditions they enforce prevent the correct exponent of $2/3$ from appearing, which would mean fluctuations of the averaged quenched distribution occur on the order of $n^{2/3}$, their result implies that the *quenched* endpoint distributions concentrate all their mass in a microscopic region of $O(1)$ diameter around a single favorite site. In the notation of this paper, if $x_n = \arg \max \rho_n(\omega_n = \cdot)$, then the result of [27] says

$$(1.11) \quad \lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \rho_n(\|\omega_n - x_n\|_1 \leq K) = 1 \quad \text{in probability.}$$

This provides an affirmative case of the so-called “favorite region” conjecture, which speculates that in strong disorder, directed polymers on the lattice localize their endpoint to a region of fixed diameter. We emphasize once more that this is a statement about quenched endpoint distributions. While little is known on annealed distributions in the general case, there has been considerable progress for the integrable LPP models. Johansson [47] proved that the rescaled location of the endpoint converges in distribution to the maximal point of an Airy_2 process minus a parabola. More recent work has expressed the density of this limit in terms of Fredholm determinants [59], or from a different approach, in terms of the Hastings–McLeod solution to the Painlevé II equation [74]; the two formulas were shown to be equivalent in [7]. Tail estimates for this density can be found in [20, 70].

For further literature on integrable directed polymers, see [10, 36, 83] and references therein. We also mention the semidiscrete model introduced by O’Connell and Yor [64], which is a positive-temperature version of Brownian LPP and has offered another provable case of $n^{1/3}$ energy fluctuations [16, 17, 76].

1.4. *An overview of results proved in this paper.* The endpoint distribution $\rho_n(\omega_n \in \cdot)$ of length- n polymers is a random probability measure on \mathbb{Z}^d . The main goal of this paper is to understand the behavior of this object as $n \rightarrow \infty$. The majority of our work is in developing methods to compute limits of endpoint distributions, or more to the point, of *functionals* of the endpoint distribution. While proving the existence of these limits is presently out of reach, by instead considering empirical averages we are able to establish results in Cesàro form. Broadly speaking, the construction consists of three components: (i) a compactification of measures on \mathbb{Z}^d ; (ii) a Markov kernel whose invariant measures are the possible limits of the endpoint distribution; and (iii) a functional that distinguishes those invariant measures with minimal energy.

1.4.1. *Compactification of measures on \mathbb{Z}^d .* One key obstacle to studying the endpoint distribution is that the standard topology of weak convergence of probability measures is inadequate for understanding its limiting behavior. Even the weaker topology of vague convergence does not provide an adequate description of the localization phenomenon that is of central interest. In the recent work of Mukherjee and Varadhan [61], this issue was addressed by exploiting translation invariance inherent in their problem to pass to a certain compactification of probability measures on \mathbb{R}^d . We have drawn inspiration from their methods, using a similar construction we will soon describe. Ultimately, this approach is enabled by the “concentration compactness” phenomenon. Before we discuss this topic further, let us motivate the discussion with two elementary examples.

Roughly speaking, the difficulty of using standard weak or vague convergence is that the endpoint distribution may manifest itself as multiple localized “blobs” that escape to infinity in different directions as $n \rightarrow \infty$. Additionally, some part of the mass may just “diffuse to zero.” As a first example, consider the following sequence of probability mass functions on \mathbb{Z} :

$$q_n(x) := \begin{cases} 1/2 & \text{if } x = n, \\ 1/(2n) & \text{if } 0 \leq x < n, \\ 0 & \text{else.} \end{cases}$$

This sequence fails to converge weakly, and its vague limit is the zero measure. However, to understand the true limiting behavior of q_n , it seems more appropriate to take the vague limit after translating the measure by n ; that is, taking $\tilde{q}_n(x) := q_n(x - n)$. The limit of the sequence \tilde{q}_n is the subprobability measure that puts mass $1/2$ at the origin and zero elsewhere, which gives a better picture of the true limiting behavior of q_n .

The situation is more complicated when multiple blobs escape to infinity in different directions. For example, consider the following sequence of probability mass functions on \mathbb{Z} :

$$r_n(x) := \begin{cases} 1/5 & \text{if } x \in \{-2n, -n, n, n+1\}, \\ 1/(5n) & \text{if } 0 \leq x < n, \\ 0 & \text{else.} \end{cases}$$

Like q_n , the sequence r_n also converges vaguely to the subprobability measure of mass zero. But in this case, no sequence of translates of r_n can fully capture its limiting behavior, because a translate can only capture the mass at one of the three blobs but not all of them simultaneously. The only recourse, it seems, is to express the limit not as a single subprobability measure, but a union of three subprobability measures on three distinct copies of \mathbb{Z} . Two of these will put mass $1/5$ at the origin in their respective copies of \mathbb{Z} , and the third will put mass $1/5$ at 0 and $1/5$ at 1 in its copy of \mathbb{Z} . The first two are the vague limits of $r_n(\cdot + 2n)$ and $r_n(\cdot + n)$, and the third is the vague limit of $r_n(\cdot - n)$. One can view this object jointly as a subprobability measure of total mass $4/5$ on the set $\{1, 2, 3\} \times \mathbb{Z}$. Of course, it is not important which copy of \mathbb{Z} gets which part of the measure, nor does it matter if a translation is applied to the subprobability measure on one of the copies. Thus, the limit object is an equivalence class of subprobability measures on $\{1, 2, 3\} \times \mathbb{Z}$ rather than a single subprobability measure.

Generalizing the above idea, we can consider equivalence classes of subprobability measures on $\mathbb{N} \times \mathbb{Z}^d$. First, define the \mathbb{N} -support of a subprobability measure f on $\mathbb{N} \times \mathbb{Z}^d$ to be the set of all $n \in \mathbb{N}$ such that f puts nonzero mass on $\{n\} \times \mathbb{Z}^d$. Next, declare two subprobability measures f and g on $\mathbb{N} \times \mathbb{Z}^d$ to be equivalent if there is a bijection σ between their \mathbb{N} -supports and a number $x_n \in \mathbb{Z}^d$ for each n in the \mathbb{N} -support of f such that

$$g(\sigma(n), \cdot) = f(n, x_n + \cdot).$$

It is easy to verify that this is indeed an equivalence relation. The equivalence classes are called *partitioned subprobability measures* on \mathbb{Z}^d in this paper, and the set of all equivalence classes is denoted by \mathcal{S} . The set \mathcal{S} is formally defined and studied in Section 2.1.

We will view probability measures on \mathbb{Z}^d (or more precisely, their mass functions) as elements of \mathcal{S} supported on a single copy of \mathbb{Z}^d . In particular, we have a natural embedding of endpoint distributions into \mathcal{S} . In Section 2.1, we define a metric d on \mathcal{S} (not to be confused with the dimension d). Due to the somewhat complicated nature of the metric, we will not reproduce the definition in this overview section, but it is constructed to ensure that certain functionals on \mathcal{S} are continuous or at least semicontinuous. In this study, these functionals are related to either localization or the free energy of the system. One of the first nontrivial results of this paper is that (\mathcal{S}, d) is a compact metric space, in analogy with the construction in [61]. This is Theorem 2.8 in Section 2.2.

REMARK. A formal comparison between our construction and the one of Mukherjee and Varadhan [61] is provided in Appendix B.

In summary, the metric d imposes a translation invariance that prevents so-called blobs from escaping to infinity. Moreover, such escape turns out to be the only obstruction to compactness. This is the fundamental observation of a more general theory known as “concentration compactness,” which builds on the notion of the concentration functions of probability measures, defined by Lévy [53]. The idea of using Lévy’s concentration functions for constructing compactifications of function spaces was introduced by Parthasarathy, Ranga Rao, and Varadhan [66], and developed into a powerful tool by Lions [54–57]. It has been applied broadly in the study of calculus of variations and PDEs. In the language of those settings, the equivalence relation described above plays the role of a “profile decomposition” (for a friendly discussion, see [81, 82]).

1.4.2. *The update map.* In Section 3, we define an “update” map $T : \mathcal{S} \rightarrow \mathcal{P}(\mathcal{S})$, where $\mathcal{P}(\mathcal{S})$ is the set of all probability measures on \mathcal{S} equipped with the Kantorovich–Rubinstein–Wasserstein distance \mathcal{W} . The map T has the property that if $f_n(\cdot) := \rho_n(\omega_n = \cdot)$ is the endpoint mass function of the length- n polymers (considered as an element of \mathcal{S} supported on a single copy of \mathbb{Z}^d), then Tf_n is the law of f_{n+1} given \mathcal{F}_n (recall the definition (1.3) of \mathcal{F}_n). In fact, it is not difficult to see that $(f_n)_{n \geq 0}$ is Markovian, and thus T is the Markov kernel generating this chain. Considerable work is done in Section 8 to show that T is a continuous map (i.e., has the Feller property). This is the conclusion of Proposition 3.2. The explicit nature of the metric d is particularly important in the proof of this result. Finally, the map T lifts to a map $\mathcal{T} : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{P}(\mathcal{S})$, defined as

$$\mathcal{T}v(dg) := \int Tf(dg)v(df).$$

The continuity of T implies that \mathcal{T} is also continuous.

In Section 4, we study the following random element of $\mathcal{P}(\mathcal{S})$:

$$\mu_n := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{f_i}.$$

Here δ_{f_i} is the unit point mass at the i th endpoint mass function f_i , considered as an element of \mathcal{S} as before. In words, μ_n is the empirical measure of the endpoint distributions up to time n . Let

$$\mathcal{K} := \{v \in \mathcal{P}(\mathcal{S}) : \mathcal{T}v = v\}$$

be the set of fixed points of $\mathcal{T} : \mathcal{P}(\mathcal{S}) \rightarrow \mathcal{P}(\mathcal{S})$. The first main result of Section 4 is Corollary 4.3, which says that

$$(1.12) \quad \lim_{n \rightarrow \infty} \inf_{\nu \in \mathcal{K}} \mathcal{W}(\mu_n, \nu) = 0 \quad \text{a.s.}$$

This result provides a heuristic connection to $(1 + 1)$ -dimensional integrable models (for instance, see [10, 36, 75, 83]), which work in part by identifying a disorder distribution and boundary conditions such that the system is stationary under spatial translations. Loosely speaking, our approach similarly recovers a stationarity property *in the limit*, even without explicit calculations. In this way, our methods replace this key feature of integrable models with a much weaker, but more general, abstract framework. This is enabled by a topology on endpoint distributions that is rich enough to capture the desired localization, yet sufficiently “compressed” to be compact.

1.4.3. The energy functional. Given the convergence (1.12), the next key observation is that the free energy F_n can be expressed in terms of the empirical measure μ_n , see (4.7). We thus define a functional $\mathcal{R} : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ so that we may concisely write $\mathbf{E}(F_n) = \mathbf{E}\mathcal{R}(\mu_n)$. This and Corollary 4.3 lead to the following variational formula for the limiting free energy, given in Theorem 4.7. For any β such that (1.1) holds,

$$p(\beta) = \lim_{n \rightarrow \infty} \mathbf{E}(F_n) = \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu).$$

The connections between this formula and those of [71, 72] are unclear.

Nevertheless, this computation allows us to improve Corollary 4.3 in an important way to yield Theorem 4.9, which says that if

$$\mathcal{M} := \left\{ \nu_0 \in \mathcal{K} : \mathcal{R}(\nu_0) = \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu) \right\},$$

then

$$\lim_{n \rightarrow \infty} \inf_{\nu \in \mathcal{M}} \mathcal{W}(\mu_n, \nu) = 0 \quad \text{a.s.}$$

In Section 5, we study the minimizing set \mathcal{M} . In particular, Theorem 5.2 says that either \mathcal{M} consists of the single element of total mass zero (which happens when $0 \leq \beta \leq \beta_c$, where β_c is the critical inverse temperature of Theorem A), or every element of \mathcal{M} has total mass one (which happens when $\beta > \beta_c$). This result is similar to the technique of identifying a phase transition as the critical point at which a recursive distributional equation begins to have a nontrivial solution; for an account of this method, we refer the reader to the survey of Aldous and Bandyopadhyay [5] and references therein. This idea is also present in work of Yoshida [88] on more general linear stochastic evolutions, although there the nontrivial solutions exist in the high temperature regime rather than the low temperature regime.

From a different perspective, the limit law of the empirical measure can be viewed as an “order parameter” for the model, whose behavior distinguishes between the high and low temperature regimes. Such order parameters arise frequently in the study of disordered systems. A prominent example is the Sherrington–Kirkpatrick (SK) model of spin glasses, where the limiting distribution of the overlap serves as the order parameter (see Panchenko [65]). Interestingly, the limiting free energy of the SK model can also be expressed as the solution of a variational problem involving the order parameter. This is the famous Parisi formula proved by Talagrand [80]. In this way, the partitioned subprobability measures might be seen as counterparts to the random overlap structures introduced by Aizenman, Sims and Starr [2], and the update map as the analog to similar stabilizing maps arising out of the cavity method for spin glasses and related competing particle systems, that were studied by Aizenman and Contucci [1], Ruzmaikina and Aizenman [73] and Arguin and Chatterjee [6]. In other ways,

however, the analogy is quite distant. For instance, the spin glass systems we speak of are mean-field models lacking any geometry from the lattice. Also, our variational formula relies on the directed nature of the problem; that is, the random environment refreshes at each time step, allowing us to exploit Markovian structure.

1.5. *Main applications.* Theorem 5.2 yields the following concrete application of our abstract theory of partitioned subprobability measures; it is later stated as Theorem 6.3. Recall the notation and terminology related to Theorem A and Theorem E.

THEOREM 1.1. *Assume (1.1).*

(a) *If $\beta > \beta_c$, then for every sequence $(\varepsilon_i)_{i \geq 0}$ tending to 0 as $i \rightarrow \infty$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^{\varepsilon_i}) = 1 \quad \text{a.s.}$$

(b) *If $0 \leq \beta \leq \beta_c$, then there exists a sequence $(\varepsilon_i)_{i \geq 0}$ tending to 0 as $i \rightarrow \infty$ such that*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^{\varepsilon_i}) = 0 \quad \text{a.s.}$$

This generalizes Theorem E, where an “in probability” version of (a) was proved under the condition $\lambda(\beta) = \infty$.

In Section 7, we apply our techniques to go further than atomic localization by considering “geometric localization.” In the low temperature phase, the endpoint distribution can not only concentrate mass on a few likely sites, but moreover have those sites close together. We make this phenomenon precise in the following manner. For $\delta \in (0, 1)$ and a nonnegative number K , let $\mathcal{G}_{\delta, K}$ denote the set of probability mass functions on \mathbb{Z}^d that assign measure greater than $1 - \delta$ to some subset of \mathbb{Z}^d having diameter at most K ; see (7.1) for a symbolic definition. We will say that the sequence $(\rho_i(\omega_i \in \cdot))_{i \geq 0}$ is *geometrically localized with positive density* if for every δ , there is $K < \infty$ and $\theta > 0$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{\rho_i(\omega_i = \cdot) \in \mathcal{G}_{\delta, K}\}} \geq \theta \quad \text{a.s.,}$$

where $\mathbb{1}_A$ denotes the indicator of the event A . That is, there are endpoint distributions with limiting density at least θ that place mass greater than $1 - \delta$ on a set of bounded diameter. We will say $(\rho_i(\omega_i \in \cdot))_{i \geq 0}$ is *geometrically localized with full density* if for every δ , there is $K < \infty$ such that

$$(1.13) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{\rho_i(\omega_i = \cdot) \in \mathcal{G}_{\delta, K}\}} \geq 1 - \delta \quad \text{a.s.}$$

The main result of Section 7 is contained in Theorem 7.3 and says the following.

THEOREM 1.2. *Assume (1.1).*

(a) *If $\beta > \beta_c$, then there is geometric localization with positive density. Moreover, the numbers K and θ are deterministic quantities that depend only on the choice of δ , the disorder distribution \mathfrak{L} , the parameter β , and the dimension d .*

(b) *If $0 \leq \beta \leq \beta_c$, then for any K and any $\delta \in (0, 1)$,*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{\rho_i(\omega_i = \cdot) \in \mathcal{G}_{\delta, K}\}} = 0 \quad \text{a.s.}$$

As mentioned in Section 1.3, the only case where a version of geometric localization has been proved is an integrable $(1 + 1)$ -dimensional model, for which Comets and Nguyen [27] proved localization and moreover computed the limit distribution of the endpoint. Similar results for one-dimensional random walk in random environment were proved by Sinai [78], Golosov [43] and Gantert, Peres and Shi [39].

The single-copy condition. In addition to the above unconditional results, we also prove a few conditional statements, which hold under the condition that every $\nu \in \mathcal{M}$ puts all its mass on those $f \in \mathcal{S}$ that are supported on a single copy of \mathbb{Z}^d . We call this the “single-copy condition.”

One consequence of the single-copy condition is geometric localization with full density, as defined in equation (1.13). Part (b) of Theorem 7.3 proves this conditional claim. A second consequence of the single-copy condition is Proposition 7.4, which gives the following Cesàro form of (1.11). For each $i \geq 0$ and $K \geq 0$, let \mathcal{C}_i^K be the set of all $x \in \mathbb{Z}^d$ that are at distance $\leq K$ from every mode of the endpoint mass function $\rho_i(\omega_i = \cdot)$. Then, assuming (1.1) and the single-copy condition, Proposition 7.4 asserts that

$$\lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{C}_i^K) = 1 \quad \text{a.s.}$$

In view of the result (1.11) of Comets and Nguyen [27], it seems plausible that the single-copy condition holds for the log-gamma polymer in $1 + 1$ dimensions. Unfortunately, we have been unable to determine whether or not the single-copy condition holds in general. Furthermore, we are not aware of any conjectures on what is true in higher dimensions. The results of [9] suggest that at least for directed polymers on b -ary trees, the single-copy condition does not hold, and full geometric localization is not valid. This may be related to the fluctuations of $\log Z_n$, which are known to be order 1 on the tree (see [38]) but conjectured to be order $n^{1/3}$ when $d = 1$.

2. Partitioned subprobability measures. In this section and in the remainder of this manuscript, we will always assume (1.1). Also, throughout, $f_i(\cdot)$ will denote the endpoint probability mass function $\rho_i(\omega_i = \cdot)$. The goals of this section are to (i) define a space of functions which contains endpoint distributions (i.e., probability measures on \mathbb{Z}^d) and their localization limits (subprobability measures on $\mathbb{N} \times \mathbb{Z}^d$); (ii) equip said space with a metric; and (iii) prove that the induced metric topology is compact.

2.1. *Definition and properties.* Let us restrict our attention to the set of functions

$$S := \{f : \mathbb{N} \times \mathbb{Z}^d \rightarrow \mathbb{R} : f \geq 0, \|f\| \leq 1\},$$

where

$$(2.1) \quad \|f\| := \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} f(u).$$

Since we regard distant point masses as nearly existing on separate copies of \mathbb{Z}^d , we consider differences between elements of $\mathbb{N} \times \mathbb{Z}^d$ in the following nonstandard way:

$$(n, x) - (m, y) := \begin{cases} x - y & \text{if } n = m, \\ \infty & \text{else.} \end{cases}$$

It then makes sense to write, for $u = (n, x)$ and $v = (m, y)$ in $\mathbb{N} \times \mathbb{Z}^d$,

$$\|u - v\|_1 := \begin{cases} \|x - y\|_1 & \text{if } n = m, \\ \infty & \text{else,} \end{cases}$$

although $\| \cdot \|_1$ is not to be thought of as a norm. If $u = (n, x)$ and $y \in \mathbb{Z}^d$, then $u \pm y$ will be understood to mean $(n, x \pm y)$. Our main tool for enabling compactness will be certain injective functions on $\mathbb{N} \times \mathbb{Z}^d$. Given a set $A \subset \mathbb{N} \times \mathbb{Z}^d$, we call a map $\phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d$ an *isometry* of degree $m \geq 1$ if for all $u, v \in A$,

$$(2.2) \quad \begin{aligned} \|u - v\|_1 < m \quad \text{or} \quad \|\phi(u) - \phi(v)\|_1 < m \\ \Rightarrow \quad \phi(u) - \phi(v) = u - v. \end{aligned}$$

The maximum m for which (2.2) holds is called the *maximum degree* of ϕ , denoted by $\text{deg}(\phi)$. To say that ϕ is an isometry of degree 1 simply means ϕ is injective. If (2.2) holds for every $m \in \mathbb{N}$, then $\text{deg}(\phi) = \infty$, meaning ϕ acts by translations. That is, each copy of \mathbb{Z}^d intersecting the domain A is translated and moved to some copy of \mathbb{Z}^d in the range, with no two copies in the domain going to the same copy in the range. Note that an isometry is necessarily injective and thus has an inverse defined on its image. Because the hypothesis in (2.2) is symmetric between domain and range, it is clear that $\text{deg}(\phi) = \text{deg}(\phi^{-1})$.

Another useful property of isometries is closure under composition, which is defined in the next lemma. It is important, especially for the proof below, to note that an isometry is permitted to have empty domain. That is, there exists the empty isometry $\phi : \emptyset \rightarrow \mathbb{N} \times \mathbb{Z}^d$, which is its own inverse and has $\text{deg}(\phi) = \infty$.

LEMMA 2.1. *Let $\phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d$ and $\psi : B \rightarrow \mathbb{N} \times \mathbb{Z}^d$ be isometries. Define $A' := \{a \in A : \phi(a) \in B\}$. Then $\theta : A' \rightarrow \mathbb{N} \times \mathbb{Z}^d$ defined by $\theta(u) = \psi(\phi(u))$ is an isometry with $\text{deg}(\theta) \geq \min(\text{deg}(\phi), \text{deg}(\psi))$.*

PROOF. Let $m := \min(\text{deg}(\phi), \text{deg}(\psi))$ so that ϕ and ψ are each isometries of degree m . For any $a_1, a_2 \in A'$,

$$\begin{aligned} \|a_1 - a_2\|_1 < m &\Rightarrow \phi(a_1) - \phi(a_2) = a_1 - a_2 \\ &\Rightarrow \|\phi(a_1) - \phi(a_2)\|_1 < m \\ &\Rightarrow \psi(\phi(a_1)) - \psi(\phi(a_2)) = \phi(a_1) - \phi(a_2) = a_1 - a_2, \end{aligned}$$

as well as

$$\begin{aligned} \|\theta(a_1) - \theta(a_2)\|_1 < m &\Rightarrow \phi(a_1) - \phi(a_2) = \psi(\phi(a_1)) - \psi(\phi(a_2)) \\ &\Rightarrow \|\phi(a_1) - \phi(a_2)\|_1 < m \\ &\Rightarrow a_1 - a_2 = \phi(a_1) - \phi(a_2) \\ &\quad = \psi(\phi(a_1)) - \psi(\phi(a_2)). \end{aligned}$$

Indeed, θ is an isometry of degree m . \square

A final observation about isometries is that they obey a certain extension property, which is proved below: By allowing the maximum degree of an isometry to be lowered by at most 2, we can expand its domain by one unit in every direction. If the maximum degree is infinite (i.e., ϕ acts by translations), then the extension can be repeated *ad infinitum* to recover the translation on all of \mathbb{Z}^d , for any copy of \mathbb{Z}^d intersecting the domain.

LEMMA 2.2. *Suppose that $\phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d$ is an isometry of degree $m \geq 3$. Then ϕ can be extended to an isometry $\Phi : B \rightarrow \mathbb{N} \times \mathbb{Z}^d$ of degree $m - 2$, where*

$$B := \{v \in \mathbb{N} \times \mathbb{Z}^d : \|u - v\|_1 \leq 1 \text{ for some } u \in A\} \supset A.$$

PROOF. For $v \in B$ such that $\|u - v\|_1 \leq 1$ with $u \in A$, define

$$\Phi(v) := \phi(u) + (v - u).$$

If $u' \in A$ also satisfies $\|u' - v\|_1 \leq 1$, then

$$\begin{aligned} \|u - u'\|_1 \leq 2 &\Rightarrow \phi(u) - \phi(u') = u - u' \\ &\Rightarrow \phi(u) + (v - u) = \phi(u') + (v - u'). \end{aligned}$$

So Φ is well defined; in particular, $\Phi(u) = \phi(u)$ for all $u \in A$. To see that Φ is an isometry of degree $m - 2$, consider any $v, v' \in B$ and take $u, u' \in A$ such that $\|u - v\|_1 \leq 1$ and $\|u' - v'\|_1 \leq 1$. If $\|v - v'\|_1 < m - 2$, then

$$\begin{aligned} \|u - u'\|_1 < m &\Rightarrow \phi(u) - \phi(u') = u - u' \\ &\Rightarrow \Phi(v) - \Phi(v') = \phi(u) + (v - u) - \phi(u') - (v' - u') \\ &= v - v'. \end{aligned}$$

Alternatively, if $\|\Phi(v) - \Phi(v')\|_1 < m - 2$, then

$$\begin{aligned} \|\phi(u) - \phi(u')\| < m &\Rightarrow u - u' = \phi(u) - \phi(u') \\ &= \Phi(v) - (v - u) - \Phi(v') + (v' - u') \\ &\Rightarrow v - v' = \Phi(v) - \Phi(v'). \end{aligned}$$

Indeed, $\text{deg}(\Phi) \geq m - 2$. \square

We can now define the desired metric on functions. Roughly speaking, an isometry allows for the comparison of the large values of two functions. The size of the isometry’s domain reflects how many of their large values are similar, while the degree of the isometry captures how similar their relative positioning is. The metric is constructed in stages.

First, given an isometry $\phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d$ with finite (for measurability reasons, see Lemma 2.11) domain $A \subset \mathbb{N} \times \mathbb{Z}^d$ and two functions $f, g \in S$, define

$$\begin{aligned} (2.3) \quad d_\phi(f, g) &:= 2 \sum_{u \in A} |f(u) - g(\phi(u))| + \sum_{u \notin A} f(u)^2 \\ &\quad + \sum_{u \notin \phi(A)} g(u)^2 + 2^{-\text{deg}(\phi)}. \end{aligned}$$

Next define

$$d(f, g) := \inf_{\phi} d_\phi(f, g),$$

where the infimum is taken over all isometries ϕ with finite domain. Since $\text{deg}(\phi^{-1}) = \text{deg}(\phi)$, it is easy to see that $d_{\phi^{-1}}(g, f) = d_\phi(f, g)$, and so the function d is symmetric:

$$(2.4) \quad d(f, g) = d(g, f) \quad \text{for all } f, g \in S.$$

In fact, d also satisfies the triangle inequality on S .

LEMMA 2.3. For any $f, g, h \in S$,

$$(2.5) \quad d(f, h) \leq d(f, g) + d(g, h).$$

PROOF. Fix $\varepsilon > 0$, and choose isometries $\phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d$ and $\psi : B \rightarrow \mathbb{N} \times \mathbb{Z}^d$ such that

$$d_\phi(f, g) < d(f, g) + \varepsilon \quad \text{and} \quad d_\psi(g, h) < d(g, h) + \varepsilon.$$

Define $\theta : A' \rightarrow \mathbb{N} \times \mathbb{Z}^d$ as in Lemma 2.1. We have

$$(2.6) \quad d_\theta(f, h) = 2 \sum_{u \in A'} |f(u) - h(\theta(u))| + \sum_{u \notin A'} f(u)^2 + \sum_{u \notin \theta(A')} h(u)^2 + 2^{-\deg(\theta)}.$$

The first sum above can be bounded as

$$(2.7) \quad \begin{aligned} \sum_{u \in A'} |f(u) - h(\theta(u))| &\leq \sum_{u \in A'} (|f(u) - g(\phi(u))| + |g(\phi(u)) - h(\psi(\phi(u)))|) \\ &= \sum_{u \in A'} |f(u) - g(\phi(u))| + \sum_{u \in B \cap \phi(A)} |g(u) - h(\psi(u))|. \end{aligned}$$

Recall that f, g and h take values in $[0, 1]$. Now, if $f(u) < g(v)$, then one trivially has $f(u)^2 \leq g(v)^2 \leq g(v)^2 + 2|f(u) - g(v)|$. On the other hand, if $f(u) \geq g(v)$, then one again has

$$\begin{aligned} f(u)^2 - g(v)^2 &= (f(u) + g(v))(f(u) - g(v)) \\ &\leq 2(f(u) - g(v)) \\ \Rightarrow f(u)^2 &\leq 2|f(u) - g(v)| + g(v)^2. \end{aligned}$$

As a result, the second sum in (2.6) satisfies

$$(2.8) \quad \begin{aligned} \sum_{u \notin A'} f(u)^2 &= \sum_{u \in A \setminus A'} f(u)^2 + \sum_{u \notin A} f(u)^2 \\ &\leq \sum_{u \in A \setminus A'} (2|f(u) - g(\phi(u))| + |g(\phi(u))|^2) + \sum_{u \notin A} f(u)^2 \\ &\leq 2 \sum_{u \in A \setminus A'} |f(u) - g(\phi(u))| + \sum_{u \notin B} g(u)^2 + \sum_{u \notin A} f(u)^2. \end{aligned}$$

Similarly, the third sum satisfies

$$(2.9) \quad \begin{aligned} \sum_{u \notin \theta(A')} h(u)^2 &= \sum_{u \in \psi(B) \setminus \theta(A')} h(u)^2 + \sum_{u \notin \psi(B)} h(u)^2 \\ &\leq \sum_{u \in B \setminus \phi(A)} (2|h(\psi(u)) - g(u)| + g(u)^2) + \sum_{u \notin \psi(B)} h(u)^2 \\ &\leq 2 \sum_{u \in B \setminus \phi(A)} |g(u) - h(\psi(u))| + \sum_{u \notin \phi(A)} g(u)^2 + \sum_{u \notin \psi(B)} h(u)^2. \end{aligned}$$

Finally, Lemma 2.1 guarantees

$$(2.10) \quad \begin{aligned} \deg(\theta) &\geq \min(\deg(\phi), \deg(\psi)) \\ \Rightarrow 2^{-\deg(\theta)} &\leq 2^{-\deg(\phi)} + 2^{-\deg(\psi)}. \end{aligned}$$

Using (2.7)–(2.10) in (2.6), we find

$$d(f, h) \leq d_\theta(f, h) \leq d_\phi(f, g) + d_\psi(g, h) < d(f, g) + d(g, h) + 2\varepsilon.$$

As ε is arbitrary, (2.5) follows. \square

From (2.4) and (2.5), we see that d is a pseudometric on S . It does not, however, separate points. Nevertheless, one can construct a metric space (\mathcal{S}, d) by taking the quotient of S with respect to the equivalence relation

$$f \equiv g \iff d(f, g) = 0.$$

We shall write without confusion the symbol f for both the equivalence class in \mathcal{S} and the representative chosen from S . When f is evaluated at a particular element $u \in \mathbb{N} \times \mathbb{Z}^d$, an explicit representative has been chosen. In the sequel, it will be important that certain properties of elements of S are invariant within the equivalence classes and thus well defined in \mathcal{S} . To identify the equivalence classes, we have the following result.

PROPOSITION 2.4. *Two functions $f, g \in S$ satisfy $d(f, g) = 0$ if and only if there is a set $B \subset \mathbb{N} \times \mathbb{Z}^d$ and a map $\psi : B \rightarrow \mathbb{N} \times \mathbb{Z}^d$ such that:*

- (i) $f(u) = g(\psi(u))$ for all $u \in B$,
- (ii) $f(u) = 0$ for all $u \notin B$,
- (iii) $g(u) = 0$ for all $u \notin \psi(B)$ and
- (iv) $\psi(u) - \psi(v) = u - v$ for all $u, v \in B$.

The proof is purely technical and thus placed in Appendix A.1. A more transparent description of equivalence under d was given in the introductory Section 1.4. We restate it here and prove equivalence to the conditions of Proposition 2.4. Recall that the \mathbb{N} -support of $f \in S$ is the set

$$H_f := \{n \in \mathbb{N} : f(n, x) > 0 \text{ for some } x \in \mathbb{Z}^d\}.$$

COROLLARY 2.5. *Let $f, g \in S$ have \mathbb{N} -supports denoted by H_f and H_g , respectively. Then $d(f, g) = 0$ if and only if there is a bijection $\sigma : H_f \rightarrow H_g$ and vectors $(x_n)_{n \in H_f} \subset \mathbb{Z}^d$ such that*

$$(2.11) \quad g(\sigma(n), x) = f(n, x + x_n) \quad \text{for all } n \in H_f, x \in \mathbb{Z}^d.$$

PROOF. First assume $d(f, g) = 0$, and take $\psi : B \rightarrow \mathbb{N} \times \mathbb{Z}^d$ as in Proposition 2.4 so that properties (i)–(iv) hold. By repeatedly applying Lemma 2.2, one may assume that the domain B is a union of copies of \mathbb{Z}^d ; that is, $B = H \times \mathbb{Z}^d$. By property (ii), we can take $H = H_f$, while property (iv) shows first that for every $n \in H_f$, there is $\sigma(n) \in \mathbb{N}$ and $x_n \in \mathbb{Z}^d$ satisfying

$$\psi(n, x) = (\sigma(n), x - x_n) \quad \text{for all } x \in \mathbb{Z}^d,$$

and second that $n \mapsto \sigma(n)$ is injective. Then (i) leads to (2.11), while (iii) guarantees that $\sigma(H_f) = H_g$. Conversely, assume $\sigma : H_f \rightarrow H_g$ and $(x_n)_{n \in H_f}$ satisfy (2.11). Then define $\psi : H_f \times \mathbb{Z}^d \rightarrow H_g \times \mathbb{Z}^d$ by $\psi(n, x) := (\sigma(n), x - x_n)$. Since the domain of ψ is $H_f \times \mathbb{Z}^d$, and its range is $H_g \times \mathbb{Z}^d$, this map satisfies (ii) and (iii). At the same time, (iv) holds by construction, and (i) follows from (2.11). \square

We can now state a clear geometric condition for a function on S to be well defined on \mathcal{S} . For the functions that will be of interest to us, it will be obvious that the hypotheses of the following corollary are satisfied.

COROLLARY 2.6. *Suppose L is a function on S satisfying the following:*

- (i) L is invariant under shifts of \mathbb{Z}^d : If there are vectors $(x_n)_{n \in \mathbb{N}}$ in \mathbb{Z}^d such that $f(n, x) = g(n, x - x_n)$, then $L(f) = L(g)$.

(ii) L is invariant under permutations of \mathbb{N} : If there is a bijection $\sigma : \mathbb{N} \rightarrow \mathbb{N}$ such that $f(n, x) = g(\sigma(n), x)$, then $L(f) = L(g)$.

(iii) L is invariant under zero-padding: If there is an increasing sequence $(n_k)_{k \geq 1}$ in \mathbb{N} such that

$$f(n, x) = \begin{cases} g(k, x) & \text{if } n = n_k, \\ 0 & \text{else,} \end{cases}$$

then $L(f) = L(g)$.

Then L is well defined on \mathcal{S} by evaluating at any representative in S .

PROOF. Suppose $f, g \in S$ are such that $d(f, g) = 0$. We wish to show that $L(f) = L(g)$. Let H_f and H_g be the \mathbb{N} -supports of f and g respectively, and let $\sigma : H_f \rightarrow H_g$ be a bijection satisfying (2.11) with vectors $(x_n)_{n \in H_f}$. If H_f is finite, then σ can be trivially extended to a bijection on \mathbb{N} , and then invariance properties (i) and (ii) are sufficient to show $L(f) = L(g)$. If H is infinite, then we enumerate the sets

$$H_f = \{n_1 < n_2 < \dots\}, \quad H_g = \{m_1 < m_2 < \dots\},$$

and define the functions $h_f \in S$ by $h_f(k, x) = f(n_k, x)$ and $h_g(\ell, x) = g(m_\ell, x)$. By (iii), we have $L(f) = L(h_f)$ and $L(g) = L(h_g)$. Furthermore, σ induces a bijection $\sigma' : \mathbb{N} \rightarrow \mathbb{N}$ by

$$\sigma : n_k \mapsto m_\ell \iff \sigma' : k \mapsto \ell.$$

We now have

$$h_f(k, x) = f(n_k, x) = g(\sigma(n_k), x - x_n) = h_g(\sigma'(k), x - x_n),$$

so that (i) and (ii) give $L(h_f) = L(h_g)$. So $L(f) = L(g)$ in this case as well. \square

In the next lemma, we discuss our first example of a function L satisfying the hypotheses of Corollary 2.6. This function will be important in defining the ‘‘update procedure’’ of Section 3.

LEMMA 2.7. *The map $\|\cdot\| : S \rightarrow [0, 1]$ defined by (2.1) is lower semicontinuous and thus measurable.*

The proof below of Lemma 2.7 is similar to the arguments for several later results. Since one encounters in these cases only variations of this proof, we will frequently spare the reader details and place them in Appendix A.

PROOF OF LEMMA 2.7. It is clear that $\|\cdot\| : S \rightarrow [0, 1]$ satisfies (i)–(iii) in Corollary 2.6, and so the map $f \mapsto \|f\|$ is well defined on \mathcal{S} . To prove lower semicontinuity, it suffices to fix $f \in S$, let $\varepsilon > 0$ be arbitrary, and find $\delta > 0$ sufficiently small that

$$d(f, g) < \delta \implies \|g\| > \|f\| - \varepsilon.$$

Upon selecting a representative $f \in S$, we can find $A \subset \mathbb{N} \times \mathbb{Z}^d$ finite but large enough that $\sum_{u \notin A} f(u) < \frac{\varepsilon}{2}$. By possibly omitting some elements of A , we may assume that f is strictly positive on A (if f is the constant zero function, this results in $A = \emptyset$).

Now take $0 < \delta < \inf_{u \in A} f(u)^2$, where an infimum over the empty set is ∞ . We will further assume $\delta < \varepsilon$. If $d(f, g) < \delta$, then there is a representative $g \in S$ and an isometry $\phi : C \rightarrow \mathbb{N} \times \mathbb{Z}^d$ such that $d_\phi(f, g) < \delta$. It follows that $A \subset C$, since otherwise we would have $d_\phi(f, g) \geq f(u)^2 > \delta$ for some $u \in A \setminus C$. Hence,

$$\begin{aligned} \|g\| &\geq \sum_{u \in \phi(A)} g(u) \geq \sum_{u \in A} f(u) - \sum_{u \in A} |f(u) - g(\phi(u))| + \sum_{u \notin A} f(u) - \sum_{u \notin A} f(u) \\ &> \|f\| - \frac{d_\phi(f, g)}{2} - \frac{\varepsilon}{2} > \|f\| - \frac{\delta}{2} - \frac{\varepsilon}{2} > \|f\| - \varepsilon. \end{aligned}$$

\square

2.2. *Compactness.* We now state the key compactness result for the metric space (\mathcal{S}, d) .

THEOREM 2.8. *Every sequence $(f_n)_{n \geq 1}$ in (\mathcal{S}, d) contains a converging subsequence. That is, there is $f \in \mathcal{S}$ and a subsequence $(f_{n_k})_{k \geq 1}$ such that*

$$\lim_{k \rightarrow \infty} d(f_{n_k}, f) = 0.$$

A continuum version of this result is Theorem 3.2 of [61]. The proof of Theorem 2.8, however, strongly uses discreteness of \mathbb{Z}^d to deal with the explicit metric d . It can be found in Appendix A.2.

The sequential compactness guaranteed by Theorem 2.8 will be used to establish results in the following metric space of probability measures. Denote by $\mathcal{P}(\mathcal{S})$ the set of Borel probability measures on \mathcal{S} , and equip this space with the *Wasserstein metric* ([86], Definition 6.1):

$$\mathcal{W}(\mu, \nu) := \inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{S} \times \mathcal{S}} d(f, g) \pi(df, dg),$$

where $\Pi(\mu, \nu)$ denotes the set of probability measures on $\mathcal{S} \times \mathcal{S}$ having μ and ν as marginals. Note that we may use dual representation of \mathcal{W} due to Kantorovich [49] when it is convenient:

$$(2.12) \quad \mathcal{W}(\mu, \nu) = \sup_{\varphi} \left(\int_{\mathcal{S}} \varphi(f) \mu(df) - \int_{\mathcal{S}} \varphi(f) \nu(df) \right),$$

where the supremum is over 1-Lipschitz functions $\varphi : \mathcal{S} \rightarrow \mathbb{R}$.

Now we recall some general facts on convergence in $\mathcal{P}(\mathcal{S})$. It is a standard result (e.g., see [86], Theorem 6.9) that for measures on a compact metric space, the Wasserstein distance metrizes the topology of weak convergence. Furthermore, this topology is again compact ([86], Remark 6.19). In the coming sections, we will employ weak convergence to prove convergence of not only continuous test functions, but also semicontinuous test functions. Therefore, we will repeatedly apply the Portmanteau lemma, which we record here so that it can be properly quoted later in the paper.

LEMMA 2.9 (Portmanteau; cf. [13], Theorem 2.1 and [84], Theorem 1.3.4). *Given a function $L : \mathcal{S} \rightarrow \mathbb{R}$, define the map $\mathcal{L} : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ by*

$$\mathcal{L}(\mu) := \int_{\mathcal{S}} L(f) \mu(df).$$

If L is lower (resp. upper) semicontinuous, then \mathcal{L} is lower (resp. upper) semicontinuous.

We conclude this section with some observations that will be needed in later arguments.

LEMMA 2.10. *For any $f, g \in \mathcal{S}$, $d(f, g) \leq 2$.*

PROOF. Pick representatives $f, g \in S$ and let $\phi : \emptyset \rightarrow \mathbb{N} \times \mathbb{Z}^d$ be the empty isometry. Then

$$d(f, g) \leq d_{\phi}(f, g) = \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} f(u)^2 + \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} g(u)^2 \leq 2. \quad \square$$

The next lemma concerns measure-theoretic properties of the spaces S and \mathcal{S} . In particular, S is considered as a subset of

$$\ell^1(\mathbb{N} \times \mathbb{Z}^d) = \{f : \mathbb{N} \times \mathbb{Z}^d \rightarrow \mathbb{R} : \|f\| < \infty\},$$

on which there is the standard ℓ^1 norm $\|\cdot\|$ that extends (2.1):

$$\|f\| := \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} |f(u)|, \quad f \in \ell^1(\mathbb{N} \times \mathbb{Z}^d).$$

LEMMA 2.11. Consider the metric spaces $\ell^1(\mathbb{N} \times \mathbb{Z}^d)$ and \mathcal{S} with their Borel σ -algebras. Then the following statements hold:

- (a) \mathcal{S} is a closed (in particular, measurable) subset of $\ell^1(\mathbb{N} \times \mathbb{Z}^d)$ and is thus itself a measurable space with the subspace σ -algebra.
- (b) The quotient map $\iota : \mathcal{S} \rightarrow \mathcal{S}$ that sends $f \in \mathcal{S}$ to its equivalence class in \mathcal{S} is measurable.

PROOF. To show $\mathcal{S} \subset \ell^1(\mathbb{N} \times \mathbb{Z}^d)$ is closed, we express \mathcal{S} as the intersection of closed sets:

$$\mathcal{S} = \{f \in \ell^1(\mathbb{N} \times \mathbb{Z}^d) : \|f\| \leq 1\} \cap \left(\bigcap_{u \in \mathbb{N} \times \mathbb{Z}^d} \{f \in \ell^1(\mathbb{N} \times \mathbb{Z}^d) : f(u) \geq 0\} \right).$$

To next show ι is measurable, it suffices to verify that the inverse image of any open ball is measurable. For $f \in \mathcal{S}$ and $r > 0$, we write

$$B_r(f) := \{g \in \mathcal{S} : d(f, g) < r\}.$$

Notice that

$$\iota^{-1}(B_r(f)) = \bigcup_{\phi} \{g \in \mathcal{S} : d_\phi(f, g) < r\},$$

where the union is over isometries with finite domains. The union occurs, therefore, over a countable set. For each ϕ , it is clear from (2.3) that $d_\phi(f, \cdot)$ is a measurable function on \mathcal{S} , and so each set in the union is measurable. Being a countable union of measurable sets, $\iota^{-1}(B_r(f))$ is measurable. \square

3. The update map. Consider the distribution of ω_n under ρ_n . In this section, we identify said distribution with an element $f_n \in \mathcal{S}$ and define the Markov kernel T that maps f_n to the law of f_{n+1} when conditioned on the random environment only up to time n . In the following, the notation $u \sim v$ is used when $u, v \in \mathbb{N} \times \mathbb{Z}^d$ satisfy $\|u - v\|_1 = 1$. The same notation will be used for adjacent elements $x, y \in \mathbb{Z}^d$.

3.1. *Definition of the update map.* To make the setup precise, we recall the polymer measure ρ_n defined by (1.2) and the endpoint probability mass function $f_n : \mathbb{Z}^d \rightarrow \mathbb{R}$ given by

$$(3.1) \quad f_n(x) := \rho_n(\omega_n = x) = \frac{(2d)^{-n}}{Z_n} \sum_{\substack{|\gamma|=n \\ \gamma(0)=0, \gamma(n)=x}} \exp\left(\beta \sum_{i=1}^n X_{i, \gamma(i)}\right),$$

where the sum is over nearest-neighbor paths $\gamma : \{0, 1, \dots, n\} \rightarrow \mathbb{Z}^d$ of length $|\gamma| = n$, starting at the origin and ending at x . Then f_n is a $[0, 1]$ -valued function on \mathbb{Z}^d and random with respect to $(\Omega_e, \mathcal{F}_n)$, where the σ -algebra \mathcal{F}_n is defined in (1.3). Its value at x gives the probability that a polymer of length n in the given random environment has x as its endpoint.

When the polymer is extended from length n to length $n + 1$, the mass function updates to

$$\begin{aligned} f_{n+1}(x) &= \frac{(2d)^{-(n+1)}}{Z_{n+1}} \sum_{y \sim x} \sum_{\substack{|\gamma|=n \\ \gamma(0)=0, \gamma(n)=y}} \exp\left(\beta \sum_{i=1}^n X_{i, \gamma(i)} + \beta X_{n+1, x}\right) \\ &= \frac{Z_n}{(2d)Z_{n+1}} \sum_{y \sim x} f_n(y) e^{\beta X_{n+1, x}} = \frac{\sum_{y \sim x} f_n(y) e^{\beta X_{n+1, x}}}{2d \cdot Z_{n+1}/Z_n}, \end{aligned}$$

where the denominator is a normalizing factor and thus equal to

$$(3.2) \quad 2d \frac{Z_{n+1}}{Z_n} = 2d \frac{Z_{n+1}}{Z_n} \sum_{x \in \mathbb{Z}^d} f_{n+1}(x) = \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f_n(y) e^{\beta X_{n+1,x}}.$$

Recall that $(X_{n+1,x})_{x \in \mathbb{Z}^d}$ is independent of \mathcal{F}_n , while f_n is measurable with respect to \mathcal{F}_n . Therefore, the distribution of $f_{n+1}(x)$ given \mathcal{F}_n is equal to the distribution of

$$(3.3) \quad F(x) := \frac{\sum_{y \sim x} f_n(y) e^{\beta Y_x}}{\sum_{z \in \mathbb{Z}^d} \sum_{y \sim z} f_n(y) e^{\beta Y_z}},$$

where $(Y_z)_{z \in \mathbb{Z}^d}$ are i.i.d. random variables distributed according to \mathcal{L} and independent of \mathcal{F}_n . Correspondingly, there is an updated free energy,

$$\begin{aligned} \log Z_{n+1} &= \log Z_n + \mathbf{E} \left[\log \frac{Z_{n+1}}{Z_n} \mid \mathcal{F}_n \right] + (\log Z_{n+1} - \mathbf{E}[\log Z_{n+1} \mid \mathcal{F}_n]) \\ &= \log Z_n + R(f_n) + d_n, \end{aligned}$$

where

$$(3.4) \quad R(f_n) := \mathbf{E} \left[\log \frac{Z_{n+1}}{Z_n} \mid \mathcal{F}_n \right] = \mathbf{E} \left[\log \left(\frac{1}{2d} \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f_n(y) e^{\beta Y_x} \right) \mid \mathcal{F}_n \right],$$

and

$$(3.5) \quad d_n := \log \frac{Z_{n+1}}{Z_n} - R(f_n)$$

is a martingale increment. That is, the Doob decomposition of $(\log Z_n)_{n \geq 0}$ is

$$(3.6) \quad \log Z_n = \sum_{i=0}^{n-1} d_i + \sum_{i=0}^{n-1} R(f_i),$$

a fact which has been frequently used in the literature (e.g., [23, 24, 28, 29]).

Next, we wish to reinterpret these observations in the space \mathcal{S} . For a given $f \in \mathcal{S}$, choose a representative (also called f) from S . Consider the law of the random variable $F \in \mathcal{S}$ whose representative is defined by

$$(3.7) \quad F(u) := \frac{\sum_{v \sim u} f(v) e^{\beta Y_u}}{\sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} f(v) e^{\beta Y_w} + 2d(1 - \|f\|) e^{\lambda(\beta)}}, \quad u \in \mathbb{N} \times \mathbb{Z}^d,$$

where $(Y_w)_{w \in \mathbb{N} \times \mathbb{Z}^d}$ are i.i.d. random variables with law \mathcal{L} . Notice that (3.7) is a generalization of (3.3) because in the latter, $\|f_n\| = 1$. The additional summand in the denominator of (3.7) is needed so that F is defined even when $f \equiv 0$; its precise value is chosen so that (3.2) is generalized in an averaged form:

$$\mathbf{E} \left[\sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} f(v) e^{\beta Y_w} + 2d(1 - \|f\|) e^{\lambda(\beta)} \right] = 2d \cdot e^{\lambda(\beta)} = 2d \cdot \mathbf{E} \left(\frac{Z_{n+1}}{Z_n} \right).$$

It is thus apparent that the correct extension of (3.4) is to define $R : \mathcal{S} \rightarrow \mathbb{R}$ by

$$(3.8) \quad R(f) := \mathbf{E} \log \left(\frac{\tilde{F}}{2d} \right), \quad \tilde{F} := \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v) e^{\beta Y_u} + 2d(1 - \|f\|) e^{\lambda(\beta)},$$

where the expectation is over the Y_u .

We must now check that the above definitions do not depend on the representative $f \in S$.

PROPOSITION 3.1. Suppose $f, g \in \mathcal{S}$ satisfy $d(f, g) = 0$. Define F and \tilde{F} as in (3.7) and (3.8), and similarly define

$$G(u) := \frac{\sum_{v \sim u} g(v) e^{\beta Z_u}}{\tilde{G}}, \quad \tilde{G} := \sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} g(v) e^{\beta Z_w} + 2d(1 - \|g\|) e^{\lambda(\beta)},$$

where the Z_w are i.i.d. with law \mathcal{L} . Then the following statements hold:

- (a) The law of $F \in \mathcal{S}$ is equal to the law of $G \in \mathcal{S}$.
- (b) The law of $\tilde{F} \in \mathbb{R}$ is equal to the law of $\tilde{G} \in \mathbb{R}$. In particular, $R(f) = R(g)$.

PROOF. To show the two claims, it suffices to exhibit a coupling of the environments $(Y_u)_{u \in \mathbb{N} \times \mathbb{Z}^d}$ and $(Z_u)_{u \in \mathbb{N} \times \mathbb{Z}^d}$ such that $F = G$ in \mathcal{S} and $\tilde{F} = \tilde{G}$ in \mathbb{R} . So we let H_f and H_g denote the \mathbb{N} -supports of f and g , respectively, and take $\sigma : H_f \rightarrow H_g$ and $(x_n)_{n \in H_f}$ as in Corollary 2.5, so that (2.11) holds. Define $\psi : H_f \times \mathbb{Z}^d \rightarrow H_g \times \mathbb{Z}^d$ by $\psi(n, x) := (\sigma(n), x - x_n)$, and let Z_u be equal to $Y_{\psi^{-1}(u)}$ whenever $u \in H_g \times \mathbb{Z}^d$. Otherwise, we may take Z_u to be an independent copy of Y_u . For $u \in H_f \times \mathbb{Z}^d$, we have by (2.11)

$$(3.9) \quad \sum_{v \sim u} f(v) e^{\beta Y_u} = \sum_{v \sim u} g(\psi(v)) e^{\beta Y_u} = \sum_{v \sim \psi(u)} g(v) e^{\beta Z_{\psi(u)}}.$$

Since $f(v) = 0$ for all $v \notin H_f \times \mathbb{Z}^d$, and similarly $g(v) = 0$ for all $v \notin H_g \times \mathbb{Z}^d$, we can sum over all of $\mathbb{N} \times \mathbb{Z}^d$, and (3.9) gives

$$\sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} f(v) e^{\beta Y_w} = \sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} g(v) e^{\beta Z_w}.$$

This identity, together with the fact that $\|f\| = \|g\|$ (cf. Lemma 2.7), shows $\tilde{F} = \tilde{G}$ and thus proves claim (b). When we further consider (3.9), we see that $F(u) = G(\psi(u))$ for all $u \in H_f \times \mathbb{Z}^d$. Hence, $d_\psi(F, G) = 0$. Moreover, for any $\varepsilon > 0$, we can find a finite subset $A \subset H_f \times \mathbb{Z}^d$ such that

$$\sum_{u \notin A} F(u)^2 + \sum_{u \notin \psi(A)} G(u)^2 < \varepsilon.$$

With $\phi := \psi|_A$, we thus have $d(F, G) \leq d_\phi(F, G) < \varepsilon$. Letting ε tend to 0 gives claim (a). □

Given Proposition 3.1, we may define $Tf \in \mathcal{P}(\mathcal{S})$ to be the law of F given $f \in \mathcal{S}$. The next step in our construction is to establish continuity. The seemingly unexciting result below requires a careful and lengthy proof that deserves a separate Section 8.

PROPOSITION 3.2.

- (a) $f \mapsto Tf$ is a continuous map from (\mathcal{S}, d) to $(\mathcal{P}(\mathcal{S}), \mathcal{W})$.
- (b) For any positive integer q , $f \mapsto \mathbf{E}(\log^q \tilde{F})$ is a continuous map from (\mathcal{S}, d) to \mathbb{R} . In particular, the case $q = 1$ implies $R(\cdot)$ is continuous.

The following consequence of part (b) is not new but will be convenient to have recorded for later use.

COROLLARY 3.3. Recall the martingale increment d_n from (3.5). For any positive integer q , there exists a constant C depending only on \mathcal{L} , β , d and q , such that $\mathbf{E}(d_n^q) \leq C$ for every n .

PROOF. If we fix $f = f_n$ and take \tilde{F} as in (3.8), then (3.2) tells us

$$\mathbf{E}\left[\log^q \frac{Z_{n+1}}{Z_n} \mid \mathcal{F}_n\right] = \mathbf{E}\left(\log^q \frac{\tilde{F}}{2d}\right).$$

By the compactness of \mathcal{S} and the continuity property of Proposition 3.2(b), the above right-hand side is bounded by a constant depending only on \mathcal{L} , β , d and q . The same is then true for $\mathbf{E}(\log^q \frac{Z_{n+1}}{Z_n})$, and thus also for $\mathbf{E}(d_n^q)$. \square

To apply $f \mapsto Tf$ to the functions of interest, namely the endpoint distributions from (3.1), we must first identify f_n with the partitioned subprobability measure having representative

$$(3.10) \quad f_n(u) = \begin{cases} f_n(x) & \text{if } u = (n, x), \\ 0 & \text{else,} \end{cases} \quad u \in \mathbb{N} \times \mathbb{Z}^d.$$

In review, f_n is a function on \mathbb{Z}^d that is random with respect to \mathcal{F}_n . It is natural (and measurable) to identify f_n with a function on $\mathbb{N} \times \mathbb{Z}^d$ that is supported on the n th copy of \mathbb{Z}^d ; this is (3.10). Finally, that function—thus far an element of \mathcal{S} —is identified with its equivalence class $\iota(f_n) \in \mathcal{S}$, where ι is the quotient map from Lemma 2.11. In accordance with our previous notation, the symbol f_n will henceforth denote the equivalence class in \mathcal{S} unless stated otherwise, while $f_n(u)$ will denote the representative defined by (3.10), evaluated at $u \in \mathbb{N} \times \mathbb{Z}^d$. By the discussion preceding (3.3), the law of $f_{n+1} \in \mathcal{S}$ given \mathcal{F}_n is equal to Tf_n . For this reason, we refer to T as the “update map.”

3.2. *Lifting the update map.* Thus, far we have only considered the random variable F in (3.7) for fixed $f \in \mathcal{S}$. If f is itself a random element of \mathcal{S} drawn from the probability measure $\mu \in \mathcal{P}(\mathcal{S})$, the resulting total law of F will be written $\mathcal{T}\mu$. That is,

$$\mathcal{T}\mu(\mathcal{A}) = \int Tf(\mathcal{A})\mu(df), \quad \text{Borel } \mathcal{A} \subset \mathcal{S}.$$

More generally, for $\varphi : \mathcal{S} \rightarrow \mathbb{R}$,

$$(3.11) \quad \int \varphi(g)\mathcal{T}\mu(dg) = \int_{\mathcal{S}} \int_{\mathcal{S}} \varphi(g)Tf(dg)\mu(df),$$

where the integral is well defined if φ is nonnegative, or if

$$\int_{\mathcal{S}} \int_{\mathcal{S}} |\varphi(g)|Tf(dg)\mu(df) < \infty.$$

We can recover the map T as $Tf = \mathcal{T}\delta_f$, where $\delta_f \in \mathcal{P}(\mathcal{S})$ is the unit point mass at f . From Proposition 3.2, one can conclude by standard arguments that $\mu \mapsto \mathcal{T}\mu$ is (uniformly) continuous. A proof is included in Appendix A.3.

PROPOSITION 3.4. $\mu \mapsto \mathcal{T}\mu$ is a continuous map $(\mathcal{P}(\mathcal{S}), \mathcal{W}) \rightarrow (\mathcal{P}(\mathcal{S}), \mathcal{W})$.

4. The empirical measure of the endpoint distribution.

4.1. *Definition and properties.* As discussed in Section 3.1, the law of the endpoint distribution f_{n+1} given \mathcal{F}_n is equal to Tf_n . Now define the empirical probability measure on \mathcal{S} generated by the f_i ,

$$(4.1) \quad \mu_n := \frac{1}{n} \sum_{i=0}^{n-1} \delta_{f_i}.$$

Then μ_n is a random element of $\mathcal{P}(\mathcal{S})$, measurable with respect to \mathcal{F}_n . While we will be interested in the quantity $\mathcal{W}(\mu_n, \mathcal{T}\mu_n)$, it is easier to replace μ_n by the shifted empirical measure,

$$\mu'_n := \frac{1}{n} \sum_{i=1}^n \delta_{f_i},$$

since $\mathcal{T}f_i$ is the distribution of f_{i+1} given \mathcal{F}_i .

Making use of the dual formulation (2.12) of Wasserstein distance, one has

$$\begin{aligned} \mathcal{W}(\mu'_n, \mathcal{T}\mu_n) &= \sup_{\varphi} \left(\int \varphi(f) \mu'_n(df) - \int \varphi(f) \mathcal{T}\mu_n(df) \right) \\ &= \sup_{\varphi} \frac{1}{n} \sum_{i=0}^{n-1} (\varphi(f_{i+1}) - \mathbf{E}[\varphi(f_{i+1}) | \mathcal{F}_i]), \end{aligned}$$

where the supremum is taken over all functions $\varphi : \mathcal{S} \rightarrow \mathbb{R}$ satisfying

$$(4.2) \quad |\varphi(f) - \varphi(g)| \leq d(f, g) \quad \text{for all } f, g \in \mathcal{S}.$$

Notice, however, that adding a constant to φ does not change the value of $\int \varphi(f) \mu'_n(df) - \int \varphi(f) \mathcal{T}\mu_n(df)$, and so the supremum can equivalently be taken over φ satisfying $\varphi(\mathbf{0}) = 0$, where $\mathbf{0}$ denotes (the equivalence class of) the constant zero function. Let us denote the set of such functions by

$$\mathcal{L} = \{ \varphi : \mathcal{S} \rightarrow \mathbb{R} : |\varphi(f) - \varphi(g)| \leq d(f, g) \text{ for all } f, g \in \mathcal{S}, \varphi(\mathbf{0}) = 0 \}.$$

Recall that the space of real-valued continuous functions on a compact metric space is equipped with the uniform norm,

$$\|\varphi\|_{\infty} := \sup_{f \in \mathcal{S}} |\varphi(f)| < \infty.$$

For $\varphi \in \mathcal{L}$, the Lipschitz condition (4.2) and Lemma 2.10 imply $\|\varphi\|_{\infty} \leq 2$. In particular, \mathcal{L} is a uniformly bounded family of continuous functions. Furthermore, since \mathcal{L} consists of Lipschitz functions whose minimal Lipschitz constants are uniformly bounded, it is both equicontinuous and closed under the topology induced by the uniform norm. By the Arzelà–Ascoli theorem (see Munkres [62], Theorem 47.1), \mathcal{L} is compact in this topology. Having made this observation, we are now ready to prove the following convergence result.

PROPOSITION 4.1. *As $n \rightarrow \infty$, $\mathcal{W}(\mu'_n, \mathcal{T}\mu_n) \rightarrow 0$ almost surely.*

We will use the following well-known fact.

LEMMA 4.2 (Burkholder–Davis–Gundy Inequality, see [22], Theorem 1.1). *Let $(M_n)_{n \geq 0}$ be a martingale, and write*

$$M_n = \sum_{i=0}^n d_i, \quad d_0 = M_0, \quad d_i = M_i - M_{i-1} \text{ for } i \geq 1.$$

Let $M_n^ := \sup_{0 \leq i \leq n} M_i$. Then for any $q \geq 1$, there are positive constants c_q and C_q such that*

$$c_q \mathbf{E} \left[\left(\sum_{i=0}^n d_i^2 \right)^{q/2} \right] \leq \mathbf{E}[(M_n^*)^q] \leq C_q \mathbf{E} \left[\left(\sum_{i=0}^n d_i^2 \right)^{q/2} \right] \quad \text{for all } n \geq 0.$$

PROOF OF PROPOSITION 4.1. We have

$$\mathcal{W}(\mu'_n, \mathcal{T}\mu_n) = \sup_{\varphi \in \mathcal{L}} \frac{1}{n} \sum_{i=0}^{n-1} (\varphi(f_{i+1}) - \mathbf{E}[\varphi(f_{i+1}) | \mathcal{F}_i]).$$

Notice that for any fixed $\varphi \in \mathcal{L}$,

$$M_n(\varphi) := \sum_{i=0}^{n-1} (\varphi(f_{i+1}) - \mathbf{E}[\varphi(f_{i+1}) | \mathcal{F}_i])$$

defines a martingale $(M_n(\varphi))_{n \geq 0}$ adapted to the filtration $(\mathcal{F}_n)_{n \geq 0}$. By Lemma 4.2, there is a constant $C = C(\varphi)$ such that

$$\begin{aligned} \mathbf{E}[M_n(\varphi)^4] &\leq \mathbf{E}[M_n^*(\varphi)^4] \\ &\leq C \mathbf{E} \left[\left(\sum_{i=0}^{n-1} (\varphi(f_{i+1}) - \mathbf{E}[\varphi(f_{i+1}) | \mathcal{F}_i]) \right)^2 \right] \leq 16Cn^2, \end{aligned}$$

where the final inequality follows from (4.2) and Lemma 2.10. A Markov bound now gives

$$\begin{aligned} \sum_{n=0}^{\infty} \mathbf{P} \left(\frac{|M_n(\varphi)|}{n} > n^{-1/5} \right) &= \sum_{n=0}^{\infty} \mathbf{P}(M_n(\varphi)^4 > n^{16/5}) \\ &\leq \sum_{n=0}^{\infty} n^{-16/5} \mathbf{E}[M_n(\varphi)^4] \\ &\leq \sum_{n=0}^{\infty} 16Cn^{-6/5} < \infty. \end{aligned}$$

By Borel–Cantelli, we may conclude

$$(4.3) \quad \lim_{n \rightarrow \infty} \frac{|M_n(\varphi)|}{n} = 0 \quad \text{a.s.}$$

As discussed above, \mathcal{L} is compact in the uniform norm topology. In particular, it is separable. Let $\varphi_1, \varphi_2, \dots$ be a countable, dense subset of \mathcal{L} . Because of (4.3), we can say that with probability one,

$$(4.4) \quad \lim_{n \rightarrow \infty} \frac{M_n(\varphi_j)}{n} = 0 \quad \text{for all } j \geq 1.$$

Assume that this almost sure event occurs. Again from (4.2) and Lemma 2.10, we know

$$\|\varphi - \varphi'\|_{\infty} < \varepsilon \quad \Rightarrow \quad \left| \frac{M_n(\varphi)}{n} - \frac{M_n(\varphi')}{n} \right| < 2\varepsilon,$$

meaning $(M_n(\cdot)/n)_{n \geq 0}$ is an equicontinuous sequence of functions on the compact metric space \mathcal{L} . The assumption (4.4) says that this family converges pointwise to 0 on a dense subset. The Arzelà–Ascoli theorem forces this convergence to be uniform. That is, for any $\varepsilon > 0$, there is N large enough that

$$n \geq N \quad \Rightarrow \quad \mathcal{W}(\mu'_n, \mathcal{T}\mu_n) = \sup_{\varphi \in \mathcal{L}} \frac{M_n(\varphi)}{n} < \varepsilon.$$

We conclude that $\mathcal{W}(\mu'_n, \mathcal{T}\mu_n)$ tends to 0 as $n \rightarrow \infty$. As this holds given the almost sure event (4.4), we are done. \square

4.2. *Convergence to fixed points of the update map.* Proposition 4.1 suggests that for large n , the empirical measure will be close to the set of fixed points of \mathcal{T} :

$$(4.5) \quad \mathcal{K} := \{v \in \mathcal{P}(\mathcal{S}) : \mathcal{T}v = v\}.$$

For $\mathcal{U} \subset \mathcal{P}(\mathcal{S})$, we will denote distance to \mathcal{U} by

$$\mathcal{W}(\mu, \mathcal{U}) := \inf_{v \in \mathcal{U}} \mathcal{W}(\mu, v), \quad \mu \in \mathcal{P}(\mathcal{S}).$$

COROLLARY 4.3. *As $n \rightarrow \infty$, $\mathcal{W}(\mu_n, \mathcal{K}) \rightarrow 0$ almost surely.*

PROOF. Recall that $\mathbf{0}$ is the element of \mathcal{S} whose unique representative in S is the constant zero function. Notice that $\mathcal{T}\delta_{\mathbf{0}} = \delta_{\mathbf{0}}$ so that \mathcal{K} is nonempty. Next observe that from Lemma 2.10, we have the trivial bound $\mathcal{W}(\mu_n, \mu'_n) \leq 2/n$, and so Proposition 4.1 immediately implies $\mathcal{W}(\mu_n, \mathcal{T}\mu_n) \rightarrow 0$ almost surely. On this almost sure event, it follows that $\mathcal{W}(\mu_n, \mathcal{K}) \rightarrow 0$, since otherwise there would exist a subsequence μ_{n_k} remaining a fixed positive distance away from \mathcal{K} . By compactness, we could assume μ_{n_k} converges to some μ , but then continuity of \mathcal{T} would force $\mathcal{W}(\mu, \mathcal{T}\mu) = 0$; that is, μ_{n_k} converges to an element of \mathcal{K} , which is a contradiction. \square

Now that the set \mathcal{K} is seen to contain all possible limits of the empirical measure, we should like to have some description of its elements. One suggestive fact proved below is that any measure in \mathcal{K} places all its mass on those elements of \mathcal{S} with norm 0 or 1. This observation will be crucial in proving our characterization of the low temperature phase in Section 5.

PROPOSITION 4.4. *If $v \in \mathcal{K}$, then*

$$v(\{f \in \mathcal{S} : 0 < \|f\| < 1\}) = 0.$$

PROOF. First take $f \in \mathcal{S}$ to be nonrandom. Then Tf is the law of the random function

$$F(u) = \frac{\sum_{v \sim u} f(v)e^{\beta Y_u}}{\sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} f(v)e^{\beta Y_w} + 2d(1 - \|f\|)e^{\lambda(\beta)}}.$$

If $\|f\| = 0$ or $\|f\| = 1$, then $\|F\| = \|f\|$. If instead $0 < \|f\| < 1$, then $\|F\|$ is random and still satisfies $0 < \|F\| < 1$. By summing over $u \in \mathbb{N} \times \mathbb{Z}^d$, we have

$$\mathbf{E}(\|F\|) = \mathbf{E} \left[\frac{\sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v)e^{\beta Y_u}}{\sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v)e^{\beta Y_u} + 2d(1 - \|f\|)e^{\lambda(\beta)}} \right].$$

Upon observing that for any constant $C > 0$, the mapping $t \mapsto \frac{t}{t+C}$ is strictly concave, we deduce from Jensen’s inequality that

$$\begin{aligned} \mathbf{E}(\|F\|) &\leq \frac{\mathbf{E}[\sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v)e^{\beta Y_u}]}{\mathbf{E}[\sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v)e^{\beta Y_u}] + 2d(1 - \|f\|)e^{\lambda(\beta)}} \\ &= \frac{2d\|f\| \cdot e^{\lambda(\beta)}}{2d \cdot e^{\lambda(\beta)}} = \|f\|, \end{aligned}$$

where equality holds if and only if $\sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v)e^{\beta Y_u}$ is an almost sure constant. Since the disorder distribution \mathcal{L} is not a point mass, this is not the case.

Now let $v \in \mathcal{P}(\mathcal{S})$, and take $f \in \mathcal{S}$ to be random with law v . If v assigns positive measure to the set $\{f \in \mathcal{S} : 0 < \|f\| < 1\}$, then the above argument gives

$$\int \|g\| \mathcal{T}v(dg) = \iint \|F\| Tf(dF)v(df) < \int \|f\| v(df).$$

But when $v \in \mathcal{K}$ so that $\mathcal{T}v = v$, the first and last expressions above are equal. We conclude that any $v \in \mathcal{K}$ must assign mass 0 to the set $\{f \in \mathcal{S} : 0 < \|f\| < 1\}$. \square

4.3. *Variational formula for the free energy.* In order to connect the results of Section 4.2 to the temperature conditions of Theorems 1.1 and 1.2, we will need to relate the free energy $F_n = \frac{1}{n} \log Z_n$ to the abstract objects we have introduced. Recall the Doob decomposition from (3.4)–(3.6):

$$\log Z_n = M_n + A_n \quad \text{where } M_n = \sum_{i=0}^{n-1} d_i, A_n = \sum_{i=0}^{n-1} R(f_i).$$

If we lift the functional $R : \mathcal{S} \rightarrow \mathbb{R}$ of (3.8) to $\mathcal{R} : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ by defining

$$(4.6) \quad \mathcal{R}(\mu) := \int R(f)\mu(df), \quad \mu \in \mathcal{P}(\mathcal{S}),$$

then we can conveniently rewrite the above decomposition as

$$(4.7) \quad F_n = \frac{\log Z_n}{n} = \frac{M_n}{n} + \mathcal{R}(\mu_n),$$

where μ_n is the empirical measure from (4.1). Our next result, which says the mean-zero martingale $(M_n)_{n \geq 0}$ has a vanishing contribution in the above expression, is merely a reinterpretation of arguments appearing in earlier works (e.g., [23, 24, 28, 29]).

PROPOSITION 4.5. *As $n \rightarrow \infty$, $|F_n - \mathcal{R}(\mu_n)| \rightarrow 0$ almost surely.*

PROOF. By Corollary 3.3, we have $\mathbf{E}(d_i^4) \leq C_1$ for some constant C_1 independent of i . From Lemma 4.2, we deduce that

$$\mathbf{E}(M_n^4) \leq C_2 \mathbf{E} \left[\left(\sum_{i=0}^{n-1} d_i^2 \right)^2 \right] \leq C_2 \sum_{i=0}^{n-1} \sum_{j=0}^{n-1} \sqrt{\mathbf{E}(d_i^4)\mathbf{E}(d_j^4)} \leq Cn^2,$$

where the constant $C = C_1 C_2$ is independent of n . It follows that $\mathbf{E}[(n^{-1}M_n)^4] \leq Cn^{-2}$. As in the proof of Proposition 4.1, an argument using Markov’s inequality and Borel–Cantelli shows

$$\lim_{n \rightarrow \infty} |F_n - \mathcal{R}(\mu_n)| = \lim_{n \rightarrow \infty} \frac{|M_n|}{n} = 0 \quad \text{a.s.} \quad \square$$

Given the convergence of μ_n to the set \mathcal{K} , one should expect $\mathbf{E}(F_n) = \mathbf{E}\mathcal{R}(\mu_n)$ to become close to $\mathcal{R}(v)$ for some $v \in \mathcal{K}$. One difficulty is that μ_n does not converge to a particular $v \in \mathcal{K}$, but rather becomes arbitrarily close to the set \mathcal{K} . Nevertheless, we can instead consider the subset

$$(4.8) \quad \mathcal{M} = \left\{ v_0 \in \mathcal{K} : \mathcal{R}(v_0) = \inf_{v \in \mathcal{K}} \mathcal{R}(v) \right\},$$

and show convergence to \mathcal{M} . By Proposition 3.2(b) and Lemma 2.9, \mathcal{R} is continuous. Therefore, since \mathcal{K} is compact (being a closed subset of a compact metric space), the set \mathcal{M} is nonempty. Moreover, \mathcal{M} is a closed subset of the compact space \mathcal{K} , and so \mathcal{M} is compact. The first step in proving the desired convergence is the following consequence of Corollary 4.3.

PROPOSITION 4.6. *Let \mathcal{K} be defined by (4.5). Let $\mathcal{R} : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ be defined by (4.6). Then*

$$(4.9) \quad \liminf_{n \rightarrow \infty} F_n \geq \inf_{v \in \mathcal{K}} \mathcal{R}(v) \quad \text{a.s.}$$

In particular,

$$(4.10) \quad \liminf_{n \rightarrow \infty} \mathbf{E}(F_n) \geq \inf_{v \in \mathcal{K}} \mathcal{R}(v).$$

PROOF. \mathcal{R} is continuous on the compact metric space $(\mathcal{P}(\mathcal{S}), \mathcal{W})$, and thus uniformly continuous. Therefore, a simple argument using Corollary 4.3 gives

$$(4.11) \quad \liminf_{n \rightarrow \infty} \mathcal{R}(\mu_n) \geq \inf_{v \in \mathcal{K}} \mathcal{R}(v) \quad \text{a.s.}$$

Since \mathcal{R} is bounded, we may apply Fatou’s lemma to obtain

$$(4.12) \quad \liminf_{n \rightarrow \infty} \mathbf{E}\mathcal{R}(\mu_n) \geq \mathbf{E}\left[\liminf_{n \rightarrow \infty} \mathcal{R}(\mu_n)\right] \geq \inf_{v \in \mathcal{K}} \mathcal{R}(v).$$

Now (4.9) follows from (4.11) and Proposition 4.5, and (4.10) follows from (4.12) and the fact that $\mathbf{E}(F_n) = \mathbf{E}\mathcal{R}(\mu_n)$. \square

Following Proposition 4.6, we naturally ask if there is a matching upper bound. The next result answers this question in the affirmative. To state the full theorem, we need to denote one element of \mathcal{S} in particular. Notice that for $f \in \mathcal{S}$ satisfying $f(u) = 1$ for some $u \in \mathbb{N} \times \mathbb{Z}^d$, Proposition 2.4 implies $d(f, g) = 0$ for $g \in \mathcal{S}$ if and only if $g(v) = 1$ for some $v \in \mathbb{N} \times \mathbb{Z}^d$. We can thus define the element $\mathbf{1} \in \mathcal{S}$ whose representatives in \mathcal{S} are the norm-1 functions supported on a single point.

THEOREM 4.7. *Let \mathcal{K} be defined by (4.5). Let $\mathcal{R} : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ be defined by (4.6). Then*

$$\limsup_{n \rightarrow \infty} \mathbf{E}(F_n) \leq \inf_{v \in \mathcal{K}} \mathcal{R}(v),$$

and so

$$(4.13) \quad \lim_{n \rightarrow \infty} F_n = \inf_{v \in \mathcal{K}} \mathcal{R}(v) \quad \text{a.s.}$$

The minimum value is equal to

$$(4.14) \quad \inf_{v \in \mathcal{K}} \mathcal{R}(v) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathcal{R}(\mathcal{T}^i \delta_{\mathbf{1}}).$$

The key ingredient in the proof of Theorem 4.7 is the following lemma. When $\|f_0\| = 1$, the result can be regarded as Jensen’s inequality applied to a weighted sum of partition functions, each one corresponding to a different starting vertex chosen at random according to f_0 .

LEMMA 4.8. *For any $f_0 \in \mathcal{S}$ and $n \geq 1$,*

$$\sum_{i=0}^{n-1} \mathcal{R}(\mathcal{T}^i \delta_{f_0}) \geq \mathbf{E} \log Z_n,$$

where $\delta_{f_0} \in \mathcal{P}(\mathcal{S})$ is the unit mass at f_0 . Equality holds if and only if $f_0 = \mathbf{1}$.

PROOF. If $f_0 = \mathbf{0}$, then $\mathcal{R}(\mathcal{T}^i \delta_{\mathbf{0}}) = \mathcal{R}(\delta_{\mathbf{0}}) = \lambda(\beta)$, and so the desired inequality is immediate from (1.4). So we may assume $f_0 \neq \mathbf{0}$. Fix a representative $f_0 \in \mathcal{S}$. Let $(Y_u^{(i)})_{u \in \mathbb{N} \times \mathbb{Z}^d}$, $1 \leq i \leq n$, be independent collections of i.i.d. random variables with law \mathcal{L} . For $1 \leq i \leq n - 1$, inductively define $f_i \in \mathcal{S}$ to have representative

$$f_i(u) = \frac{\sum_{v \sim u} f_{i-1}(v) e^{\beta Y_u^{(i)}}}{\sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} f_{i-1}(v) e^{\beta Y_w^{(i)}} + 2d(1 - \|f_{i-1}\|) e^{\lambda(\beta)}},$$

so that the law of f_i is equal to \mathcal{T} applied to the law of f_{i-1} . By induction, then, the law of f_i is $\mathcal{T}^i \delta_{f_0}$. By definitions (4.6) and (3.8),

$$\mathcal{R}(\mathcal{T}^i \delta_{f_0}) = \mathbf{E} \log \tilde{F}_i - \log 2d,$$

where

$$\tilde{F}_i := \sum_{u_{i+1} \in \mathbb{N} \times \mathbb{Z}^d} \sum_{u_i \sim u_{i+1}} f_i(u_i) e^{\beta Y_{u_{i+1}}^{(i+1)}} + 2d(1 - \|f_i\|) e^{\lambda(\beta)},$$

and the expectation is over both f_i and the $Y^{(i+1)}$ variables. Repeatedly rewriting f_i in terms of f_{i-1} leads to the identity

$$\begin{aligned} & \tilde{F}_0 \tilde{F}_1 \cdots \tilde{F}_{n-1} \\ &= \sum_{u_n \in \mathbb{N} \times \mathbb{Z}^d} \sum_{u_0 \sim u_1 \sim \cdots \sim u_n} f_0(u_0) e^{\beta \sum_{i=1}^n Y_{u_i}^{(i)}} + (2d)^n (1 - \|f_0\|) e^{n\lambda(\beta)} \\ &= \sum_{u_0 \in \mathbb{N} \times \mathbb{Z}^d} \frac{f_0(u_0)}{\|f_0\|} \left[\|f_0\| \sum_{u_n \sim \cdots \sim u_0} e^{\beta \sum_{i=1}^n Y_{u_i}^{(i)}} + (2d)^n (1 - \|f_0\|) \mathbf{E}(Z_n) \right]. \end{aligned}$$

Using the concavity of the log function in three consecutive steps, we deduce

$$\begin{aligned} & \log \tilde{F}_0 \tilde{F}_1 \cdots \tilde{F}_{n-1} \\ & \geq \sum_{u_0} \frac{f_0(u_0)}{\|f_0\|} \log \left[\|f_0\| \sum_{u_n \sim \cdots \sim u_0} e^{\beta \sum_{i=1}^n Y_{u_i}^{(i)}} + (2d)^n (1 - \|f_0\|) \mathbf{E}(Z_n) \right] \\ & \geq \sum_{u_0} \frac{f_0(u_0)}{\|f_0\|} \left[\|f_0\| \log \sum_{u_n \sim \cdots \sim u_0} e^{\beta \sum_{i=1}^n Y_{u_i}^{(i)}} + (1 - \|f_0\|) \log((2d)^n \mathbf{E}(Z_n)) \right] \\ & \geq \sum_{u_0} \frac{f_0(u_0)}{\|f_0\|} \left[\|f_0\| \log \sum_{u_n \sim \cdots \sim u_0} e^{\beta \sum_{i=1}^n Y_{u_i}^{(i)}} + (1 - \|f_0\|) \mathbf{E} \log((2d)^n Z_n) \right], \end{aligned}$$

where equality holds throughout if and only if $f_0(u_0) = 1$ for some $u_0 \in \mathbb{N} \times \mathbb{Z}^d$. Since the random variable $\sum_{u_n \sim \cdots \sim u_0} \exp(\beta \sum_{i=1}^n Y_{u_i}^{(i)})$ is equal in law to $(2d)^n Z_n$ for any fixed $u_0 \in \mathbb{N} \times \mathbb{Z}^d$, taking expectation yields

$$\begin{aligned} & \mathbf{E} \log \tilde{F}_0 \tilde{F}_1 \cdots \tilde{F}_{n-1} \\ & \geq \sum_{u_0} \frac{f_0(u_0)}{\|f_0\|} (\|f_0\| \mathbf{E} \log((2d)^n Z_n) + (1 - \|f_0\|) \mathbf{E} \log((2d)^n Z_n)) \\ & = \mathbf{E} \log((2d)^n Z_n). \end{aligned}$$

It follows that

$$\begin{aligned} \sum_{i=0}^{n-1} \mathcal{R}(\mathcal{T}^i \delta_{f_0}) &= \sum_{i=0}^{n-1} (\mathbf{E} \log \tilde{F}_i - \log 2d) \\ &= \mathbf{E} \log(\tilde{F}_0 \tilde{F}_1 \cdots \tilde{F}_{n-1}) - \log(2d)^n \geq \mathbf{E} \log Z_n, \end{aligned}$$

with equality if and only if $f_0 = \mathbf{1}$. \square

PROOF OF THEOREM 4.7. Consider any $\nu \in \mathcal{K}$. Using the fact that $\mathcal{T}\nu = \nu$ and Lemma 4.8, we have

$$\begin{aligned} \mathcal{R}(\nu) &= \frac{1}{n} \sum_{i=0}^{n-1} \mathcal{R}(\mathcal{T}^i \nu) = \int \frac{1}{n} \sum_{i=0}^{n-1} \mathcal{R}(\mathcal{T}^i \delta_f) \nu(\mathrm{d}f) \\ &\geq \int \frac{1}{n} \mathbf{E} \log Z_n \nu(\mathrm{d}f) = \mathbf{E}(F_n). \end{aligned}$$

We note that linearity of \mathcal{T} and Fubini’s theorem (3.11) have been applied, which is permissible since Proposition 3.2(b) shows that R is continuous on \mathcal{S} and hence bounded.

As the above estimate holds for every $\nu \in \mathcal{K}$ and every n , we have

$$(4.15) \quad \limsup_{n \rightarrow \infty} \mathbf{E}(F_n) \leq \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu).$$

It now follows from (4.10) and (4.15) that

$$(4.16) \quad \lim_{n \rightarrow \infty} \mathbf{E}(F_n) = \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu).$$

Then (1.5) improves (4.16) to (4.13). Finally, equation (4.14) follows from the final statement in Lemma 4.8. \square

We now turn to strengthening Corollary 4.3 by proving convergence not only to \mathcal{K} , but to the smaller set \mathcal{M} , defined in (4.8).

THEOREM 4.9. *As $n \rightarrow \infty$, $\mathcal{W}(\mu_n, \mathcal{M}) \rightarrow 0$ almost surely.*

PROOF. As $n \rightarrow \infty$, $\mathcal{W}(\mu_n, \mathcal{K}) \rightarrow 0$ almost surely (by Corollary 4.3), and $\mathcal{R}(\mu_n)$ converges almost surely to $\inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu)$ (by Proposition 4.5 and Theorem 4.7). Therefore, by continuity of \mathcal{R} and compactness of $\mathcal{P}(\mathcal{S})$, we have $\mathcal{W}(\mu_n, \mathcal{K}_\delta) \rightarrow 0$ almost surely, where δ is any positive number, and

$$\mathcal{K}_\delta := \left\{ \nu_0 \in \mathcal{K} : \mathcal{R}(\nu_0) < \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu) + \delta \right\}.$$

Now, given any $\varepsilon > 0$, we can choose δ such that $\sup_{\nu \in \mathcal{K}_\delta} \mathcal{W}(\nu, \mathcal{M}) < \varepsilon$. (Indeed, if this were not the case, then one could find a sequence $(\nu_k)_{k \geq 1}$ in \mathcal{K} such that $\mathcal{R}(\nu_k) \searrow \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu)$ as $k \rightarrow \infty$, but $\mathcal{W}(\nu_k, \mathcal{M}) \geq \varepsilon$ for all k . Since \mathcal{K} is compact, we may pass to a subsequence and assume ν_k converges to some $\nu_0 \in \mathcal{K}$. In particular, $\mathcal{W}(\nu_0, \mathcal{M}) \geq \varepsilon$. But the continuity of \mathcal{R} implies $\mathcal{R}(\nu_0) = \inf_{\nu \in \mathcal{K}} \mathcal{R}(\nu)$, meaning $\nu \in \mathcal{M}$, a contradiction.) As $\mathcal{W}(\mu_n, \mathcal{K}_\delta) \rightarrow 0$ almost surely and $\varepsilon > 0$ is arbitrary, we are done. \square

5. Limits of empirical measures. In the first part of this section, we give a characterization of the low temperature regime in terms of the fixed points of the update map \mathcal{T} . This is stated as Theorem 5.2(b). We know from Theorem 4.9 that those fixed points minimizing the energy functional \mathcal{R} constitute the possible limits of the empirical measure μ_n , and so this characterization will allow us to study the Cesàro asymptotics of endpoint distributions.

In Section 5.2, we make this connection more concrete by outlining the general steps by which the theory developed in the previous three sections can be used to prove statements about directed polymers. Indeed, the strategy described will be employed in Sections 6 and 7. But to provide a more brief, instructional example in the present section, we give a new proof of Theorem D, which offered a first characterization of the low temperature phase. Once appropriate definitions are made, only a short argument is needed to prove the statement.

5.1. *Characterizing high and low temperature phases.* Recall that $\mathbf{0}$ is the element of \mathcal{S} whose unique representative in S is the constant zero function, and $\mathbf{1}$ is the element of \mathcal{S} whose representatives in S are the norm-1 functions supported on a single point.

LEMMA 5.1. *The function $R : \mathcal{S} \rightarrow \mathbb{R}$ from (3.8) achieves a unique minimum $R(\mathbf{1}) = \mathbf{E} \log Z_1$, and a unique maximum $R(\mathbf{0}) = \lambda(\beta)$.*

PROOF. The first statement is immediate from Lemma 4.8 by taking $n = 1$. The second statement follows from the Jensen inequality $R(f) \leq \log \mathbf{E}(\tilde{F}) - \log 2d = \lambda(\beta)$, where \tilde{F} is defined in (3.8). Moreover, the inequality is strict whenever $f \neq \mathbf{0}$, since then \tilde{F} is not an almost sure constant. \square

Now we characterize the high and low temperature regimes by the elements of \mathcal{K} and \mathcal{M} , defined by (4.5) and (4.8), respectively. Notice that δ_0 is always an element of \mathcal{K} ; the temperature regime is determined by whether it is also an element of \mathcal{M} .

THEOREM 5.2. *Assume (1.1). The following statements hold:*

- (a) *If $0 \leq \beta \leq \beta_c$, then $\mathcal{K} = \mathcal{M} = \{\delta_0\}$.*
- (b) *If $\beta > \beta_c$, then $\nu(\{f \in \mathcal{S} : \|f\| = 1\}) = 1$ for every $\nu \in \mathcal{M}$, and so \mathcal{T} has more than one fixed point.*

PROOF. Since the hypotheses of (a) and (b) are complementary, it suffices to prove their converses. We always have $\mathcal{T}\delta_0 = \delta_0$. If \mathcal{K} contains no other elements of $\mathcal{P}(\mathcal{S})$, then $\mathcal{M} = \{\delta_0\}$ and Theorem 4.7 shows

$$p(\beta) = \lim_{n \rightarrow \infty} \mathbf{E}(F_n) = \mathcal{R}(\delta_0) = R(\mathbf{0}) = \lambda(\beta).$$

That is, $0 \leq \beta \leq \beta_c$ when $\mathcal{K} = \mathcal{M} = \{\delta_0\}$.

If instead there exists $\nu \in \mathcal{K}$ distinct from δ_0 , then Proposition 4.4 implies that ν assigns positive mass to the set

$$\mathcal{U} := \{f \in \mathcal{S} : \|f\| = 1\},$$

which is measurable by Lemma 2.7. Moreover, Proposition 4.4 guarantees $\nu(\{\mathbf{0}\}) = 1 - \nu(\mathcal{U})$, and Lemma 5.1 shows $R(f) < R(\mathbf{0})$ for all $f \in \mathcal{U}$. It follows that

$$\lim_{n \rightarrow \infty} \mathbf{E}(F_n) \leq \mathcal{R}(\nu) = \int R(f)\nu(df) < R(\mathbf{0}).$$

In this case we have $p(\beta) < \lambda(\beta)$, meaning $\beta > \beta_c$. Furthermore, we can consider the probability measure

$$\nu_{\mathcal{U}}(\mathcal{A}) := \frac{\nu(\mathcal{A} \cap \mathcal{U})}{\nu(\mathcal{U})}, \quad \text{Borel } \mathcal{A} \subset \mathcal{S}.$$

Notice that \mathcal{U} and $\mathcal{S} \setminus \mathcal{U}$ are both closed under \mathcal{T} :

$$\begin{aligned} f \in \mathcal{U} &\Rightarrow \mathcal{T}f(\mathcal{U}) = 1, \\ f \notin \mathcal{U} &\Rightarrow \mathcal{T}f(\mathcal{U}) = 0. \end{aligned}$$

Indeed, this observation was made in the proof of Proposition 4.4. Therefore, $\mathcal{T}\nu = \nu$ implies $\nu_{\mathcal{U}}$ is also an element of \mathcal{K} . If $\nu(\mathcal{U}) < 1$, then $\nu_{\mathcal{U}}$ satisfies

$$\begin{aligned} \mathcal{R}(\nu_{\mathcal{U}}) &= \frac{1}{\nu(\mathcal{U})} \int_{\mathcal{U}} R(f)\nu(df) = \int_{\mathcal{U}} R(f)\nu(df) + \frac{1 - \nu(\mathcal{U})}{\nu(\mathcal{U})} \int_{\mathcal{U}} R(f)\nu(df) \\ &< \int_{\mathcal{U}} R(f)\nu(df) + (1 - \nu(\mathcal{U}))R(\mathbf{0}) = \mathcal{R}(\nu). \end{aligned}$$

It follows that $\nu \in \mathcal{M}$ only if $\nu(\mathcal{U}) = 1$. \square

5.2. *An illustrative application.* To demonstrate how the abstract setup can be employed to prove results on directed polymers, we will use it to prove Theorem D, which is restated here.

THEOREM 5.3. *Assume (1.1). Then the following statements hold:*

(a) *If $0 \leq \beta \leq \beta_c$, then*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \max_{x \in \mathbb{Z}^d} f_i(x) = 0 \quad \text{a.s.}$$

(b) *If $\beta > \beta_c$, then there exists $c > 0$ such that*

$$(5.1) \quad \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \max_{x \in \mathbb{Z}^d} f_i(x) \geq c \quad \text{a.s.}$$

The first step is to define the functional(s) relevant to the problem. At first, definitions are made on the space S . For instance, from (5.1) we are motivated to define $\max : S \rightarrow [0, 1]$ by

$$\max(f) := \max_{u \in \mathbb{N} \times \mathbb{Z}^d} f(u).$$

By appealing to Corollary 2.6, we see that the functional is well defined on the quotient space \mathcal{S} . This should be true in order for the abstract machinery to be applicable. Notice that when f_i is seen as an element of \mathcal{S} ,

$$(5.2) \quad \frac{1}{n} \sum_{i=0}^{n-1} \max_{x \in \mathbb{Z}^d} f_i(x) = \int \max(f) \mu_n(df).$$

The second step is to prove a continuity condition so that Portmanteau (Lemma 2.9) may be applied. Therefore, it is important to understand the topology induced by the metric d . Both here and in later sections, this step is the most technical part of the process, but is generally straightforward and follows a similar proof strategy.

LEMMA 5.4. *The function $\max : \mathcal{S} \rightarrow [0, 1]$ given by*

$$\max(f) := \max_{u \in \mathbb{N} \times \mathbb{Z}^d} f(u)$$

is continuous and thus measurable.

PROOF. Let $f \in \mathcal{S}$ be given, and fix a representative in S . Choose $u \in \mathbb{N} \times \mathbb{Z}^d$ such that $f(u) = \max_{v \in \mathbb{N} \times \mathbb{Z}^d} f(v)$. If $f(u) = 0$, then f is identically zero, and

$$\begin{aligned} d(f, g) < \delta &\Rightarrow \exists \phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d, \quad \sum_{v \in A} g(\phi(v)) + \sum_{v \notin \phi(A)} g(v)^2 < \delta \\ &\Rightarrow \max_{v \in \mathbb{N} \times \mathbb{Z}^d} g(v) < \max\{\delta, \sqrt{\delta}\}. \end{aligned}$$

Otherwise, for any $\delta \in (0, f(u)^2)$, we have

$$\begin{aligned} d(f, g) < \delta &\Rightarrow \exists \phi : A \rightarrow \mathbb{N} \times \mathbb{Z}^d, \quad d_\phi(f, g) < \delta < f(u)^2 \\ &\Rightarrow u \in A \text{ and } |f(v) - g(\phi(v))| < \delta \text{ for all } v \in A, \\ &\quad \text{and } g(v)^2 < \delta \text{ for all } v \notin \phi(A) \\ &\Rightarrow \left| \max_{v \in \mathbb{N} \times \mathbb{Z}^d} g(v) - \max_{v \in \mathbb{N} \times \mathbb{Z}^d} f(v) \right| < \max\{\delta, \sqrt{\delta}\}. \end{aligned}$$

From these two cases, we conclude that $f \mapsto \max f$ is continuous on \mathcal{S} . \square

After the appropriate functionals have been defined, their limiting behavior (in a Cesàro sense) can be determined by applying Theorem 5.2. Consequently, the next step in our program is to study how the functionals of interest behave according to the elements of \mathcal{M} . This will depend on the value of β . For instance, in the high temperature phase, \mathcal{M} consists of only the point mass $\delta_{\mathbf{0}}$ at the zero function $\mathbf{0}$, and so trivially we have

$$(5.3) \quad 0 \leq \beta \leq \beta_c \quad \Rightarrow \quad \int \max(f)v(df) = \max(\mathbf{0}) = 0 \quad \text{for all } v \in \mathcal{M}.$$

On the other hand, in the low temperature phase, every $v \in \mathcal{M}$ is supported on those $f \in \mathcal{S}$ with $\|f\| = 1$. So clearly

$$\beta > \beta_c \quad \Rightarrow \quad \int \max(f)v(df) > 0 \quad \text{for all } v \in \mathcal{M}.$$

Furthermore, because of Lemma 2.9, Lemma 5.4 shows $v \mapsto \int \max(f)v(df)$ is continuous on the compact set \mathcal{M} , and so we actually have

$$(5.4) \quad \beta > \beta_c \quad \Rightarrow \quad \int \max(f)v(df) \geq c \quad \text{for all } v \in \mathcal{M},$$

for some $c > 0$. Now we are poised to prove the desired result. The final step is to interpret the above observations in terms of the directed polymer model, via the almost sure convergence to \mathcal{M} that is stated in Theorem 4.9. It is useful to remember that \mathcal{S} and $\mathcal{P}(\mathcal{S})$ are compact spaces.

PROOF OF THEOREM 5.3. By the (uniform) continuity of the map

$$\mu \mapsto \int \max(f)\mu(df)$$

on the compact space $\mathcal{P}(\mathcal{S})$, we can find for any $\varepsilon > 0$ some $\delta > 0$ such that

$$\begin{aligned} \mathcal{W}(\mu, \mathcal{M}) < \delta \quad \Rightarrow \quad & \inf_{v \in \mathcal{M}} \int \max(f)v(df) - \varepsilon \\ & \leq \int \max(f)\mu(df) \leq \sup_{v \in \mathcal{M}} \int \max(f)v(df) + \varepsilon. \end{aligned}$$

Thus Theorem 4.9 implies that almost surely,

$$\begin{aligned} \inf_{v \in \mathcal{M}} \int \max(f)v(df) & \leq \liminf_{n \rightarrow \infty} \int \max(f)\mu_n(df) \\ & \leq \limsup_{n \rightarrow \infty} \int \max(f)\mu_n(df) \leq \sup_{v \in \mathcal{M}} \int \max(f)v(df). \end{aligned}$$

When $0 \leq \beta \leq \beta_c$, (5.3) says that the infimum and supremum appearing above are both equal to 0, and so

$$(5.5) \quad \lim_{n \rightarrow \infty} \int \max(f)\mu_n(df) = 0 \quad \text{a.s.}$$

When $\beta > \beta_c$, (5.4) shows that the infimum is bounded below by $c > 0$, in which case we have

$$(5.6) \quad \liminf_{n \rightarrow \infty} \int \max(f)\mu_n(df) \geq c \quad \text{a.s.}$$

Recalling (5.2), we see that (5.5) and (5.6) are exactly what we wanted to show. \square

6. Asymptotic pure atomicity. Following the approach outlined in Section 5.2, we will prove that directed polymers are asymptotically purely atomic if and only if the parameter β falls in the low temperature regime. We say that the sequence of endpoint probability mass functions $(f_i)_{i \geq 0}$ is *asymptotically purely atomic* if for every sequence $(\varepsilon_i)_{i \geq 0}$ tending to 0 as $i \rightarrow \infty$, we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^{\varepsilon_i}) = 1 \quad \text{a.s.,}$$

where

$$\mathcal{A}_i^\varepsilon := \{x \in \mathbb{Z}^d : f_i(x) > \varepsilon\}, \quad i \geq 0, \varepsilon > 0.$$

In [85], Vargas defines asymptotic pure atomicity by the same convergence condition, but in probability rather than almost surely. A priori, the definition considered here is a stronger one, but since (6.1) implies the negation of the “in probability” definition, Theorem 6.3 shows the two definitions are equivalent given (1.1).

We begin by making the necessary definitions in the abstract setting.

6.1. *Definitions of relevant functionals.* Observe that the quantity of interest, $\rho_i(\omega_i \in \mathcal{A}_i^\varepsilon)$, is a function of the mass function f_i . Specifically,

$$\rho_i(\omega_i \in \mathcal{A}_i^\varepsilon) = \sum_{x \in \mathbb{Z}^d : f_i(x) > \varepsilon} f_i(x).$$

We are thus motivated to define $\|\cdot\|_\varepsilon : \mathcal{S} \rightarrow [0, 1]$ by

$$\|f\|_\varepsilon := \sum_{u \in \mathbb{N} \times \mathbb{Z}^d : f(u) > \varepsilon} f(u).$$

For any $\varepsilon \in (0, 1)$, the map $f \mapsto \|f\|_\varepsilon$ satisfies (i)–(iii) of Corollary 2.6 and thus induces a well-defined function on \mathcal{S} . To establish asymptotic pure atomicity, it will be important that this function is lower semicontinuous, a fact we prove in the lemma below. Another useful functional will be $\mathbf{I}_\varepsilon : \mathcal{S} \rightarrow \{0, 1\}$ given by

$$\mathbf{I}_\varepsilon(f) = \begin{cases} 1 & \text{if } \max_{u \in \mathbb{N} \times \mathbb{Z}^d} f(u) \geq \varepsilon, \\ 0 & \text{else.} \end{cases}$$

Once more, it is clear that \mathbf{I}_ε satisfies the hypotheses of Corollary 2.6, and so it is well defined on \mathcal{S} .

LEMMA 6.1. *For any $\varepsilon \in (0, 1)$, the following statements hold:*

- (a) *The map $\|\cdot\|_\varepsilon : \mathcal{S} \rightarrow [0, 1]$ is lower semicontinuous and thus measurable.*
- (b) *The map $\mathbf{I}_\varepsilon : \mathcal{S} \rightarrow \{0, 1\}$ is upper semicontinuous and thus measurable.*

PROOF. The proof of (a) is similar to that of Lemma 2.7; see Appendix A.4. For claim (b) we need only to consider the case when $(f_n)_{n \geq 1}$ is a sequence in \mathcal{S} satisfying $\mathbf{I}_\varepsilon(f_n) = 1$ for infinitely many n , and $d(f_n, f) \rightarrow 0$ as $n \rightarrow \infty$. We must show $\mathbf{I}_\varepsilon(f) = 1$. By passing to a subsequence, we may assume $\mathbf{I}_\varepsilon(f_n) = 1$ for all n . That is, $\max(f_n) \geq \varepsilon$ for all n , and so Lemma 5.4 forces $\max(f) \geq \varepsilon$, meaning $\mathbf{I}_\varepsilon(f) = 1$. \square

6.2. *Proof of asymptotic pure atomicity at low temperature.* First, we simplify the problem by providing a sufficient condition for asymptotic pure atomicity. The statement below can be easily verified by the reader.

LEMMA 6.2. *If for every $c < 1$, there is $\varepsilon > 0$ such that*

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^\varepsilon) > c \quad \text{a.s.},$$

then $(f_i)_{i \geq 0}$ is asymptotically purely atomic.

We are now ready to prove Theorem 1.1. For the convenience of the reader, we will restate the result here before giving the proof.

THEOREM 6.3. *Assume (1.1). Then the following statements hold:*

- (a) *If $\beta > \beta_c$, then $(f_i)_{i \geq 0}$ is asymptotically purely atomic.*
- (b) *If $0 \leq \beta \leq \beta_c$, then there is a sequence $(\varepsilon_i)_{i \geq 0}$ tending to 0 as $i \rightarrow \infty$, such that*

$$(6.1) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^{\varepsilon_i}) = 0 \quad \text{a.s.}$$

While the functionals $\|\cdot\|_\varepsilon$ and $\mathcal{I}_\varepsilon(\cdot)$ are not continuous, their semicontinuity will be sufficient to make a limiting argument, thanks to the following generalization of Dini’s first theorem.

LEMMA 6.4. *Let (\mathcal{X}, τ) be a compact metric space. If $(F_n)_{n \geq 1}$ is a nondecreasing sequence of lower semicontinuous functions $\mathcal{X} \rightarrow \mathbb{R}$ converging pointwise to an upper semicontinuous function F , then the convergence is uniform.*

PROOF. Let $G_n := F - F_n$ so that $G_n \searrow 0$ pointwise on \mathcal{X} , and G_n is upper semicontinuous. For given $\varepsilon > 0$, consider the set

$$U_n := \{x \in \mathcal{X} : G_n(x) < \varepsilon\},$$

which is open because G_n is upper semicontinuous. Since $G_n(x) \rightarrow 0$ for every $x \in \mathcal{X}$, we have $\bigcup_n U_n = \mathcal{X}$. By compactness of \mathcal{X} , there is a finite list $n_1 < n_2 < \dots < n_k$ such that $\bigcup_{j=1}^k U_{n_j} = \mathcal{X}$. But the monotonicity assumption guarantees U_n is an ascending sequence. Hence, $\mathcal{X} = \bigcup_{j=1}^k U_{n_j} = U_{n_k} = U_n$ for all $n \geq n_k$, meaning

$$\begin{aligned} n \geq n_k &\Rightarrow U_n = \mathcal{X} \\ &\Rightarrow |F(x) - F_n(x)| = F(x) - F_n(x) < \varepsilon \quad \text{for all } x \in \mathcal{X}. \end{aligned}$$

That is, $F_n \nearrow F$ uniformly on \mathcal{X} . \square

PROOF OF THEOREM 6.3. We first prove (a). For $\varepsilon > 0$, define $\|\cdot\|_\varepsilon$ on $\mathcal{P}(\mathcal{S})$ by

$$\|v\|_\varepsilon := \int \|f\|_\varepsilon v(df) = \int \sum_{u: f(u) > \varepsilon} f(u) v(df).$$

In light of Lemma 2.9, Lemma 6.1(a) implies $\|\cdot\|_\varepsilon : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ is lower semicontinuous. For any $f \in \mathcal{S}$ such that $\|f\| = 1$, we have $\|f\|_\varepsilon \nearrow 1$ as $\varepsilon \rightarrow 0$. In particular, under the

assumption that $\beta > \beta_c$, Theorem 5.2(b) and monotone convergence imply that for any $\nu \in \mathcal{M}$,

$$\|\nu\|_\varepsilon \nearrow 1 \quad \text{as } \varepsilon \rightarrow 0.$$

Since \mathcal{M} is compact, Lemma 6.4 strengthens this pointwise convergence to uniform convergence. That is, for any $c < 1$, there is $\varepsilon > 0$ such that $\|\nu\|_\varepsilon > c$ for all $\nu \in \mathcal{M}$. Furthermore, by compactness of $\mathcal{P}(\mathcal{S})$ and lower semicontinuity of $\|\cdot\|_\varepsilon$, we can find $\delta > 0$ such that for any $\mu \in \mathcal{P}(\mathcal{S})$,

$$\mathcal{W}(\mu, \mathcal{M}) < \delta \quad \Rightarrow \quad \|\mu\|_\varepsilon > c.$$

Now Theorem 4.9 implies

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{A}_i^\varepsilon) = \liminf_{n \rightarrow \infty} \|\mu_n\|_\varepsilon \geq c \quad \text{a.s.}$$

We have thus verified the hypothesis of Lemma 6.2, and so $(f_i)_{i \geq 0}$ is asymptotically purely atomic.

For (b), we assume $0 \leq \beta \leq \beta_c$. For $\varepsilon > 0$, define $\mathcal{I}_\varepsilon : \mathcal{P}(\mathcal{S}) \rightarrow \mathbb{R}$ by

$$\mathcal{I}_\varepsilon(\nu) := \int \mathbf{I}_\varepsilon(f) \nu(\mathrm{d}f) = \int \mathbb{1}_{\{\max_{u \in \mathbb{N} \times \mathbb{Z}^d} f(u) \geq \varepsilon\}} \nu(\mathrm{d}f).$$

Considering Lemma 2.9, we see from Lemma 6.1(b) that \mathcal{I}_ε is an upper semicontinuous map. By Theorem 5.2 and Theorem 4.9, $\mathcal{W}(\mu_n, \delta_0) \rightarrow 0$ almost surely as $n \rightarrow \infty$, and so

$$\limsup_{n \rightarrow \infty} \mathcal{I}_\varepsilon(\mu_n) \leq \mathcal{I}_\varepsilon(\delta_0) = \mathbf{I}_\varepsilon(\mathbf{0}) = 0 \quad \text{a.s. for any } \varepsilon > 0.$$

In particular, for any $j \in \mathbb{N}$,

$$\lim_{N \rightarrow \infty} \mathbf{P} \left(\bigcap_{n=N}^{\infty} \{\mathcal{I}_\varepsilon(\mu_n) < 2^{-j}\} \right) = 1.$$

Now let $(\kappa_j)_{j \geq 1}$ be any decreasing sequence tending to 0 as $j \rightarrow \infty$. Set $M_0 := 0$. By taking complements in the above display, we can inductively choose $M_j > M_{j-1}$ such that

$$(6.2) \quad \mathbf{P} \left(\bigcup_{n=M_j}^{\infty} \{\mathcal{I}_{\kappa_{j+1}}(\mu_n) \geq 2^{-(j+1)}\} \right) < 2^{-(j+1)}, \quad j \geq 1.$$

Define $\varepsilon_i := \kappa_j$ when $M_{j-1} \leq i \leq M_j - 1$. For $M_{\ell-1} \leq n \leq M_\ell$ and any $k < \ell$, the monotonicity of κ_j gives

$$\begin{aligned} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{I}_{\varepsilon_i}(f_i) &= \frac{1}{n} \left[\sum_{i=0}^{M_k-1} \mathbf{I}_{\varepsilon_i}(f_i) + \sum_{j=k+1}^{\ell-1} \sum_{i=M_{j-1}}^{M_j-1} \mathbf{I}_{\kappa_j}(f_i) + \sum_{i=M_{\ell-1}}^{n-1} \mathbf{I}_{\kappa_\ell}(f_i) \right] \\ &\leq \frac{1}{n} \left[\sum_{i=0}^{M_k-1} \mathbf{I}_{\kappa_k}(f_i) + \sum_{j=k+1}^{\ell} \sum_{i=0}^{n-1} \mathbf{I}_{\kappa_j}(f_i) \right] \\ &\leq \sum_{j=k}^{\ell} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{I}_{\kappa_j}(f_i) = \sum_{j=k}^{\ell} \mathcal{I}_{\kappa_j}(\mu_n), \end{aligned}$$

where we have identified f_i with an element of \mathcal{S} (this identification is measurable by Lemma 2.11; cf. the discussion following Corollary 3.3). Writing a more general inequality, we can say that for all $n \geq M_k$,

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbf{I}_{\varepsilon_i}(f_i) \leq \sum_{j \geq k: n \geq M_{j-1}} \mathcal{I}_{\kappa_j}(\mu_n).$$

It follows from a union bound and (6.2) that

$$\begin{aligned} & \mathbf{P} \left(\bigcup_{n=M_k}^{\infty} \left\{ \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{I}_{\varepsilon_i}(f_i) \geq \sum_{j \geq k} 2^{-j} \right\} \right) \\ & \leq \mathbf{P} \left(\bigcup_{n=M_{k-1}}^{\infty} \bigcup_{j \geq k: n \geq M_{j-1}} \{ \mathcal{I}_{\kappa_j}(\mu_n) \geq 2^{-j} \} \right) \\ & = \mathbf{P} \left(\bigcup_{j \geq k} \bigcup_{n=M_{j-1}}^{\infty} \{ \mathcal{I}_{\kappa_j}(\mu_n) \geq 2^{-j} \} \right) \\ & \leq \sum_{j \geq k} \mathbf{P} \left(\bigcup_{n=M_{j-1}}^{\infty} \{ \mathcal{I}_{\kappa_j}(\mu_n) \geq 2^{-j} \} \right) < \sum_{j \geq k} 2^{-j} = 2^{-k+1}. \end{aligned}$$

By Borel–Cantelli, the following is true with probability one: For only finitely many k is

$$\frac{1}{n} \sum_{i=0}^{n-1} \mathbf{I}_{\varepsilon_i}(f_i) \geq 2^{-k+1} \quad \text{for some } n \geq M_k.$$

This implies

$$(6.3) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbf{I}_{\varepsilon_i}(f_i) = 0 \quad \text{a.s.}$$

Finally, note that $\rho_i(\omega_i \in \mathcal{A}_i^\varepsilon)$ is nonzero only when $\mathcal{A}_i^\varepsilon$ is nonempty, in which case $\mathbf{I}_\varepsilon(f_i) = 1$. Hence

$$\rho_i(\omega_i \in \mathcal{A}_i^\varepsilon) \leq \mathbf{I}_\varepsilon(f_i) \quad \text{for any } i \geq 0, \varepsilon > 0,$$

and so (6.1) follows from (6.3). \square

7. Geometric localization. Recall the following definition from Section 1.4. We say that the sequence $(f_i)_{i \geq 0}$ exhibits *geometric localization with positive density* if for every $\delta > 0$, there is $K < \infty$ and $\theta > 0$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{G}_{\delta, K}\}} \geq \theta \quad \text{a.s.,}$$

where

$$(7.1) \quad \mathcal{G}_{\delta, K} := \left\{ f : \mathbb{Z}^d \rightarrow [0, 1] \mid \|f\| = 1, \exists D \subset \mathbb{Z}^d \text{ such that} \right. \\ \left. \text{diam}(D) \leq K \text{ and } \sum_{x \in D} f(x) > 1 - \delta \right\}$$

and

$$\text{diam}(D) := \sup\{\|x - y\|_1 : x, y \in D\}.$$

If K can always be chosen so that $\theta \geq 1 - \delta$, then the sequence is “geometrically localized with full density.” The main goal of this section is to establish that under (1.1), there is positive density geometric localization if and only if $\beta > \beta_c$.

7.1. *Definitions of relevant functionals.* To employ the theory of partitioned subprobability measures, we make a few definitions. For $f \in \mathcal{S}$, let H_f denote \mathbb{N} -support of f :

$$H_f := \{n \in \mathbb{N} : f(n, x) > 0 \text{ for some } x \in \mathbb{Z}^d\}.$$

The first functional of interest is the size of the \mathbb{N} -support of f ,

$$N(f) := |H_f|,$$

which we call the *support number* of f . Second, for $f \in \mathcal{S}$ and $\delta \in (0, 1)$, it is natural in studying geometric localization to consider the quantity

$$W_\delta(f) := \inf \left\{ \text{diam}(D) : D \subset \mathbb{Z}^d, \sum_{x \in D} f(n, x) > 1 - \delta \text{ for some } n \in \mathbb{N} \right\}.$$

In words, $W_\delta(f)$ is the smallest K such that a region of diameter K in $\mathbb{N} \times \mathbb{Z}^d$ has f -mass strictly greater than $1 - \delta$. Here we follow the convention that the infimum of an empty set is infinity, meaning $W_\delta(f) = \infty$ whenever there is no one copy of \mathbb{Z}^d on which f has mass greater than $1 - \delta$. We will write

$$\mathcal{V}_{\delta, K} := \{f \in \mathcal{S} : W_\delta(f) \leq K\},$$

so that we can naturally identify $\mathcal{G}_{\delta, K}$ in \mathcal{S} by

$$(7.2) \quad \mathcal{G}_{\delta, K} = \mathcal{V}_{\delta, K} \cap \{f \in \mathcal{S} : N(f) = 1, \|f\| = 1\}.$$

Next define $q_n : \mathcal{S} \rightarrow [0, 1]$ by

$$q_n(f) := \sum_{x \in \mathbb{Z}^d} f(n, x),$$

and let

$$m(f) := \max_{n \in \mathbb{N}} q_n(f).$$

Finally, it will also be useful to analyze the function

$$Q(f) := \sum_{n \in \mathbb{N}} \frac{q_n(f)}{1 - q_n(f)},$$

where $1/0 = \infty$. Observe that $Q(f) = \infty$ if and only if $N(f) = 1$ and $\|f\| = 1$.

It is easy to see from Corollary 2.6 that N, W_δ, m and Q are well defined on \mathcal{S} . The fact that $N : \mathcal{S} \rightarrow \mathbb{N} \cup \{0, \infty\}$ is measurable is surprisingly nontrivial. The proof, however, is peripheral to our present focus, and so we defer this argument to Appendix A.5. A more central part of our methods is the semicontinuity of relevant functionals, which is the content of the next lemma.

LEMMA 7.1. *The following statements hold:*

- (a) *For any $\delta \in (0, 1)$, $W_\delta : \mathcal{S} \rightarrow \mathbb{N} \cup \{0, \infty\}$ is upper semicontinuous and thus measurable.*
- (b) *$m : \mathcal{S} \rightarrow [0, 1]$ is lower semicontinuous and thus measurable.*
- (c) *$Q : \mathcal{S} \rightarrow [0, \infty]$ is lower semicontinuous and thus measurable.*

These continuity properties can be understood intuitively. For instance, (a) is simply a consequence of the fact that if a measure on \mathbb{Z}^d places mass greater than $1 - \delta$ on a compact set K , then any sufficiently similar measure (in the weak or vague topologies) must do the same. On the other hand, (b) and (c) come from the observation that a converging sequence of partitioned subprobability measures can divide mass on one copy of \mathbb{Z}^d between several copies in the limit when large parts of the mass are drifting away from each other. But the reverse is not true, because vertices in distinct copies of \mathbb{Z}^d are considered infinitely far apart. The proofs that make these ideas precise are tedious and thus postponed to Appendix A.6. They are similar in spirit the proof of Lemma 2.7 but substantially more involved.

An observation that will be useful is that at low temperature, the functional Q must have infinite expectation according to any element of \mathcal{M} .

LEMMA 7.2. *Assume $\beta > \beta_c$. Then for any $\nu \in \mathcal{M}$,*

$$\int Q(f) \nu(df) = \infty.$$

PROOF. Consider any $\nu \in \mathcal{M}$. By Theorem 5.2, the assumption of $\beta > \beta_c$ implies $\nu(\{f \in \mathcal{S} : \|f\| = 1\}) = 1$. Suppose toward a contradiction that $\nu \in \mathcal{M}$ satisfies

$$(7.3) \quad \int Q(f) \nu(df) < \infty.$$

It must then be the case that $\nu(\{f \in \mathcal{S} : N(f) = 1\}) = 0$, since $Q(f) = \infty$ whenever $N(f) = 1$ and $\|f\| = 1$. Let $f \in \mathcal{S}$ be random with law ν . By the previous observation, $1 - q_n(f) > 0$ for all $n \in \mathbb{N}$ with ν -probability 1. Consider an environment $(Y_u)_{u \in \mathbb{N} \times \mathbb{Z}^d}$ of i.i.d. random variables with law \mathcal{L} , that is independent of f . When viewed as an element of \mathcal{S} , the law of the function

$$F(u) := \frac{\sum_{v \sim u} f(v) e^{\beta Y_u}}{\sum_{w \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim w} f(v) e^{\beta Y_w}}, \quad u \in \mathbb{N} \times \mathbb{Z}^d,$$

is $\mathcal{T} \nu = \nu$. Observe that because $\|F\| = 1$,

$$\begin{aligned} Q(F) &= \sum_{n \in \mathbb{N}} \frac{q_n(F)}{1 - q_n(F)} = \sum_{n \in \mathbb{N}} \frac{\sum_{x \in \mathbb{Z}^d} F(n, x)}{\sum_{k \neq n} \sum_{x \in \mathbb{Z}^d} F(k, x)} \\ &= \sum_{n \in \mathbb{N}} \frac{\sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(n, y) e^{\beta Y_{n,x}}}{\sum_{k \neq n} \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(k, y) e^{\beta Y_{k,x}}}. \end{aligned}$$

Conditioned on f , the numerator and denominator of

$$\frac{\sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(n, y) e^{\beta Y_{n,x}}}{\sum_{k \neq n} \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(k, y) e^{\beta Y_{k,x}}}$$

are independent. Hence,

$$\begin{aligned} \mathbf{E}[Q(F) \mid f] &= \sum_{n \in \mathbb{N}} \mathbf{E} \left[\frac{\sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(n, y) e^{\beta Y_{n,x}}}{\sum_{k \neq n} \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(k, y) e^{\beta Y_{k,x}}} \mid f \right] \\ &= \sum_{n \in \mathbb{N}} \mathbf{E} \left[\sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(n, y) e^{\beta Y_{n,x}} \mid f \right] \\ &\quad \cdot \mathbf{E} \left[\frac{1}{\sum_{k \neq n} \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(k, y) e^{\beta Y_{k,x}}} \mid f \right] \end{aligned}$$

$$\begin{aligned} &> \sum_{n \in \mathbb{N}} \frac{\mathbf{E}[\sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(n, y) e^{\beta Y_{n,x}} \mid f]}{\mathbf{E}[\sum_{k \neq n} \sum_{x \in \mathbb{Z}^d} \sum_{y \sim x} f(k, y) e^{\beta Y_{k,x}} \mid f]} \\ &= \sum_{n \in \mathbb{N}} \frac{2d \cdot e^{\lambda(\beta)} \cdot q_n(f)}{2d \cdot e^{\lambda(\beta)} (1 - q_n(f))} = Q(f), \end{aligned}$$

where the strict inequality is due to the strict convexity of $t \mapsto 1/t$ on $(0, \infty)$ and the nondegeneracy of \mathcal{L} . But now (7.3) allows us to write

$$\begin{aligned} \int Q(f) \nu(df) &= \iint Q(F) T f(dF) \nu(df) \\ &= \int \mathbf{E}[Q(F) \mid f] \nu(df) > \int Q(f) \nu(df), \end{aligned}$$

yielding the desired contradiction. \square

7.2. *Proof of geometric localization with positive density.* We can now prove the main theorem of Section 7. This theorem, which encompasses Theorem 1.2, shows that positive density geometric localization is equivalent to $\beta > \beta_c$. It also proves that the single-copy condition (stated as (7.4) below) implies full density geometric localization.

THEOREM 7.3. *Assume (1.1). Let $f_i(\cdot) := \rho_i(\omega_i = \cdot)$. Then the following statements hold:*

(a) *If $\beta > \beta_c$, then $(f_i)_{i \geq 0}$ is geometrically localized with positive density. Moreover, the quantities K and θ in the definition of positive density geometric localization are deterministic, and depend only on the choice of δ , as well as \mathcal{L} , β and d .*

(b) *If $\beta > \beta_c$ and*

$$(7.4) \quad \nu(\mathcal{U}_1) = \nu(\{f \in \mathcal{S} : N(f) = 1\}) = 1 \quad \text{for all } \nu \in \mathcal{M},$$

then $(f_i)_{i \geq 0}$ is geometrically localized with full density.

(c) *If $0 \leq \beta \leq \beta_c$, then for any $\delta \in (0, 1)$ and any K ,*

$$(7.5) \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{G}_{\delta, K}\}} = 0 \quad \text{a.s.}$$

PROOF. Fix $\delta \in (0, 1)$ throughout. Recall that

$$\mathcal{V}_{\delta, K} = \{f \in \mathcal{S} : W_\delta(f) \leq K\} = \{f \in \mathcal{S} : W_\delta(f) < K + 1\},$$

which is open by the upper semicontinuity of W_δ . For (a) and (b), we assume $\beta > \beta_c$. Given $\delta > 0$, define the set

$$\mathcal{U}_\delta := \{f \in \mathcal{S} : m(f) > 1 - \delta\} = \bigcup_{K=0}^\infty \mathcal{V}_{\delta, K}.$$

For (b) only, note the following consequence of Theorem 5.2(b):

$$(7.4) \quad \Rightarrow \quad \nu(\mathcal{U}_\delta) = 1 \quad \text{for all } \nu \in \mathcal{M}.$$

Otherwise, we refer to Lemma 7.2 which tells us that for any $\nu \in \mathcal{M}$,

$$\int Q(f) \nu(df) = \infty.$$

It follows that $\nu(\mathcal{U}_\delta) > 0$, since otherwise the ν -essential supremum of $m(f)$ would be strictly less than 1, forcing the ν -essential supremum of $Q(f)$ to be finite. By Lemma 7.1(b), \mathcal{U}_δ is an open set. Consequently, the map $\mu \mapsto \mu(\mathcal{U}_\delta)$ is lower semicontinuous on $\mathcal{P}(\mathcal{S})$. Since a lower semicontinuous function on a compact set attains its minimum value, we must have

$$\Theta := \inf_{\nu \in \mathcal{M}} \nu(\mathcal{U}_\delta) > 0.$$

For any $f \in \mathcal{U}_\delta$, the quantity $W_\delta(f)$ is finite. Therefore, for any $\nu \in \mathcal{M}$,

$$\nu(\mathcal{V}_{\delta,K}) \wedge \Theta \nearrow \nu(\mathcal{U}_\delta) \wedge \Theta = \Theta \quad \text{as } K \nearrow \infty.$$

Since $\nu \mapsto \nu(\mathcal{V}_{\delta,K}) \wedge \Theta$ is lower semicontinuous, Lemma 6.4 upgrades this convergence to be uniform on the compact set \mathcal{M} . That is, for any $\theta < \Theta$, we can find K large enough that $\nu(\mathcal{V}_{\delta,K}) > \theta$ for every $\nu \in \mathcal{M}$. Furthermore, by compactness of $\mathcal{P}(\mathcal{S})$ there exists $\xi > 0$ such that

$$\mathcal{W}(\mu, \mathcal{M}) < \xi \quad \Rightarrow \quad \mu(\mathcal{V}_{\delta,K}) > \theta.$$

We may now conclude from Theorem 4.9 and (7.2) that

$$\begin{aligned} \theta &\leq \liminf_{n \rightarrow \infty} \mu_n(\mathcal{V}_{\delta,K}) = \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{V}_{\delta,K}\}} \\ &= \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{G}_{\delta,K}\}} \quad \text{a.s.,} \end{aligned}$$

which completes the proof of (a) and (b).

For claim (c), suppose $0 \leq \beta \leq \beta_c$ so that Theorem 5.2 gives $\mathcal{M} = \{\delta_0\}$. Fix $K > 0$. Recall from the proof of Theorem 5.3 that in this high temperature case,

$$(7.6) \quad \lim_{n \rightarrow \infty} \int \max(f) \mu_n(df) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \max_{x \in \mathbb{Z}^d} f_i(x) = 0 \quad \text{a.s.}$$

Notice that if $\varepsilon > 0$ is sufficiently small that

$$(7.7) \quad D \subset \mathbb{Z}^d, \quad \text{diam}(D) \leq K \quad \Rightarrow \quad \varepsilon|D| < 1 - \delta,$$

then

$$(7.8) \quad \frac{1}{n} \sum_{i=0}^{n-1} \max_{x \in \mathbb{Z}^d} f_i(x) < \varepsilon^2 \quad \Rightarrow \quad \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{G}_{\delta,K}\}} < \varepsilon.$$

Indeed, the hypothesis in (7.8) implies there are fewer than εn numbers i between 0 and $n - 1$ such that $\max_{x \in \mathbb{Z}^d} f_i(x) \geq \varepsilon$, and (7.7) implies all the remaining i must satisfy $f_i \notin \mathcal{G}_{\delta,K}$. Therefore, (7.8) is true, and so (7.6) implies (7.5). \square

7.3. Localization in a favorite region. In this final section, we prove that single-copy condition (7.4) implies localization in a “favorite region” of size $O(1)$. Recall that the mode of a probability mass function is a location where the function attains its maximum. For any $n \geq 0$ and $K \geq 0$, let \mathcal{C}_n^K be the set of all $x \in \mathbb{Z}^d$ that are within ℓ^1 distance K from every mode of the random probability mass function f_n . Note that \mathcal{C}_n^K is a set with diameter at most $2K$. The following theorem establishes localization of the endpoint in \mathcal{C}_n^K , if the single-copy condition holds.

PROPOSITION 7.4. *Assume (1.1) and (7.4). Then*

$$(7.9) \quad \lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{C}_i^K) = 1 \quad \text{a.s.}$$

PROOF. For the sake of completeness, we first verify that $\rho_i(\omega_i \in \mathcal{C}_i^K)$ is a measurable function (with respect to \mathcal{F}_i). For any Borel set $B \subset \mathbb{R}$, the event $\{\rho_i(\omega_i \in \mathcal{C}_i^K) \in B\}$ can be expressed as

$$\bigcup_{\substack{A \subset \mathbb{Z}^d \\ |A| \leq (2d)^i}} \left[\left(\bigcap_{x \in A} \{f_i(x) = \max_{y \in \mathbb{Z}^d} f_i(y)\} \right) \cap \left\{ \sum_{y \in \mathbb{Z}^d: \|x-y\|_1 \leq K \forall x \in A} f_i(y) \in B \right\} \right].$$

It is clear that the above display is a measurable event, and so $\rho_i(\omega_i \in \mathcal{C}_i^K)$ is measurable.

Assume (7.4), so that $(f_i)_{i \geq 0}$ is geometrically localized with full density. (In particular, by Theorem 7.3(c), we must have $\beta > \beta_c$.) As in the proof of Theorem 6.3(a), the assumption $\beta > \beta_c$ ensures that the conclusion of Lemma 6.2 holds. Consequently, given any $\theta < 1$, we can choose $\varepsilon \in (0, 1 - \theta)$ such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{\mathcal{A}_i^\varepsilon \text{ is nonempty}\}} > \theta \quad \text{a.s.}$$

And by geometric localization with full density, there is K such that

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{G}_{\varepsilon, K}\}} > \theta \quad \text{a.s.}$$

Therefore, there almost surely exists some N satisfying

$$n \geq N \quad \Rightarrow \quad \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{\mathcal{A}_i^\varepsilon \text{ is nonempty}\}} > \theta \quad \text{and} \quad \frac{1}{n} \sum_{i=0}^{n-1} \mathbb{1}_{\{f_i \in \mathcal{G}_{\varepsilon, K}\}} > \theta.$$

So for $n \geq N$, there are at least $(2\theta - 1)n$ numbers i between 0 and $n - 1$ for which both of the following statements are true: First, $f_i(x) > \varepsilon$ for some (and thus any) mode $x \in \mathbb{Z}^d$ of f_i . Second, $\rho_i(\omega_i \in D_i) > 1 - \varepsilon$ for some $D_i \subset \mathbb{Z}^d$ with $\text{diam}(D_i) \leq K$. For such i , all modes must belong to D_i , which implies $D_i \subset \mathcal{C}_i^K$. In particular,

$$\rho_i(\omega_i \in \mathcal{C}_i^K) \geq \rho_i(\omega_i \in D_i) > 1 - \varepsilon > \theta,$$

and so

$$n \geq N \quad \Rightarrow \quad \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{C}_i^K) > \theta(2\theta - 1).$$

As $\rho_i(\omega_i \in \mathcal{C}_i^L) \geq \rho_i(\omega_i \in \mathcal{C}_i^K)$ for $L \geq K$, we in fact have

$$\lim_{K \rightarrow \infty} \liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \rho_i(\omega_i \in \mathcal{C}_i^K) \geq \theta(2\theta - 1) \quad \text{a.s.}$$

As $\theta < 1$ is arbitrary, we can take a countable sequence $\theta_k \rightarrow 1$ to conclude (7.9). \square

8. Proof of Proposition 3.2. Since (\mathcal{S}, d) is a compact metric space, uniform continuity of $f \mapsto Tf$ and $f \mapsto \mathbf{E}(\log^q \tilde{F})$ will be implied by continuity. So it suffices to prove continuity at a fixed $f \in \mathcal{S}$. Given $\varepsilon > 0$, choose $\kappa_1 > 0$ small enough that

$$(8.1) \quad ((2d)^{-1}\kappa_1 + d^{-1/2}\sqrt{2\kappa_1})\sqrt{e^{\lambda(2\beta)}e^{\lambda(-2\beta)}} < \frac{\varepsilon}{8}.$$

Next choose $\kappa_2 > 0$ small enough that all of the following statements are true:

$$(8.2a) \quad \kappa_2 e^{\lambda(2\beta)} e^{\lambda(-2\beta)} < \frac{\varepsilon}{4},$$

$$(8.2b) \quad (4d + 1)\kappa_2 < \kappa_1,$$

$$(8.2c) \quad 2d\sqrt{\kappa_2} < \kappa_1.$$

Given a representative $f \in \mathcal{S}$, we may take $A \subset \mathbb{N} \times \mathbb{Z}^d$ to be finite but large enough that

$$(8.3) \quad \sum_{u \notin A} f(u) < \kappa_2.$$

By possibly omitting some elements of A , we may assume $f(u) > 0$ for all $u \in A$, and then take $\delta > 0$ small enough that

$$(8.4) \quad f(u)^2 > \delta \quad \text{for all } u \in A.$$

Considering (8.2), we may choose δ so that all of the following statements are also true:

$$(8.5a) \quad (\delta + \kappa_2)e^{\lambda(2\beta)} e^{\lambda(-2\beta)} < \frac{\varepsilon}{4},$$

$$(8.5b) \quad (4d + 1)(\delta + \kappa_2) < \kappa_1,$$

$$(8.5c) \quad 2d\sqrt{\delta + \kappa_2} < \kappa_1.$$

In addition, we will assume

$$(8.5d) \quad (2d)^2|A|\sqrt{\delta} < \kappa_2,$$

and finally

$$(8.5e) \quad \delta < \min\left(\frac{\varepsilon}{16}, 2^{-3}\right).$$

Now suppose $g \in \mathcal{S}$ satisfies $d(f, g) < \delta$. We will show $\mathcal{W}(Tf, Tg) < \varepsilon$.

Given a representative $g \in \mathcal{S}$, we can choose an isometry $\psi : C \rightarrow \mathbb{N} \times \mathbb{Z}^d$ such that $d_\psi(f, g) < \delta$. From (8.4), we deduce that $A \subset C$, since otherwise $d_\psi(f, g) \geq f(u)^2 > \delta$ for some $u \in A \setminus C$. Let ϕ be the restriction of ψ to A . It follows that

$$(8.6) \quad \sum_{v \in A} |f(v) - g(\phi(v))| \leq \sum_{v \in C} |f(v) - g(\psi(v))| < \delta.$$

Since $d_\psi(f, g) < \delta < 2^{-3}$, we necessarily have $\deg(\phi) \geq \deg(\psi) \geq 4$. Let $\Phi : B \rightarrow \mathbb{N} \times \mathbb{Z}^d$ be defined as in Lemma 2.2 so that $\deg(\Phi) \geq 2$. Now take $(Y_u)_{u \in \mathbb{Z}^d}$ to be i.i.d. random variables with shared law \mathfrak{L} . Define $Z_u := Y_{\Phi^{-1}(u)}$ if $u \in \Phi(B)$; otherwise let Z_u be an independent copy of Y_u . Define $F, G \in \mathcal{S}$ as in Proposition 3.1 so that the laws of F and G are Tf and Tg , respectively, and we may consider the coupling (F, G) to determine an upper bound on $\mathcal{W}(Tf, Tg)$. That is, $\mathcal{W}(Tf, Tg) \leq \mathbf{E}[d(F, G)] \leq \mathbf{E}[d_\Phi(F, G)]$.

To simplify notation, we will write

$$\begin{aligned} \tilde{f}(u) &= \sum_{v \sim u} f(v)e^{\beta Y_u}, & \tilde{F} &= \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \tilde{f}(u) + 2d(1 - \|f\|)e^{\lambda(\beta)}, \\ \tilde{g}(u) &= \sum_{v \sim u} g(v)e^{\beta Z_u}, & \tilde{G} &= \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \tilde{g}(u) + 2d(1 - \|g\|)e^{\lambda(\beta)} \end{aligned}$$

so that $F(u) = \tilde{f}(u)/\tilde{F}$ and $G(u) = \tilde{g}(u)/\tilde{G}$. For any $u \in B$,

$$\begin{aligned} \tilde{f}(u) - \tilde{g}(\Phi(u)) &= \sum_{v \sim u} f(v)e^{\beta Y_u} - \sum_{v \sim \Phi(u)} g(v)e^{\beta Z_{\Phi(u)}} \\ (8.7) \qquad &= e^{\beta Y_u} \left(\sum_{v \sim u} f(v) - \sum_{v \sim \Phi(u)} g(v) \right). \end{aligned}$$

Since $\text{deg}(\Phi) \geq 2$, the two sets

$$R_u := \{v : v \sim u, v \in A\} \quad \text{and} \quad R'_u := \{v : v \sim \Phi(u), v \in \phi(A)\}$$

satisfy $\phi(R_u) = R'_u$. Indeed, if $v \in R_u$, then $\Phi(u) - \phi(v) = \Phi(u) - \Phi(v) = u - v$, and so $\phi(v)$ belongs to R'_u . Conversely, if $\phi(v) = \Phi(v) \in R'_u$, then $u - v = \Phi(u) - \Phi(v) = \Phi(u) - \phi(v)$, and so v belongs to R_u . We also define

$$S_u := \{v : v \sim u, v \notin A\} \quad \text{and} \quad S'_u := \{v : v \sim \Phi(u), v \notin \psi(C)\},$$

and finally

$$T'_u := \{v : v \sim \Phi(u), v \in \psi(C \setminus A)\}.$$

We thus have

$$\begin{aligned} \left| \sum_{v \sim u} f(v) - \sum_{v \sim \Phi(u)} g(v) \right| &\leq \sum_{v \in R_u} |f(v) - g(\phi(v))| + \sum_{v \in S_u} f(v) \\ (8.8) \qquad &+ \sum_{v \in S'_u} g(v) + \sum_{v \in T'_u} g(v). \end{aligned}$$

When summed over $u \in B$, each of the four terms on the right-hand side of (8.8) can be separately bounded from above. We do so in the next paragraph.

Any $v \in \mathbb{N} \times \mathbb{Z}^d$ is the neighbor of at most $2d$ elements of B . Consequently, the inclusion $R_u \subset A$ and the assumption (8.6) together imply

$$(8.9) \qquad \sum_{u \in B} \sum_{v \in R_u} |f(v) - g(\phi(v))| \leq 2d \cdot \delta.$$

Next, $S_u \subset A^c$ and (8.3) imply

$$(8.10) \qquad \sum_{u \in B} \sum_{v \in S_u} f(v) \leq 2d \cdot \kappa_2.$$

Since $|B| \leq 2d|A|$, the inclusion $S'_u \subset \psi(C)^c$ and (8.5d) imply

$$\begin{aligned} \sum_{u \in B} \sum_{v \in S'_u} g(v) &\leq (2d)^2|A| \sup_{v \notin \psi(C)} g(v) \leq (2d)^2|A| \sqrt{\sum_{v \notin \psi(C)} g(v)^2} \\ (8.11) \qquad &\leq (2d)^2|A| \sqrt{d_\psi(f, g)} < (2d)^2|A| \sqrt{\delta} < \kappa_2. \end{aligned}$$

Finally, one more application of (8.3) gives

$$\begin{aligned}
 \sum_{u \in B} \sum_{v \in T'_u} g(v) &\leq \sum_{u \in B} \sum_{v \in T'_u} (|f(\psi^{-1}(v)) - g(v)| + f(\psi^{-1}(v))) \\
 (8.12) \qquad &\leq 2d \sum_{u \in C \setminus A} |f(u) - g(\psi(u))| + 2d \sum_{u \notin A} f(u) \\
 &< 2d(d_\psi(f, g) + \kappa_2) < 2d(\delta + \kappa_2).
 \end{aligned}$$

Having established (8.8)–(8.12) and assumed (8.5b), we have shown

$$(8.13) \qquad \sum_{u \in B} \left| \sum_{v \sim u} f(v) - \sum_{v \sim \Phi(u)} g(v) \right| < (4d + 1)(\delta + \kappa_2) < \kappa_1.$$

This inequality will be pivotal in obtaining an upper bound on $\mathbf{E}[d_\Phi(F, G)]$.

Four other inequalities stated below will also be useful, namely (8.14), (8.17), (8.18) and (8.19). Let us now prove each of them. First, for any $u \notin B$, an application of Cauchy–Schwarz gives

$$\sum_{u \notin B} \tilde{f}(u)^2 = \sum_{u \notin B} \left(\sum_{v \sim u} f(v) e^{\beta Y_u} \right)^2 \leq \sum_{u \notin B} 2d \cdot e^{2\beta Y_u} \sum_{v \sim u} f(v)^2.$$

Notice that if $u \notin B$, and $v \sim u$, then $v \notin A$. It follows that

$$\begin{aligned}
 (8.14) \qquad \mathbf{E} \left[\sum_{u \notin B} \tilde{f}(u)^2 \right] &\leq 2d \cdot e^{\lambda(2\beta)} \sum_{u \notin B} \sum_{v \sim u} f(v)^2 \\
 &\leq (2d)^2 \cdot e^{\lambda(2\beta)} \sum_{v \notin A} f(v)^2 \\
 &\leq (2d)^2 \cdot e^{\lambda(2\beta)} \sum_{v \notin A} f(v) < \kappa_2 (2d)^2 e^{\lambda(2\beta)},
 \end{aligned}$$

where we have used (8.3) in the last inequality. Since $\text{deg}(\Phi) \geq 2$, the same reasoning as above gives

$$(8.15) \qquad \mathbf{E} \left[\sum_{u \notin \Phi(B)} \tilde{g}(u)^2 \right] \leq (2d)^2 e^{\lambda(2\beta)} \sum_{v \notin \Phi(A)} g(v)^2,$$

and we can bound the last sum by again applying (8.3):

$$\begin{aligned}
 \sum_{v \notin \Phi(A)} g(v)^2 &\leq \sum_{v \in \psi(C \setminus A)} g(v) + \sum_{v \notin \psi(C)} g(v)^2 \\
 (8.16) \qquad &\leq \sum_{v \in C \setminus A} (|f(v) - g(\psi(v))| + f(v)) + \sum_{v \notin \psi(C)} g(v)^2 \\
 &\leq \sum_{v \in C} |f(v) - g(\psi(v))| + \sum_{v \notin \psi(C)} g(v)^2 + \sum_{v \notin A} f(v) \\
 &< d_\psi(f, g) + \kappa_2 < \delta + \kappa_2.
 \end{aligned}$$

In light of (8.16), now (8.15) becomes

$$(8.17) \qquad \mathbf{E} \left[\sum_{u \notin \Phi(B)} \tilde{g}(u)^2 \right] < (\delta + \kappa_2) (2d)^2 e^{\lambda(2\beta)}.$$

The third and fourth inequalities to show consider the quantities \tilde{F} and \tilde{G} . Suppose $q \in (-\infty, 0] \cup [1, \infty]$. Since

$$\sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v) + 2d(1 - \|f\|) = 2d,$$

Jensen’s inequality applied to the convex function $t \mapsto t^q$ gives

$$\begin{aligned} \mathbf{E}(\tilde{F}^q) &\leq (2d)^q \mathbf{E} \left[\frac{1}{2d} \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v) e^{q\beta Y_u} + (1 - \|f\|) e^{q\lambda(\beta)} \right] \\ (8.18) \quad &= (2d)^q (\|f\| e^{\lambda(q\beta)} + (1 - \|f\|) e^{q\lambda(\beta)}) \leq (2d)^q e^{\lambda(q\beta)}. \end{aligned}$$

By the same argument with g in place of f , we obtain the fourth desired inequality:

$$(8.19) \quad \mathbf{E}(\tilde{G}^q) \leq (2d)^q e^{\lambda(q\beta)}.$$

We will only use (8.18) and (8.19) when $|q| \leq 2$, in accordance with our assumption (1.1).

We now turn in earnest to bounding $\mathbf{E}[d_\Phi(F, G)]$, which is the sum of four expectations corresponding to the four summands in definition (2.3). Seeking an estimate on the first summand, we observe that for $u \in B$,

$$\begin{aligned} |F(u) - G(\Phi(u))| &= \left| \frac{\tilde{f}(u)}{\tilde{F}} - \frac{\tilde{g}(\Phi(u))}{\tilde{G}} \right| \\ (8.20) \quad &\leq \left| \frac{\tilde{f}(u)}{\tilde{F}} - \frac{\tilde{g}(\Phi(u))}{\tilde{F}} \right| + \left| \frac{\tilde{g}(\Phi(u))}{\tilde{F}} - \frac{\tilde{g}(\Phi(u))}{\tilde{G}} \right| \\ &= \frac{|\tilde{f}(u) - \tilde{g}(\Phi(u))|}{\tilde{F}} + \frac{\tilde{g}(\Phi(u))}{\tilde{G}} \cdot \left| \frac{\tilde{G}}{\tilde{F}} - 1 \right|. \end{aligned}$$

Summing over B and taking expectation, we have for the first term in (8.20)

$$\begin{aligned} \mathbf{E} \sum_{u \in B} \frac{|\tilde{f}(u) - \tilde{g}(\Phi(u))|}{\tilde{F}} &= \sum_{u \in B} \mathbf{E} \left[\frac{|\tilde{f}(u) - \tilde{g}(\Phi(u))|}{\tilde{F}} \right] \\ (8.21) \quad &\leq \sum_{u \in B} \sqrt{\mathbf{E}(|\tilde{f}(u) - \tilde{g}(\Phi(u))|^2) \mathbf{E}(\tilde{F}^{-2})} \\ &\leq (2d)^{-1} \sum_{u \in B} \sqrt{e^{\lambda(2\beta)} e^{\lambda(-2\beta)}} \left| \sum_{v \sim u} f(v) - \sum_{v \sim \Phi(u)} g(v) \right| \\ &< (2d)^{-1} \kappa_1 \sqrt{e^{\lambda(2\beta)} e^{\lambda(-2\beta)}}, \end{aligned}$$

where we have applied Cauchy–Schwarz to obtain the first inequality, (8.7) and (8.18) to obtain the second, and (8.13) for the third. Meanwhile, the second term in (8.20) satisfies

$$\begin{aligned} \mathbf{E} \left[\sum_{u \in B} \frac{\tilde{g}(\Phi(u))}{\tilde{G}} \cdot \left| \frac{\tilde{G}}{\tilde{F}} - 1 \right| \right] &\leq \mathbf{E} \left| \frac{\tilde{G}}{\tilde{F}} - 1 \right| = \mathbf{E} \left| \frac{\tilde{G} - \tilde{F}}{\tilde{F}} \right| \\ (8.22) \quad &\leq \sqrt{\mathbf{E}(\tilde{F}^{-2}) \mathbf{E}(|\tilde{G} - \tilde{F}|^2)} \\ &\leq (2d)^{-1} \sqrt{e^{\lambda(-2\beta)} \mathbf{E}(|\tilde{G} - \tilde{F}|^2)}, \end{aligned}$$

where we have again applied (8.18). Notice that

$$\tilde{F} - 2d \cdot e^{\lambda(\beta)} = \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \tilde{f}(u) - 2d \cdot e^{\lambda(\beta)} \|f\|$$

$$\begin{aligned}
 &= \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v) e^{\beta Y_u} - \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v) e^{\lambda(\beta)} \\
 &= \sum_{u \in \mathbb{N} \times \mathbb{Z}^d} \sum_{v \sim u} f(v) [e^{\beta Y_u} - e^{\lambda(\beta)}] \\
 &= \sum_{u \in B} \sum_{v \sim u} f(v) [e^{\beta Y_u} - e^{\lambda(\beta)}] + \sum_{u \notin B} \sum_{v \sim u} f(v) [e^{\beta Y_u} - e^{\lambda(\beta)}],
 \end{aligned}$$

and similarly

$$\begin{aligned}
 &\tilde{G} - 2d \cdot e^{\lambda(\beta)} \\
 &= \sum_{u \in \Phi(B)} \sum_{v \sim u} g(v) [e^{\beta Z_u} - e^{\lambda(\beta)}] + \sum_{u \notin \Phi(B)} \sum_{v \sim u} g(v) [e^{\beta Z_u} - e^{\lambda(\beta)}] \\
 &= \sum_{u \in B} \sum_{v \sim \Phi(u)} g(v) [e^{\beta Y_u} - e^{\lambda(\beta)}] + \sum_{u \notin \Phi(B)} \sum_{v \sim u} g(v) [e^{\beta Z_u} - e^{\lambda(\beta)}].
 \end{aligned}$$

Hence,

$$\begin{aligned}
 \tilde{F} - \tilde{G} &= \sum_{u \in B} \left(\sum_{v \sim u} f(v) - \sum_{v \sim \Phi(u)} g(v) \right) [e^{\beta Y_u} - e^{\lambda(\beta)}] \\
 &\quad + \sum_{u \notin B} \sum_{v \sim u} f(v) [e^{\beta Y_u} - e^{\lambda(\beta)}] - \sum_{u \notin \Phi(B)} \sum_{v \sim u} g(v) [e^{\beta Z_u} - e^{\lambda(\beta)}].
 \end{aligned}$$

We thus consider the following i.i.d. random variables:

$$\begin{aligned}
 Z_u &:= e^{\beta Y_u} - e^{\lambda(\beta)}, & u \in B, \\
 Z'_u &:= e^{\beta Y_u} - e^{\lambda(\beta)}, & u \notin B, \\
 Z''_u &:= e^{\beta Z_u} - e^{\lambda(\beta)}, & u \notin \Phi(B).
 \end{aligned}$$

Furthermore, their coefficients

$$\begin{aligned}
 \alpha_u &:= \sum_{v \sim u} f(v) - \sum_{v \sim \Phi(u)} g(v), & u \in B, \\
 \alpha'_u &:= \sum_{v \sim u} f(v), & u \notin B, \\
 \alpha''_u &:= - \sum_{v \sim u} g(v), & u \notin \Phi(B)
 \end{aligned}$$

satisfy

$$\begin{aligned}
 |\alpha_u| &< \kappa_1 \quad \text{by (8.13),} \\
 |\alpha'_u| = \alpha'_u &\leq \sum_{v \notin A} f(v) < \kappa_2 < \kappa_1 \quad \text{by (8.3) and (8.2b),} \\
 |\alpha''_u| = -\alpha''_u &\leq 2d \sup_{v \notin \phi(A)} g(v) \leq 2d \sqrt{\delta + \kappa_2} < \kappa_1 \quad \text{by (8.16) and (8.5c).}
 \end{aligned}$$

Since $\|f\|, \|g\| \leq 1$, one also has

$$\sum_{u \in B} |\alpha_u| + \sum_{u \notin B} \alpha'_u - \sum_{u \notin \Phi(B)} \alpha''_u \leq (1 + 1) \cdot 2d + 2d + 2d = 8d.$$

Therefore,

$$\begin{aligned} \mathbf{E}[|\tilde{F} - \tilde{G}|^2] &= \mathbf{E}\left[\left(\sum_{u \in B} \alpha_u Z_u + \sum_{u \notin B} \alpha'_u Z'_u + \sum_{u \notin \Phi(B)} \alpha''_u Z''_u\right)^2\right] \\ &= \left(\sum_{u \in B} \alpha_u^2 + \sum_{u \notin B} (\alpha'_u)^2 + \sum_{u \notin \Phi(B)} (\alpha''_u)^2\right) (e^{\lambda(2\beta)} - e^{2\lambda(\beta)}) \\ &\leq \kappa_1 \left(\sum_{u \in B} |\alpha_u| + \sum_{u \notin B} \alpha'_u - \sum_{u \notin \Phi(B)} \alpha''_u\right) (e^{\lambda(2\beta)} - e^{2\lambda(\beta)}) \\ &\leq \kappa_1 \cdot 8d \cdot e^{\lambda(2\beta)}. \end{aligned}$$

Upon using the above display in (8.22), we obtain

$$(8.23) \quad \mathbf{E}\left[\sum_{u \in B} \frac{\tilde{g}(\Phi(u))}{\tilde{G}} \cdot \left|\frac{\tilde{G}}{\tilde{F}} - 1\right|\right] \leq \mathbf{E}\left|\frac{\tilde{G}}{\tilde{F}} - 1\right| \leq d^{-1/2} \sqrt{2\kappa_1 e^{\lambda(2\beta)} e^{\lambda(-2\beta)}}.$$

Then applying (8.21) and (8.23) with (8.20), we find

$$\mathbf{E}\left[\sum_{u \in B} |F(u) - G(\Phi(u))|\right] \leq ((2d)^{-1} \kappa_1 + d^{-1/2} \sqrt{2\kappa_1}) \sqrt{e^{\lambda(2\beta)} e^{\lambda(-2\beta)}}.$$

Now (8.1) gives us control on the first summand in $\mathbf{E}[d_\Phi(F, G)]$:

$$(8.24) \quad \mathbf{E}\left[2 \sum_{u \in B} |F(u) - G(\Phi(u))|\right] < \frac{\varepsilon}{4}.$$

From what we have argued, the second and third summands in $\mathbf{E}[d_\Phi(F, G)]$ are not difficult to bound. Notice that for any $u \in \mathbb{N} \times \mathbb{Z}^d$, the random variables $\tilde{f}(u)$, $\tilde{g}(u)$, \tilde{F} and \tilde{G} are nondecreasing functions of Y_u or Z_u . Therefore, we can use the FKG inequality. For the second summand, we apply FKG, then (8.14) and (8.18), and finally (8.2a):

$$(8.25) \quad \mathbf{E} \sum_{u \notin B} F(u)^2 \leq \mathbf{E}\left[\sum_{u \notin B} \tilde{f}(u)^2\right] \mathbf{E}(\tilde{F}^{-2}) < \kappa_2 e^{\lambda(2\beta)} e^{\lambda(-2\beta)} < \frac{\varepsilon}{4}.$$

For the third summand, we appeal to FKG, then to (8.17) and (8.19), and finally to (8.5a):

$$(8.26) \quad \begin{aligned} \mathbf{E} \sum_{u \notin \Phi(B)} G(u)^2 &\leq \mathbf{E}\left[\sum_{u \notin \Phi(B)} \tilde{g}(u)^2\right] \mathbf{E}(\tilde{G}^{-2}) \\ &< (\delta + \kappa_2) e^{\lambda(2\beta)} e^{\lambda(-2\beta)} < \frac{\varepsilon}{4}. \end{aligned}$$

The fourth and final summand in $\mathbf{E}[d_\Phi(F, G)]$ depends on the maximum degree of the isometry Φ . From (8.5e) we know $d_\psi(f, g) < \delta < \varepsilon/16$, and so $2^{-\deg(\psi)} < \varepsilon/16$. Since ϕ is a restriction of ψ , we have $\deg(\phi) \geq \deg(\psi)$. And by Lemma 2.2, $\deg(\Phi) \geq \deg(\phi) - 2$. As a result,

$$(8.27) \quad 2^{-\deg(\Phi)} \leq 2^{-\deg(\phi)+2} \leq 4 \cdot 2^{-\deg(\psi)} < \frac{\varepsilon}{4}.$$

Summing the left-hand sides of (8.24)–(8.27), we reach the desired conclusion for part (a):

$$\mathcal{W}(Tf, Tg) \leq \mathbf{E}[d_\Phi(F, G)] < \varepsilon.$$

For part (b), we assume q is a positive integer. By several applications of Cauchy–Schwarz, we have

$$\begin{aligned}
 (8.28) \quad & |\mathbf{E}(\log^q \tilde{F}) - \mathbf{E}(\log^q \tilde{G})| \leq \mathbf{E}|\log^q \tilde{F} - \log^q \tilde{G}| \\
 & = \mathbf{E} \left| \left(\log \frac{\tilde{F}}{\tilde{G}} \right) \sum_{i=0}^{q-1} (\log^{q-1-i} \tilde{F})(\log^i \tilde{G}) \right| \\
 & \leq \sqrt{\mathbf{E} \log^2 \frac{\tilde{F}}{\tilde{G}}} \sqrt{\mathbf{E} \left[\left(\sum_{i=0}^{q-1} (\log^{q-1-i} \tilde{F})(\log^i \tilde{G}) \right)^2 \right]} \\
 & \leq \sqrt{\mathbf{E} \log^2 \frac{\tilde{F}}{\tilde{G}}} \sqrt{q \sum_{i=0}^{q-1} \sqrt{\mathbf{E}(\log^{4(q-1-i)} \tilde{F}) \mathbf{E}(\log^{4i} \tilde{G})}}.
 \end{aligned}$$

Now, by the inequality $\log^2 x \leq 2(|x - 1| + |x^{-1} - 1|)$ for $x > 0$, we have

$$\mathbf{E} \log^2 \frac{\tilde{F}}{\tilde{G}} \leq 2 \left(\mathbf{E} \left| \frac{\tilde{F}}{\tilde{G}} - 1 \right| + \mathbf{E} \left| \frac{\tilde{G}}{\tilde{F}} - 1 \right| \right).$$

From (8.23) and (8.1), we know $\mathbf{E}|\frac{\tilde{G}}{\tilde{F}} - 1| < \varepsilon/8$. A parallel analysis—exchange $\mathbf{E}(\tilde{F}^{-2})$ and $\mathbf{E}(\tilde{G}^{-2})$ in the second expression of (8.22), and then apply (8.19) instead of (8.18)—also gives $\mathbf{E}|\frac{\tilde{F}}{\tilde{G}} - 1| < \varepsilon/8$. On the other hand, for each integer $i \geq 0$, there exists a large enough constant C_i so that $\log^{4i} x \leq C_i(x + x^{-1})$ for all $x > 0$. Therefore, we can appeal to (8.18) and (8.19) with $q = \pm 1$ in order to upper bound the second square root in the final expression of (8.28). In summary,

$$|\mathbf{E}(\log^q \tilde{F}) - \mathbf{E}(\log^q \tilde{G})| \leq C \frac{\sqrt{\varepsilon}}{2},$$

where C is a constant that depends only on \mathfrak{L} , β , d and q . The continuity of the map $f \mapsto \mathbf{E}(\log^q \tilde{F})$ is now clear.

9. Open problems. Given the subsequent applications [15, 51] of Mukherjee and Varadhan’s construction in [61] and of a variant construction in [60], as well as the success of the concentration compactness phenomenon in general, we are optimistic that the ideas presented in this study can fuel future work. In fact, after the preprint release of this manuscript, our “cavity approach” inspired similar works by Bröker and Mukherjee for the stochastic heat equation and Gaussian multiplicative chaos [21], and by Bakhtin and Seo [8] for directed polymers with a continuous-space reference walk.

Below we mention some remaining questions concerning the methods of this manuscript. For a broader discussion of open problems on directed polymers, we refer the reader to Section 12.9 of [37].

1. For the $(1 + 1)$ -dimensional log-gamma model, localization occurs around a single favorite region [27]. Does this hold more generally, specifically in the low temperature regime? We showed via Theorem 7.3(b) and Proposition 7.4 that the answer is yes if the single-copy condition holds. Nevertheless, it is challenging to rule out the possibility that the endpoint distribution maintains an unbounded number of favorite sites, so that some $\nu \in \mathcal{M}$ may put mass on partitioned subprobability measures supported on infinitely many copies of \mathbb{Z}^d . Such behavior has been observed for other types of Gibbs measures [9, 26].

2. The set \mathcal{M} defined by (4.8) is closed and convex in $\mathcal{P}(\mathcal{S})$. By Theorem 5.2, \mathcal{M} is a singleton at high temperature. Is the same true in the low temperature phase? If true, this would imply that the empirical measure converges to a deterministic limit, instead of the subsequential convergence that is established in this paper. If false, can one describe the extreme points of \mathcal{M} ?

3. Does the endpoint distribution converge in law? In other words, can we go beyond the Cesàro averages and prove limiting results for the actual endpoint distribution? Note that if this is true, then the limiting law must be an element of \mathcal{M} .

4. Can the methods of this paper be extended to understand the distinction between (SD) and (VSD)? Such an extension may lead to the resolution of some longstanding questions in this area, such as the following. (a) For $d \geq 3$, do the critical values β_c and $\tilde{\beta}_c$ coincide? This is known to be the case for $d = 1$ [30] and for $d = 2$ [52]. If the answer is no, then there would exist *critical strong disorder*, in which (SD) holds but not (VSD). (b) For $d \geq 3$, is there strong disorder at inverse temperature $\tilde{\beta}_c$? For analogous models on self-similar trees, the answer is yes [48].

Acknowledgments. We thank David Aldous, Alexei Borodin, Persi Diaconis, Jafar Jafarov, Chiranjib Mukherjee and Clifford Weil for fruitful discussions. We are also grateful to Erwin Bolthausen, Francis Comets, Tom Spencer, Srinivasa Varadhan, Vincent Vargas and Zhengqing Zhou for useful comments on our preliminary draft. We are also grateful to the referees for many useful suggestions and corrections.

The first author was supported in part by NSF Grant DGE-114747.

The second author was supported in part by NSF Grants DMS-1441513 and DMS-1608249.

SUPPLEMENTARY MATERIAL

Appendix A: Remaining technical details (DOI: [10.1214/19-AOP1376SUPPA](https://doi.org/10.1214/19-AOP1376SUPPA); .pdf). This appendix contains the proofs of Proposition 2.4, Theorem 2.8, Proposition 3.4, Lemma 6.1(a), measurability of the support number, Lemma 7.1, and the equivalence of two notions of asymptotic pure atomicity.

Appendix B: Comparison to the Mukherjee–Varadhan topology (DOI: [10.1214/19-AOP1376SUPPB](https://doi.org/10.1214/19-AOP1376SUPPB); .pdf). This appendix proves that the topology introduced by Mukherjee and Varadhan [61], when adapted to the discrete setting, is equivalent to the one constructed in this manuscript.

REFERENCES

- [1] AIZENMAN, M. and CONTUCCI, P. (1998). On the stability of the quenched state in mean-field spin-glass models. *J. Stat. Phys.* **92** 765–783. MR1657840 <https://doi.org/10.1023/A:1023080223894>
- [2] AIZENMAN, M., SIMS, R. and STARR, S. L. (2007). Mean-field spin glass models from the cavity-ROST perspective. In *Prospects in Mathematical Physics. Contemp. Math.* **437** 1–30. Am. Math. Soc., Providence, RI. MR2354653 <https://doi.org/10.1090/conm/437/08422>
- [3] ALBERTS, T., KHANIN, K. and QUASTEL, J. (2014). The intermediate disorder regime for directed polymers in dimension $1 + 1$. *Ann. Probab.* **42** 1212–1256. MR3189070 <https://doi.org/10.1214/13-AOP858>
- [4] ALBEVERIO, S. and ZHOU, X. Y. (1996). A martingale approach to directed polymers in a random environment. *J. Theoret. Probab.* **9** 171–189. MR1371075 <https://doi.org/10.1007/BF02213739>
- [5] ALDOUS, D. J. and BANDYOPADHYAY, A. (2005). A survey of max-type recursive distributional equations. *Ann. Appl. Probab.* **15** 1047–1110. MR2134098 <https://doi.org/10.1214/105051605000000142>
- [6] ARGUIN, L.-P. and CHATTERJEE, S. (2013). Random overlap structures: Properties and applications to spin glasses. *Probab. Theory Related Fields* **156** 375–413. MR3055263 <https://doi.org/10.1007/s00440-012-0431-6>

- [7] BAIK, J., LIECHTY, K. and SCHEHR, G. (2012). On the joint distribution of the maximum and its position of the Airy₂ process minus a parabola. *J. Math. Phys.* **53** 083303. MR3012644 <https://doi.org/10.1063/1.4746694>
- [8] BAKHTIN, Y. and SEO, D. Localization of directed polymers in continuous space. Preprint. Available at [arXiv:1905.00930](https://arxiv.org/abs/1905.00930).
- [9] BARRAL, J., RHODES, R. and VARGAS, V. (2012). Limiting laws of supercritical branching random walks. *C. R. Math. Acad. Sci. Paris* **350** 535–538. MR2929063 <https://doi.org/10.1016/j.crma.2012.05.013>
- [10] BARRAQUAND, G. and CORWIN, I. (2017). Random-walk in beta-distributed random environment. *Probab. Theory Related Fields* **167** 1057–1116. MR3627433 <https://doi.org/10.1007/s00440-016-0699-z>
- [11] BATES, E. and CHATTERJEE, S. (2020). Supplement to “The endpoint distribution of directed polymers.” <https://doi.org/10.1214/19-AOP1376SUPPA>, <https://doi.org/10.1214/19-AOP1376SUPPB>.
- [12] BEZERRA, S., TINDEL, S. and VIENS, F. (2008). Superdiffusivity for a Brownian polymer in a continuous Gaussian environment. *Ann. Probab.* **36** 1642–1675. MR2440919 <https://doi.org/10.1214/07-AOP363>
- [13] BILLINGSLEY, P. (1999). *Convergence of Probability Measures*, 2nd ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. Wiley, New York. MR1700749 <https://doi.org/10.1002/9780470316962>
- [14] BOLTHAUSEN, E. (1989). A note on the diffusion of directed polymers in a random environment. *Comm. Math. Phys.* **123** 529–534. MR1006293
- [15] BOLTHAUSEN, E., KÖNIG, W. and MUKHERJEE, C. (2017). Mean-field interaction of Brownian occupation measures II: A rigorous construction of the Pekar process. *Comm. Pure Appl. Math.* **70** 1598–1629. MR3666565 <https://doi.org/10.1002/cpa.21682>
- [16] BORODIN, A. and CORWIN, I. (2014). Macdonald processes. *Probab. Theory Related Fields* **158** 225–400. MR3152785 <https://doi.org/10.1007/s00440-013-0482-3>
- [17] BORODIN, A., CORWIN, I. and FERRARI, P. (2014). Free energy fluctuations for directed polymers in random media in 1 + 1 dimension. *Comm. Pure Appl. Math.* **67** 1129–1214. MR3207195 <https://doi.org/10.1002/cpa.21520>
- [18] BORODIN, A., CORWIN, I. and REMENIK, D. (2013). Log-gamma polymer free energy fluctuations via a Fredholm determinant identity. *Comm. Math. Phys.* **324** 215–232. MR3116323 <https://doi.org/10.1007/s00220-013-1750-x>
- [19] BORODIN, A. and PETROV, L. (2014). Integrable probability: From representation theory to Macdonald processes. *Probab. Surv.* **11** 1–58. MR3178541 <https://doi.org/10.1214/13-PS225>
- [20] BOTHNER, T. and LIECHTY, K. (2013). Tail decay for the distribution of the endpoint of a directed polymer. *Nonlinearity* **26** 1449–1472. MR3056134 <https://doi.org/10.1088/0951-7715/26/5/1449>
- [21] BRÖKER, Y. and MUKHERJEE, C. (2018). Localization of the Gaussian multiplicative chaos in the Wiener space and the stochastic heat equation in strong disorder. Preprint. Available at [arXiv:1808.05202](https://arxiv.org/abs/1808.05202).
- [22] BURKHOLDER, D. L., DAVIS, B. J. and GUNDY, R. F. (1972). Integral inequalities for convex functions of operators on martingales. In *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability (Univ. California, Berkeley, Calif., 1970/1971), Vol. II: Probability Theory* 223–240. MR0400380
- [23] CARMONA, P. and HU, Y. (2002). On the partition function of a directed polymer in a Gaussian random environment. *Probab. Theory Related Fields* **124** 431–457. MR1939654 <https://doi.org/10.1007/s004400200213>
- [24] CARMONA, P. and HU, Y. (2006). Strong disorder implies strong localization for directed polymers in a random environment. *ALEA Lat. Am. J. Probab. Math. Stat.* **2** 217–229. MR2249669
- [25] COMETS, F. (2017). *Directed Polymers in Random Environments. Lecture Notes in Math.* **2175**. Springer, Cham. MR3444835 <https://doi.org/10.1007/978-3-319-50487-2>
- [26] COMETS, F. and DEMBO, A. (2001). Ordered overlaps in disordered mean-field models. *Probab. Theory Related Fields* **121** 1–29. MR1857106 <https://doi.org/10.1007/PL00008794>
- [27] COMETS, F. and NGUYEN, V.-L. (2016). Localization in log-gamma polymers with boundaries. *Probab. Theory Related Fields* **166** 429–461. MR3547743 <https://doi.org/10.1007/s00440-015-0662-4>
- [28] COMETS, F., SHIGA, T. and YOSHIDA, N. (2003). Directed polymers in a random environment: Path localization and strong disorder. *Bernoulli* **9** 705–723. MR1996276 <https://doi.org/10.3150/bj/1066223275>
- [29] COMETS, F., SHIGA, T. and YOSHIDA, N. (2004). Probabilistic analysis of directed polymers in a random environment: A review. In *Stochastic Analysis on Large Scale Interacting Systems. Adv. Stud. Pure Math.* **39** 115–142. Math. Soc. Japan, Tokyo. MR2073332 <https://doi.org/10.2969/aspm/03910115>
- [30] COMETS, F. and VARGAS, V. (2006). Majorizing multiplicative cascades for directed polymers in random media. *ALEA Lat. Am. J. Probab. Math. Stat.* **2** 267–277. MR2249671
- [31] COMETS, F. and YOSHIDA, N. (2004). Some new results on Brownian directed polymers in random environment. *RIMS Kokyuroku* **1386** 50–66.

- [32] COMETS, F. and YOSHIDA, N. (2005). Brownian directed polymers in random environment. *Comm. Math. Phys.* **254** 257–287. MR2117626 <https://doi.org/10.1007/s00220-004-1203-7>
- [33] COMETS, F. and YOSHIDA, N. (2006). Directed polymers in random environment are diffusive at weak disorder. *Ann. Probab.* **34** 1746–1770. MR2271480 <https://doi.org/10.1214/009117905000000828>
- [34] CORWIN, I. (2012). The Kardar–Parisi–Zhang equation and universality class. *Random Matrices Theory Appl.* **1** 1130001. MR2930377 <https://doi.org/10.1142/S2010326311300014>
- [35] CORWIN, I., O’CONNELL, N., SEPPÄLÄINEN, T. and ZYGOURAS, N. (2014). Tropical combinatorics and Whittaker functions. *Duke Math. J.* **163** 513–563. MR3165422 <https://doi.org/10.1215/00127094-2410289>
- [36] CORWIN, I., SEPPÄLÄINEN, T. and SHEN, H. (2015). The strict-weak lattice polymer. *J. Stat. Phys.* **160** 1027–1053. MR3373650 <https://doi.org/10.1007/s10955-015-1267-0>
- [37] DEN HOLLANDER, F. (2009). *Random Polymers. Lecture Notes in Math.* **1974**. Springer, Berlin. MR2504175 <https://doi.org/10.1007/978-3-642-00333-2>
- [38] DERRIDA, B. and SPOHN, H. (1988). Polymers on disordered trees, spin glasses, and traveling waves. *J. Stat. Phys.* **51** 817–840. MR0971033 <https://doi.org/10.1007/BF01014886>
- [39] GANTERT, N., PERES, Y. and SHI, Z. (2010). The infinite valley for a recurrent random walk in random environment. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 525–536. MR2667708 <https://doi.org/10.1214/09-AIHP205>
- [40] GEORGIU, N., RASSOUL-AGHA, F. and SEPPÄLÄINEN, T. (2016). Variational formulas and cocycle solutions for directed polymer and percolation models. *Comm. Math. Phys.* **346** 741–779. MR3535900 <https://doi.org/10.1007/s00220-016-2613-z>
- [41] GEORGIU, N., RASSOUL-AGHA, F., SEPPÄLÄINEN, T. and YILMAZ, A. (2015). Ratios of partition functions for the log-gamma polymer. *Ann. Probab.* **43** 2282–2331. MR3395462 <https://doi.org/10.1214/14-AOP933>
- [42] GEORGIU, N. and SEPPÄLÄINEN, T. (2013). Large deviation rate functions for the partition function in a log-gamma distributed random potential. *Ann. Probab.* **41** 4248–4286. MR3161474 <https://doi.org/10.1214/12-AOP768>
- [43] GOLOSOV, A. O. (1984). Localization of random walks in one-dimensional random environments. *Comm. Math. Phys.* **92** 491–506. MR0736407
- [44] HUSE, D. A. and HENLEY, C. L. (1985). Pinning and roughening of domain walls in Ising systems due to random impurities. *Phys. Rev. Lett.* **54** 2708–2711. <https://doi.org/10.1103/PhysRevLett.54.2708>
- [45] IMBRIE, J. Z. and SPENCER, T. (1988). Diffusion of directed polymers in a random environment. *J. Stat. Phys.* **52** 609–626. MR0968950 <https://doi.org/10.1007/BF01019720>
- [46] JOHANSSON, K. (2000). Shape fluctuations and random matrices. *Comm. Math. Phys.* **209** 437–476. MR1737991 <https://doi.org/10.1007/s002200050027>
- [47] JOHANSSON, K. (2003). Discrete polynuclear growth and determinantal processes. *Comm. Math. Phys.* **242** 277–329. MR2018275 <https://doi.org/10.1007/s00220-003-0945-y>
- [48] KAHANE, J.-P. and PEYRIÈRE, J. (1976). Sur certaines martingales de Benoit Mandelbrot. *Adv. Math.* **22** 131–145. MR0431355 [https://doi.org/10.1016/0001-8708\(76\)90151-1](https://doi.org/10.1016/0001-8708(76)90151-1)
- [49] KANTOROVITCH, L. (1942). On the translocation of masses. *C. R. (Dokl.) Acad. Sci. URSS* **37** 199–201. MR0009619
- [50] KIFER, Y. (1997). The Burgers equation with a random force and a general model for directed polymers in random environments. *Probab. Theory Related Fields* **108** 29–65. MR1452549 <https://doi.org/10.1007/s004400050100>
- [51] KÖNIG, W. and MUKHERJEE, C. (2017). Mean-field interaction of Brownian occupation measures. I: Uniform tube property of the Coulomb functional. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 2214–2228. MR3729652 <https://doi.org/10.1214/16-AIHP788>
- [52] LACOIN, H. (2010). New bounds for the free energy of directed polymers in dimension $1 + 1$ and $1 + 2$. *Comm. Math. Phys.* **294** 471–503. MR2579463 <https://doi.org/10.1007/s00220-009-0957-3>
- [53] LÉVY, P. (1954). *Théorie de L’addition des Variables Aléatoires*. Gauthier-Villars, Paris.
- [54] LIONS, P.-L. (1984). The concentration-compactness principle in the calculus of variations. The locally compact case. I. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **1** 109–145. MR0778970
- [55] LIONS, P.-L. (1984). The concentration-compactness principle in the calculus of variations. The locally compact case. II. *Ann. Inst. H. Poincaré Anal. Non Linéaire* **1** 223–283. MR0778974
- [56] LIONS, P.-L. (1985). The concentration-compactness principle in the calculus of variations. The limit case. I. *Rev. Mat. Iberoam.* **1** 145–201. MR0834360 <https://doi.org/10.4171/RMI/6>
- [57] LIONS, P.-L. (1985). The concentration-compactness principle in the calculus of variations. The limit case. II. *Rev. Mat. Iberoam.* **1** 45–121. MR0850686 <https://doi.org/10.4171/RMI/12>

- [58] MEJANE, O. (2004). Upper bound of a volume exponent for directed polymers in a random environment. *Ann. Inst. Henri Poincaré Probab. Stat.* **40** 299–308. MR2060455 [https://doi.org/10.1016/S0246-0203\(03\)00072-4](https://doi.org/10.1016/S0246-0203(03)00072-4)
- [59] MORENO FLORES, G., QUASTEL, J. and REMENIK, D. (2013). Endpoint distribution of directed polymers in $1 + 1$ dimensions. *Comm. Math. Phys.* **317** 363–380. MR3010188 <https://doi.org/10.1007/s00220-012-1583-z>
- [60] MUKHERJEE, C. (2017). Gibbs measures on mutually interacting Brownian paths under singularities. *Comm. Pure Appl. Math.* **70** 2366–2404. MR3716856 <https://doi.org/10.1002/cpa.21700>
- [61] MUKHERJEE, C. and VARADHAN, S. R. S. (2016). Brownian occupation measures, compactness and large deviations. *Ann. Probab.* **44** 3934–3964. MR3572328 <https://doi.org/10.1214/15-AOP1065>
- [62] MUNKRES, J. R. (2000). *Topology*. Prentice Hall, Upper Saddle River, NJ. MR3728284
- [63] O’CONNELL, N. (2003). Random matrices, non-colliding processes and queues. In *Séminaire de Probabilités, XXXVI. Lecture Notes in Math.* **1801** 165–182. Springer, Berlin. MR1971584 https://doi.org/10.1007/978-3-540-36107-7_3
- [64] O’CONNELL, N. and YOR, M. (2001). Brownian analogues of Burke’s theorem. *Stochastic Process. Appl.* **96** 285–304. MR1865759 [https://doi.org/10.1016/S0304-4149\(01\)00119-3](https://doi.org/10.1016/S0304-4149(01)00119-3)
- [65] PANCHENKO, D. (2013). *The Sherrington–Kirkpatrick Model. Springer Monographs in Mathematics*. Springer, New York. MR3052333 <https://doi.org/10.1007/978-1-4614-6289-7>
- [66] PARTHASARATHY, K. R., RANGA RAO, R. and VARADHAN, S. R. S. (1962). On the category of indecomposable distributions on topological groups. *Trans. Amer. Math. Soc.* **102** 200–217. MR0153041 <https://doi.org/10.2307/1993674>
- [67] PETERMANN, M. (2000). Superdiffusivity of polymers in random environment. Ph.D. thesis, Univ. Zürich.
- [68] PIZA, M. S. T. (1997). Directed polymers in a random environment: Some results on fluctuations. *J. Stat. Phys.* **89** 581–603. MR1484057 <https://doi.org/10.1007/BF02765537>
- [69] QUASTEL, J. and REMENIK, D. (2014). Airy processes and variational problems. In *Topics in Percolative and Disordered Systems. Springer Proc. Math. Stat.* **69** 121–171. Springer, New York. MR3229288 https://doi.org/10.1007/978-1-4939-0339-9_5
- [70] QUASTEL, J. and REMENIK, D. (2015). Tails of the endpoint distribution of directed polymers. *Ann. Inst. Henri Poincaré Probab. Stat.* **51** 1–17. MR3300961 <https://doi.org/10.1214/12-AIHP525>
- [71] RASSOUL-AGHA, F., SEPPÄLÄINEN, T. and YILMAZ, A. (2013). Quenched free energy and large deviations for random walks in random potentials. *Comm. Pure Appl. Math.* **66** 202–244. MR2999296 <https://doi.org/10.1002/cpa.21417>
- [72] RASSOUL-AGHA, F., SEPPÄLÄINEN, T. and YILMAZ, A. (2017). Variational formulas and disorder regimes of random walks in random potentials. *Bernoulli* **23** 405–431. MR3556777 <https://doi.org/10.3150/15-BEJ747>
- [73] RUZMAIKINA, A. and AIZENMAN, M. (2005). Characterization of invariant measures at the leading edge for competing particle systems. *Ann. Probab.* **33** 82–113. MR2118860 <https://doi.org/10.1214/009117904000000865>
- [74] SCHEHR, G. (2012). Extremes of N vicious walkers for large N : Application to the directed polymer and KPZ interfaces. *J. Stat. Phys.* **149** 385–410. MR2992794 <https://doi.org/10.1007/s10955-012-0593-8>
- [75] SEPPÄLÄINEN, T. (2012). Scaling for a one-dimensional directed polymer with boundary conditions. *Ann. Probab.* **40** 19–73. MR2917766 <https://doi.org/10.1214/10-AOP617>
- [76] SEPPÄLÄINEN, T. and VALKÓ, B. (2010). Bounds for scaling exponents for a $1 + 1$ dimensional directed polymer in a Brownian environment. *ALEA Lat. Am. J. Probab. Math. Stat.* **7** 451–476. MR2741194
- [77] SINAI, Y. G. (1995). A remark concerning random walks with random potentials. *Fund. Math.* **147** 173–180. MR1341729 <https://doi.org/10.4064/fm-147-2-173-180>
- [78] SINAI, YA. G. (1982). The limit behavior of a one-dimensional random walk in a random environment. *Teor. Veroyatn. Primen.* **27** 247–258. MR0657919
- [79] SONG, R. and ZHOU, X. Y. (1996). A remark on diffusion of directed polymers in random environments. *J. Stat. Phys.* **85** 277–289. MR1413246 <https://doi.org/10.1007/BF02175566>
- [80] TALAGRAND, M. (2006). The Parisi formula. *Ann. of Math. (2)* **163** 221–263. MR2195134 <https://doi.org/10.4007/annals.2006.163.221>
- [81] TAO, T. (2008). Concentration compactness and the profile decomposition. *What’s New (blog Entry)*. Available at <https://terrytao.wordpress.com/2008/11/05/concentration-compactness-and-the-profile-decomposition/>.
- [82] TAO, T. (2010). Concentration compactness via nonstandard analysis. *What’s New (blog Entry)*. Available at <https://terrytao.wordpress.com/2010/11/29/concentration-compactness-via-nonstandard-analysis/>.
- [83] THIERY, T. and LE DOUSSAL, P. (2015). On integrable directed polymer models on the square lattice. *J. Phys. A* **48** 465001. MR3418005 <https://doi.org/10.1088/1751-8113/48/46/465001>

- [84] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer Series in Statistics. Springer, New York. MR1385671 <https://doi.org/10.1007/978-1-4757-2545-2>
- [85] VARGAS, V. (2007). Strong localization and macroscopic atoms for directed polymers. *Probab. Theory Related Fields* **138** 391–410. MR2299713 <https://doi.org/10.1007/s00440-006-0030-5>
- [86] VILLANI, C. (2009). *Optimal Transport: Old and New*. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences] **338**. Springer, Berlin. MR2459454 <https://doi.org/10.1007/978-3-540-71050-9>
- [87] WÜTHRICH, M. V. (1998). Superdiffusive behavior of two-dimensional Brownian motion in a Poissonian potential. *Ann. Probab.* **26** 1000–1015. MR1634412 <https://doi.org/10.1214/aop/1022855742>
- [88] YOSHIDA, N. (2008). Phase transitions for the growth rate of linear stochastic evolutions. *J. Stat. Phys.* **133** 1033–1058. MR2462010 <https://doi.org/10.1007/s10955-008-9646-4>