Comparison Theorems for Gibbs Measures

Patrick Rebeschini · Ramon van Handel

Abstract The Dobrushin comparison theorem is a powerful tool to bound the difference between the marginals of high-dimensional probability distributions in terms of their local specifications. Originally introduced to prove uniqueness and decay of correlations of Gibbs measures, it has been widely used in statistical mechanics as well as in the analysis of algorithms on random fields and interacting Markov chains. However, the classical comparison theorem requires validity of the Dobrushin uniqueness criterion, essentially restricting its applicability in most models to a small subset of the natural parameter space. In this paper we develop generalized Dobrushin comparison theorems in terms of influences between blocks of sites, in the spirit of Dobrushin–Shlosman and Weitz, that substantially extend the range of applicability of the classical comparison theorem. Our proofs are based on the analysis of an associated family of Markov chains. We develop in detail an application of our main results to the analysis of sequential Monte Carlo algorithms for filtering in high dimension.

Keywords Dobrushin comparison theorem · Gibbs measures · Filtering algorithms

1 Introduction

The canonical description of a high-dimensional random system is provided by a probability measure ρ on a (possibly infinite) product space $\mathbb{S} = \prod_{i \in I} \mathbb{S}^i$: each site $i \in I$ represents a single degree of freedom, or dimension, of the model. When I is the set of vertices of a graph, the measure ρ defines a graphical model or a random field. Models of this type are ubiquitous in statistical mechanics, combinatorics, computer science, statistics, and in many other areas of science and engineering.

P. Rebeschini · R. van Handel (✉)
Sherrerd Hall, Princeton University, Princeton, NJ 08544, USA
e-mail: rvan@princeton.edu

P. Rebeschini
e-mail: prebesch@princeton.edu

Let ρ and $\tilde{\rho}$ be two such models that are defined on the same space $\mathbb{S}$. We would like to address the following basic question: when is $\tilde{\rho}$ a good approximation of ρ? Such questions arise at a basic level both in understanding the properties of random systems themselves, and in the analysis of the algorithms that are used to investigate and approximate these systems. Of course, probability theory provides numerous methods to evaluate the difference between arbitrary probability measures, but the high-dimensional setting brings specific challenges: any approximation of practical utility in high dimension must yield error bounds that do not grow, or at least grow sufficiently slowly, in the model dimension. We therefore seek quantitative methods that allow to establish dimension-free bounds on high-dimensional distributions.

A general method to address precisely this problem was developed by Dobrushin [5] in the context of statistical mechanics. In the approach pioneered by Dobrushin, Lanford, and Ruelle, an infinite-dimensional system of interacting particles is defined by its *local* characteristics: for finite sets of sites $J \subset I$, the conditional distribution $\rho(dx^J | x^{I \setminus J})$ of the configuration in J is specified given that the particles outside J are frozen in a fixed configuration. This local description is a direct consequence of the physical parameters of the problem. The model ρ is then defined as a probability measure (called a Gibbs measure) that is compatible with the given system of local conditional distributions; see Sect. 2.1. This setting gives rise to many classical questions in statistical mechanics [12,21]; for example, the Gibbs measure may or may not be unique, reflecting the presence of a phase transition.

The Dobrushin comparison theorem [5, Theorem 3] provides a powerful tool to obtain dimension-free estimates on the difference between the marginals of Gibbs measures ρ and $\tilde{\rho}$ in terms of the single site conditional distributions $\rho(dx^j | x^{I \setminus \{j\}})$ and $\tilde{\rho}(dx^j | x^{I \setminus \{j\}})$. In its simplified form due to Föllmer [10], this result has become standard textbook material, cf. [12, Theorem 8.20], [21, Theorem V.2.2]. It is widely used to establish numerous properties of Gibbs measures, including uniqueness, decay of correlations, global Markov properties, and analyticity [11,12,21], as well as functional inequalities and concentration of measure properties [13,15,26], and has similarly proved to be useful in the analysis of algorithms on random fields and interacting Markov chains [2,18,22,27].

Despite this broad array of applications, the assumptions required by the Dobrushin comparison theorem prove to be rather stringent. This can already be seen in the easiest qualitative consequence of this result: the comparison theorem implies uniqueness of the Gibbs measure under the well-known Dobrushin uniqueness criterion [5]. Unfortunately, this criterion is restrictive: even in models where uniqueness can be established by explicit computation, the Dobrushin uniqueness criterion holds only in a small subset of the natural parameter space (see, e.g., [25] for examples). This suggests that the Dobrushin comparison theorem is a rather blunt tool. On the other hand, it is also known that the Dobrushin uniqueness criterion can be substantially improved: this was accomplished in Dobrushin and Shlosman [4] by considering a local description in terms of larger blocks $\rho(dx^J | x^{I \setminus J})$ instead of the single site specification $\rho(dx^j | x^{I \setminus \{j\}})$. In this manner, it is often possible to capture a large part of or even the entire uniqueness region. The uniqueness results of Dobrushin and Shlosman were further generalized by Weitz [25], who developed general combinatorial criteria for uniqueness. However, while the proofs of Dobrushin–Shlosman and Weitz also provide some information on decay of correlations, they do not provide an analogue of the powerful machinery for obtaining quantitative estimates on the difference between the marginal distributions of different Gibbs measures ρ and $\tilde{\rho}$ that the Dobrushin comparison theorem yields in its more restrictive setting.

The aim of the present paper is to fill this gap. Our main results (Theorems 2.3 and 2.9) provide a direct extension of the Dobrushin comparison theorem to the much more

general setting considered by Weitz [25], substantially extending the range of applicability of the classical comparison theorem. While the classical comparison theorem is an immediate consequence of our main result (Corollary 2.4), the usual proof does not appear to extend easily beyond the single site setting. We therefore develop a different, though certainly related, method of proof that systematically exploits the connection of Markov chains. In particular, our main results are derived from a more general comparison theorem for Markov chains that is applied to a suitably defined family of Gibbs samplers, cf. Sect. 3 below.

Our original motivation for developing generalized comparison theorems was the investigation of algorithms for filtering in high dimension. Filtering—the computation of the conditional distributions of a random dynamical system given observed data—is a problem that arises in a wide array of applications in science and engineering. Modern filtering algorithms utilize sequential Monte Carlo methods to efficiently approximate the conditional distributions [1]. Unfortunately, such algorithms suffer heavily from the curse of dimensionality, making them largely useless in complex data assimilation problems that arise in high-dimensional applications such as weather forecasting (the state-of-the-art in such applications is still dominated by *ad-hoc* methods). Motivated by such problems, we have begun to investigate in [18] a class of regularized filtering algorithms that can, in principle, exhibit dimension-free performance in models that possess decay of correlations. For the simplest possible algorithm of this type, dimension-free error bounds are proved in [18] by systematic application of the Dobrushin comparison theorem.

In order to ensure decay of correlations, [18] imposes a weak interactions assumption that is dictated by the Dobrushin comparison theorem. As will be explained in Sect. 4, however, this assumption is unsatisfactory already at the *qualitative* level: it limits not only the spatial interactions (as is needed to ensure decay of correlations) but also the dynamics in time. Overcoming this unnatural restriction requires a generalized comparison theorem, which provided the motivation for our main results. As an illustration of our main results, and as a problem of interest in its own right, the application to filtering algorithms will be developed in detail in Sect. 4.

The remainder of this paper is organized as follows. Section 2 introduces the basic setup and notation to be used throughout the paper, and states our main results. While the comparison theorem, being quantitative in nature, is already of interest in the finite setting card $I < \infty$ (unlike the qualitative uniqueness questions that are primarily of interest when card $I = \infty$), we will develop our main results in a general setting that admits infinite-range interactions. The proofs of the main results are given in Sect. 3. The application to filtering algorithms is finally developed in Sect. 4.

2 Main Results

2.1 Setting and Notation

We begin by introducing the basic setting that will be used throughout this section.

Sites and Configurations

Let I be a finite or countably infinite set of *sites*. Each subset $J \subseteq I$ is called a *region*; the set of finite regions will be denoted as

$$\mathfrak{I} := \{J \subseteq I : \text{card } J < \infty\}.$$

To each site $i \in I$ is associated a measurable space $\mathbb{S}^i$, the *local state space*. A *configuration* is an assignment $x_i \in \mathbb{S}^i$ to each site $i \in I$. The set of all configurations $\mathbb{S}$, and the set $\mathbb{S}^J$ of configurations in a given region $J \subseteq I$, are defined as

$$\mathbb{S} := \prod_{i \in I} \mathbb{S}^i, \quad \mathbb{S}^J := \prod_{i \in J} \mathbb{S}^i.$$

For $x = (x_i)_{i \in I} \in \mathbb{S}$, we denote by $x^J := (x_i)_{i \in J} \in \mathbb{S}^J$ the natural projection on $\mathbb{S}^J$. When $J \cap K = \varnothing$, we define $z = x^J y^K \in \mathbb{S}^{J \cup K}$ such that $z^J = x^J$ and $z^K = y^K$.

Local Functions

A function $f : \mathbb{S} \to \mathbb{R}$ is said to be *J-local* if $f(x) = f(z)$ whenever $x^J = z^J$, that is, if $f(x)$ depends on x^J only. The function f is said to be *local* if it is J-local for some finite region $J \in \mathfrak{I}$. When I is a finite set, every function is local. When I is infinite, however, we will frequently restrict attention to local functions. More generally, we will consider a class of "nearly" local functions to be defined presently.

Given any function $f : \mathbb{S} \to \mathbb{R}$, define for $J \in \mathfrak{I}$ and $x \in \mathbb{S}$ the J-local function

$$f_x^J(z) := f(z^J x^{I \setminus J}).$$

f is called *quasilocal* if it can be approximated pointwise by the local functions f_x^J:

$$\lim_{J \in \mathfrak{I}} |f_x^J(z) - f(z)| = 0 \quad \text{for all } x, z \in \mathbb{S},$$

where $\lim_{J \in \mathfrak{I}} a_J$ denotes the limit of the net $(a_J)_{J \in \mathfrak{I}}$ where $\mathfrak{I}$ is directed by inclusion $\subseteq$ (equivalently, $a_J \to 0$ if and only if $a_{J_i} \to 0$ for every sequence $J_1, J_2, \ldots \in \mathfrak{I}$ such that $J_1 \subseteq J_2 \subseteq \cdots$ and $\bigcup_i J_i = I$). Let us note that this notion is slightly weaker than the conventional notion of quasilocality used, for example, in [12].

Metrics

In the sequel, we fix for each $i \in I$ a metric η_i on $\mathbb{S}^i$ (we assume throughout that η_i is measurable as a function on $\mathbb{S}^i \times \mathbb{S}^i$). We will write $\|\eta_i\| = \sup_{x,z} \eta_i(x, z)$.

Given a function $f : \mathbb{S} \to \mathbb{R}$ and $i \in I$, we define

$$\delta_i f := \sup_{x,z \in \mathbb{S}: x^{I \setminus \{i\}} = z^{I \setminus \{i\}}} \frac{|f(x) - f(z)|}{\eta_i(x_i, z_i)}.$$

The quantity $\delta_i f$ measures the variability of $f(x)$ with respect to the variable x_i.

Matrices

The calculus of possibly infinite nonnegative matrices will appear repeatedly in the sequel. Given matrices $A = (A_{ij})_{i,j \in I}$ and $B = (B_{ij})_{i,j \in I}$ with nonnegative entries $A_{ij} \geq 0$ and $B_{ij} \geq 0$, the matrix product is defined as usual by

$$(AB)_{ij} = \sum_{k \in I} A_{ik} B_{kj}.$$

This quantity is well defined as the terms in the sum are all nonnegative, but $(AB)_{ij}$ may possibly take the value $+\infty$. As long as we consider only nonnegative matrices, all the usual

rules of matrix multiplication extend to infinite matrices provided that we allow entries with the value $+\infty$ and that we use the convention $+\infty \cdot 0 = 0$ (this follows from the Fubini–Tonelli theorem, cf. [6, Chap. 4]). In particular, the matrix powers A^k, $k \geq 1$ are well defined, and we define $A^0 = I$ where $I := (\mathbf{1}_{i=j})_{i,j \in I}$ denotes the identity matrix. We will write $A < \infty$ if the nonnegative matrix A satisfies $A_{ij} < \infty$ for every $i, j \in I$.

Kernels, Covers, Local Characteristics

A *transition kernel* γ from a measurable space $(\Omega, \mathcal{F})$ to a measurable space $(\Omega', \mathcal{F}')$ is a map $\gamma : \Omega \times \mathcal{F}' \to \mathbb{R}$ such that $\omega \mapsto \gamma_\omega(A)$ is a measurable function for each $A \in \mathcal{F}'$ and $\gamma_\omega(\cdot)$ is a probability measure for each $\omega \in \Omega$, cf. [14]. Given a probability measure μ on Ω and function f on Ω', we define as usual the probability measure $(\mu\gamma)(A) = \int \mu(d\omega)\gamma_\omega(A)$ on Ω' and function $(\gamma f)(\omega) = \int \gamma_\omega(d\omega') f(\omega')$ on Ω. A transition kernel γ between product spaces is called *quasilocal* if γf is quasilocal for every bounded and measurable quasilocal function f.

Our interest throughout this paper is in models of random configurations, described by a probability measure μ on $\mathbb{S}$. We would like to understand the properties of such models based on their *local* characteristics. A natural way to express the local structure in a finite region $J \in \mathcal{I}$ is to consider the conditional distribution $\gamma_x^J(dz^J) = \mu(dz^J | x^{I \setminus J})$ of the configuration in J given a fixed configuration $x^{I \setminus J}$ for the sites outside J: conceptually, γ^J describes how the sites in J "interact" with the sites outside J. The conditional distribution γ^J is a transition kernel from $\mathbb{S}$ to $\mathbb{S}^J$. To obtain a complete local description of the model, we must consider a class of finite regions J that covers the entire set of sites I. Let us call a collection of regions $\mathcal{J} \subseteq \mathcal{I}$ a *cover* of I if every site $i \in I$ is contained in at least one element of $\mathcal{J}$ (note that, by definition, a cover contains only finite regions). Given any cover $\mathcal{J}$, the collection $(\gamma^J)_{J \in \mathcal{J}}$ provides a local description of the model.

In fact, our main results will hold in a more general setting than is described above. Let μ be a probability measure on $\mathbb{S}$ and γ^J be transition kernel from $\mathbb{S}$ to $\mathbb{S}^J$. We say that μ is γ^J-*invariant* if for every bounded measurable function f

$$\int \mu(dx)\, f(x) = \int \mu(dx)\, \gamma_x^J(dz^J)\, f(z^J x^{I \setminus J});$$

by a slight abuse of notation, we will also write $\mu f = \mu\gamma^J f^J$. This means that if the configuration x is drawn according to μ, its distribution is unchanged if we replace the configuration x^J inside the region J by a random sample from the distribution γ_x^J, keeping the configuration $x^{I \setminus J}$ outside J fixed. Our main results will be formulated in terms of a collection of transition kernels $(\gamma^J)_{J \in \mathcal{J}}$ such that $\mathcal{J}$ is a cover of I and such that μ is γ^J-invariant for every $J \in \mathcal{J}$. If we choose $\gamma_x^J(dz^J) = \mu(dz^J | x^{I \setminus J})$ as above, then the γ^J-invariance of μ holds by construction [14, Theorem 6.4]; however, any family of γ^J-invariant kernels will suffice for the validity of our main results.

Remark 2.1 The idea that the collection $(\gamma^J)_{J \in \mathcal{J}}$ provides a natural description of high-dimensional probability distributions is prevalent in many applications. Statistical mechanical models are usually *defined* in terms of such a family: to this end, one fixes a priori a family of transition kernels $(\gamma^J)_{J \in \mathcal{I}}$, called a specification, that describes the local characteristics of the model. The definition of γ^J is done directly in terms of the parameters of the problem (the potentials that define the physical interactions). A measure μ on $\mathbb{S}$ is called a *Gibbs measure* for the given specification if $\mu(dz^J | x^{I \setminus J}) = \gamma_x^J(dz^J)$ for every $J \in \mathcal{I}$, cf. [12,21]. The construction of Gibbs measures from specifications is not essential for the validity or

applicability of our results, however: for example, in the application in Sect. 4 the model is defined by certain conditional distributions of a high-dimensional Markov chain, and the collection $(\gamma^J)_{J \in \mathcal{J}}$ is merely a device that is used in the proof. For this reason, we have not phrased our results in terms of Gibbs measures, but rather in the more general setting of a probability measure together with a family of invariant kernels.

2.2 Main Result

Let ρ and $\tilde{\rho}$ be probability measures on the space of configurations $\mathbb{S}$. Our main result, Theorem 2.3 below, provides a quantitative bound on the difference between ρ and $\tilde{\rho}$ in terms of their local characteristics. Before we can state our results, we must first introduce some basic notions. Our terminology is inspired by Weitz [25].

As was explained above, the local description of a probability measure ρ on $\mathbb{S}$ will be provided in terms of a family of transition kernels. We formalize this as follows.

Definition 2.1 A *local update rule* for ρ is a collection $(\gamma^J)_{J \in \mathcal{J}}$ where $\mathcal{J}$ is a cover of I, γ^J is a transition kernel from $\mathbb{S}$ to $\mathbb{S}^J$ and ρ is γ^J-invariant for every $J \in \mathcal{J}$.

In order to compare two measures ρ and $\tilde{\rho}$ on the basis of their local update rules $(\gamma^J)_{J \in \mathcal{J}}$ and $(\tilde{\gamma}^J)_{J \in \mathcal{J}}$, we must quantify two separate effects. On the one hand, we must understand how the two models differ locally: that is, we must quantify how γ_x^J and $\tilde{\gamma}_x^J$ differ when acting on the same configuration x. On the other hand, we must understand how perturbations to the local update rule in different regions interact: to this end, we will quantify the extent to which γ_x^J and γ_z^J differ for different configurations x, z. Both effects will be addressed by introducing a suitable family of couplings. Recall that a probability measure Q on a product space $\Omega \times \Omega$ is called a *coupling* of two probability measures μ, ν on Ω if its marginals coincide with μ, ν, that is, if $Q(\cdot \times \Omega) = \mu$ and $Q(\Omega \times \cdot) = \nu$.

Definition 2.2 A *coupled update rule* for $(\rho, \tilde{\rho})$ is a collection $(\gamma^J, \tilde{\gamma}^J, Q^J, \hat{Q}^J)_{J \in \mathcal{J}}$, where $\mathcal{J}$ is a cover of I, such that the following properties hold:

1. $(\gamma^J)_{J \in \mathcal{J}}$ and $(\tilde{\gamma}^J)_{J \in \mathcal{J}}$ are local update rules for ρ and $\tilde{\rho}$, respectively.
2. $Q_{x,z}^J$ is a coupling of γ_x^J, γ_z^J for all $J \in \mathcal{J}$, $x, z \in \mathbb{S}$ with $\mathrm{card}\{i : x_i \neq z_i\} = 1$.
3. $\hat{Q}_x^J$ is a coupling of $\gamma_x^J, \tilde{\gamma}_x^J$ for all $J \in \mathcal{J}$, $x \in \mathbb{S}$.

We can now state our main result.

Theorem 2.3 *Let $\mathcal{J}$ be a cover of I, let $(w_J)_{J \in \mathcal{J}}$ be a family of strictly positive weights, and let $(\gamma^J, \tilde{\gamma}^J, Q^J, \hat{Q}^J)_{J \in \mathcal{J}}$ be a coupled update rule for $(\rho, \tilde{\rho})$. Define*

$$W_{ij} := \mathbf{1}_{i=j} \sum_{J \in \mathcal{J}: i \in J} w_J,$$

$$R_{ij} := \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \frac{1}{\eta_j(x_j, z_j)} \sum_{J \in \mathcal{J}: i \in J} w_J \, Q_{x,z}^J \eta_i,$$

$$a_j := \sum_{J \in \mathcal{J}: j \in J} w_J \int^* \tilde{\rho}(dx) \, \hat{Q}_x^J \eta_j$$

for $i, j \in I$. Assume that γ^J is quasilocal for every $J \in \mathcal{J}$, and that

$$W_{ii} \leq 1 \quad \text{and} \quad \lim_{n \to \infty} \sum_{j \in I} (I - W + R)_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j = 0 \quad \text{for all } i \in I. \tag{1}$$

Then we have

$$|\rho f - \tilde{\rho} f| \le \sum_{i,j \in I} \delta_i f \, D_{ij} \, W_{jj}^{-1} a_j \quad where \quad D := \sum_{n=0}^{\infty} (W^{-1} R)^n,$$

for any bounded measurable quasilocal function f such that $\delta_i f < \infty$ for all $i \in I$.

Remark 2.2 While it is essential in the proof that γ^J and $\tilde{\gamma}^J$ are transition kernels, we do not require that Q^J and $\hat{Q}^J$ are transition kernels in Definition 2.2, that is, the couplings $Q_{x,z}^J$ and $\hat{Q}_x^J$ need not be measurable in x, z. It is for this reason that a_j are defined in terms of an outer integral rather than an ordinary integral [23]:

$$\int^* f(x) \, \rho(dx) := \inf \left\{ \int g(x) \, \rho(dx) : f \le g, \ g \text{ is measurable} \right\}.$$

When $x \mapsto \hat{Q}_x^J \eta_j$ is measurable this issue can be disregarded. In practice measurability will hold in all but pathological cases, but may not always be trivial to prove. We therefore allow for nonmeasurable couplings for sake of technical convenience, so that it is not necessary to check measurability when applying Theorem 2.3.

The proof of Theorem 2.3 will be given in Sects. 3.1–3.3. At the heart of the proof lies a general but rather elementary comparison theorem for the invariant measures ρ and $\tilde{\rho}$ of two given Markov kernels G and $\tilde{G}$, which is proved in Sect. 3.1. If $\sum_J w_J < \infty$, then $\rho, \tilde{\rho}$ are invariant for $G \propto \sum_J w_J \gamma^J$ and $\tilde{G} \propto \sum_J w_J \tilde{\gamma}^J$ by construction, and the application of the general comparison theorem of Sect. 3.1 is straightforward. In this case, an essentially complete proof is obtained in Sect. 3.2. The main complication in Theorem 2.3 is that we have not assumed summability of the weights; the analysis of the general case is performed in Sect. 3.3.

In the remainder of this section, we formulate a number of special cases and extensions of Theorem 2.3. A detailed application is developed in Sect. 4.

2.3 The Classical Comparison Theorem

The original comparison theorem of Dobrushin [5, Theorem 3] and its formulation due to Föllmer [10] correspond to the special case of Theorem 2.3 where the cover $\mathfrak{J} = \mathfrak{J}_s := \{\{i\} : i \in I\}$ consists of single sites. For example, the main result of [10] follows readily from Theorem 2.3 under a mild regularity assumption. To formulate it, recall that the *Wasserstein distance* $d_\eta(\mu, \nu)$ between probability measures μ and ν on a measurable space Ω with respect to a measurable metric η is defined as

$$d_\eta(\mu, \nu) := \inf_{\substack{Q(\cdot \times \Omega)=\mu \\ Q(\Omega \times \cdot)=\nu}} Q\eta,$$

where the infimum is taken over probability measures Q on $\Omega \times \Omega$ with the given marginals μ and ν. We now obtain the following (cf. [10] and [11, Remark 2.17]).

Corollary 2.4 ([10]) *Assume $\mathbb{S}^i$ is Polish and η_i is lower-semicontinuous for all $i \in I$. Let $(\gamma^{\{i\}})_{i \in I}$ and $(\tilde{\gamma}^{\{i\}})_{i \in I}$ be local update rules for ρ and $\tilde{\rho}$, respectively, and let*

$$C_{ij} := \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \frac{d_{\eta_i}(\gamma_x^{\{i\}}, \gamma_z^{\{i\}})}{\eta_j(x_j, z_j)}, \quad b_j := \int^* \tilde{\rho}(dx) \, d_{\eta_j}(\gamma_x^{\{j\}}, \tilde{\gamma}_x^{\{j\}}).$$

Assume that $\gamma^{\{i\}}$ is quasilocal for every $i \in I$, and that

$$\lim_{n \to \infty} \sum_{j \in I} C_{ij}^n (\rho \otimes \tilde{\rho}) \eta_j = 0 \quad \text{for all } i \in I.$$

Then we have

$$|\rho f - \tilde{\rho} f| \leq \sum_{i, j \in I} \delta_i f \, D_{ij} \, b_j \quad \text{where} \quad D := \sum_{n=0}^{\infty} C^n,$$

for any bounded measurable quasilocal function f such that $\delta_i f < \infty$ for all $i \in I$.

If $Q_{x,z}^{\{i\}}$, $\hat{Q}_x^{\{i\}}$ are minimizers in the definition of $d_{\eta_i}(\gamma_x^{\{i\}}, \gamma_z^{\{i\}})$, $d_{\eta_i}(\gamma_x^{\{i\}}, \tilde{\gamma}_x^{\{i\}})$, respectively, and if we let $\mathcal{J} = \mathcal{J}_s$ and $w_{\{i\}} = 1$ for all $i \in I$, then Corollary 2.4 follows immediately from Theorem 2.3. For simplicity, we have imposed the mild topological regularity assumption on $\mathbb{S}^i$ and η_i to ensure the existence of minimizers [24, Theorem 4.1] (when minimizers do not exist, it is possible to obtain a similar result by using near-optimal couplings). Note that when $\eta_i(x, z) = \mathbf{1}_{x \neq z}$ is the trivial metric, the Wasserstein distance reduces to the total variation distance

$$d_\eta(\mu, \nu) = \frac{1}{2}\|\mu - \nu\| := \frac{1}{2} \sup_{f : \|f\| \leq 1} |\mu f - \nu f| \quad \text{when} \quad \eta(x, z) = \mathbf{1}_{x \neq z},$$

and an optimal coupling exists in any measurable space [5, p. 472]. Thus in this case no regularity assumptions are needed, and Corollary 2.4 reduces to the textbook comparison theorem that appears, e.g., in [12, Theorem 8.20] or [21, Theorem V.2.2].

While the classical comparison theorem of Corollary 2.4 follows from our main result, it should be emphasized that the single site assumption $\mathcal{J} = \mathcal{J}_s$ is a significant restriction. The general statement of Theorem 2.3 constitutes a crucial improvement that substantially extends the range of applicability of the comparison method, as we will see below and in Sect. 4. Let us also note that the proofs in [5, 10], based on the "method of estimates," do not appear to extend easily beyond the single site setting.

2.4 Alternative Assumptions

The key assumption of Theorem 2.3 is (1). The aim of the present section is to obtain useful alternatives to assumption (1) that are easily verified in practice.

We begin by defining the notion of a tempered measure [11, Remark 2.17].

Definition 2.5 A probability measure μ on $\mathbb{S}$ is called $x^\star$-*tempered* if

$$\sup_{i \in I} \int \mu(dx) \, \eta_i(x_i, x_i^\star) < \infty.$$

In the sequel $x^\star \in \mathbb{S}$ will be considered fixed and μ will be called *tempered*.

It is often the case in practice that the collection of metrics is uniformly bounded, that is, $\sup_i \|\eta_i\| < \infty$. In this case, every probability measure on $\mathbb{S}$ is trivially tempered. However, the restriction to tempered measures may be essential when the spaces $\mathbb{S}^i$ are noncompact (see [5, Sect. 5] for an illuminating example).

Let us recall that a norm $\|\cdot\|$ defined on an algebra of square (possibly infinite) matrices is called a *matrix norm* if $\|AB\| \leq \|A\| \|B\|$. We also recall that the matrix norms $\|\cdot\|_\infty$ and

$\| \cdot \|_1$ are defined for nonnegative matrices $A = (A_{ij})_{i,j \in I}$ as

$$\|A\|_\infty := \sup_{i \in I} \sum_{j \in I} A_{ij}, \quad \|A\|_1 := \sup_{j \in I} \sum_{i \in I} A_{ij}.$$

The following result collects various alternatives to (1). It is proved in Sect. 3.4.

Corollary 2.6 *Suppose that ρ and $\tilde{\rho}$ are tempered. Then the conclusion of Theorem 2.3 remains valid when the assumption (1) is replaced by one of the following:*

1. card $I < \infty$ and $D < \infty$.
2. card $I < \infty$, $R < \infty$, and $\|(W^{-1}R)^n\| < 1$ *for some $n \geq 1$, matrix norm $\| \cdot \|$.*
3. $\sup_i W_{ii} < \infty$ and $\|W^{-1}R\|_\infty < 1$.
4. $\sup_i W_{ii} < \infty$, $\|RW^{-1}\|_\infty < \infty$, and $\|(RW^{-1})^n\|_\infty < 1$ *for some $n \geq 1$.*
5. $\sup_i W_{ii} < \infty$, $\sum_i \|\eta_i\| < \infty$, and $\|RW^{-1}\|_1 < 1$.
6. $\sup_i W_{ii} < \infty$, *there exists a metric m on I such that* $\sup\{m(i,j) : R_{ij} > 0\} < \infty$ *and* $\sup_i \sum_j e^{-\beta m(i,j)} < \infty$ *for all $\beta > 0$, and* $\|RW^{-1}\|_1 < 1$.

The conditions of Corollary 2.6 are closely related to the uniqueness of Gibbs measures. Suppose that the collection of quasilocal transition kernels $(\gamma^J)_{J \in \mathcal{J}}$ is a local update rule for ρ. It is natural to ask whether ρ is the *unique* measure that admits $(\gamma^J)_{J \in \mathcal{J}}$ as a local update rule (see Remark 2.1). We now observe that uniqueness is a necessary condition for the conclusion of Theorem 2.3. Indeed, let $\tilde{\rho}$ be another measure that admits the same local update rule. If (1) holds, we can apply Theorem 2.3 with $\tilde{\gamma}^J = \gamma^J$ and $a_j = 0$ to conclude that $\tilde{\rho} = \rho$. In particular, $\sum_j (I - W + R)^n_{ij} \to 0$ in Theorem 2.3 implies uniqueness in the class of tempered measures.

Of course, the point of Theorem 2.3 is that it provides a quantitative tool that goes far beyond qualitative uniqueness questions. It is therefore interesting to note that this single result nonetheless captures many uniqueness conditions that appear in the literature. In Corollary 2.6, Condition 3 is precisely the "influence on a site" condition of Weitz [25, Theorem 2.5] (our setting is even more general in that we do not require bounded-range interactions). Conditions 5 and 6 constitute a slight strengthening of the "influence of a site" condition of Weitz [25, Theorem 2.7] under summable metric or subexponential graph assumptions, in the spirit of the classical uniqueness condition of Dobrushin and Shlosman [4]. In the finite setting with single site updates, Condition 2 is in the spirit of [8] and Condition 4 is in the spirit of [7].

On the other hand, we can now see that Theorem 2.3 provides a crucial improvement over the classical comparison theorem. The single site setting of Corollary 2.4 corresponds essentially to the original Dobrushin uniqueness regime [5]. It is well known that this setting is restrictive, in that it captures only a small part of the parameter space where uniqueness of Gibbs measures holds. It is precisely for this reason that Dobrushin and Shlosman introduced their improved uniqueness criterion in terms of larger blocks [4], which in many cases allows to capture a large part of or even the entire uniqueness region; see [25, Sect. 5] for examples. The generalized comparison Theorem 2.3 in terms of larger blocks can therefore be fruitfully applied to a much larger and more natural class of models than the classical comparison theorem. This point will be further emphasized in the context of the application in Sect. 4.

Remark 2.3 The "influence of a site" condition $\|RW^{-1}\|_1 < 1$ that appears in Corollary 2.6 is slightly stronger than the corresponding condition of Dobrushin–Shlosman [4] and Weitz

[25, Theorem 2.7]. Writing out the definition of R, our condition reads

$$\|RW^{-1}\|_1 = \sup_{j \in I} W_{jj}^{-1} \sum_{i \in I} \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \frac{1}{\eta_j(x_j, z_j)} \sum_{J \in \mathfrak{J}: i \in J} w_J \, Q_{x,z}^J \eta_i < 1,$$

while the condition of [25, Theorem 2.7] (which extends the condition of [4]) reads

$$\sup_{j \in I} W_{jj}^{-1} \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \frac{1}{\eta_j(x_j, z_j)} \sum_{i \in I} \sum_{J \in \mathfrak{J}: i \in J} w_J \, Q_{x,z}^J \eta_i < 1.$$

The latter is slightly weaker as the sum over sites i appears inside the supremum over configurations x, z. While the distinction between these conditions is inessential in many applications, there do exist situations in which the weaker condition yields an essential improvement, see, e.g., [25, Sect. 5.3]. In such problems, Theorem 2.3 is not only limited by the stronger uniqueness condition but could also lead to poor quantitative bounds, as the comparison bound is itself expressed in terms of the uniform influence coefficients R_{ij}. It could therefore be of interest to develop comparison theorems that are able to exploit the finer structure that is present in the weaker uniqueness condition. In fact, the proof of Theorem 2.3 already indicates a natural approach to such improved bounds. However, the resulting comparison theorems are necessarily *nonlinear* in that the action of the matrix R is replaced by a nonlinear operator R. The nonlinear expressions are somewhat difficult to handle in practice, and as we do not at present have a compelling application for such bounds we do not pursue this direction here. However, for completeness, we will briefly sketch how such bounds can be obtained in Remark 3.1 below.

2.5 A One-Sided Comparison Theorem

As was discussed in Sect. 2.1, it is natural in many applications to describe high-dimensional probability distributions in terms of local conditional probabilities of the form $\mu(dz^J | x^{I \setminus J})$. This is in essence a static picture, where we describe the behavior of each local region J given that the configuration of the remaining sites $I \setminus J$ is frozen. In models that possess dynamics, this description is not very natural. In this setting, each site $i \in I$ occurs at a given time $\tau(i)$, and its state is only determined by the configuration of sites $j \in I$ in the past and present $\tau(j) \leq \tau(i)$, but not by the future. For example, the model might be defined as a high-dimensional Markov chain whose description is naturally given in terms of one-sided conditional probabilities (see, e.g., [9] and the application in Sect. 4). It is therefore interesting to note that the original comparison theorem of Dobrushin [5] is actually more general than Corollary 2.4 in that it is applicable both in the static and dynamic settings. We presently develop an analogous generalization to Theorem 2.3.

For the purposes of this section, we assume that we are given a function $\tau : I \to \mathbb{Z}$ that assigns to each site $i \in I$ an integer index $\tau(i)$. We define

$$I_{\leq k} := \{i \in I : \tau(i) \leq k\}, \quad \mathbb{S}_{\leq k} := \mathbb{S}^{I_{\leq k}},$$

and for any probability measure ρ on $\mathbb{S}$ we denote by $\rho_{\leq k}$ the marginal on $\mathbb{S}_{\leq k}$.

Definition 2.7 A *one-sided local update rule* for ρ is a collection $(\gamma^J)_{J \in \mathfrak{J}}$ where

1. $\mathfrak{J}$ is a cover of I such that $\min_{i \in J} \tau(i) = \max_{i \in J} \tau(i) =: \tau(J)$ for every $J \in \mathfrak{J}$.
2. γ^J is a transition kernel from $\mathbb{S}_{\leq \tau(J)}$ to $\mathbb{S}^J$.
3. $\rho_{\leq \tau(J)}$ is γ^J-invariant for every $J \in \mathfrak{J}$.

The canonical example of a one-sided local update rule is to consider the one-sided conditional distributions $\gamma_x^J(dz^J) = \rho(dz^J|x^{I_{\leq \tau(J)} \setminus J})$. This situation is particularly useful in the investigation of interacting Markov chains or probabilistic cellular automata, cf. [5,9], where $\tau(j)$ denotes the time index of the site j and we condition only on the past and present, but not on the future.

Definition 2.8 A *one-sided coupled update rule* for $(\rho, \tilde{\rho})$ is a collection of transition kernels $(\gamma^J, \tilde{\gamma}^J, Q^J, \hat{Q}^J)_{J \in \mathcal{J}}$ such that the following hold:

1. $(\gamma^J)_{J \in \mathcal{J}}$ and $(\tilde{\gamma}^J)_{J \in \mathcal{J}}$ are one-sided local update rules for ρ and $\tilde{\rho}$, respectively.
2. $Q_{x,z}^J$ is a coupling of γ_x^J, γ_z^J for $J \in \mathcal{J}$, $x, z \in \mathbb{S}_{\leq \tau(J)}$ s.t. $\mathrm{card}\{i : x_i \neq z_i\} = 1$.
3. $\hat{Q}_x^J$ is a coupling of $\gamma_x^J, \tilde{\gamma}_x^J$ for $J \in \mathcal{J}$, $x \in \mathbb{S}_{\leq \tau(J)}$.

We can now state a one-sided counterpart to Theorem 2.3.

Theorem 2.9 *Let $(\gamma^J, \tilde{\gamma}^J, Q^J, \hat{Q}^J)_{J \in \mathcal{J}}$ be a one-sided coupled update rule for $(\rho, \tilde{\rho})$ and let $(w_J)_{J \in \mathcal{J}}$ be a family of strictly positive weights. Define the matrices W and R and the vector a as in Theorem 2.3. Assume that γ^J is quasilocal for every $J \in \mathcal{J}$,*

$$\sum_{j \in I} D_{ij} \, (\rho \otimes \tilde{\rho}) \eta_j < \infty \quad \text{for all } i \in I \quad \text{where} \quad D := \sum_{n=0}^{\infty} (W^{-1}R)^n, \tag{2}$$

and that (1) holds. Then we have

$$|\rho f - \tilde{\rho} f| \leq \sum_{i,j \in I} \delta_i f \, D_{ij} \, W_{jj}^{-1} a_j$$

for any bounded measurable quasilocal function f such that $\delta_i f < \infty$ for all $i \in I$.

Note that the result of Theorem 2.9 is formally the same as that of Theorem 2.3, except that we have changed the nature of the update rules used in the coefficients. We also require a further assumption (2) in addition to assumption (1) of Theorem 2.3, but this is not restrictive in practice: in particular, it is readily verified that the conclusion of Theorem 2.9 holds under any of the conditions of Corollary 2.6.

3 Proofs

3.1 General Comparison Principle

The proof of Theorem 2.3 is derived from a general comparison principle for Markov chains that will be formalized in this section. The basic idea behind this principle is to consider two transition kernels G and $\tilde{G}$ on $\mathbb{S}$ such that $\rho G = G$ and $\tilde{\rho} \tilde{G} = \tilde{\rho}$. One should think of G as the transition kernel of a Markov chain that admits ρ as its invariant measure, and similarly for $\tilde{G}$. The comparison principle of this section bounds the difference between the invariant measures ρ and $\tilde{\rho}$ in terms of the transition kernels G and $\tilde{G}$. In the following sections, we will apply this principle to a specific choice of G and $\tilde{G}$ that is derived from the coupled update rule.

We first recall a standard notion in the analysis of high-dimensional Markov chains [9] (note that our indices are reversed as compared to the definition in [9]).

Definition 3.1 $(V_{ij})_{i,j \in I}$ is a *Wasserstein matrix* for a transition kernel G on $\mathbb{S}$ if

$$\delta_j Gf \le \sum_{i \in I} \delta_i f \, V_{ij}$$

for every $j \in I$ and bounded and measurable quasilocal function f.

We now state our general comparison principle.

Proposition 3.2 *Let G and $\tilde{G}$ be transition kernels on $\mathbb{S}$ such that $\rho G = \rho$ and $\tilde{\rho} \tilde{G} = \tilde{\rho}$, and let Q_x be a coupling between the measures G_x and $\tilde{G}_x$ for every $x \in \mathbb{S}$. Assume that G is quasilocal, and let V be a Wasserstein matrix for G. Then we have*

$$|\rho f - \tilde{\rho} f| \le \sum_{i,j \in I} \delta_i f \, N_{ij}^{(n)} \int^* \tilde{\rho}(dx) \, Q_x \eta_j + \sum_{i,j \in I} \delta_i f \, V_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j,$$

where we defined

$$N^{(n)} := \sum_{k=0}^{n-1} V^k,$$

for any bounded and measurable quasilocal function f and $n \ge 1$.

Theorem 2.3 will be derived from this result. Roughly speaking, we will design the kernel G such that $V = I - W + R$ is a Wasserstein matrix; then assumption (1) implies that the second term in Proposition 3.2 vanishes as $n \to \infty$, and the result of Theorem 2.3 reduces to some matrix algebra (as will be explained below, however, a more complicated argument is needed to obtain Theorem 2.3 in full generality).

To prove Proposition 3.2 we require a simple lemma.

Lemma 3.3 *Let Q be a coupling of probability measures μ, ν on $\mathbb{S}$. Then*

$$|\mu f - \nu f| \le \sum_{i \in I} \delta_i f \, Q \eta_i$$

for every bounded and measurable quasilocal function f.

Proof Let $J \in \mathcal{I}$. Enumerate its elements arbitrarily as $J = \{j_1, \ldots, j_r\}$, and define $J_k = \{j_1, \ldots, j_k\}$ for $1 \le k \le r$ and $J_0 = \varnothing$. Then we can evidently estimate

$$|f_x^J(z) - f_x^J(\tilde{z})| \le \sum_{k=1}^{r} |f_x^J(z^{J_k} \tilde{z}^{J \setminus J_k}) - f_x^J(z^{J_{k-1}} \tilde{z}^{J \setminus J_{k-1}})| \le \sum_{j \in J} \delta_j f \, \eta_j(z_j, \tilde{z}_j).$$

As f is quasilocal, we can let $J \uparrow I$ to obtain

$$|f(z) - f(\tilde{z})| \le \sum_{i \in I} \delta_i f \, \eta_i(z_i, \tilde{z}_i).$$

The result follows readily as $|\mu f - \nu f| \le \int |f(z) - f(\tilde{z})| \, Q(dz, d\tilde{z})$. $\square$

We now proceed to the proof of Proposition 3.2.

Proof (Proposition 3.2) We begin by writing

$$|\rho f - \tilde{\rho} f| = |\rho G^n f - \tilde{\rho} \tilde{G}^n f|$$

$$\leq \sum_{k=0}^{n-1} |\tilde{\rho} \tilde{G}^{n-k-1} G^{k+1} f - \tilde{\rho} \tilde{G}^{n-k} G^k f| + |\rho G^n f - \tilde{\rho} G^n f|$$

$$= \sum_{k=0}^{n-1} |\tilde{\rho} G G^k f - \tilde{\rho} \tilde{G} G^k f| + |\rho G^n f - \tilde{\rho} G^n f|.$$

As G is assumed quasilocal, $G^k f$ is quasilocal, and thus Lemma 3.3 yields

$$|\tilde{\rho} G G^k f - \tilde{\rho} \tilde{G} G^k f| \leq \int \tilde{\rho}(dx) |G_x G^k f - \tilde{G}_x G^k f|$$

$$\leq \int^* \tilde{\rho}(dx) \sum_{j \in I} \delta_j G^k f \, Q_x \eta_j$$

$$\leq \sum_{i,j \in I} \delta_i f \, V_{ij}^k \int^* \tilde{\rho}(dx) \, Q_x \eta_j.$$

Similarly, as $\rho \otimes \tilde{\rho}$ is a coupling of $\rho, \tilde{\rho}$, we obtain by Lemma 3.3

$$|\rho G^n f - \tilde{\rho} G^n f| \leq \sum_{j \in I} \delta_j G^n f \, (\rho \otimes \tilde{\rho}) \eta_j \leq \sum_{i,j \in I} \delta_i f \, V_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j.$$

Thus the proof is complete.   □

3.2 Gibbs Samplers

To put Proposition 3.2 to good use, we must construct kernels G and $\tilde{G}$ for which ρ and $\tilde{\rho}$ are invariant, and that admit tractable estimates for the quantities in the comparison theorem in terms of the coupled update rule $(\gamma^J, \tilde{\gamma}^J, Q^J, \hat{Q}^J)_{J \in \mathcal{J}}$ and weights $(w_J)_{J \in \mathcal{J}}$. To this end, we will use a standard construction called the *Gibbs sampler*: at each time, we draw a region $J \in \mathcal{J}$ with probability $v_J \propto w_J$, and then apply the transition kernel γ^J to the current configuration. This defines a transition kernel G for which ρ is G-invariant (as ρ is γ^J-invariant for every $J \in \mathcal{J}$). The construction for $\tilde{G}$ is identical. As will be explained below, this is not the most natural construction for the proof of our results; however, it will form the basis for further computations.

We fix throughout this section a coupled update rule $(\gamma^J, \tilde{\gamma}^J, Q^J, \hat{Q}^J)_{J \in \mathcal{J}}$ for $(\rho, \tilde{\rho})$ and weights $(w_J)_{J \in \mathcal{J}}$ satisfying the assumptions of Theorem 2.3. Let $\mathbf{v} = (v_J)_{J \in \mathcal{J}}$ be nonnegative weights such that $\sum_J v_J \leq 1$. We define the Gibbs samplers

$$G_x^{\mathbf{v}}(A) := \left(1 - \sum_{J \in \mathcal{J}} v_J\right) \mathbf{1}_A(x) + \sum_{J \in \mathcal{J}} v_J \int \mathbf{1}_A(z^J x^{I \setminus J}) \, \gamma_x^J(dz^J),$$

$$\tilde{G}_x^{\mathbf{v}}(A) := \left(1 - \sum_{J \in \mathcal{J}} v_J\right) \mathbf{1}_A(x) + \sum_{J \in \mathcal{J}} v_J \int \mathbf{1}_A(z^J x^{I \setminus J}) \, \tilde{\gamma}_x^J(dz^J).$$

Evidently $G^{\mathbf{v}}$ and $\tilde{G}^{\mathbf{v}}$ are transition kernels on $\mathbb{S}$, and $\rho G^{\mathbf{v}} = \rho$ and $\tilde{\rho} \tilde{G}^{\mathbf{v}} = \tilde{\rho}$ by construction. To apply Proposition 3.2, we must establish some basic properties.

Lemma 3.4 *Assume that γ^J is quasilocal for every $J \in \mathcal{J}$. Then $G^{\mathbf{v}}$ is quasilocal.*

Proof Let $f : \mathbb{S} \to \mathbb{S}$ be a bounded and measurable quasilocal function. It evidently suffices to show that $\gamma^J f^J$ is quasilocal for every $J \in \mathcal{J}$. To this end, let us fix $J \in \mathcal{J}$, $x, z \in \mathbb{S}$, and $J_1, J_2, \ldots \in \mathcal{J}$ such that $J_1 \subseteq J_2 \subseteq \cdots$ and $\bigcup_i J_i = I$. Then we have

$$\gamma^J_{z^{J_i} x^{I \setminus J_i}} \xrightarrow{i \to \infty} \gamma^J_z \quad \text{setwise}$$

as γ^J is quasilocal. On the other hand, we have

$$f^J_{z^{J_i} x^{I \setminus J_i}} \xrightarrow{i \to \infty} f^J_z \quad \text{pointwise}$$

as f is quasilocal. Thus by [19, Proposition 18, p. 270] we obtain

$$\gamma^J_{z^{J_i} x^{I \setminus J_i}} f^J_{z^{J_i} x^{I \setminus J_i}} \xrightarrow{i \to \infty} \gamma^J_z f^J_z.$$

As the choice of x, z and $(J_i)_{i \geq 1}$ is arbitrary, the result follows. □

Lemma 3.5 *Assume that γ^J is quasilocal for every $J \in \mathcal{J}$, and define*

$$W^{\mathbf{v}}_{ij} := \mathbf{1}_{i=j} \sum_{J \in \mathcal{J}: i \in J} v_J,$$

$$R^{\mathbf{v}}_{ij} := \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \frac{1}{\eta_j(x_j, z_j)} \sum_{J \in \mathcal{J}: i \in J} v_J \, Q^J_{x,z} \eta_i.$$

Then $V^{\mathbf{v}} = I - W^{\mathbf{v}} + R^{\mathbf{v}}$ is a Wasserstein matrix for $G^{\mathbf{v}}$.

Proof Let $f : \mathbb{S} \to \mathbb{S}$ be a bounded measurable quasilocal function, and let $x, z \in \mathbb{S}$ be configurations that differ at a single site $\mathrm{card}\{i \in I : x_i \neq z_i\} = 1$. Note that

$$\gamma^J_x f^J_x = (\gamma^J_x \otimes \delta_{x^{I \setminus J}}) f, \quad \gamma^J_z f^J_z = (\gamma^J_z \otimes \delta_{z^{I \setminus J}}) f.$$

As $Q^J_{x,z}$ is a coupling of γ^J_x and γ^J_z by construction, the measure $Q^J_{x,z} \otimes \delta_{x^{I \setminus J}} \otimes \delta_{z^{I \setminus J}}$ is a coupling of $\gamma^J_x \otimes \delta_{x^{I \setminus J}}$ and $\gamma^J_z \otimes \delta_{z^{I \setminus J}}$. Thus Lemma 3.3 yields

$$|\gamma^J_x f^J_x - \gamma^J_z f^J_z| \leq \sum_{i \in I} \delta_i f \, (Q^J_{x,z} \otimes \delta_{x^{I \setminus J}} \otimes \delta_{z^{I \setminus J}}) \eta_i$$

$$= \sum_{i \in J} \delta_i f \, Q^J_{x,z} \eta_i + \sum_{i \in I \setminus J} \delta_i f \, \eta_i(x_i, z_i).$$

In particular, we obtain

$$|G^{\mathbf{v}} f(x) - G^{\mathbf{v}} f(z)| \leq \left(1 - \sum_{J \in \mathcal{J}} v_J\right) |f(x) - f(z)| + \sum_{J \in \mathcal{J}} v_J \, |\gamma^J_x f^J_x - \gamma^J_z f^J_z|$$

$$\leq \left(1 - \sum_{J \in \mathcal{J}} v_J\right) \sum_{i \in I} \delta_i f \, \eta_i(x_i, z_i)$$

$$+ \sum_{J \in \mathcal{J}} v_J \left(\sum_{i \in J} \delta_i f \, Q^J_{x,z} \eta_i + \sum_{i \in I \setminus J} \delta_i f \, \eta_i(x_i, z_i)\right)$$

$$= \sum_{i \in I} \delta_i f \, \{1 - W^{\mathbf{v}}_{ii}\} \eta_i(x_i, z_i) + \sum_{i \in I} \delta_i f \sum_{J \in \mathcal{J}: i \in J} v_J \, Q^J_{x,z} \eta_i.$$

Now suppose that $x^{I\setminus\{j\}} = z^{I\setminus\{j\}}$ (and $x \neq z$). Then by definition

$$\sum_{J \in \mathcal{J}: i \in J} v_J \, Q_{x,z}^J \eta_i \leq R_{ij}^{\mathbf{v}} \, \eta_j(x_j, v_j),$$

and we obtain

$$\frac{|G^{\mathbf{v}} f(x) - G^{\mathbf{v}} f(z)|}{\eta_j(x_j, z_j)} \leq \delta_j f \{1 - W_{jj}^{\mathbf{v}}\} + \sum_{i \in I} \delta_i f \, R_{ij}^{\mathbf{v}}.$$

Thus $V^{\mathbf{v}} = I - W^{\mathbf{v}} + R^{\mathbf{v}}$ satisfies Definition 3.1.   □

Using Lemmas 3.4 and 3.5, we can now apply Proposition 3.2.

Corollary 3.6 *Assume that γ^J is quasilocal for every $J \in \mathcal{J}$. Then*

$$|\rho f - \tilde{\rho} f| \leq \sum_{i,j \in I} \delta_i f \, N_{ij}^{\mathbf{v}(n)} a_j^{\mathbf{v}} + \sum_{i,j \in I} \delta_i f \, (I - W^{\mathbf{v}} + R^{\mathbf{v}})_{ij}^n (\rho \otimes \tilde{\rho}) \eta_j$$

for every $n \geq 1$ and bounded and measurable quasilocal function f, where

$$N^{\mathbf{v}(n)} := \sum_{k=0}^{n-1} (I - W^{\mathbf{v}} + R^{\mathbf{v}})^k$$

and the coefficients $(a_j^{\mathbf{v}})_{j \in I}$ are defined by $a_j^{\mathbf{v}} := \sum_{J \in \mathcal{J}: j \in J} v_J \int^* \tilde{\rho}(dx) \, \hat{Q}_x^J \eta_j$.

Proof Let $G = G^{\mathbf{v}}, \tilde{G} = \tilde{G}^{\mathbf{v}}, V = I - W^{\mathbf{v}} + R^{\mathbf{v}}$ in Proposition 3.2. The requisite assumptions are verified by Lemmas 3.4 and 3.5, and it remains to show that there is a coupling Q_x of G_x and $\tilde{G}_x$ such that $\int^* \tilde{\rho}(dx) \, Q_x \eta_j \leq a_j$ for every $j \in I$. But

$$Q_x g := \left(1 - \sum_{J \in \mathcal{J}} v_J\right) g(x, x) + \sum_{J \in \mathcal{J}} v_J \int \hat{Q}_x^J(dz^J, d\tilde{z}^J) \, g(z^J x^{I\setminus J}, \tilde{z}^J x^{I\setminus J})$$

is easily verified to satisfy the necessary properties.   □

In order for the Gibbs sampler to make sense, the weights v_J must be probabilities. This imposes the requirement $\sum_J v_J \leq 1$. If we were to assume that $\sum_J w_J \leq 1$, we could apply Corollary 3.6 with $v_J = w_J$. Then assumption (1) guarantees that the second term in Corollary 3.6 vanishes as $n \to \infty$, which yields

$$|\rho f - \tilde{\rho} f| \leq \sum_{i,j \in I} \delta_i f \, N_{ij} a_j \quad \text{with} \quad N := \sum_{k=0}^{\infty} (I - W + R)^k.$$

The proof of Theorem 2.3 would now be complete after we establish the identity

$$N = \sum_{k=0}^{\infty} (I - W + R)^k = \sum_{k=0}^{\infty} (W^{-1} R)^k \, W^{-1} = D W^{-1}.$$

This simple matrix identity will be proved in the next section. The assumption that the weights w_J are summable is restrictive, however, when I is infinite: in Theorem 2.3 we only assume that $W_{ii} \leq 1$ for all i, so we cannot set $v_J = w_J$.

When the weights w_j are not summable, it is not natural to interpret them as probabilities. In this setting, a much more natural construction would be to consider a *continuous*

time counterpart of the Gibbs sampler called *Glauber dynamics*. To define this process, one attaches to each region $J \in \mathcal{J}$ an independent Poisson process with rate w_J, and applies the transition kernel γ^J at every jump time of the corresponding Poisson process. Thus w_J does not represent the probability of selecting the region J in one time step, but rather the frequency with which region J is selected in continuous time. Once this process has been defined, one would choose the transition kernel G to be the transition semigroup of the continuous time process on any fixed time interval. Proceeding with this construction we expect, at least formally, to obtain Theorem 2.3 under the stated assumptions.

Unfortunately, there are nontrivial technical issues involved in implementing this approach: it is not evident *a priori* that the continuous time construction defines a well-behaved Markov semigroup, so that it is unclear when the above program can be made rigorous. The existence of a semigroup has typically been established under more restrictive assumptions than we have imposed in the present setting [17]. In order to circumvent such issues, we will proceed by an alternate route. Formally, the Glauber dynamics can be obtained by an appropriate scaling limit of discrete time Gibbs samplers. We will also utilize this scaling, but instead of applying Proposition 3.2 to the limiting dynamics we will take the scaling limit directly in Corollary 3.6. Thus, while our intuition comes from the continuous time setting, we avoid some technicalities inherent in the construction of the limit dynamics. Instead, we now face the problem of taking limits of powers of infinite matrices. The requisite matrix algebra will be worked out in the following section.

Remark 3.1 Let us briefly sketch how the previous results can be sharpened to obtain a *nonlinear* comparison theorem that could lead to sharper bounds in some situations. Assume for simplicity that $\sum_J w_J \leq 1$. Then $V = I - W + R$ is a Wasserstein matrix for G by Lemma 3.5. Writing out the definitions, this means $\delta(Gf) \leq \delta(f)V$ where

$$(\beta V)_j = \sum_{i \in I} \beta_i \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \left\{ \mathbf{1}_{i=j} \left(1 - \sum_{J: i \in J} w_J \right) + \frac{1}{\eta_j(x_j, z_j)} \sum_{J: i \in J} w_J \, Q_{x,z}^J \eta_i \right\}$$

(here we interpret $\beta = (\beta_i)_{i \in I}$ and $\delta(f) = (\delta_i f)_{i \in I}$ as row vectors). However, from the proof of Lemma 3.5 we even obtain the sharper bound $\delta(Gf) \leq \mathsf{V}[\delta(f)]$ where

$$\mathsf{V}[\beta]_j := \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \sum_{i \in I} \beta_i \left\{ \mathbf{1}_{i=j} \left(1 - \sum_{J: i \in J} w_J \right) + \frac{1}{\eta_j(x_j, z_j)} \sum_{J: i \in J} w_J \, Q_{x,z}^J \eta_i \right\}$$

is defined with the supremum over configurations outside the sum. The nonlinear operator V can now be used much in the same way as the Wasserstein matrix V. In particular, following the identical proof as for Proposition 3.2, we immediately obtain

$$|\rho f - \tilde{\rho} f| \leq \sum_{j \in I} \sum_{k=0}^{n-1} \mathsf{V}^k[\delta(f)]_j \int^* \tilde{\rho}(dx) \, Q_x \eta_j + \sum_{j \in I} \mathsf{V}^n[\delta(f)]_j \, (\rho \otimes \tilde{\rho}) \eta_j,$$

where V^k denotes the kth iterate of the nonlinear operator V. Proceeding along these lines, one can develop nonlinear comparison theorems under Dobrushin–Shlosman type conditions (see Remark 2.3). The nonlinear expressions are somewhat tedious to handle, however, and we do not develop this idea further in this paper.

3.3 Proof of Theorem 2.3

Throughout this section, we work under the assumptions of Theorem 2.3. The main idea of the proof is the following continuous scaling limit of Corollary 3.6.

Proposition 3.7 *Let $t > 0$. Define the matrices*

$$N := \sum_{k=0}^{\infty} (I - W + R)^k, \quad V^{[t]} := \sum_{k=0}^{\infty} \frac{t^k e^{-t}}{k!} (I - W + R)^k.$$

Then we have, under the assumptions of Theorem 2.3,

$$|\rho f - \tilde{\rho} f| \leq \sum_{i,j \in I} \delta_i f \, N_{ij} \, a_j + \sum_{i,j \in I} \delta_i f \, V_{ij}^{[t]} \, (\rho \otimes \tilde{\rho}) \eta_j$$

for every bounded measurable quasilocal function f such that $\delta_i f < \infty$ for all $i \in I$.

Proof Without loss of generality, we will assume that f is a local function (so that only finitely many $\delta_i f$ are nonzero). The extension to quasilocal f follows readily by applying the local result to f_x^J and letting $J \uparrow I$ as in the proof of Lemma 3.3.

As the cover $\mathfrak{J}$ is at most countable (because $\mathfrak{I}$ is countable), we can enumerate its elements arbitrarily as $\mathfrak{J} = \{J_1, J_2, \ldots\}$. Define the weights $\mathbf{v}^r = (v_J^r)_{J \in \mathfrak{J}}$ as

$$v_J^r := \begin{cases} w_J & \text{when } J = J_k \text{ for } k \leq r, \\ 0 & \text{otherwise.} \end{cases}$$

For every $r \in \mathbb{N}$, the weight vector $u\mathbf{v}^r$ evidently satisfies $\sum_J u v_J^r \leq 1$ for all $u > 0$ sufficiently small (depending on r). The main idea of the proof is to apply Corollary 3.6 to the weight vector $\mathbf{v} = (t/n)\mathbf{v}^r$, then let $n \to \infty$, and finally $r \to \infty$.

Let us begin by considering the second term in Corollary 3.6. We can write

$$
\begin{aligned}
(I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})^n &= \left(\left(1 - \frac{t}{n}\right) I + \frac{t}{n} (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r}) \right)^n \\
&= \sum_{k=0}^{n} \binom{n}{k} \left(1 - \frac{t}{n}\right)^{n-k} \left(\frac{t}{n}\right)^k (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^k \\
&= \mathbf{E}[(I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^{Z_n}],
\end{aligned}
$$

where we defined the Binomial random variables $Z_n \sim \text{Bin}(n, t/n)$. The random variables Z_n converge weakly as $n \to \infty$ to the Poisson random variable $Z_\infty \sim \text{Pois}(t)$. To take the limit of the above expectation, we need a simple estimate.

Lemma 3.8 *Let $(c_j)_{j \in I}$ be any nonnegative vector. Then*

$$\sum_{j \in I} (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})_{ij}^k c_j \leq 2^k \max_{0 \leq \ell \leq k} \sum_{j \in I} (I - W + R)_{ij}^\ell c_j$$

for every $i \in I$ and $k \geq 0$.

Proof As $R^{\mathbf{v}}$ is nondecreasing in $\mathbf{v}$ we obtain the elementwise estimate

$$I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r} \leq I + R \leq I + (I - W + R),$$

where we have used $W_{ii} \leq 1$. We therefore have

$$\sum_{j \in I} (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})_{ij}^k \, c_j \leq \sum_{j \in I} (I + \{I - W + R\})_{ij}^k \, c_j$$

$$= \sum_{\ell=0}^{k} \binom{k}{\ell} \sum_{j \in I} (I - W + R)_{ij}^{\ell} \, c_j,$$

and the proof is easily completed. □

Define the random variables

$$X_n = g(Z_n) \quad \text{with} \quad g(k) = \sum_{i,j \in I} \delta_i f \, (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})_{ij}^k \, (\rho \otimes \tilde{\rho}) \eta_j.$$

Then $X_n \to X_\infty$ weakly by the continuous mapping theorem. On the other hand, applying Lemma 3.8 with $c_j = (\rho \otimes \tilde{\rho}) \eta_j$ we estimate $g(k) \leq C 2^k$ for some finite constant $C < \infty$ and all $k \geq 0$, where we have used (1) and that f is local. As

$$\limsup_{u \to \infty} \sup_{n \geq 1} \mathbf{E}[2^{Z_n} \mathbf{1}_{2^{Z_n} \geq u}] \leq \lim_{u \to \infty} u^{-1} \sup_{n \geq 1} \mathbf{E}[4^{Z_n}] = \lim_{u \to \infty} u^{-1} e^{3t} = 0,$$

it follows that the random variables $(X_n)_{n \geq 1}$ are uniformly integrable. We therefore conclude that $\mathbf{E}[X_n] \to \mathbf{E}[X_\infty]$ as $n \to \infty$ (cf. [14, Lemma 4.11]). In particular,

$$\lim_{n \to \infty} \sum_{i,j \in I} \delta_i f \, (I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j = \sum_{i,j \in I} \delta_i f \, V_{ij}^{r[t]} \, (\rho \otimes \tilde{\rho}) \eta_j,$$

where

$$V^{r[t]} = \sum_{k=0}^{\infty} \frac{t^k e^{-t}}{k!} (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^k.$$

Now let $r \to \infty$. Note that $W^{\mathbf{v}^r} \uparrow W$ and $R^{\mathbf{v}^r} \uparrow R$ elementwise and, as in the proof of Lemma 3.8, we have $I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r} \leq I + (I - W + R)$ elementwise where

$$\sum_{k=0}^{\infty} \sum_{i,j \in I} \frac{t^k e^{-t}}{k!} \delta_i f \, \{I + (I - W + R)\}_{ij}^k \, (\rho \otimes \tilde{\rho}) \eta_j$$

$$\leq e^t \sup_{\ell \geq 0} \sum_{i,j \in I} \delta_i f \, (I - W + R)_{ij}^{\ell} \, (\rho \otimes \tilde{\rho}) \eta_j$$

is finite by (1) and as f is local. We therefore obtain by dominated convergence

$$\lim_{r \to \infty} \lim_{n \to \infty} \sum_{i,j \in I} \delta_i f \, (I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j = \sum_{i,j \in I} \delta_i f \, V_{ij}^{[t]} \, (\rho \otimes \tilde{\rho}) \eta_j.$$

That is, the second term in Corollary 3.6 with $\mathbf{v} = (t/n)\mathbf{v}^r$ converges as $n \to \infty$ and $r \to \infty$ to the second term in statement of the present result.

It remains to establish the corresponding conclusion for the first term in Corollary 3.6, which proceeds much along the same lines. We begin by noting that

$$
\frac{1}{n}\sum_{k=0}^{n-1}(I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})^k = \frac{1}{n}\sum_{k=0}^{n-1}\left(\left(1 - \frac{t}{n}\right)I + \frac{t}{n}(I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})\right)^k
$$

$$
= \frac{1}{n}\sum_{k=0}^{n-1}\sum_{\ell=0}^{k}\binom{k}{\ell}\left(1 - \frac{t}{n}\right)^{k-\ell}\left(\frac{t}{n}\right)^\ell (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^\ell
$$

$$
= \sum_{\ell=0}^{n-1}p_\ell^{(n)}(I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^\ell,
$$

where we have defined

$$
p_\ell^{(n)} = \frac{1}{n}\sum_{k=\ell}^{n-1}\binom{k}{\ell}\left(1 - \frac{t}{n}\right)^{k-\ell}\left(\frac{t}{n}\right)^\ell
$$

$$
= \frac{1}{t}\int_{\ell t/n}^{t}\binom{\lfloor sn/t\rfloor}{\ell}\left(1 - \frac{t}{n}\right)^{\lfloor sn/t\rfloor-\ell}\left(\frac{t}{n}\right)^\ell ds
$$

for $\ell < n$. An elementary computation yields

$$
\sum_{\ell=0}^{n-1}p_\ell^{(n)} = 1 \quad\text{and}\quad p_\ell^{(n)} \xrightarrow{n\to\infty} p_\ell^{(\infty)} = \frac{1}{t}\int_0^t \frac{s^\ell e^{-s}}{\ell!}\,ds.
$$

We can therefore introduce $\{0, 1, \ldots\}$-valued random variables Y_n with $\mathbf{P}[Y_n = \ell] = p_\ell^{(n)}$ for $\ell < n$, and we have shown above that $Y_n \to Y_\infty$ weakly and that

$$
\frac{1}{n}\sum_{k=0}^{n-1}(I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})^k = \mathbf{E}[(I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^{Y_n}].
$$

The first term in Corollary 3.6 with $\mathbf{v} = (t/n)\mathbf{v}^r$ can be written as

$$
\sum_{i,j\in I}\delta_i f\sum_{k=0}^{n-1}(I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})_{ij}^k\, a_j^{(t/n)\mathbf{v}^r} = t\,\mathbf{E}[h(Y_n)],
$$

where we have defined

$$
h(k) = \sum_{i,j\in I}\delta_i f\,(I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})_{ij}^k\, a_j^{\mathbf{v}^r}.
$$

We now proceed essentially as above. We can assume without loss of generality that

$$
\sup_{\ell\geq 0}\sum_{i,j\in I}\delta_i f\,(I - W + R)_{ij}^\ell\, a_j < \infty,
$$

as otherwise the right-hand side in the statement of the present result is infinite and the estimate is trivial. It consequently follows from Lemma 3.8 that $h(k) \leq C2^k$ for some finite constant $C < \infty$ and all $k \geq 0$. A similar computation as above shows that $(h(Y_n))_{n\geq 0}$ is uniformly integrable, and thus $\mathbf{E}[h(Y_n)] \to \mathbf{E}[h(Y_\infty)]$. In particular, the first term in

Corollary 3.6 with $\mathbf{v} = (t/n)\mathbf{v}^r$ converges as $n \to \infty$ to

$$\lim_{n\to\infty} \sum_{i,j\in I} \delta_i f \sum_{k=0}^{n-1} (I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})_{ij}^k \, a_j^{(t/n)\mathbf{v}^r} = \sum_{i,j\in I} \delta_i f \, N_{ij}^r \, a_j^{\mathbf{v}^r},$$

where

$$N^r = \sum_{k=0}^{\infty} \int_0^t \frac{s^k e^{-s}}{k!} \, ds \, (I - W^{\mathbf{v}^r} + R^{\mathbf{v}^r})^k.$$

Similarly, letting $r \to \infty$ and repeating exactly the arguments used above for the second term of Corollary 3.6, we obtain by dominated convergence

$$\lim_{r\to\infty} \lim_{n\to\infty} \sum_{i,j\in I} \delta_i f \sum_{k=0}^{n-1} (I - W^{(t/n)\mathbf{v}^r} + R^{(t/n)\mathbf{v}^r})_{ij}^k \, a_j^{(t/n)\mathbf{v}^r} = \sum_{i,j\in I} \delta_i f \, \tilde{N}_{ij} \, a_j,$$

where

$$\tilde{N} = \sum_{k=0}^{\infty} \int_0^t \frac{s^k e^{-s}}{k!} \, ds \, (I - W + R)^k.$$

To conclude, we have shown that applying Corollary 3.6 to the weight vector $\mathbf{v} = (t/n)\mathbf{v}^r$ and taking the limit as $n \to \infty$ and $r \to \infty$, respectively, yields the estimate

$$|\rho f - \tilde{\rho} f| \le \sum_{i,j\in I} \delta_i f \, \tilde{N}_{ij} \, a_j + \sum_{i,j\in I} \delta_i f \, V_{ij}^{[t]} \, (\rho \otimes \tilde{\rho}) \eta_j.$$

It remains to note that $t^k e^{-t}/k!$ is the density of a Gamma distribution (with shape $k + 1$ and scale 1), so $\int_0^t s^k e^{-s}/k! \, ds \le 1$ and thus $\tilde{N} \le N$ elementwise. □

We can now complete the proof of Theorem 2.3.

Proof (Theorem 2.3) Once again, we will assume without loss of generality that f is a local function (so that only finitely many $\delta_i f$ are nonzero). The extension to quasilocal f follows readily by localization as in the proof of Lemma 3.3.

We begin by showing that the second term in Proposition 3.7 vanishes as $t \to \infty$. Indeed, for any $n \ge 0$, we can evidently estimate the second term as

$$\sum_{k=0}^{\infty} \frac{t^k e^{-t}}{k!} \sum_{i,j\in I} \delta_i f \, (I - W + R)_{ij}^k \, (\rho \otimes \tilde{\rho}) \eta_j \le \sup_{\ell\ge 0} \sum_{i,j\in I} \delta_i f \, (I - W + R)_{ij}^\ell \, (\rho \otimes \tilde{\rho}) \eta_j \sum_{k=0}^{n} \frac{t^k e^{-t}}{k!}$$

$$+ \sup_{\ell>n} \sum_{i,j\in I} \delta_i f \, (I - W + R)_{ij}^\ell \, (\rho \otimes \tilde{\rho}) \eta_j.$$

By assumption (1) and as f is local, the two terms on the right vanish as $t \to \infty$ and $n \to \infty$, respectively. Thus second term in Proposition 3.7 vanishes as $t \to \infty$.

We have now proved the estimate

$$|\rho f - \tilde{\rho} f| \le \sum_{i,j\in I} \delta_i f \, N_{ij} \, a_j.$$

To complete the proof of Theorem 2.3, it remains to establish the identity $N = DW^{-1}$. This is an exercise in matrix algebra. By the definition of the matrix product

$$(I - W + R)^p = \sum_{k=0}^{p} \sum_{\substack{n_0,\dots,n_k \geq 0 \\ n_0+\cdots+n_k=p-k}} (I - W)^{n_k} R \cdots (I - W)^{n_1} R (I - W)^{n_0}.$$

We can therefore write

$$\sum_{p=0}^{\infty} (I - W + R)^p = \sum_{k=0}^{\infty} \sum_{n_0,\dots,n_k \geq 0} \sum_{p=0}^{\infty} \mathbf{1}_{n_0+\cdots+n_k=p-k} \mathbf{1}_{k \leq p} (I - W)^{n_k} R \cdots (I - W)^{n_1} R (I - W)^{n_0}$$

$$= \sum_{k=0}^{\infty} \sum_{n_0,\dots,n_k \geq 0} (I - W)^{n_k} R \cdots (I - W)^{n_1} R (I - W)^{n_0}$$

$$= \sum_{k=0}^{\infty} (W^{-1} R)^k W^{-1},$$

where we used $W^{-1} = \sum_{n=0}^{\infty} (I - W)^n$ as W is diagonal with $0 < W_{ii} \leq 1$.   □

3.4 Proof of Corollary 2.6

Note that $\sup_i W_{ii} < \infty$ in all parts of Corollary 2.6 (either by assumption or as card $I < \infty$). Moreover, it is easily seen that all parts of Corollary 2.6 as well as the conclusion of Theorem 2.3 are unchanged if all the weights are multiplied by the same constant. We may therefore assume without loss of generality that $\sup_i W_{ii} \leq 1$.

Next, we note that as ρ and $\tilde{\rho}$ are tempered, we have

$$\sup_{i \in I} (\rho \otimes \tilde{\rho}) \eta_i \leq \sup_{i \in I} \rho \, \eta_i(\cdot, x_i^\star) + \sup_{i \in I} \tilde{\rho} \, \eta_i(x_i^\star, \cdot) < \infty$$

by the triangle inequality. To verify (1), it therefore suffices to show that

$$\lim_{k \to \infty} \sum_{j \in I} (I - W + R)_{ij}^k = 0 \quad \text{for all } i \in I. \tag{3}$$

We now proceed to verify this condition in the different cases of Corollary 2.6.

Proof (Corollary 2.6.1) It was shown at the end of the proof of Theorem 2.3 that

$$\sum_{k=0}^{\infty} (I - W + R)^k = \sum_{k=0}^{\infty} (W^{-1} R)^k W^{-1} = DW^{-1}.$$

As W^{-1} has finite entries, $D < \infty$ certainly implies that $(I - W + R)^k \to 0$ as $k \to \infty$ elementwise. But this trivially yields (3) when card $I < \infty$.   □

Proof (Corollary 2.6.2) Note that we can write

$$D = \sum_{k=0}^{\infty} (W^{-1} R)^k = \sum_{p=0}^{n-1} (W^{-1} R)^p \sum_{k=0}^{\infty} (W^{-1} R)^{nk}.$$

Therefore, if $R < \infty$ and $\|(W^{-1} R)^n\| < 1$, we can estimate

$$\|D\| \leq \left\| \sum_{p=0}^{n-1} (W^{-1} R)^p \right\| \sum_{k=0}^{\infty} \|(W^{-1} R)^n\|^k < \infty.$$

Thus $D < \infty$ and we conclude by the previous part. □

Proof (Corollary 2.6.3) We give a simple probabilistic proof (a more complicated matrix-analytic proof could be given along the lines of [20, Theorem 3.21]). Let $P = W^{-1}R$. As $\|P\|_\infty < 1$, the infinite matrix P is substochastic. Thus P is the transition probability matrix of a killed Markov chain $(X_n)_{n\geq 0}$ with $\mathbf{P}[X_n = j|X_{n-1} = i] = P_{ij}$ and $\mathbf{P}[X_n \text{ is dead}|X_{n-1} = i] = 1 - \sum_j P_{ij}$ (once the chain dies, it stays dead). Denote by $\zeta = \inf\{n : X_n \text{ is dead}\}$ the killing time of the chain. Then we obtain

$$\mathbf{P}[\zeta > n|X_0 = i] = \mathbf{P}[X_n \text{ is not dead}|X_0 = i] = \sum_{j\in I} P_{ij}^n \leq \|P^n\|_\infty \leq \|P\|_\infty^n.$$

Therefore, as $\|P\|_\infty < 1$, we find by letting $n \to \infty$ that $\mathbf{P}[\zeta = \infty|X_0 = i] = 0$. That is, the chain dies eventually with unit probability for any initial condition.

Now define $\tilde{P} = I - W + R = I - W + WP$. As $\sup_i W_{ii} \leq 1$, the matrix $\tilde{P}$ is also substochastic and corresponds to the following transition mechanism. If $X_{n-1} = i$, then at time n we flip a biased coin that comes up heads with probability W_{ii}. In case of heads we make a transition according to the matrix P, but in case of tails we leave the current state unchanged. From this description, it is evident that we can construct a Markov chain $(\tilde{X}_n)_{n\geq 0}$ with transition matrix $\tilde{P}$ by modifying the chain $(X_n)_{n\geq 0}$ as follows. Conditionally on $(X_n)_{n\geq 0}$, draw independent random variables $(\xi_n)_{n\geq 0}$ such that ξ_n is geometrically distributed with parameter $W_{X_n X_n}$. Now define the process $(\tilde{X}_n)_{n\geq 0}$ such that it stays in state X_0 for the first ξ_0 time steps, then is in state X_1 for the next ξ_1 time steps, etc. By construction, the resulting process is Markov with transition matrix $\tilde{P}$. Moreover, as $\zeta < \infty$ a.s., we have $\tilde{\zeta} := \inf\{n : \tilde{X}_n \text{ is dead}\} < \infty$ a.s. also. We therefore obtain

$$\lim_{n\to\infty} \sum_{j\in I}(I - W + R)_{ij}^n = \lim_{n\to\infty} \mathbf{P}[\tilde{\zeta} > n|X_0 = i] = 0$$

for every $i \in I$. We have therefore established (3). □

Proof (Corollary 2.6.4) We begin by writing as above

$$\sum_{k=0}^\infty (I - W + R)^k = \sum_{k=0}^\infty (W^{-1}R)^k W^{-1} = \sum_{k=0}^\infty W^{-1}(RW^{-1})^k,$$

where the last identity is straightforward. Arguing as in Corollary 2.6.2, we obtain

$$W_{ii}\sum_{k=0}^\infty \sum_{j\in I}(I - W + R)_{ij}^k = \sum_{j\in I}\sum_{k=0}^\infty (RW^{-1})_{ij}^k \leq \left\|\sum_{k=0}^\infty (RW^{-1})^k\right\|_\infty$$

$$\leq \sum_{p=0}^{n-1}\|RW^{-1}\|_\infty^p \sum_{k=0}^\infty \|(RW^{-1})^n\|_\infty^k < \infty.$$

It follows immediately that (3) holds. □

Proof (Corollary 2.6.5) Note that

$$\sum_{j\in I}(RW^{-1})_{ij}^k\|\eta_j\| \leq \|(RW^{-1})^k\|_1 \sum_{j\in I}\|\eta_j\| \leq \|RW^{-1}\|_1^k \sum_{j\in I}\|\eta_j\|.$$

Thus $\sum_j \|\eta_j\| < \infty$ and $\|RW^{-1}\|_1 < 1$ yield

$$\sum_{k=0}^{\infty} \sum_{j \in I} (I - W + R)_{ij}^k \|\eta_j\| = W_{ii}^{-1} \sum_{k=0}^{\infty} \sum_{j \in I} (RW^{-1})_{ij}^k \|\eta_j\| < \infty,$$

which evidently implies

$$\lim_{k \to \infty} \sum_{j \in I} (I - W + R)_{ij}^k (\rho \otimes \tilde{\rho}) \eta_j = 0 \quad \text{for all } i \in I.$$

We have therefore established (1).   □

Proof (Corollary 2.6.6) Let $r = \sup\{m(i, j) : R_{ij} > 0\}$ (which is finite by assumption), and choose $\beta > 0$ such that $\|RW^{-1}\|_1 < e^{-\beta r}$. Then we can estimate

$$\|RW^{-1}\|_{1,\beta m} := \sup_{j \in I} \sum_{i \in I} e^{\beta m(i,j)} (RW^{-1})_{ij} \le e^{\beta r} \|RW^{-1}\|_1 < 1.$$

As m is a pseudometric, it satisfies the triangle inequality and it is therefore easily seen that $\|\cdot\|_{1,\beta m}$ is a matrix norm. In particular, we can estimate

$$e^{\beta m(i,j)} (RW^{-1})_{ij}^n \le \|(RW^{-1})^n\|_{1,\beta m} \le \|RW^{-1}\|_{1,\beta m}^n$$

for every $i, j \in I$. But then

$$\|(RW^{-1})^n\|_{\infty} = \sup_{i \in I} \sum_{j \in I} (RW^{-1})_{ij}^n \le \|RW^{-1}\|_{1,\beta m}^n \sup_{i \in I} \sum_{j \in I} e^{-\beta m(i,j)} < \infty$$

for all n. We therefore have $\|RW^{-1}\|_{\infty} < \infty$, and we can choose n sufficiently large that $\|(RW^{-1})^n\|_{\infty} < 1$. The conclusion now follows from Corollary 2.6.4.   □

3.5 Proof of Theorem 2.9

In the case of one-sided local updates, the measure $\rho_{\le k}$ is γ^J-invariant for $\tau(J) = k$ (but not for $\tau(J) < k$). The proof of Theorem 2.9 proceeds by induction on k. In each stage of the induction, we apply the logic of Theorem 2.3 to the partial local updates $(\gamma^J)_{J \in \mathcal{J}:\tau(J)=k}$, and use the induction hypothesis to estimate the remainder.

Throughout this section, we work in the setting of Theorem 2.9. Define

$$I_{\le k} := \{i \in I : \tau(i) \le k\}, \quad I_k := \{i \in I : \tau(i) = k\}.$$

We can assume without loss of generality that $R_{ij} = 0$ when $\tau(j) > \tau(i)$. Indeed, the local update rule γ_x^J does not depend on x_j for $\tau(j) > \tau(J)$, so we can trivially choose the coupling $Q_{x,z}^J$ for $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$ such that $Q_{x,z}^J \eta_i = 0$ for all $i \in J$. On the other hand, the choice $R_{ij} = 0$ evidently yields the smallest bound in Theorem 2.9. In the sequel, we will always assume that $R_{ij} = 0$ whenever $\tau(j) > \tau(i)$.

The key induction step is formalized by the following result.

Proposition 3.9 *Assume (1). Let $(\beta_i)_{i \in I_{\le k-1}}$ be nonnegative weights such that*

$$|\rho_{\le k-1} g - \tilde{\rho}_{\le k-1} g| \le \sum_{i \in I_{\le k-1}} \delta_i g \, \beta_i$$

for every bounded measurable quasilocal function g on $\mathbb{S}_{\leq k-1}$ s.t. $\delta_i g < \infty$ $\forall i$. Then

$$|\rho_{\leq k} f - \tilde{\rho}_{\leq k} f| \leq \sum_{j \in I_{\leq k-1}} \left\{ \delta_j f + \sum_{i,l \in I_k} \delta_i f \, D_{il} \, (W^{-1} R)_{lj} \right\} \beta_j + \sum_{i,j \in I_k} \delta_i f \, D_{ij} \, W_{jj}^{-1} a_j$$

for every bounded measurable quasilocal function f on $\mathbb{S}_{\leq k}$ with $\delta_i f < \infty$ $\forall i$.

Proof We fix throughout the proof a bounded and measurable local function $f : \mathbb{S}_{\leq k} \to \mathbb{R}$ such that $\delta_i f < \infty$ for all $i \in I_{\leq k}$. The extension of the conclusion to quasilocal functions f follows readily by localization as in the proof of Lemma 3.3.

Let $G^{\mathbf{v}}$ and $\tilde{G}^{\mathbf{v}}$ be the Gibbs samplers defined in section 3.2. Enumerate the partial cover $\{J \in \mathcal{J} : \tau(J) = k\}$ as $\{J_1, J_2, \ldots\}$, and define weights $\mathbf{v}^r$ as in the proof of Proposition 3.7. By the definition of the one-sided local update rule, $\rho_{\leq k}$ is $G^{u\mathbf{v}^r}$-invariant and $\tilde{\rho}_{\leq k}$ is $\tilde{G}^{u\mathbf{v}^r}$-invariant for every r, u such that $\sum_J u v_J^r \leq 1$. Thus

$$|\rho_{\leq k} f - \tilde{\rho}_{\leq k} f| \leq \sum_{i,j \in I_{\leq k}} \delta_i f \, N_{ij}^{u\mathbf{v}^r(n)} a_j^{u\mathbf{v}^r} + |\rho_{\leq k}(G^{u\mathbf{v}^r})^n f - \tilde{\rho}_{\leq k}(G^{u\mathbf{v}^r})^n f|$$

as in the proof of Corollary 3.6, with the only distinction that we refrain from using the Wasserstein matrix to expand the second term in the proof of Proposition 3.2. We now use the induction hypothesis to obtain an improved estimate for the second term.

Lemma 3.10 *We can estimate*

$$|\rho_{\leq k} g - \tilde{\rho}_{\leq k} g| \leq \sum_{i \in I_{\leq k-1}} \delta_i g \, \beta_i + 3 \sum_{i \in I_k} \delta_i g \, (\rho \otimes \tilde{\rho}) \eta_i$$

for any bounded measurable quasilocal function $g : \mathbb{S}_{\leq k} \to \mathbb{R}$ with $\delta_i g < \infty$ $\forall i$.

Proof For any $x \in \mathbb{S}_{\leq k}$ we can estimate

$$|\rho_{\leq k} g - \tilde{\rho}_{\leq k} g| \leq |\rho_{\leq k-1} \hat{g}_x - \tilde{\rho}_{\leq k-1} \hat{g}_x| + |\rho_{\leq k}(g - \hat{g}_x)| + |\tilde{\rho}_{\leq k}(g - \hat{g}_x)|,$$

where we defined $\hat{g}_x(z) := g(z^{I_{\leq k-1}} x^{I_k})$. By Lemma 3.3 we have

$$|g(z) - \hat{g}_x(z)| \leq \sum_{i \in I_k} \delta_i g \, \eta_i(z_i, x_i).$$

We can therefore estimate using the induction hypothesis and the triangle inequality

$$|\rho_{\leq k} g - \tilde{\rho}_{\leq k} g| \leq \sum_{i \in I_{\leq k-1}} \delta_i g \, \beta_i + \sum_{i \in I_k} \delta_i g \, \{\rho \eta_i(\,\cdot\,, \tilde{X}_i) + \eta_i(\tilde{X}_i, x_i) + \tilde{\rho} \eta_i(\,\cdot\,, x_i)\}$$

for all $x, \tilde{X} \in \mathbb{S}_{\leq k}$. Now integrate this expression with respect to $\rho(dx) \tilde{\rho}(d\tilde{X})$. $\square$

To lighten the notation somewhat we will write $\mathbf{v} = u\mathbf{v}^r$ until further notice. Note that by construction $a_j^{\mathbf{v}} = 0$ whenever $\tau(j) < k$, while $R_{ij}^{\mathbf{v}} = 0$ whenever $\tau(j) > \tau(i)$ by assumption. Thus we obtain using Lemma 3.10 and Lemma 3.5

$$|\rho_{\leq k} f - \tilde{\rho}_{\leq k} f| \leq \sum_{i,j \in I_k} \delta_i f \, N_{ij}^{\mathbf{v}(n)} a_j^{\mathbf{v}} + 3 \sum_{i,j \in I_k} \delta_i f \, (I - W^{\mathbf{v}} + R^{\mathbf{v}})_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j$$

$$+ \sum_{i \in I_{\leq k}} \sum_{j \in I_{\leq k-1}} \delta_i f \, (I - W^{\mathbf{v}} + R^{\mathbf{v}})_{ij}^n \, \beta_j,$$

provided that $\sum_i \delta_i f \, (I - W^{\mathbf{v}} + R^{\mathbf{v}})_{ij}^n < \infty$ for all j.

As $v_J = 0$ for $\tau(J) < k$, we have $R_{ij}^{\mathbf{v}} = W_{ij}^{\mathbf{v}} = 0$ for $i \in I_{\leq k-1}$. Thus

$$V^{\mathbf{v}} = I - W^{\mathbf{v}} + R^{\mathbf{v}} = \begin{pmatrix} \check{V}^{\mathbf{v}} & \check{R}^{\mathbf{v}} \\ 0 & I \end{pmatrix},$$

where $\check{V}^{\mathbf{v}} := (V_{ij}^{\mathbf{v}})_{i,j \in I_k}$ and $\check{R}^{\mathbf{v}} := (R_{ij}^{\mathbf{v}})_{i \in I_k, j \in I_{\leq k-1}}$. In particular,

$$(I - W^{\mathbf{v}} + R^{\mathbf{v}})^n = \begin{pmatrix} (\check{V}^{\mathbf{v}})^n & \sum_{k=0}^{n-1} (\check{V}^{\mathbf{v}})^k \check{R}^{\mathbf{v}} \\ 0 & I \end{pmatrix}.$$

Moreover, as $R_{ij}^{\mathbf{v}} = 0$ whenever $\tau(j) > \tau(i)$, we evidently have $(\check{V}^{\mathbf{v}})_{ij}^k = (V^{\mathbf{v}})_{ij}^k$ for $i, j \in I_k$. Substituting into the above expression, we obtain

$$|\rho_{\leq k} f - \tilde{\rho}_{\leq k} f| \leq \sum_{i,j \in I_k} \delta_i f \, N_{ij}^{\mathbf{v}(n)} \, a_j^{\mathbf{v}} + 3 \sum_{i,j \in I_k} \delta_i f \, (I - W^{\mathbf{v}} + R^{\mathbf{v}})_{ij}^n \, (\rho \otimes \tilde{\rho}) \eta_j$$

$$+ \sum_{j \in I_{\leq k-1}} \left\{ \delta_j f + \sum_{i,l \in I_k} \delta_i f \, N_{il}^{\mathbf{v}(n)} \, R_{lj}^{\mathbf{v}} \right\} \beta_j$$

provided that $\sum_i \delta_i f \, (I - W^{\mathbf{v}} + R^{\mathbf{v}})_{ij}^n < \infty$ for all j. But the latter is easily verified using (1) and Lemma 3.8, as f is local and $\delta_i f < \infty$ for all i by assumption.

The rest of the proof now proceeds precisely as in the proof of Proposition 3.7 and Theorem 2.3. We set $\mathbf{v} = (t/n)\mathbf{v}^r$, let $n \to \infty$ and then $r \to \infty$. The arguments for the first two terms are identical to the proof of Proposition 3.7, while the argument for the third term is essentially identical to the argument for the first term. The proof is then completed as in Theorem 2.3. We leave the details for the reader. □

We now proceed to complete the proof of Theorem 2.9.

Proof (Theorem 2.9) Consider first the case that $k_- := \inf_{i \in I} \tau(i) > -\infty$. In this setting, we say that the comparison theorem holds for a given $k \geq k_-$ if we have

$$|\rho_{\leq k} f - \tilde{\rho}_{\leq k} f| \leq \sum_{i,j \in I_{\leq k}} \delta_i f \, D_{ij} \, W_{jj}^{-1} a_j$$

for every bounded measurable quasilocal function f on $\mathbb{S}_{\leq k}$ such that $\delta_i f < \infty \, \forall i$. Theorem 2.3 shows that the comparison theorem holds for k_-. We will now use Proposition 3.9 to show that if the comparison theorem holds for $k - 1$, then it holds for k also. Then the comparison theorem holds for every $k \geq k_-$ by induction, so the conclusion of Theorem 2.9 holds whenever f is a local function. The extension to quasilocal f follows readily by localization as in the proof of Lemma 3.3.

We now complete the induction step. When the comparison theorem holds for $k - 1$ (the induction hypothesis), we can apply Proposition 3.9 with

$$\beta_i = \sum_{j \in I_{\leq k-1}} D_{ij} \, W_{jj}^{-1} a_j.$$

This gives

$$|\rho_{\leq k} f - \tilde{\rho}_{\leq k} f| \leq \sum_{j,q \in I_{\leq k-1}} \sum_{i,l \in I_k} \delta_i f \, D_{il} \, (W^{-1} R)_{lq} \, D_{qj} \, W_{jj}^{-1} a_j$$

$$+ \sum_{i,j \in I_{\leq k-1}} \delta_i f \, D_{ij} \, W_{jj}^{-1} a_j + \sum_{i,j \in I_k} \delta_i f \, D_{ij} \, W_{jj}^{-1} a_j$$

for every bounded measurable quasilocal function f on $\mathbb{S}_{\leq k}$ so that $\delta_i f < \infty \; \forall i$. To complete the proof, it therefore suffices to show that we have

$$D_{ij} = \sum_{q \in I_{\leq k-1}} \sum_{l \in I_k} D_{il} \, (W^{-1}R)_{lq} \, D_{qj} \qquad \text{for } i \in I_k, \; j \in I_{\leq k-1}.$$

To see this, note that as $R_{ij} = 0$ for $\tau(i) < \tau(j)$, we can write

$$D_{ij} = \sum_{p=1}^{\infty} \sum_{\substack{j_1,\ldots,j_{p-1} \in I: \\ \tau(j) \leq \tau(j_1) \leq \cdots \leq \tau(j_{p-1}) \leq k}} (W^{-1}R)_{ij_{p-1}} \cdots (W^{-1}R)_{j_2 j_1} (W^{-1}R)_{j_1 j}$$

$$= \sum_{p=1}^{\infty} \sum_{n=1}^{p} \sum_{l \in I_k} \sum_{q \in I_{\leq k-1}} (W^{-1}R)_{il}^{n-1} (W^{-1}R)_{lq} (W^{-1}R)_{qj}^{p-n}$$

for $i \in I_k$ and $j \in I_{\leq k-1}$, where we have used that when $\tau(j_1) \leq \cdots \leq \tau(j_{p-1}) \leq k$ there exists $1 \leq n \leq p$ such that $j_1, \ldots, j_{p-n} \in I_{\leq k-1}$ and $j_{p-n+1}, \ldots, j_{p-1} \in I_k$. Rearranging yields the desired identity for D_{ij}, completing the proof for the case $k_- > -\infty$ (note that in this case the additional assumption (2) was not needed).

We now turn to the case $k_- = -\infty$. Let us say that $(\beta_i)_{i \in I_{\leq k}}$ is a k-estimate if

$$|\rho_{\leq k} g - \tilde{\rho}_{\leq k} g| \leq \sum_{i \in I_{\leq k}} \delta_i g \, \beta_i$$

for every bounded measurable quasilocal function g on $\mathbb{S}_{\leq k}$ such that $\delta_i g < \infty \; \forall i$. Then the conclusion of Proposition 3.9 can be reformulated as follows: if $(\beta_i)_{i \in I_{\leq k-1}}$ is a $(k-1)$-estimate, then $(\beta'_i)_{i \in I_{\leq k}}$ is a k-estimate with $\beta'_i = \beta_i$ for $i \in I_{\leq k-1}$ and

$$\beta'_i = \sum_{j \in I_{\leq k-1}} \sum_{l \in I_k} D_{il} \, (W^{-1}R)_{lj} \, \beta_j + \sum_{j \in I_k} D_{ij} \, W_{jj}^{-1} a_j$$

for $i \in I_k$. Thus we can repeatedly apply Proposition 3.9 to extend an initial estimate. In particular, if we fix $k \in \mathbb{Z}$ and $n \geq 1$, and if $(\beta_i)_{i \in I_{\leq k-n}}$ is a $(k-n)$-estimate, we obtain a k-estimate $(\beta'_i)_{i \in I_{\leq k}}$ by iterating Proposition 3.9 n times. We claim that

$$\beta'_i = \sum_{s=k-n+1}^{k-r} \left\{ \sum_{j \in I_{\leq k-n}} \sum_{l \in I_s} D_{il} \, (W^{-1}R)_{lj} \, \beta_j + \sum_{j \in I_s} D_{ij} \, W_{jj}^{-1} a_j \right\}$$

for $0 \leq r \leq n-1$ and $i \in I_{k-r}$. To see this, we proceed again by induction. As $(\beta_i)_{i \in I_{\leq k-n}}$ is a $(k-n)$-estimate, the expression is valid for $r = n-1$ by Proposition 3.9. Now suppose the expression is valid for all $u < r \leq n-1$. Then we obtain

$$\beta'_i = \sum_{j \in I_{\leq k-n}} \sum_{l \in I_{k-u}} D_{il} \, (W^{-1}R)_{lj} \, \beta_j + \sum_{j \in I_{k-u}} D_{ij} \, W_{jj}^{-1} a_j$$

$$+ \sum_{s=k-n+1}^{k-u-1} \sum_{j \in I_s} \sum_{l \in I_{k-u}} \sum_{t=k-n+1}^{s} \sum_{q \in I_{\leq k-n}} \sum_{p \in I_t} D_{il} \, (W^{-1}R)_{lj} \, D_{jp} \, (W^{-1}R)_{pq} \, \beta_q$$

$$+ \sum_{s=k-n+1}^{k-u-1} \sum_{j \in I_s} \sum_{l \in I_{k-u}} \sum_{t=k-n+1}^{s} \sum_{q \in I_t} D_{il} \, (W^{-1}R)_{lj} \, D_{jq} \, W_{qq}^{-1} a_q$$

for $i \in I_{k-u}$ by Proposition 3.9. Rearranging the sums yields

$$
\beta_i' = \sum_{j \in I_{\leq k-n}} \sum_{l \in I_{k-u}} D_{il} (W^{-1}R)_{lj} \beta_j + \sum_{j \in I_{k-u}} D_{ij} W_{jj}^{-1} a_j
$$
$$
+ \sum_{t=k-n+1}^{k-u-1} \left\{ \sum_{q \in I_{\leq k-n}} \sum_{p \in I_t} \bar{D}_{ip} (W^{-1}R)_{pq} \beta_q + \sum_{p \in I_t} \bar{D}_{ip} W_{pp}^{-1} a_p \right\},
$$

for $i \in I_{k-u}$, where we have defined

$$
\bar{D}_{ij} := \sum_{\ell=s}^{t-1} \sum_{q \in I_\ell} \sum_{l \in I_t} D_{il} (W^{-1}R)_{lq} D_{qj}
$$

whenever $i \in I_t$ and $j \in I_s$ for $s < t$. But as $D_{qj} = 0$ when $\tau(q) < \tau(j)$, we have

$$
\bar{D}_{ij} = \sum_{q \in I_{\leq t-1}} \sum_{l \in I_t} D_{il} (W^{-1}R)_{lq} D_{qj} = D_{ij} \qquad \text{for } i \in I_t, \ j \in I_{\leq t-1}
$$

using the identity used in the proof for the case $k_- > -\infty$, and the claim follows.

We can now complete the proof for the case $k_- = -\infty$. It suffices to prove the theorem for a given local function f (the extension to quasilocal f follows readily as in the proof of Lemma 3.3). Let us therefore fix a K-local function f for some $K \in \mathbb{J}$, and let $k = \max_{i \in K} \tau(i)$ and $n \geq 1$. By Lemma 3.3, we find that $(\beta_i)_{i \in I_{\leq k-n}}$ is trivially a $(k-n)$-estimate if we set $\beta_i = (\rho \otimes \tilde{\rho})\eta_i$ for $i \in I_{\leq k-n}$. Therefore

$$
|\rho f - \tilde{\rho} f| \leq \sum_{i,j \in I} \delta_i f\, D_{ij}\, W_{jj}^{-1} a_j + \sum_{i \in I} \sum_{j \in I_{\leq k-n}} \delta_i f\, D_{ij}\, (\rho \otimes \tilde{\rho})\eta_j
$$

from the k-estimate $(\beta_i')_{i \in I_{\leq k}}$ derived above, where we have used that $DW^{-1}R \leq D$. But as f is local and $\delta_i f < \infty$ for all i by assumption, the second term vanishes as $n \to \infty$ by assumption (2). This completes the proof for the case $k_- = -\infty$. □

4 Application: Particle Filters

Our original motivation for developing generalized comparison theorems was the investigation of algorithms for filtering in high dimension. In this section we will develop one such application in detail. Our result answers a question raised in [18], and also serves as a concrete illustration of the utility of our main results.

4.1 Introduction and Main Result

Let $(X_n, Y_n)_{n \geq 0}$ be a Markov chain. We interpret X_n as the unobserved component of the model and Y_n as the observed component. A fundamental problem in this setting is to track the state of the unobserved component X_n given the history of observed data $Y_1, \ldots, Y_n$. Such problems are ubiquitous in a wide variety of applications, ranging from classical tracking problems in navigation and robotics to large-scale forecasting problems such as weather prediction, and are broadly referred to as *filtering* or *data assimilation* problems.

In principle, the optimal solution to the tracking problem is provided by the *filter*

$$
\pi_n := \mathbf{P}[X_n \in \cdot\, | Y_1, \ldots, Y_n].
$$

If the conditional distribution π_n can be computed, in yields not only a least mean square estimate of the unobserved state X_n but also a complete representation of the uncertainty in this estimate. Unfortunately, when X_n takes values in a continuous (or finite but large) state space, the filter is rarely explicitly computable and approximations become necessary. In practice, filtering is widely implemented by a class of sequential Monte Carlo algorithms called *particle filters*, cf. [1], that have been very successful in classical scientific and engineering applications.

The major problem with particle filtering algorithms is that they typically require an exponential number of samples in the model dimension to achieve a fixed approximation error. Such algorithms are therefore largely useless in high-dimensional problems that arise in complex applications such as weather forecasting. We refer to [18] for a detailed discussion of these issues and for further references. In many applications, the high-dimensional nature of the problem is due to the presence of spatial degrees of freedom: X_n and Y_n at each time n are themselves random fields that evolve dynamically over time. In practice, such models are typically expected to exhibit decay of correlations. We have started in [18] to explore the possibility that such properties could be exploited to beat the curse of dimensionality by including a form of spatial localization in the filtering algorithm. In particular, the initial analysis in [18] has yielded dimension-free error bounds for the simplest possible class of local particle filtering algorithms, called *block particle filters*, under strong (but dimension-free) model assumptions that ensure the presence of decay of correlations.

It should be noted that the block particle filtering algorithm exhibits some significant drawbacks that could potentially be resolved by using more sophisticated algorithms. These issues are discussed at length in [18], but are beyond the scope of this paper. In the sequel, we will reconsider the same algorithm that was introduced in [18], but we provide an improved analysis of its performance on the basis of the generalized comparison Theorem 2.3. We will see that the use of Theorem 2.3 already yields a *qualitative* improvement over the main result of [18].

In the remainder of this section we recall the setting of [18] and state our main result on block particle filters. The following sections are devoted to the proofs.

Dynamical Model

Let $(X_n, Y_n)_{n \geq 0}$ be a Markov chain that takes values in the product space $\mathbb{X} \times \mathbb{Y}$, and whose transition probability P can be factored as

$$P((x, y), A) = \int \mathbf{1}_A(x', y') \, p(x, x') \, g(x', y') \, \psi(dx') \, \varphi(dy').$$

Such processes are called *hidden Markov models* [1]. The definition implies that the unobserved process $(X_n)_{n \geq 0}$ is a Markov chain in its own right with transition density p (with respect to the reference measure ψ), while each observation Y_n is a noisy function of X_n with observation density g (with respect to the reference measure φ).

Our interest is in high-dimensional hidden Markov models that possess spatial structure. To this end, we introduce a finite undirected graph $G = (V, E)$ that determines the spatial degrees of freedom. The state (X_n, Y_n) at each time n is itself a random field $(X_n^v, Y_n^v)_{v \in V}$ indexed by the vertices of G. In particular, we choose

$$\mathbb{X} = \prod_{v \in V} \mathbb{X}^v, \quad \mathbb{Y} = \prod_{v \in V} \mathbb{Y}^v, \quad \psi = \bigotimes_{v \in V} \psi^v, \quad \varphi = \bigotimes_{v \in V} \varphi^v,$$
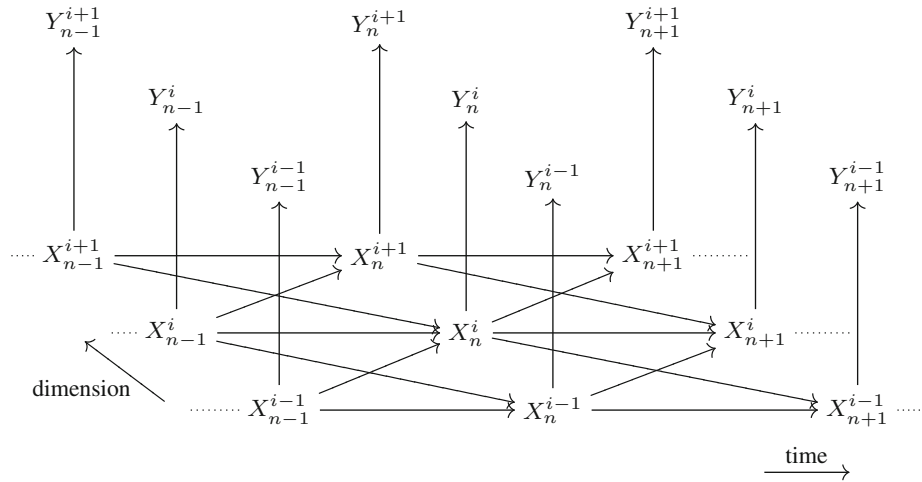
Fig. 1 Dependency graph of a high-dimensional filtering model of the type considered in Sect. 4

where $(\mathbb{X}^v, \psi^v)$ and $(\mathbb{Y}^v, \varphi^v)$ are measure spaces for every $v \in V$. To define the dynamics of the model, we introduce for each $v \in V$ a local transition density $p^v : \mathbb{X} \times \mathbb{X}^v \to \mathbb{R}_+$ and local observation density $g^v : \mathbb{X}^v \times \mathbb{Y}^v \to \mathbb{R}_+$, and we set

$$p(x, z) = \prod_{v \in V} p^v(x, z^v), \quad g(x, y) = \prod_{v \in V} g^v(x^v, y^v).$$

Therefore, each observation Y_n^v at location v is a noisy function of the unobserved state X_n^v at location v, and the current state X_n^v is determined by the configuration X_{n-1} at the previous time step. We will assume throughout that the dynamics is *local* in the sense that $p^v(x, z^v) = p^v(\tilde{X}, z^v)$ whenever $x^{N(v)} = \tilde{X}^{N(v)}$, where $N(v) = \{v' \in V : d(v, v') \le r\}$ denotes a neighborhood of radius r around the vertex v with respect to the graph distance d. That is, the state X_n^v at time n and vertex v depends only on the past $X_0, \ldots, X_{n-1}$ through the states $X_{n-1}^{N(v)}$ in an r-neighborhood of v in the previous time step; the interaction radius r is fixed throughout. The dependence structure of our general model is illustrated schematically in Fig. 1 (in the simplest case of a linear graph G with $r = 1$).

Block Particle Filters

An important property of the filter $\pi_n = \mathbf{P}[X_n \in \cdot | Y_1, \ldots, Y_n]$ is that it can be computed recursively. To this end, define for every probability measure ρ on $\mathbb{X}$ the probability measures $\mathsf{P}\rho$ and $\mathsf{C}_n\rho$ as follows:

$$\mathsf{P}\rho(dx') := \psi(dx') \int p(x, x') \, \rho(dx), \quad \mathsf{C}_n\rho(dx) := \frac{g(x, Y_n) \, \rho(dx)}{\int g(z, Y_n) \, \rho(dz)}.$$

Then it is an elementary consequence of Bayes' formula that [1]

$$\pi_n = \mathsf{F}_n\pi_{n-1} := \mathsf{C}_n\mathsf{P}\pi_{n-1} \quad \text{for every } n \ge 1,$$

where the initial condition is $\pi_0 = \mu := \mathbf{P}[X_0 \in \cdot]$. In the sequel we will often write $\pi_n^\mu := \mathsf{F}_n \cdots \mathsf{F}_1\mu$ to make explicit the initial condition of the filtering recursion.

To obtain approximate algorithms, we insert additional steps in the filtering recursion that enable a tractable implementation. The classical particle filtering algorithm inserts a random sampling step S^N in the filtering recursion that replaces the current filtering distribution by the empirical measure of N independent samples: that is,

$$\mathsf{S}^N \rho := \frac{1}{N} \sum_{i=1}^{N} \delta_{x(i)} \quad \text{where} \quad (x(i))_{i=1,\ldots,N} \text{ are i.i.d. samples} \sim \rho.$$

This yields a sequential Monte Carlo algorithm that maintains N approximate samples from the current filtering distribution at any time; cf. [1,18] and the references therein. As $N \to \infty$, the sampling error vanishes by the law of large numbers and the particle filter converges to the exact filter. Unfortunately, the number of samples N needed to achieve a fixed error is typically exponential in the dimension card V.

To alleviate the curse of dimensionality we must localize the algorithm so that it can benefit from the decay of correlations of the underlying model. The simplest possible algorithm of this type inserts an additional localization step in the filtering recursion in the following manner. Fix a partition $\mathcal{K}$ of the vertex set V into nonoverlapping blocks, and define for any probability ρ on $\mathbb{X}$ the blocking operator $\mathsf{B}\rho$ as

$$\mathsf{B}\rho := \bigotimes_{K \in \mathcal{K}} \mathsf{B}^K \rho,$$

where we denote by $\mathsf{B}^J \rho$ for $J \subseteq V$ the marginal of ρ on $\mathbb{X}^J$. That is, the blocking operation forces the underlying measure to be independent across different blocks in the partition $\mathcal{K}$. The *block particle filter* $\hat{\pi}_n^\mu$ is now defined by the recursion

$$\hat{\pi}_n^\mu := \hat{\mathsf{F}}_n \cdots \hat{\mathsf{F}}_1 \mu, \quad \hat{\mathsf{F}}_k := \mathsf{C}_k \mathsf{B} \mathsf{S}^N \mathsf{P}.$$

This algorithm is very straightforward to implement in practice, cf. [18].

Analysis of [18]

We first introduce some notation. Recall that in our model, each vertex v interacts only with vertices in an r-neighborhood $N(v)$ in the previous time step, where the interaction radius r is fixed throughout. Given $J \subseteq V$, define the r-inner boundary as

$$\partial J := \{v \in J : N(v) \nsubseteq J\}.$$

Thus ∂J is the subset of vertices in J that can interact with vertices outside J in one time step of the dynamics. We also define the quantities

$$|\mathcal{K}|_\infty := \max_{K \in \mathcal{K}} \operatorname{card} K,$$
$$\Delta := \max_{v \in V} \operatorname{card}\{v' \in V : d(v, v') \leq r\},$$
$$\Delta_\mathcal{K} := \max_{K \in \mathcal{K}} \operatorname{card}\{K' \in \mathcal{K} : d(K, K') \leq r\},$$

where $d(J, J') := \min_{v \in J} \min_{v' \in J'} d(v, v')$ for $J, J' \subseteq V$. Thus $|\mathcal{K}|_\infty$ is the maximal size of a block, while Δ ($\Delta_\mathcal{K}$) is the maximal number of vertices (blocks) that interact with a single vertex (block) in one time step. Note that $r, \Delta, \Delta_\mathcal{K}$ are *local* quantities that depend

on the geometry but not on the size of the graph G. We finally define, for each $J \subseteq I$, the following local distance between random measures ρ, ρ':

$$|||\rho - \rho'|||_J := \sup_{f \in \mathcal{X}^J : |f| \leq 1} \mathbf{E}[|\rho f - \rho' f|^2]^{1/2}$$

where $\mathcal{X}^J$ is the set of J-local measurable functions f on $\mathbb{X}$. We will write $\pi_n^x := \pi_n^{\delta_x}$ and $\hat{\pi}_n^x := \hat{\pi}_n^{\delta_x}$ when the filtering recursions are initialized at a point $x \in \mathbb{X}$.

We can now recall the main result of [18].

Theorem 4.1 (Theorem 2.1 in [18]) *There exists a constant $0 < \varepsilon_0 < 1$, depending only on the local quantities Δ and $\Delta_{\mathcal{K}}$, such that the following holds.*

Suppose there exist $\varepsilon_0 < \varepsilon < 1$ and $0 < \kappa < 1$ such that

$$\varepsilon \leq p^v(x, z^v) \leq \varepsilon^{-1}, \quad \kappa \leq g^v(x^v, y^v) \leq \kappa^{-1} \quad \forall v \in V, \ x, z \in \mathbb{X}, \ y \in \mathbb{Y}.$$

Then for every $n \geq 0$, $\sigma \in \mathbb{X}$, $K \in \mathcal{K}$ and $J \subseteq K$ we have

$$|||\pi_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \alpha \, \mathrm{card} \, J \left[e^{-\beta_1 d(J, \partial K)} + \frac{e^{\beta_2 |\mathcal{K}|_\infty}}{N^{\frac{1}{2}}} \right],$$

where the constants $0 < \alpha, \beta_1, \beta_2 < \infty$ depend only on $\varepsilon, \kappa, r, \Delta,$ and $\Delta_{\mathcal{K}}$.

The error bound in Theorem 4.1 contains two terms. The first term quantifies the bias introduced by the localization B, which decreases when we take larger blocks. The second term quantifies the variance introduced by the sampling S^N, which decreases with increasing sample size N but grows exponentially in the block size. Traditional particle filtering algorithms correspond to the choice of a single block $\mathcal{K} = \{V\}$, and in this case the error grows exponentially in the dimension card V. To avoid this curse of dimensionality, we must tune the block size so as to optimize the tradeoff between bias and variance. As all the constants in Theorem 4.1 depend only on local quantities, the optimal block size results in a dimension-free error bound. We refer to [18] for a full discussion.

Main Result

The intuition behind the block particle filtering algorithm is that the localization controls the sampling error (as it replaces the model dimension card V by the block size $|\mathcal{K}|_\infty$), while the decay of correlations property of the model controls the localization error (as it ensures that the effect of the localization decreases in the distance to the block boundary). This intuition is clearly visible in the conclusion of Theorem 4.1. It is however not automatically the case that our model does indeed exhibit decay of correlations: when there are strong interactions between the vertices, phase transitions can arise and the decay of correlations can fail much as for standard models in statistical mechanics [16], in which case we cannot expect to obtain dimension-free performance for the block particle filter. Such phenomena are ruled out in Theorem 4.1 by the assumption that $\varepsilon \leq p^v \leq \varepsilon^{-1}$ for $\varepsilon > \varepsilon_0$, which ensures that the interactions in our model are sufficiently weak.

It is notoriously challenging to obtain sharp quantitative results for interacting models, and it is unlikely that one could obtain realistic values for the constants in Theorem 4.1 at the level of generality considered here. More concerning, however, is that the weak interaction assumption of Theorem 4.1 is already unsatisfactory at the *qualitative* level. Note that there is no interaction between the vertices in the extreme case $\varepsilon = 1$; the assumption $\varepsilon > \varepsilon_0$ should be viewed as a perturbation of this situation (i.e., weak interactions). However, setting $\varepsilon = 1$

turns off not only the interaction between different vertices, but also the interaction between the same vertex at different times: in this setting the dynamics of the model become trivial. In contrast, one would expect that it is only the strength of the spatial interactions, and not the local dynamics, that is relevant for dimension-free errors, so that Theorem 4.1 places an unnatural restriction on our understanding of block particle filters.

Our main result resolves this qualitative deficiency of Theorem 4.1. Rather than assuming $p^v(x, z^v) \approx 1$ as in Theorem 4.1, we will assume only that the spatial interactions are weak in the sense that $p^v(x, z^v) \approx q^v(x^v, z^v)$, where the transition density q^v describes the local dynamics at the vertex v in the absence of interactions.

Theorem 4.2 *For any $0 < \delta < 1$ there exists $0 < \varepsilon_0 < 1$, depending only on δ and Δ, so that the following holds. Suppose there exist $\varepsilon_0 < \varepsilon < 1$ and $0 < \kappa < 1$ so that*

$$\varepsilon q^v(x^v, z^v) \leq p^v(x, z^v) \leq \varepsilon^{-1} q^v(x^v, z^v),$$

$$\delta \leq q^v(x^v, z^v) \leq \delta^{-1},$$

$$\kappa \leq g^v(x^v, y^v) \leq \kappa^{-1}$$

for every $v \in V$, $x, z \in \mathbb{X}$, $y \in \mathbb{Y}$, where $q^v : \mathbb{X}^v \times \mathbb{X}^v \to \mathbb{R}_+$ is a transition density with respect to ψ^v. Then for every $n \geq 0$, $\sigma \in \mathbb{X}$, $K \in \mathcal{K}$ and $J \subseteq K$ we have

$$|||\pi_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \alpha \, \text{card} \, J \left[e^{-\beta_1 d(J, \partial K)} + \frac{e^{\beta_2 |\mathcal{K}|_\infty}}{N^\gamma} \right],$$

where $0 < \gamma \leq \frac{1}{2}$ and $0 < \alpha, \beta_1, \beta_2 < \infty$ depend only on δ, ε, κ, r, Δ, and $\Delta_\mathcal{K}$.

In Theorem 4.2, the parameter ε controls the spatial correlations while the parameter δ controls the temporal correlations (in contrast to Theorem 4.1, where both are controlled simultaneously by the single parameter ε). The key point is that δ can be arbitrary, and only ε must lie above the threshold ε_0. That the threshold ε_0 depends on δ is natural: the more ergodic the dynamics, the more spatial interactions can be tolerated without losing decay of correlations.

The proof of Theorem 4.1 was based on repeated application of the classical Dobrushin comparison theorem (Corollary 2.4). While there are some significant differences between the details of the proofs, the essential improvement that makes it possible to prove Theorem 4.2 is that the generalized comparison theorem (Theorem 2.3) enables us to treat spatial and temporal degrees of freedom on a different footing.

Organization of the Proof

As in [18], we consider three recursions

$$\pi_n^\mu := \mathsf{F}_n \cdots \mathsf{F}_1 \mu, \quad \tilde{\pi}_n^\mu := \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \mu, \quad \hat{\pi}_n^\mu := \hat{\mathsf{F}}_n \cdots \hat{\mathsf{F}}_1 \mu,$$

where $\mathsf{F}_n := \mathsf{C}_n \mathsf{P}$, $\tilde{\mathsf{F}}_n := \mathsf{C}_n \mathsf{BP}$, and $\hat{\mathsf{F}}_n := \mathsf{C}_n \mathsf{BS}^N \mathsf{P}$. The filter π_n^μ and the block particle filter $\hat{\pi}_n^\mu$ were already defined above. The block filter $\tilde{\pi}_n^\mu$ is intermediate: it inserts only the localization but not the sampling step in the filtering recursion. This allows to decompose the approximation error into two terms

$$|||\pi_n^\mu - \hat{\pi}_n^\mu|||_J \leq |||\pi_n^\mu - \tilde{\pi}_n^\mu|||_J + |||\tilde{\pi}_n^\mu - \hat{\pi}_n^\mu|||_J$$

by the triangle inequality. In the proof of Theorem 4.2, each of these terms will be considered separately. The first term, which quantifies the bias due to the localization, will be bounded

in Sect. 4.2. The second term, which quantifies the sampling variance, will be bounded in Sect. 4.3. Combining these bounds completes the proof.

4.2 Bounding the Bias

The goal of this section is to bound the bias term $\|\pi_n^\sigma - \tilde{\pi}_n^\sigma\|_J$, where we denote by

$$\|\mu - \nu\|_J := \sup_{f \in \mathcal{X}^J : |f| \le 1} |\mu f - \nu f|$$

the local total variation distance on the set of sites J. [Note that $\|\mu - \nu\|_J \le K$ for some $K \in \mathbb{R}$ evidently implies $|||\mu - \nu|||_J \le K$; the random measure norm $||| \cdot |||_J$ will be essential to bound the sampling error, but is irrelevant for the bias term.]

Let us first give an informal outline of the ideas behind the proof of the bias bound. While π_n^σ is itself a high-dimensional distribution (defined on the set of sites V), we do not know how to obtain a tractable local update rule for it. Thus we cannot apply Theorem 2.3 directly. Instead, we will consider the *smoothing* distribution

$$\rho = \mathbf{P}^\sigma[X_1, \dots, X_n \in \cdot |Y_1, \dots, Y_n],$$

defined on the extended set of sites $I = \{1, \dots, n\} \times V$ and configuration space $\mathbb{S} = \mathcal{X}^n$. As $(X_k^v, Y_k^v)_{(k,v) \in I}$ is a Markov random field (as is illustrated in Fig. 1), we can read off a local update rule for ρ from the model definition. At the same time, as $\pi_n^\sigma = \mathbf{P}^\sigma[X_n \in \cdot |Y_1, \dots, Y_n]$ is a marginal of ρ, we immediately obtain estimates for π_n^σ from estimates for ρ.

This basic idea relies on the probabilistic definition of the filter as a conditional distribution of a Markov random field: the filtering recursion (which was only introduced for computational purposes) plays no role in the analysis. The block filter $\tilde{\pi}_n^\sigma$, on the other hand, is *defined* in terms of a recursion and does not have an intrinsic probabilistic interpretation. In order to handle the block filter, we will artificially cook up a probability measure $\tilde{\mathbf{P}}$ on $\mathbb{S}$ such that the block filter satisfies $\tilde{\pi}_n^\sigma = \tilde{\mathbf{P}}[X_n \in \cdot |Y_1, \dots, Y_n]$, and choose the corresponding smoothing distribution

$$\tilde{\rho} = \tilde{\mathbf{P}}[X_1, \dots, X_n \in \cdot |Y_1, \dots, Y_n].$$

This implies in particular that

$$\|\pi_n^\sigma - \tilde{\pi}_n^\sigma\|_J = \|\rho - \tilde{\rho}\|_{\{n\} \times J},$$

and we can now bound the bias term by applying Theorem 2.3.

To apply the comparison theorem we must choose a good cover $\mathcal{J}$. It is here that the full flexibility of Theorem 2.3, as opposed to the classical comparison theorem, comes into play. If we were to apply Theorem 2.3 with the singleton cover $\mathcal{J}_s = \{\{i\} : i \in I\}$, we would recover the result of Theorem 4.1: in this case both the spatial and temporal interactions must be weak in order to ensure that $D = \sum_n (W^{-1} R)^n < \infty$. To avoid this problem, we work instead with larger blocks in the temporal direction. That is, our blocks $J \in \mathcal{J}$ will have the form $J = \{k + 1, \dots, k + q\} \times \{v\}$ for an appropriate choice of the block length q. The local update γ_x^J now behaves as q time steps of an ergodic Markov chain in $\mathcal{X}^v$: the temporal interactions decay geometrically with q, and can therefore be made arbitrarily small even if the interaction in one time step is arbitrarily strong. On the other hand, when we increase q there will be more nonzero terms in the matrix $W^{-1} R$. We must therefore ultimately tune the block length q appropriately to obtain the result of Theorem 4.2.

Remark 4.1 The approach used here to bound the bias directly using the comparison theorem is different than the one used in [18], which exploits the recursive property of the filter. The latter approach has a broader scope, as it does not rely on the ability to express the approximate filter as the marginal of a random field: this could be essential for the analysis of more sophisticated algorithms that do not admit such a representation. For the purpose of this paper, however, the present approach yields a shorter proof that is well adapted to the analysis of block particle filters.

Remark 4.2 The problem under investigation is based on an interacting Markov chain model, and is therefore certainly dynamical in nature. Nonetheless, our proofs use Theorem 2.3 and not the one-sided Theorem 2.9. If we were to approximate the dynamics of the Markov chain X_n itself, it would be much more convenient to apply Theorem 2.9 as the model is already defined in terms of one-sided conditional distributions $p(x, z) \psi(dz)$. Unfortunately, when we condition on the observations Y_n, the one-sided conditional distributions take a complicated form that incorporates all the information in the future observations, whereas conditioning on all variables outside a block $J \in \mathcal{J}$ gives rise to relatively tractable local expressions. For this reason, the static "space-time" picture remains the most convenient approach for the investigation of high-dimensional filtering problems.

We now turn to the proof details. We first state the main result of this section.

Theorem 4.3 (Bias term) *Suppose there exist $0 < \varepsilon, \delta < 1$ such that*

$$\varepsilon q^v(x^v, z^v) \le p^v(x, z^v) \le \varepsilon^{-1} q^v(x^v, z^v),$$
$$\delta \le q^v(x^v, z^v) \le \delta^{-1}$$

for every $v \in V$ and $x, z \in \mathbb{X}$, where $q^v : \mathbb{X}^v \times \mathbb{X}^v \to \mathbb{R}_+$ is a transition density with respect to ψ^v. Suppose also that we can choose $q \in \mathbb{N}$ and $\beta > 0$ such that

$$c := 3q\Delta^2 e^{\beta(q+2r)}(1 - \varepsilon^{2(\Delta+1)}) + e^\beta(1 - \varepsilon^2\delta^2) + e^{\beta q}(1 - \varepsilon^2\delta^2)^q < 1.$$

Then we have

$$\|\pi_n^\sigma - \tilde{\pi}_n^\sigma\|_J \le \frac{2e^{\beta r}}{1 - c}(1 - \varepsilon^{2(q+1)\Delta}) \operatorname{card} J \, e^{-\beta d(J, \partial K)}$$

for every $n \ge 0$, $\sigma \in \mathbb{X}$, $K \in \mathcal{K}$ and $J \subseteq K$.

In order to use the comparison theorem, we must have a method to construct couplings. Before we proceed to the proof of Theorem 4.3, we begin by formulating two elementary results that will provide us with the necessary tools for this purpose.

Lemma 4.4 *If probabilities μ, ν, γ satisfy $\mu(A) \ge \alpha\gamma(A)$ and $\nu(A) \ge \alpha\gamma(A)$ for every measurable set A, there is a coupling Q of μ, ν with $\int \mathbf{1}_{x \ne z} Q(dx, dz) \le 1 - \alpha$.*

Proof Define $\tilde{\mu} = (\mu - \alpha\gamma)/(1 - \alpha)$, $\tilde{\nu} = (\nu - \alpha\gamma)/(1 - \alpha)$, and let

$$Qf = \alpha \int f(x, x) \gamma(dx) + (1 - \alpha) \int f(x, z) \tilde{\mu}(dx) \tilde{\nu}(dz).$$

The claim follows readily. □

Lemma 4.5 *Let $P_1, \ldots, P_q$ be transition kernels on a measurable space $\mathbb{T}$, and let*

$$\mu_x(d\omega_1, \ldots, d\omega_q) = P_1(x, d\omega_1) P_2(\omega_1, d\omega_2) \cdots P_q(\omega_{q-1}, d\omega_q).$$

Suppose there exist probability measures $\nu_1, \ldots, \nu_q$ on $\mathbb{T}$ so that $P_i(x, A) \geq \alpha \nu_i(A)$ for every measurable set A, $x \in \mathbb{T}$, and $i \leq q$. Then there exists for every $x, z \in \mathbb{T}$ a coupling $Q_{x,z}$ of μ_x, μ_z so that $\int \mathbf{1}_{\omega_i \neq \omega_i'} Q_{x,z}(d\omega, d\omega') \leq (1 - \alpha)^i$ for all $i \leq q$.

Proof Define the transition kernels $\tilde{P}_i = (P_i - \alpha \nu_i)/(1 - \alpha)$ and

$$\tilde{Q}_i f(x, z) = \alpha \int f(x', x') \, \nu_i(dx') + (1 - \alpha) \mathbf{1}_{x \neq z} \int f(x', z') \, \tilde{P}_i(x, dx') \, \tilde{P}_i(z, dz')$$

$$+ (1 - \alpha) \mathbf{1}_{x=z} \int f(x', x') \, \tilde{P}_i(x, dx').$$

Then $\tilde{Q}_i(x, z, \cdot)$ is a coupling of $P_i(x, \cdot)$ and $P_i(z, \cdot)$. Now define

$$Q_{x,z}(d\omega_1, d\omega_1', \ldots, d\omega_q, d\omega_q') = \tilde{Q}_1(x, z, d\omega_1, d\omega_1') \cdots \tilde{Q}_q(\omega_{q-1}, \omega_{q-1}', d\omega_q, d\omega_q').$$

The result follows once we note that $\int \mathbf{1}_{x' \neq z'} \tilde{Q}_i(x, z, dx', dz') \leq (1 - \alpha) \mathbf{1}_{x \neq z}$. $\qquad\square$

We can now proceed to the proof of Theorem 4.3.

Proof (Theorem 4.3) We begin by constructing a measure $\tilde{\mathbf{P}}$ that allows to describe the block filter $\tilde{\pi}_n^\sigma$ as a conditional distribution. We fix the initial condition $\sigma \in \mathbb{X}$ throughout the proof (the dependence of various quantities on σ is implicit).

To construct $\tilde{\mathbf{P}}$, define for $K \in \mathcal{K}$ and $n \geq 1$ the function

$$h_n^K(x, z^{\partial K}) := \int \tilde{\pi}_{n-1}^\sigma(d\omega) \prod_{v \in \partial K} p^v(x^K \omega^{V \setminus K}, z^v).$$

Evidently h_n^K is a transition density with respect to $\bigotimes_{v \in \partial K} \psi^v$. Let

$$\tilde{P}_n(x, z) := \prod_{K \in \mathcal{K}} h_n^K(x, z^{\partial K}) \prod_{v \in K \setminus \partial K} p^v(x, z^v),$$

and define $\tilde{\mathbf{P}}_n \mu(dx') := \psi(dx') \int \tilde{P}_n(x, x') \, \mu(dx)$. Then $\tilde{\mathbf{P}}_n \tilde{\pi}_{n-1}^\sigma = \mathsf{BP} \tilde{\pi}_{n-1}^\sigma$ by construction for every $n \geq 1$, as $\tilde{\pi}_{n-1}^\sigma$ is a product measure across blocks. Thus

$$\pi_n^\sigma = \mathsf{C}_n \mathsf{P} \cdots \mathsf{C}_1 \mathsf{P} \delta_\sigma, \quad \tilde{\pi}_n^\sigma = \mathsf{C}_n \tilde{\mathsf{P}}_n \cdots \mathsf{C}_1 \tilde{\mathsf{P}}_1 \delta_\sigma,$$

that is, the filter and the block filter satisfy the same recursion with different transition densities p and $\tilde{P}_n$. Thus we can interpret the block filter as the filter corresponding to a time-inhomogeneous Markov chain with transition densities $\tilde{P}_n$: that is, if we set

$$\tilde{\mathbf{P}}[(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \in A]$$

$$:= \int \mathbf{1}_A(x_1, \ldots, x_n, y_1, \ldots, y_n) \, \tilde{P}_1(\sigma, x_1) \prod_{k=2}^n \tilde{P}_k(x_{k-1}, x_k) \, g(x_k, y_k) \, \psi(dx_k) \, \varphi(dy_k)$$

(note that $\mathbf{P}^\sigma$ satisfies the same formula where $\tilde{P}_k$ is replaced by p), we can write

$$\tilde{\pi}_n^\sigma = \tilde{\mathbf{P}}[X_n \in \cdot \, | Y_1, \ldots, Y_n].$$

Let us emphasize that the transition densities $\tilde{P}_k$ and operators $\tilde{\mathsf{P}}_k$ themselves depend on the initial condition σ, which is certainly not the case for the regular filter. However, since σ is fixed throughout the proof, this is irrelevant for our computations.

From now on we fix $n \geq 1$ in the remainder of the proof. Let

$$\rho = \mathbf{P}^\sigma[X_1, \ldots, X_n \in \cdot \,|Y_1, \ldots, Y_n], \quad \tilde{\rho} = \tilde{\mathbf{P}}[X_1, \ldots, X_n \in \cdot \,|Y_1, \ldots, Y_n].$$

Then ρ and $\tilde{\rho}$ are probability measures on $\mathbb{S} = \mathbb{X}^n$, which is indexed by the set of sites $I = \{1, \ldots, n\} \times V$ (the observation sequence on which we condition is arbitrary and is fixed throughout the proof). The proof now proceeds by applying Theorem 2.3 to $\rho, \tilde{\rho}$, the main difficulty being the construction of a coupled update rule.

Fix $q \geq 1$. We first specify the cover $\mathcal{J} = \{J_l^v : 1 \leq l \leq \lceil n/q \rceil, v \in V\}$ by

$$J_l^v := \{(l-1)q + 1, \ldots, lq \wedge n\} \times \{v\} \quad \text{for } 1 \leq l \leq \lceil n/q \rceil, \ v \in V.$$

We choose the local updates $\gamma_x^J(dz^J) = \rho(dz^J|x^{I \setminus J})$ and $\tilde{\gamma}_x^J(dz^J) = \tilde{\rho}(dz^J|x^{I \setminus J})$, and postpone the construction of the coupled updates $Q_{x,z}^J$ and $\hat{Q}_x^J$ to be done below. Now note that the cover $\mathcal{J}$ is in fact a partition of I; thus Theorem 2.3 yields

$$\|\pi_n^\sigma - \tilde{\pi}_n^\sigma\|_J = \|\rho - \tilde{\rho}\|_{\{n\} \times J} \leq 2 \sum_{i \in \{n\} \times J} \sum_{j \in I} D_{ij} b_j$$

provided that $D = \sum_{k=0}^\infty C^k < \infty$ (cf. Corollary 2.6), where

$$C_{ij} = \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \int \mathbf{1}_{\omega_i \neq \omega_i'} \, Q_{x,z}^{J(i)}(d\omega, d\omega'),$$

$$b_i = \sup_{x \in \mathbb{S}} \int \mathbf{1}_{\omega_i \neq \omega_i'} \, \hat{Q}_x^{J(i)}(d\omega, d\omega'),$$

and where we write $J(i)$ for the unique block $J \in \mathcal{J}$ that contains $i \in I$. To proceed, we must introduce coupled updates $Q_{x,z}^J$ and $\hat{Q}_x^J$ and estimate C_{ij} and b_j.

Fix until further notice a block $J = J_l^v \in \mathcal{J}$. We will consider first the case that $1 < l < \lceil n/q \rceil$; the cases $l = 1, \lceil n/q \rceil$ will follow subsequently using the identical proof. Let $s = (l-1)q$. Then we can compute explicitly the local update rule

$$\gamma_x^J(A) = \frac{\int \mathbf{1}_A(x^J) \, p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}$$

by Bayes' formula, the definition of $\mathbf{P}^\sigma$ (in the same form as the definition of $\tilde{\mathbf{P}}$), and as $p^v(x, z^v)$ depends only on $x^{N(v)}$. We now construct couplings $Q_{x,z}^J$ of γ_x^J, γ_z^J where x, z differ only at the site $j = (k, w) \in I$. We distinguish the following cases:

1. $k = s, \ w \in N(v)\setminus\{v\}$;
2. $k = s, \ w = v$;
3. $k \in \{s+1, \ldots, s+q\}, \ w \in \bigcup_{u \in N(v)} N(u)\setminus\{v\}$;
4. $k = s + q + 1, \ w \in N(v)\setminus\{v\}$;
5. $k = s + q + 1, \ w = v$.

It is easily verified that γ_x^J does not depend on x_k^w except in one of the above cases. Thus when j satisfies none of the above conditions, we can set $C_{ij} = 0$ for $i \in J$.

Case 1. Note that

$$\gamma_x^J(A) \geq \varepsilon^2 \frac{\int \mathbf{1}_A(x^J) \, q^v(x_s^v, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int q^v(x_s^v, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}$$

and the right hand side does not depend on x_s^w for $w \neq v$. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$ for $j = (s, w)$ with $w \in N(v) \setminus \{v\}$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.4 such that $C_{ij} \leq 1 - \varepsilon^2$ for every $i \in J$.

Case 2. Define the transition kernels on $\mathbb{X}^v$

$$P_{k,x}(\omega, A) = \frac{\int \mathbf{1}_A(x_k^v) \, p^v(\omega x_{k-1}^{V \setminus \{v\}}, x_k^v) \prod_{m=k}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int p^v(\omega x_{k-1}^{V \setminus \{v\}}, x_k^v) \prod_{m=k}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}$$

for $k = s+1, \ldots, s+q$. As $P_{k,x}(x_{k-1}^v, dx_k^v) = \gamma_x^J(dx_k^v | x_{s+1}^v, \ldots, x_{k-1}^v)$ by construction, we are in the setting of Lemma 4.5. Moreover, we can estimate

$$P_{k,x}(\omega, A) \geq \varepsilon^2 \delta^2 \frac{\int \mathbf{1}_A(x_k^v) \prod_{m=k}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int \prod_{m=k}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} p^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)},$$

where the right hand side does not depend on ω. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$ for $j = (s, v)$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.5 such that $C_{ij} \leq (1 - \varepsilon^2 \delta^2)^{k-s}$ for $i = (k, v)$ with $k = s+1, \ldots, s+q$.

Case 3. Fix $k \in \{s+1, \ldots, s+q\}$ and $u \neq v$. Note that

$$\gamma_x^J(A) \geq \varepsilon^{2(\Delta+1)} \times \frac{\int \mathbf{1}_A(x^J) \, p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_m^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_m^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)},$$

where we define $\beta_m^w(x_m, x_{m+1}^w) = q^w(x_m^w, x_{m+1}^w)$ if either $m = k$ or $m = k-1$ and $w = u$, and we define $\beta_m^w(x_m, x_{m+1}^w) = p^w(x_m, x_{m+1}^w)$ otherwise. The right hand side of this expression does not depend on x_k^u as the terms $q^w(x_m^w, x_{m+1}^w)$ for $w \neq v$ cancel in the numerator and denominator. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$ for $j = (k, u)$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.4 such that $C_{ij} \leq 1 - \varepsilon^{2(\Delta+1)}$ for every $i \in J$.

Case 4. Let $u \in N(v) \setminus v$. Note that

$$\gamma_x^J(A) \geq \varepsilon^2 \frac{\int \mathbf{1}_A(x^J) \, p^v(x_s^v, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_m^w(x_m, x_{m+1}^w) \psi^v(dx_m^v)}{\int p^v(x_s^v, x_{s+1}^v) \prod_{m=s+1}^{s+q} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_m^w(x_m, x_{m+1}^w) \psi^v(dx_m^v)},$$

where we set $\beta_m^w(x_m, x_{m+1}^w) = q^w(x_m^w, x_{m+1}^w)$ if $m = s+q$ and $w = u$, and we set $\beta_m^w(x_m, x_{m+1}^w) = p^w(x_m, x_{m+1}^w)$ otherwise. The right hand side does not depend on x_{s+q+1}^u as the term $q^u(x_{s+q}^u, x_{s+q+1}^u)$ cancels in the numerator and denominator. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$ for $j = (s+q+1, u)$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.4 such that $C_{ij} \leq 1 - \varepsilon^2$ for every $i \in J$.

Case 5. Define for $k = s+1, \ldots, s+q$ the transition kernels on $\mathbb{X}^v$

$$P_{k,x}(\omega, A) = \frac{\int \mathbf{1}_A(x_k^v) \, p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^{k} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_{m,\omega}^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^{k} g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_{m,\omega}^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)},$$

where $\beta_{m,\omega}^w(x_m, x_{m+1}^w) = p^v(x_k, \omega)$ if $m = k$ and $w = v$, and $\beta_{m,\omega}^w(x_m, x_{m+1}^w) = p^w(x_m, x_{m+1}^w)$ otherwise. As $P_{k,x}(x_{k+1}^v, dx_k^v) = \gamma_x^J(dx_k^v | x_{k+1}^v, \ldots, x_{s+q}^v)$ by construction, we are in the setting of Lemma 4.5. Moreover, we can estimate

$$P_{k,x}(\omega, A) \geq \varepsilon^2 \delta^2 \times \frac{\int \mathbf{1}_A(x_k^v) \, p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^k g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_m^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)}{\int p^v(x_s, x_{s+1}^v) \prod_{m=s+1}^k g^v(x_m^v, Y_m^v) \prod_{w \in N(v)} \beta_m^w(x_m, x_{m+1}^w) \, \psi^v(dx_m^v)},$$

where we define $\beta_m^w(x_m, x_{m+1}^w) = 1$ if $m = k$ and $w = v$ and $\beta_m^w(x_m, x_{m+1}^w) = p^w(x_m, x_{m+1}^w)$ otherwise. Note that the right hand side does not depend on ω. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$ for $j = (s+q+1, v)$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.5 such that $C_{ij} \leq (1 - \varepsilon^2 \delta^2)^{s+q+1-k}$ for $i = (k, v)$ with $k = s+1, \ldots, s+q$.

We have now constructed coupled updates $Q_{x,z}^J$ for every pair $x, z \in \mathbb{S}$ that differ only at one point. Collecting the above bounds on C_{ij}, we can estimate

$$\sum_{(k',v') \in I} e^{\beta\{|k-k'| + d(v,v')\}} C_{(k,v)(k',v')}$$
$$\leq 2e^{\beta(q+r)}(1 - \varepsilon^2)\Delta + e^{\beta(q+2r)}(1 - \varepsilon^{2(\Delta+1)})\Delta^2 q$$
$$+ e^{\beta(k-s)}(1 - \varepsilon^2 \delta^2)^{k-s} + e^{\beta(s+q+1-k)}(1 - \varepsilon^2 \delta^2)^{s+q+1-k}$$
$$\leq 3q\Delta^2 e^{\beta(q+2r)}(1 - \varepsilon^{2(\Delta+1)}) + e^\beta(1 - \varepsilon^2 \delta^2) + e^{\beta q}(1 - \varepsilon^2 \delta^2)^q =: c,$$

when $(k, v) \in J$. In the last line, we have used that $\alpha^{x+1} + \alpha^{q-x}$ is a convex function of $x \in [0, q-1]$, and thus attains its maximum on the endpoints $x = 0, q - 1$.

Up to this point we have considered an arbitrary block $J = J_l^v \in \mathcal{J}$ with $1 < l < \lceil n/q \rceil$. It is however evident that the identical proof holds for the boundary blocks $l = 1, \lceil n/q \rceil$, except that for $l = 1$ we only need to consider Cases 3–5 above and for $l = \lceil n/q \rceil$ we only need to consider Cases 1–3 above. As all the estimates are otherwise identical, the corresponding bounds on C_{ij} are at most as large as those in the case $1 < l < \lceil n/q \rceil$. We therefore obtain

$$\|C\|_{\infty,\beta m} := \max_{i \in I} \sum_{j \in I} e^{\beta m(i,j)} C_{ij} \leq c,$$

where we define the metric $m(i, j) = |k - k'| + d(v, v')$ for $(k, v), (k', v') \in I$.

Next, we construct couplings $\hat{Q}_x^J$ of γ_x^J and $\tilde{\gamma}_x^J$ and to estimate the coefficients b_i. To this end, let us first note that $h_n^K(x, z^{\partial K})$ depends only on $x^{\partial^2 K}$, where

$$\partial^2 K := \bigcup_{w \in \partial K} N(w) \cap K$$

is the subset of vertices in K that can interact with vertices outside K in two time steps. It is easily seen that $\gamma_x^J = \tilde{\gamma}_x^J$, so we can set $b_i = 0$ for $i \in J$, unless $J = J_l^v$ with $v \in \partial^2 K$ for some $K \in \mathcal{K}$. In the latter case we obtain by Bayes' formula

$$\tilde{\gamma}_x^J(A) = \frac{\int \mathbf{1}_A(x^J) \prod_{t=s}^{s+q} g^v(x_t^v, Y_t^v) h_{t+1}^K(x_t, x_{t+1}^{\partial K}) \prod_{w \in N(v) \cap K \setminus \partial K} p^w(x_t, x_{t+1}^w) \, d\psi^J}{\int \prod_{t=s}^{s+q} g^v(x_t^v, Y_t^v) h_{t+1}^K(x_t, x_{t+1}^{\partial K}) \prod_{w \in N(v) \cap K \setminus \partial K} p^w(x_t, x_{t+1}^w) \, d\psi^J}$$

for $1 < l < \lceil n/q \rceil$, where $s = (l-1)q$ and $d\psi^J = \bigotimes_{(k,v) \in J} \psi^v(dx_k^v)$. Note that

$$\prod_{w \in N(v) \setminus (K \setminus \partial K)} p^w(x, z^w) \geq \varepsilon^\Delta \prod_{w \in N(v) \setminus (K \setminus \partial K)} q^w(x^w, z^w),$$

while

$$h_m^K(x, z^{\partial K}) \geq \varepsilon^\Delta \prod_{w \in N(v) \cap \partial K} q^w(x^w, z^w) \int \tilde{\pi}_{m-1}^\sigma(d\omega) \prod_{w \in \partial K \setminus N(v)} p^w(x^K \omega^{V \setminus K}, z^w).$$

Thus we can estimate $\gamma_x^J(A) \geq \varepsilon^{2(q+1)\Delta} \Gamma(A)$ and $\tilde{\gamma}_x^J(A) \geq \varepsilon^{2(q+1)\Delta} \Gamma(A)$ with

$$\Gamma(A) = \frac{\int \mathbf{1}_A(x^J) \prod_{m=s}^{s+q} g^v(x_m^v, Y_m^v) \, \beta(x_m^v, x_{m+1}^v) \prod_{w \in N(v) \cap K \backslash \partial K} p^w(x_m, x_{m+1}^w) \, d\psi^J}{\int \prod_{m=s}^{s+q} g^v(x_m^v, Y_m^v) \, \beta(x_m^v, x_{m+1}^v) \prod_{w \in N(v) \cap K \backslash \partial K} p^w(x_m, x_{m+1}^w) \, d\psi^J},$$

where we define $\beta(x, z) = q^v(x, z)$ if $v \in \partial K$ and $\beta(x, z) = 1$ if $v \in \partial^2 K \backslash \partial K$. We can therefore construct a coupling $\hat{Q}_x^J$ using Lemma 4.4 such that $b_i \leq 1 - \varepsilon^{2(q+1)\Delta}$ for all $i \in J$ in the case $1 < l < \lceil n/q \rceil$. The same conclusion follows for $l = 1, \lceil n/q \rceil$ by the identical proof.

We are now ready to put everything together. As $\| \cdot \|_{\infty, \beta m}$ is a matrix norm,

$$\|D\|_{\infty, \beta m} \leq \sum_{k=0}^{\infty} \|C\|_{\infty, \beta m}^k \leq \frac{1}{1 - c} < \infty.$$

Thus $D < \infty$, to we can apply the comparison theorem. Moreover,

$$\sup_{i \in J} \sum_{j \in J'} D_{ij} = \sup_{i \in J} e^{-\beta m(i, J')} \sum_{j \in J'} e^{\beta m(i, J')} D_{ij} \leq e^{-\beta m(J, J')} \|D\|_{\infty, \beta m}.$$

Thus we obtain

$$\|\pi_n^\sigma - \tilde{\pi}_n^\sigma\|_J \leq 2(1 - \varepsilon^{2(q+1)\Delta}) \sum_{i \in \{n\} \times J} \sum_{j \in \{1, \ldots, n\} \times \partial^2 K} D_{ij}$$

$$\leq \frac{2}{1 - c} (1 - \varepsilon^{2(q+1)\Delta}) \operatorname{card} J \, e^{-\beta d(J, \partial^2 K)}.$$

But clearly $d(J, \partial^2 K) \geq d(J, \partial K) - r$, and the proof is complete. $\square$

Remark 4.3 In the proof of Theorem 4.3 (and similarly for Theorem 4.8 below), we apply the comparison theorem with a nonoverlapping cover $\{(l-1)q + 1, \ldots, lq \wedge n\}, l \leq \lceil n/q \rceil$. Working with overlapping blocks $\{s + 1, \ldots, s + q\}, s \leq n - q$ would give somewhat better estimates at the expense of even more tedious computations.

4.3 Bounding the Variance

We now turn to bounding the variance term $|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J$. We will follow the basic approach taken in [18], where a detailed discussion of the requisite ideas can be found. In this section we develop the necessary changes to the proof in [18].

At the heart of the proof lies a stability result for the block filter [18, Proposition 4.15]. This result must be modified in the present setting to account for the different assumptions on the spatial and temporal correlations. This will be done next, using the generalized comparison Theorem 2.3 much as in the proof of Theorem 4.3.

Proposition 4.6 *Suppose there exist $0 < \varepsilon, \delta < 1$ such that*

$$\varepsilon q^v(x^v, z^v) \leq p^v(x, z^v) \leq \varepsilon^{-1} q^v(x^v, z^v),$$

$$\delta \leq q^v(x^v, z^v) \leq \delta^{-1}$$

for every $v \in V$ and $x, z \in \mathbb{X}$, where $q^v : \mathbb{X}^v \times \mathbb{X}^v \to \mathbb{R}_+$ is a transition density with respect to ψ^v. Suppose also that we can choose $q \in \mathbb{N}$ and $\beta > 0$ such that

$$c := 3q \Delta^2 e^{\beta q} (1 - \varepsilon^{2(\Delta+1)}) + e^\beta (1 - \varepsilon^2 \delta^2) + e^{\beta q} (1 - \varepsilon^2 \delta^2)^q < 1.$$

Then we have

$$\|\tilde{F}_n \cdots \tilde{F}_{s+1} \delta_\sigma - \tilde{F}_n \cdots \tilde{F}_{s+1} \delta_{\tilde{\sigma}}\|_J \leq \frac{2}{1-c} \text{ card } J \, e^{-\beta(n-s)}$$

for every $s < n$, $\sigma, \tilde{\sigma} \in \mathbb{X}$, $K \in \mathcal{K}$ and $J \subseteq K$.

Proof We fix $n > 0$, $K \in \mathcal{K}$, and $J \subseteq K$ throughout. We will also assume for notational simplicity that $s = 0$. As $\tilde{F}_k$ differ for different k only by their dependence on different observations Y_k, and as the conclusion of the Proposition is independent of the observations, the conclusion for $s = 0$ extends trivially to any $s < n$.

As in Theorem 4.3, the idea behind the proof is to introduce a Markov random field ρ of which the block filter is a marginal, followed by an application of the generalized comparison theorem. Unfortunately, the construction in the proof of Theorem 4.3 is not appropriate in the present setting, as there all the local interactions depend on the initial condition σ. That was irrelevant in Theorem 4.3 where the initial condition was fixed, but is fatal in the present setting where we aim to understand a perturbation to the initial condition. Instead, we will use a more elaborate construction of ρ introduced in [18], called the *computation tree*, that we presently recall.

Define for $K' \in \mathcal{K}$ the block neighborhood $N(K') := \{K'' \in \mathcal{K} : d(K', K'') \leq r\}$ (we recall that card $N(K') \leq \Delta_{\mathcal{K}}$). We can evidently write

$$\mathsf{B}^{K'} \tilde{\mathsf{F}}_s \bigotimes_{K'' \in \mathcal{K}} \mu^{K''} = \mathsf{C}_s^{K'} \mathsf{P}^{K'} \bigotimes_{K'' \in N(K')} \mu^{K''},$$

where we define for any probability η on $\mathbb{X}^{K'}$

$$(\mathsf{C}_s^{K'} \eta)(A) := \frac{\int \mathbf{1}_A(x^{K'}) \prod_{v \in K'} g^v(x^v, Y_s^v) \, \eta(dx^{K'})}{\int \prod_{v \in K'} g^v(x^v, Y_s^v) \, \eta(dx^{K'})},$$

and for any probability η on $\mathbb{X}^{\bigcup_{K'' \in N(K')} K''}$

$$(\mathsf{P}^{K'} \eta)(A) := \int \mathbf{1}_A(x^{K'}) \prod_{v \in K'} p^v(z, x^v) \, \psi^v(dx^v) \, \eta(dz).$$

Iterating this identity yields

$$\mathsf{B}^K \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_\sigma = \mathsf{C}_n^K \mathsf{P}^K \bigotimes_{K_{n-1} \in N(K)} \left[\cdots \mathsf{C}_2^{K_2} \mathsf{P}^{K_2} \bigotimes_{K_1 \in N(K_2)} \left[\mathsf{C}_1^{K_1} \mathsf{P}^{K_1} \bigotimes_{K_0 \in N(K_1)} \delta_{\sigma^{K_0}} \right] \cdots \right].$$

The nested products can be naturally viewed as defining a tree.

To formalize this idea, define the tree index set (we write $K_n := K$ for simplicity)

$$T := \{[K_u \cdots K_{n-1}] : 0 \leq u < n, \; K_s \in N(K_{s+1}) \text{ for } u \leq s < n\} \cup \{[\varnothing]\}.$$

The root of the tree $[\varnothing]$ represents the block K at time n, while $[K_u \cdots K_{n-1}]$ represents a duplicate of the block K_u at time u that affects the root along the branch $K_u \to K_{u+1} \to \cdots \to K_{n-1} \to K$. The set of sites corresponding to the tree is

$$I = \{[K_u \cdots K_{n-1}]v : [K_u \cdots K_{n-1}] \in T, \; v \in K_u\} \cup \{[\varnothing]v : v \in K\},$$

and the corresponding configuration space is $\mathbb{S} = \prod_{i \in I} \mathbb{X}^i$ with $\mathbb{X}^{[t]v} := \mathbb{X}^v$. The following tree notation will be used throughout the proof. Define for vertices of the tree T the depth $d([K_u \cdots K_{n-1}]) := u$ and $d([\varnothing]) := n$. For every site $[t]v \in I$, we define the associated

vertex $v(i) := v$ and depth $d(i) := d([t])$. Define also the sets $I_+ := \{i \in I : d(i) > 0\}$ and $T_0 := \{[t] \in T : d([t]) = 0\}$ of non-leaf sites and leaf vertices, respectively. Define the set of children $c(i)$ of site $i \in I$ as

$$c([K_u \cdots K_{n-1}]v) := \{[K_{u-1} \cdots K_{n-1}]v' : K_{u-1} \in N(K_u), \ v' \in N(v)\},$$

and similarly for $c([\varnothing]v)$. Finally, we will often identify a vertex $[K_u \cdots K_{n-1}] \in T$ with the set of sites $\{[K_u \cdots K_{n-1}]v : v \in K_u\}$, and analogously for $[\varnothing]$.

Having introduced the tree structure, we now define probabilities $\rho, \tilde{\rho}$ on $\mathbb{S}$ by

$$\rho(A) = \frac{\int \mathbf{1}_A(x) \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^i) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \delta_{\sigma^{[t]}}(dx^{[t]})}{\int \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^i) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \delta_{\sigma^{[t]}}(dx^{[t]})},$$

$$\tilde{\rho}(A) = \frac{\int \mathbf{1}_A(x) \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^i) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \delta_{\tilde{\sigma}^{[t]}}(dx^{[t]})}{\int \prod_{i \in I_+} p^{v(i)}(x^{c(i)}, x^i) g^{v(i)}(x^i, Y^i) \psi^{v(i)}(dx^i) \prod_{[t] \in T_0} \delta_{\tilde{\sigma}^{[t]}}(dx^{[t]})},$$

where we write $\sigma^{[K_0 \cdots K_{n-1}]} := \sigma^{K_0}$ and $Y^i := Y_{d(i)}^{v(i)}$ for simplicity. By construction, $\mathsf{B}^K \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_\sigma$ coincides with the marginal of ρ on the root of the computation tree, while $\mathsf{B}^K \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_{\tilde{\sigma}}$ coincides with the marginal of $\tilde{\rho}$ on the root of the tree: this is easily seen by expanding the above nested product identity. In particular, we obtain

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_\sigma - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_{\tilde{\sigma}}\|_J = \|\rho - \tilde{\rho}\|_{[\varnothing]J},$$

and we aim to apply the comparison theorem to estimate this quantity.

The construction of the computation tree that we have just given is identical to the construction in [18]. We deviate from the proof of [18] from this point onward, since we must use Theorem 2.3 instead of the classical Dobrushin comparison theorem to account for the distinction between temporal and spatial correlations.

Fix $q \geq 1$. In analogy with the proof of Theorem 4.3, we consider a cover $\mathcal{J}$ consisting of blocks of sites $i \in I$ such that $(l-1)q < d(i) \leq lq \wedge n$ and $v(i) = v$. Here, however, the same vertex v is duplicated many times in the tree, so we end up with many connected blocks of different lengths. To keep track of these, define

$$I_0 := \{i \in I : d(i) = 0\}, \quad I_l := \{i \in I : d(i) = (l-1)q + 1\}$$

for $1 \leq l \leq \lceil n/q \rceil$, and let

$$\ell([K_u, \ldots, K_{n-1}]v) := \max\{s \geq u : K_u = K_{u+1} = \cdots = K_s\}.$$

We now define the cover $\mathcal{J}$ as

$$\mathcal{J} = \{J_l^i : 0 \leq l \leq \lceil n/q \rceil, \ i \in I_l\},$$

where

$$J_0^i := \{i\}, \quad J_l^i := \{[K_u \cdots K_{n-1}]v : (l-1)q + 1 \leq u \leq lq \wedge \ell(i)\}$$

for $i = [K_{(l-1)q+1} \cdots K_{n-1}]v \in I_l$ and $1 \leq l \leq \lceil n/q \rceil$. It is easily seen that $\mathcal{J}$ is in fact a partition of the computation tree I into linear segments.

Having defined the cover $\mathcal{J}$, we must consider a coupled update rule. We choose the natural local updates $\gamma_x^J(dz^J) = \rho(dz^J | x^{I \setminus J})$ and $\tilde{\gamma}_x^J(dz^J) = \tilde{\rho}(dz^J | x^{I \setminus J})$, with the coupled updates $Q_{x,z}^J$ and $\hat{Q}_x^J$ to be constructed below. Theorem 2.3 yields

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_\sigma - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_{\tilde{\sigma}}\|_J = \|\rho - \tilde{\rho}\|_{[\varnothing]J} \leq 2 \sum_{i \in [\varnothing]J} \sum_{j \in I} D_{ij} b_j$$

provided that $D = \sum_{k=0}^{\infty} C^k < \infty$ (cf. Corollary 2.6), where

$$C_{ij} = \sup_{\substack{x,z \in \mathbb{S}: \\ x^{I \setminus \{j\}} = z^{I \setminus \{j\}}}} \int \mathbf{1}_{\omega_i \neq \omega'_i} \, Q_{x,z}^{J(i)}(d\omega, d\omega'),$$

$$b_i = \sup_{x \in \mathbb{S}} \int \mathbf{1}_{\omega_i \neq \omega'_i} \, \hat{Q}_x^{J(i)}(d\omega, d\omega'),$$

and where we write $J(i)$ for the unique block $J \in \mathcal{J}$ that contains $i \in I$. To proceed, we must introduce coupled updates $Q_{x,z}^J$ and $\hat{Q}_x^J$ and estimate C_{ij} and b_j.

Fix until further notice a block $J = J_l^i \in \mathcal{J}$ with $i = [K_{(l-1)q+1} \cdots K_{n-1}]v \in I_l$ and $1 \leq l \leq \lceil n/q \rceil$. From the definition of ρ, we can compute explicitly

$$\gamma_x^J(A) = \frac{\int \mathbf{1}_A(x^J) \, p^v(x^{c(i)}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \, \psi^v(dx^b)}{\int p^v(x^{c(i)}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \, \psi^v(dx^b)}$$

using the Bayes formula. We now proceed to construct couplings $Q_{x,z}^J$ of γ_x^J and γ_z^J for $x, z \in \mathbb{S}$ that differ only at the site $j \in I$. We distinguish the following cases:

1. $d(j) = (l-1)q$ and $v(j) \neq v$;
2. $d(j) = (l-1)q$ and $v(j) = v$;
3. $(l-1)q + 1 \leq d(j) \leq lq \wedge \ell(i)$ and $v(j) \neq v$;
4. $d(j) = lq \wedge \ell(i) + 1$ and $v(j) \neq v$;
5. $d(j) = lq \wedge \ell(i) + 1$ and $v(j) = v$.

It is easily seen that γ_x^J does not depend on x^j except in one of the above cases. Thus when j satisfies none of the above conditions, we can set $C_{aj} = 0$ for $a \in J$.

Case 1. In this case, we must have $j \in c(i)$ with $v(j) \neq v$. Note that

$$\gamma_x^J(A) \geq \varepsilon^2 \times \frac{\int \mathbf{1}_A(x^J) \, q^v(x^{i-}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \psi^v(dx^b)}{\int q^v(x^{i-}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \psi^v(dx^b)},$$

where $i_- \in c(i)$ is the (unique) child of i such that $v(i_-) = v(i)$. As the right hand side does not depend on x^j, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.4 such that $C_{aj} \leq 1 - \varepsilon^2$ for every $a \in J$ and $x, z \in \mathbb{S}$ such that $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$.

Case 2. In this case $j = i_-$. Let us write $J = \{i_1, \ldots, i_u\}$ where $u = \text{card } J$ and $d(i_k) = (l-1)q + k$ for $k = 1, \ldots, u$. Thus $i_1 = i$, and we define $i_0 = i_-$. Let us also write $\tilde{J}_k = \{i_k, \ldots, i_u\}$. Then we can define the transition kernels on $\mathbb{X}^v$

$$P_{k,x}(\omega, A) = \frac{\int \mathbf{1}_A(x^{i_k}) \, p^v(\omega x^{c(i_k) \setminus i_{k-1}}, x^{i_k}) \prod_{\tilde{J}_k \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \psi^v(dx^b)}{\int p^v(\omega x^{c(i_k) \setminus i_{k-1}}, x^{i_k}) \prod_{\tilde{J}_k \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \psi^v(dx^b)}$$

for $k = 1, \ldots, u$. By construction, $P_{k,x}(x^{i_{k-1}}, dx^{i_k}) = \gamma_x^J(dx^{i_k} | x^{i_1}, \ldots, x^{i_{k-1}})$, so we are in the setting of Lemma 4.5. Moreover, we can estimate

$$P_{k,x}(\omega, A) \geq \varepsilon^2 \delta^2 \frac{\int \mathbf{1}_A(x^{i_k}) \prod_{\tilde{J}_k \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \, \psi^v(dx^b)}{\int \prod_{\tilde{J}_k \cap c(a) \neq \varnothing} p^{v(a)}(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \, \psi^v(dx^b)}.$$

Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I \setminus \{j\}} = z^{I \setminus \{j\}}$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.5 such that $C_{i_k j} \leq (1 - \varepsilon^2 \delta^2)^k$ for every $k = 1, \ldots, u$.

Case 3. In this case $j \in \bigcup_{a \in I_+ : J \cap c(a) \neq \varnothing} c(a)$ or $J \cap c(j) \neq \varnothing$, with $v(j) \neq v$. We remark for future reference that there are at most $q\Delta^2$ such sites j. Note that

$$\gamma_x^J(A) \geq \varepsilon^{2(\Delta+1)} \times \frac{\int \mathbf{1}_A(x^J) \, p^v(x^{c(i)}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} \beta^a(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \, \psi^v(dx^b)}{\int p^v(x^{c(i)}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} \beta^a(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \, \psi^v(dx^b)},$$

where $\beta^a(x^{c(a)}, x^a) = q^{v(a)}(x^{a-}, x^a)$ when $j = a$ or $j \in c(a)$ and $\beta^a(x^{c(a)}, x^a) = p^{v(a)}(x^{c(a)}, x^a)$ otherwise. The right hand side of this expression does not depend on x^j as the terms $q^{v(a)}(x^{a-}, x^a)$ for $v(a) \neq v$ cancel in the numerator and denominator. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I\backslash\{j\}} = z^{I\backslash\{j\}}$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.4 such that $C_{aj} \leq 1 - \varepsilon^{2(\Delta+1)}$ for every $a \in J$.

Case 4. In this case $J \cap c(j) \neq \varnothing$ with $v(j) \neq v$. Note that

$$\gamma_x^J(A) \geq \varepsilon^2 \frac{\int \mathbf{1}_A(x^J) \, p^v(x^{c(i)}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} \beta^a(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \, \psi^v(dx^b)}{\int p^v(x^{c(i)}, x^i) \prod_{a \in I_+ : J \cap c(a) \neq \varnothing} \beta^a(x^{c(a)}, x^a) \prod_{b \in J} g^v(x^b, Y^b) \, \psi^v(dx^b)},$$

where we have defined $\beta^a(x^{c(a)}, x^a) = q^{v(a)}(x^{a-}, x^a)$ whenever $j = a$, and we define $\beta^a(x^{c(a)}, x^a) = p^{v(a)}(x^{c(a)}, x^a)$ otherwise. The right hand side does not depend on x^j as the term $q^{v(j)}(x^{j-}, x^j)$ cancels in the numerator and denominator. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I\backslash\{j\}} = z^{I\backslash\{j\}}$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.4 such that $C_{aj} \leq 1 - \varepsilon^2$ for every $a \in J$.

Case 5. In this case $j_- \in J$. Note that the existence of such j necessarily implies that $\ell(i) > lq$ by the definition of J. Thus we can write $J = \{i_1, \ldots, i_q\}$ where $d(i_k) = lq - k + 1$ for $k = 1, \ldots, q$, and we define $i_0 = j$. Let us also define the sets $\tilde{J}_k = \{i_k, \ldots, i_q\}$. Then we can define for $k = 1, \ldots, q$ the transition kernels on $\mathbb{X}^v$

$$P_{k,x}(\omega, A) = \frac{\int \mathbf{1}_A(x^{i_k}) \, p^v(x^{c(i_q)}, x^{i_q}) \prod_{a \in I_+ : \tilde{J}_k \cap c(a) \neq \varnothing} \beta^a_\omega(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \, \psi^v(dx^b)}{\int p^v(x^{c(i_q)}, x^{i_q}) \prod_{a \in I_+ : \tilde{J}_k \cap c(a) \neq \varnothing} \beta^a_\omega(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \, \psi^v(dx^b)}$$

with $\beta^a_\omega(x^{c(a)}, x^a) = p^v(x^{c(a)}, \omega)$ if $a = i_{k-1}$ and $\beta^a_\omega(x^{c(a)}, x^a) = p^{v(a)}(x^{c(a)}, x^a)$ otherwise. By construction $P_{k,x}(x^{i_{k-1}}, dx^{i_k}) = \gamma_x^J(dx^{i_k} | x^{i_1}, \ldots, x^{i_{k-1}})$, so we are in the setting of Lemma 4.5. Moreover, we can estimate

$$P_{k,x}(\omega, A) \geq \varepsilon^2 \delta^2$$
$$\times \frac{\int \mathbf{1}_A(x^{i_k}) \, p^v(x^{c(i_q)}, x^{i_q}) \prod_{a \in I_+ : \tilde{J}_k \cap c(a) \neq \varnothing} \beta^a(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \, \psi^v(dx^b)}{\int p^v(x^{c(i_q)}, x^{i_q}) \prod_{a \in I_+ : \tilde{J}_k \cap c(a) \neq \varnothing} \beta^a(x^{c(a)}, x^a) \prod_{b \in \tilde{J}_k} g^v(x^b, Y^b) \, \psi^v(dx^b)},$$

where $\beta^a(x^{c(a)}, x^a) = 1$ if $a = i_{k-1}$ and $\beta^a(x^{c(a)}, x^a) = p^{v(a)}(x^{c(a)}, x^a)$ otherwise. Thus whenever $x, z \in \mathbb{S}$ satisfy $x^{I\backslash\{j\}} = z^{I\backslash\{j\}}$, we can construct a coupling $Q_{x,z}^J$ using Lemma 4.5 such that $C_{i_k j} \leq (1 - \varepsilon^2\delta^2)^k$ for every $k = 1, \ldots, q$.

We have now constructed coupled updates $Q_{x,z}^J$ for every pair $x, z \in \mathbb{S}$ that differ only at one point. Collecting the above bounds on the matrix C, we can estimate

$$\sum_{j \in I} e^{\beta|d(a) - d(j)|} C_{aj} \leq 3q\Delta^2 e^{\beta q} (1 - \varepsilon^{2(\Delta+1)}) + e^\beta (1 - \varepsilon^2\delta^2) + e^{\beta q} (1 - \varepsilon^2\delta^2)^q =: c$$

whenever $a \in J$, where we have used the convexity of the function $\alpha^{x+1} + \alpha^{q-x}$.

Up to this point we have considered an arbitrary block $J = J_l^i \in \mathcal{J}$ with $1 \leq l \leq \lceil n/q \rceil$. However, in the remaining case $l = 0$ it is easily seen that $\gamma_x^J = \delta_{\sigma^J}$ does not depend on x,

so we can evidently set $C_{aj} = 0$ for $a \in J$. Thus we have shown that

$$\|C\|_{\infty, \beta m} := \max_{i \in I} \sum_{j \in I} e^{\beta m(i,j)} C_{ij} \leq c,$$

where we define the pseudometric $m(i, j) = |d(i) - d(j)|$. On the other hand, in the present setting it is evident that $\gamma_x^J = \tilde{\gamma}_x^J$ when $J = J_l^i \in \mathcal{J}$ with $1 \leq l \leq \lceil n/q \rceil$. We can therefore choose couplings $\hat{Q}_x^J$ such that $b_i \leq \mathbf{1}_{d(i)=0}$ for all $i \in I$. Substituting into the comparison theorem and arguing as in the proof of Theorem 4.3 yields

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_\sigma - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_1 \delta_{\tilde{\sigma}}\|_J \leq \frac{2}{1-c} \text{ card } J \, e^{-\beta n}.$$

Thus the proof is complete.   □

Proposition 4.6 provides control of the block filter as a function of time but not as a function of the initial conditions. The dependence on the initial conditions can however be incorporated *a posteriori* as in the proof of [18, Proposition 4.17]. This yields the following result, which forms the basis for the proof of Theorem 4.8 below.

Corollary 4.7 (Block filter stability) *Suppose there exist $0 < \varepsilon, \delta < 1$ such that*

$$\varepsilon q^v(x^v, z^v) \leq p^v(x, z^v) \leq \varepsilon^{-1} q^v(x^v, z^v),$$

$$\delta \leq q^v(x^v, z^v) \leq \delta^{-1}$$

for every $v \in V$ and $x, z \in \mathbb{X}$, where $q^v : \mathbb{X}^v \times \mathbb{X}^v \to \mathbb{R}_+$ is a transition density with respect to ψ^v. Suppose also that we can choose $q \in \mathbb{N}$ and $\beta > 0$ such that

$$c := 3q\Delta^2 e^{\beta q}(1 - \varepsilon^{2(\Delta+1)}) + e^\beta(1 - \varepsilon^2\delta^2) + e^{\beta q}(1 - \varepsilon^2\delta^2)^q < 1.$$

Let μ and ν be (possibly random) probability measures on $\mathbb{X}$ of the form

$$\mu = \bigotimes_{K \in \mathcal{K}} \mu^K, \quad \nu = \bigotimes_{K \in \mathcal{K}} \nu^K.$$

Then we have

$$\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \mu - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \nu\|_J \leq \frac{2}{1-c} \text{ card } J \, e^{-\beta(n-s)},$$

as well as

$$\mathbf{E}[\|\tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \mu - \tilde{\mathsf{F}}_n \cdots \tilde{\mathsf{F}}_{s+1} \nu\|_J^2]^{1/2}$$
$$\leq \frac{2}{1-c} \frac{1}{(\varepsilon\delta)^{2|\mathcal{K}|_\infty}} \text{ card } J \, (e^{-\beta} \Delta_{\mathcal{K}})^{n-s} \max_{K \in \mathcal{K}} \mathbf{E}[\|\mu^K - \nu^K\|^2]^{1/2},$$

for every $s < n$, $K \in \mathcal{K}$ and $J \subseteq K$.

Proof The proof is a direct adaptation of [18, Proposition 4.17].   □

The block filter stability result in [18] is the only place in the proof of the variance bound where the inadequacy of the classical comparison theorem plays a role. Having exploited the generalized comparison Theorem 2.3 to extend the stability results in [18] to the present setting, we would therefore expect that the remainder of the proof of the variance bound follows verbatim from [18]. Unfortunately, however, there is a complication: the result of Corollary 4.7 is not as powerful as the corresponding result in [18]. Note that the first (uniform)

bound in Corollary 4.7 decays exponentially in time n, but the second (initial condition dependent) bound only decays in n if it happens to be the case that $e^{-\beta}\Delta_{\mathcal{K}} < 1$. As in [18] both the spatial and temporal interactions were assumed to be sufficiently weak, we could assume that the latter was always the case. In the present setting, however, it is possible that $e^{-\beta}\Delta_{\mathcal{K}} \geq 1$ no matter how weak are the spatial correlations.

To surmount this problem, we will use a slightly different error decomposition than was used in [18] to complete the proof. The present approach is inspired by [3]. The price we pay is that the variance bound scales in the number of samples as $N^{-\gamma}$ where γ may be less than the optimal (by the central limit theorem) rate $\frac{1}{2}$. It is likely that a more sophisticated method of proof would yield the optimal $N^{\frac{1}{2}}$ rate. However, let us note that in order to put the block particle filter to good use we must optimize over the size of the blocks in $\mathcal{K}$, and optimizing the error bound in Theorem 4.2 yields at best a rate of order $N^{-\alpha}$ for some constant α depending on the constants β_1, β_2. As the proof of Theorem 4.2 is not expected to yield realistic values for the constants β_1, β_2, the suboptimality of the variance rate γ does not significantly alter the practical conclusions that can be drawn from Theorem 4.2.

We now state the variance bound. The following is the main result of this section.

Theorem 4.8 (Variance term) *Suppose there exist $0 < \varepsilon, \delta, \kappa < 1$ such that*

$$\varepsilon q^v(x^v, z^v) \leq p^v(x, z^v) \leq \varepsilon^{-1} q^v(x^v, z^v),$$

$$\delta \leq q^v(x^v, z^v) \leq \delta^{-1},$$

$$\kappa \leq g^v(x^v, y^v) \leq \kappa^{-1}$$

for every $v \in V$, $x, z \in \mathbb{X}$, and $y \in \mathbb{Y}$, where $q^v : \mathbb{X}^v \times \mathbb{X}^v \to \mathbb{R}_+$ is a transition density with respect to ψ^v. Suppose also that we can choose $q \in \mathbb{N}$ and $\beta > 0$ so that

$$c := 3q\Delta^2 e^{\beta q}(1 - \varepsilon^{2(\Delta+1)}) + e^{\beta}(1 - \varepsilon^2\delta^2) + e^{\beta q}(1 - \varepsilon^2\delta^2)^q < 1.$$

Then for every $n \geq 0$, $\sigma \in \mathbb{X}$, $K \in \mathcal{K}$ and $J \subseteq K$, the following hold:

1. *If $e^{-\beta}\Delta_{\mathcal{K}} < 1$, we have*

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \text{card } J \, \frac{32\Delta_{\mathcal{K}}}{1-c} \, \frac{2 - e^{-\beta}\Delta_{\mathcal{K}}}{1 - e^{-\beta}\Delta_{\mathcal{K}}} \, \frac{(\varepsilon\delta\kappa^{\Delta_{\mathcal{K}}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}}.$$

2. *If $e^{-\beta}\Delta_{\mathcal{K}} = 1$, we have*

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \text{card } J \, \frac{16\beta^{-1}\Delta_{\mathcal{K}}}{1-c} \, (\varepsilon\delta\kappa^{\Delta_{\mathcal{K}}})^{-4|\mathcal{K}|_\infty} \, \frac{3 + \log N}{N^{\frac{1}{2}}}.$$

3. *If $e^{-\beta}\Delta_{\mathcal{K}} > 1$, we have*

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \text{card } J \, \frac{32\Delta_{\mathcal{K}}}{1-c} \left\{ \frac{1}{e^{-\beta}\Delta_{\mathcal{K}} - 1} + 2 \right\} \frac{(\varepsilon\delta\kappa^{\Delta_{\mathcal{K}}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{\beta}{2\log\Delta_{\mathcal{K}}}}}.$$

The proof of Theorem 4.8 combines the stability bounds of Corollary 4.7 and one-step bounds on the sampling error, [18, Lemma 4.19 and Proposition 4.22], that can be used verbatim in the present setting. We recall the latter here.

Proposition 4.9 (Sampling error) *Suppose there exist $0 < \varepsilon, \delta, \kappa < 1$ such that*

$$\varepsilon q^v(x^v, z^v) \leq p^v(x, z^v) \leq \varepsilon^{-1} q^v(x^v, z^v),$$

$$\delta \leq q^v(x^v, z^v) \leq \delta^{-1},$$

$$\kappa \leq g^v(x^v, y^v) \leq \kappa^{-1}$$

for every $v \in V$, $x, z \in \mathbb{X}$, *and* $y \in \mathbb{Y}$. *Then we have*

$$\max_{K \in \mathcal{K}} |||\tilde{F}_n \hat{\pi}_{n-1}^\sigma - \hat{F}_n \hat{\pi}_{n-1}^\sigma|||_K \leq \frac{2\kappa^{-2|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}}$$

and

$$\max_{K \in \mathcal{K}} \mathbf{E}[\|\tilde{F}_{s+1} \tilde{F}_s \hat{\pi}_{s-1}^\sigma - \tilde{F}_{s+1} \hat{F}_s \hat{\pi}_{s-1}^\sigma\|_K^2]^{1/2} \leq \frac{16 \Delta_{\mathcal{K}} (\varepsilon\delta)^{-2|\mathcal{K}|_\infty} \kappa^{-4|\mathcal{K}|_\infty \Delta_{\mathcal{K}}}}{N^{\frac{1}{2}}}$$

for every $0 < s < n$ *and* $\sigma \in \mathbb{X}$.

Proof Immediate from [18, Lemma 4.19 and Proposition 4.22] replacing ε by $\varepsilon\delta$. □

We can now prove Theorem 4.8.

Proof (Theorem 4.8) We fix for the time being an integer $t \geq 1$ (we will optimize over t at the end of the proof). We argue differently when $n \leq t$ and when $n > t$.

Suppose first that $n \leq t$. In this case, we estimate

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J = |||\tilde{F}_n \cdots \tilde{F}_1 \delta_\sigma - \hat{F}_n \cdots \hat{F}_1 \delta_\sigma|||_J$$

$$\leq \sum_{k=1}^n |||\tilde{F}_n \cdots \tilde{F}_{k+1} \tilde{F}_k \hat{\pi}_{k-1}^\sigma - \tilde{F}_n \cdots \tilde{F}_{k+1} \hat{F}_k \hat{\pi}_{k-1}^\sigma|||_J$$

using a telescoping sum and the triangle inequality. The term $k = n$ in the sum is estimated by the first bound in Proposition 4.9, while the remaining terms are estimated by the second bound of Corollary 4.7 and Proposition 4.9. This yields

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \operatorname{card} J \, \frac{32 \Delta_{\mathcal{K}}}{1-c} \frac{(\varepsilon\delta\kappa^{\Delta_{\mathcal{K}}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}} \left\{ \frac{(e^{-\beta} \Delta_{\mathcal{K}})^{n-1} - 1}{e^{-\beta} \Delta_{\mathcal{K}} - 1} + 1 \right\}$$

(in the case $e^{-\beta} \Delta_{\mathcal{K}} = 1$, the quantity between the brackets $\{ \cdot \}$ equals n).

Now suppose that $n > t$. Then we decompose the error as

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq |||\tilde{F}_n \cdots \tilde{F}_{n-t+1} \tilde{\pi}_{n-t}^\sigma - \tilde{F}_n \cdots \tilde{F}_{n-t+1} \hat{\pi}_{n-t}^\sigma|||_J$$

$$+ \sum_{k=n-t+1}^n |||\tilde{F}_n \cdots \tilde{F}_{k+1} \tilde{F}_k \hat{\pi}_{k-1}^\sigma - \tilde{F}_n \cdots \tilde{F}_{k+1} \hat{F}_k \hat{\pi}_{k-1}^\sigma|||_J,$$

that is, we develop the telescoping sum for t steps only. The first term is estimated by the first bound in Corollary 4.7, the sum is estimated as in the case $n \leq t$. This yields

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \leq \frac{\operatorname{card} J}{1-c} \left[2e^{-\beta t} + \frac{32 \Delta_{\mathcal{K}} (\varepsilon\delta\kappa^{\Delta_{\mathcal{K}}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}} \left\{ \frac{(e^{-\beta} \Delta_{\mathcal{K}})^{t-1} - 1}{e^{-\beta} \Delta_{\mathcal{K}} - 1} + 1 \right\} \right]$$

(in the case $e^{-\beta} \Delta_{\mathcal{K}} = 1$, the quantity between the brackets $\{ \cdot \}$ equals t).

We now consider separately the three cases in the statement of the Theorem.

Case 1. In this case we choose $t = n$, and note that

$$\frac{(e^{-\beta} \Delta_{\mathcal{K}})^{n-1} - 1}{e^{-\beta} \Delta_{\mathcal{K}} - 1} + 1 \leq \frac{2 - e^{-\beta} \Delta_{\mathcal{K}}}{1 - e^{-\beta} \Delta_{\mathcal{K}}} \quad \text{for all } n \geq 1.$$

Thus the result follows from the first bound above.

Case 2. In this case we have

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \le \frac{\text{card } J}{1-c} \left[2e^{-\beta t} + \frac{32\Delta_\mathcal{K}(\varepsilon\delta\kappa^{\Delta_\mathcal{K}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}} t \right]$$

for all $t, n \ge 1$. Now choose $t = \lceil (2\beta)^{-1} \log N \rceil$. Then

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \le \frac{\text{card } J}{1-c} \left[16\beta^{-1}\Delta_\mathcal{K}(\varepsilon\delta\kappa^{\Delta_\mathcal{K}})^{-4|\mathcal{K}|_\infty} \frac{\log N}{N^{\frac{1}{2}}} + \frac{34\Delta_\mathcal{K}(\varepsilon\delta\kappa^{\Delta_\mathcal{K}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}} \right],$$

which readily yields the desired bound.

Case 3. In this case we have

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \le \frac{\text{card } J}{1-c} \left[2e^{-\beta t} + \frac{32\Delta_\mathcal{K}(\varepsilon\delta\kappa^{\Delta_\mathcal{K}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{1}{2}}} \left\{ \frac{(e^{-\beta}\Delta_\mathcal{K})^{t-1} - 1}{e^{-\beta}\Delta_\mathcal{K} - 1} + 1 \right\} \right]$$

for all $t, n \ge 1$. Now choose $t = \left\lceil \frac{\log N}{2\log\Delta_\mathcal{K}} \right\rceil$. Then

$$|||\tilde{\pi}_n^\sigma - \hat{\pi}_n^\sigma|||_J \le \text{card } J \frac{32\Delta_\mathcal{K}}{1-c} \left\{ \frac{1}{e^{-\beta}\Delta_\mathcal{K} - 1} + 2 \right\} \frac{(\varepsilon\delta\kappa^{\Delta_\mathcal{K}})^{-4|\mathcal{K}|_\infty}}{N^{\frac{\beta}{2\log\Delta_\mathcal{K}}}},$$

and the proof is complete.   □

The conclusion of Theorem 4.2 now follows readily from Theorems 4.3 and 4.8. We must only check that the assumptions Theorems 4.3 and 4.8 are satisfied. The assumption of Theorem 4.3 is slightly stronger than that of Theorem 4.8, so it suffices to consider the former. To this end, fix $0 < \delta < 1$, and choose $q \in \mathbb{N}$ such that

$$1 - \delta^2 + (1 - \delta^2)^q < 1.$$

Then we may evidently choose $0 < \varepsilon_0 < 1$, depending on δ and Δ only, such that

$$3q\Delta^2(1 - \varepsilon^{2(\Delta+1)}) + 1 - \varepsilon^2\delta^2 + (1 - \varepsilon^2\delta^2)^q < 1$$

for all $\varepsilon_0 < \varepsilon \le 1$. This is the constant ε_0 that appears in the statement of Theorem 4.2. Finally, we can clearly choose $\beta > 0$ sufficiently close to zero (depending on $\delta, \varepsilon, r, \Delta$ only) such that $c < 1$. Thus the proof of Theorem 4.2 is complete.

References

1. Cappé, O., Moulines, E., Rydén, T.: Inference in Hidden Markov Models. Springer Series in Statistics. Springer, New York (2005)
2. Dai Pra, P., Scoppola, B., Scoppola, E.: Sampling from a Gibbs measure with pair interaction by means of PCA. J. Stat. Phys. **149**(4), 722–737 (2012)
3. Del Moral, P., Guionnet, A.: On the stability of interacting processes with applications to filtering and genetic algorithms. Ann. Inst. H. Poincaré Probab. Stat. **37**(2), 155–194 (2001)
4. Dobrushin, R.L., Shlosman, S.B.: Constructive criterion for the uniqueness of Gibbs field. In: Statistical physics and dynamical systems (Köszeg, 1984), Program Physics, vol. 10, pp. 347–370. Birkhäuser Boston, Boston (1985)
5. Dobrušin, R.L.: Definition of a system of random variables by means of conditional distributions. Teor. Verojatnost. i Primenen. **15**, 469–497 (1970)

6. Dudley, R.M.: Real Analysis and Probability, Cambridge Studies in Advanced Mathematics, vol. 74. Cambridge University Press, Cambridge (2002) (Revised reprint of the 1989 original)
7. Dyer, M., Goldberg, L.A., Jerrum, M.: Dobrushin conditions and systematic scan. Comb. Probab. Comput. **17**(6), 761–779 (2008)
8. Dyer, M., Goldberg, L.A., Jerrum, M.: Matrix norms and rapid mixing for spin systems. Ann. Appl. Probab. **19**(1), 71–107 (2009)
9. Föllmer, H.: Tail structure of Markov chains on infinite product spaces. Z. Wahrsch. Verw. Gebiete **50**(3), 273–285 (1979)
10. Föllmer, H.: A covariance estimate for Gibbs measures. J. Funct. Anal. **46**(3), 387–395 (1982)
11. Föllmer, H.: Random fields and diffusion processes. In: École d'Été de Probabilités de Saint-Flour XV–XVII, 1985–1987, Lecture Notes in Mathematics, vol. 1362, pp. 101–203. Springer, Berlin (1988)
12. Georgii, H.O.: Measures and Phase Transitions, de Gruyter Studies in Mathematics, vol. 9, 2nd edn. Walter de Gruyter & Co., Berlin (2011)
13. Guionnet, A., Zegarlinski, B.: Lectures on Logarithmic Sobolev Inequalities. In: Séminaire de Probabilités, XXXVI, Lecture Notes in Mathematics, vol. 1801, pp. 1–134. Springer, Berlin (2003)
14. Kallenberg, O.: Foundations of Modern Probability. Probability and its Applications (New York), 2nd edn. Springer-Verlag, New York (2002)
15. Külske, C.: Concentration inequalities for functions of Gibbs fields with application to diffraction and random Gibbs measures. Commun. Math. Phys. **239**(1–2), 29–51 (2003)
16. Lebowitz, J.L., Maes, C., Speer, E.R.: Statistical mechanics of probabilistic cellular automata. J. Stat. Phys. **59**(1–2), 117–170 (1990)
17. Liggett, T.M.: Interacting Particle Systems. Classics in Mathematics. Springer-Verlag, Berlin (2005) (Reprint of the 1985 original)
18. Rebeschini, P., van Handel, R.: Can Local Particle Filters Beat the Curse of Dimensionality? Preprint arxiv:1301.6585 (2013)
19. Royden, H.L.: Real Analysis, 3rd edn. Macmillan Publishing Company, New York (1988)
20. Rue, T.D.L., Fernández, R., Sokal, A.D.: How to clean a dirty floor: probabilistic potential theory and the Dobrushin uniqueness theorem. Markov Process Relat. Fields **14**(1), 1–78 (2008)
21. Simon, B.: The Statistical Mechanics of Lattice Gases. Princeton Series in Physics, vol. I. Princeton University Press, Princeton (1993)
22. Tatikonda, S.C.: Convergence of the sum-product algorithm. In: Information Theory Workshop, 2003. Proceedings. 2003 IEEE, pp. 222–225 (2003)
23. van der Vaart, A.W., Wellner, J.A.: Weak Convergence and Empirical Processes. Springer Series in Statistics. With Applications to Statistics. Springer-Verlag, New York (1996)
24. Villani, C.: Optimal Transport, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 338. Springer-Verlag, Berlin (2009). Old and new
25. Weitz, D.: Combinatorial criteria for uniqueness of Gibbs measures. Random Struct. Algorithms **27**(4), 445–475 (2005)
26. Wu, L.: Poincaré and transportation inequalities for Gibbs measures under the Dobrushin uniqueness condition. Ann. Probab. **34**(5), 1960–1989 (2006)
27. Younes, L.: Parametric inference for imperfectly observed Gibbsian fields. Probab. Theory Relat. Fields **82**(4), 625–645 (1989)