Nonlinear Dynamics for the Ising Model

Pietro Caputo
University of Rome III
Rome, Italy
pietro.caputo@uniroma3.it

Alistair Sinclair
University of California
Berkeley, USA
sinclair@cs.berkeley.edu

ABSTRACT

We introduce and analyze a natural class of nonlinear dynamics for spin systems such as the Ising model. This class of dynamics is based on the framework of mass action kinetics, which models the evolution of systems of entities under pairwise interactions, and captures a number of important nonlinear models from various fields, including chemical reaction networks, Boltzmann's model of an ideal gas, recombination in population genetics, and genetic algorithms. In the context of spin systems, it is a natural generalization of linear dynamics based on Markov chains, such as Glauber dynamics and block dynamics, which are by now well understood. However, the inherent nonlinearity makes the dynamics much harder to analyze, and rigorous quantitative results so far are limited to processes which converge to essentially trivial stationary distributions that are product measures.

In this paper we provide the first quantitative convergence analysis for natural nonlinear dynamics in a combinatorial setting where the stationary distribution contains non-trivial correlations, namely spin systems at high temperatures. We prove that nonlinear versions of both the Glauber dynamics and the block dynamics converge to the Gibbs distribution of the Ising model (with given external fields) in times $O(n \log n)$ and $O(\log n)$ respectively, where n is the size of the underlying graph (number of spins). Given the lack of general analytical methods for such nonlinear systems, our analysis is unconventional, and combines tools such as information percolation (due in the linear setting to Lubetzky and Sly), a novel coupling of the Ising model with Erdős-Rényi random graphs, and non-traditional branching processes augmented by a "fragmentation" process. Our results extend immediately to any spin system with a finite number of spins and bounded interactions.

CCS CONCEPTS

• **Theory of computation** → **Random walks and Markov chains**; *Generating random combinatorial structures*; • **Mathematics of computing** → **Stochastic processes**; *Probabilistic algorithms*; • **Applied computing** → **Physics**; Population genetics.

KEYWORDS

Mass action kinetics; Nonlinear Markov chains; Mixing time; Ising model; Spin systems; Branching processes

STOC '24, June 24–28, 2024, Vancouver, BC, Canada

ACM ISBN 979-8-4007-0383-6/24/06
https://doi.org/10.1145/3618260.3649759

ACM Reference Format:
Pietro Caputo and Alistair Sinclair. 2024. Nonlinear Dynamics for the Ising Model. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing (STOC '24), June 24–28, 2024, Vancouver, BC, Canada.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3618260.3649759

1 INTRODUCTION

Mass action kinetics is a general framework for studying systems of interacting entities. The framework emerged in the study of chemical reaction networks, dating back at least to the seminal work of Horn and Jackson in the 1970s [11], and has seen a resurgence of activity in recent years; see the monograph [7]. However, it also captures a wide range of processes that are of interest in other fields, including Boltzmann's model of an ideal gas [1], classical models of population genetics [10, 22], genetic algorithms in combinatorial optimization [9, 14], and random sampling [18, 20].

We describe mass action kinetics in the special case where all interactions are pairwise and homogeneous; this captures most of the complexity of general systems while keeping notation and technicalities to a minimum. Let Ω denote a finite set of *types*. A (quadratic) mass action system is described by a directed graph whose vertices are unordered pairs of types (σ, σ'), and a directed edge from (σ, σ') to (τ, τ') indicates the presence of a *reaction* in which types σ, σ' combine to produce types τ, τ'. Reactions are governed by a *collision kernel* $Q(\sigma, \sigma' ; \cdot, \cdot)$, where $Q(\sigma, \sigma' ; \tau, \tau')$ is the probability that the outcome of a reaction involving the pair (σ, σ') is the pair (τ, τ').

The state of the system at any time t is fully described by the vector p_t, where $p_t(\sigma)$ is the mass of type σ at time t, normalized so that $\sum_{\sigma \in \Omega} p_t(\sigma) = 1$ (i.e., the $p_t(\sigma)$ can be viewed as concentrations, or probabilities). According to the so-called "mass action" principle, each reaction $(\sigma, \sigma') \to (\tau, \tau')$ takes place at a rate determined by the *product* of the current masses of types σ, σ'. The dynamics of the system is now described by the following set of equations, one for each type $\tau \in \Omega$:

$$p_{t+1}(\tau) = \sum_{\sigma, \sigma', \tau'} p_t(\sigma) p_t(\sigma') Q(\sigma, \sigma' ; \tau, \tau'). \tag{1.1}$$

At this level of generality such systems can be arbitrarily badly behaved (e.g., chaotic), so it is necessary to impose standard regularity conditions. A mass action system is said to be *reversible*[1] or *detailed balanced* if there exists a strictly positive mass vector $\mu = (\mu(\sigma)) > 0$ such that, for all $\sigma, \sigma', \tau, \tau'$,

$$\mu(\sigma)\mu(\sigma') Q(\sigma, \sigma' ; \tau, \tau') = \mu(\tau)\mu(\tau') Q(\tau, \tau' ; \sigma, \sigma'). \tag{1.2}$$

[1] In the mass action kinetics literature, the term "reversible" has unfortunately been used to denote the weaker property that $Q(\sigma, \sigma' ; \tau, \tau') \neq 0$ iff $Q(\tau, \tau' ; \sigma, \sigma') \neq 0$, whereas in physics reversibility is synonymous with detailed balance. In this paper, we shall use the terms "detailed balanced" and "reversible" interchangeably to denote the stronger condition (1.2).

It is easy to check that any such μ is necessarily an *equilibrium* or *stationary* point for the dynamics (1.1). A mass action system may have many positive equilibrium points, but if any one of them satisfies the detailed balance condition then they all do. We stress that we do *not* require the kernel Q to be irreducible (i.e., the directed graph describing it need not be strongly connected), and that the dynamics will in general have *conserved quantities*.

The mass action system defined in (1.1) can be viewed as a natural nonlinear analog of a reversible Markov chain, whose dynamics takes the form $p_{t+1}(\tau) = \sum_\sigma p_t(\sigma)Q(\sigma\,;\,\tau)$, where now $Q(\sigma\,;\,\tau)$ is the transition matrix of the chain and the reversibility condition is $\mu(\sigma)Q(\sigma\,;\,\tau) = \mu(\tau)Q(\tau\,;\,\sigma)$ for all σ, τ. In the linear setting, there are well known criteria for convergence to stationarity and there is a vast literature on mixing times of reversible Markov chains and their algorithmic applications to sampling, approximate counting and integration, statistical physics, etc. By contrast, in the nonlinear setting even the most basic questions are still open: for example, the Global Attractor Conjecture [7] asserts that any detailed balanced[2] mass action system converges to a fixed point when started from any initial point p_0 with full support. (In fact many stationary points may exist, but only one is consistent with any given p_0.) And even in particular cases of interest where convergence has been proved, almost nothing is known about the *rate* of convergence (the analog of the *mixing time* for Markov chains).

In the discrete combinatorial setting, one of very few examples for which useful bounds on the convergence rate are known is the classical Hardy-Weinberg model of genetic recombination [10, 22]. Here the types are bit strings $\Omega = \{0,1\}^n$ (each bit representing an allele on a chromosome), and a reaction between two strings σ, σ' involves picking a "crossover" subset $\Lambda \subseteq \{1,\dots,n\}$ of positions according to some probabilistic rule and exchanging the bits in Λ between σ and σ' to obtain two new strings τ, τ'. (For example, one classical rule is to pick $i \in \{0,\dots,n\}$ u.a.r. and let Λ consist of the first i bits.) It is well known that this dynamics converges to the distribution μ in which all bits are *independent*, with the marginal probabilities of a 1 at each position given by those in the initial distribution p_0. In [4, 17], the rate of convergence was related precisely to the rate at which the strings are *fragmented* by the repeated random cuts Λ, thus enabling very precise estimates of the convergence time for any choice of crossover rule.

The above analysis relies crucially on the fact that in the equilibrium distribution all bits are independent. When there is even a small amount of correlation, there appear to be no techniques available to obtain useful bounds on convergence rates. In this paper, we address this question for arguably the most natural example in which correlations arise, namely the Ising model of statistical physics, a very simple model of nearest-neighbor interactions on a graph. Here the types are spin configurations $\sigma \in \Omega = \{\pm 1\}^V$ which assign one of two possible spin values ± 1 to each vertex of a graph $G = (V, E)$. The probability of any configuration σ is given by the *Gibbs distribution*:

$$\mu_{\mathbf{J},\mathbf{h}}(\sigma) = \frac{1}{Z_{\mathbf{J},\mathbf{h}}} \exp\left\{\frac{1}{2}\sum_{x,y\in V} J_{xy}\sigma_x\sigma_y + \sum_{x\in V} h_x\sigma_x\right\}, \qquad (1.3)$$

where $\mathbf{h} = \{h_x\}_{x\in V}$ is a vector of real numbers whose entry h_x represents the *external field* at vertex x, and $\mathbf{J} = \{J_{xy}\}_{x,y\in V}$ is a symmetric real matrix whose entry J_{xy} represents the *interaction* between spins at adjacent vertices x, y. (When there is no edge between x and y, $J_{xy} = 0$.) The normalizing factor $Z_{\mathbf{J},\mathbf{h}}$ is the *partition function*. Note that we allow the interactions J_{xy} to be either positive (favoring neighbors with *aligned* spins) or negative (favoring *non-aligned* spins), and the fields h_x to be either positive (favoring $+1$ spins) or negative (favoring -1 spins). Setting $\mathbf{J} = \beta A$, where $\beta > 0$ and A is the adjacency matrix of G, corresponds to the standard ferromagnetic Ising model on G at inverse temperature β. Note also that the trivial case $\mathbf{J} = 0$ corresponds to independent spins as in the Hardy-Weinberg model above. We emphasize that, although for simplicity we develop our results for the specific case of the Ising model, they hold equally for any spin system with a constant number of different spins and bounded pairwise interactions (such as the q-state Potts model); see Section 5 for more detail.

The classical *Glauber dynamics* for the Ising model picks a random vertex $x \in V$ at each step and resamples the spin at x according to the correct conditional distribution given its neighboring spins; this Markov chain converges to the Gibbs distribution (1.3) from any initial configuration. The analogous nonlinear mass action kinetics is defined by equation (1.1) with the following kernel: Given two configurations σ, σ', pick a random vertex x and exchange the spins σ_x, σ'_x, obtaining two new configurations τ, τ'. The transition probabilities $Q(\sigma, \sigma'\,;\,\tau, \tau')$ are chosen to satisfy the detailed balance condition (1.2), where $\mu = \mu_{\mathbf{J},\mathbf{h}}$ is the Gibbs distribution. We emphasize that, in contrast to Glauber dynamics, here the system is evolving *endogenously* via pairwise interactions between configurations, rather than via exogenously applied spin updates. Our first result shows that this dynamics converges to the Gibbs distribution (1.3), where the fields $\mathbf{h}$ are determined by the marginal probabilities of the spins at each vertex in the initial distribution. (The fact that the marginals determine a unique vector of fields $\mathbf{h}$ follows from standard convexity arguments; see, e.g., [6].)

Theorem 1.1. *Let p_t denote the distribution at time t for the above mass action kinetics for the Ising model with interactions $\mathbf{J}$ starting from any initial distribution p_0, and let $\mathbf{h}$ be the unique choice of external fields such that the marginal probabilities at each vertex $x \in V$ in $\mu_{\mathbf{J},\mathbf{h}}$ are the same as those in p_0. Then p_t converges to $\mu_{\mathbf{J},\mathbf{h}}$ as $t \to \infty$.*

Note that, unlike the standard Glauber dynamics, the nonlinear dynamics has conserved quantities—namely, the marginal probabilities of the spins at each vertex—and the values of these invariants determine which of the family of stationary points the dynamics converges to. This phenomenon is typical in mass action kinetics. The key to our proof of Theorem 1.1 is establishing an *irreducibility* property: along any trajectory, the probability of any configuration eventually remains uniformly bounded away from zero.

We pause to briefly mention some features of mass action kinetics that make its analysis much more complex than that of Glauber dynamics, and which explain the lack of quantitative convergence results. First, as noted above, there are in general multiple equilibrium points, which are characterized by conserved quantities. Second, unlike the linear case, the total variation distance to stationarity is *not* monotonically decreasing [2, Remark 2.7] and there are

[2] Actually, this property is conjectured to hold under the weaker condition known as "complex balance" [7].

no simple coupling arguments to rely upon. Finally, the nonlinearity means that we do not have at our disposal a spectral theory and other functional analysis tools that have proved so powerful in the analysis of Markov chains. As usual in kinetic theory, a natural way to study convergence to stationarity here is to use relative entropy, which provides a monotonically decreasing functional; however, quantitative analysis of this quantity is a notoriously difficult problem in the nonlinear setting, which has so far been solved only in the non-interacting case (genetic recombination) [3, 4].

Our main result establishes tight bounds on the rate of convergence for this nonlinear dynamics in the so-called "high-temperature" regime, when the interactions are non-trivial but relatively weak. Specifically, the condition we require is that $\max_{x\in V}\sum_{y\in V}|J_{xy}| \le \delta_0$ for some absolute constant $\delta_0 > 0$, i.e., the aggregated strength of all interactions at any given vertex is not too large. This condition mirrors the standard Dobrushin condition for Glauber dynamics, which gives a non-trivial sufficient condition for rapid mixing (see, e.g., [23]). We state this result in the following theorem.

Theorem 1.2. *In the scenario of Theorem 1.1, with the additional assumption that* $\max_{x\in V}\sum_{y\in V}|J_{xy}| \le \delta_0$ *for an absolute constant* $\delta_0 > 0$*, the rate of convergence of* p_t *to* $\mu_{\mathbf{J},\mathbf{h}}$ *is given by*

$$\|p_t - \mu_{\mathbf{J},\mathbf{h}}\|_{\mathrm{TV}} \le Cne^{-ct/n}$$

for absolute constants $C, c > 0$*, where* $\|\cdot\|_{\mathrm{TV}}$ *denotes total variation distance. Thus in particular the time required to achieve variation distance* ε *is* $t = O(n\log(n/\varepsilon))$.

We note that this upper bound on convergence time is (up to constants) the same as for the genetic recombination model discussed above in which just a single allele is exchanged between the strings at each step [17], and is therefore also tight by virtue of the lower bound in the same paper. That model is equivalent to the trivial case of the Ising model in which there are no interactions ($J_{xy} = 0$ for all x, y), with the h_x determined by the marginal probabilities at each allele x (and the spins ±1 identified with the bits $1, 0$). However, as we explain in Section 1.1 below, the correlations present in the Ising model make the analysis much more challenging.

We also consider a "block" version of the nonlinear dynamics, in which σ, σ' exchange spins at a random subset $\Lambda \subseteq V$ of vertices (rather than just at a single randomly chosen vertex). The kernel $Q(\sigma, \sigma'; \tau, \tau')$ is again determined by the detailed balance condition (1.2), and the basic convergence result in Theorem 1.1 still holds. Under the same Dobrushin-type condition on the interactions as in Theorem 1.2, we again obtain a tight bound on the convergence rate:

Theorem 1.3. *With the same notation and assumptions as in Theorem 1.2, the variation distance of the block version of the mass action kinetics for the Ising model satisfies*

$$\|p_t - \mu_{\mathbf{J},\mathbf{h}}\|_{\mathrm{TV}} \le Cn^2e^{-ct}$$

for absolute constants $C, c > 0$*. Thus in particular the time required to achieve variation distance* ε *is* $t = O(\log(n/\varepsilon))$.

Note that convergence here is exponentially faster than in the single-vertex version of Theorem 1.2, reflecting the fact that this version is non-local and changes large portions of the configurations at each step. Again, the bound of Theorem 1.3 matches the lower bound for zero interaction [17].

We stress that the goal of this paper is *not* to design an efficient algorithm for sampling configurations of the Ising model. Such algorithms, based on standard linear Glauber dynamics, are already known throughout the high-temperature regime. Rather, our goal is to analyze the rate of convergence of a natural *nonlinear* dynamics, for the first time in a model with correlations. We view this as a first step towards a better understanding of such dynamics and the techniques needed to understand them; in addition to their inherent interest, these techniques may lead to algorithmic applications in future. Our work can be viewed as an extension of the successful application of a TCS lens in the analysis of mixing times of linear dynamics (Markov chains), which, as is well known, has seen both mathematical and algorithmic applications over many years. However, we point out that our convergence analysis in Theorems 1.2 and 1.3 actually does yield polynomial time sampling algorithms based on simulation of the respective nonlinear dynamics. We outline these algorithms, together with some associated open questions, in Section 5.

We also point out a further interesting algorithmic aspect of our results. Recall that our nonlinear processes sample from an Ising Gibbs measure $\mu_{\mathbf{J},\mathbf{h}}$, where the fields $\mathbf{h}$ are determined *implicitly* by the marginals at each site. It is these marginals (not the fields) that are specified by the initial distribution p_0. (As far as we are aware, all existing sampling algorithms for the Ising model require the specification of the fields $\mathbf{h}$ rather than the marginals.) Additionally, our processes can be used to learn the fields $\mathbf{h}$ corresponding to given marginals, an inverse problem that is also of independent interest (see [15] for a survey of such inverse problems): given samples from $\mu_{\mathbf{J},\mathbf{h}}$ produced by the nonlinear dynamics, standard methods can be used to infer the field vector $\mathbf{h}$.

1.1 Techniques

We begin by describing the earlier approach of [17] to analyzing the rate of convergence of the simpler population genetics dynamics, which corresponds to the trivial case of the Ising model with no interactions ($\mathbf{J} = 0$). Since the equilibrium distribution here consists of independent bits, the analysis is relatively straightforward given the following insight. The derivation of an individual σ at time t can be viewed as a binary tree going backwards in time, in which each individual inherits a random subset of its bits from each of its two parents according to the random crossover subset Λ. We may therefore follow the derivation of the n bits in σ back in time, until each of these bits is derived from a *distinct* individual at time 0. At that point we can deduce that the bits of σ are independently sampled from their respective marginal distributions, so σ is in equilibrium. The analysis therefore reduces to the question of how many steps are needed until all the n bits are separated, or "fragmented", under the repeated action of partition by the random crossover subset Λ, which in turn is a straightforward combinatorial calculation.

In the case where correlations are present in the equilibrium distribution, as in the Ising model, the above analysis breaks down because it is no longer sufficient to consider only fragmentation of the bits: indeed, the process must involve not only the breakdown of correlations in the initial distribution, but also, crucially, the

creation of the correct equilibrium correlations as mandated by the Gibbs distribution (1.3). Moreover, the process by which an individual inherits bits from its parents is no longer independent of the parents, but dictated by a complex function of both parents.

To account for this, we appeal to the *information percolation* framework developed by Lubetzky and Sly [13] in the context of Glauber dynamics for the Ising model. This framework suggests that we keep track of a "dependence cluster" going back in time, which records the neighboring spins that have influenced each spin in our current configuration. In the linear setting of [13], it can be shown (under a similar high-temperature assumption to ours) that this cluster is dominated by a subcritical branching process and thus will die out with large probability: the equilibrium correlations are then implicitly encoded by the history of this process. The time until the process dies out gives a bound on the mixing time.

In our nonlinear setting, the dependence clusters are no longer describable in terms of a simple branching process, but rather by a new type of process that combines branching with fragmentation, a process we refer to as "fragmentation plus noise." The first main ingredient of our analysis is the precise construction of such a process and the proof that it encodes the dependence structure of the nonlinear dynamics. The second main ingredient is the proof that, under the high-temperature assumption $\max_{x\in V}\sum_{y\in V}|J_{xy}| \le \delta_0$, the fragmentation plus noise process is subcritical and therefore dies out with large probability on a suitable time scale. To establish this latter fact, we introduce a non-standard form of "high-temperature expansion" for the dependence structure obtained by a coupling with non-uniform Erdős-Rényi random graphs. We refer the reader to Section 3 for a more technical high-level description of these ideas.

1.2 Related Work

Due to space constraints, we refer the reader to the full version of the paper [5] for a detailed account of related work, including more background on mass action kinetics, the Boltzmann equation, and other classes of nonlinear Monte Carlo processes.

1.3 Organization of the Paper

In Section 2 we formally define both of our nonlinear dynamics and establish some of their basic properties, including a proof of our convergence result, Theorem 1.1. In Section 3 we analyze the convergence rate for the nonlinear block dynamics, culminating in a proof of Theorem 1.3; prior to embarking on the details, we provide in Section 3.1 a more technical, high-level sketch of our approach. In Section 4 we apply a similar approach, though substantially different in detail, to analyze the nonlinear Glauber dynamics and prove Theorem 1.2. We conclude with some additional observations, extensions and open problems in Section 5. In light of space constraints, we defer many proofs and technical details to the full version [5].

2 PRELIMINARIES

2.1 The Ising Model

We recall from the introduction the definition (1.3) of the Ising model via its Gibbs distribution $\mu_{\mathbf{J},\mathbf{h}}$. Note that we allow arbitrary edge-dependent interactions $\mathbf{J} = \{J_{xy}\}_{x,y\in V}$ and arbitrary external fields $\mathbf{h} = \{h_x\}_{x\in V}$. When all the external fields are zero we write simply $\mu_{\mathbf{J}}$. We identify the set of vertices (or sites) V with $[n]$, and denote the set of spin configurations by $\Omega = \{\pm1\}^{[n]}$.

Remark 2.1. *For simplicity we have taken a model with no boundary conditions. However, there is no difficulty in extending our results to the case of a Gibbs measure with arbitrary boundary conditions, i.e., when the spins in some subset $V_0 \subseteq V$ are pinned to given values ±1. This generalization is achieved by taking limits $h_x \to \pm\infty$ for all $x \in V_0$ that are pinned to the values ±1, respectively.*

2.2 The Nonlinear Dynamics

Let $\mathcal{P}(\Omega)$ denote the set of probability measures on Ω. We define the nonlinear (mass action) dynamics as the dynamical system $p \mapsto T_t(p)$, where $T_0(p) = p \in \mathcal{P}(\Omega)$ is the initial distribution, $T_t(p) \in \mathcal{P}(\Omega)$ is the distribution after t steps, and $T_t(p) = T_{t-1}(p)\circ T_{t-1}(p)$. Here one step of the dynamics is defined, as in equation (1.1) of the introduction, by

$$p \mapsto p\circ p := \sum_{\sigma,\sigma',\tau'} p(\sigma)p(\sigma')Q(\sigma,\sigma';\tau,\tau'), \tag{2.1}$$

where, for each $\sigma,\sigma' \in \Omega$, $Q(\sigma,\sigma';\cdot,\cdot) \in \mathcal{P}(\Omega\times\Omega)$ is a probability measure satisfying the reversibility condition

$$\mu(\sigma)\mu(\sigma')Q(\sigma,\sigma';\tau,\tau') = \mu(\tau)\mu(\tau')Q(\tau,\tau';\sigma,\sigma'),\ \forall\,\sigma,\sigma',\tau,\tau', \tag{2.2}$$

for some $\mu \in \mathcal{P}(\Omega)$ with $\mu > 0$. It will be convenient later to write (2.1) in the equivalent form

$$p \mapsto p\circ p := \sum_{\sigma,\sigma'} p(\sigma)p(\sigma')Q(\cdot\,|\,\sigma,\sigma'), \tag{2.3}$$

where, for fixed $\sigma,\sigma' \in \Omega$, the distribution $Q(\cdot\,|\,\sigma,\sigma') \in \mathcal{P}(\Omega)$ is defined by

$$Q(\tau\,|\,\sigma,\sigma') := \sum_{\tau'\in\Omega} Q(\sigma,\sigma';\tau,\tau'). \tag{2.4}$$

In this paper we take $\mu = \mu_{\mathbf{J},\mathbf{h}}$ as the Ising measure (1.3) and consider two natural choices of the kernel Q that satisfy (2.2), which we now describe.

2.2.1 Nonlinear Block Dynamics. The first model, which we refer to as the *nonlinear block dynamics*, corresponds to interactions in which a pair of configurations (σ,σ') exchange their spins at an arbitrary, randomly chosen subset $\Lambda \subseteq [n]$ of sites, i.e.,

$$(\sigma,\sigma') \mapsto (\sigma_\Lambda\sigma'_{\Lambda^c},\sigma'_\Lambda\sigma_{\Lambda^c}), \tag{2.5}$$

where $\sigma_\Lambda\sigma'_{\Lambda^c}$ denotes the element of Ω with entries σ_x for $x \in \Lambda$ and σ'_x for $x \in \Lambda^c = [n]\setminus\Lambda$. Here, to ensure reversibility, the set Λ is chosen with probability proportional to $\mu(\sigma_\Lambda\sigma'_{\Lambda^c})\mu(\sigma'_\Lambda\sigma_{\Lambda^c})$. Thus the associated kernel is defined as

$$Q_{\mathbf{J}}(\sigma,\sigma';\tau,\tau') = \frac{\sum_{\Lambda\subseteq V}\mu(\sigma_\Lambda\sigma'_{\Lambda^c})\mu(\sigma'_\Lambda\sigma_{\Lambda^c})\mathbf{1}_{\tau=\sigma_\Lambda\sigma'_{\Lambda^c}}\mathbf{1}_{\tau'=\sigma'_\Lambda\sigma_{\Lambda^c}}}{\sum_{A\subseteq V}\mu(\sigma_A\sigma'_{A^c})\mu(\sigma'_A\sigma_{A^c})}. \tag{2.6}$$

Note that transitions of the form (2.5) can only produce pairs (τ,τ') that belong to the equivalence class

$$C(\sigma,\sigma') = \left\{(\sigma_\Lambda\sigma'_{\Lambda^c},\sigma'_\Lambda\sigma_{\Lambda^c}),\ \Lambda\subseteq V\right\}.$$

Thus the kernel (2.6) defines a (linear) Markov chain on the pair space $\Omega\times\Omega$ that is in general not irreducible, and whose communicating classes are precisely $C(\sigma,\sigma')$. Note also that the kernel $Q_{\mathbf{J}}$

depends on $\mu = \mu_{\mathbf{J},\mathbf{h}}$ only through the interaction $\mathbf{J}$ and is insensitive to the choice of fields $\mathbf{h}$. Indeed, once σ, σ' are given, then for any $(\tau, \tau') \in C(\sigma, \sigma')$ and $(\eta, \eta') \in C(\sigma, \sigma')$ one has

$$\frac{\mu_{\mathbf{J},\mathbf{h}}(\tau)\mu_{\mathbf{J},\mathbf{h}}(\tau')}{\mu_{\mathbf{J},\mathbf{h}}(\eta)\mu_{\mathbf{J},\mathbf{h}}(\eta')} = \frac{\mu_{\mathbf{J},\mathbf{h}'}(\tau)\mu_{\mathbf{J},\mathbf{h}'}(\tau')}{\mu_{\mathbf{J},\mathbf{h}'}(\eta)\mu_{\mathbf{J},\mathbf{h}'}(\eta')}, \qquad \forall\, \mathbf{h}, \mathbf{h}' \in \mathbb{R}^n.$$

Thus, w.l.o.g., we may take $\mathbf{h} = 0$, and $\mu = \mu_{\mathbf{J}} = \mu_{\mathbf{J},\mathbf{0}}$, in the definition of the kernel (2.6). The kernel $Q_{\mathbf{J}}$ in (2.6) is an example of a so-called "folding" transformation [21].

Observe that, for all $\mathbf{h} \in \mathbb{R}^n$, the reversibility condition (2.2) holds in the form

$$\mu_{\mathbf{J},\mathbf{h}}(\sigma)\mu_{\mathbf{J},\mathbf{h}}(\sigma')Q_{\mathbf{J}}(\sigma, \sigma'\,;\, \tau, \tau') = \mu_{\mathbf{J},\mathbf{h}}(\tau)\mu_{\mathbf{J},\mathbf{h}}(\tau')Q_{\mathbf{J}}(\tau, \tau'\,;\, \sigma, \sigma'), \tag{2.7}$$

for all $\sigma, \sigma', \tau, \tau' \in \Omega$. Thus, for a fixed interaction $\mathbf{J}$, the kernel $Q_{\mathbf{J}}$ is reversible w.r.t. all measures $\{\mu_{\mathbf{J},\mathbf{h}}, \mathbf{h} \in \mathbb{R}^n\}$. In particular, all these measures are stationary for the dynamics (2.1), i.e.,

$$\mu_{\mathbf{J},\mathbf{h}} \circ \mu_{\mathbf{J},\mathbf{h}} = \mu_{\mathbf{J},\mathbf{h}}, \qquad \forall\, \mathbf{h} \in \mathbb{R}^n, \tag{2.8}$$

as can be easily checked from reversibility.

We note that in the case $\mathbf{J} = 0$, the nonlinear block dynamics reduces to the uniform crossover model from population genetics [3, 17]. In this case, the stationary distributions $\mu_{\mathbf{J},\mathbf{h}}$ are just product measures over spins with marginals determined by $\mathbf{h}$.

2.2.2 Nonlinear Glauber Dynamics. In our second model, the configurations σ, σ' exchange their spins at a *single* randomly chosen site $x \in [n]$, i.e.,

$$(\sigma, \sigma') \mapsto (\sigma_x \sigma'_{[n]\setminus\{x\}}, \sigma'_x \sigma_{[n]\setminus\{x\}}).$$

By analogy with the familiar Glauber dynamics (a Markov chain that updates the spin at one site in each step), we refer to this as the *nonlinear Glauber dynamics*. As usual, to ensure reversibility w.r.t. $\mu_{\mathbf{J},\mathbf{h}}$, we need to perform such an exchange with an appropriate probability $\alpha_x(\sigma, \sigma')$. Specifically, we use the generic dynamics (2.1) with the kernel

$$Q_{\mathbf{J}}(\sigma, \sigma'\,;\, \tau, \tau') = \frac{1}{n}\sum_{x \in V} Q_{\mathbf{J},x}(\sigma, \sigma'\,;\, \tau, \tau'), \tag{2.9}$$

where

$$\begin{aligned} Q_{\mathbf{J},x}(\sigma, \sigma'\,;\, \tau, \tau') \;:=\; & \alpha_x(\sigma, \sigma')\mathbf{1}_{\tau = \sigma_x \sigma'_{[n]\setminus\{x\}}}\mathbf{1}_{\tau' = \sigma'_x \sigma_{[n]\setminus\{x\}}} \\ & + (1 - \alpha_x(\sigma, \sigma'))\mathbf{1}_{\tau=\sigma}\mathbf{1}_{\tau'=\sigma'}, \end{aligned}$$

and

$$\alpha_x(\sigma, \sigma') := \frac{\mu(\sigma_x \sigma'_{[n]\setminus\{x\}})\mu(\sigma'_x \sigma_{[n]\setminus\{x\}})}{\mu(\sigma_x \sigma'_{[n]\setminus\{x\}})\mu(\sigma'_x \sigma_{[n]\setminus\{x\}}) + \mu(\sigma)\mu(\sigma')}.$$

Once again, the Markov chain on pairs $\Omega \times \Omega$ defined by the kernel (2.9) is not irreducible, and the kernel $Q_{\mathbf{J}}$ depends on $\mu = \mu_{\mathbf{J},\mathbf{h}}$ only through the interaction $\mathbf{J}$. As in (2.7)-(2.8), reversibility and stationarity of all measures $\mu_{\mathbf{J},\mathbf{h}}$ can be easily checked.

2.3 Conservation Laws

In both dynamics defined above, the map $p \mapsto p \circ p$ conserves the marginal probabilities of spins at every vertex, i.e., for every $x \in [n]$, and for any $p \in \mathcal{P}(\Omega)$, one has

$$(p \circ p)_x = p_x, \tag{2.10}$$

where $p_x(a) := p(\sigma_x = a)$, $a \in \{-1, 1\}$, denotes the marginal of p at x. It is convenient to state the following stronger property. Let us define the commutative convolution product of two distributions $p, q \in \mathcal{P}(\Omega)$ by

$$p \circ q := \frac{1}{2}\sum_{\sigma,\sigma'} \big(p(\sigma)q(\sigma') + p(\sigma')q(\sigma)\big)Q(\cdot \mid \sigma, \sigma'), \tag{2.11}$$

where Q is defined by (2.4) and (2.6) for the nonlinear block dynamics and by (2.4) and (2.9) for the nonlinear Glauber dynamics, respectively. Note that the notation (2.11) is consistent with (2.3).

Lemma 2.2. *Both of the above dynamics satisfy*

$$(p \circ q)_x = \tfrac{1}{2}\,(p_x + q_x).$$

In particular, the conservation law (2.10) *holds.*

The proof, which is deferred to the full version, follows easily from the fact that, in both dynamics, spins are simply exchanged between σ, σ', as well as from the symmetry of $Q_{\mathbf{J}}$ w.r.t. σ, σ'.

2.4 The Derivation Tree and Fragmentation

Throughout the paper, the following view of the nonlinear dynamics will be central. By definition, $T_t(p)$ is the result of repeated pairwise interactions and can be represented as the distribution at the root of a binary "derivation" tree, where each leaf is equipped with the distribution $p \in \mathcal{P}(\Omega)$, and recursively, starting from the leaves, each internal node is assigned the distribution $p_1 \circ p_2$ where p_1, p_2 represent the distributions assigned to the left and right descendants of that node; see Figure 1 for a schematic picture of the case $t = 2$.

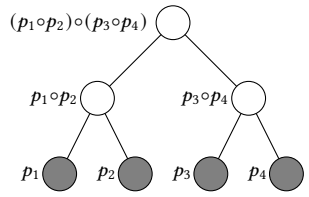

Figure 1: Graphical representation of the distribution $(p_1 \circ p_2) \circ (p_3 \circ p_4)$ at time $t = 2$ when each leaf $i = 1, \ldots, 4$ is equipped with distribution p_i. When $p_i = p$ $\forall i$, the distribution at the root is $T_2(p)$.

We focus now on the simple case $\mathbf{J} = 0$, i.e., no correlations between spins. Under block dynamics, the configuration τ at the root of the tree (at time t) is constructed according to the random partition (Λ, Λ^c) of V, which is equivalent to drawing each spin τ_x from the configuration at the left or right child node with probability $\frac{1}{2}$, independently for each site $x \in V$. Continuing down the tree in the same fashion, we see that each spin at the root is drawn from one of the 2^t leaves (at time 0), independently and uniformly at random. Thus this process induces a *partition* of the sites V into 2^t disjoint subsets (some of which may be empty), where the ℓth subset consists of those sites that draw their spin from the configuration at leaf ℓ.

Now let $\mathcal{A}$ denote the event that none of the 2^t subsets in this partition contains more than one site; equivalently, each spin in τ is drawn from a *distinct* leaf. We call $\mathcal{A}$ the "complete fragmentation"

event. Note that, conditional on $\mathcal{A}$, the distribution of the configuration τ at the root is just the product $\pi := \otimes_{x\in V} p_x$, since there are no remaining correlations between spins. Hence we may write

$$T_t(p) = \nu(\mathcal{A})\,\pi + \nu(\mathcal{A}^c)\,q(t), \tag{2.12}$$

for some other distribution $q(t)$, where ν denotes the uniform distribution over all $2^t - 1$ independent random subsets Λ occurring in the tree.

We can use (2.12) to obtain an upper bound on the convergence time for the nonlinear block dynamics when $\mathbf{J} = 0$, as was done in [17]. First, we claim that $\nu(\mathcal{A}^c) \le \binom{n}{2} 2^{-t}$. To see this, note that for any given pair of distinct sites $x, y \in V$, the probability that x, y are not separated after t levels of the successive partitioning process is 2^{-t}, and then take a union bound over pairs. Hence by (2.12), taking $t = O(\log(n/\varepsilon))$ ensures that $\|\overline{T}_t(p) - \pi\|_{\mathrm{TV}} \le \varepsilon$, so the convergence time is $O(\log(n/\varepsilon))$.

For the nonlinear Glauber dynamics with $\mathbf{J} = 0$ a similar analysis applies, except that now each node in the tree chooses one random spin from the left child and the remainder from the right child. The complete fragmentation event $\mathcal{A}$ now corresponds to isolating each of the n bits in this process, which is just a coupon-collecting event for $n = |V|$ coupons. Thus we have $\nu(\mathcal{A}^c) \le n(1 - \frac{1}{n})^t \le ne^{-t/n}$ (where now ν denotes the uniform distribution over all $2^t - 1$ independent choices of random spins occurring in the tree), which by (2.12) implies a convergence time of $O(n \log(n/\varepsilon))$.

When $\mathbf{J} \neq 0$, so that correlations are present, it is no longer possible to reduce the analysis of convergence to the fragmentation event, because in the stationary distribution spins are not independent. Thus to prove Theorems 1.2 and 1.3 we will need to augment the simple derivation process above to obtain a more complex process that we call "fragmentation with noise" (see Sections 3 and 4).

2.5 Convergence to Stationarity

In this subsection we prove convergence of both nonlinear dynamics to the stationary distribution $\mu_{\mathbf{J},\mathbf{h}}$, as claimed in Theorem 1.1.

As we have seen, reversibility implies that, for a fixed interaction matrix $\mathbf{J}$, the Ising measures $\mu_{\mathbf{J},\mathbf{h}}$ defined in (1.3) are all stationary, regardless of the choice of $\mathbf{h}$. In fact, these are the *only* stationary distributions, as proved in [4]:

Lemma 2.3. *[4, Lemma 3.2] For both the above dynamics, for any fixed interaction matrix $\mathbf{J}$, a distribution $\mu \in \mathcal{P}(\Omega)$ is stationary for (2.3) if and only if μ has the form (1.3) for some choice of the fields $\mathbf{h}$.*

We now state a general convergence criterion for reversible mass action dynamics, based on the notion of irreducibility.

Definition 2.4. *We say that an initial distribution $p \in \mathcal{P}(\Omega)$ is* irreducible *for a given mass action system with kernel Q if there exist $\varepsilon > 0$ and t_0 such that $T_t(p)(\tau) \ge \varepsilon$ for all $t \ge t_0$ and all $\tau \in \Omega$.*

Thus irreducibility says that the trajectory of the dynamics starting from p eventually remains bounded away from the boundary of the simplex.

Theorem 2.5. *Suppose the kernel Q is reversible w.r.t. some distribution $\mu > 0$ and has non-zero diagonals (i.e., $Q(\sigma, \sigma'; \sigma, \sigma') > 0$ for all σ, σ'). Then for any initial distribution p that is irreducible, the general nonlinear dynamics defined in (1.1) satisfies $T_t(p) \to \nu$ as $t \to \infty$, where $\nu > 0$ is stationary.*

A continuous time version of this convergence criterion is already known in the reaction networks literature [19] for the wider class of "complex balanced" mass action systems. In the full version of this paper [5], we present a self-contained proof in discrete time for complex balanced systems. The key idea in the proof is to show that the relative entropy $\mathcal{D}(T_t(p)\|\nu)$ w.r.t. any stationary $\nu > 0$ is monotonically strictly decreasing unless p is stationary. Moreover, by standard compactness and continuity arguments, any trajectory has a convergent subsequence whose limit ν is stationary; and, by irreducibility, it must be the case that $\nu > 0$. Finally, we can deduce that $\mathcal{D}(T_t(p)\|\nu) \to \mathcal{D}(\nu\|\nu) = 0$, giving the required convergence.

In light of Theorem 2.5, to prove convergence it suffices to prove that our particular mass action systems are irreducible.

We say that a measure $p \in \mathcal{P}(\Omega)$ has *nondegenerate marginals* if there exists $\delta > 0$ such that

$$\min_{x\in V} \min_{a\in\{-1,+1\}} p_x(a) \ge \delta > 0. \tag{2.13}$$

In what follows we shall assume that the initial distribution p satisfies this condition. This is actually no loss of generality since one can restrict to the nondegenerate spins and consider the degenerate spins as a fixed boundary condition, or pinning; see Remark 2.1.

Lemma 2.6. *For any Ising interaction matrix $\mathbf{J}$, any initial distribution $p \in \mathcal{P}(\Omega)$ with nondegenerate marginals is irreducible for both the nonlinear block dynamics and the nonlinear Glauber dynamics.*

We defer the proof of this lemma to the full version. The main idea of the proof is that every interaction can be decomposed as a convex combination of a pure product kernel (in which spins are exchanged between input configurations σ, σ' independently, i.e., according to the $\mathbf{J} = 0$ dynamics) and some other more complex kernel. Crucially, by virtue of the fragmentation analysis in (2.12), and using (2.13), the product kernel eventually gives rise to a uniformly positive distribution, and this property is preserved over all future times.

We note that Theorem 1.1 provides no quantitative estimate at all on the *rate* of convergence to stationarity. In particular, there is no explicit dependence on the size of the system n. In analogy with the mixing time analysis for linear Markov chains, in the remainder of the paper we will study the rate of convergence to equilibrium under the assumption that the interactions in $\mathbf{J}$ are sufficiently weak (usually referred to as the "high temperature" regime).

3 THE NONLINEAR BLOCK DYNAMICS

Let $T_t(p)$, $t \in \mathbb{N}$, denote the evolution of the initial distribution $p \in \mathcal{P}(\Omega)$ under the nonlinear block dynamics (2.6). From Theorem 1.1 we know that for any fixed interaction matrix $\mathbf{J}$, and any $p \in \mathcal{P}(\Omega)$, one has the convergence $T_t(p) \to \mu_{\mathbf{J},\mathbf{h}}$ as $t \to \infty$, where $\mathbf{h}$ is the unique vector of external fields such that $\mu_{\mathbf{J},\mathbf{h}}$ and the initial state p have the same marginals at x, for all $x \in V$. Our main result for the nonlinear block dynamics (Theorem 1.3 in the introduction) establishes a tight bound on the rate of convergence as a function of the cardinality $n = |V|$, under the Dobrushin-type high-temperature condition on the interaction matrix $\mathbf{J}$, namely

$$\max_x \sum_{y\in V} |J_{xy}| \le \delta_0\,, \tag{3.1}$$

for an absolute constant $\delta_0 > 0$.

3.1 Main Ideas of the Proof

Before embarking on the details of the proof, we give a high level description of the main steps. By symmetry we may rewrite the operator (2.11) in the form

$$(p \circ q)(\tau) = \sum_{\sigma,\sigma'} p(\sigma) q(\sigma') \sum_{\Lambda \subseteq V} \gamma(\Lambda \,|\, \sigma, \sigma') \mathbf{1}_{\tau = \sigma_\Lambda \sigma'_{\Lambda^c}},$$

where

$$\gamma(\Lambda \,|\, \sigma, \sigma') = \frac{\mu(\sigma_\Lambda \sigma'_{\Lambda^c}) \mu(\sigma'_\Lambda \sigma_{\Lambda^c})}{\sum_{A \subseteq V} \mu(\sigma_A \sigma'_{A^c}) \mu(\sigma'_A \sigma_{A^c})}. \tag{3.2}$$

Thus, for each $\sigma, \sigma' \in \Omega$, $\gamma(\cdot \,|\, \sigma, \sigma')$ is a probability measure over subsets $\Lambda \subseteq V$. It will be convenient to view the distribution $\gamma(\cdot \,|\, \sigma, \sigma')$ as a spin system, i.e., a probability measure over spin configurations $\eta \in \{-1, +1\}^n$, by identifying $\eta_x = +1$ with $x \in \Lambda$ and $\eta_x = -1$ with $x \notin \Lambda$. Recall that in the non-interacting case $\mathbf{J} = 0$, the distribution $\gamma(\cdot \,|\, \sigma, \sigma')$ does not depend on the pair (σ, σ'), and is simply the product of Bernoulli measures with parameter $1/2$. As described in Section 2.4, the dynamics is then entirely governed by the fragmentation process that starts with the set V and recursively splits sets of vertices uniformly at random until it reaches a collection of singletons. The simple argument given in that section then gives a convergence time of $O(\log(n/\varepsilon))$; see [3, 17] for a detailed analysis of the non-interacting case.

When there is a nontrivial interaction $\mathbf{J} \neq 0$, this straightforward analysis breaks down. Our proof of Theorem 1.3 is based on a coupling argument that allows us to reduce the problem to the analysis of a more general process in which the fragmentation mechanism is perturbed by a "local growth" process arising from the correlations inherent in the interactions. The main idea is that if the local growth is sufficiently sparse, then the underlying fragmentation dominates and eventually the memory of the initial distribution (except for the marginals) is lost.

The first step in the proof is to couple the above random variable η with distribution $\gamma(\cdot \,|\, \sigma, \sigma')$ with a random subgraph G of the complete graph K_n having a suitable distribution ν, i.e., we shall write

$$\gamma(\cdot \,|\, \sigma, \sigma') = \sum_G \nu(G) \, \gamma_G(\cdot \,|\, \sigma, \sigma'), \tag{3.3}$$

where the sum extends over all possible subgraphs $G \subseteq K_n$, and $\gamma_G(\cdot \,|\, \sigma, \sigma')$ is a probability measure on Ω for each realization G. The key features of this coupling are:

- the distribution ν does not depend on the pair (σ, σ');
- the distribution $\gamma_G(\cdot \,|\, \sigma, \sigma')$ depends on the pair (σ, σ') only through the spins
$$\sigma_{V_G} = \{\sigma_x, \ x \in V_G\}, \qquad \sigma'_{V_G} = \{\sigma'_x, \ x \in V_G\},$$
where V_G denotes the vertex set of G; and
- under $\gamma_G(\cdot \,|\, \sigma, \sigma')$, the random variables $\{\eta_y, \ y \in V \setminus V_G\}$ are i.i.d. Bernoulli with parameter $1/2$.

Actually, it will be crucial that ν can be taken to be the inhomogeneous Erdős-Rényi random graph with edge weights proportional to

$$\lambda_{xy} := e^{4|J_{xy}|} - 1.$$

This ensures that, under the assumption (3.1), the graph G will be sufficiently sparse and the size of the connected components will satisfy good tail bounds. Note that the expression (3.3) can be seen as a form of *high-temperature expansion* [8] for the measure $\gamma(\cdot \,|\, \sigma, \sigma')$. However, a standard high-temperature expansion would produce an expression of the form (3.3) with real-valued coefficients $\nu(G)$ which depend on (σ, σ'), while it is crucial for our coupling argument that ν be a *probability measure independent of* (σ, σ').

Armed with the coupling (3.3), we consider all $2^t - 1$ interactions in the derivation tree of Section 2.4 that produce the final distribution $T_t(p)$. For each interaction we use a realization of the graph G and we specify a realization B of the Bernoulli random variables with parameter $1/2$ which determine η_y for $y \in V \setminus V_G$. We then compute the resulting distribution. Letting $(\vec{G}, \vec{B}) = (G_1, B_1), \ldots, (G_{2^t-1}, B_{2^t-1})$ denote the vector of all such realizations, we may then write

$$T_t(p) = \sum_{(\vec{G}, \vec{B})} \widehat{\nu}(\vec{G}, \vec{B}) \, T_t(p \,|\, \vec{G}, \vec{B}), \tag{3.4}$$

where $\widehat{\nu}$ is a suitable distribution over the realizations $(\vec{G}, \vec{B})$ and $T_t(p \,|\, \vec{G}, \vec{B}) \in \mathcal{P}(\Omega)$ represents the distribution at time t conditional on the realizations $(\vec{G}, \vec{B})$. The important point here is that $\widehat{\nu}$ is independent of the initial conditions, and therefore all correlations in the initial distribution appear only in the measures $T_t(p \,|\, \vec{G}, \vec{B})$. Moreover, the measure $\widehat{\nu}$ can naturally be interpreted as a stochastic process that combines fragmentation with local growth.

The second main ingredient in the proof of Theorem 1.3 is the identification of an event $\mathcal{E}_t$ for this process, roughly representing the fact that within time t all fragments have reached their minimum size, and such that

$$\widehat{\nu}(\mathcal{E}_t) \geq 1 - A n^2 e^{-b t} \tag{3.5}$$

for some absolute constants $A, b > 0$. The nature of the event $\mathcal{E}_t$ will be such that

$$T_t(p \,|\, \vec{G}, \vec{B}) = T_t(p' \,|\, \vec{G}, \vec{B}), \qquad (\vec{G}, \vec{B}) \in \mathcal{E}_t \tag{3.6}$$

for all $p, p' \in \mathcal{P}(\Omega)$ which have the same marginals at every vertex $x \in V$. Once these facts are established, (3.4), (3.5) and (3.6) imply that for any such $p, p' \in \mathcal{P}(\Omega)$ one has

$$\|T_t(p) - T_t(p')\|_{\mathrm{TV}} \leq A n^2 e^{-b t}.$$

This implies the result of Theorem 1.3 by taking $p' = \mu_{\mathbf{J},\mathbf{h}}$.

We now turn to proofs of the various claims sketched above.

3.2 Coupling with Inhomogeneous Erdős-Rényi Random Graphs

We start by observing that for every fixed $\sigma, \sigma' \in \Omega$, there is an interaction matrix $\tilde{\mathbf{J}} = \tilde{\mathbf{J}}(\sigma, \sigma')$ such that the set of spins to be exchanged, $\gamma(\cdot \,|\, \sigma, \sigma')$ from (3.2), is itself an Ising Gibbs measure $\mu_{\tilde{\mathbf{J}}}$ as defined in (1.3).

Lemma 3.1. *For any $\sigma, \sigma' \in \Omega$, we have $\gamma(\cdot \,|\, \sigma, \sigma') = \mu_{\tilde{\mathbf{J}}}$, i.e.,*

$$\gamma(\eta \,|\, \sigma, \sigma') \propto \exp\Big\{\frac{1}{2} \sum_{x,y \in V} \tilde{J}_{xy} \eta_x \eta_y\Big\}, \qquad \eta \in \Omega, \tag{3.7}$$

where the interaction matrix $\tilde{\mathbf{J}} = \tilde{\mathbf{J}}(\sigma, \sigma')$ is given by

$$\tilde{J}_{xy} := 2 J_{xy} \sigma_x \sigma_y \mathbf{1}_{x \in D(\sigma,\sigma')} \mathbf{1}_{y \in D(\sigma,\sigma')}$$

and $D(\sigma, \sigma') := \{z \in V : \sigma_z \neq \sigma'_z\}$.

The proof is an elementary calculation and is deferred.

It follows from Lemma 3.1 that for any fixed pair $\sigma, \sigma' \in \Omega$, if $\eta \in \Omega$ is distributed according to $\gamma(\cdot \mid \sigma, \sigma')$, then $\{\eta_x, x \in V \setminus D(\sigma, \sigma')\}$ is the Bernoulli measure with parameter $1/2$ and, independently, $\{\sigma_x \eta_x, x \in D(\sigma, \sigma')\}$ is the Ising measure on $D(\sigma, \sigma')$ with zero external fields and interaction $2\mathbf{J}$.

For our purposes, the problem with the representation of $\gamma(\cdot \mid \sigma, \sigma')$ in Lemma 3.1 is that it is structurally highly dependent on the configurations σ, σ' through the set $D(\sigma, \sigma')$. Our goal in this subsection is to formulate an alternative representation, in Lemma 3.2 below, that overcomes this problem.

Let $\mathcal{G}$ be the set of all subgraphs of the complete graph K_n over $V \subseteq [n]$ with isolated vertices removed, and write $\mathcal{P}(\mathcal{G})$ for the set of probability measures over $\mathcal{G}$. Thus $G \in \mathcal{G}$ can be viewed as a collection of unordered pairs $\{x, y\}$ for $x, y \in V$. Note that $G \in \mathcal{G}$ need not be connected and can be the empty graph (with no vertices). We write V_G, E_G for the vertex and edge set of $G \in \mathcal{G}$, respectively.

Lemma 3.2. *Let $\nu_{\mathbf{J}}$ be the inhomogeneous Erdős-Rényi measure associated with the weights $\lambda_{xy} = e^{4|J_{xy}|} - 1$, i.e.,*

$$\nu_{\mathbf{J}}(G) \propto \prod_{\{x,y\} \in E_G} (e^{4|J_{xy}|} - 1)\,. \tag{3.8}$$

Then

$$\gamma(\cdot \mid \sigma, \sigma') = \sum_{G \in \mathcal{G}} \nu_{\mathbf{J}}(G)\, \mu_G(\cdot \mid \sigma_{V_G}, \sigma'_{V_G}) \otimes \mathrm{Be}_{V \setminus V_G}(\tfrac{1}{2})\,,$$

where, for any $G \in \mathcal{G}$, $\mu_G(\cdot \mid \sigma_{V_G}, \sigma'_{V_G})$ is a probability measure on $\{-1,+1\}^{V_G}$ that depends on σ, σ' only through the spins $\sigma_{V_G}, \sigma'_{V_G}$ and $\mathrm{Be}_{V \setminus V_G}(\frac{1}{2})$ is the Bernoulli probability measure on $\{-1,+1\}^{V \setminus V_G}$, which assigns independently the values ± 1 with probability $1/2$ to each $x \in V \setminus V_G$. Moreover, the probability measure $\mu_G(\cdot \mid \sigma_{V_G}, \sigma'_{V_G})$ has the product structure

$$\mu_G(\cdot \mid \sigma_{V_G}, \sigma'_{V_G}) = \otimes_{i=1}^{k} \mu_{G_i}(\cdot \mid \sigma_{V_{G_i}}, \sigma'_{V_{G_i}})\,,$$

where $G_1, \dots, G_k$ are the connected components of G.

We defer the rather technical proof of this lemma to the full version [5].

So far we have not used the weak interaction assumption (3.1), so Lemma 3.2 holds for arbitrary coefficients J_{xy}. Next, we observe that the condition (3.1) implies a strong sparsity property of the measure $\nu_{\mathbf{J}}$ in Lemma 3.2. From the definition (3.8), this measure is the inhomogeneous Erdős-Rényi random graph where each edge $e = \{x, y\}$, $x, y \in V$, is included independently with probability

$$p_{xy} = 1 - e^{-4|J_{xy}|}. \tag{3.9}$$

In particular, using $(e^z - 1)e^{-z} \le z$ for $z \ge 0$, one has $p_{xy} \le 4|J_{xy}|$, and if (3.1) holds, then for any $x \in V$,

$$\nu_{\mathbf{J}}(V_G \ni x) \le \sum_{y:\, y \neq x} p_{xy} \le 4\delta_0.$$

Moreover, if δ_0 is sufficiently small in (3.1), then for any given $x \in V$ the size of the connected component of G at x has an exponential tail. We make this precise in Lemma 3.6 below.

3.3 Fragmentation with Noise

We now develop the main construction behind the convergence result in Theorem 1.3. It is based on a *perturbed* fragmentation process, i.e., a process that combines the random fragmentation of the non-interacting case (as described in Section 3.1) with some competing noise represented by the random graphs encoding dependencies.

Given a set $A \subseteq [n]$ and a random graph $G \in \mathcal{G}$ with distribution $\nu_{\mathbf{J}}$ as in Lemma 3.2, we define the random set A' as the vertex set of the union of all connected components of G that have non-empty intersection with A. More formally, write $G = \cup_{i=1}^{\ell} G_i$ where G_i are the connected components of G and let

$$G(A) = \bigcup_{i:\, V_{G_i} \cap A \neq \emptyset} G_i.$$

Then we set $A' = V_{G(A)}$. We may sample A' starting from A by a breadth-first search, i.e., by revealing sequentially for each $x \in A$ the neighborhood of x in G, then recursively the neighborhood of each vertex revealed at the previous step, and so on until there are no more neighbors to reveal. Clearly, A' may contain sites that are not in A. However, if $x \in A \cap V_G^c$ then $x \notin A'$. If $A \setminus A' \neq \emptyset$, for every $x \in A \setminus A'$, we independently declare x to be *in* or *out* by a fair coin flip. We thus obtain two random sets A_{in} and A_{out}, such that

$$A_{\mathrm{in}} \cup A_{\mathrm{out}} = A \setminus A'\,, \qquad A_{\mathrm{in}} \cap A_{\mathrm{out}} = \emptyset.$$

Next, for any $A \subseteq [n]$, we define two sets $\Phi_0(A), \Phi_1(A)$ by

$$\Phi_0(A) = \Phi_1(A) = \emptyset\,, \qquad \text{if } |A| \le 1\,, \tag{3.10}$$

and

$$\Phi_0(A) = A' \cup A_{\mathrm{in}}\,, \quad \Phi_1(A) = A' \cup A_{\mathrm{out}}\,, \quad \text{if } |A| \ge 2\,. \tag{3.11}$$

Definition 3.3. *The* fragmentation plus noise process *$\mathcal{F}_t$, $t = 0, 1, \dots$ is the random process defined as follows. For each $t \in \mathbb{N}$, $\mathcal{F}_t$ consists of 2^t labeled fragments, i.e., (possibly empty) subsets $F_1^{(t)}, \dots, F_{2^t}^{(t)}$, $F_i^{(t)} \subseteq [n]$, obtained by repeated application of the following rule. At time zero we have $\mathcal{F}_0 = [n]$, i.e., $F_1^{(0)} = [n]$. At time $t \in \mathbb{N}$, if $\mathcal{F}_{t-1} = (F_1^{(t-1)}, \dots, F_{2^{t-1}}^{(t-1)})$, then for each i independently, we replace $F_i^{(t-1)}$ by $(\Phi_0(F_i^{(t-1)}), \Phi_1(F_i^{(t-1)}))$ where Φ_0, Φ_1 are the random maps defined by (3.10)-(3.11), so that*

$$\begin{aligned}\mathcal{F}_t &= (F_1^{(t)}, \dots, F_{2^t}^{(t)}) \\ &= (\Phi_0(F_1^{(t-1)}), \Phi_1(F_1^{(t-1)}), \dots, \Phi_0(F_{2^{t-1}}^{(t-1)}), \Phi_1(F_{2^{t-1}}^{(t-1)}))\,.\end{aligned}$$

We say that the process dies out *if there is a time t such that all fragments are empty, i.e., $F_i^{(t)} = \emptyset$ for all $i = 1, \dots, 2^t$. With slight abuse of notation, we write $\mathcal{F}_t = \emptyset$ for the latter event.*

Remark 3.4. *In the non-interacting case $\mathbf{J} = 0$ one has $p_{xy} = 0$ for all $\{x, y\}$, and thus $A' = \emptyset$, $\Phi_0(A) = A_{\mathrm{in}}, \Phi_1(A) = A_{\mathrm{out}}$ and $A \mapsto (\Phi_0(A), \Phi_1(A))$ is one step of a pure fragmentation process, where the set A is partitioned into two subsets using independent fair coin flips for each vertex. In this case $F_i^{(t)} \cap F_j^{(t)} = \emptyset$ for all $i, j = 1, \dots, 2^t$ and for all $t \in \mathbb{N}$. In particular, it is not hard to see that in this case*

$$\mathbb{P}(\mathcal{F}_t \neq \emptyset) \le n(n-1)\, 2^{-t}\,, \qquad t = 1, 2, \dots \tag{3.12}$$

Indeed, by construction $\mathcal{F}_t \neq \emptyset$ implies that there exist two vertices $x, y \in [n]$ that belong to the same fragment up to time $t-1$. Thus (3.12) follows as in the analysis in Section 2.4. In the interacting case, there is a first stage where the set A grows according to the local branching at every $x \in A$, and the fragmentation occurs only on those vertices that have an empty neighborhood in G. Our main technical result below establishes that the fragmentation plus noise process also satisfies a bound of the form (3.12), with a slightly weaker exponential decay rate, provided (3.1) holds for a suitably small $\delta_0 > 0$.

Lemma 3.5. *For any $\delta \in (0,1)$, there exists $\delta_0 > 0$, and a constant $C_\delta > 0$ such that if (3.1) holds with constant δ_0 then*

$$\mathbb{P}(\mathcal{F}_t \neq \emptyset) \leq n^2 C_\delta \, (2-\delta)^{-t}, \qquad t = 1, 2, \dots$$

Proof. The event $\mathcal{F}_t \neq \emptyset$ implies that there exists a fragment $F_i^{(t-1)}$ at time $t-1$ that has cardinality $|F_i^{(t-1)}| \geq 2$. Indeed, by construction, if all fragments have size at most 1 at time $t-1$, then $\mathcal{F}_t = \emptyset$; see (3.10). Since there are 2^{t-1} fragments at time $t-1$, and since all the $F_i^{(t-1)}$ have the same distribution, by a union bound it suffices to show that for any $\delta \in (0,1)$,

$$\mathbb{P}(|F_1^{(t-1)}| \geq 2) \leq n^2 C_\delta \, (4-\delta)^{-t}, \qquad t = 1, 2, \dots, \quad (3.13)$$

provided (3.1) holds with a sufficiently small $\delta_0 > 0$. To prove (3.13) we shall use a stochastic domination argument that bounds the evolution of the fragment $F_1^{(t-1)}$, $t \geq 1$ by means of independent *labeled branching processes*.

Consider n independent processes $X^y := \{X_\ell^y, \ \ell = 0, 1, \dots\}$, $y \in [n]$, such that for each $y \in [n]$, X^y is the labeled branching process with $X_0^y = \{y\}$ and such that, at time $t \in \mathbb{N}$, each individual with label x in the $(t-1)$-th generation independently gives birth to the set of individuals $U \subseteq [n]$ with offspring distribution

$$\mu_x(U) = \begin{cases} \frac{1}{2}(1-\rho_x) & U = \emptyset \ \text{ or } \ U = \{x\}; \\ \sum_{G \in \mathcal{G}} \nu_{\mathbf{J}}(G) \, \mathbf{1}_{G(x)=U} & |U| \geq 2, \end{cases} \quad (3.14)$$

where $G(x)$ denotes the connected component of G containing x and, for any $x \in [n]$,

$$\rho_x := 1 - \prod_{z \in [n] \setminus \{x\}} (1 - p_{xz}) = \sum_{G \in \mathcal{G}} \nu_{\mathbf{J}}(G) \, \mathbf{1}_{G(x) \neq \emptyset}$$

is the probability that x has a non-empty neighborhood in the random graph defined by (3.9). Notice that by definition either $G(x)$ is empty or $|G(x)| \geq 2$, and therefore (3.14), for any $x \in [n]$, defines a probability measure on subsets of $[n]$: a sample U from μ_x is obtained by first sampling the neighborhood $G(x)$ from $\nu_{\mathbf{J}}$; if $|G(x)| \geq 2$ then we set $U = G(x)$; if $G(x) = \emptyset$ then we flip a fair coin and set $U = \emptyset$ if heads and $U = \{x\}$ if tails.

Let $N^y(t)$ denote the size of the whole population of the labeled branching process X^y at time $t-1$, i.e., the total number of individuals generated up to time $t-1$. The proof of Lemma 3.5 is based on the following bound on the exponential moment of the random variable $N^y(t)$, whose proof is deferred to the full version of the paper.

Lemma 3.6. *For all $a \in (0,1)$, there exists $\delta_0 > 0$ and $C_a > 0$ such that if (3.1) holds with constant δ_0 then, for all $y \in [n]$ and all $t \in \mathbb{N}$,*

$$\mathbb{E}[2^{a N^y(t)}] \leq C_a \, .$$

Now denote by X_ℓ^y the set of all individuals generated at the ℓ-th step, i.e., the ℓ-th generation of the process X^y, and let $|X_\ell^y|$ denote its cardinality. With this notation, an inspection of the definition of the fragmentation plus noise process shows that $F_1^{(t-1)}$ is stochastically dominated by the union of the X^y's, i.e., $\{F_1^{(\ell-1)}, \ \ell \in \mathbb{N}\}$ and the independent processes $\{X^y, \ y \in [n]\}$ can be coupled so that, for any $\ell = 1, 2, \dots,$

$$|F_1^{(\ell-1)}| \leq \sum_{y \in [n]} |X_{\ell-1}^y| \, . \quad (3.15)$$

The main observation at this point is that the event $|F_1^{(t-1)}| \geq 2$, by the definition of fragmentation plus noise, implies that $|F_1^{(\ell-1)}| \geq 2$ for all $1 \leq \ell \leq t$, which by the domination (3.15) implies that at all times $1 \leq \ell \leq t-1$ one has

$$\sum_{y \in [n]} |X_\ell^y| \geq 2.$$

For this to happen there must be two processes X^y, X^z and a time $s < t$ such that both X^y, X^z are alive up to time s and such that X^y has at least two individuals in each generation from time $s+1$ to time $t-1$. For example, taking $s = t-1$, this includes the case where both X^y, X^z are alive up to time $t-1$, while taking $s = 0$ it includes the case where all processes die at the first time step, except for X^y which has $|X_\ell^y| \geq 2$ for all $1 \leq \ell \leq t-1$. We refer to Figure 2 for an illustrative example.

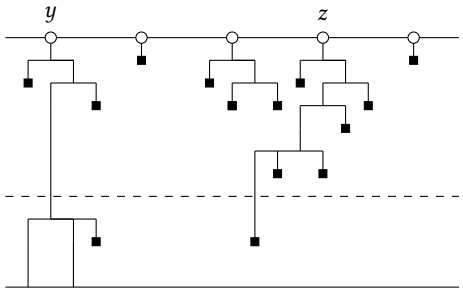

Figure 2: Illustration of the event in (3.16). The dashed line represents a time s up to which both processes X^y and X^z have cardinality at least 1 and after which the process X^y has cardinality at least 2.

Using the independence of X^y, X^z and the union bound, one has the estimate

$$\mathbb{P}(|F_1^{(t-1)}| \geq 2) \quad (3.16)$$
$$\leq \sum_{y,z \in [n]} \sum_{s=0}^{t-1} \mathbb{P}\big(|X_\ell^y| \geq 1, \ \forall\, 0 \leq \ell \leq s, |X_\ell^y| \geq 2, \ \forall s < \ell < t\big) \times \mathbb{P}\big(|X_\ell^z| \geq 1, \ \forall\, 0 \leq \ell \leq s\big).$$

Letting $N^y(t)$ denote the size of the whole population of X^y up to time $t-1$, one has

$$N^y(t) = \sum_{\ell=1}^{t} |X_{\ell-1}^y| = 1 + \sum_{\ell=1}^{t-1} |X_\ell^y|.$$

We note that for any $s \geq 0$, the event $\{|X_\ell^y| \geq 1, \ \forall\, 0 \leq \ell \leq s\}$ implies $N^y(s+1) \geq s+1$, and the events $\{|X_\ell^y| \geq 1, \ \forall\, 0 \leq \ell \leq s\}$

and $\{|X_\ell^y| \geq 2\,,\ \forall s < \ell < t\}$ together imply

$$N^y(t) \geq 2(t-s-1)+s+1 = 2t-s-1.$$

Therefore, for any fixed $a \in (0,1)$, from Lemma 3.6 and Markov's inequality, we obtain that each term in the sum in (3.16) is bounded by $C_a^2\, 2^{-2at}$. In conclusion, we deduce that

$$\mathbb{P}(|F_1^{(t-1)}| \geq 2) \leq n^2 t\, C_a^2\, 2^{-2at}.$$

Since a can be taken arbitrarily close to 1, this proves the desired estimate (3.13) by taking $t2^{-2at} \leq (4-\delta)^{-t}$ for all t large enough and adjusting the value of the constant C_δ in order to cover all values of $t \in \mathbb{N}$. This completes the proof of Lemma 3.5. □

3.4 Proof of Theorem 1.3

We now have the tools to conclude the proof of Theorem 1.3, our main result for nonlinear block dynamics. Recall the construction of $T_t(p)$ in terms of the binary derivation treein Section 2.4.

By the invariance property (2.8), the target measure $\mu_{\mathbf{J},\mathbf{h}}$ can be obtained at the root by taking the distribution $\mu_{\mathbf{J},\mathbf{h}}$ on each leaf. Thus, Theorem 1.3 says that when each leaf is given a distribution p with the same marginals as $\mu_{\mathbf{J},\mathbf{h}}$, the two distributions $T_t(p)$ and $\mu_{\mathbf{J},\mathbf{h}}$ can be coupled with an error at most Cn^2e^{-ct} for any t. We shall actually prove the following stronger result. Let $\vec{p} = (p_1, \dots, p_{2^t})$ and $\vec{q} = (q_1, \dots, q_{2^t})$ be arbitrary vectors of distributions in $\mathcal{P}(\Omega)$ whose marginals on σ_x at all sites x agree, i.e., $p_i, q_i \in \mathcal{P}(\Omega)$ satisfy

$$(p_i)_x = (p_j)_x = (q_i)_x = (q_j)_x, \quad i,j = 1,\dots,2^t,\ x \in V. \tag{3.17}$$

Let $T_t(\vec{p})$ (resp., $T_t(\vec{q})$) denote the distribution at the root of the binary tree of depth t, where the leaf labeled $i = 1,\dots,2^t$ is equipped with the distribution p_i (resp., q_i); recall Figure 1 for the case $t = 2$.

Theorem 3.7. *There exist absolute constants $\delta_0 > 0$, $c > 0$ and $C > 0$ such that, if (3.1) holds with constant δ_0 then for any choice of initial distributions $\vec{p}, \vec{q}$ as in (3.17),*

$$\|T_t(\vec{p}) - T_t(\vec{q})\|_{\mathrm{TV}} \leq Cn^2e^{-ct}, \qquad t \in \mathbb{N}. \tag{3.18}$$

Clearly, Theorem 3.7 implies Theorem 1.3 since we may take $p_i \equiv p \in \mathcal{P}(\Omega)$ and $q_i \equiv \mu_{\mathbf{J},\mathbf{h}}$, where the external fields $\mathbf{h}$ are chosen in such a way that p and $\mu_{\mathbf{J},\mathbf{h}}$ have the same marginals.

We shall prove Theorem 3.7 by analyzing the interaction history backwards in time, i.e., from the root to the leaves. This is reminiscent of the coupling from the past approach for linear Markov chains [16], and to some extent our proof is inspired by ideas that have been developed in that context. In particular, our proof is related to the information percolation framework developed by Lubetzky and Sly in [13].

Each internal node of the tree is associated with an interaction. For each such interaction we reveal the realization of the graph G and of the Bernoulli variables B in $V \setminus V_G$ that are used in sampling the random set Λ with distribution $\gamma(\cdot\,|\,\sigma,\sigma')$; see Lemma 3.2. In this way, starting from the root, we have a pair (G,B), $G \in \mathcal{G}$ and a subset $B \subseteq V \setminus V_G$ is identified with the set where $\eta_x = +1$ for $x \in V \setminus V_G$. Suppose the descendants of the root have distributions p and q respectively, as in Figure 3. Then, according to Lemma 3.2, the distribution at the root is given by

$$p \circ q = \sum_{(G,B)} \nu(G,B)\, T(p,q\,|\,G,B)\,,$$

where $\nu(G,B) = 2^{-|V\setminus V_G|}\nu_{\mathbf{J}}(G)$, and, for every realization (G,B), $T(p,q\,|\,G,B) \in \mathcal{P}(\Omega)$ is the distribution

$$T(p,q\,|\,G,B)(\tau) = \tag{3.19}$$
$$\sum_{\sigma,\sigma'} p(\sigma)q(\sigma') \sum_{\eta_{V_G}} \mu_G(\eta_{V_G}\,|\,\sigma_{V_G},\sigma'_{V_G})\ \mathbf{1}_{\tau\sim(\sigma,\sigma',\eta_{V_G},B)}.$$

Here, for $\tau \in \Omega$, the notation $\tau \sim (\sigma,\sigma',\eta_{V_G},B)$ is shorthand for

$$\tau \sim (\sigma,\sigma',\eta_{V_G},B) \ \Leftrightarrow\ \begin{cases} \tau_x = \sigma_x\,, & x \in V:\ \eta_x = +1 \\ \tau_x = \sigma'_x\,, & x \in V:\ \eta_x = -1 \end{cases} \tag{3.20}$$

with the understanding that the value of η on $V \setminus V_G$ is specified by $B \subseteq V \setminus V_G$, i.e., $\eta_y = +1$ for $y \in B$ and $\eta_y = -1$ for $y \notin B$.

For every $G \in \mathcal{G}$, $\mu_G(\cdot\,|\,\sigma_{V_G},\sigma'_{V_G})$ depends on (σ,σ') only through $(\sigma_{V_G},\sigma'_{V_G})$, and for every $\eta_{V_G} \in \{-1,1\}^{V_G}$, $\tau \in \{-1,1\}^n$ and $B \subseteq V \setminus V_G$, the condition (3.20) depends on (σ,σ') only through

$$\{\sigma_x,\ x \in V_G \cup B\}\,, \qquad \{\sigma'_x,\ x \in V_G \cup ((V\setminus V_G)\setminus B)\}\,.$$

Therefore, to compute the distribution $T(p,q\,|\,G,B)$ we only need the marginals $p_{V_G\cup B}$ and $q_{V_G\cup B'}$, where $B' = (V \setminus V_G) \setminus B$. Note that we may identity $V_G \cup B$ with the set $\Phi_0([n])$ and $V_G \cup B'$ with the set $\Phi_1([n])$, where Φ_0, Φ_1 are the maps defined in (3.11). Indeed, by definition of the measure $\nu_{\mathbf{J}}$, these random sets have the same distribution since, when $A = [n]$, V_G is equivalent to A' and B is equivalent to A_{in}. One way to rephrase this is to say that, as far as the distribution $T(p,q\,|\,G,B)$ is concerned, the only relevant information about the distribution p is contained in the set $\Phi_0([n])$ and the only relevant information about the distribution q is contained in the set $\Phi_1([n])$.

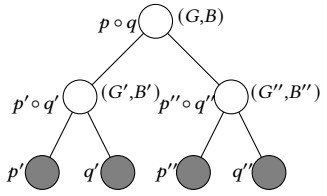

Figure 3: Graphical representation of the distribution at the root, when $t = 2$ and the leaves are equipped with distributions p', q', p'', q''. Each internal node is equipped with a realization of the random pair (G,B), where $G \in \mathcal{G}$ and $B \subseteq V \setminus V_G$.

Next, we move one step backwards in time and consider the interaction which produced the distribution p from the previous computation. Suppose that p', q' are the distributions at the two descendants of p respectively, so that $p = p' \circ q'$, as in Figure 3. Suppose we revealed the realization (G',B') of the graph and Bernoulli variables associated with this interaction. Note that we can use the same expressions (3.19)-(3.20) to compute $T(p',q'\,|\,G',B')$, provided we replace (p,q) by (p',q') and (G,B) by (G',B'). However, the key point is that we now only need the marginal of p on the set $\Phi_0([n])$, and therefore when we compute $T(p',q'\,|\,G',B')(\tau)$ as above we can sum away all τ_y, $y \notin \Phi_0([n])$ so that the indicator function (3.20) is relevant only for sites $x \in \Phi_0([n])$, and the distribution of η_x, $x \in V_{G'} \cap \Phi_0([n])$, under $\mu_{G'}(\cdot\,|\,\sigma_{V_{G'}},\sigma'_{V_{G'}})$ is only influenced by the spins σ_y, σ'_y for $y \in G'(\Phi_0([n]))$, where we recall that $G'(A)$ is the union of all connected components of G' that have

nonempty intersection with a set A. The latter property is a key consequence of the product structure of the measure $\mu_{G'}(\cdot \mid \sigma_{V_{G'}}, \sigma'_{V_{G'}})$, see Lemma 3.2. Hence we can neglect all connected components of G' that do not intersect $\Phi_0([n])$ and we can discard the information about all Bernoulli variables at sites $y \in V \setminus V_{G'}$ such that $y \notin \Phi_0([n])$. A close inspection of our definition of the maps Φ_0 and Φ_1 then reveals that the only information about the distributions p', q' that is needed to compute $T(p', q' \mid G', B')$ is contained in the marginals $p'_{\Phi_0(\Phi_0([n]))}$ and $q'_{\Phi_1(\Phi_0([n]))}$.

Similarly, considering the interaction which produced the distribution $q = p'' \circ q''$ (see Figure 3), we may fix a realization (G'', B'') of the graph and Bernoulli variables and, repeating the above reasoning, one has that the only information about the distributions p'', q'' that is needed to compute $T(p'', q'' \mid G'', B'')$ is contained in the marginals $p''_{\Phi_0(\Phi_1([n]))}$ and $q''_{\Phi_1(\Phi_1([n]))}$. Thus after two steps of the evolution, conditional on the realizations of the variables (G, B), (G', B'), (G'', B''), we have obtained a probability measure at the root depending only on (G, B), (G', B'), (G'', B'')

$$p'_{\Phi_0(\Phi_0([n])))}, q'_{\Phi_1(\Phi_0([n]))}, p''_{\Phi_0(\Phi_1([n]))}, q''_{\Phi_1(\Phi_1([n]))}.$$

Equivalently, using the notation introduced in Definition 3.3, after two steps we have that correlations of the initial distributions at the leaves are entirely contained in the fragments

$$\begin{aligned}\mathcal{F}_2 &= (F_1^{(2)}, F_2^{(2)}, F_3^{(2)}, F_4^{(2)}) \\ &= (\Phi_0(\Phi_0([n])), \Phi_1(\Phi_0([n])), \Phi_0(\Phi_1([n])), \Phi_1(\Phi_1([n]))).\end{aligned}$$

Example 3.8. *Consider the following simple example with $n = 4$. Suppose $G = \{\{1, 2\}\}$, i.e., G consists of the single edge $\{1, 2\}$ and suppose that $B = \emptyset$, i.e., both 3 and 4 are* out. *This gives $F_1^{(1)} = \Phi_0([n]) = \{1, 2\}$, $F_2^{(1)} = \Phi_1([n]) = \{1, 2, 3, 4\}$. Suppose also that $G' = \{\{3, 4\}\}$, and $B' = \{1\}$. Thus $F_1^{(2)} = \{1\}$, and $F_2^{(2)} = \{2\}$. Suppose finally that $G'' = \emptyset$, $B'' = \{1, 2\}$. Then $F_3^{(2)} = \{1, 2\}$, and $F_4^{(2)} = \{3, 4\}$. Thus, one has the following fragmentation after two steps*

$$\begin{aligned}\mathcal{F}_0 &= \{1, 2, 3, 4\}, \\ \mathcal{F}_1 &= (\{1, 2\}, \{1, 2, 3, 4\}), \\ \mathcal{F}_2 &= (\{1\}, \{2\}, \{1, 2\}, \{3, 4\}).\end{aligned}$$

Note that in this example, conditional on the given realizations of the variables (G, B), (G', B'), and (G'', B''), the distribution $p \circ q$ at the root in Figure 3 can be computed only using the marginals

$$(p')_{\{1\}}, (q')_{\{2\}}, (p'')_{\{1,2\}}, (q'')_{\{3,4\}}$$

of the input distributions p', q', p'', q''.

Repeating the above procedure one has that, after t steps, conditional on the realizations of all graphs and Bernoulli variables involved in all $2^t - 1$ interactions, the only information needed from leaf i is contained in the marginal of the distribution at that leaf on the fragment $F_i^{(t)}$, for each $i = 1, \dots, 2^t$, as defined in Definition 3.3. The crucial observation is that, as soon as a fragment becomes either empty or contains one site only, then the information carried by the corresponding leaf is irrelevant. Indeed, if it is empty this is obvious, while if it contains one site only then the marginal at that site is irrelevant since all marginals are fixed. This explains why we introduced the killing step (3.10) in our definition of the fragmentation plus noise process $\mathcal{F}_t$, which in turn is crucial to the probability of extinction we are able to establish in Lemma 3.5.

The above discussion shows that, if we denote by

$$(\vec{G}, \vec{B}) = (G^{(1)}, B^{(1)}, \dots, G^{(2^t-1)}, B^{(2^t-1)})$$

the vector of realizations of the random graphs and Bernoulli variables involved in each interaction at the $2^t - 1$ internal nodes of the binary tree of depth t, one can write

$$T_t(\vec{p}) = \sum_{(\vec{G}, \vec{B})} \widehat{\nu}(\vec{G}, \vec{B})\, T_t(\vec{p} \mid \vec{G}, \vec{B})$$

where

$$\widehat{\nu}(\vec{G}, \vec{B}) = \prod_{i=1}^{2^t-1} \nu_{\mathbf{J}}(G^{(i)})\, 2^{-|V \setminus V_{G^{(i)}}|} \tag{3.21}$$

is the distribution of the random graphs and Bernoulli variables, while $T_t(\vec{p} \mid \vec{G}, \vec{B})$ is some probability measure that may depend on $(\vec{G}, \vec{B})$ and on $\vec{p} = (p_1, \dots, p_{2^t})$ in a complicated way but has the property that its dependence on the distribution p_i from the i-th leaf occurs only through the marginal of p_i on the fragment $F_i^{(t)}$. In particular, if $\mathcal{F}_t = \emptyset$, i.e., $F_i^{(t)} = \emptyset$ for all i, then $T_t(\vec{p} \mid \vec{G}, \vec{B}) = T_t(\vec{q} \mid \vec{G}, \vec{B})$. Thus we have

$$\|T_t(\vec{p}) - T_t(\vec{q})\|_{\mathrm{TV}} \le \sum_{(\vec{G}, \vec{B}) \notin \{\mathcal{F}_t = \emptyset\}} \widehat{\nu}(\vec{G}, \vec{B})\,.$$

Next, we note that

$$\sum_{(\vec{G}, \vec{B}) \notin \{\mathcal{F}_t = \emptyset\}} \widehat{\nu}(\vec{G}, \vec{B}) = \mathbb{P}(\mathcal{F}_t \neq \emptyset)\,, \qquad t \in \mathbb{N}, \tag{3.22}$$

where the latter is the probability estimated in Lemma 3.5. Indeed, (3.22) is a consequence of our definition of the fragmentation plus noise process: we have already observed that each step of the fragmentation

$$F_i^{(t-1)} \longrightarrow (\Phi_0(F_i^{(t-1)}), \Phi_1(F_i^{(t-1)}))$$

is produced with the correct distribution, and the product structure (3.21) of the measure $\widehat{\nu}(\vec{G}, \vec{B})$ guarantees that all such steps are performed independently. From Lemma 3.5 we thus conclude that, for any $\delta \in (0, 1)$, there exists a constant $\delta_0 > 0$ in (3.1), and a constant $C_\delta > 0$ such that

$$\|T_t(\vec{p}) - T_t(\vec{q})\|_{\mathrm{TV}} \le n^2 C_\delta\, (2 - \delta)^{-t}\,, \qquad t = 1, 2, \dots$$

This implies (3.18) (and in fact shows that we can take the constant c as close as we wish to $\log 2$ provided δ_0 is taken suitably small). This ends the proof of Theorem 3.7 and thus also of Theorem 1.3. □

4 NONLINEAR GLAUBER DYNAMICS

In this section we prove Theorem 1.2, a $O(n \log n)$ bound on the convergence time for the nonlinear Glauber dynamics. The proof follows a similar overall strategy to that of Theorem 1.3, but with some significant technical differences. Due to lack of space, we defer the proof to the full version and list here only the main differences.

First, the coupling of the spin exchange process with inhomogeneous Erdös-Rényi graphs in Section 3.2 is replaced by a coupling with *inhomogeneous random star graphs* centered at a site x. This reflects the fact that only a single spin is updated at each step, so the dependencies in the dynamics are restricted to its neighbors.

Second, in the single-site dynamics, the fragmentation event becomes the event that every site is hit at least once (i.e., all spins have been updated), which is just coupon collecting. Thus the "fragmentation with noise" process of Section 3.3 is replaced by "coupon collecting with noise", where the noise again models the growth of dependencies, governed now by the random star graphs.

Finally, as in Section 3, if (3.1) holds with sufficiently small δ_0, the resulting branching process can be shown to be subcritical. The proof of this fact is again quite technical and involves a detour into continuous time branching processes.

5 REMARKS AND OPEN QUESTIONS

We conclude by briefly mentioning some interesting directions for further research, along with some additional observations.

(1) Can we prove exponential decay (even with an exponentially bad dependence on n) for the nonlinear Ising dynamics for *arbitrary* interactions $\mathbf{J}$, without the high-temperature condition? Recall that we did prove convergence for arbitrary $\mathbf{J}$ in Theorem 1.1, but unlike its analogs for linear Markov chains our proof apparently gives no useful rate information. It is in fact possible that the rate of convergence is polynomial for arbitrary $\mathbf{J}$, though of course the high-temperature methods we use, based on limiting the spread of dependencies, will not work there.

(2) Can our high-temperature condition $\max_x \sum_{y\in V} |J_{xy}| \le \delta_0$ be relaxed to an optimal value for δ_0, in particular in the case of the complete graph (mean-field or Curie-Weiss model)? Here all interactions $J_{xy} = \beta/n$, and $\beta = 1$ marks the phase transition. Thus one might hope to sharpen our results to require only $\delta_0 < 1$ in this case. One might also hope to replace the ℓ_1 condition on $\mathbf{J}$ by a spectral condition, as in recent work by Chen and Eldan [6].

(3) As mentioned in the introduction, our results for the Ising model generalize to any spin system with a constant number of spins at each vertex and bounded pairwise interactions, including the q-state Potts model, under an analogous Dobrushin-type condition (3.1). (In that condition, the sum is now over the maximum absolute values of all interactions involving any given site x.) This follows from the fact that, as can readily be checked, the representation of the measure $\gamma(\cdot \mid \sigma, \sigma')$ in terms of a two-spin system as described in Lemma 3.1 still holds in this more general setting, and beyond that point the rest of the analysis depends only on that representation.

(4) In contrast to linear Markov chains, nonlinear dynamics do not immediately provide an efficient algorithm for sampling from the stationary distribution, even when the convergence time is short: to obtain a single sample from the time-t distribution $T_t(p)$, we would naively need to simulate the entire derivation tree of depth t, beginning with 2^t independent random samples from p at the leaves. However, for both of the dynamics we consider here, our analysis does in fact provide a polynomial time algorithm for sampling from the stationary distribution $\mu_{\mathbf{J},\mathbf{h}}$. This is immediately obvious for the nonlinear block dynamics: since Theorem 1.3 establishes that $t = O(\log n)$ steps are enough for convergence, the entire tree is of polynomial size and thus can be constructed in polynomial time. (A detail here is that each interaction in the tree requires sampling the set of sites Λ to be exchanged according to the distribution (3.7). However, this itself is a high-temperature Ising Gibbs measure, and thus can be sampled from efficiently by other means.)

An interesting related question is whether a more explicit simulation is possible: namely, starting with a finite population of size $N = \text{poly}(n)$ sampled independently from the initial distribution p, evolve it via interactions between randomly chosen pairs. The finite population introduces statistical deviations from $T_t(p)$ that need to be controlled. (This process, which is actually a Markov chain on a very large state space, is known as the "Kac model" in kinetic theory [12], and the convergence to the true population is referred to as "propagation of chaos".) In the non-interacting (population genetics) case, it was shown that a low-degree polynomial population size does in fact suffice [3, 17]. It would be interesting to see if this argument can be extended to the Ising model at high temperatures.

ACKNOWLEDGEMENTS

PC was supported in part by the Miller Institute at UC Berkeley. AS was supported in part by NSF grants CCF-1815328 and CCF-2231095. Part of this work was done while AS was visiting EPFL, Lausanne and while PC was visiting UC Berkeley.

REFERENCES

[1] L. Boltzmann. 1974. The second law of thermodynamics. In *Populäre Schriften, Essay 3,* 1886; reprinted in *Ludwig Boltzmann, Theoretical Physics and Philosophical Problems*, S.G. Brush (Trans.). Reidel, Boston.
[2] P. Caputo, C. Labbé, and H. Lacoin. 2024. Cutoff phenomenon in nonlinear recombinations. *Preprint arXiv:2402.11396* (2024).
[3] P. Caputo and D. Parisi. 2022. Nonlinear recombinations and generalized random transpositions. *Preprint arXiv:2207.04775* (2022).
[4] P. Caputo and A. Sinclair. 2018. Entropy production in nonlinear recombination models. *Bernoulli* 24, 4B (2018), 3246–3282.
[5] P. Caputo and A. Sinclair. 2023. Nonlinear dynamics for the Ising model. *Preprint arXiv:2305.18788* (2023).
[6] Y. Chen and R. Eldan. 2022. Localization schemes: A framework for proving mixing times for Markov chains. *Preprint arXiv:2203.04163* (2022).
[7] M. Feinberg. 2019. *Foundations of Chemical Reaction Network Theory.* Springer Nature Switzerland.
[8] S. Friedli and Y. Velenik. 2017. *Statistical Mechanics of Lattice Systems: A Concrete Mathematical Introduction.* Cambridge University Press.
[9] D.E. Goldberg. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning.* Addison-Wesley, Reading MA.
[10] G.H. Hardy. 1908. Mendelian proportions in a mixed population. *Science* 28 (1908), 49–50.
[11] F.J.M. Horn and R. Jackson. 1972. General mass action kinetics. *Arch. Rational Mech. Anal.* 49 (1972), 81–116.
[12] M. Kac. 1956. Foundations of kinetic theory. In *Proceedings of The 3rd Berkeley Symposium on Mathematical Statistics & Probability Vol. 3.* 171–197.
[13] E. Lubetzky and A. Sly. 2017. Universality of cutoff for the Ising model. *The Annals of Probability* 45, 6A (2017), 3664–3696.
[14] M. Mitchell. 1998. *An Introduction to Genetic Algorithms.* MIT Press, Cambridge, MA.
[15] H.C. Nguyen, R. Zecchina, and J. Berg. 2017. Inverse statistical problems: from the inverse Ising problem to data science. *Advances in Physics* 66 (2017), 197–261.
[16] J.G. Propp and D.B. Wilson. 1996. Exact sampling with coupled Markov chains and applications to statistical mechanics. *Random Struct. Algor.* 9 (1996), 223–252.
[17] Y. Rabani, Y. Rabinovich, and A. Sinclair. 1998. A computational view of population genetics. *Random Struct. Algor.* 12, 4 (1998), 313–334.
[18] Y. Rabinovich, A. Sinclair, and A. Wigderson. 1992. Quadratic dynamical systems. In *Proceedings of the 33rd IEEE FOCS.* IEEE, 304–313.
[19] D. Siegel and D. MacLean. 2000. Global stability of complex balanced mechanisms. *Journal of Mathematical Chemistry* 27 (2000), 89–110.
[20] M. Singh and N. Vishnoi. 2014. Entropy, optimization and counting. *Proceedings of the 46th ACM STOC* (2014), 50–59.
[21] J. van den Berg and A. Gandolfi. 2013. BK-type inequalities and generalized random-cluster representations. *Probab. Theory Rel. Fields* 157 (2013), 157–181.
[22] W. Weinberg. 1908. Über den Nachweis der Vererbung beim Menschen. *Jarhes-hefte des Vereins für vaterländische Naturkunde in Württemberg* 64 (1908), 368–382.
[23] D. Weitz. 2005. Combinatorial criteria for uniqueness of Gibbs measures. *Random Struct. Algor.* 27 (2005), 445–475.

Received 13-NOV-2023; accepted 2024-02-11