Information Storage in the Stochastic Ising Model

Ziv Goldfeld, Guy Bresler and Yury Polyanskiy

Abstract

Most information systems store data by modifying the local state of matter, in the hope that atomic (or sub-atomic) local interactions would stabilize the state for a sufficiently long time, thereby allowing later recovery. In this work we initiate the study of information retention in locally-interacting systems. The evolution in time of the interacting particles is modeled via the stochastic Ising model (SIM). The initial spin configuration X_0 serves as the user-controlled input. The output configuration X_t is produced by running t steps of the Glauber chain. Our main goal is to evaluate the information capacity $I_n(t) \triangleq \max_{p_{X_0}} I(X_0; X_t)$ when the time t scales with the size of the system n. For the zero-temperature SIM on the two-dimensional $\sqrt{n} \times \sqrt{n}$ grid and free boundary conditions, it is easy to show that $I_n(t) = \Theta(n)$ for $t = O(n)$. In addition, we show that on the order of $\sqrt{n}$ bits can be stored for infinite time (and even with zero error) in horizontally or vertically striped configurations. The $\sqrt{n}$ achievability is optimal when $t \to \infty$ and n is fixed.

One of the main results of this work is an achievability scheme that stores more than $\sqrt{n}$ bits (in orders of magnitude) for superlinear (in n) times. The analysis of the scheme decomposes the system into $\Omega(\sqrt{n})$ independent Z-channels whose crossover probability is found via the (recently rigorously established) Lifshitz law of phase boundary movement. Lastly for the zero-temperature case, two order optimal characterizations of $I_n(t)$, for all t, are given for the grid dynamics with an external magnetic field and for the dynamics over the honeycomb lattice. In both these setups $I_n(t) = \Theta(n)$, for all t, suggesting their superiority over the grid without an external field for storage purposes. We also provide results for the positive but small temperature regime. We show that an initial configuration drawn according to the Gibbs measure cannot retain more than a single bit for $t \geq e^{cn^{\frac{1}{4}+\epsilon}}$. On the other hand, when scaling time with β, the stripe-based coding scheme (that stores for infinite time at zero temperature) is shown to retain its bits for time that is exponential in β.

Index Terms

Glauber dynamics, information capacity, Markov chains, stochastic Ising model, storage.

This work of Z. Goldfeld and Y. Polyanskiy was supported in part by the National Science Foundation CAREER award under grant agreement CCF-12-53205, by the Center for Science of Information (CSoI), an NSF Science and Technology Center under grant agreement CCF-09-39370, and a grant from Skoltech–MIT Joint Next Generation Program (NGP). The work of Z. Goldfeld was also supported by the Rothschild postdoc fellowship. The work of G. Bresler was supported by the ONR N00014-17-1-2147, DARPA W911NF-16-1-0551, and NSF CCF-1565516.
This paper will be presented in part at the 2018 IEEE International Symposium on Information Theory (ISIT-2018), Vail, Colorado, US.
Z. Goldfeld, G. Bresler and Y. Polyanskiy are with the Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139 USA (e-mails: zivg@mit.edu, guy@mit.edu, yp@mit.edu).

CONTENTS

I. Introduction

A. Storing Information Inside Matter

The predominant technology for long-term digital storage is based on physical effects such as magnetization of domains, or changes of meta-stable states of organic molecules. Data is written to the system by perturbing the local state of matter, e.g., by magnetizing particles to take one of two possible spins. In the time between writing and reading, the stored data degrades due to interparticle interactions driven by quantum/thermal fluctuations. We aim to capture this key physical phenomena in our model and study its influence on the duration of (reliable) storage. The physical model of an Ising spin system is adopted to describe the interplay between particles. Data is written into the initial state of the system by manipulating spins at the particle level. The features of this model set the ground for the study of the fundamental notion of data storage inside matter, isolated from any particular technology-driven application. The current paper deals with the question of how much information can be stored in a planar medium for a given time-duration.

Typical models for the study of reliable data storage overlook the inherent time-dependent evolution of the system, and, in particular, the interactions between particles. Motivated by application based on magnetic/flash storage devices, interactions are partially accounted for by run-length limited (RLL) codes [1], [2], which mitigate unwanted patterns. On top of that, a certain (stationary, with respect to the sequence of input bits) error model is

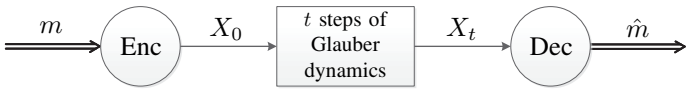


Fig. 1: A new model for storage: The encoder maps the data m into the initial Ising configuration X_0. The channel applies t steps of Glauber dynamics. The decoder tries to recover the data from the resulting configuration X_t.

adopted and governs the relation between input and output. Error-correcting codes are used to combat these errors. Maximum distance separable (MDS) codes, such as Reed-Solomon codes [3] (and many more), are preferable from several aspects such as their resilience to erasures. In this way, the storage problem is reduced to that of coding over a memoryless noisy channel. While the conversion to the well-understood channel coding problem is helpful, this approach fails to capture central physical phenomena concerning the system's evolution in time.

We propose a new model for the study of storage in locally interacting systems, which accounts for the system's fluctuations in time. The time-evolution is modeled by the Glauber dynamics [4] for the Ising model [5], also known as the *stochastic Ising model* (SIM). This is a widely accepted model for nonequilibrium ferromagnetic statistical mechanics [6] that describes interparticle interaction in real-life substances. In the proposed setup (Fig. 1), the encoder controls the initial configuration X_0, and the system undergoes t steps of the Glauber dynamics. This produces the output configuration X_t, from which the decoder tries to recover the message. The fundamental quantity we consider is the information capacity

$$\max_{p_{X_0}} I(X_0; X_t). \tag{1}$$

The information capacity is shown to have the desired operational meaning: it characterizes (at least approximately) the number of bits that can be stored for t time-units in the storage medium. Our main focus is to evaluate (1) when t scales with the size of the system n according to various rates.

B. The Stochastic Ising Model

The SIM at inverse temperature $\beta > 0$ is a reversible Markov chain (MC) whose stationary distribution is the Gibbs measure [7, Chapter 15]. At each step, the discrete-time chain picks a site uniformly at random and generates a new spin for that site according to the Gibbs measure conditioned on the rest of the system. Consequently, spins have a tendency to align, i.e., spins at adjacent sites favor having the same value. The lower the temperature, the stronger the influence of neighboring spins on one another.

The SIM has attracted intensive attention since its introduction by Glauber in 1963 [4]. Works on the topic are highly diverse, varying between questions on mixing times [8], [9] (see also [7, Chapter 15]), phase transition [5], [10], [11], metastability [12]–[15], and many more.

Of particular relevance to our paper is the so-called zero-temperature SIM on the two-dimensional (2D) square lattice. Taking the limit of $\beta \to \infty$, the transition rule amounts to a majority update: the updated site takes the same spin as the majority of its neighbors, or, in case of a tie, chooses its spin according to a fair coin toss. This

process has been much studied in the physics literature as a model of "domain coarsening" (see, e.g., [16]): clusters of constant sign (either $+1$ or -1) shrink, grow, split or coalesce as their boundaries evolve. One of the prominent rigorous results on this model concerns the disappearance time of an all-plus droplet in a sea of minuses. In [17] it was shown that any such convex droplet disappears within time proportional to its area. This result, known as the Lifshitz law, is central to analyzing one of the coding schemes in this work.

C. The Storage Problem and Contributions

The SIM's underlying graph describes the interactions between particles and corresponds to the topology of the storage medium. Putting practical interpretations aside for a moment, a natural first case-study is the complete graph K_n on n vertices, also known as the Curie-Weiss model. In the high-temperature regime, when $\beta < 1$, the Glauber chain exhibits fast mixing of time $O(n \log n)$ [18]. Hence, not even 1 bit can be stored in such a system for a long (e.g., exponential) time. In the low temperature regime, when $\beta > 1$, prima facie, the situation seems better: an exponential mixing time for Glauber dynamics on K_n was established in [19]. The equilibrium distribution in this case concentrates around two values, one corresponds to positive magnetization[1] and the other one to negative magnetization. The exponential mixing time is a consequence of the bottleneck in between these two phases. In [20], it was shown that if the chain is restricted to states of non-negative magnetization, then mixing happens in time $O(n \log n)$. From our perspective, this means that while a single bit (corresponding to positive or negative magnetization) can be stored for exponential time, any data beyond that dissolves after order of $n \log n$ updates. Once again, a pessimistic conclusion.

Planar typologies correspond to the structure of real-life storage units. In this work we focus on latices, and in particular on the 2D square grid of n vertices. At infinite temperature ($\beta = 0$), interactions are eliminated and, upon selection, particles flip with probability $\frac{1}{2}$, independently of their locality. Taking $t = cn$, the grid essentially becomes an n-fold binary-symmetric channel (BSC) with flip probability $\frac{1}{2}\left(1 - e^{-c/4}\right)$, which is arbitrarily close to $\frac{1}{2}$ for large c. Thus, the per-site capacity is almost zero. Understanding whether introducing interactions (i.e., $\beta > 0$), enhances the capacity of the system is one of our main interests. Classical results on the 2D Ising model phase transition and mixing times [21] imply the following: for $\beta < \beta_c$, where $\beta_c = \frac{1}{2}\log\left(1 + \sqrt{2}\right)$, we have $I_n\big(\mathrm{poly}(n)\big) = 0$, while for $\beta > \beta_c$, $I_n\big(\exp(\sqrt{n})\big) \geq 1$.[2]

To better understand whether anything beyond a single bit can be stored in the SIM on the 2D grid at $\beta > \beta_c$ consider two (stochastic) trajectories with one started from an all-plus state and another one from a fixed state σ. Evolving them jointly using the standard synchronous coupling, we let $p_t(\sigma)$ be the probability that the trajectories have coupled by time t. [22] shows that $p_t(\sigma)$ averaged over all σ sampled from the Gibbs distribution conditioned on positive magnetization converges to one for $t \geq e^{cn^{\frac{1}{4}+\epsilon}}$. This time corresponds to the mixing time of a SIM on a 2D grid with a plus boundary condition, recently improved by [8] to $n^{O(\log n)}$, and conjectured to be order n (possibly up to logarithmic terms). We point out that this does not resolve the question we posed, but suggests

[1]The magnetization of a configuration is the normalized sum of all the spins.

[2]The 2D SIM on the grid mixes within $O(n \log n)$ time when $\beta < \beta_c$, and exhibits exponential mixing time of $e^{\Omega(\sqrt{n})}$, when $\beta > \beta_c$ [21].

that if a state that does not couple with the all-plus trajectory exists, it should not be a typical one with respect to the Gibbs distribution. To quantify this statement, we use the result from [22, Proposition 5.2] to show that for sufficiently large $\beta > 0$ (i.e., at low temperature) $I(X_0; X_t) \leq \log 2 + o(1)$, if X_0 is distributed according to the Gibbs measure and $t \geq e^{cn^{\frac{1}{4}+\epsilon}}$. This result is stated and proven in Section VII. The positive temperature regime also enables to study to scale time with β (instead of n). We show that $\sqrt{n}$ bits stored into monochromatic horizontal or vertical stripes, will be decodable via majority decoding after $t \sim e^{c\beta}$ (Section VIII). Key to this observation is a new result on the survival time of a single plus-labeled stripe in a sea of minuses. Even with these understandings, however, a general analysis of the information capacity for $\beta > 0$ seems very challenging.

As a first step towards the general $\beta > 0$ case, we consider the limit $\beta \to \infty$. The information capacity, denoted by $I_n(t)$, is first studied when n is fixed and $t \to \infty$. Beyond answering the question of 'how much information can be stored in the system forever?', this limiting quantity also lower bounds $I_n(t)$, for all t. We characterize the zero-temperature SIM as an absorbing MC, and identify the set of absorbing states (referred to as 'stable configurations'). A configuration is stable if and only if it has a (horizontal or vertical) striped pattern, with stripes of width at least two. Since the number of stripes is of order $\sqrt{n}$, we obtain $\lim_{t\to\infty} I_n(t) = \Theta(\sqrt{n})$. Achievability follows by coding only over the stripes; the converse uses the MC's absorbing nature.

Next, the Gilbert-Varshamov existence claim easily shows that up to linear times of approximately $\frac{n}{4}$, one can store a linear number of bits (which is order optimal). The main challenge, therefore, becomes understanding what happens in between these two regimes, where t is superlinear but finite. A coding scheme that stores more than $\sqrt{n}$ bits (in order of magnitude) for superlinear times is of particular interest. We devise a scheme based on arranging monochromatic droplets (i.e., of all-plus or all-minus spins) in a sea of minuses. By growing the size of the droplets $a(n)$ as any $o(n)$ function, one can reliably store $\frac{n}{a(n)}$ bits for times up to $a(n) \cdot n$. We analyze the continuous-time version of the dynamics (easily shown to be equivalent for our purposes to the discrete-time version). The configurations in our codebook decouple into $\frac{n}{a(n)}$ independent MCs, and for each one, thresholding the number of pluses results in a Z-channel with positive capacity. Particularizing the Lifshitz law result from [17] to the case of an $a(n)$-sized square droplet, implies that its disappearance time is of order $a(n) \cdot n$. We provide an alternative and simple proof for this special case of the Lifshitz law based on stochastic domination arguments and Hopf's Umlaufsatz. The droplet's erosion time and a tenzorization step are used to conclude the analysis of the droplet-based coding scheme. Our main results for the zero-temperature grid dynamics are summarized in Table I.

The importance of storage for superlinear time is best viewed by translating the discrete-time dynamics to the 'physical' time domain. Our choice to state the problem in terms of discrete-time is merely for the sake of simplicity (and because its equivalence to the appropriate continuous-time chain). 'Physical' time, however, corresponds to a continuous-time SIM where the spin at each site is refreshed according to a Poisson clock of rate 1, independent of the system's size n. Since, on average, the discrete-time dynamics updates each site once in every n steps, discrete time is a stretched version of physical time by a factor of n. Thus, linear time in our discrete-time SIM corresponds to *constant* physical time. Storage for superlinear time, on the other hand, translates into a system that, in the physical time domain, stores for longer times as its size increases. A storage medium whose stability benefits from increased size is highly desirable from any practical aspect.

Time	Information Capacity	Comments
$t=0$	$I_n(t)=n$	An upper bound for all t
$t=O(n)$	$I_n(t)=\Theta(n)$	Linear discrete-time = Constant 'physical' time
$t=a(n)\cdot n$ where $a(n)$ is $o(n)$	$I_n(t)=\Omega\left(\frac{n}{a(n)}\right)$	$t=n\log n \implies I_n(t)=\Omega\left(\frac{n}{\log n}\right)$ $t=n^{1+\alpha},\ \alpha\in\left(0,\frac{1}{2}\right) \implies I_n(t)=\Omega(n^{1-\alpha})$
$t\to\infty$ independent of n	$I_n(t)=\Theta(\sqrt{n})$	A lower bound for all t

TABLE I: Summary of main results for data storage in the zero-temperature SIM on the 2D grid

Finally, we highlight two modifications to the zero-temperature dynamics for which storage performance significantly improves. Introducing an external magnetic field to the grid dynamics, results in a tie-braking rule for updating sites with a balanced neighborhood. This increases the size of the stable set from order of $\sqrt{n}$ to $\Theta(n)$, which implies that $I_n(t)=\Theta(n)$, uniformly in t. The same holds, without an external field, when the grid is replaced with the honeycomb lattice. The sites in its interior all have an odd degree (namely, 3), which makes ties impossible. Similarly to the grid with an external field, this enables a tiling-based achievability of $\Omega(n)$ bits for infinite time. We conclude that for storage purposes, the architecture of the honeycomb lattice is favorable over the grid when no external field is applied. If, on the other hand, the grid dynamics are endowed with an (however small) external field, then its infinite-time storage capacity abruptly grows from $\Theta(\sqrt{n})$ to $\Theta(n)$.

D. Organization

The remainder of this paper is organized as follows. Section II provides notations and the formal definition of the SIM. Section III sets up the operational definition of the storage problem and connects the maximal size of reliable codes and $I_n(t)$. Section IV studies the asymptotics of $I_n(t)$ when $t\to\infty$ independently of n. In this section the Glauber chain is proven to be absorbing and its absorbing set is identified. In Section V we focus on reliable storage for superlinear times. The droplet-based achievability scheme is constructed and analyzed in that section. This section also states the droplet disappearance time result, for which a simple and new proof is given in Appendix F. Section VI considers the grid dynamics with an external field and the dynamics over the honeycomb lattice; in both cases, an order optimal characterization of $I_n(t)=\Theta(n)$, for all t, is given. Sections VII and VIII give preliminary results on information capacity at positive (low) temperatures. Finally, Section IX discusses the main insights of the paper and appealing future directions.

II. The Model

A. Notation

Given two integers $k\le\ell$, we set $[k:\ell]\triangleq\left\{i\in\mathbb{Z}\middle| k\le i\le\ell\right\}$; if $k=1$, the shorthand $[\ell]$ is used. Let $G_n=(\mathcal{V}_n,\mathcal{E}_n)$ be a square grid of side length $\sqrt{n}\in\mathbb{N}$, where $\mathcal{V}_n=\left\{(i,j)\right\}_{i,j\in[\sqrt{n}]}$. [3] We interchangeably use

[3]For convenience, we assume $\sqrt{n}\in\mathbb{N}$; if $\sqrt{n}\notin\mathbb{N}$, simple modification of some of the subsequent statements using ceiling and/or floor operations are needed. Regardless, our focus is on the asymptotic regime as $n\to\infty$, and the assumption that $\sqrt{n}\in\mathbb{N}$ has no affect on the asymptotic behavior.

$v \in \mathcal{V}_n$ or $(i,j) \in \mathcal{V}_n$ when referring to a vertex from $\mathcal{V}_n$. We write $v \sim w$ if $\{v,w\} \in \mathcal{E}_n$ and, for any $v \in \mathcal{V}_n$, we denote the set of v-s neighbors by $\mathcal{N}_v \triangleq \{w \in \mathcal{V}_n | w \sim v\}$. The graph distance corresponding to G_n is denoted by $d : \mathcal{V}_n \times \mathcal{V}_n \to \mathbb{N}_0$, where $\mathbb{N}_0 \triangleq \mathbb{N} \cup \{0\}$. For two sets $\mathcal{U}, \mathcal{W} \subseteq \mathcal{V}_n$, we define $d(\mathcal{U}, \mathcal{W}) \triangleq \min_{\substack{u \in \mathcal{U}, \\ w \in \mathcal{W}}} d(u,w)$.

B. The Stochastic Ising Model

Fix $\sqrt{n} \in \mathbb{N}$ and let $\Omega_n \triangleq \{-1,+1\}^{\mathcal{V}_n}$. For every $\sigma \in \Omega_n$ and $v \in \mathcal{V}_n$, $\sigma(v)$ denotes the value of σ at v. The Hamiltonian associated with the Ising model on G_n is given by

$$\mathcal{H}(\sigma) \triangleq - \sum_{\{u,v\} \in \mathcal{E}_n} \sigma(u)\sigma(v), \tag{2}$$

where $\sigma \in \Omega_n$. The Gibbs measure over G_n at inverse temperature $\beta > 0$ and free boundary conditions is

$$\pi(\sigma) = \frac{1}{Z(\beta)} e^{-\beta H(\sigma)}, \tag{3}$$

where $Z(\beta)$ is the partition function (the normalizing constant).

The Glauber dynamics for the Ising model on G_n at inverse temperature $\beta > 0$ (and free boundary conditions) is a discrete-time Markov chain on the state space $\Omega_n \triangleq \{-1,+1\}^{\mathcal{V}_n}$ that is reversible with respect to π from (3). The dynamics are described as follows. At each time step, a vertex $v \in \mathcal{V}_n$ is chosen uniformly at random (independent from everything else); the spin at v is refreshed by sampling a new value from $\{-1,+1\}$ according to the conditional Gibbs measure

$$\pi_v(s) \triangleq \pi\big(s \big| \{\sigma(u)\}_{u \neq v}\big), \tag{4}$$

where $s \in \{-1,+1\}$.

The discrete-time chain defined above can also be set up in continuous-time, which we do in Section V-E to simplify some of the derivations.

C. Zero-Temperature Dynamics and Further Definitions

The zero-temperature SIM over G_n is obtained by taking the limit as $\beta \to \infty$ of the general dynamics described above. This results in a simple majority update of the uniformly chosen site. Before giving a formal description of the transition kernel, some further definitions are needed.

Given a configuration $\sigma \in \Omega_n$ and $v \in \mathcal{V}_n$, σ^v denotes the configuration that agrees with σ everywhere except v. More precisely,

$$\sigma^v(u) \triangleq \begin{cases} \sigma(u), & u \neq v \\ -\sigma(u), & u = v \end{cases}.$$

We denote the all-plus and the all-minus configurations by $\boxplus$ and $\boxminus$, respectively. Let $m_v(\sigma) \triangleq \big|\big\{w \in \mathcal{N}_v | \sigma(w) = \sigma(v)\big\}\big|$ be the number of neighbors of v whose spin agrees with that of v. Also, set $\ell_v(\sigma) \triangleq |\mathcal{N}_v| - m_v(\sigma)$ as the number of neighbors of v disagreeing with it.

Given a configuration $\sigma \in \Omega_n$, the zero-temperature SIM on G_n with free boundary conditions evolves according to the following rule:

1) Pick a vertex $v \in \mathcal{V}$ uniformly at random.
2) Modify σ at v as follows:
   - If $m_v(\sigma) > \ell_v(\sigma)$, keep the value of $\sigma(v)$;
   - If $m_v(\sigma) < \ell_v(\sigma)$, flip the value of $\sigma(v)$;
   - Otherwise, draw $\sigma(v)$ uniformly from $\{-1, +1\}$ independently of everything else.

Denote the corresponding transition kernel by P. With respect to P, we define a *path* as a sequence $\omega = (\omega_1, \omega_2, \ldots, \omega_k)$, $k \in \mathbb{N}$, $\omega_i \in \Omega_n$ for $i \in [k]$, such that $P(\omega_i, \omega_{i+1}) > 0$, for all $i \in [k-1]$. For $\sigma, \sigma' \in \Omega_n$, we write $\sigma \rightsquigarrow \sigma'$ if σ' is reachable from σ, i.e., if there exists a path ω (say of length k) with $\omega_1 = \sigma$ and $\omega_k = \sigma'$. To specify that a path ω goes from σ to σ' we write $\omega : \sigma \rightsquigarrow \sigma'$. For $\sigma \in \Omega_n$ and $\mathcal{A} \subseteq \Omega_n$, we write $\omega : \sigma \rightsquigarrow \mathcal{A}$ if $\omega : \sigma \rightsquigarrow \sigma'$ for some $\sigma' \in \mathcal{A}$.

Let $(X_t)_{t \in \mathbb{N}_0}$ denote the Markov chain with transition kernel P. We use $\mathbb{P}$ and $\mathbb{E}$ to denote the corresponding probability measure and expectation, respectively, while $\mathbb{P}_\sigma$ and $\mathbb{E}_\sigma$ indicate a conditioning on $X_0 = \sigma$. Given an initial distribution $X_0 \sim p_{X_0}$ and any $t \in \mathbb{N}_0$, the pair (X_0, X_t) is distributed according to

$$p_{X_0,X_t}(\sigma, \eta) \triangleq \mathbb{P}(X_0 = \sigma, X_t = \eta) = p_{X_0}(\sigma) P^t(\sigma, \eta), \tag{5}$$

where P^t is the t-step transition kernel. The X_t-marginal of p_{X_0,X_t} is denoted by p_{X_t}. To stress that a certain (finite) tuple of random variables, say (X_r, X_s, X_t) with $0 \le r < s < t$, forms a Markov chain we write $X_r - X_s - X_t$. The mutual information between X_0 and X_t, denoted as $I(X_0; X_t)$, is taken with respect to p_{X_0,X_t}. The entropy of X_t is $H(X_t)$.

III. Operational versus Information Capacity

Our main focus is on the asymptotic behaviour of the information capacity

$$I_n(t) \triangleq \max_{p_{X_0}} I(X_0; X_t). \tag{6}$$

While the study of $I_n(t)$ for the Ising model is of independent interest, we are also motivated by coding for storage. This section describes the operational problem and establishes $I_n(t)$ as a fundamental quantity in the study thereof. The rest of the paper deals only with $I_n(t)$ without referring back to the operational setup.

For a fixed $\sqrt{n} \in \mathbb{N}$ and $t \in \mathbb{N}_0$, the t-th power of the transition matrix P of the zero-temperature SIM on G_n constitutes a channel from X_0 to X_t. This channel is referred to as the Stochastic Ising Channel of time t, which we abbreviate by $\mathsf{SIC}_n(t)$. As illustrated in Fig. 2, the input X_0 is controlled by the encoder, while X_t serves as the output observed by the decoder. This models a storage unit of size n (i.e., a unit with n cells), with t being the time between writing and reading. The goal is to maintain reliable communication over the $\mathsf{SIC}_n(t)$ of a message $m \in [M]$ with the largest possible alphabet size.

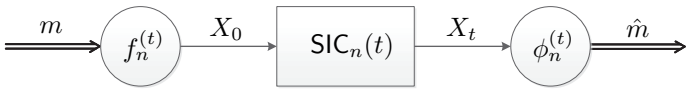


Fig. 2: The stochastic Ising channel $\mathsf{SIC}_n(t)$: The channel transition probability is P^t, where P is the one step transition kernel of the SIM on the $\sqrt{n} \times \sqrt{n}$ two-dimensional grid at zero-temperature. The functions $f_n^{(t)}$ and $\phi_n^{(t)}$ are, respectively, the encoder and the decoder.

Definition 1 (Code) *An (M, n, t, ϵ)-code $c_n^{(t)}$ for the $\mathsf{SIC}_n(t)$ is a pair of maps, the encoder $f_n^{(t)} : [M] \to \Omega_n$ and the decoder $\phi_n^{(t)} : \Omega_n \to [M]$, satisfying*

$$\frac{1}{M} \sum_{m \in [M]} \mathbb{P}_{f_n^{(t)}(m)} \left(\phi_n^{(t)}(X_t) \neq m \right) \leq \epsilon. \tag{7}$$

Definition 2 (Maximal Code Size) *The maximal code size for the $\mathsf{SIC}_n(t)$ attaining probability of error at most ϵ is*

$$M^\star(n, t, \epsilon) = \max \Big\{ M \,\Big|\, \exists \text{ an } (M, n, t, \epsilon)\text{-code for } \mathsf{SIC}_n(t) \Big\}. \tag{8}$$

The next proposition relates $M^\star(n, t, \epsilon)$ and $I_n(t)$; see Appendix A for its proof.

Proposition 1 (Maximal Code Size and Information Capacity) *The following upper and lower bounds on $M^\star(n, t, \epsilon)$ hold:*

1) *Upper Bound: For any $n \in \mathbb{N}$, $t \geq 0$ and $\epsilon > 0$, we have*

$$\log M^\star(n, t, \epsilon) \leq \frac{1}{1 - \epsilon} \Big(I_n(t) + h(\epsilon) \Big), \tag{9}$$

   where $h : [0, 1] \to [0, 1]$ is the binary entropy function.

2) *Lower Bound: Let $n_1 = o(n)$. For any $t \geq 0$ and $\epsilon > 0$, we have*

$$\frac{1}{n} \log M^\star \left(n + o\left(\frac{n}{\sqrt{n_1}} \right), t, \epsilon \right) \geq \frac{1}{n_1} I_{n_1}(t) - \sqrt{\frac{n_1}{n(1 - \epsilon)}}. \tag{10}$$

To interpret item (2), let $\alpha \in (0, 1)$ and $n_1 = n^{1-\alpha}$. This gives

$$\frac{1}{n} \log M^\star \left(n + o\left(n^{\frac{1+\alpha}{2}} \right), t, \epsilon \right) \geq \frac{1}{n^{1-\alpha}} I_{n^{1-\alpha}}(t) - \sqrt{\frac{1}{n^\alpha (1 - \epsilon)}}, \tag{11}$$

and approximates the normalized largest code size by the normalized information capacity of $G_{n^{1-\alpha}}$, for α however small. The upper and lower bounds from (9)-(10) show the importance of $I_n(t)$ for the study of coding over the $\mathsf{SIC}_n(t)$. Understanding the asymptotic behaviour of $I_n(t)$ as n and t are scaled and simultaneously grow is of particular interest. This is the focus of the rest of this work.

IV. Infinite-Time Capacity

We start by studying the limit of $I_n(t)$ as $t \to \infty$ and $n \in \mathbb{N}$ is fixed. By the Data Processing Inequality (DPI), $I_n^{(\infty)} \triangleq \lim_{s \to \infty} I_n(s)$ is a lower bound on $I_n(t)$, for all t. Namely, the DPI implies that for any $0 \leq t_1 < t_2$, we

have $I_n(t_1) \geq I_n(t_2)$ (since for all $X_0 \sim p_{X_0}$, $X_0 - X_{t_1} - X_{t_2}$ forms a Markov chain). Thus, $I_n(t) \geq I_n^{(\infty)}$, for all $t \in \mathbb{N}_0$.

To characterize $I_n^{(\infty)}$, we identify $(X_t)_{t\in\mathbb{N}_0}$ as an absorbing Markov chain. Absorbing/stable configurations are characterized as being striped. The number of striped configurations is shown to be $2^{\Theta(\sqrt{n})}$, which allows to conclude that $I_n^{(\infty)} = \Theta(\sqrt{n})$. Achievability follows by coding only over the striped configurations, while the converse uses the absorbing nature of the chain.

A. Characterization of Stable Configurations

We begin by defining the absorbing/stable set.

Definition 3 (Stable Configurations) *A configuration $\sigma \in \Omega_n$ is called stable if $P(\sigma,\sigma) = 1$. The set of all stable configurations is denoted by $\mathcal{S}_n$.*

The chain $(X_t)_{t\in\mathbb{N}_0}$ clearly has absorbing states since $P(\sigma,\sigma) = 1$, if, e.g., $\sigma \in \{\boxminus, \boxplus\}$. However, $\mathcal{S}_n$ contains many more configurations than the two ground states. We show that $\mathcal{S}_n$ coincides with the set of all (vertically or horizontally) striped configurations. A proper striped configuration partitions the $\sqrt{n} \times \sqrt{n}$ grid into horizontal or vertical monochromatic stripes of width at least 2. The formal definition is as follows.

Definition 4 (Striped Configuration) *A configuration $\sigma \in \Omega_n$ is* horizontally striped *if there exist $\ell \in \mathbb{N}$ and integers $1 < j_1 < j_2 < \ldots < j_{\ell-1} < j_\ell = \sqrt{n}$, with $j_{k+1} \geq j_k + 2$ for every $k \in [\ell - 1]$, such that*

$$\sigma\big((i,j)\big) = \sigma\big((1,j_k)\big), \quad \forall i \in \left[\sqrt{n}\right],\ j_{k-1}+1 \leq j \leq j_k,\ k \in [\ell], \tag{12}$$

where we set $j_0 = 0$.

Similarly, a configuration $\sigma \in \Omega_n$ is vertically striped *if there exist $\ell \in \mathbb{N}$ and integers $1 < i_1 < i_2 < \ldots < i_{\ell-1} < i_\ell = \sqrt{n}$, with $i_{k+1} \geq i_k + 2$ for every $k \in [\ell - 1]$, such that*

$$\sigma\big((i,j)\big) = \sigma\big((i_k,1)\big), \quad \forall i_{k-1}+1 \leq i \leq i_k,\ j \in \left[\sqrt{n}\right],\ k \in [\ell], \tag{13}$$

where $i_0 = 0$.

Finally, a striped *configuration is either vertically or horizontally striped. The set of all striped configurations is denoted by $\mathcal{A}_n$.*

The next proposition counts the number of striped configurations; its proof is given in Appendix B.

Proposition 2 (Number of Striped Configurations) *For the zero-temperature SIM on the $\sqrt{n}\times\sqrt{n}$ grid, we have $|\mathcal{A}_n| = 4f_{\sqrt{n}-1}$, where $\{f_k\}_{k\in\mathbb{N}_0}$ is the Fibonacci sequence on the indexing where $f_0 = 0$ and $f_1 = 1$. Namely,*

$$|\mathcal{A}_n| = \frac{4}{\sqrt{5}}(\phi^{\sqrt{n}-1} - \psi^{\sqrt{n}-1}), \tag{14}$$

with $\phi = \frac{1+\sqrt{5}}{2}$ and $\psi = -\frac{1}{\phi}$.

The main result of this subsection characterizes the stable configurations as striped:

Theorem 1 (Characterization of Stable Configurations) *A configuration $\sigma \in \Omega_n$ is stable if and only if it is striped, i.e., $\mathcal{S}_n = \mathcal{A}_n$.*

Theorem 1 is proved in Appendix C.

B. Characterization of Infinite-Time Capacity

Based on Theorem 1 we are now ready to state the main result for $I_n^{(\infty)}$.

Theorem 2 (Infinite Time) *For the $\sqrt{n} \times \sqrt{n}$ zero-temperature SIM, we have*

$$\log |\mathcal{S}_n| \leq I_n^{(\infty)} \leq \log |\mathcal{S}_n| + 1, \tag{15}$$

and, in particular, $I_n^{(\infty)} = \Theta(\sqrt{n})$.

The proof of Theorem 2 relies on $(X_t)_{t\in\mathbb{N}_0}$ being an absorbing Markov chain. Namely, if $\mathcal{T}_n \triangleq \Omega_n \setminus \mathcal{S}_n$, we say that $(X_t)_{t\in\mathbb{N}_0}$ is an absorbing Markov chain if for any $\rho \in \mathcal{T}_n$ there exist $t(\rho) \in \mathbb{N}$ such that $P^{t(\rho)}(\rho, \mathcal{S}_n) = \sum_{\sigma\in\mathcal{S}} P^{t(\rho)}(\rho, \sigma) > 0$.

Lemma 1 (Absorbing Markov Chain) *$(X_t)_{t\in\mathbb{N}_0}$ is an absorbing Markov chain, and consequently,*

$$\lim_{t\to\infty} \max_{\sigma\in\Omega_n} \mathbb{P}_\sigma\big(X_t \notin \mathcal{S}_n\big) = 0. \tag{16}$$

That $(X_t)_{t\in\mathbb{N}_0}$ is an absorbing Markov chain is proven in Appendix D. The proof constructs paths from any $\rho \in \mathcal{T}_n$ to some $\sigma \in \mathcal{S}_n$. The convergence in probability from (16) is a well-known property of absorbing Markov chains (see, e.g., [23, Chapter 11]). We are now ready to prove Theorem 2.

Proof of Theorem 2: For the lower bound, let X_0 be uniformly distributed over the set of stable/striped configurations $\mathcal{S}_n$. Since $P(\sigma, \sigma) = 1$ for all $\sigma \in \mathcal{S}_n$, we have

$$I_n(t) \geq H(X_0) = \log |\mathcal{S}_n|, \quad \forall t \in \mathbb{N}_0. \tag{17}$$

To upper bound $I_n(t)$, let $E_t = \mathbb{1}_{\{X_t\in\mathcal{S}_n\}}$ be the indicator on the event that X_t is stable. For any p_{X_0} and $t \in \mathbb{N}_0$, we have

$$\begin{aligned}
I(X_0; X_t) &= I(X_0; E_t, X_t) \\
&= I(X_0; E_t) + I(X_0; X_t|E_t) \\
&\leq 1 + \mathbb{P}(X_t \in \mathcal{S}_n) I(X_0; X_t|X_t \in \mathcal{S}_n) + \mathbb{P}(X_t \notin \mathcal{S}_n) I(X_0; X_t|X_t \notin \mathcal{S}_n) \\
&\overset{(a)}{\leq} 1 + \log |\mathcal{S}_n| + n \cdot \mathbb{P}(X_t \notin \mathcal{S}_n) \\
&\leq 1 + \log |\mathcal{S}_n| + n \cdot \max_{\sigma\in\Omega} \mathbb{P}_\sigma(X_t \notin \mathcal{S}_n)
\end{aligned} \tag{18}$$

where (a) uses $H(X_t|X_t \in \mathcal{S}_n) \leq \log|\mathcal{S}_n|$ and $H(X_t|X_t \notin \mathcal{S}_n) \leq \log|\Omega_n|$.

Combining (17) and (18), we obtain

$$\log|\mathcal{S}_n| \leq I_n(t) \leq 1 + \log|\mathcal{S}_n| + n \cdot \max_{\sigma\in\Omega} \mathbb{P}_\sigma(X_t \notin \mathcal{S}_n). \tag{19}$$

Taking the limit as $t \to \infty$ while using (16) and Proposition 2 shows that $I_n^{(\infty)} = \Theta(\sqrt{n})$. ■

V. Storing for Superlinear Time

The previous section showed that for any $t \in \mathbb{N}_0$, $I_n(t) = \Theta(\sqrt{n})$ is achievable by coding only over striped configurations. The question, thus, becomes: can we achieve a higher information capacity and, if yes, at what time scales? We start with the following observation:

Proposition 3 (Order Optimality for Linear Time) *Fix $\epsilon > 0$. For any $t < \left(\frac{1}{4} - \epsilon\right) n$, we have $I_n(t) = \Theta(n)$.*

Proof: First, observe that $I_n(0) = n$, which follows by choosing the initial distribution as a symmetric Bernoulli product measure (i.e., such that the initial spins are i.i.d. uniformly over $\{-1, +1\}$). By the DPI this implies that the upper bound $I_n(t) \leq n$, for all $t \geq 0$.

To show that $I_n(t) = \Omega(n)$, for any $t < \left(\frac{1}{4} - \epsilon\right) n$ and $\epsilon > 0$, we use the Gilbert-Varshamov bound, It states that there exist error-correcting codes of rate $1 - h\left(\frac{1}{4} - \epsilon\right) > 0$ and minimum distance $d > \left(\frac{1}{4} + \epsilon\right) n$. This minimum distance is greater than the number of elapsed time-steps and hence also greater than the number of spin flips, so decoding is always possible. Thus,

$$I_n(t) > n \cdot \left[1 - h\left(\frac{1}{4} - \epsilon\right) - o(1)\right] = \Omega(n), \tag{20}$$

for all $t < \left(\frac{1}{4} - \epsilon\right)$. ■

A. Beyond Linear Times - A Droplet-Based Scheme

We propose a coding scheme that decomposes G_n into independent sub-squares, each capable of reliably storing a bit for $\omega(n)$ time. The decomposition separates the sub-squares by all-minus stripes of width 2. This disentangles the dynamics inside the sub-squares and enables a tenzorization-based analysis.

Consider the evolution of a single square droplet of positive spins surrounded by a sea of minuses. If the droplet has side-length $\ell \in \mathbb{N}$, then it survives for a time of order $n\ell^2$. For instance, the erosion time of a droplet of side length $n^{\frac{\alpha}{2}}$ is $\Theta(n^{1+\alpha})$, where $\alpha \in (0, 1)$. Since the $\sqrt{n} \times \sqrt{n}$ grid can be padded with roughly $n^{1-\alpha}$ appropriately spaced such droplets, we will deduce that $I_n(t) = \Omega(n^{1-\alpha})$, for $t = O(n^{1+\alpha})$. Other droplet growth rates are of interest as well.

B. Erosion Time of Monochromatic Droplets

Fix $\ell \leq \sqrt{n} - 2$, and let $\mathcal{R}_\ell \subset \Omega_n$ be the set of configurations with a single square droplet (Fig. 3): all spins are -1 except for those inside an $\ell \times \ell$ square, which are +1. The graph distance between the square and the boundary of the grid is assumed to be at least one.

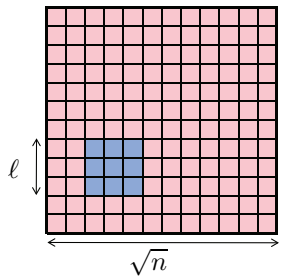


Fig. 3: A square droplet configuration of size $\ell \times \ell$. All the spins inside the square are $+1$ (represented by blue squares), while those outside of it are -1 (depicted in red).

Consider a system initiated at $X_0 = \rho \in \mathcal{R}_\ell$ and let τ be the hitting time of the all-minus configuration $\boxminus$, i.e.,

$$\tau \triangleq \inf\{t \in \mathbb{N}_0 | X_t = \boxminus\}. \tag{21}$$

Observe that the plus-labeled vertices cannot exceed the original $\ell \times \ell$ box as the system evolves with time. The gaps between the borders of the square and the grid's boundary ensure that the only stable configuration reachable from $\mathcal{R}_\ell$ is $\boxminus$. Thus, the original square shrinks until it is eliminated. As long as the graph distance between the square droplet and the boundary of the grid is at least one, the actual size of the grid is of no consequence.

To approximate τ (with high probability), it is convenient to consider the continuous-time version of the zero-temperature dynamics. As described next, the continuous-time results translate back to discrete time via a Poisson approximation.

C. Continuous-Time Zero-Temperature Dynamics

The continuous-time chain that corresponds to the discrete-time zero-temperature SIM is defined next. This is a standard definition, which we set up similarly to [7, Chapter 20]. Recall that P is the transition kernel of the discrete-time dynamics $(X_t)_{t\in\mathbb{N}_0}$. The continuous-time process $\left(X_t^{(c)}\right)_{t\geq 0}$ on Ω_n is constructed by letting $(N_t)_{t\geq 0}$ be a Poisson process of rate 1 independent of $(X_t)_{t\in\mathbb{N}_0}$, setting $X_0^{(c)} = X_0$ and

$$X_t^{(c)} = X_{N_t}, \quad t > 0. \tag{22}$$

The process $\left(X_t^{(c)}\right)_{t\geq 0}$ is updated at the jump times of the Poisson process; in between the jumps, the state stays fixed. The transition probabilities of the continuous-time dynamics are described through its heat-kernel H_t, given by

$$H_t(\sigma, \eta) \triangleq \sum_{k=0}^{\infty} \frac{e^{-t}t^k}{k!} P^k(\sigma, \eta). \tag{23}$$

For an $m \times m$ matrix M, defining the $m \times m$ matrix $e^M \triangleq \sum_{k=0}^{\infty} \frac{M^k}{k!}$, Equation (23) can be written in matrix form as $H_t = e^{t(P-I)}$.

To interpret this definition, note that

$$\mathbb{P}\Big(X_t^{(c)} = \eta \Big| N_t = k, X_0^{(c)} = \sigma\Big) = P^k(\sigma, \eta), \tag{24}$$

and therefore,

$$H_t(\sigma, \eta) = \mathbb{P}_\sigma\Big(X_t^{(c)} = \eta\Big). \tag{25}$$

For any $\sqrt{n} \in \mathbb{N}$ and $t \geq 0$, let $I_n^{(c)}(t) \triangleq \max_{p_{X_0^{(c)}}} I\left(X_0^{(c)}; X_t^{(c)}\right)$. The following proposition states that $I_n^{(c)}(t)$ is a good proxy for $I_n(t)$, whenever t scales sufficiently fast with n.

Proposition 4 (Relation Between Discrete and Continuous Time Dynamics) *For any $\epsilon \in (0,1)$, we have*

$$I_n^{(c)}\big((1+o(1))t\big) - n \cdot e^{-\Theta(t^\epsilon)} \leq I_n(t) \leq \left(1 - e^{-\Theta(t^\epsilon)}\right)^{-1} \left[I_n^{(c)}\big((1-o(1))t\big) + h\left(e^{-\Theta(t^\epsilon)}\right)\right], \tag{26}$$

where $h : [0,1] \to [0,1]$ is the binary entropy function.

The proof is relegated to Appendix E. The proposition shows that as long as t scales super-logarithmically with n (in particular, linear scaling and above will do), the discrete-time and the continuous-time mutual information terms are of the same order.

D. Equivalent Representation of the Continuous-Time Dynamics

A standard equivalent description of the continuous-time dynamics, which is easier to analyze, is given next. Any continuous-time Markov chain $(Y_t)_{t\geq 0}$ on a finite state space Ξ with kernel Q, is characterized by its transition rates $\{c_{x,y}\}_{\substack{x,y\in\mathcal{S}\\ x\neq y}}$, which satisfy

$$Q\big(Y_{t+h} = y | Y_t = x\big) = c_{x,y}h + o(h). \tag{27}$$

The generator $\mathcal{L}_Y$ associated with such a Markov chain operates on functions $f : \Xi \to \mathbb{R}$ as

$$\mathcal{L}_Y f(x) = \sum_{y\in\Xi} c_{x,y}\big(f(y) - f(x)\big). \tag{28}$$

For Glauber dynamics, this representation simplifies. By the definition of the continuous-time dynamics from the previous section, for any $\sigma, \eta \in \Omega_n$ with $\sigma \neq \eta$, we have

$$\mathbb{P}\Big(X_{t+h}^{(c)} = \eta \Big| X_t^{(c)} = \sigma\Big) = \mathbb{P}_\sigma\Big(X_h^{(c)} = \eta\Big) = \sum_{k=1}^{\infty} \frac{e^{-h}h^k}{k!} P^k(\sigma, \eta) = P(\sigma, \eta) \cdot h + o(h) \tag{29}$$

Now recall that the discrete-time chain is such that $P(\sigma, \eta) > 0$ if and only if the exists $v \in \mathcal{V}_n$, such that $\eta = \sigma^v$ (i.e., η and σ differ at v). Furthermore,

$$P(\sigma, \sigma^v) = \begin{cases} \frac{1}{n}, & m_v(\sigma) < \ell_v(\sigma) \\ \frac{1}{2n}, & m_v(\sigma) = \ell_v(\sigma) \\ 0, & m_v(\sigma) > \ell_v(\sigma). \end{cases} \tag{30}$$

Denoting $c_{v,\sigma} \triangleq P(\sigma, \sigma^v)$, for each $\sigma \in \Omega_n$ and $v \in \mathcal{V}_n$, the associated generator $\mathcal{L}$ of $\left(X_t^{(c)}\right)_{t\geq 0}$ simplifies to

$$\mathcal{L}f(\sigma) = \sum_{v\in\mathcal{V}_n} c_{v,\sigma}\big[f(\sigma^v) - f(\sigma)\big]. \tag{31}$$

Accordingly, the continuous-time dynamics are described as follows: when in state $\sigma \in \Omega_n$, each site $v \in \mathcal{V}_n$ is assigned with an independent Poisson clock of rate $c_{v,\sigma}$. When the clock at site v rings, the spin at this site flips. Any v whose spin disagrees with the spin of the majority of its neighbors (i.e., when $m_v(\sigma) < \ell_v(\sigma)$) flips independently with rate $\frac{1}{n}$; sites with a balanced locality ($m_v(\sigma) = \ell_v(\sigma)$) flip with rate $\frac{1}{2n}$; sites whose spin agrees with the spins of the majority of their neighbors ($m_v(\sigma) > \ell_v(\sigma)$) cannot flip, at least until one of the neighbors is updated.

$I_n^{(c)}(t)$ is easier to analyze with this description because the independent Poisson clocks decorrelate non-interacting pieces of the grid. This, in turn, tenzorizes the mutual information.

Remark 1 (Speedup or Slowdown of Flip Rates) *Let $\left(\bar{X}_t^{(c)}\right)_{t\geq 0}$ be a continuous-time zero-temperature dynamics with flip rate $r_{v,\sigma} = n \cdot c_{v,\sigma}$, $v \in \mathcal{V}_n$, and $\sigma \in \Omega_n$. In words, $\left(\bar{X}_t^{(c)}\right)_{t\geq 0}$ is a speedup by a factor n of the original continuous-time dynamics $\left(X_t^{(c)}\right)_{t\geq 0}$. A straightforward consequence of the Poisson process properties is that for any p_{X_0}, where $X_0^{(c)} = \bar{X}_0^{(c)} \sim p_{X_0}$, we have*

$$I\left(\bar{X}_0^{(c)}; \bar{X}_t^{(c)}\right) = I\left(X_0^{(c)}; X_{nt}^{(c)}\right). \tag{32}$$

Furthermore, the flip rates of the speedup dynamics $\left(\bar{X}_t^{(c)}\right)_{t\geq 0}$ are independent of the size of the system ($r_{v,\sigma} \in \left\{0, \frac{1}{2}, 1\right\}$, for all $v \in \mathcal{V}_n$ and $\sigma \in \Omega_n$.). This is probably the most natural physical scaling of the dynamics.

E. Erosion Time of a Square Droplet

Consider the process $\left(\bar{X}_t^{(c)}\right)_{t\geq 0}$ defined in Remark 1, with an initial configuration $\bar{X}_0^{(c)} = \rho \in \mathcal{R}_\ell$ (i.e., where ρ has $+1$ spins inside an $\ell \times \ell$ square and -1 spins everywhere else, and such that the square has a graph distance at least one from the borders of the grid). The droplet dynamics are agnostic of the size of the system. Therefore, with some abuse of notation, we assume that $\left(\bar{X}_t^{(c)}\right)_{t\geq 0}$ has state space $\Omega_{\mathbb{Z}^2} \triangleq \{-1, +1\}^{\mathbb{Z}^2}$. Define the hitting time to the all-minus configuration:

$$\bar{\tau} \triangleq \inf\left\{t \geq 0 \middle| \bar{X}_t^{(c)} = \boxminus\right\}. \tag{33}$$

A landmark result from the zero-temperature SIM literature (known as the Lifshitz law) is that, with high probability, $\bar{\tau}$ is of order linear in the size of the droplet ℓ^2 [17, Theorem 2.2]. The result from [17] is more general and shows that the disappearance time of any convex droplet is proportional to its area. In Appendix F we give a simple new proof of the square droplet disappearance time (stated next) based on stochastic domination and Hopf's Umlaufsatz.

Theorem 3 (Erosion Time of a Square Droplet) *There exist positive constants $c, C, \gamma > 0$, such that*

$$\mathbb{P}_\rho\Big(c\ell^2 \leq \bar{\tau} \leq C\ell^2\Big) \geq 1 - e^{-\gamma\ell}, \tag{34}$$

for every $\ell \geq 1$.

The next remark discusses the validity of Theorem 3 for an additional case.

Remark 2 (Droplets with Boundary Conditions) *As the bounds from* (34) *depend only on droplet's size the result of Theorem 3 remains valid for a droplet with boundary conditions. Namely, consider an* $\ell \times \ell$ *all-plus droplet that constitutes the entire system but with all-minus boundary conditions around it. Let* $\mathcal{U}_\ell \triangleq [0:\ell+1]\times[0:\ell+1]$ *be the entire vertex set,* $\mathcal{V}_\ell = [\ell]\times[\ell]$ *be (as before) the region where the dynamics evolve, and*

$$\mathcal{B}_\ell \triangleq \mathcal{U}_\ell \setminus \Big\{\mathcal{V}_\ell \cup \big\{(0,0),(0,\ell+1),(\ell+1,0),(\ell+1,\ell+1)\big\}\Big\}$$

be the boundary. Consider the continuous-time chain with state space $\Omega_\ell \triangleq \{-1,+1\}^{\mathcal{V}_\ell}$*, and (all-minus) boundary conditions* $\tau \in \Gamma_\ell \triangleq \{-1,+1\}^{\mathcal{B}_\ell}$*, where* $\tau(v) = -1$*, for all* $v \in \mathcal{B}_\ell$*. The flip rate at site* $v \in \mathcal{V}_\ell$ *when in configuration* $\sigma \in \Omega_\ell$ *and boundary conditions* $\tau \in \Gamma_\ell$ *is given by*

$$b_{v,\sigma}^{(\tau)} = \begin{cases} 1, & m_v(\sigma,\tau) < \ell_v(\sigma,\tau) \\ \frac{1}{2}, & m_v(\sigma,\tau) = \ell_v(\sigma,\tau) \\ 0, & m_v(\sigma,\tau) > \ell_v(\sigma,\tau) \end{cases} \tag{35}$$

where $m_v(\sigma,\tau)$ *and* $\ell_v(\sigma,\tau)$ *are defined like* $m_v(\sigma)$ *and* $\ell_v(\sigma)$*, respectively, but with the neighborhood of a site* v *being* $\mathcal{N}_v \triangleq \big\{w \in \mathcal{V}_\ell \cup \mathcal{B}_\ell \big| ||v-w||_1 = 1\big\}$*, where* $||\cdot||_1$ *is the* L^1 *norm in* $\mathbb{R}^2$*. The generator of these dynamics is the same as* (31)*, but with* $b_{v,\sigma}^{(\tau)}$ *in the role of* $c_{v,\sigma}$*.*

Denote the hitting time of the all-minus configuration in the original continuous-time dynamics $\left(X_t^{(c)}\right)_{t\geq 0}$ by

$$\tau^{(c)} \triangleq \inf\Big\{t \geq 0 \Big| X_t^{(c)} = \boxminus\Big\}. \tag{36}$$

Combining the result of Theorem 3 and the observation from Remark 1 gives

$$\mathbb{P}_\rho\left(cn\ell^2 \leq \tau^{(c)} \leq Cn\ell^2\right) \geq 1 - e^{-\gamma\ell}. \tag{37}$$

This inequality plays a central role in the analysis of the droplet-based coding scheme presented next.

F. Storing for Superlinear Time

The main result of this section is the following.

Theorem 4 (Storing for Superlinear Time) *Let* $a(n) \geq 1$ *be a function that scales with* n *slower than* $\sqrt{n}$*, i.e.,* $a(n) = o\left(\sqrt{n}\right)$*. Then there exists* $c > 0$*, such that for all* $t \leq c \cdot a(n) \cdot n$*, we have*

$$I_n(t) = \Omega\left(\frac{n}{a(n)}\right). \tag{38}$$

Proof: First we move from discrete-time to continuous-time using Proposition 4. As the time scale $a(n)\cdot n$ is superlinear in n, the proposition implies that the discrete-time $I_n(t)$ and the continuous-time $I_n^{(c)}(t)$ are of the same order. For the latter, we construct the distribution of $X_0^{(c)}$ as follows.

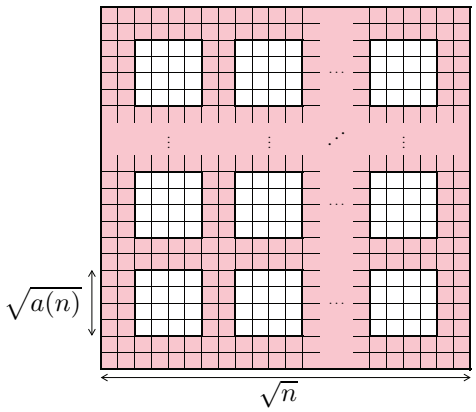


Fig. 4: A partitioning of the $\sqrt{n} \times \sqrt{n}$ grid into square droplet regions of side $\sqrt{a(n)}$ separated by stripes of negative spins of width at least 2. Red squares represent negative spins, while white squares stand for unspecified spins.

Tile the $\sqrt{n} \times \sqrt{n}$ grid with monochromatic sub-squares of side length $\sqrt{a(n)}$ (whose spins are to be specified later) separated by all-minus stripes of width exactly 2. Such a partitioning of the grid contains a total number of $\Theta(n)$ square droplets. The tiling is illustrated in Fig. 4.

Let $\mathcal{C}_n$ be the collection of all configurations whose topology corresponds to Fig. 4 with monochromatic spin assignments to each of the $\Theta\left(\frac{n}{a(n)}\right)$ sub-squares. The formal definition is as follows. Let $K = \left\lfloor \frac{\sqrt{n}-2}{\sqrt{a(n)}+2} \right\rfloor$ and for $i, j \in [K]$, let $\rho_{i,j} \in \mathcal{R}_{\sqrt{a(n)}}$ be the configuration with positive spins inside and only inside the $\sqrt{a(n)} \times \sqrt{a(n)}$ sub-square whose bottom-right corner is $\left(1 + i\big(\sqrt{a(n)}+2\big), 1 + i\big(\sqrt{a(n)}+2\big)\right) \in \mathcal{V}_n$. Namely, $\rho_{i,j}\big((x,y)\big) = +1$ if and only if $(x,y) \in \mathcal{D}_{i,j}$, where

$$\mathcal{D}_{i,j} \triangleq \left\{ (x,y) \in \mathcal{V}_n \middle| \begin{array}{l} 1 + i\big(\sqrt{a(n)}+2\big) - \sqrt{a(n)} \le x \le 1 + i\big(\sqrt{a(n)}+2\big), \\ 1 + j\big(\sqrt{a(n)}+2\big) - \sqrt{a(n)} \le y \le 1 + j\big(\sqrt{a(n)}+2\big) \end{array} \right\}. \tag{39}$$

We also denote the external boundary of each $\mathcal{D}_{i,j}$ region as

$$\mathcal{B}_{i,j} \triangleq \left\{ v \in \mathcal{V}_n \setminus \mathcal{D}_{i,j} \middle| d(v, \mathcal{D}_{i,j}) = 1 \right\}. \tag{40}$$

Letting $\Xi \triangleq \bigcup_{i,j \in [K]} \{\rho_{i,j}\}$, we set

$$\mathcal{C}_n = \bigcup_{\mathcal{A} \in 2^{\Xi}} \bigvee_{\rho \in \mathcal{A} \cup \{\boxminus\}} \rho, \tag{41}$$

where 2^{Ξ} is the power set of Ξ, while $\bigvee_{\sigma \in \mathcal{S}} \sigma$ means taking the 'or' operation of all the configuration in a set $\mathcal{S}$ (the 'or' operation $\sigma \vee \eta$ on two configurations σ and η is $(\sigma \vee \eta)(v) = -1$ if and only if $\sigma(v) = \eta(v) = -1$,). The collection $\mathcal{C}_n$ can be thought of as our codebook, where each sub-square stores a bit. Accordingly, one may encode order of $K^2 = \Theta\left(\frac{n}{a(n)}\right)$ bits into the initial configuration by mapping bits to droplets. We next show that $I_n^{(c)}(t) = \Omega(K^2)$, for $t \le C \cdot a(n) \cdot n$ (where $C > 0$ is some numerical constant).

Let $X_0^{(c)} \sim p_{X_0}$ with $\mathrm{supp}(p_{X_0}) = \mathcal{C}_n$ and such that p_{X_0} is an i.i.d. Bernoulli measure in each block (the parameter will be understood later), and note that by the nature of the continuous-time dynamics, $X_0^{(c)}$ and $X_t^{(c)}$ decompose into K^2 independent components, each corresponding to a different $\sqrt{a(n)} \times \sqrt{a(n)}$ sub-square surrounded by all-minus boundary conditions. This follows by the independent Poisson clocks that define the flip rates of $\left(X_t^{(c)}\right)_{t\geq 0}$ and because when initiating the Markov chain at any $\sigma \in \mathcal{C}_n$, interactions are confined to the $\left\{\mathcal{D}_{i,j}\right\}_{i,j\in[K]}$ regions (the white regions in Fig. 4). Letting $X_0^{(c)}(i,j)$ and $X_t^{(c)}(i,j)$ be the restriction of X_0 and X_t to $\mathcal{D}_{i,j} \cup \mathcal{B}_{i,j}$, we have that $\left\{ \left(X_0^{(c)}(i,j), X_t^{(c)}(i,j)\right) \right\}_{i,j\in[K]}$ are i.i.d. Consequently,

$$I_n^{(c)}(t) \geq \sum_{i,j\in[K]} \max_{p_{i,j}} I\left(X_0^{(c)}(i,j); X_t^{(c)}(i,j)\right) = K^2 \cdot \max_{p_{1,1}} I\left(X_0^{(c)}(1,1); X_t^{(c)}(1,1)\right). \tag{42}$$

Each mutual information term inside the sum above corresponds to a continuous-time dynamics with state space $\{-1,+1\}^{\mathcal{D}_{i,j}}$, an initial configuration $X_0^{(c)}(i,j)\Big|_{\mathcal{D}_{i,j}} \sim p_{i,j}$ with $\mathrm{supp}\,(p_{i,j}) = \{\boxminus, \boxplus\} \subset \{-1,+1\}^{\mathcal{D}_{i,j}}$, and boundary conditions $X_0^{(c)}(i,j)\Big|_{\mathcal{B}_{i,j}} = \boxminus \in \{-1,+1\}^{\mathcal{B}_{i,j}}$. Here $X_0^{(c)}(i,j)\Big|_{\mathcal{A}}$, for $\mathcal{A} \subseteq \mathcal{D}_{i,j} \cup \mathcal{B}_{i,j}$, denotes the restriction of $X_0^{(c)}(i,j)$ to $\mathcal{A}$. The equality to the RHS of (42) follows because the pairs of random variables inside each mutual information term are identically distributed.

Based on (42), to prove Theorem 4 it suffices to show that there exists $c > 0$ such that $\max_{p_{1,1}} I\left(X_0^{(c)}(1,1); X_t^{(c)}(1,1)\right) > 0$, for all $t \leq c \cdot a(n) \cdot n$. This follows because $X_0^{(c)}(1,1)$ and $X_t^{(c)}(1,1)$ are related through a Z-channel of positive capacity. To see this, define the map $\phi : \{-1,+1\}^{\mathcal{D}_{1,1}} \to \{-1,+1\}^{\mathcal{D}_{1,1}}$ as

$$\phi(\sigma) = \boxplus \cdot \mathbb{1}_{\{\sigma \neq \boxminus\}} + \boxminus \cdot \mathbb{1}_{\{\sigma = \boxminus\}}. \tag{43}$$

In words, given a sub-square with at least one positive spin, the function ϕ sets all the spins inside it to $+1$; if the original sub-square has only negative spins, ϕ is the identity map.

Note that if $X_0^{(c)}(1,1)\Big|_{\mathcal{D}_{i,j}} = \boxminus$, then $\phi\left(X_t^{(c)}(1,1)\right) = X_t^{(c)}(1,1) = X_0^{(c)}(1,1)$ almost surely (due to the all-minus boundary conditions). On the other hand, if $X_0^{(c)}(1,1)\Big|_{\mathcal{D}_{i,j}} = \boxplus$, then $\phi\left(X_t^{(c)}(1,1)\right) = X_0^{(c)}(1,1)$ with probability $q_n \triangleq \mathbb{P}\left(\tau_{1,1}^{(c)} \leq c \cdot a(n) \cdot n\right)$, and $\phi\left(X_t^{(c)}(1,1)\right) = \boxminus$ with the complement probability. Here $\tau_{1,1}^{(c)} \triangleq \inf\left\{t \geq 0 \middle| X_t^{(c)}(1,1) = \boxminus\right\}$ is the hitting time of $\boxminus$. This correspondence to the Z-channel is shown in Fig. 5.

By Theorem 3, there exist $c, \gamma > 0$, such that $q_n \leq e^{-\gamma \cdot a(n)}$. The capacity of the Z-channel with crossover probability q_n is

$$\max_{p_{1,1}} I\left(X_0^{(c)}(1,1); X_t^{(c)}(1,1)\right) = \log\left(1 + (1-q_n) q_n^{\frac{q_n}{1-q_n}}\right), \tag{44}$$

which is non-zero for all $t < c \cdot a(n) \cdot n$. ∎

VI. Tight Information Capacity Results

We consider two additional scenarios: the 2D grid with an external field and the honeycomb lattice. In both cases we show that $I_n(t) = \Theta(n)$, for all t. Introducing an external field (however small) or switching to the Honeycomb lattice is highly beneficial for the purpose of storing data in the SIM. It ensures that a linear (with n) number of

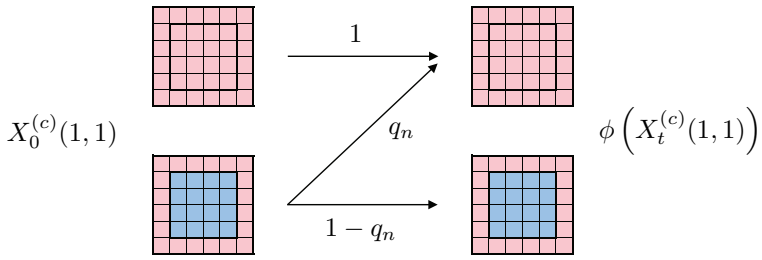


Fig. 5: The Z-channel between $X_0^{(c)}(1,1)$ and $\phi\left(X_t^{(c)}(1,1)\right)$, where the crossover probability is $q_n \triangleq \mathbb{P}\left(\tau_{1,1}^{(c)} \leq c \cdot a(n) \cdot n\right)$. Red and blue squares correspond to negative and positive spins, respectively.

bits can be stored in the system for indefinite time. This improves upon the asymptotic storage ability of the model without of external field, which is of order $\sqrt{n}$ (Theorem 2).

A. 2D Grid with External Field

In the presence of an external magnetic field $h > 0$, the Ising model Hamiltonian from (2) is replaced with

$$\mathcal{H}(\sigma) \triangleq -\sum_{\{u,v\}\in\mathcal{E}_n} \sigma(u)\sigma(v) - h\sum_{v\in\mathcal{V}_n} \sigma(v). \tag{45}$$

Introducing a positive magnetic field, however small, to the zero-temperature dynamics serves as a tie-breaker: now if we update a configuration $\sigma \in \Omega_n$ at a vertex v with $m_v(\sigma) = \ell_v(\sigma)$ (i.e., whose local environment is a tie), the spin at v is updated to $+1$ with probability 1. If the external field is negative, i.e., $h < 0$, the the role of $+1$ and -1 are reversed.

We have the following tight characterization of the information capacity.

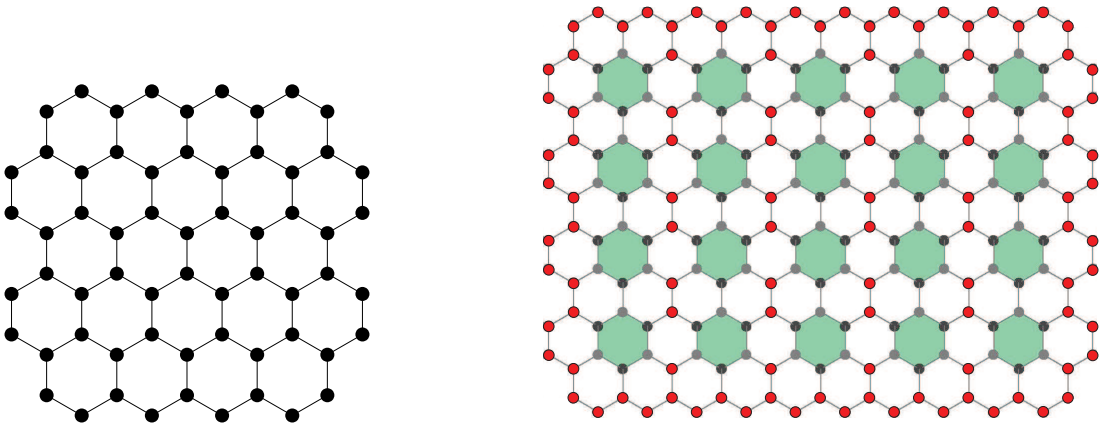
Theorem 5 (Information Capacity with an External Field) *For the zero-temperature SIM on G_n with an external field $h > 0$, we have $I_n(t) = \Theta(n)$, for all $t \geq 0$.*

Proof: The upper bound follows from $I_n(t) \leq I_n(0) = n$, for all $t \geq 0$. For the lower bound, note that with a positive external field, the droplet configurations comprising the set $\mathcal{C}_n$ from (41) (in the proof of Theorem 4), with droplet size $a(n) = 2$ (any larger constant will do), are all stable. Hence, by taking p_{X_0} to be the uniform distribution over $\mathcal{C}_n$, we get $I_n(t) = \Omega(n)$. ■

B. The honeycomb Lattice

Another instance of a zero-temperature dynamics for which the result of Theorem 5 holds is when the graph is the honeycomb Lattice on n vertices (Fig. 6), denoted by H_n.

Theorem 6 (Information Capacity on the Honeycomb Lattice) *For the zero-temperature SIM on the honeycomb lattice H_n, we have $I_n(t) = \Theta(n)$, for all $t \geq 0$.*



(a) (b)

Fig. 6: (a) A portion of the 2D honeycomb lattice; (b) Storing $\Omega(n)$ stable bits in the honeycomb lattice: Each green region corresponds to a bit, which is written by assigning the same spin to all the vertices on the border of that region; the rest of the vertices are assigned with negative spins, shown by the red circles. Each such spin assignment results in a stable configuration.

Key to the proof of Theorem 6 is that all the vertices in the interior of the finite lattice have an odd degree. This property of H_n implies that no ties can occur in the zero-temperature dynamics (similar to the effect of an external magnetic field on grid dynamics). This suggests that when no external field is applied on the system, the honeycomb lattice is preferable (over the grid) for storage purposes.

Proof: As before, we need only to prove the lower bound. To avoid tediously formalizing a simple idea, we show in Fig. 6(b) a certain tiling defined by the edges of H_n. A stable bit (which suffers no erosion in time) can be stored in each colored region. To write a bit, all the sites at the border of that region are assigned with the same spin (say, a negative spin for '$0'$ and a positive one for '$1'$). The rest of the spins in the system are set to -1 (depicted by the red circles in Fig. 6(b)). Any such configuration is stable. As the colored regions are of constant size and are within constant gaps from one another, their total number is $\Omega(n)$. Taking X_0 to be uniformly distributed over this set of configuration proves the lower bound. ■

VII. 1-Bit Upper Bound under Gibbs Initialization

This section and the next provide initial observations on the information capacity of the $\sqrt{n} \times \sqrt{n}$ 2D grid at low but positive temperature. The first result shows that at sufficiently large inverse temperature $\beta > \beta_c$, by drawing the initial configuration from the Gibbs distribution one cannot store more than a single bit in the system for times $t \geq e^{cn^{\frac{1}{4}+\epsilon}}$. This observation relies on the result of [22, Proposition 5.2] which we restate as follows.[4]

Fix $\beta \in (0, \infty)$ and let $(X_t)_{t \geq 0}$ be the discrete-time dynamics at inverse temperature β on the state space Ω_n (see Section II-B) with an initial configuration $X_0 \sim \pi$, where π is the corresponding Gibbs measure (see (2)-(3)). We also use $(X_t^\sigma)_{t \geq 0}$ as the notation of the dynamics when it is initiated at $X_0 = \sigma$. The case when the

[4]The original claim from [22] considers the continuous-time dynamics. Our restatement is for the discrete-time version described in Section II-B.

staring configuration is the all-plus state $\boxplus$ is distinguished by denoting $Y_t \triangleq X_t^{\boxplus}$, for all $t \geq 0$. Finally, for any configuration $\sigma \in \Omega_n$, denote its magnetization by

$$m(\sigma) = \frac{1}{n} \sum_{v \in \mathcal{V}_n} \sigma(v). \tag{46}$$

To state the result of [22, Proposition 5.2] we first need to set up the Markov chains $\left\{(X_t^\sigma)_{t\geq 0}\right\}_{\sigma\in\Omega_n}$ at inverse temperature $\beta > 0$ over the same probability space. The construction is similar to the one in the proof of Theorem 3 in Appendix F. Let $\left\{V_t\right\}_{t\in\mathbb{N}}$ be an i.i.d. process of random variables uniformly distributed over the vertex set $\mathcal{V}_n$. Also let $\left\{U_t\right\}_{t\in\mathbb{N}}$ be an i.i.d. process with $U_t \sim \mathsf{Unif}[0,1]$. For each initial configuration $\sigma \in \Omega_n$, we construct $(X_t^\sigma)_{t\geq 0}$ as follows. At each time $t \in \mathbb{N}$, $X_t^\sigma(\mathcal{V}_n \setminus \{V_t\}) = X_{t-1}^\sigma(\mathcal{V}_n \setminus \{V_t\})$, while for spin at V_t we set

$$X_t^\sigma(V_t) = \begin{cases} +1, & \text{if } U_t \leq \pi\big(+1 \big| X_t(\mathcal{V}_N \setminus \{V_t\})\big) \\ -1, & \text{otherwise} \end{cases}. \tag{47}$$

In the above, for any $\mathcal{A} \subseteq \mathcal{V}_n$, we used the notation $X_t(\mathcal{A})$ to denote the restriction of X_t to $\mathcal{A}$. Denote the probability measure associated with this joint probability space by $\mathbb{P}$.

Proposition 5 (Coupling with All-Plus Phase [22]) *Fix $\sqrt{n} \in \mathbb{N}$, $\epsilon \in \left(0, \frac{1}{2}\right)$ and $\gamma > 0$. There exist $\beta_0, C < \infty$ such that for any $\beta \geq \beta_0$ and $t \geq n \cdot e^{C\beta n^{\frac{1}{4}+\epsilon}}$, we have*

$$\sum_{\substack{\sigma\in\Omega_n: \\ m(\sigma)>0}} \pi(\sigma)\mathbb{P}\Big(X_t^\sigma \neq Y_t\Big) \leq e^{-\gamma\sqrt{n}}. \tag{48}$$

In words, the proposition states that if the initial configuration is distributed according to the restriction of the Gibbs measure to $\left\{\sigma \in \Omega_n \middle| m(\sigma) > 0\right\}$, by time $t \geq n \cdot e^{C\beta n^{\frac{1}{4}+\epsilon}}$, the dynamics become indistinguishable from those initiated at the all-plus state. Thus, in this time scale, the only thing the chain remembers is the positive magnetization of the initial configuration, rather than the exact starting point.

Based on Proposition 5, we show that for $X_0 \sim \pi$, the mutual information $I(X_0; X_t)$ is at most 1 bit for the aforementioned time scale.

Corollary 1 (1-bit Upper Bound on Mutual Information) *Let ϵ, γ, β_0 and C be as in Proposition 5. For any $\beta > \beta_0$ there exist $c(\beta) > 0$ such that*

$$I(X_0; X_t) \leq \log 2 + \epsilon_n(\beta), \tag{49}$$

for all $t \geq n \cdot e^{C\beta n^{\frac{1}{4}+\epsilon}}$, where

$$\epsilon_n(\beta) \triangleq n e^{-c(\beta)\sqrt{n}} + n\frac{2e^{-\gamma\sqrt{n}}}{1-e^{-c(\beta)\sqrt{n}}} + h\Big(e^{-c(\beta)\sqrt{n}}\Big) + 2h\left(\frac{2e^{-\gamma\sqrt{n}}}{1-e^{-c(\beta)\sqrt{n}}}\right), \tag{50}$$

and $h : [0,1] \to [0,1]$ is the binary entropy function. In particular, $\lim_{n\to\infty} \epsilon_n(\beta) = 0$ for all β values as above.

Proof: We first recall the classic result of Schonmann [24] stating that in the phase coexistence regime (i.e., when $\beta > \beta_c$), zero magnetization is highly atypical with respect to the Gibbs distribution. Namely, for each $\beta > \beta_c$

there exists $c(\beta) > 0$ such that

$$\pi\Big(\big\{\sigma\big|m(\sigma) = 0\big\}\Big) \le e^{-c(\beta)\sqrt{n}}. \tag{51}$$

Now, fix $\beta > \beta_0$ (which, in particular, is larger than β_c) and define $E = \mathbb{1}_{\{m(X_0)=0\}}$. Let $t \ge n \cdot e^{C\beta n^{\frac{1}{4}+\epsilon}}$ and consider

$$\begin{aligned} I(X_0; X_t) &= I(X_0, E; X_t) \\ &\le H(E) + I(X_0; X_t|E) \\ &\le h\Big(e^{-c(\beta)\sqrt{n}}\Big) + ne^{-c(\beta)\sqrt{n}} + I(X_0; X_t|E = 0) \end{aligned} \tag{52}$$

where the last inequality uses (51) twice. To upper bound the mutual information term on the RHS of (52), let $S = \mathsf{sign}\big(m(X_0)\big)$. We have

$$\begin{aligned} I(X_0; X_t|E = 0) &\le H(S|E = 0) + I(X_0; X_t|S, E = 0) \\ &\overset{(a)}{\le} \log 2 + \mathbb{P}\big(S = 1\big|E = 0\big) I(X_0; X_t|S = 1, E = 0) + \mathbb{P}\big(S = -1\big|E = 0\big) I(X_0; X_t|S = -1, E = 0) \\ &\overset{(b)}{\le} \log 2 + 2\mathbb{P}\big(S = 1\big|E = 0\big) I(X_0; X_t|S = 1, E = 0) \\ &\overset{(c)}{\le} \log 2 + I\big(X_0; X_t\big|m(X_0) > 0\big) \end{aligned} \tag{53}$$

where (a) is because $\mathbb{P}\big(S = 0\big|E = 0\big) = 0$ by definition, (b) uses the symmetry with respect to a global spin flip of the Gibbs measure and the SIM with free boundary conditions, while (c) is because $\{S = 1\} \cap \{E = 0\} = \big\{m(X_0) > 0\big\}$ and $\mathbb{P}\big(S = 1\big|E = 0\big) \le \frac{1}{2}$.

Next, let $G = \mathbb{1}_{\{X_t = Y_t\}}$ and approximate the mutual information from the RHS of (53) as

$$\begin{aligned} I\big(X_0; X_t\big|m(X_0) > 0\big) &\le I\big(X_0; X_t, G\big|m(X_0) > 0\big) \\ &\le H\big(G\big|m(X_0) > 0\big) + \mathbb{P}\big(G = 0\big|m(X_0) > 0\big) I\big(X_0; X_t\big|G = 0, m(X_0) > 0\big) \\ &\qquad + \mathbb{P}\big(G = 1\big|m(X_0) > 0\big) I\big(X_0; X_t\big|G = 1, m(X_0) > 0\big) \\ &\le h\Big(\mathbb{P}\big(X_t = Y_t\big|m(X_0) > 0\big)\Big) + \mathbb{P}\big(X_t \ne Y_t\big|m(X_0) > 0\big) I\big(X_0; X_t\big|X_t \ne Y_t, m(X_0) > 0\big) \\ &\qquad + \mathbb{P}\big(X_t = Y_t\big|m(X_0) > 0\big) I\big(X_0; X_t\big|X_t = Y_t, m(X_0) > 0\big). \end{aligned} \tag{54}$$

The conditional probability of interest is controlled using (48) and (51) as follows:

$$\begin{aligned} \mathbb{P}\big(X_t \ne Y_t\big|m(X_0) > 0\big) &= \sum_{\sigma \in \Omega_n} \mathbb{P}\big(X_0 = \sigma, X_t \ne Y_t\big|m(X_0) > 0\big) \\ &\overset{(a)}{=} \frac{1}{\mathbb{P}\big(m(X_0) > 0\big)} \sum_{\substack{\sigma \in \Omega_n: \\ m(\sigma) > 0}} \pi(\sigma)\big(X_t^\sigma \ne Y_t\big) \\ &\overset{(b)}{\le} \frac{1}{\frac{1}{2}\left(1 - e^{-c(\beta)\sqrt{n}}\right)} \sum_{\substack{\sigma \in \Omega_n: \\ m(\sigma) > 0}} \pi(\sigma)\big(X_t^\sigma \ne Y_t\big) \end{aligned}$$

$$\overset{(b)}{\leq} \frac{2e^{-\gamma\sqrt{n}}}{1-e^{-c(\beta)\sqrt{n}}} \tag{55}$$

where (a) uses the Markov property, (b) is due to $\pi(m>0)=\frac{1}{2}\pi(m\neq 0)$ and (51), and (c) uses (48). Denoting the RHS of (55) by p_n and inserting it back into (54) gives

$$\begin{aligned}
&I\big(X_0;X_t\big|m(X_0)>0\big)\\
&\leq h(p_n)+p_n I\big(X_0;X_t\big|X_t\neq Y_t,m(X_0)>0\big)+\mathbb{P}\big(X_t=Y_t\big|m(X_0)>0\big)I\big(X_0;X_t\big|X_t=Y_t,m(X_0)>0\big)\\
&\overset{(a)}{\leq} h(p_n)+np_n+\mathbb{P}\big(X_t=Y_t\big|m(X_0)>0\big)I\big(X_0;Y_t\big|X_t=Y_t,m(X_0)>0\big)\\
&\qquad\qquad +\mathbb{P}\big(X_t\neq Y_t\big|m(X_0)>0\big)I\big(X_0;Y_t\big|X_t\neq Y_t,m(X_0)>0\big)\\
&= h(p_n)+np_n+I\big(X_0;Y_t\big|G,m(X_0)>0\big)\\
&\leq h(p_n)+np_n+H\big(G\big|m(X_0)>0\big)+I\big(X_0;Y_t\big|m(X_0)>0\big)\\
&\leq 2h(p_n)+np_n+I\big(X_0;Y_t\big|m(X_0)>0\big)
\end{aligned} \tag{56}$$

(a) uses the non-negativity of mutual information.

To conclude the proof, we show that $I\big(X_0;Y_t\big|m(X_0)>0\big)=0$. Indeed, Y_t can be represented as a deterministic function of $Y_0=\boxplus$, the i.i.d. site-selection process $\big\{V_t\big\}_{t\in\mathbb{N}}$ and i.i.d. uniform process $\big\{U_t\big\}_{t\in\mathbb{N}}$. However, $\Big(Y_0,\big\{V_t\big\}_{t\in\mathbb{N}},\big\{U_t\big\}_{t\in\mathbb{N}}\Big)$ is independent of X_0, and therefore, so is Y_t. Thus,

$$I\big(X_0;X_t\big|m(X_0)>0\big)\leq 2h(p_n)+np_n, \tag{57}$$

which together with (52) and (53) concludes the proof.

■

VIII. Storing for Exponential Time in Inverse Temperature

At the positive but small temperature regime one may consider asymptotics not only in n, but also in β. Accordingly, we next show that a single stripe of plus-labeled sites (in a sea of minuses) at the bottom of an $\sqrt{n}\times\sqrt{n}$ grid retains at least half of its original $\sqrt{n}$ pluses for at least $e^{c\beta}$ time, where c is a numerical constant. This result is later generalized to stripes of width 2 inside the grid (i.e., not at the borders). Together, these two claims enable encoding information into sufficiently separated monochromatic stripes and extracting the written data after $e^{c\beta}$ time via majority decoding. As stated in Theorem 8 below, this achievability scheme implies that the information capacity is $\Omega(\sqrt{n})$, for all $t\leq e^{c\beta}$.

A. Single Stripe at the Bottom

We start with the single stripe analysis. Consider the continuous-time dynamics at inverse temperature $\beta>0$ with Poisson clocks of rate 1 associated with each $v\in\mathcal{V}_n$. Denoting the dynamics by $(X_t)_{t\geq 0}$, each time a clock rings, say at site v, the spin at v is refreshed according to the conditional Gibbs measure

$$\pi_v(s)=\pi\big(s\big|X_t(\mathcal{N}_v)\big), \tag{58}$$

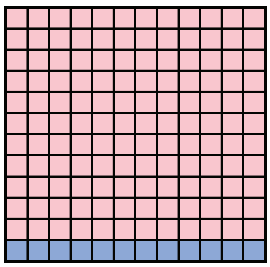


Fig. 7: Initial configuration, with blue and red squares marking plus- and minus-labeled sites, respectively.

where $s \in \{-1, +1\}$ and $\mathcal{N}_v$ is the neighborhood of v in $\mathcal{G}_n$ (see (4)). Calculating the update distribution, if the current configuration is σ and vertex v is selected, then the chance the spin at v is updated to -1 is

$$p(\sigma, v) = \phi\big(S(\sigma, v)\big), \tag{59}$$

where $\phi(a) \triangleq \frac{e^{a\beta}}{e^{a\beta}+e^{-a\beta}}$ and $S(\sigma, v) \triangleq \sum_{w:w\sim v} \sigma(w) = \big|\{w \sim v | \sigma(w) = -1\}\big| - \big|\{w \sim v | \sigma(w) = +1\}\big|$.

Consider initializing the system at $X_0 = \sigma$, with $\sigma\big((i, 1)\big) = +1$, for all $i \in [\sqrt{n}]$, and $\sigma(v) = -1$ otherwise; the initial configuration is shown in Fig. 7. Let $\mathcal{B} \triangleq \big\{(i, 1)\big\}_{i\in[\sqrt{n}]}$, define $N_1^{(+)}(\sigma) \triangleq \big|\{v \in \mathcal{B} | \sigma(v) = +1\}\big|$ and set $N_1^{(+)}(t) = N_1^{(+)}(X_t)$, for all $t \geq 0$. In words, $N_1^{(+)}(t)$ is the number of plus-labeled sites at the bottom strip after time t. We have the following lower bound on the expected value of $N_1^{(+)}(e^{c\beta})$.

Theorem 7 (Exponential Survival Time at the Bottom) *Fix any $c,' c \in (0, 1)$. There exist $\beta_0 > 0$ such that for any $\beta \geq \beta_0$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, we have*

$$\mathbb{E}N_1^{(+)}(t_0) \geq c'\sqrt{n}, \tag{60}$$

for $t_0 = e^{c\beta}$.

The full proof is given in Appendix G. Below we outline the main ideas in the proof, which comprises two parts: the *phase separation part* and the *analysis part*. To explain each, we start by highlighting the challenges in analyzing the evolution of $N_1^{(+)}(t)$. First, since $\beta < \infty$, pluses may in general spread out to the portion of G_n above the bottom stripe. However, since β is large, such flips are highly unlikely (exponentially small probability in β). Therefore, we circumvent this complication by simply restricting minus-labeled vertices from flipping. Doing so can only speed up the shrinking of $N_1^{(+)}(t)$, and thus, establishing (60) under this restricted dynamics suffices.

Even with this simplification, the main difficulty in the analysis of $N_1^{(+)}(t)$ is the unordered fashion in which flips of pluses into minuses at the bottom stripe occur. We distinguish between two main types of flips:

- **Sprinkle:** A flip of a plus-labeled vertex whose all horizontal neighbours are also pluses. The probability of a sprinkle update in the bulk of the bottom strip (i.e., excluding the two corners) is $\phi(-1)$.
- **Erosion:** A flip of a plus-labeled vertex with at least one horizontal neighbor that is a minus. The probability of an erosion update in the bulk of the bottom strip is either $\phi(1)$ or $\phi(3)$, depending on whether the updated

site has one or zero plus-labeled neighbors, respectively.

These two types of updates are interleaved with one another as the dynamics evolve. However, noting that sprinkles have exponentially small probability (in β), while erosion updates have probability exponentially close to 1, one would expect that during a certain initial time interval the system stays close to X_0 with occasional occurrences of sprinkles. Each sprinkle in the bulk results in two contiguous runs of pluses (abbreviated as a 'contigs') to its left and right. After a sufficient number of sprinkles, the drift of $N_1^{(+)}(t)$ is dominated by the erosion of the formed contigs. Arguing that the interleaved dynamics can be indeed separated into two pure phases of *sprinkling* and *erosion* is the first main ingredient of our proof. Once we are in the phase-separated dynamics, the analysis of $N_1^{(+)}(t)$ first identifies the typical length and number of contigs, and then studies how fast these contiguous pieces are eaten up.

B. Width-2 Stripe in the Bulk

Our coding scheme also needs a result similar to Theorem 7, but when the system is initialized at a configuration with one monochromatic stripe of width 2 somewhere in the bulk of the grid. Specifically, consider the same dynamics as before but with $X_0 = \sigma$, where $\sigma\big((i, j_0)\big) = \sigma\big((i, j_0 + 1)\big) = -1$, for all $i \in [\sqrt{n}]$ and some $j_0 \in [2 : \sqrt{n} - 2]$. Letting $N_{j_0}^{(+)}(t)$ be the number of pluses in this stripe at time t, we have the following corollary.

Corollary 2 (Exponential Survival Time in the Bulk) *Fix any $c,' c \in (0, 1)$. There exist $\beta_1 > 0$ such that for any $\beta \geq \beta_1$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, we have*

$$\mathbb{E}N_{j_0}^{(+)}(t_0) \geq c' \sqrt{n}, \tag{61}$$

for $t_0 = e^{c\beta}$.

This result follows by the same analysis as in the proof of Theorem 7 (see Appendix G), but with an additional preliminary step. This preliminary step is described next and the rest of the proof is omitted. The idea is to speed up the dynamics so that they correspond to the evolution of the system when initiated with a single bottom stripe (as in Fig. 7). As before, we start by prohibiting minus-labeled sites to flip. To circumvent the need to deal with the width dimension, every time a site is flipped we immediately also flip its vertical plus-labeled neighbor. Thus, flips occur in vertical pairs, with sprinkling and (modified) erosion probabilities as described below:

- **Sprinkling:** The sprinkling rate is $\phi(-2)$ (each site has 3 plus neighbors and one minus). Note that flipping the vertical neighbor of a sprinkle-flipped site only speeds up the elimination of pluses, without affecting the time scale. This is since the sites that are manually flipped have flip probability $\phi(0) = \frac{1}{2}$ (balanced neighborhood) in the original dynamics, which is much higher than $\phi(-2)$. Sprinkling in the modified dynamics is similar to sprinkling in the bottom stripe dynamics, as they both happen with probability exponentially small in β (namely, $\phi(-2)$ and $\phi(-1)$, respectively).
- **Erosion:** Every sprinkle produces two contigs. However, the sites at the borders of these contigs have flip rate $\phi(0)$, which is too slow compared to the bottom stripe case, where the erosion flip probability was $\phi(1)$.

Since we may keep speeding up the dynamics, we simply replace these $\phi(0)$ rates with $\phi(1)$. This modification produces erosion rates similar to the bottom stripe case.

Modifying the original dynamics as described above produces a new dynamics that behaves similarly to the bottom stripe ones. Then, one may repeat the arguments from the proof of Theorem 7 to produce the result of Corollary 2.

C. Stripe-Based Achievability Scheme

We are almost ready to state our achievability result. Beforehand, we need a one last corollary. Theorem 7 and Corollary 2 state their results in terms of expected values. Chebyshev's inequality readily translates them to claims on probability of a successful majority decoding.

Corollary 3 (Exponential Survival Time Probability) *For sufficiently large β and n (taken from Theorem 7 and Corollary 2), we have*

$$\mathbb{P}\left(N_{j_0}^{(+)}(t_0) > \frac{\sqrt{n}}{2}\right) \geq \frac{2}{3}, \quad \forall j_0 \in [1 : \sqrt{n} - 2]. \tag{62}$$

By symmetry to the $j_0 = 1$ case, this also holds for a horizontal stripe of width 1 at the top of the grid.

Proof: We only prove the statement for $j_0 = 1$. The expected value of $N_1^{(+)}(t_0)$ is controlled by Theorem 7. Since $N_1^{(+)}(t_0) \leq \sqrt{n}$ almost surely, we also have $\mathsf{var} \leq \left(1 - c'^2\right)\sqrt{n}$. Applying Chebyshev produces

$$\mathbb{P}\left(N_1^{(+)}(t_0) \leq \frac{\sqrt{n}}{2}\right) \leq \frac{4(1 - c'^2)}{2c' - 1}, \tag{63}$$

which can be made smaller than $\frac{1}{3}$ by taking c' as a constant close enough to 1. ■

The following is the information-theoretic result of this section. It shows that by coding over stripes and using majority decoding, one may store $\Theta(\sqrt{n})$ bits in the SIM at positive but small temperature for time $e^{c\beta}$. Let $I_n^{(\beta)}(t)$ be the information capacity of the $\sqrt{n} \times \sqrt{n}$ grid at inverse temperature β.

Theorem 8 (Storing $\sqrt{n}$ Bits) *There exist $\beta^\star > 0$ such that for any $\beta \geq \beta^\star$ there exists $n_0 \in \mathbb{N}$ such that for all $n > n_0$, we have*

$$I_n^{(\beta)}(t) = \Omega(\sqrt{n}), \tag{64}$$

for all $t \leq e^{c\beta}$, where $c \in (0, 1)$.

Proof: Partition the grid into monochromatic horizontal stripes (whose spins are specified later) such that:

1) the top and bottom stripes are of width 1;
2) intermediate stripes are of width 2;
3) the stripes are separated by all-minus walls of width 2 (larger widths are allowed only at the separation between the top stripe and the one right below it, in order to preserve a minimal distance of at least 2; for simplicity on notation, throughout the proof we assume that all these widths are exactly 2).

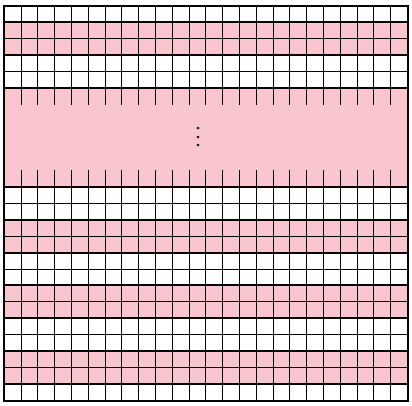


Fig. 8: A partitioning of the $\sqrt{n} \times \sqrt{n}$ grid into monochromatic striped regions of width 1 on the top and bottom and width 2 in between (shown in white). The regions are separated by all-minus walls of width 2. Red squares represent negative spins, while white squares stand for unspecified spins.

The partitioning is illustrated in Fig. 8. Let K be the total number of such stripes, and associate an index $j \in [K]$ with each from bottom to top. Clearly, $K = \Theta(\sqrt{n})$. For $j \in [K]$, let $\mathcal{S}_j$ be the set of vertices in the j-th stripe (the white squares in Fig. 8). Further set

$$\mathcal{D}_j \triangleq \mathcal{S}_j \cup \big\{u \in \mathcal{V}_n \big| d(u, \mathcal{S}_j) = 1\big\}. \tag{65}$$

Let $\mathcal{C}_n$ be the collection of all configurations whose topology corresponds to Fig. 8 with monochromatic spin assignments to each of the K stripes. Let $X_0 \sim p_{X_0}$ with $\mathrm{supp}(p_{X_0}) = \mathcal{C}_n$ and such that p_{X_0} is an i.i.d. Bernoulli $\frac{1}{2}$ measure on each stripe. For each $\sigma \in \Omega_n$, denote by $\sigma^{(j)}$ the restriction of σ to $\mathcal{S}_j$. For $\mathcal{J} \subseteq [K]$, we write $\sigma^{(\mathcal{J})}$ for $\big(\sigma^{(j)}\big)_{j \in \mathcal{J}}$. Similarly, we write $\bar{\sigma}^{(j)}$ for the restriction of σ to $\mathcal{D}_j$, and define $\bar{\sigma}^{(\mathcal{J})}$, for $\mathcal{J} \subseteq [K]$, analogously. With some abuse of notation, let $N_j^{(+)}(\sigma)$ be the number of plus labeled sites inside S_j. Furthermore, for each $j \in [K]$, let $\psi_j : \Omega_n \to \mathcal{S}_j$ be the majority decoder inside $\mathcal{S}_j$, i.e.,

$$\psi_j(\sigma) = \begin{cases} +1. & N_j^{(+)}(\sigma) \geq \frac{\sqrt{n}}{2} \\ -1, & N_j^{(+)}(\sigma) < \frac{\sqrt{n}}{2} \end{cases}. \tag{66}$$

Observe that the relation between X_0 and $\psi_j(X_t)$ is described by a binary channel which inputs a monochromatic stripe $X_0^{(j)}$ (drawn to be all-minus or all-plus with probability $\frac{1}{2}$ each), and outputs $+1$ if $N_j^{(+)}(t) \triangleq N_j^{(+)}(X_t) \geq \frac{\sqrt{n}}{2}$ and -1 if $N_j^{(+)}(t) < \frac{\sqrt{n}}{2}$. If $X_0 = \sigma$ with $\sigma^{(j)} = \boxplus \in \{-1, +1\}^{\mathcal{S}_j}$, then the crossover probability is $p_+^{(j)}(\sigma, t) \triangleq \mathbb{P}_\sigma\left(N_j^{(+)}(t) < \frac{\sqrt{n}}{2}\right)$, while if $X_0 = \sigma'$ with $\sigma'^{(j)} = \boxminus$, then it is $p_-^{(j)}(\sigma', t) \triangleq \mathbb{P}_{\sigma'}\left(N_j^{(+)}(t) \geq \frac{\sqrt{n}}{2}\right)$. Note that the transition probabilities are specified by the initial configuration through the entire region outside of $\mathcal{S}_j$. Thus, for each $j \in [K]$, any $\sigma_j^{\mathsf{out}} \triangleq \sigma^{[K]\setminus\{j\}} \in \{-1, +1\}^{\mathcal{V}\setminus\mathcal{S}_j}$ defines a binary (in general, asymmetric) channel from $\{\boxminus, \boxplus\} \subset \{-1, +1\}^{\mathcal{S}_j}$ to $\{-1, +1\}$ with the crossover probabilities given above.

For a each $j \in [K]$, let $\mathsf{T}_j : \{-1,+1\}^{\mathcal{S}_j} \to \Omega_n$ be a transformation defined by

$$\left(\mathsf{T}_j \sigma^{(j)}\right)(v) = \begin{cases} \sigma^{(j)}(v), & v \in \mathcal{S}_j, \\ -\sigma^{(j)}(u), & v \notin \mathcal{S}_j \end{cases} . \tag{67}$$

where u is the bottom left vertex in $\mathcal{S}_j$.[5] By monotonicity of the SIC (see beginning of Appendix F), we have that for any $j \in [K]$, $t \geq 0$ and $\sigma \in \mathcal{C}_n$, if $\sigma^{(j)} = \boxplus$ then $p_+^{(j)}(\sigma, t) \leq p_+^{(j)}(\mathsf{T}_j \sigma^{(j)}, t)$, while if $\sigma^{(j)} = \boxminus$ then $p_-^{(j)}(\sigma, t) \leq p_-^{(j)}(\mathsf{T}_j \sigma^{(j)}, t)$. This means that among all configurations $\sigma \in \mathcal{C}_n$ that agree on $\sigma^{(j)}$, $\mathsf{T}_j \sigma^{(j)}$ is the one that induces the highest crossover probability in the corresponding binary channel when $\sigma^{(j)}$ is transmitted. Furthermore, Corollary 3 states that for $t = t_0 \triangleq e^{c\beta}$, where $c \in (0,1)$, both these probabilities are upper bounded by $\frac{1}{3}$, which implies that the worst binary channel for each $\mathcal{S}_j$ region is the BSC with crossover probability $\frac{1}{3}$. The latter has a positive capacity of $\mathsf{C}_{\mathsf{BSC}}\left(\frac{1}{3}\right) = 1 - h\left(\frac{1}{3}\right)$, with $\mathsf{Ber}\left(\frac{1}{2}\right)$ as the capacity achieving distribution.

Combining these pieces we have

$$\begin{aligned} I_n^{(\beta)}(t) &\geq I(X_0; X_{t_0}) \\ &\overset{(a)}{=} \sum_{j \in [K]} I\left(\bar{X}_0^{(j)}; X_{t_0} \middle| \bar{X}_0^{[j-1]}\right) \\ &\overset{(b)}{\geq} \sum_{j \in [K]} I\left(\bar{X}_0^{(j)}; \psi_j(X_{t_0}) \middle| \bar{X}_0^{[j-1]}\right) \\ &\overset{(c)}{\geq} K \cdot \mathsf{C}_{\mathsf{BSC}}\left(\frac{1}{3}\right) \end{aligned} \tag{68}$$

where (a) is the mutual information chain rule, (b) uses $I(A;B) \geq I\big(A; f(B)\big)$ for any deterministic function f, while (c) is because the capacity of a binary (asymmetric) channel is a monotone decreasing function of both its crossover probabilities. Recalling that $K = \Theta(\sqrt{n})$ concludes the proof.

■

IX. Discussion and Future Directions

This work proposed a new model for the study of information storage in matter, which accounts for the interparticle interactions. The model relates the written and read configurations through the SIM (under the Glauber dynamics). The fundamental quantity we considered was the information capacity, $I_n(t)$, from (6). It was first shown that $I_n(t)$ is directly related to the maximal size of a reliable code for the operational setup.

Focusing on the 2D $\sqrt{n} \times \sqrt{n}$ grid at zero-temperature, the asymptotic behaviour of $I_n(t)$, when n and t grow together, was studied. First, we established that the infinite-time capacity of a finite grid is $\lim_{t\to\infty} I_n(t) = \Theta(\sqrt{n})$. Thus, order of $\sqrt{n}$ bits can be stored in the system indefinitely, while any information beyond that dissolves as $t \to \infty$. The key observation here was that striped configuration, of which there are $2^{\Theta(\sqrt{n})}$ many, are stable. By coding only over stripes one initiates the SIM at a state that it can not leave, and hence the information is recoverable

[5]This choice is arbitrary because subsequently T_j is only applied on portions of configuration from $\mathcal{C}_n$ and for $\sigma \in \mathcal{C}_n$, the $\mathcal{S}_j$ regions are monochromatic.

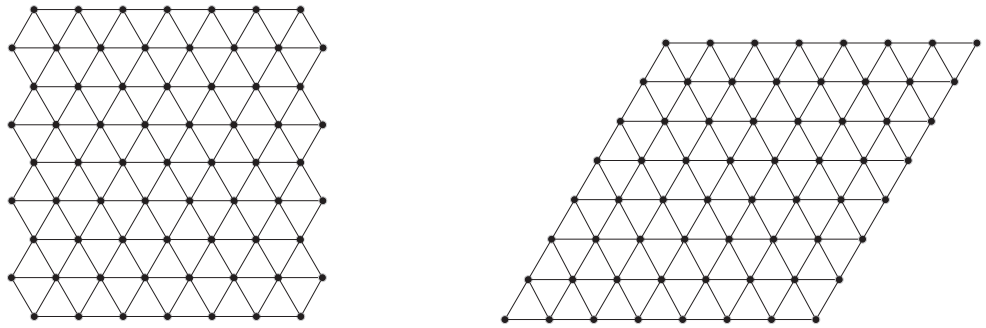
after arbitrarily large times. On the other hand, by establishing the zero-temperature dynamics as an absorbing MC, whose absorbing set consists of the striped configurations, we deduce that no non-striped pattern can survive forever. It would be very interesting to understand when the chain actually gets absorbed into stripes (with high probability). More precisely, how should t scale with n so as to ensure that $S_\sigma(t) \triangleq \mathbb{P}\big(X_t \text{ is not a stripe} \big| X_0 = \sigma\big)$ is close to 0, for any initial configuration σ? When the initial configuration is chosen according to an i.i.d. symmetric Bernoulli measure, this quantity is called the *survival probability* in the statistical physics literature. While some heuristics and numerical results are available for the survival probability (see, e.g., [16]), to the best of our knowledge, there are currently no rigorous bounds on it. Finding any time scale with respect to which $\max_\sigma S_\sigma(t)$ decays at least as fast as $\frac{1}{n}$ would immediately translate (via arguments similar to those in the converse proof Theorem 2) into an $O(\sqrt{n})$ upper bound on $I_n(t)$.

For finite time, simple arguments showed that up to $t \sim \frac{n}{4}$, the information capacity is $\Theta(n)$. A droplet-based coding scheme for superlinear time that reliably stores $\omega(\sqrt{n})$ bits ($\sqrt{n}$ can always be achieve by coding only over striped configuration) was then proposed. The idea was to partition the grid into appropriately spaced sub-squares of area $a(n) = o(n)$, and writing a bit onto each by setting all its spins to positive or negative. Using the Lifshitz law established in [17], each of the $\frac{n}{a(n)}$ sub-squares is capable of retaining its bit up to time $t \sim a(n) \cdot n$. Thus, we concluded that $I_n(t) = \Omega\left(\frac{n}{a(n)}\right)$ for $t = O\big(a(n) \cdot n\big)$. Note that this schemes improves upon the stripe-based scheme (that achieves order of $\sqrt{n}$ bits for all t) only up to $t \sim n^{\frac{3}{2}}$; indeed, taking $a(n) = \sqrt{n}$, we have $\frac{n}{a(n)} = \sqrt{n}$. The question remains whether or not it is possible to store more than $\sqrt{n}$ bits for times longer than $n^{\frac{3}{2}}$. A negative answer would require a converse bound, while a positive answer calls for a different achievability scheme.

We have examined the possibility of nesting the original sub-squares in the droplet-based scheme with secondary smaller squares, which would allow writing 2 bits into each isolated region (rather than 1, as in our original construction). While this would not result in an increased order of the number of stored bits (just a multiplicative factor of 2) it opens the door to an intriguing line of thought. Consider letting the number of nested squares also grow with n, where the smaller the square is, the slower the growth rate of its area. Such a scheme seems cumbersome to analyze, but there is a possibility that it enables storing more bits (in order of magnitude) than the vanilla droplet-based scheme proposed in this paper.

It was also shown that applying an arbitrarily small external field on the grid dynamics significantly improves its storage ability. The infinite-time capacity grows from $\Theta(\sqrt{n})$ to $\Theta(n)$, which is a consequence of a tie-braking rule induced by the external field. With this tie-braking rule (appropriately spaced) droplets of side length at least 2, of which there are $\Theta(n)$ many, are stable. Since n is always an upper bound on $I_n(t)$, the information capacity is characterized as $I_n(t) = \Theta(n)$, for all t. The same holds without an external field when the grid is replaced with the honeycomb lattice, whose structure prohibits ties in the first place. This suggests that both these architectures are superior to the grid without an external field for storage purposes.

For the triangular lattice on n vertices, where each vertex in the interior has degree 6, the size of the stable set is even smaller than that of the grid; in fact, it also depends on how the triangles are arranged in the plane. Specifically, considering the topology shown in Fig. 9(a), only horizontally striped configurations with stripes of width at least two are stable. The stable set is, thus, effectively half the size of the stable set of the corresponding



(a) (b)

Fig. 9: Two possible arrangements of the triangular lattice in the plane.

grid. For the topology from Fig. 9(b) matters are even worse: it can be shown that only the two ground states are stable (this is a consequence of the even degree of the vertices on the boundary). We conclude that lattices with odd and small degrees are preferable for storage. The *coordination number* is the chemical term that corresponds to the degree of the vertices in a graph. It refers to the number of the nearest neighbors of a central atom in a molecule. The observations above suggest that the coordination number of the molecules comprising the storage medium significantly impacts capacity.

We believe that the proposed model captures the essential features for the study of the fundamental limits of data storage inside matter. From a practical point of view, however, the study of the information capacity at positive small temperatures (below criticality) is an essential next step, which we are currently pursuing. While most of this work focused on the zero-temperature dynamics, we also provided preliminary results for the positive but small temperature regime. First, based on Proposition 5.2 from [22] we established a 1-bit upper bound on the information that can be stored into a initial configuration drawn according to the Gibbs measure for more than $e^{cn^{\frac{1}{4}+\epsilon}}$ time. This suggests that resilient configuration are not ones that are typical with respect to the Gibbs measure. Finding such configuration is of great interest to us. However, even their effect is limited as time progresses. It is known that in a finite box of side length L of the d-dimensional lattice $\mathbb{Z}^d$ with free boundary conditions, the relaxation to the Gibbs measure occurs on a time scale exponentially large in the surface L^{d-1} [25], [26]. Thus, for the 2D grid, the chain mixes after $e^{\Theta(\sqrt{n})}$ time, at which point all information stored in the system is lost. While no configuration is stable at non-zero temperature, stripes turn out to be a robust design methodology when scaling t with β. We proved that a stripe-based coding scheme (very similar to the scheme for indefinite storage in the zero-temperature dynamics) can retain its $\sqrt{n}$ bits for $t \sim e^{c\beta}$. This relies on a new result establishing that a monochromatic stripe in a sea of opposite spins retains at least half of its original spins for the aforementioned time-scale. Extensions to 3-dimensional topologies are of interest as well.

ACKNOWLEDGEMENTS

The authors would like to thank Yuzhou Gu for suggesting a simplified proof for Theorem 1, as appears in the paper. We also thank Elchanan Mossel for bringing to our attention the work of F. Martinelli [22].

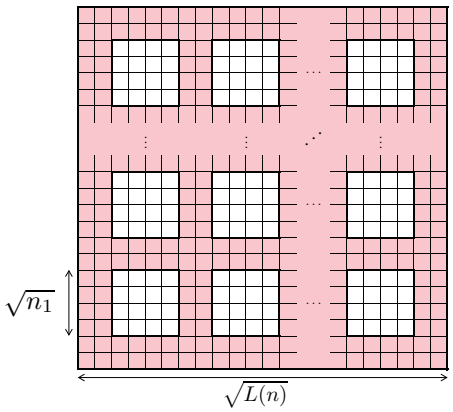


Fig. 10: A partition of the $\sqrt{L(n)} \times \sqrt{L(n)}$ grid into $\sqrt{n}$ smaller $\sqrt{n} \times \sqrt{n}$ grids separated by stripes of negative spins of width at least 2 (depicted in red).

APPENDIX A
PROOF OF PROPOSITION 1

For the upper bound, let $c_n^{(t)}$ be an $\big(M^\star(n,t,\epsilon), n, t, \epsilon\big)$-code for the $\mathsf{SIC}_n(t)$. We have

$$\log M^\star(n,t,\epsilon) = H(M) \overset{(a)}{\leq} I(X_0;X_t) + H(M|X_t) \overset{(b)}{\leq} I_n(t) + h(\epsilon) + \epsilon \log M^\star(n,t,\epsilon), \tag{69}$$

where (a) follows by the DPI as $M - X_0 - X_t$ forms a Markov chain, and (b) uses Fano's inequality. Rearranging (69) produces the upper bound from (9).

For the lower bound, we show how the $\mathsf{SIC}_n(t)$ (which is by no means a memoryless channel) can be used to support roughly $\frac{n}{n_1}$ uses of a DMC with capacity $I_{n_1}(t)$. Having this, we invoke the finite-blocklength achievability bound from [27] to obtain the desired lower bound.

The subsequent derivation can be thought of a the spatial analog of the classic argument that splits the transmission over a DMC into temporal blocks. In contrast to the DMC scenario, splitting the transmission over the $\mathsf{SIC}_n(t)$ into temporal blocks is impossible. This is since the encoder $f_n^{(t)}$ controls only the initial configuration X_0 and cannot access or influence the channel in subsequent times. A block-based transmission of an alternative nature is, therefore, required, which brings us to the spatial splitting presented next.

We outline the main ideas of the proof and refer the reader to the proof of Theorem 4, where a similar construction is used, for which a full formalization is provided. Let the total grid be of side length $\sqrt{L(n)} \triangleq \sqrt{\frac{n}{n_1}}(\sqrt{n}_1+2)+2$, so $L(n) = n + o\left(\frac{n}{\sqrt{n_1}}\right)$. An $\sqrt{L(n)} \times \sqrt{L(n)}$ grid can accommodate $\frac{n}{n_1}$ sub-grids of size n_1 separated from one another and from the boundaries by (horizontal and vertical) walls of width at least 2 of all-minus spins (see Fig. 10).

The separating walls decorrelate the different $\sqrt{n_1} \times \sqrt{n_1}$ sub-grids.[6] Each sub-grid constitutes a spatial block,

[6]For this statement to be precise one should consider the continuous-time version of the zero-temperature dynamics, as defined in Section V-C, rather than the discrete-time one. However, as the discrete-time and the continuous-time chains are equivalent in any sense of interest (see, Proposition 4), and because we only outline the main ideas of the proof of Proposition 1, this subtlety is ignored here.

of which there are $\frac{n}{n_1}$ many. Each block can be viewed as an access to the DMC $\mathsf{SIC}_{n_1}(t)$ (over the super-alphabet Ω_{n_1}). From the finite-block length achievability result found in [27], where $\frac{n}{n_1}$ is the number of channel uses and $I_{n_1}(t)$ is the capacity, we have

$$\log M^\star\big(L(n),t,\epsilon\big) \geq \frac{n}{n_1} I_{n_1}(t) - \sqrt{\frac{nV_{n_1}(t)}{n_1(1-\epsilon)}}. \tag{70}$$

Here $V_{n_1}(t) = \mathsf{var}\left(\log \frac{p_{X_0,X_t}(X_0,X_t)}{p_{X_0}(X_0)p_{X_t}(X_t)}\right)$ is the dispersion of the $\mathsf{SIC}_{n_1}(t)$. Using [27, Theorem 50], the dispersion is upper bounded as $V_{n_1}(t) \leq 2n_1^2$. Inserting this into (70) and normalizing by n produces (10).

APPENDIX B
PROOF OF PROPOSITION 2

Denote $k = \sqrt{n}$ and let a_k be the number of valid horizontally striped configuration. Due to symmetry, we have $|\mathcal{A}_{k^2}| = 2a_k$. Note that there is a bijection between horizontally striped configurations and binary strings of length k, where all runs on 0's or 1's are of length at least two. Henceforth, a_k is thought of as the number of such binary strings.

To count a_k, observe that a valid string of length $k+2$ either ends in a run of length exactly 2 or a run of length greater than 2. In the former case, the $(k+2)$-lengthed string can be any valid string of length k, followed by two characters that are opposite to the last character of the k-lengthed string. In the latter case, the string of length $k+2$ can be any valid string of length $k+1$ with its last character duplicated. Hence, we have the Fibonacci recursive relation

$$a_{k+2} = a_{k+1} + a_k, \quad \forall k \geq 2. \tag{71}$$

Along with the initial conditions $a_2 = 2$ and $a_3 = 2$, the above relation implies that a_k is $(k-1)$-th Fibonacci number (on the indexing where $f_0 = 0$ and $f_1 = 1$, respectively). Recalling that

$$f_k = \frac{\phi^k - \psi^k}{\sqrt{5}}, \tag{72}$$

where $\phi = \frac{1+\sqrt{5}}{2}$ and $\psi = -\frac{1}{\phi}$, we have that $|\mathcal{A}_{k^2}| = 4f_{k-1}$.

APPENDIX C
PROOF OF THEOREM 1

The inclusion $\mathcal{A}_n \subseteq \mathcal{S}_n$ is straightforward, as any striped configuration $\sigma \in \mathcal{A}_n$ satisfies $P(\sigma,\sigma) = 1$, whence it is stable. This follows because if $\sigma \in \mathcal{A}_n$, then for any $v \in \mathcal{V}_n$, $m_v(\sigma) > \ell_v(\sigma)$, i.e., $\sigma(v)$ coincides with the majority of its neighbors.

For the opposite inclusion, fix a stable configuration $\sigma \in \mathcal{S}_n$. We construct a certain subgraph $\mathcal{G}_n^{(\sigma)}$ of the dual lattice that depends on σ and use it to establish the inclusion. First, let $\tilde{\mathcal{G}}_n^{(\sigma)} = \left(\mathcal{V}_n, \tilde{\mathcal{E}}_n^{(\sigma)}\right)$ be a subgraph of $\mathcal{G}_n$, with the same vertex set and an edge set

$$\tilde{\mathcal{E}}_n^{(\sigma)} = \Big\{\{v,u\} \in \mathcal{E}_n \big| \sigma(u) \neq \sigma(v)\Big\}. \tag{73}$$

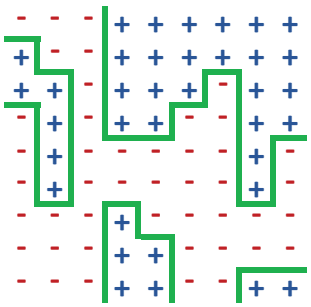


Fig. 11: For a given configuration $\sigma \in \Omega_n$, the edges of the graph $\mathcal{G}_n^{(\sigma)}$ are shown as the green lines. These edges form curves that separate regions of opposite spins in σ.

Namely, $\tilde{\mathcal{E}}_n^{(\sigma)}$ contains only edges connecting sites with opposite spins in σ.

From $\tilde{\mathcal{G}}_n^{(\sigma)}$ we construct the $\mathcal{G}_n^{(\sigma)} = \left(\hat{\mathcal{V}}_n, \mathcal{E}_n^{(\sigma)}\right)$, where $\hat{\mathcal{V}}_n$ is the $(\sqrt{n}+1) \times (\sqrt{n}+1)$ grid in the dual lattice that covers $\mathcal{V}_n$, i.e., $\hat{\mathcal{V}}_n = \bigcup_{i,j\in\{-1,+1\}} \left\{\mathcal{V}_n + \left(\frac{i}{2}, \frac{j}{2}\right)\right\}$. The edge set of $\mathcal{G}_n^{(\sigma)}$ is as follows: for any $x, y \in \hat{\mathcal{V}}_n$ with $||x-y||_1 = 1$, $\{x,y\} \in \mathcal{E}_n^{(\sigma)}$ if and only if there exists $e \in \tilde{\mathcal{E}}_n^{(\sigma)}$ that crosses $\{x,y\}$. In other words, viewing edges as segments in $\mathbb{R}^2$ of length 1, the above condition is equivalent to $\{x,y\} \cap e \neq \emptyset$. In particular, this construction implies that vertices on the boundary of $\hat{\mathcal{V}}_n$ are never connected to one another. The edges of $\mathcal{G}_n^{(\sigma)}$ separate regions of opposite spins in σ (see Fig. 11 for an illustration).

Denote the degree of a vertex x of $\mathcal{G}_n^{(\sigma)}$ by $\deg_\sigma(x)$. A first trivial observation is that for any $\sigma \in \Omega_n$ (not necessarily stable) and $x \in \hat{\mathcal{V}}_n$ that is not on the boundary of $\hat{\mathcal{V}}_n$, we have that $\deg_\sigma(x)$ is even (just consider the possible cases for local spin labels in σ). Furthermore, since $\sigma \in \mathcal{S}_n$ we have that the degree of any vertex as above cannot be 4, i.e., $\deg_\sigma(x) \in \{0,2\}$. To see this note that $\deg_\sigma(x) = 4$ is possible only if $\sigma\left(x + \left(\frac{1}{2}, \frac{1}{2}\right)\right) = \sigma\left(x - \left(\frac{1}{2}, \frac{1}{2}\right)\right) \neq \sigma\left(x + \left(\frac{1}{2}, -\frac{1}{2}\right)\right) = \sigma\left(x + \left(-\frac{1}{2}, \frac{1}{2}\right)\right)$. This labeling, however, contradicts the stability of σ (each of these four site, if selected, has probability of at least $\frac{1}{2}$ to flip).

Next, note that for a vertex $x \in \hat{\mathcal{V}}_n$ with $\deg_\sigma(x) = 2$, the two edges containing x must have the same orientation. Namely, both edges must be vertical or horizontal. Indeed, if, for instance, the edges containing x are $\{x,y\}, \{x,z\} \in \mathcal{E}_n^{(\sigma)}$ with $y = x+(1,0)$ and $z = x+(0,1)$, then we have that $\sigma\left(x + \left(\frac{1}{2}, \frac{1}{2}\right)\right) \neq \sigma\left(x + \left(\frac{1}{2}, -\frac{1}{2}\right)\right) = \sigma\left(x + \left(-\frac{1}{2}, \frac{1}{2}\right)\right)$, which contradicts stability. The three other cases of edges with disagreeing orientation are treated similarly.

One last observation regarding $\mathcal{E}_n^{(\sigma)}$ is that it contains no two parallel edges such that the shortest (Euclidean) distance between them is 1. Assume to the contrary that $e_1, e_2 \in \mathcal{E}_n^{(\sigma)}$ are parallel with minimal distance 1. Without loss of generality assume that both these edges are horizontal (otherwise repeat the same argument with a 90° rotation of the grid) and that e_2 is above e_1; this means that $e_1 = \{x,y\}$ and $e_2 = \{x+(0,1), y+(0,1)\}$, where $x \in \hat{\mathcal{V}}_n$ and $y = x+(1,0)$. As a result, we have that $\sigma\left(x + \left(\frac{1}{2}, \frac{1}{2}\right)\right) \neq \sigma\left(x - \left(\frac{1}{2}, \frac{1}{2}\right)\right) = \sigma\left(x + (0,1) + \left(\frac{1}{2}, \frac{1}{2}\right)\right)$, which compromises the stability of σ.

Combining the above three observations we see that each connected component of $\mathcal{G}_n^{(\sigma)}$ is a straight (horizontal or vertical) crossing of vertices in $\hat{\mathcal{V}}_n$ that extends between two opposite boundaries. Clearly, horizontal and vertical

crossings cannon coexists, as otherwise the vertex x which is contained in them both has $\deg_\sigma(v) = 4$, a case that was already ruled out. Having this, we deduce that σ must have the structure of monochromatic horizontal or vertical stripes of width at least two, except, possibly, the first and last stripes that may still have width 1. To see that this is impossible, assume that σ is horizontally striped but that the stripe at height $j = 1$ is of width 1. Assume further that $\sigma\left(\{(i,1)\}_{i\in[n]}\right) = +1$ and that $\sigma\left(\{(i,2)\}_{i\in[n]}\right) = -1$ (otherwise, switch the roles of $+1$ and -1). Noting that $|\mathcal{N}_{(1,1)}| = 2$, but, $m_{(1,1)}(\sigma) = 1$, implies that $(1,1)$ is an unstable site.

APPENDIX D
PROOF OF LEMMA 1

For any $\sigma \in \mathcal{T}_n$ we construct a path to $\mathcal{S}_n$ with a finite number of steps. First, for each $\sigma \in \Omega_n$ and $v \in \mathcal{V}_n$, let

$$\mathcal{C}_v(\sigma) \triangleq \left\{u \in \mathcal{V}_n \middle| \exists m \in \mathbb{N},\ \exists\{u_k\}_{k=0}^m \subseteq \mathcal{V}_n,\ u_0 = v,\ u_m = u,\ u_{k-1} \sim u_k \text{ and } \sigma(u_k) = \sigma(v),\ \forall k \in [m]\right\} \quad (74)$$

be the monochromatic connected component of v in σ. Define the *external boundary* of $\mathcal{C}_v(\sigma)$ as

$$\partial_{\mathsf{Ext}}\mathcal{C}_v(\sigma) \triangleq \left\{u \in \mathcal{V}_n \setminus \mathcal{C}_v(\sigma) \middle| d\big(u, \mathcal{C}_v(\sigma)\big) = 1\right\}. \quad (75)$$

We start with some operations on configurations that constitute the building blocks of our proof.

A. Expansion Connected Components

We describe an operation on configurations through a finite sequence of single-site flips of positive probability in the zero-temperature dynamics. To do so, for any $\sigma \in \Omega_n$, define the sets

$$\mathcal{N}_a(\sigma) \triangleq \left\{v \in \mathcal{V}_n \middle| m_v(\sigma) > \frac{|\mathcal{N}_v|}{2}\right\} \quad (76a)$$

$$\mathcal{N}_d(\sigma) \triangleq \left\{v \in \mathcal{V}_n \middle| m_v(\sigma) < \frac{|\mathcal{N}_v|}{2}\right\}. \quad (76b)$$

These sets contain, respectively, all the vertices in $\mathcal{V}_n$ that agree or disagree with the majority of their neighbors under the configuration σ. Also let $\mathcal{N}_u(\sigma) \triangleq \mathcal{V}_n \setminus \left\{\mathcal{N}_a(\sigma) \cup \mathcal{N}_d(\sigma)\right\}$ be the set of undecided vertices, i.e., all $v \in \mathcal{V}_n$ (with an even number of neighbors) such that $m_v(\sigma) = \frac{|\mathcal{N}_v|}{2}$. Denote the sizes of the above sets by:

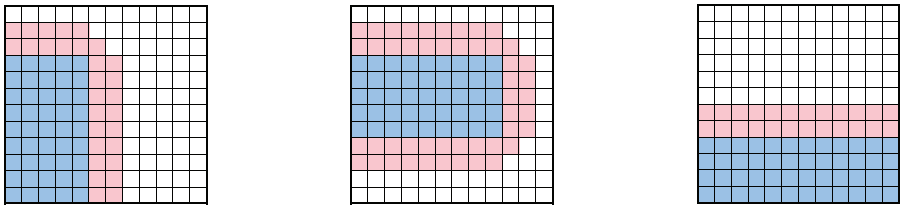
$$N_a(\sigma) \triangleq \left|\mathcal{N}_a(\sigma)\right| \ \ ; \ \ N_d(\sigma) \triangleq \left|\mathcal{N}_d(\sigma)\right| \ \ ; \ \ N_u(\sigma) \triangleq \left|\mathcal{N}_u(\sigma)\right|. \quad (77)$$

Let $\sigma \in \Omega_n$ be any configuration and $v \in \mathcal{V}_n$ be a vertex on the left border of $\mathcal{V}_n$, i.e., $v \in \mathcal{L} \triangleq \left\{(1,j)\right\}_{j\in[\sqrt{n}]}$. Consider the following pseudo-algorithm that takes σ and v as inputs and outputs a new configuration η reachable from σ by a finite path.

Since the grid in finite, Algorithm 1 terminates in a finite number of steps. Once it does, $\partial_{\mathsf{Ext}}\mathcal{C}_v(\eta)$ contains no unstable nor disagreeing vertices. Denoting the mapping described by the algorithm by $\mathsf{E} : \Omega_n \times \mathcal{L} \to \Omega_n$ (it can be verified that the mapping is well defined despite the random choice of w in Step 4), for any $\sigma \in \Omega_n$ and $v \in \mathcal{L}$,

Algorithm 1 Single-flip expansion of $\mathcal{C}_v(\sigma)$

$\zeta \leftarrow \sigma$
while $\partial_{\mathsf{Ext}}\mathcal{C}_v(\zeta) \cap \Big\{\mathcal{N}_u(\zeta) \cup \mathcal{N}_d(\zeta)\Big\} \neq \emptyset$ **do**
    Draw $w \sim \mathsf{Unif}\left(\partial_{\mathsf{Ext}}\mathcal{C}_v(\zeta) \cap \Big\{\mathcal{N}_u(\zeta) \cup \mathcal{N}_d(\zeta)\Big\}\right)$
    $\zeta \leftarrow \zeta^w$
$\eta \leftarrow \zeta$



(a) (b) (c)

Fig. 12: Rectangular configurations, where blue and red squares represent opposite spins, while white squares stand for unspecified spins.

the connected component $\mathcal{C}_v(\eta)$, where $\eta = \mathsf{E}(\sigma, v)$, has a rectangular shape. Namely, if $v = (1, j) \in \mathcal{L}$, then there exist $1 \leq j_f^{(1)} \leq j \leq j_f^{(2)} \leq \sqrt{n}$ and an $i_f \in [\sqrt{n}]$ such that

$$\mathcal{C}_v(\eta) = \Big\{(i, j) \in \mathcal{V}_n \Big| i \in [i_f],\ j \in [j_f^{(1)} : j_f^{(2)}]\Big\}. \tag{78}$$

This is so because any non-rectangular connected component has flippable spins on its external boundary, which contradicts the fact that $\partial_{\mathsf{Ext}}\mathcal{C}_v(\eta) \cap \Big\{\mathcal{N}_u(\eta) \cup \mathcal{N}_d(\eta)\Big\} = \emptyset$.

Furthermore, observe that $\eta = \mathsf{E}(\sigma, v)$ is such that any $u \in \mathcal{V}_n$ with $d\big(u, \mathcal{C}_v(\eta)\big) \leq 2$ satisfies $\eta(u) = -\eta(v)$. Assuming $\mathcal{C}_v(\eta) \neq \mathcal{V}_n$, this statement is trivial for sites u at graph distance 1 from $\mathcal{C}_v(\eta)$ (just by the definition of $\mathsf{E}(\sigma, v)$). To see why this holds for sites with $d\big(u, \mathcal{C}_v(\eta)\big) = 2$, let u be such a site but with $\eta(u) = \eta(v)$. If there exists $w \in \mathcal{C}_v(\eta)$ with $w = u - (0, 2)$ (i.e., u is right above $\mathcal{C}_v(\eta)$), then $w + (0, 1) \in \partial_{\mathsf{Ext}}\mathcal{C}_v(\eta) \cap \Big\{\mathcal{N}_u(\eta) \cup \mathcal{N}_d(\eta)\Big\}$, in contradiction to the termination rule of Algorithm 1. Similarly, if $u = w + (2, 0)$, for some $w \in \mathcal{C}_v(\eta)$, then the site $w + (1, 0)$ disagrees with at least half of its neighbors. Finally, if $u = w + (1, 1)$ for some $w \in \mathcal{C}_v(\eta)$ (the case when $u = w - (1, 1)$ is treated similarly), then both $w + (0, 1)$ and $w + (1, 0)$ are unstable, which again leads to a contradiction. Fig. 12 shows three examples of possible shapes of $\mathcal{C}_v(\eta)$ with their corresponding surroundings.

B. Flipping Rectangles

Let $v = (1, j_v) \in \mathcal{L}$, $h \in [\sqrt{n} - j_v - 1]$ and $\ell \in [\sqrt{n} - 2]$. Define $\mathcal{R}_{h,\ell}^{(v)}$ as the set of all configurations $\sigma \in \Omega_n$ satisfying:

i) $\mathcal{C}_v(\sigma) = \big\{v + (i, j) \big| i \in [0 : \ell],\ j \in [0 : h]\big\}$;
ii) $\mathcal{C}_v(\sigma) \cap \mathcal{N}_u(\sigma) \neq \emptyset$.

Namely, $\mathcal{R}_{h,\ell}^{(v)}$ is the set of all configuration in which the connected monochromatic component of v is an $h \times \ell$ rectangle with v as it bottom-left corner, and such that at least two of the rectangle's sides are bounded away from

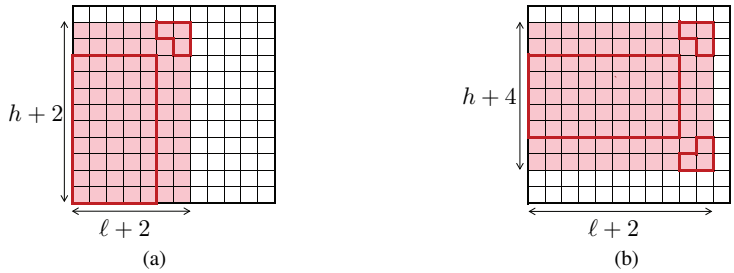


Fig. 13: Figs. (a) and (b) show, respectively, the operation of $\mathsf{F}_{h,\ell}^{(v)}$ on the configurations from Figs. 12(a) and 12(b). The regions outlined in red are the only ones that $\mathsf{F}_{h,\ell}^{(v)}$ affected.

the border of the grid (which is condition (ii) in the definition of $\mathcal{R}_{h,\ell}^{(v)}$). Further let $\bar{\mathcal{R}}_{h,\ell}^{(v)}$ and $\underline{\mathcal{R}}_{h,\ell}^{(v)}$ be the subsets of $\mathcal{R}_{h,\ell}^{(v)}$ in which the spin at v is $+1$ or -1, respectively.

Let $\mathsf{F}_{h,\ell}^{(v)} : \mathcal{R}_{h,\ell}^{(v)} \to \Omega_n$ be a mapping from $\mathcal{R}_{h,\ell}^{(v)}$ to Ω_n that flips all the spins inside $\mathcal{C}_v(\sigma)$ and sets the spins at the missing corners of $\mathcal{C}_v(\sigma) \cup \partial_{\mathsf{Ext}}\mathcal{C}_v(\sigma)$ to $-\sigma(v)$. Formally, let

$$\mathcal{M}_v(\sigma) \triangleq \Big\{\mathcal{C}_v(\sigma) \cap \mathcal{N}_u(\sigma)\Big\} + \big\{(1,2),(2,1),(2,2),(1,-2),(2,-1),(2,-2)\big\} \tag{79}$$

be a set (defined as a Minkowski sum) that contains the missing corners of $\mathcal{C}_v(\sigma)$ (and some other sites). The mapping $\mathsf{F}_{h,\ell}^{(v)}$ is given by

$$\left(\mathsf{F}_{h,\ell}^{(v)}\sigma\right)(w) = \begin{cases} -\sigma(v), & w \in \mathcal{C}_v(\sigma) \cup \mathcal{M}_v(\sigma) \\ \sigma(u), & \text{otherwise} \end{cases} . \tag{80}$$

Fig. 13 shows the outcome of applying $\mathsf{F}_{h,\ell}^{(v)}$ on the configurations from Figs. 12(a)-(b). Note that if $\sigma \in \mathcal{R}_{h,\ell}^{(v)}$, then $\mathcal{C}_v\left(\mathsf{F}_{h,\ell}^{(v)}\sigma\right)$ contains a rectangle of size at least $(h+2)\times(\ell+2)$.

Finally, note the for any $\sigma \in \mathcal{R}_{h,\ell}^{(v)}$ we have $\sigma \rightsquigarrow \mathsf{F}_{h,\ell}^{(v)}\sigma$, i.e., there exists a finite path between σ and $\mathsf{F}_{h,\ell}^{(v)}\sigma$. A path $\omega : \sigma \rightsquigarrow \mathsf{F}_{h,\ell}^{(v)}\sigma$ can be constructed by first flipping any unstable corner of the $h \times \ell$ rectangle, then flipping the neighbor of that flipped corner (any neighbor will do), followed by flipping their neighbors, etc. This is done until an $(h-1)\times(\ell-1)$ rectangle is reached and the process repeats itself until the entire original rectangle is flipped. Finally, if needed, the spins sites $u \in \mathcal{M}_v(\sigma)$ are flipped as well.

We describe a construction of such a path $\omega : \sigma \rightsquigarrow \mathsf{F}_{h,\ell}^{(v)}\sigma$ for a configuration $\sigma \in \mathcal{R}_{h,\ell}^{(v)}$, with $v = (1,1)$ and $h,\ell \in [\sqrt{n}-2]$ (σ is as shown in Fig. 12(a)). Set $\omega_0 = \sigma$ and

$$\omega_{j\ell+i+1} = \omega_{j\ell+i}^{(\ell-i,h-j)}, \quad i \in [0:\ell-1],\ j \in [0:h-1], \tag{81}$$

with (possibly) additional three steps $\omega_{h\ell+1} = \omega_{h\ell}^{(\ell+1,h+2)}$, $\omega_{h\ell+2} = \omega_{h\ell+1}^{(\ell+2,h+1)}$ and $\omega_{h\ell+3} = \omega_{h\ell+2}^{(\ell+2,h+2)}$, for any of these three sites whose initial spin was $-\sigma(v)$. By similar path constructions it follows that there exists a finite path (that includes at most $h\ell+3$ single-site flips, each with positive probability in the zero-temperature dynamics) $\omega : \sigma \rightsquigarrow \mathsf{F}_{h,\ell}^{(v)}\sigma$, for all $\sigma \in \mathcal{R}_{h,\ell}^{(v)}$, where $v \in \mathcal{L}$ is any vertex on the left border of $\mathcal{V}_n$ and h,ℓ are any valid lengths.

C. Proof of Lemma 1

Let $\sigma \in \mathcal{T}_n$ and consider the following pseudo-algorithm that transforms σ to a new configuration $\eta \in \mathcal{S}_n$. The transformation is described only by means of the mappings E, and $\{\mathsf{F}_{h,\ell}^{(v)}\}$ defined above. Since for each of these mappings there exists a path of single-spin flips (of positive probability) from the input to the output, Algorithm 2 establishes the connectivity argument from Lemma 1.

Given a configuration $\sigma \in \Omega_n$ and a vertex $v \in \mathcal{V}_n$ such that the monochromatic connected component $\mathcal{C}_v(\sigma)$ is a rectangle (stripes are considered rectangles), let $\mathsf{h}(v,\sigma)$ $\mathsf{l}(v,\sigma)$ be the functions that return the height and length of $\mathcal{C}_v(\sigma)$; if $\mathcal{C}_v(\sigma)$ is not a rectangle, set $\mathsf{h}(v,\sigma) = \mathsf{l}(v,\sigma) = 0$.

Algorithm 2 Path construction from $\mathcal{T}$ to $\mathcal{S}$

1: $\zeta \leftarrow \sigma$
2: $v \leftarrow (1,1)$
3: **while** $\zeta \notin \mathcal{S}_n$ **do**
4: $\quad \zeta \leftarrow \mathsf{E}(\zeta, \mathsf{v})$ ▷ Expand $\mathcal{C}_v(\zeta)$ to the maximal possible rectangle
5: $\quad h = \mathsf{h}(v,\zeta)$ ▷ Height of the rectangle
6: $\quad \ell = \mathsf{l}(v,\zeta)$ ▷ Length of the rectangle
7: $\quad$ **if** $\max(h,\ell) = \sqrt{n}$ and $\min(h,\ell) < \sqrt{n}$ **then** ▷ If $\mathcal{C}_v(\zeta)$ is a stripe
8: $\quad\quad$ **if** $\ell < \sqrt{n}$ **then** ▷ Horizontal stripe
9: $\quad\quad\quad v \leftarrow (1, h+1)$ ▷ Move v above stripe
10: $\quad\quad$ **if** $\ell < \sqrt{n}$ **then** ▷ Vertical stripe
11: $\quad\quad\quad v \leftarrow (\ell+1, 1)$ ▷ Move v to the right of the stripe
12: $\quad$ **if** $\max(h,\ell) < \sqrt{n}$ **then** ▷ If $\mathcal{C}_v(\zeta)$ is a rectangle bounded away from the top and right borders
13: $\quad\quad \zeta \leftarrow \mathsf{F}_{h,\ell}^{(v)}\zeta$ ▷ Flip the rectangle to create a larger rectangle with an opposite spin
14: **if** $\mathsf{h}\big(\zeta,(1,1)\big) = 1$ or $\mathsf{l}\big(\zeta,(1,1)\big) = 1$ **then** ▷ If the first stripe is of width 1
15: $\quad$ **if** $\mathsf{h}\big(\zeta,(1,1)\big) = 1$ **then** ▷ Horizontal stripe of height 1
16: $\quad\quad \zeta(i,1) \leftarrow -\zeta(i,1), \quad \forall i \in [\sqrt{n}]$ ▷ Flip the spins at the bottom horizontal stripe of width 1
17: $\quad$ **if** $\mathsf{l}\big(\zeta,(1,1)\big) = 1$ **then** ▷ Vertical stripe of length 1
18: $\quad\quad \zeta(1,j) \leftarrow -\zeta(1,j), \quad \forall j \in [\sqrt{n}]$ ▷ Flip the spins at the leftmost vertical stripe of width 1
19: $\eta \leftarrow \zeta$ ▷ Output

Algorithm 2 transforms the original $\sigma \in \mathcal{T}_n$ to $\eta \in \mathcal{S}_n$ by first expanding the connected component of $v = (1,1)$ an $h \times \ell$ rectangle. If the rectangle is a stripe, i.e., if $h = \sqrt{n}$ or $\ell = \sqrt{n}$ (in case that both equalities hold then $\mathsf{E}\big(\sigma,(1,1)\big)$ is either $\boxplus$ or $\boxminus$ and we are done), then v is shifted to be the bottom-left corner of the rest of the grid and the process starts over. Otherwise, i.e., if $\mathsf{E}\big(\sigma,(1,1)\big)$ is a rectangle bounded away from the top and right borders of the grid, then we use map $\mathsf{F}_{h,\ell}^{(1,1)}$ to flip the rectangle and get a new configuration where the connected component of $(1,1)$ contains a rectangle of size at least $(h+2)\times(\ell+2)$, and the process repeats itself. The finiteness of the grid $\mathcal{V}_n$ ensures that the **while** loop in Algorithm 2 (Steps (3)-(13)) terminates after a finite number of steps.

Once the loop terminates, the output configuration ζ is striped. However, the connected component of $(1,1)$ in ζ may be a stripe of width 1 (if, e.g., in the initial configuration $\mathcal{C}_{(1,1)}(\sigma) = \big\{(i,1)\big\}_{i\in[\sqrt{n}]}$ and $\sigma\big((i',j)\big) = -\sigma(v)$ for all $i' \in [\sqrt{n}]$ and $j \in [2:3]$, then $\mathsf{E}(\sigma,v) = \sigma$), which is unstable. To fix this, Steps (14)-(18) check if this is the case and flip the spins of that stripe to correspond with those in the adjacent stripe (directly on top or to

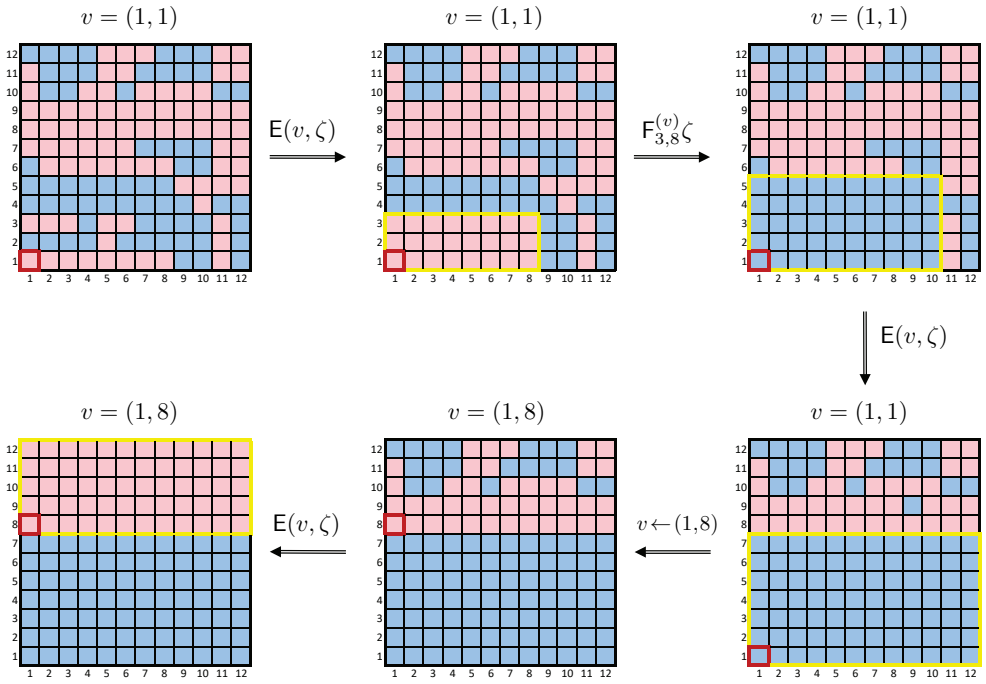


Fig. 14: Illustration of Algorithm 2. In each figure, the square outlined in red represents the current v (which is also indicated on the top of each grid). The yellow line encloses the portion of the grid that was modified from the previous step.

the right of $\mathcal{C}_{(1,1)}(\zeta)$). Doing so also involves only single-spin flips of positive probability, and results in the final configuration $\eta \in \mathcal{S}_n$. The evolution of the initial configuration according to Algorithm 2 is illustrated in Fig. 14.

Denoting by $\mathsf{A} : \Omega_n \to \Omega_n$ the mapping described by Algorithm 2, we have that for each $\sigma \in \mathcal{T}_n$, $\mathsf{A}\sigma \in \mathcal{S}_n$ and $\sigma \rightsquigarrow \mathsf{A}\sigma$. Thus, for each σ, there exists a $t \triangleq t(\mathsf{A}\sigma) \in \mathbb{N}$ such that $P^t(\sigma, \mathsf{A}\sigma) > 0$, which establishes Lemma 1.

Appendix E
Proof of Proposition 4

Fix $\sqrt{n} \in \mathbb{N}$ and $t \geq 0$. Recall that $\mathbb{E}N_t = t$ and let $E_{\delta,t} = \mathbb{1}_{\{|N_t - t| \leq \delta t\}}$. To prove the lower bound, let $p^\star_{X_0}$ attain the maximum in $I_n^{(c)}(t)$ and set $X_0^{(c)} \sim p^\star_{X_0}$. We have

$$\begin{aligned}
I_n^{(c)}(t) &= I\left(X_0^{(c)}; X_t^{(c)}\right) \\
&\leq I\left(X_0^{(c)}; X_t^{(c)}, E_{\delta,t}\right) \\
&\overset{(a)}{=} I\left(X_0^{(c)}; X_t^{(c)} \middle| E_{\delta,t}\right) \\
&= \mathbb{P}\Big(|N_t - t| \leq \delta t\Big) I\left(X_0^{(c)}; X_t^{(c)} \middle| E_{\delta,t} = 1\right) + \mathbb{P}\Big(|N_t - t| > \delta t\Big) I\left(X_0^{(c)}; X_t^{(c)} \middle| E_{\delta,t} = 0\right) \\
&\overset{(b)}{\leq} I\left(X_0; X_{N_t} \middle| E_{\delta,t} = 1\right) + n \cdot \mathbb{P}\Big(|N_t - t| > \delta t\Big) \\
&\overset{(c)}{\leq} I\left(X_0; X_{\lceil (1-\delta)t \rceil}\right) + n \cdot \mathbb{P}\Big(|N_t - t| > \delta t\Big),
\end{aligned} \tag{82}$$

where (a) follows by the independence of $X_0^{(c)}$ and $(N_t)_{t\geq 0}$, (b) uses $I\left(X_0^{(c)};X_t^{(c)}\middle|E_{\delta,t}=0\right)\leq \log|\Omega_n|$ and the definition of the continuous-time dynamics, while (c) is a consequence of the DPI (which implies that our mutual information cannot grow with time) and the independence of $(X_t)_{t\in\mathbb{N}_0}$ and $(N_t)_{t\geq 0}$.

Fix $\epsilon\in(0,1)$ and set $\delta=t^{-\frac{1-\epsilon}{2}}$. Bounding the tail probability of a Poisson distribution (via, e.g., a Chernoff Bound), we have

$$\mathbb{P}\Big(|N_t-t|>\delta t\Big)\leq 2e^{-\frac{\sqrt{t}}{2(\sqrt{t}+1)}t^{\epsilon}}=e^{-\Theta(t^{\epsilon})}. \tag{83}$$

Combining (82) and (83) and maximizing the RHS of (82) over all p_{X_0}, we have

$$I_n(t)\geq I_n^{(c)}\big((1+o(1))t\big)-n\cdot e^{-\Theta(t^{\epsilon})},\quad \forall t\in\mathbb{N}_0. \tag{84}$$

For the upper bound, let $\tilde{p}_{X_0}$ attain the maximum in $I_n\big((1+\delta)t\big)$, where δ is as above. Letting $X_0^{(c)}\sim\tilde{p}_{X_0}$, consider

$$\begin{aligned}
I_n^{(c)}(t)&\geq I\left(X_0^{(c)};X_t^{(c)}\right)\\
&=I\left(X_0^{(c)};X_t^{(c)}\middle|E_{\delta,t}\right)-I\left(X_0^{(c)};E_{\delta,t}\middle|X_t^{(c)}\right)\\
&\overset{(a)}{\geq}\mathbb{P}\Big(|N_t-t|\leq\delta t\Big)I\left(X_0;X_{N_t}\middle|E_{\delta,t}=1\right)-h\Big(\mathbb{P}\big(E_{\delta,t}=0\big)\Big)\\
&\overset{(b)}{\geq}\mathbb{P}\Big(|N_t-t|\leq\delta t\Big)I\left(X_0;X_{\lfloor(1+\delta)t\rfloor}\right)-h\Big(\mathbb{P}\big(E_{\delta,t}=0\big)\Big)\\
&=\mathbb{P}\Big(|N_t-t|\leq\delta t\Big)I_n\big(\lfloor(1+\delta)t\rfloor\big)-h\Big(\mathbb{P}\big(E_{\delta,t}=0\big)\Big),
\end{aligned} \tag{85}$$

where (a) is because $I\left(X_0^{(c)};E_{\delta,t}\middle|X_t^{(c)}\right)\leq H(E_{\delta,t})=h\Big(\mathbb{P}\big(E_{\delta,t}\big)\Big)$ with h denoting the binary entropy, (b) uses the DPI and the independence of the discrete-time dynamics and the Poisson process. Using (83), this becomes

$$I_n\big(t\big)\leq\Big(1-e^{-\Theta(t^{\epsilon})}\Big)^{-1}\left[I_n^{(c)}\big((1-o(1))t\big)+h\left(e^{-\Theta(t^{\epsilon})}\right)\right],\quad\forall t\in\mathbb{N}_0, \tag{86}$$

which concludes the proof.

APPENDIX F
PROOF OF THEOREM 3

Before starting the proof we formalize the considered framework. A la Remark 1, we consider the continuous-time Markov process on $\Omega_{\mathbb{Z}^2}\triangleq\{-1,+1\}^{\mathbb{Z}^2}$, with generator $\mathcal{L}$ acting on functions $f:\Omega_{\mathbb{Z}^2}\to\mathbb{R}$ as

$$\mathcal{L}f(\sigma)=\sum_{v\in\mathbb{Z}^2}r_{v,\sigma}\big[f(\sigma^v)-f(\sigma)\big], \tag{87}$$

where $r_{v,\sigma}$ is the flip rate at vertex v when the system is in state σ, given by

$$r_{v,\sigma}=\begin{cases}1, & m_v(\sigma)<\ell_v(\sigma)\\ \frac{1}{2}, & m_v(\sigma)=\ell_v(\sigma)\\ 0, & m_v(\sigma)>\ell_v(\sigma),\end{cases} \tag{88}$$

where, with some abuse of notation, $m_v(\sigma)$ and $\ell_v(\sigma)$ are defined as in Section II-A but with the neighbourhood $\mathcal{N}(v)$ of a vertex v defined with respect to the infinite 2D grid $\mathbb{Z}^2$ (rather than the $\sqrt{n} \times \sqrt{n}$ grid). To simplify notation, we reuse $(X_t)_{t\geq 0}$ for the considered Markov process and τ for the stopping time of interest (rather than $\big(\bar{X}_t^{(c)}\big)_{t\geq 0}$ and $\bar{\tau}$, respectively).

An important property of Glauber dynamics that is used several times throughout the proof is its so-called monotonicity. It enables comparing the evolution of the underlying Markov chain when initiated at different starting configurations. Namely, for each $\sigma \in \Omega_{\mathbb{Z}^2}$, let $(X_t^\sigma)_{t\geq 0}$ be the continuous-time Markov chain described by (87)-(88) initiated from $X_0 = \sigma$. To state the monotonicity property, we first explain how to construct the different $(X_t^\sigma)_{t\geq 0}$, for $\sigma \in \Omega_{\mathbb{Z}^2}$, over the same probability space.[7] Let $\big\{T^{(v)}\big\}_{v\in\mathbb{Z}^2}$ be a family of independent Poisson clock processes, i.e., for each $v \in \mathbb{Z}^2$, $T^{(v)} = \big\{T_n^{(v)}\big\}_{n\in\mathbb{N}_0}$, where $T_0^{(v)} = 0$ and $T_{n+1}^{(v)} - T_n^{(v)}$ are i.i.d. exponentially distributed random variables with mean 1. Also let $\big\{U^{(v)}\big\}_{v\in\mathbb{Z}^2}$ be a set of i.i.d. Bernoulli processes with values in $\{-1, +1\}$ and parameter $\frac{1}{2}$, i.e., $U^{(v)} = \big\{U_n^{(v)}\big\}_{n\in\mathbb{N}_0}$ is an i.i.d. with $U_n^{(v)} \sim \mathsf{Ber}\big(\frac{1}{2}\big)$, for all $n \in \mathbb{N}_0$. Now, given an initial configuration $\sigma \in \Omega_{\mathbb{Z}^2}$, we construct $(X_t^\sigma)_{t\geq 0}$ as follows:

- For each $v \in \mathbb{Z}^2$, $\big(X_t^\sigma(v)\big)_{t\geq 0}$ is constant on the intervals $\Big\{\big[T_n^{(v)}, T_{n+1}^{(v)}\big)\Big\}_{n\in\mathbb{N}_0}$
- For each $v \in \mathbb{Z}^2$ and $n \in \mathbb{N}$, $X^\sigma_{T_n^{(v)}}(v)$ is set to the spin of the majority of its neighbours, if a strict majority exists, and to $X^\sigma_{T_n^{(v)}}(v) = U_n^{(v)}$ otherwise.[8]

One may readily verify that the above set up of the dynamics conforms with (87)-(88). We reuse $\mathbb{P}$ to denote the probability measure associated with the above described coupling; $\mathbb{E}$ is the corresponding expectation.

Over this joint probability space, monotonicity is stated as follows. Consider the partial ordering of $\Omega_{\mathbb{Z}^2}$ where $\sigma \geq \sigma'$ if and only if $\sigma(v) \geq \sigma'(v)$, for all $v \in \mathbb{Z}^2$. If $\sigma \geq \sigma'$, we have $\mathbb{P}$-almost surely that $X_t^\sigma \geq X_t^{\sigma'}$, for all $t \geq 0$. In words, initiating the dynamics at two comparable configurations $\sigma \geq \sigma'$, the above described coupling preserves this ordering between the two chains as time progresses.

A. Upper Bound

For the upper bound we show that for any $c > 4$,

$$\mathbb{P}_\rho\big(\tau \leq c\ell^2\big) \geq 1 - e^{-\left(\frac{c}{12} - \frac{1}{3}\right)\ell} \tag{89}$$

holds. Since the hitting time of the all-minus configuration is invariant to the placement of the initial $\ell \times \ell$ droplet in the infinite lattice, we set initial configuration $X_0 = \rho \in \Omega_{\mathbb{Z}^2}$ as the one with $\rho(v) = +1$ for all $v \in [\ell]^2$, and $\rho(v) = -1$ otherwise.

To simplify the analysis we resort to a slightly modified dynamics. The idea is to stochastically dominate $(X_t)_{t\geq 0}$ by a process that erodes the initial droplet only from a single corner (rather than the four-corner-erosion in the original dynamics). To set up the modified dynamics we first need some definitions.

[7] Subsequently, we discard the superscript σ indicating the initial configuration whenever it is clear from the context.

[8] Note that, almost surely, two different sites will not update at the same time.

For $\sigma \in \Omega_{\mathbb{Z}^2}$ and $i \in \mathbb{Z}$, define

$$\sigma^{i\uparrow}\big((i',j')\big) = \begin{cases} -1, & i' = i \text{ and } j' \geq \ell + 1 \\ \sigma\big((i',j')\big), & \text{otherwise} \end{cases} \tag{90}$$

as the configuration that coincides with σ everywhere possibly except on $\{(i,j)\}_{j \geq \ell+1}$, where it is set to -1. Analogously, for $j \in \mathbb{Z}$, define

$$\sigma^{j\rightarrow}\big((i',j')\big) = \begin{cases} -1, & j' = j \text{ and } i' \geq \ell + 1 \\ \sigma\big((i',j')\big), & \text{otherwise} \end{cases} . \tag{91}$$

Let $\mathsf{Q} : \Omega_{\mathbb{Z}^2} \to \Omega_{\mathbb{Z}^2}$ be a transformation described by

$$\mathsf{Q}(\sigma) = \begin{cases} \sigma^{i\uparrow}, & \text{if } \sigma((i,\ell)) = -1 \\ \sigma^{j\rightarrow}, & \text{if } \sigma((\ell,j)) = -1 \\ \sigma, & \text{otherwise} \end{cases} \tag{92}$$

In words, Q takes a configuration and inspects the line $\big\{(i,\ell)\big\}_{i\in[\ell]}$. Every site with a negative spin gets vertically continued to an all-minus semi-infinite stripe upwards. An analogous (horizontal) operation is preformed on the minus-labeled sites in $\big\{(\ell,j)\big\}_{j\in[\ell]}$.

Now consider a modified dynamics $(\bar{Y}_t)_{t\geq 0}$ on $\Omega_{\mathbb{Z}^2}$ with initial condition $\bar{Y}_0 = \bar{\rho} \in \Omega$, where $\bar{\rho}$ is the configuration that assigns $+1$ spins to all the sites in the 1st quadrant of $\mathbb{Z}^2$, and -1 spins otherwise (i.e., $\bar{\rho}(v) = +1$ if and only if $v = (i,j)$ with $i,j \geq 1$). Let $\bar{\tau}$ be the first time when the spin at (ℓ,ℓ) is flipped from the original $+1$ to -1, and set $\bar{X}_t = \bar{Y}_{t\wedge\bar{\tau}}$. Whenever the Poisson clock at site $v \in \mathbb{Z}^2$ rings (before $(\bar{X}_t)_{t\geq 0}$ freezes into the all-minus configuration), first the current spin at v is refreshed, and then the transformation Q is applied to the obtained configuration. Thus, the $\bar{X}_t$ dynamics coincide with the original X_t dynamics, when initiating the latter at $X_0 = \bar{\rho}$, except for those times when a spin at $v = (i,j)$ with $i = \ell$ or $j = \ell$ is flipped. When this happens, the $\bar{X}_t$ dynamics transforms all the spins above or to the right of v to -1 (depending on whether $j = \ell$ or $i = \ell$, respectively). Let $\bar{\mathbb{P}}$, $\bar{\mathbb{E}}$ and $\bar{\mathcal{L}}$ be the probability measure, expectation and generator associated with $(\bar{X}_t)_{t\geq 0}$.

Observe that in $(\bar{X}_t)_{t\geq 0}$, if for some $t \geq 0$ we have $\bar{X}_t\big((i,j)\big) = -1$, for $i,j \in [\ell]^2$ (i.e., the spin at (i,j) was flipped from its original $+1$), then $\bar{X}_t\big((i',j')\big) = -1$, for all $i' \leq i$ and $j' \leq j$. This implies that $X_{\bar{\tau}}(v) = -1$, for all $v \in [\ell]^2$.

Note that one can couple $(X_t)_{t\geq 0}$ and $(\bar{X}_t)_{t\geq 0}$ so that $\bar{X}_t \geq X_t$, for all $t \geq 0$. Clearly, $\bar{X}_0 = \bar{\rho} \geq \rho = X_0$, while for $t > 0$, the coupling merely uses the same Poisson clocks for all sites with a coinciding neighborhood in both dynamics, and independent clocks otherwise. It is clear that if up to time $t > 0$ only sites inside $[\ell - 1]^2$ were updated, then $\bar{X}_t \geq X_t$ due to $\bar{X}_0 \geq X_0$ and monotonicity. Therefore, one has to verify that monotonicity is preserved only after sites $(i,j) \in [\ell]^2$, with $i = \ell$ or $j = \ell$, were updated in the $(\bar{X}_t)_{t\geq 0}$ dynamics (after which the transformation Q is applied). Indeed, assume that at time $t > 0$ the spin $\bar{X}_t(i,\ell)$, for some $i \in [\ell]$, is flipped. The aforementioned coupling implies that $\bar{X}_t\big((i',j')\big) = X_t\big((i',j')\big) = -1$, for all $i' \leq i$ and $j' \leq \ell$. Since

$X_t\big((i,j')\big) = -1$, for all $j \geq \ell + 1$ and $t \geq 0$, even after the transformation Q, the relation $\bar{X}_t \geq X_t$ still holds. The argument accounting for spin flips at sites (ℓ, i), for some $j \in [\ell]$, is symmetric.

Consequently, $\tau \leq \bar{\tau}$ almost surely, with respect to the coupling between $(X_t)_{t\geq 0}$ and $(\bar{X}_t)_{t\geq 0}$, which implies

$$\mathbb{P}\big(\tau > c\ell^2\big) \leq \bar{\mathbb{P}}\big(\bar{\tau} > c\ell^2\big). \tag{93}$$

The simplification in analyzing $\bar{\tau}$ can be intuitively understood as a restriction on the original dynamics (initiated at ρ) to erose the square droplet only from its bottom-left corner. This prohibits several unwanted phenomena that require a delicate analysis without affecting the time scale (the restriction slows the disappearance time by a constant factor of).

For $\sigma \in \Omega_{\mathbb{Z}^2}$, let $\mathcal{C}(\sigma) \triangleq \big\{v \in [\ell]^2 \big| \sigma(v) = +1\big\}$ be the set of plus-labeled vertices inside $[\ell]^2$, and define $N : \Omega_{\mathbb{Z}^2} \to \mathbb{N}_0$ by $N(\sigma) \triangleq \big|\mathcal{C}(\sigma)\big|$. To evaluate $\bar{\mathbb{E}}\bar{\tau}$, we study the quantity $\phi(t) \triangleq \bar{\mathbb{E}}N(t)$. Recalling that $\bar{\mathcal{L}}$ is the generator of $(\bar{X}_t)_{t\geq 0}$, we have

$$\frac{d\phi(t)}{dt} = \bar{\mathbb{E}}\bar{\mathcal{L}}N(\bar{X}_t). \tag{94}$$

Furthermore, since $N(\sigma)$ depends only on the values of σ inside $[\ell]^2$, while the transformation Q may alter configurations only outside $[\ell]^2$, for any $\sigma \in \Omega_{\mathbb{Z}^2}$, we have

$$\bar{\mathcal{L}}N(\sigma) = \sum_{v\in[\ell]^2} r_{v,\sigma}\big[N(\sigma^v) - N(\sigma)\big]. \tag{95}$$

To evaluate the RHS above, let

$$\mathcal{D}(\sigma) \triangleq \bigcup_{v\in\mathcal{C}(\sigma)} \mathcal{B}_\infty\left(v, \frac{1}{2}\right) \tag{96}$$

where $\mathcal{B}_p(v, r)$ is the ball, with respect to the L^p norm, of radius r centered at v. Denote the boundary of $\mathcal{D}(\sigma)$ by $\gamma(\sigma)$ and note that the for all configurations σ reachable from $\bar{\rho}$ in the $(\bar{X}_t)_{t\geq 0}$ dynamics (except the final configuration $\bar{X}_{\bar{\tau}}$, for which we account separately) $\gamma(\sigma)$ is a simple path. Furthermore, if $v_1(\sigma)$ and $v_2(\sigma)$ are, respectively, the bottom-right and top-left vertices in $\mathcal{C}(\sigma)$, then the shape of $\gamma(\sigma)$ is as follows: it goes directly up from $v_1(\sigma) + \big(\frac{1}{2}, -\frac{1}{2}\big)$ to $\big(\ell + \frac{1}{2}, \ell + \frac{1}{2}\big)$, from which it goes directly left to $v_2(\sigma) + \big(-\frac{1}{2}, \frac{1}{2}\big)$, and finally it descends from $v_2(\sigma) + \big(-\frac{1}{2}, \frac{1}{2}\big)$ to $v_1(\sigma) + \big(\frac{1}{2}, -\frac{1}{2}\big)$ by downwards and rightwards moves over segments of positive integer lengths. A typical shape of $\gamma(\sigma)$ is given in Fig. 15.

We use $\gamma(\sigma)$ in a counting argument for the number of flippable plus and minus spins. This will allow us to upper bound $\bar{\mathcal{L}}N(\sigma)$. Similarly to (76), for any $\sigma \in \Omega_{\mathbb{Z}^2}$ define

$$\mathcal{N}_a(\sigma) \triangleq \left\{v \in \mathbb{Z}^2 \middle| m_v(\sigma) > \frac{|\mathcal{N}_v|}{2}\right\} \tag{97a}$$

$$\mathcal{N}_d(\sigma) \triangleq \left\{v \in \mathbb{Z}^2 \middle| m_v(\sigma) < \frac{|\mathcal{N}_v|}{2}\right\}. \tag{97b}$$

The sets $\mathcal{N}_a(\sigma)$ and $\mathcal{N}_d(\sigma)$ contain, respectively, all the vertices in $\mathbb{Z}^2$ that agree or disagree with the majority of their neighbors. Also let $\mathcal{N}_u(\sigma) \triangleq \mathbb{Z}^2 \setminus \Big\{\mathcal{N}_a(\sigma) \cup \mathcal{N}_d(\sigma)\Big\}$ be the set of undecided vertices, i.e., all $v \in \mathbb{Z}^2$ with

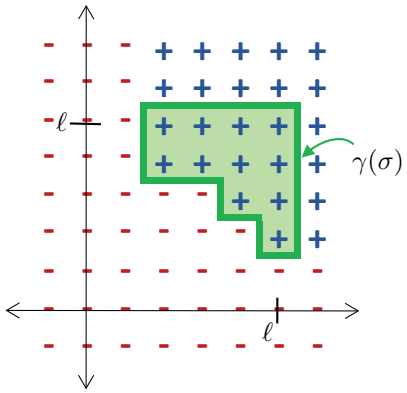


Fig. 15: A typical shape of $\gamma(\sigma)$ is depicted by the green line. The region $\mathcal{D}(\sigma)$ is colored in light green.

$m_v(\sigma) = \frac{|\mathcal{N}_v|}{2}$. Denote the sizes of the above sets by:

$$N_a(\sigma) \triangleq \big|\mathcal{N}_a(\sigma)\big| \quad ; \quad N_d(\sigma) \triangleq \big|\mathcal{N}_d(\sigma)\big| \quad ; \quad N_u(\sigma) \triangleq \big|\mathcal{N}_u(\sigma)\big|. \tag{98}$$

We also use $\mathcal{N}_a^+(\sigma) = \big\{v \in \mathcal{N}_a(\sigma)\big|\sigma(v) = +1\big\}$, $\mathcal{N}_a^-(\sigma) \triangleq \mathcal{N}_a(\sigma) \setminus \mathcal{N}_a^+(\sigma)$ and analogously define $\mathcal{N}_d^+(\sigma)$, $\mathcal{N}_d^-(\sigma)$, $\mathcal{N}_u^+(\sigma)$ and $\mathcal{N}_u^-(\sigma)$; the cardinalities of these sets are denoted by $N_a^+(\sigma)$, $N_a^-(\sigma)$, etc.

The following lemma, which a consequence of Hopf's Umlaufsatz, states a general result using which we use to characterize the relation between the above-defined cardinalities.

Lemma 2 *Let $\sigma \in \Omega_{\mathbb{Z}^2}$ be a configuration with $L+1$ connected monochromatic components $\big\{\mathcal{C}_i(\sigma)\big\}_{i\in[0:L]}$ of size at least 2, such that $\cup_{i\in[0:L]}\mathcal{C}_i(\sigma) = \mathbb{Z}^2$ and $\mathcal{C}_0(\sigma)$ is the unique infinite component which is labeled -1. Assume that for each $i,j \in [L]$, $d\big(\mathcal{C}_i(\sigma),\mathcal{C}_j(\sigma)\big) \geq 2$, where $d : \mathbb{Z}^2 \times \mathbb{Z}^2 \to \mathbb{N}_0$ is the graph distance. Let $\mathcal{E}(\sigma)$ be the collection of edges that separate cells of opposite polarity. The configuration σ is said to be shard-free if there does not exist two parallel edges in $\mathcal{E}(\sigma)$ at L^1 distance 1. Assuming that σ is shard-free, we have $N_u^+(\sigma) - N_u^-(\sigma) + 2N_d^+(\sigma) - 2N_d^-(\sigma) = 4L$.*

The proof of Lemma 2 is given in Appendix H. A corollary of the above lemma for the framework of interest is as follows.

Corollary 4 (Relation between $N_u^+(\bar{X}_t)$ and $N_u^-(\bar{X}_t)$) *For any $t \geq 0$, $N_d^+(\bar{X}_t) = N_d^-(\bar{X}_t) = 0$ almost surely. Also, given $t < \bar{\tau}$, we have $N_u^+(\bar{X}_t) - N_u^-(\bar{X}_t) = 1$, almost surely.*

We proceed with upper bounding the RHS of (93). For any $\lambda > 0$, we have

$$\bar{\mathbb{P}}\big(\bar{\tau} > c\ell^2\big) = \bar{\mathbb{P}}\Big(N\big(\bar{X}_{c\ell^2}\big) \geq 1\Big) \leq e^{-\lambda}\bar{\mathbb{E}}e^{\lambda N\left(\bar{X}_{c\ell^2}\right)}. \tag{99}$$

Denoting $f \triangleq e^{\lambda N}$ and $g(t) \triangleq \bar{\mathbb{E}}f(\bar{X}_t)$, we next show that $g(c\ell^2)$ is upper bounded by an exponentially decaying

(in ℓ) function. This is achieved using the relation

$$\frac{dg(t)}{dt} = \bar{\mathbb{E}}\bar{\mathcal{L}}f(\bar{X}_t) = \bar{\mathbb{E}}\left[\sum_{v\in[\ell]^2} r_{v,\bar{X}_t}\big[f(\bar{X}_t^v) - f(\bar{X}_t)\big]\right]. \tag{100}$$

For any $t \geq 0$, we have

$$\begin{aligned}
\bar{\mathbb{E}}\bar{\mathcal{L}}f(\bar{X}_t) &\overset{(a)}{=} \bar{\mathbb{E}}\left[\sum_{v\in[\ell]^2} r_{v,\bar{X}_t}\big[f(\bar{X}_t^v) - f(\bar{X}_t)\big]\mathbb{1}_{\{\bar{\tau}>t\}}\right] \\
&= \bar{\mathbb{E}}\left[\left(N_u^+(\bar{X}_t)\cdot f(\bar{X}_t)\frac{e^{-\lambda}-1}{2} + N_u^-(\bar{X}_t)\cdot f(\bar{X}_t)\frac{e^{\lambda}-1}{2}\right)\mathbb{1}_{\{\bar{\tau}>t\}}\right] \\
&\overset{(b)}{\leq} \bar{\mathbb{E}}\left[f(\bar{X}_t)\left(\frac{\lambda}{2}\big(N_u^-(\bar{X}_t) - N_u^+(\bar{X}_t)\big) + \frac{\lambda^2}{4}N_u^+(\bar{X}_t) + \frac{\lambda^2}{2(1-\lambda)}N_u^-(\bar{X}_t)\right)\mathbb{1}_{\{\bar{\tau}>t\}}\right] \\
&\overset{(c)}{\leq} \left(-\frac{\lambda}{2} + \frac{\lambda^2\ell}{4} + \frac{\lambda^2(\ell-1)}{2(1-\lambda)}\right)g(t)
\end{aligned} \tag{101}$$

where (a) is because $r_{v,\bar{X}_t}\mathbb{1}_{\{\bar{\tau}\leq t\}} = 0$ almost surely, (b) uses $e^{-\lambda} \leq 1-\lambda+\frac{\lambda^2}{2}$ and $e^{\lambda} \leq 1+\lambda+\frac{\lambda^2}{1-\lambda}$, while (c) invokes Corollary 4 and the fact that $N_u^+(\bar{X}_t) \leq \ell$ and $N_u^-(\bar{X}_t) \leq \ell-1$, almost surely.

Substituting $\lambda = \frac{1}{3\ell}$ into the RHS of (101), further gives

$$\bar{\mathbb{E}}\bar{\mathcal{L}}f(\bar{X}_t) \leq -\frac{1}{12\ell}g(t), \tag{102}$$

which, when combined with (100), produces

$$\frac{dg(t)}{dt} \leq -\frac{1}{12\ell}g(t). \tag{103}$$

On account of (103) and $g(0) = e^{\frac{\ell}{3}}$, we obtain

$$g(t) \leq e^{-\frac{t}{12\ell}+\frac{\ell}{3}}. \tag{104}$$

Combining (104) with (93) and (99) gives

$$\mathbb{P}\big(\tau > c\ell^2\big) \leq e^{-\frac{1}{3\ell}}g(c\ell^2) \leq e^{-\left(\frac{c}{12}-\frac{1}{3}\right)\ell}, \tag{105}$$

and concludes the proof of the upper bound.

B. Lower Bound

We next show that there exist $c, \gamma > 0$ such that

$$\mathbb{P}_\rho\big(\tau \geq c\ell^2\big) \geq 1 - e^{-\gamma\ell}. \tag{106}$$

Analogously to the proof of the upper bound, to approximate $\mathbb{P}_\rho\big(\tau < c\ell^2\big)$ we may switch to an alternative dynamics that speeds up the erosion of the square droplet. Furthermore, we may replace τ with any earlier stopping time. Both these ideas are subsequently used to simplify the analysis.

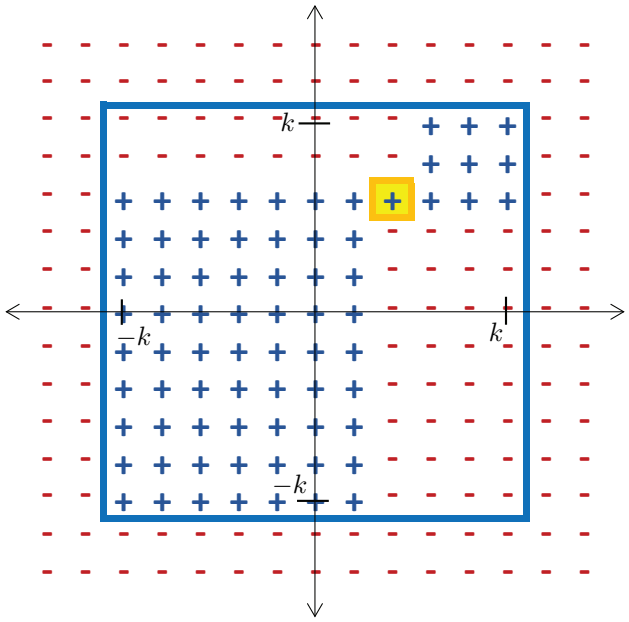


Fig. 16: An illustration of η_{v_0}, for $k = 5$ and $v_0 = (2, 3)$. The unstable site v_0 is highlighted in yellow. The blue line is the contour of the original square droplet of pluses. If the Poisson clock at v_0 rings, the set of all plus-labeled vertices seizes to be connected.

For convenience, assume that $\ell = 2k+1$, for some $k \in \mathbb{N}$ (otherwise, simple rounding arguments adjust the proof to even values of ℓ). Furthermore, due to translation invariance of the droplet dynamics over $\mathbb{Z}^2$, assume that the droplet $X_0 = \rho$ is centered at $(0, 0) \in \mathbb{Z}^2$. Let $\mathcal{C}(\sigma) \triangleq \{v \in [-k : k]^2 | \sigma(v) = +1\}$ be the set of all plus-labeled sites of σ in $[-k : k]^2$. The main challenge in analyzing $(X_t)_{t\geq 0}$ is that while $\mathcal{C}(\rho)$ is a connected components, it can split into two disconnected components as the dynamics evolve. For instance, for any $v_0 = (i_0, j_0) \in [-(k-2) : k-2]^2$ the configuration η_{v_0}, where

$$\eta_{v_0}(v) = \begin{cases} +1, & v \in \{[i_0 + 1 : k] \times [j_0 : k]\} \cup \{[-k : i_0 - 1] \times [-k : j_0]\} \cup \{(i_0, j_0)\} \\ -1, & \text{otherwise} \end{cases}, \tag{107}$$

is reachable from ρ (see Fig. 16). Since $v_0 \in \mathcal{N}_u^+(\eta_{v_0})$, the Poisson clock at v_0 has a positive rate $r_{v_0,\eta_{v_0}} = \frac{1}{2} > 0$. If that Poisson clock rings, the connected component of pluses splits into two disconnected components.

We circumvent this issue by introducing several modifications to the original dynamics. First, define $\tilde{\rho} \in \Omega_{\mathbb{Z}^2}$ as

$$\tilde{\rho}(v) = \begin{cases} +1, & v \in [-k : k]^2 \cap \mathcal{B}_1(0, k+1) \\ -1, & \text{otherwise.} \end{cases} \tag{108}$$

Next, let $m = \lfloor \frac{3}{4}k \rfloor$ and define

$$\mathcal{T}_m = \{(i, 0)\}_{\substack{i\in[-m:m] \\ j\in[-1:1]}} \cup \{(0, j)\}_{\substack{i\in[-1:1] \\ j\in[-m:m]}}. \tag{109}$$

The last ingredient needed to defined the auxiliary dynamics is an analog of the mapping E from the proof of Lemma 1 from Appendix D, which eliminates unstable sites by flipping their spins. The features of the droplet dynamics allows to simplify the definition of E for the purpose of the current proof. With some abuse of notation, we (re)define $\mathsf{E} : \Omega_{\mathbb{Z}^2} \to \Omega_{\mathbb{Z}^2}$ as follows:

$$\mathsf{E}(\sigma) = \begin{cases} \sigma^v, & \text{if } \exists!\ v \in \mathcal{N}_d(\sigma) \\ \sigma\ , & \text{otherwise} \end{cases}, \tag{110}$$

i.e., $\mathsf{E}(\sigma)$ flips the spin of the unique site in $\mathcal{N}_d(\sigma)$ if such a site exists, and does nothing otherwise.

Consider an auxiliary dynamics $(\tilde{Y}_t)_{t\geq 0}$ on $\Omega_{\mathbb{Z}^2}$ initiated at $\bar{Y}_0 = \tilde{\rho} \in \Omega_{\mathbb{Z}^2}$. For $t > 0$, whenever the Poisson clock at site $v \in \mathbb{Z}^2$ rings, these dynamics first update the current spin at v, and then apply the transformation E to the obtained configuration. Let $\tilde{\tau}$ be the first time when the spin of any of the site in $\mathcal{T}_m$ is flipped, and set $\tilde{X}_t = \tilde{Y}_{t\wedge\tilde{\tau}}$, for all $t \geq 0$. The probability measure, expectation and generator associated with $(\tilde{Y}_t)_{t\geq 0}$ are denoted by $\tilde{\mathbb{P}}$, $\tilde{\mathbb{E}}$ and $\tilde{\mathcal{L}}$, respectively. Since $\tilde{\rho} \leq \rho$ and $\tilde{\tau} < \tau$ $\mathbb{P}$-almost surely, we have

$$\mathbb{P}_\rho\big(\tau < c\ell^2\big) \leq \mathbb{P}_{\tilde{\rho}}\big(\tilde{\tau} < c\ell^2\big). \tag{111}$$

Furthermore, getting the desired upper bound on the RHS above with k in the role of ℓ is sufficient as $2k \leq \ell \leq 3k$.

Now, note that under the $(\tilde{Y}_t)_{t\geq 0}$ dynamics the erosion of the plus-labeled component happens faster than under the original dynamics as a result of the transformation E. Through a coupling argument similar to that presented in the proof of (89), we can define both chains over the same probability space, such that $\tilde{Y}_t \leq X_t$, for any $t \geq 0$.

Consequently, to prove (106), it suffices to show that

$$\tilde{\mathbb{P}}\big(\tilde{\tau} < ck^2\big) \leq e^{-\gamma k}. \tag{112}$$

This requires controlling the difference between the number of unstable plus- and minus-labeled sites. Like before, we achieve this using Hopf's Umlaufsatz. Let $\mathcal{C}(\sigma) \triangleq \big\{v \in [-k:k]^2 \big| \sigma(v) = +1\big\}$ and define

$$\mathcal{D}(\sigma) \triangleq \bigcup_{v\in\mathcal{C}(\sigma)} \mathcal{B}_\infty\left(v, \frac{1}{2}\right). \tag{113}$$

Denote the boundary of $\mathcal{D}(\sigma)$ by $\gamma(\sigma)$. A certain regularity claim on the evolution of $\gamma(\tilde{X}_t)$ with time is required in order to apply Hopf's Umlaufsatz. To state it, for any $\sigma \in \Omega_{\mathbb{Z}^2}$, we split the set of unstable sites of positive spins, $\mathcal{N}_u^+(\sigma)$, into two non-intersecting subsets. Define

$$\mathcal{N}_1^+(\sigma) \triangleq \Big\{v \in \mathcal{N}_u^+(\sigma)\Big| \sigma\big(v+(1,0)\big) = \sigma\big(v-(1,0)\big) = +1 \text{ or } \sigma\big(v+(0,1)\big) = \sigma\big(v-(0,1)\big) = +1\Big\} \tag{114a}$$

$$\mathcal{N}_2^+(\sigma) \triangleq \mathcal{N}_u^+(\sigma) \setminus \mathcal{N}_1^+(\sigma). \tag{114b}$$

The set $\mathcal{N}_1^+(\sigma)$ contains all the sites of positive spin whose plus-labeled neighbors are either above and below them or to the right and to the left of them. All the other unstable plus-labeled sites are in $\mathcal{N}_2^+(\sigma)$. Furthermore,

for any $\sigma \in \Omega_{\mathbb{Z}^2}$, let

$$\mathcal{M}_u^+(\sigma) \triangleq \left\{ v \in \mathcal{N}_u^+(\sigma) \middle| \mathcal{N}_d(\sigma^v) \neq \emptyset \right\} \tag{115}$$

be the set of sites in σ whose flipping creates at least one unstable site in the obtained configuration. We have the following lemma.

Lemma 3 (Properties of $\mathcal{C}(\tilde{X}_t)$) *For any $t \geq 0$, the following holds almost surely:*

1) $\mathcal{N}_d^+(\tilde{X}_t) = \mathcal{N}_d^-(\sigma) = \emptyset$;

2) $\gamma(\tilde{X}_t)$ is a simple and closed curve in $\mathbb{R}^2$;

3) For every $v \in \mathcal{M}_u^+(\tilde{X}_t)$, we have $\left|\mathcal{C}\big(\mathsf{E}(\tilde{X}_t^v)\big)\right| = |\mathcal{C}(\tilde{X}_t)| - 2$. Furthermore, $M_u^+(\tilde{X}_t) \triangleq \left|\mathcal{M}_u^+(\tilde{X}_t)\right| \leq 8$;

4) $\mathcal{N}_1^+(\tilde{X}_t) = \emptyset$.

The proof of this lemma is relegated to Appendix J. Lemma 3 together with Lemma 2 imply that

$$N_u^+(\tilde{X}_t) - N_u^-(\tilde{X}_t) = 4, \tag{116}$$

almost surely, for all $t \geq 0$.

Recall that the dynamics $(\tilde{X}_t)_{t\geq 0}$ stops when any spin from $\mathcal{T}_m$ is flipped. The evolution of $(\tilde{X}_t)_{t\geq 0}$ implies that the first site from $\mathcal{T}_m$ to flip must be inside $\big\{(i,j)\big\}_{\substack{i=-1,1\\ j=-m,m}}$. Assume without loss of generality that $(1,m)$ is the first of the above eight options to become unstable. For this to happen, at least all the plus-labeled spins at sites inside $\mathcal{G}_l = \big\{(i,j)\big| i \leq 0, j \geq m\big\}$ or $\mathcal{G}_r = \big\{(i,j)\big| i \geq 2, j \geq m\big\}$ must already be flipped. Consequently, if $\tilde{\tau} < ck^2$, then

$$N(\tilde{X}_0) - N(\tilde{X}_{ck^2}) \geq |\mathcal{G}_r| \geq \frac{\left(\frac{1}{5}k\right)^2}{2} = \frac{k^2}{50}, \tag{117}$$

and we have

$$\tilde{\mathbb{P}}\big(\tilde{\tau} < ck^2\big) \leq \tilde{\mathbb{P}}\left(N(\tilde{X}_0) - N(\tilde{X}_{ck^2}) \geq \frac{k^2}{50}\right) \leq e^{-\frac{\lambda k^2}{50}} \tilde{\mathbb{E}} e^{\lambda\left(N(\tilde{\rho}) - N(\tilde{X}_{ck^2})\right)}, \tag{118}$$

for any $\lambda > 0$. Setting $f_\lambda(\sigma) \triangleq e^{\lambda\left(N(\tilde{\rho}) - N(\sigma)\right)}$ and $g_\lambda(t) \triangleq \tilde{\mathbb{E}} f_\lambda(\tilde{X}_t)$, we again approximate the expected value using the relation

$$\frac{dg_\lambda(t)}{dt} = \tilde{\mathbb{E}}\tilde{\mathcal{L}} f_\lambda(\tilde{X}_t). \tag{119}$$

Expanding the RHS, we have

$$\begin{aligned}
\tilde{\mathbb{E}}\tilde{\mathcal{L}} f_\lambda(\tilde{X}_t) &\overset{(a)}{=} \tilde{\mathbb{E}}\left[\sum_{v\in\mathbb{Z}^2} r_{v,\tilde{X}_t}\big[f_\lambda(\tilde{X}_t^v) - f_\lambda(\tilde{X}_t)\big]\mathbb{1}_{\{\bar{\tau}>t\}}\right] \\
&= \tilde{\mathbb{E}}\left[f_\lambda(\tilde{X}_t)\left(M_u^+(\tilde{X}_t)\frac{e^{2\lambda}-1}{2} + \big(N_u^+(\tilde{X}_t) - M_u^+(\tilde{X}_t)\big)\frac{e^{\lambda}-1}{2} + N_u^-(\tilde{X}_t)\frac{e^{-\lambda}-1}{2}\right)\mathbb{1}_{\{\bar{\tau}>t\}}\right] \\
&\overset{(b)}{=} \tilde{\mathbb{E}}\left[f_\lambda(\tilde{X}_t)\left(\frac{\lambda\big(M_u^+(\tilde{X}_t) + N_u^+(\tilde{X}_t) - N_u^-(\tilde{X}_t)\big)}{2} + \frac{\lambda^2\big(M_u^+(\tilde{X}_t) + N_u^+(\tilde{X}_t) + N_u^-(\tilde{X}_t)\big)}{2(1-\lambda)}\right)\mathbb{1}_{\{\bar{\tau}>t\}}\right] \\
&\overset{(c)}{\leq} \tilde{\mathbb{E}}\left[f_\lambda(\tilde{X}_t)\left(6\lambda + \frac{\lambda^2\big(M_u^+(\tilde{X}_t) + N_u^+(\tilde{X}_t) + N_u^-(\tilde{X}_t)\big)}{2(1-\lambda)}\right)\mathbb{1}_{\{\bar{\tau}>t\}}\right]
\end{aligned}$$

$$\overset{(d)}{\leq} \left(6\lambda + \frac{4\lambda^2(k+2)}{1-\lambda}\right) g_\lambda(t). \tag{120}$$

where:

(a) is because $c_{v,\tilde{X}_t} \mathbb{1}_{\{\bar{\tau} \leq t\}} = 0$;

(b) follows by $e^{2\lambda} \leq 1 + 2\lambda + \frac{2\lambda^2}{1-\lambda}$ and $e^{-\lambda} \leq e^{\lambda} \leq 1 + \lambda + \frac{\lambda^2}{1-\lambda}$;

(c) follows by Claim 3) from Lemma 3 and (116);

(d) is since the total number of unstable sites is at most linear with k and, in particular, is upper bounded $N_u(\tilde{X}_t) \leq \left|\partial_{\mathsf{Int}}\mathcal{C}(\tilde{\rho})\right| + \left|\partial_{\mathsf{Ext}}\mathcal{C}(\tilde{\rho})\right| \leq 8(k+1)$.

Inserting $\lambda = \frac{1}{k}$ into the RHS of (120) and denoting $g(t) \triangleq g_{\frac{1}{k}}(t)$, we see the there exists $c' > 0$ such that

$$\frac{dg(t)}{dt} \leq \frac{c'}{k} g(t). \tag{121}$$

Integrating both sides of (121) and using $g(0) = 1$ produces

$$g(t) \leq e^{\frac{c'}{k}t}. \tag{122}$$

Using the proxy from (122) to upper bound the RHS of (118), we obtain

$$\tilde{\mathbb{P}}\big(\tilde{\tau} < ck^2\big) \leq e^{-\frac{k}{50}} g(ck^2) \leq e^{-\left(\frac{1}{50} - cc'\right)k}. \tag{123}$$

Taking $c \in \left(0, \frac{1}{50c'}\right)$ and denoting $\gamma \triangleq \frac{1}{50} - cc' > 0$ concludes the proof.

APPENDIX G
PROOF OF THEOREM 7

An alternative yet equivalent description of the continuous-time dynamics uses the Poisson thinning property. The original description considers Poisson clocks of rate 1 that select sites for update; a new spin for the selected site is then randomly drawn according to the update distribution from (59). Instead, the update probability may be accounted for by associating with each vertex a time-varying Poisson clock whose rate depends on the instantaneous neighborhood of each site. More precisely, with each $v \in \mathcal{V}_n$, we associate a Poisson process $\left(N_t^{(v)}\right)_{t \geq 0}$ whose rate $\lambda_v(t)$ is given by

$$\lambda_v(t) = \begin{cases} \phi\big(S(X_t, v)\big), & X_t(v) = +1 \\ 1 - \phi\big(S(X_t, v)\big), & X_t(v) = -1 \end{cases}. \tag{124}$$

With these thinned rates, each time $\left(N_t^{(v)}\right)_{t \geq 0}$ jumps, the spin at v flips. To see the equivalence between the two descriptions, just consider the generator of each dynamics.

Our first step is to restrict minus-labeled sites from flipping. To do so, with some abuse of notation, we replace (124) with

$$\lambda_v(t) = \phi\big(S(X_t, v)\big) \mathbb{1}_{\left\{X_t(v) = +1\right\}}, \tag{125}$$

i.e., minus-labeled sites have flip rate 0. This modification means that until the spin at any $v \in \mathcal{B}$ flips to -1, $\lambda_v(t)$ is a piece-wise constant monotonically non-decreasing process that jumps at the times when v's horizontal

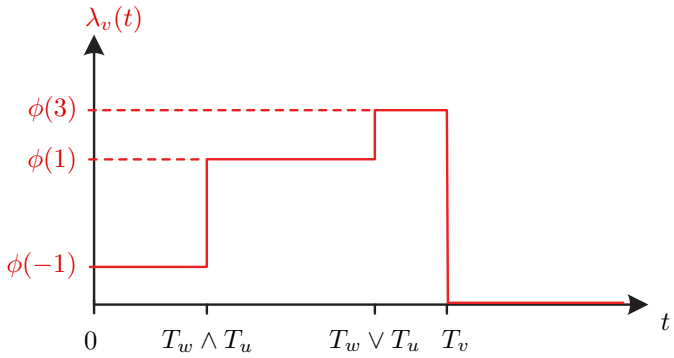


Fig. 17: An illustration of the flip rate process $\lambda_v(t)$ of a vertex v with two horizontal neighbors u and w. The flip times (from $+1$ to -1) of u, v and w are denoted by T_u, T_v and T_w. The illustration assumes $T_v > T_u \vee T_w$.

neighbors flip their spins. For instance, an explicit description of $\lambda_v(t)$, for $v \in \mathcal{B}_{\mathsf{Int}} \triangleq \mathcal{B} \setminus \big\{(1,1), (\sqrt{n}, 1)\big\}$, is

$$\lambda_v(t) = \begin{cases} \phi(-1), & N_v^{(-)}(t) = 1 \text{ and } X_t(v) = +1 \\ \phi(1), & N_v^{(-)}(t) = 2 \text{ and } X_t(v) = +1 \\ \phi(3), & N_v^{(-)}(t) = 3 \text{ and } X_t(v) = +1 \\ 0, & X_t(v) = -1 \end{cases} \tag{126}$$

where $N_v^{(-)}(t) = \big|\big\{u \in \mathcal{N}_v \big| X_t(u) = -1\big\}\big|$ is the number of v's neighbors with a -1 spin. An illustration of the evolution of $\lambda_v(t)$ for the case when v flips after both its horizontal neighbors is shown in Fig. 17. With some abuse of notation, we still use $(X_t)_{t\geq 0}$ for the restricted dynamics under which flips of minus-labeled sites are forbidden. By monotonicity of the SIM, the expected number of pluses in $\mathcal{B}$ with respect to the restricted dynamics cannot be higher than the original one.

Our next step is to replace $(X_t)_{t\geq 0}$ with a new dynamics with two pure (sprinkling and erosion) phases, at least up to a certain time. To have some terminology, any plus-labeled site with $N_v^{(-)}(t) = i$, where $i = 1, 2, 3$, is referred to as a site of Type i. When the spin at v is -1, its Type is 0. Accordingly, a sprinkling is a flip of a Type 1 site; flips of Type 2 or 3 sites are erosion flips; Type 0 sites never flip. Our goal is, for any fixed $s > 0$, to construct from $\big(X_t\big)_{t\geq 0}$ a new process $\big(X_t^{(s)}\big)_{t\geq 0}$ over the same probability space, such that

1) $\forall t \in [0, s)$ only Type 1 sites are flipped in $X_t^{(s)}$ (sprinkling phase).
2) $\forall t \in [s, 2s)$ only Type 2 or 3 sites are flipped in $X_t^{(s)}$ (erosion phase).
3) $X_s \geq X_{2s}^{(s)}$.

Here $s > 0$ is a design parameter that can be tuned for the purpose of the proof. Provided that such a process exists, by setting $s = t_0 = e^{c\beta}$ and letting $\hat{N}_{\mathcal{B}}^{(+)}(t) \triangleq \big|\big\{v \in \mathcal{B}\big| X_t^{(t_0)}(v) = +1\big\}\big|$ as the number of pluses in $\mathcal{B}$ with respect to $\big(X_t^{(t_0)}\big)_{t\geq 0}$, we will be able to lower bound the expected value of interest as

$$\mathbb{E}N_1^{(+)}(t_0) \geq \mathbb{E}\hat{N}_{\mathcal{B}}^{(+)}(2t_0). \tag{127}$$

Since the RHS above corresponds to the phase-separated dynamics, it allows a per-phase analysis - a significantly simpler task.

Construction of Phase-Separated Dynamics: Fix $s > 0$, and consider first an intermediate dynamics $\left(\tilde{X}_t^{(s)}\right)_{t\in[0,2s)}$ constructed as follows. Recall that $(N_t^{(v)})_{t\geq 0}$, for $v \in \mathcal{B}$, are the Poisson processes whose (first) jumps define the flips of the sites in $\mathcal{B}$. For $t \in [0, 2s)$ and any $v \in \mathcal{V}_n$, define

$$\tilde{N}_t^{(v)} = \begin{cases} N_t^{(v)}, & t \in [0, s) \\ N_{t-s}^{(v)}, & t \in [s, 2s) \end{cases}, \tag{128}$$

i.e, $\left(\tilde{N}_t^{(v)}\right)_{t\in[0,2s)}$ repeats the sequence of clock rings specified by $\left(N_t^{(v)}\right)_{t\in[0,s)}$ twice in a consecutive manner. The jumps of the processes $\left(\tilde{N}_t^{(v)}\right)_{t\in[0,2s)}$, $v \in \mathcal{V}_n$, can be thought of as the clock rings in the $\left(\tilde{X}_t^{(s)}\right)_{t\in[0,2s)}$ dynamics, however, as explained next, not all clock rings result in a flip.

Specifically, fix a sequence of times and vertices for update that occurred in $(X_t)_{t\geq 0}$ during $[0, s)$. Let k_s be the total number of such updates, and denote the k_s clock ring times by $t_1 < t_2 < \ldots < t_{k_s}$ and the corresponding sites by $v_1, v_2, \ldots, v_{k_s}$. Thus, for all $i \in [k_s]$, the clock at v_i rang at time t_i, which is when the original $+1$ spin at v_i flipped to -1. Note that there are no repetitions of the updated sites since only plus-labeled sites may update, and once updated, their flip rate drops to 0. These sequences of times and sites determine the sequence of clock rings in the $\left(\tilde{X}_t^{(s)}\right)_{t\in[0,2s)}$ dynamics. To differentiate the notation associated with each dynamics, define for the latter the times $\{\tau_j\}_{j\in[2k_s]}$ and the sites $\{u_j\}_{j\in[2k_s]}$, where

$$\tau_j = \begin{cases} t_j, & j \in [k_s] \\ t_{j-k_s} + s, & j \in [k_s + 1 : 2k_s] \end{cases} \tag{129}$$

and

$$u_j = \begin{cases} v_j, & j \in [k_s] \\ v_{j-k_s}, & j \in [k_s + 1 : 2k_s] \end{cases}. \tag{130}$$

Thus, $0 < \tau_1 < \ldots < \tau_{k_s} < s < \tau_{k_s+1} < \ldots < \tau_{2k_s} < 2s$ and each τ_j, for $j \in [2k_s]$, corresponds to a clock ring at site u_j. Some of the flips that correspond to these clock rings in the $\left(\tilde{X}_t^{(s)}\right)_{t\in[0,2s)}$ dynamics are blocked as described next:

- For $j \in [k_s]$, if in the original dynamics $N_{u_j}^{(-)}(\tau_j) \in \{2, 3\}$, i.e., the j-th update was an erosion flip in the $\left(X_t\right)_{t\in[0,s)}$ dynamics, then it is blocked with probability 1. Sprinkling updates, for which $N_{u_j}^{(-)}(\tau_j) = 1$, are allowed.
- For $j \in [k_s+1 : k_s]$, if in the original dynamics $N_{u_j}^{(-)}(\tau_j - s) = 1$ (sprinkle), then it is blocked with probability 1. Erosion flips, for which $N_{u_i}^{(-)}(\tau_j - s) \in \{2, 3\}$, go through.

By definition, $\left(\tilde{X}_t^{(s)}\right)_{t\in[0,2s)}$ satisfies the requirements for a phase-separated dynamics; it is merely a reordering of the flips in the $\left(X_t\right)_{t\in[0,s)}$ dynamics in a stretched time interval, so that only Type 1 updates occur during $[0, s)$, while only Type 2 or 3 updates happen during $[s, 2s)$. Clearly, $X_s = \tilde{X}_{2s}^{(s)}$.

The two phases of $\left(\tilde{X}_t^{(s)}\right)_{t\in[0,2s)}$ have the same exact clock rings almost surely. However, the rates at which the

clocks ring during the 2nd phase are incorrect, in the sense that they do not necessarily satisfy (125) with $\tilde{X}_t^{(s)}$ in the role of X_t therein, for $t \in [s, 2s)$. This is since $\tilde{X}_s^{(s)} \geq X_0$, and therefore, the flip rates of $\big(N_t^{(v)}\big)_{t\in[0,s)}$, for each $v \in \mathcal{B}$, are (possibly) too low to reflect the neighborhood of v at times $t \in [s, 2s)$.

Our final step in the construction is to fix the flip rates at the 2nd phase. Consider a third dynamics $\big(X_t^{(s)}\big)_{t\in[0,2s)}$, with a ring process $\big(M_t^{(v)}\big)_{t\in[0,2s)}$ defined as follows. Let $\Big\{\big(L_t^{(v)}\big)_{t\geq 0}\Big\}_{v\in\mathcal{B}}$ be a collection of independent Poisson processes (amongst themselves and independent of $\Big\{\big(N_t^{(v)}\big)_{t\geq 0}\Big\}_{v\in\mathcal{B}}$), each with rate

$$\eta_v(t) \triangleq \Big[\phi\big(S(\tilde{X}_{t+s}^{(s)}, v)\big) - \phi\big(S(X_t, v)\big)\Big]\mathbb{1}_{\left\{\tilde{X}_{t+s}(v)=+1\right\}}. \tag{131}$$

Since $\tilde{X}_s^{(s)} \geq X_0$, and because as long as v has a positive spin its flip rate is monotonically non-decreasing with t, we have that $\eta_v(t) \geq 0$. Furthermore, $\eta_v(t) + \lambda_v(t) = \phi\big(S(\tilde{X}_{t+s}^{(s)}, v)\big)\mathbb{1}_{\left\{\tilde{X}_{t+s}(v)=+1\right\}}$ (see (125)), which is the correct rate of clock rings for v in the 2nd phase (i.e., for $t \in [s, 2s)$) of the $\big(\tilde{X}_t^{(s)}\big)_{t\in[0,2s)}$ dynamics. Now, define $\big(M_t^{(v)}\big)_{t\in[0,2s)}$ as

$$M_t^{(v)} = \begin{cases} N_t^{(v)}, & t \in [0, s) \\ N_{t-s}^{(v)} + L_t^{(v)}, & t \in [s, 2s) \end{cases}. \tag{132}$$

Since the sum of two independent Poisson processes in a Poisson process whose rate is the sum of the two rates of its independent components, we have that $\big(M_t^{(v)}\big)_{t\in[0,2s)}$ is Poisson with rate

$$\mu_v(t) = \phi\big(S(\tilde{X}_t^{(s)}, v)\big)\mathbb{1}_{\left\{\tilde{X}_t(v)=+1\right\}}, \tag{133}$$

for all $t \in [0, 2s)$, as we wanted. Finally, to preserve the phase-separated property, we apply the same blocking rule on $\big(X_t^{(s)}\big)_{t\in[0,2s)}$ as we did for $\big(\tilde{X}_t^{(s)}\big)_{t\in[0,2s)}$. Namely, during $t \in [0, s)$, flips associated with clock rings of Type 2 or Type 3 sites are blocked, while for $t \in [s, 2s)$, we block flips of Type 1 sites.

Observe that by definition $M_t^{(v)} \geq \tilde{N}_t^{(v)}$ for all $t \in [0, 2s)$, i.e., the new dynamics $\big(X_t^{(s)}\big)_{t\in[0,2s)}$ has faster clock ring rates than $\big(\tilde{X}_t^{(s)}\big)_{t\in[0,2s)}$. Because the same blocking rule was applied on both dynamics, we conclude that $\big(X_t^{(s)}\big)_{t\in[0,2s)}$ is a phase-separated dynamics that satisfies $X_{2s}^{(s)} \leq \tilde{X}_{2s}^{(s)} = X_s$, as needed. As mentioned before, setting $s = t_0 = e^{c\beta}$ and letting $\hat{N}_{\mathcal{B}}^{(+)}(t) \triangleq \big|\big\{v \in \mathcal{B}\big|X_t^{(t_0)}(v) = +1\big\}\big|$, we have

$$\mathbb{E}N_1^{(+)}(t_0) \geq \mathbb{E}\hat{N}_{\mathcal{B}}^{(+)}(2t_0), \tag{134}$$

which implies that lower bounding the RHS above is sufficient.

Analysis of Phase-Separated Dynamics:

1) Sprinkling Phase: Replace the positive spins that the initial configuration $X_0^{(t_0)}$ has at $(1, 1)$ and $(\sqrt{n}, 1)$ with a negative spin. This can only further decrease the expectation $\mathbb{E}\hat{N}_{\mathcal{B}}^{(+)}(2t_0)$. The reason for doing so is that now $(2, 1)$ and $(\sqrt{n}-1, 1)$ are blocked from flipping during the 1st phase (as they have -1 neighbours), while for every $v \in \hat{\mathcal{B}} \triangleq \mathcal{B}_{\mathsf{Int}} \setminus \big\{(2, 1), (\sqrt{n}-1, 1)\big\}$, the initial neighborhood is the same (in terms of the number of neighbors

and their spins).

Right at the end of Phase 1, i.e., at time $t = t_0$, $X_{t_0}^{(t_0)}(\mathcal{B})$ comprises runs of plus-labeled vertices separated by minus-labeled sprinkles that occurred during the first phase. We refer to these contiguous runs of pluses as contigs. For a vertex $v = (i, 1) \in \mathcal{B}$ with $X_{t_0}^{(t_0)}(v) = -1$, we denote by L_i the length of the contig that begins at v (including v), namely:

$$L_i \triangleq 1 + \max\left\{ j \in [i+1 : \sqrt{n}] \middle| \forall \ell \in [i+1 : j]\ X_{t_0}^{(t_0)}\big((\ell, 1)\big) = +1 \text{ and } X_{t_0}^{(t_0)}\big((j+1, 1)\big) = -1 \right\}. \tag{135}$$

Now, for any $v \in \hat{\mathcal{B}}$, as long as both of v's horizontal neighbors are plus-labeled, the time until v flips is an exponential random variable with parameter $\phi(-1)$. However, if any of v's horizontal neighbors flip to -1 before v itself, the spin at v stays $+1$ until the end of Phase 1. Consequently, by allowing the sites in $\hat{\mathcal{B}}$ to flip independently during the 1st phase of $\big(X_t^{(t_0)}\big)_{t\in[0,2t_0)}$ after $\mathsf{Exp}\big(\phi(-1)\big)$ time (regardless of whether their horizontal neighbours have flipped or not), again, only speeds up the disappearance of pluses. With this relaxation, the probability of each site to flip until time t_0 is

$$\mathbb{P}\Big(\mathsf{Exp}\big(\phi(-1)\big) \le t_0\Big) = 1 - e^{-\phi(-1)t_0} \triangleq p(\beta), \tag{136}$$

where the dependence of $p(\beta)$ on β is through the exponent $\phi(-1)t_0 = \frac{e^{(c-1)\beta}}{e^{-\beta}+e^{\beta}}$. As $c < 1$ (in fact, $c < 2$ suffices here) we have that $p(\beta)$ can be made arbitrarily small by increasing β.

With this relaxation, the contigs' lengths are mutually independent, and the distribution of the i-th length L_i is

$$\mathbb{P}\big(L_i = k\big) = \begin{cases} \big(1-p(\beta)\big)^{k-1} p(\beta), & k \in [1 : \sqrt{n} - 1 - i] \\ 0, & \text{otherwise} \end{cases}. \tag{137}$$

This is a Geometric distribution but with a finite number of trials. Consider the total lengths of the first M contigs (M to be specified later), i.e., $L_M = \sum_{i=1}^{M} L_i$. We first show that for an appropriate choice of M, the probability that L exceeds the right corner $v_r = (\sqrt{n}, 1)$ of $\mathcal{B}$ can be made arbitrarily small with n, for any $\beta > 0$. To do so, observe that replacing each L_i with a pure Geometric random variable $G_i \sim \mathsf{Geo}\big(p(\beta)\big)$ can only increase its length. Namely, by setting $L_i = G_i \wedge (\sqrt{n} - 1 - i)$, it is coupled with G_i so that $G_i \ge L_i$ almost surely. Let $(G_i)_{i\in[M]}$ be i.i.d. $\mathsf{Geo}\big(p(\beta)\big)$ and denote by $\mu \triangleq \frac{M}{p(\beta)}$ the expected value of their sum. Consider the probability that L_M exceeds $\sqrt{n}$:

$$\mathbb{P}\Big(L_M \ge \sqrt{n}\Big) \le \mathbb{P}\left(\sum_{i=1}^{M} G_i \ge \sqrt{n}\right) = \mathbb{P}\left(\sum_{i=1}^{M} G_i \ge \frac{p(\beta)\sqrt{n}}{M}\mu\right). \tag{138}$$

Setting $M = \frac{p(\beta)}{2}\sqrt{n}$ and bounding the tail probability of a sum of geometric random variables (see, e.g., [28, Theorem 2.1]) we obtain,

$$\mathbb{P}\Big(L_M \ge \sqrt{n}\Big) \le e^{-\frac{1}{2}p(\beta)(1-\ln 2)\sqrt{n}} \triangleq \theta_n(\beta). \tag{139}$$

For any $\beta > 0$, $\theta_n(\beta)$ can be made arbitrarily small by taking n large enough. Thus, with high probability, we

have that the first $M = \frac{p(\beta)}{2}\sqrt{n}$ contigs ends before exceeding v_r:

$$\mathbb{P}\Big(L_M \le \sqrt{n} - 1\Big) \ge 1 - \theta_n(\beta). \tag{140}$$

The above bound alleviates the need to deal with boundary effect when considering the first M contigs formed after the sprinkling phase.

We next analyze the length of each such contig. Intersected with the event that $\big\{L_M \le \sqrt{n} - 1\big\}$, the length L_i has a geometric distribution with parameter $p(\beta)$. Let $G \sim \mathsf{Geo}\big(p(\beta)\big)$ and denote $\ell(\beta) = \mathbb{E}G = \frac{1}{p(\beta)}$. The Paley-Zygmund inequality states that for any $\theta \in (0,1)$, we have

$$\mathbb{P}\Big(N > \theta\mathbb{E}N\Big) \ge (1-\theta)^2 \frac{\big(\mathbb{E}N\big)^2}{\mathbb{E}[N^2]}. \tag{141}$$

Fixing any $\alpha \in (0,1)$, this gives

$$\mathbb{P}\Big(G \ge \alpha\ell(\beta)\Big) \ge \frac{(1-\alpha)^2}{2 - p(\beta)}. \tag{142}$$

We related the above back to L_i by

$$\begin{aligned}
\mathbb{P}(\Big(L_i \ge \alpha\ell(\beta)\Big) &\ge \mathbb{P}\left(\Big\{L_i > \alpha\ell(\beta)\Big\} \cap \Big\{L_M \le \sqrt{n} - 1\Big\}\right) \\
&\ge \mathbb{P}\left(\Big\{G > \alpha\ell(\beta)\Big\} \cap \Big\{L_M \le \sqrt{n} - 1\Big\}\right) \\
&\ge \big(1 - \theta_n(\beta)\big)\frac{(1-\alpha)^2}{2 - p(\beta)}.
\end{aligned} \tag{143}$$

Our last step in the analysis of Phase 1, is to show that (143) implies that there are many contigs of length at least $\ell(\beta)$. Let us extend the notation of L_i to all $i \in [\sqrt{n}]$, by defining $L_i = 0$ for all i with $X_{t_0}^{(t_0)}\big((i,1)\big) = +1$. In words, if no contig starts at site $(i,1)$, then the length associated with i is zero. The number of contigs whose length is at least $\alpha\ell(\beta)$ is $N \triangleq \sum_{i=1}^{\sqrt{n}} \mathbb{1}_{\left\{L_i > \alpha\ell(\beta)\right\}}$. By (143), we have

$$\mathbb{E}N = \sum_{i=1}^{\sqrt{n}} \mathbb{P}\Big(L_i > \alpha\ell(\beta)\Big) \ge \frac{\big(1 - \theta_n(\beta)\big)(1-\alpha)^2}{2 - p(\beta)}\sqrt{n}. \tag{144}$$

Furthermore, since $N \le \frac{\sqrt{n}}{\alpha\ell(\beta)}$ almost surely, $\mathbb{E}[N^2]$ is upper bounded by $\frac{p^2(\beta)n}{\alpha^2}$. Using the Paley-Zygmund inequality again, we have that

$$\mathbb{P}\left(N > \gamma\frac{\big(1 - \theta_n(\beta)\big)(1-\alpha)^2}{2 - p(\beta)}\sqrt{n}\right) \ge (1-\gamma)^2 \frac{\alpha^2(1-\alpha)^4}{p^2(\beta)\big(2 - p(\beta)\big)^2}, \tag{145}$$

for any $\gamma \in (0,1)$.

Summarizing the derivation up until this point, we see that, with high probability, the sprinkling phase of the $\big(X_t^{(t_0)}\big)_{t\in[t_0,2t_0)}$ dynamics terminates with many contigs each of length at least $\alpha\ell(\beta)$. This allows us to approximate from below the expected number of pluses at time $2t_0$ as

$$\mathbb{E}\hat{N}_{\mathcal{B}}^{(+)}(2t_0) \ge \mathbb{P}\Big(N \ge \gamma\mathbb{E}N\Big)\mathbb{E}\Big[\hat{N}_{\mathcal{B}}^{(+)}(2t_0)\Big|N > \gamma\mathbb{E}N\Big]$$

$$\geq \mathbb{P}\Big(N \geq \gamma \mathbb{E}N\Big) \cdot \gamma \mathbb{E}N \cdot \mathbb{E}\Big[\hat{N}_{\ell_0}^{(+)}(t_0)\Big], \tag{146}$$

where $\hat{N}_{t_0}^{(+)}\big(\ell(\beta)\big)$ is the number of pluses that survived for t_0 time in a single contig of length $\ell_0 \triangleq \alpha\ell(\beta)$ during the 2nd phase of our dynamics. Plugging (144)-(145) into (146), we obtain

$$\begin{aligned}\mathbb{E}\hat{N}_{\mathcal{B}}^{(+)}(2t_0) &\geq (1-\gamma)^2 \frac{\alpha^2(1-\alpha)^4}{p^2(\beta)\big(2-p(\beta)\big)^2} \cdot \gamma \frac{\big(1-\theta_n(\beta)\big)(1-\alpha)^2}{2-p(\beta)} \sqrt{n} \cdot \mathbb{E}\Big[\hat{N}_{\ell_0}^{(+)}\big(t_0)\big)\Big] \\ &= \gamma(1-\gamma)^2 \frac{\alpha^2(1-\alpha)^6\big(1-\theta_n(\beta)\big)}{p^2(\beta)\big(2-p(\beta)\big)^3} \sqrt{n} \cdot \mathbb{E}\Big[\hat{N}_{\ell_0}^{(+)}\big(t_0)\big)\Big]. \end{aligned} \tag{147}$$

2) Erosion Phase: It remains to analyze the expected value of $\hat{N}_{\ell_0}^{(+)}\big(t_0\big)$. During the second phase, each contig is eaten from both its side with speed $\phi(1)$. For any $\lambda \in (0,1)$, consider the probability:

$$\mathbb{P}\left(\hat{N}_{\ell_0}^{(+)}\big(t_0\big) > \lambda\ell_0\right) \geq 1 - \mathbb{P}\left(\left\lceil \frac{1-\lambda}{2}\ell_0 \right\rceil \text{ pluses eaten from each side}\right). \tag{148}$$

The event that exactly $k_0 \triangleq \left\lceil \frac{1-\lambda}{2}\ell_0 \right\rceil$ pluses from a given side of the contig have flipped in t_0 time is

$$\left\{\sum_{i=1}^{k_0} T_i \leq t_0\right\}, \tag{149}$$

where the T_i's are i.i.d. exponential random variables with $\mathbb{E}T_i = \frac{1}{\phi(1)}$. Observe that

$$\frac{1}{2}k_0\mathbb{E}T_1 \geq \frac{1-\lambda}{4}\ell_0\big(1+e^{-2\beta}\big) = \frac{(1-\lambda)\alpha}{4p(\beta)}\big(1+e^{-2\beta}\big) \overset{(a)}{\geq} \frac{(1-\lambda)\alpha}{4}e^{(2-c)\beta}\big(1+e^{-2\beta}\big), \tag{150}$$

where (a) follows because $t_0 = e^{c\beta}$ and $\frac{1}{p(\beta)} = \frac{1}{1-e^{-\phi(-1)t_0}} \geq e^{(2-c)\beta}$. Recalling that $c<1$, we see that for any β sufficiently large $t_0 \leq \frac{1}{2}k_0\mathbb{E}T_1$. We now have

$$\mathbb{P}\left(\left\lceil \frac{(1-\lambda)}{2}\ell_0 \right\rceil \text{ pluses eaten from each side}\right) \leq \mathbb{P}\left(\sum_{i=1}^{k_0} T_i \leq t_0\right) \leq \mathbb{P}\left(\sum_{i=1}^{k_0} T_i \leq \frac{1}{2}k_0\mathbb{E}T_1\right) \overset{(a)}{\leq} e^{-\kappa\frac{\alpha(1-\lambda)}{2p(\beta)}}. \tag{151}$$

where (a) bounds the tail probability if a sum of exponential random variables (see, e.g., [28, Theorem 5.1 Part (iii)]) and $\kappa = \big(\ln 2 - \frac{1}{2}\big) > 0$. Denoting the RHS of (151) by $s(\beta)$ (which exponentially fast to 0 with β) and inserting back to (148), we conclude that

$$\mathbb{P}\left(\hat{N}_{\ell_0}^{(+)}\big(t_0\big) > \lambda\ell_0\right) \geq 1 - s(\beta). \tag{152}$$

Collecting the pieces and proceeding from (147), we have

$$\begin{aligned}\mathbb{E}\hat{N}_{\mathcal{B}}^{(+)}(2t_0) &\geq \gamma(1-\gamma)^2 \frac{\alpha^2(1-\alpha)^6\big(1-\theta_n(\beta)\big)}{p^2(\beta)\big(2-p(\beta)\big)^3} \sqrt{n} \cdot \mathbb{E}\Big[\hat{N}_{\ell_0}^{(+)}\big(t_0\big)\Big] \\ &\overset{(a)}{\geq} \gamma(1-\gamma)^2 \frac{\alpha^2(1-\alpha)^6\big(1-\theta_n(\beta)\big)}{p^2(\beta)\big(2-p(\beta)\big)^3} \sqrt{n} \cdot \big(1-s(\beta)\big)\lambda\ell_0 \\ &\overset{(b)}{\geq} \gamma(1-\gamma)^2 \frac{\alpha^2(1-\alpha)^6\big(1-\theta_n(\beta)\big)}{8p^3(\beta)\big(2-p(\beta)\big)^3} \big(1-s(\beta)\big)\lambda\sqrt{n} \end{aligned} \tag{153}$$

where (a) uses (152), while (b) substitutes $\ell_0 = \frac{\alpha}{p(\beta)}$ and uses $(2-p(\beta)) < 2$. To conclude the proof we tune the

parameters α, γ and λ as follows. First, let $\gamma = p^2(\beta)$, which is a valid choice since we also requires $\gamma \in (0,1)$ (see (145)). Furthermore, set $\lambda = \frac{8}{\alpha^2}p(\beta)$ and increase β sufficiently so that $\lambda \in (0,1)$. Since λ can be made arbitrarily small with β, it is readily verified that with this choice $t_0 \leq \frac{1}{2}k_0\mathbb{E}T_1$ still holds (see (150)). Together with (134), the bound from (153) gives

$$\mathbb{E}N_1^{(+)}(t_0) \geq \big(1-p^2(\beta)\big)^2\big(1-\theta_n(\beta)\big)\big(1-s(\beta)\big)(1-\alpha)^6\sqrt{n}, \tag{154}$$

for $t_0 = e^{c\beta}$, with any $c < 1$.

Fix any $c' \in (0,1)$. To complete the proof of (60) , i.e., that

$$\mathbb{E}N_1^{(+)}(t_0) \geq c'\sqrt{n}, \tag{155}$$

recall that $p(\beta)$ and $s(\beta)$ decay to 0 as β increases (independently of n), and $\lim_{n\to\infty}\theta_n(\beta) = 0$ for all $\beta > 0$. Setting $\alpha \in (0,1)$ as a sufficiently small constant, taking β large enough and adjusting n produces the result.

APPENDIX H
PROOF OF LEMMA 2

Since the monochromatic connected components $\big\{\mathcal{C}_i(\sigma)\big\}_{i\in[L]}$ are separated by graph distance of at least 2, to prove the lemma it suffices to show that

$$N_u^+(\sigma_i) - N_u^-(\sigma_i) + 2N_d^+(\sigma_i) - 2N_d^-(\sigma_i) = 4 \tag{156}$$

for each $i \in [L]$, where $\sigma_i \in \Omega_{\mathbb{Z}^2}$ is the configuration that agrees with σ on $\mathcal{C}_i(\sigma)$ and is -1 everywhere else.

To evaluate the LHS of (156) let $\gamma_i(s)$, $s \in [0, s_{\mathsf{f}}]$, be a smooth counter-clockwise parameterization of $\gamma(\sigma_i)$ which slightly rounds its corners. Since $\gamma(\sigma_i)$ is a closed simple curve we have $\gamma_i(0) = \gamma_i(s_{\mathsf{f}})$. [9] Thus, $\gamma_i(s)$ is a continuously differentiable curve.

Consider the 2-dimensional Gauss map from $\gamma_i(s)$ to the unit circle, given by the trajectory of the unit vector $v(s) = \frac{\gamma_i'(s)}{|\gamma_i'(s)|}$. As one travels along $\gamma_i(s)$, the $v(s)$ vector travels on the circle. The trajectory of $v(s)$ is mostly constant with quick sweeps happening when we approach the slightly smoothed corners of $\gamma_i(s)$. Sweeps of $v(s)$ are reveal unstable or disagreeing plus- or minus-labeled sites in σ_i as follows.

- a single sweep in the counter-clockwise direction corresponds to a site in $N_u^+(\sigma_i)$;
- a double sweep in the counter-clockwise direction corresponds to a site in $N_d^+(\sigma_i)$;
- a single sweep in the clockwise direction corresponds to a site in $N_u^-(\sigma_i)$;
- a double sweep in the clockwise direction corresponds to a site in $N_d^-(\sigma_i)$.

The only way a triple sweep can occur is if $\mathcal{C}_i(\sigma)$ is just a single site of positive spin (in which case we have four counter-clockwise consecutive sweeps). This case is excluded by the assumption that $\big|\mathcal{C}_i(\sigma)\big| \geq 2$.

Since the index of the simple closed curve is 1, overall the vector $v(t)$ has to complete exactly one full rotation on a unit circle. Hopf's Umlaufsatz thus produces (156), which, in turn, implies the result of Lemma 2.

[9] The particular parameterization in use is of no consequence here.

APPENDIX I
PROOF OF COROLLARY 4

First observe that the result is trivial for the final configuration (i.e., when $(\bar{X}_t)_{t\geq 0}$ stops). Namely, denoting the final configuration by $\zeta \in \Omega_{\mathbb{Z}^2}$, where ζ is the configuration for which $\zeta\big((i,j)\big) = +1$ if and only if $i,j \geq \ell+1$, we have $\mathcal{N}_u^+(\zeta) = \big\{(\ell+1,\ell+1)\big\}$ and $\mathcal{N}_d^+(\zeta) = \mathcal{N}_d^-(\zeta) = \mathcal{N}_u^-(\zeta) = \emptyset$.

Now, assuming $\bar{X}_t = \sigma \neq \zeta$, define $\bar{\sigma} \in \Omega_{\mathbb{Z}^2}$ as

$$\bar{\sigma}\big((i,j)\big) = \begin{cases} \sigma\big((i,j)\big), & i,j \leq \ell \\ -1, & \text{otherwise} \end{cases}. \tag{157}$$

For $\bar{\sigma}$, we apply Lemma 2 with $L=1$, $\mathcal{C}_1(\bar{\sigma}) = \mathcal{C}(\bar{\sigma}) = \mathcal{C}(\bar{\sigma}) \cap [\ell]^2$ and $\mathcal{C}_0(\bar{\sigma}) = \mathbb{Z}^2 \setminus \mathcal{C}_1(\bar{\sigma})$ (the assumptions of the lemma are readily satisfied here). This gives

$$N_u^+(\bar{\sigma}) - N_u^-(\bar{\sigma}) + 2N_d^+(\bar{\sigma}) - 2N_d^-(\bar{\sigma}) = 4. \tag{158}$$

Let $v_1(\bar{\sigma})$ and $v_2(\bar{\sigma})$ be the bottom-right and top-left vertices in $\mathcal{C}_1(\bar{\sigma})$. We decompose $\gamma(\bar{\sigma})$ into three parts $\gamma_1(\bar{\sigma})$, $\gamma_2(\bar{\sigma})$ and $\gamma_3(\bar{\sigma})$. The first part $\gamma_1(\bar{\sigma})$ is a straight vertical segment from $v_1(\bar{\sigma}) + \big(\frac{1}{2}, -\frac{1}{2}\big)$ to $\big(\ell+\frac{1}{2}, \ell+\frac{1}{2}\big)$. The second part $\gamma_2(\bar{\sigma})$ is a straight horizontal segment between $\big(\ell+\frac{1}{2}, \ell+\frac{1}{2}\big)$ and $v_2(\bar{\sigma}) + \big(-\frac{1}{2}, \frac{1}{2}\big)$. The third and final part $\gamma_3(\bar{\sigma})$ connects $v_2(\bar{\sigma}) + \big(-\frac{1}{2}, \frac{1}{2}\big)$ back to $v_1(\bar{\sigma}) + \big(\frac{1}{2}, -\frac{1}{2}\big)$ by downwards and rightwards moves over segments of positive integer length. Note that every flippable site in $\bar{\sigma}$ must be adjacent from the top of from the right to a portion of $\gamma_3(\bar{\sigma})$. In particular, the shape of γ_3 implies that $\mathcal{N}_d^+(\bar{\sigma}) = \mathcal{N}_d^-(\bar{\sigma}) = \emptyset$, which further gives

$$\mathcal{N}_d^+(\sigma) = \mathcal{N}_d^-(\sigma) = \emptyset. \tag{159}$$

Finally, note that $N_u^+(\bar{\sigma}) = N_u^+(\sigma) + 3$. This follows by the definition of $\bar{\sigma}$ which relabels all positive spins of σ in the 1st quadrant outside of $[\ell]^2$ as -1. Doing so adds 3 sites (namely, $v_1(\sigma)$, $v_2(\sigma)$ and (ℓ,ℓ)) to $\mathcal{N}_u^+(\bar{\sigma})$ that did not exist in the original $\mathcal{N}_u^+(\sigma)$. Inserting that, along with (159), into (158) completes the proof.

APPENDIX J
PROOF OF LEMMA 3

Assertion (1), that $\mathcal{N}_d^+(\tilde{X}_t) = \mathcal{N}_d^-(\tilde{X}_t) = \emptyset$, is a direct consequence of the definition of the auxiliary dynamics. In particular, $\mathcal{N}_d^+(\tilde{X}_t) = \emptyset$ follows due to the application of E after each flip, while $\mathcal{N}_d^-(\tilde{X}_t) = \emptyset$ holds as a result of $\mathcal{N}_d^-(\tilde{\rho}) = \emptyset$ and the evolution pattern.

To prove the 2nd statement, we show that for any $t \geq 0$, if $\gamma(\tilde{X}_t)$ is simple and closed then $\gamma(\tilde{X}_t^v)$ is simple and closed, for any flippable site $v \in \mathcal{N}_u(\tilde{X}_t)$. First note that $\gamma(\tilde{X}_0) = \gamma(\tilde{\rho})$ is simple and closed. Next, fix any $t \geq 0$ and assume $\tilde{X}_t = \sigma$ is such that $\gamma(\sigma)$ is simple and closed. Let $v \in \mathcal{N}_u(\sigma)$ and consider the following.

If $v \in \mathcal{N}_u^-(\sigma)$, we have $\mathcal{D}(\sigma) \subseteq \mathcal{D}(\sigma^v)$, which implies that $\gamma(\sigma^v)$ stays simple and closed. To account for $v \in \mathcal{N}_u^+$, we need some definitions. Define, respectively, the top, bottom, right and left frames of $\mathcal{C}(\tilde{X}_0)$ as:

$$\mathcal{F}_t \triangleq \big\{(i,k+1)\big\}_{i\in[-k:k]} \tag{160a}$$

$$\mathcal{F}_b \triangleq \big\{(i,-k-1)\big\}_{i\in[-k:k]} \tag{160b}$$

$$\mathcal{F}_r \triangleq \big\{(k+1,j)\big\}_{j\in[-k:k]} \tag{160c}$$

$$\mathcal{F}_l \triangleq \big\{(-k-1,j)\big\}_{j\in[-k:k]}. \tag{160d}$$

Furthermore, set

$$\mathcal{T}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| v = \operatorname*{argmin}_{u\in\mathcal{C}(\sigma)} d(u,\mathcal{F}_t) \right\} \tag{161a}$$

$$\mathcal{B}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| v = \operatorname*{argmin}_{u\in\mathcal{C}(\sigma)} d(u,\mathcal{F}_b) \right\} \tag{161b}$$

$$\mathcal{R}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| v = \operatorname*{argmin}_{u\in\mathcal{C}(\sigma)} d(u,\mathcal{F}_r) \right\} \tag{161c}$$

$$\mathcal{L}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| v = \operatorname*{argmin}_{u\in\mathcal{C}(\sigma)} d(u,\mathcal{F}_l) \right\}. \tag{161d}$$

These sets contain the sites in $\mathcal{C}(\sigma)$ that are closest to each of the above defined frames. The corners of these sets are defined as

$$v_t^{(r)} \triangleq \operatorname*{argmin}_{u\in\mathcal{T}(\sigma)} d\big(u,\mathcal{R}(\sigma)\big) \quad ; \quad v_t^{(l)} \triangleq \operatorname*{argmin}_{u\in\mathcal{T}(\sigma)} d\big(u,\mathcal{L}(\sigma)\big) \tag{162a}$$

$$v_b^{(r)} \triangleq \operatorname*{argmin}_{u\in\mathcal{B}(\sigma)} d\big(u,\mathcal{R}(\sigma)\big) \quad ; \quad v_b^{(l)} \triangleq \operatorname*{argmin}_{u\in\mathcal{B}(\sigma)} d\big(u,\mathcal{L}(\sigma)\big) \tag{162b}$$

$$v_r^{(t)} \triangleq \operatorname*{argmin}_{u\in\mathcal{R}(\sigma)} d\big(u,\mathcal{T}(\sigma)\big) \quad ; \quad v_r^{(b)} \triangleq \operatorname*{argmin}_{u\in\mathcal{R}(\sigma)} d\big(u,\mathcal{B}(\sigma)\big) \tag{162c}$$

$$v_l^{(t)} \triangleq \operatorname*{argmin}_{u\in\mathcal{L}(\sigma)} d\big(u,\mathcal{T}(\sigma)\big) \quad ; \quad v_l^{(b)} \triangleq \operatorname*{argmin}_{u\in\mathcal{L}(\sigma)} d\big(u,\mathcal{B}(\sigma)\big) \tag{162d}$$

Note that since $\mathcal{N}_d^+(\sigma)=\emptyset$, the above defined corners are unique and different from one another.

Through the quantities from (162d) we consider four particular pieces of the curve $\gamma(\sigma)$. Let:

- $\gamma_{r\to t}(\sigma)$ be the portion of $\gamma(\sigma)$ that extends from $v_r^{(t)}+\big(-\frac{1}{2},\frac{1}{2}\big)$ to $v_t^{(r)}+\big(\frac{1}{2},-\frac{1}{2}\big)$.
- $\gamma_{t\to l}(\sigma)$ be the portion of $\gamma(\sigma)$ that extends from $v_t^{(l)}-\big(\frac{1}{2},\frac{1}{2}\big)$ to $v_l^{(t)}+\big(\frac{1}{2},\frac{1}{2}\big)$.
- $\gamma_{l\to b}(\sigma)$ be the portion of $\gamma(\sigma)$ that extends from $v_l^{(b)}+\big(\frac{1}{2},-\frac{1}{2}\big)$ to $v_b^{(l)}+\big(-\frac{1}{2},\frac{1}{2}\big)$.
- $\gamma_{b\to r}(\sigma)$ be the portion of $\gamma(\sigma)$ that extends from $v_b^{(r)}+\big(\frac{1}{2},\frac{1}{2}\big)$ to $v_r^{(b)}-\big(\frac{1}{2},\frac{1}{2}\big)$.

See Fig. 18 for an illustration of these portions of $\gamma(\sigma)$.

The last definition we need is of the sets of vertices that are outlined by each piece of $\gamma(\sigma)$. We set:

$$\mathcal{C}_{r\to t}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| \left| \mathcal{B}_\infty\left(v,\frac{1}{2}\right)\cap\gamma_{r\to l}(\sigma)\right| > 1 \right\} \tag{163a}$$

$$\mathcal{C}_{t\to l}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| \left| \mathcal{B}_\infty\left(v,\frac{1}{2}\right)\cap\gamma_{t\to l}(\sigma)\right| > 1 \right\} \tag{163b}$$

$$\mathcal{C}_{l\to b}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| \left| \mathcal{B}_\infty\left(v,\frac{1}{2}\right)\cap\gamma_{l\to b}(\sigma)\right| > 1 \right\} \tag{163c}$$

$$\mathcal{C}_{b\to r}(\sigma) \triangleq \left\{ v\in\mathcal{C}(\sigma) \middle| \left| \mathcal{B}_\infty\left(v,\frac{1}{2}\right)\cap\gamma_{b\to r}(\sigma)\right| > 1 \right\}. \tag{163d}$$

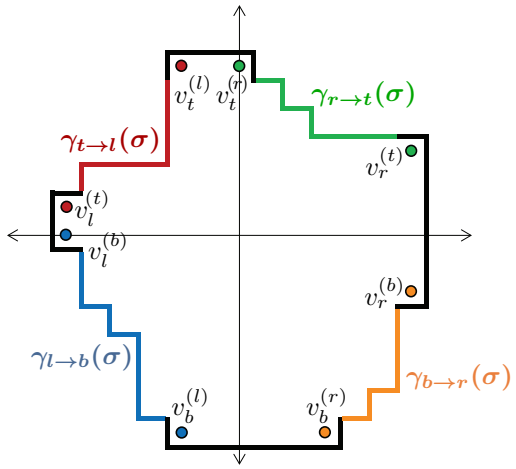


Fig. 18: The portions $\gamma_{r\to t}(\sigma)$, $\gamma_{t\to l}(\sigma)$, $\gamma_{l\to b}(\sigma)$ and $\gamma_{b\to r}(\sigma)$ are shown by the green, red, blue and orange curves, respectively. The remaining pieces of $\gamma(\sigma)$ are colored in black. The corner of $\mathcal{R}(\sigma)$, $\mathcal{T}(\sigma)$, $\mathcal{L}(\sigma)$ and $\mathcal{B}(\sigma)$ that define the aforementioned pieces of $\gamma(\sigma)$ are shown by circles colored as the corresponding piece of $\gamma(\sigma)$.

Defining the above sets through the condition that the cardinality of the intersection between the L^∞ ball and the corresponding piece of $\gamma(\sigma)$ is grater than 1 is of pure technical nature. To clarify, the set $\mathcal{C}_{r\to t}(\sigma)$ should contain all vertices whose surrounding L^∞ ball of radius $\frac{1}{2}$ contributes at least one side to $\gamma_{r\to t}(\sigma)$. In such a case, the cardinality of the intersection is the cardinality of the continuum. We do not want, e.g., $v_t^{(r)}$ (or $v_r^{(t)}$) to be inside $\mathcal{C}_{r\to t}(\sigma)$. Noting that $\mathcal{B}_\infty\left(v_t^{(r)},\frac{1}{2}\right)\cap\gamma_{r\to t}(\sigma)=\left\{v_t^{(r)}+\left(\frac{1}{2},-\frac{1}{2}\right)\right\}$, the above definition excludes the corners from being in these sets. With respect to the above definitions, one can uniquely decompose the internal boundary of $\mathcal{C}(\sigma)$ as

$$\partial_{\mathsf{Int}}\mathcal{C}(\sigma)=\mathcal{C}_{r\to t}(\sigma)\uplus\mathcal{T}(\sigma)\uplus\mathcal{C}_{t\to l}(\sigma)\uplus\mathcal{L}(\sigma)\uplus\mathcal{C}_{l\to b}(\sigma)\uplus\mathcal{B}(\sigma)\uplus\mathcal{C}_{b\to r}(\sigma)\uplus\mathcal{R}(\sigma). \tag{164}$$

We are now ready to prove that $\gamma(\sigma^v)$ is simple and connected, for any $v\in\mathcal{N}_u^+(\sigma)$. Fix $v\in\mathcal{N}_u^+(\sigma)$ and assume, without loss of generality, that $v\in\mathcal{C}_{r\to t}(\sigma)\uplus\mathcal{T}(\sigma)$.

We consider two cases. First, assume $v\in\mathcal{T}(\sigma)$ and $|\mathcal{T}(\sigma)|=2$. Recalling that $\mathcal{T}(\sigma)$ always contains $v_t^{(r)}$ and $v_t^{(l)}$, we deduce that $\mathcal{T}(\sigma)=\left\{v_t^{(r)},v_t^{(l)}\right\}$. Assume $v=v_t^{(r)}$ and, for brevity, denote $u\triangleq v_t^{(l)}$. Flipping v will cause u, which has a positive spin in σ, to disagree with the negative spin (in σ^v) of 3 of its 4 neighbors. As $\mathcal{N}_d^+(\sigma)=\emptyset$, we get $\mathcal{N}_d^+(\sigma^v)=\{u\}$. By definition of the auxiliary dynamics, since $\mathcal{N}_d^+(\sigma^v)$ is non-empty and contains a single element, the transformation E is then be applied. This gives $\mathsf{E}(\sigma^v)=(\sigma^v)^u$, i.e., the spin of u will be flipped (from positive to negative). Thus, the next state of the dynamics is $(\sigma^v)^u$, for which $\mathcal{C}\left((\sigma^v)^u\right)=\mathcal{C}(\sigma)\setminus\mathcal{T}(\sigma)$. The resulting curve $\gamma\left((\sigma^v)^u\right)$ remains connected and simple.

The remaining case is when $v\in\mathcal{T}(\sigma)$ with $|v\in\mathcal{T}(\sigma)|>3$ or $v\in\mathcal{C}_{r\to t}(\sigma)$. These two scenarios can be treated as one because both satisfy $\mathsf{E}(\sigma^v)=\sigma^v$. To see that $\gamma(\sigma^v)$ is simple and connected, assume to the contrary that it is not. This can happen only if $\sigma\big(v-(1,1)\big)=-1$ (any other situation would contradict the fact that the original curve $\gamma(\sigma)$ is closed and simple). Since $\sigma(v)=+1$, having a negative spin at $v-(1,1)$ implies that $\sigma(u)=-1$,

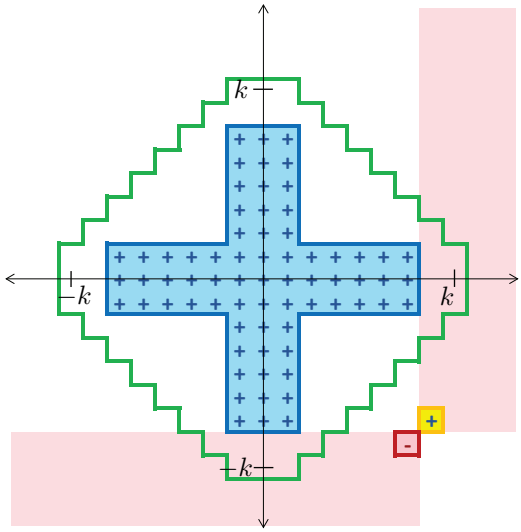


Fig. 19: An illustration of the obtained contradiction for $k = 8$, $m = \frac{3}{4}k = 6$.The site v (here shown for case (i) when $i_v \geq m+1$ and $j_v \leq -m$) and u are outlined by the yellow and red borders, respectively. The pink regions enclose sites that must have a negative spin in order for $v \in \mathcal{N}_u^+(\sigma)$ and $\sigma(u) = -1$ to hold. The original rhombus-shaped plus-labeled component $\mathcal{C}(\tilde{\rho})$ is enclosed by the green curve. The set $\mathcal{T}_m$ is shown in blue.

for all $u = v - (i,j)$ with $i,j \geq 1$. Furthermore, since $v \in \mathcal{N}_u^+(\sigma)$ with $\sigma\big(v+(1,0)\big) = \sigma\big(v+(0,1)\big) = -1$, we have that $\sigma(w) = -1$ for all $w \in \big\{v + (i,j) \big| i,j \geq 0\big\} \setminus \{v\}$.

Now, by definition of the $(\tilde{X}_t)_{t\geq 0}$ dynamics, we have that $\mathcal{T}_m \subseteq \mathcal{C}(\sigma)$.[10] Denoting the coordinates of v by $v = (i_v, j_v)$, the fact that $\mathcal{T}_m \subseteq \mathcal{C}(\sigma)$ implies that one of the following two assertions hold:

(i) $i_v \geq m+1$ and $j_v \leq -m$;

(ii) $i_v \leq -m$ and $j_v \geq m+1$.

However, as $m = \frac{3}{4}k$, in both cases we get that $v \notin \mathcal{C}(\tilde{\rho})$, which is a contradiction. This idea is illustrated in Fig. 19. This concludes the proof that if $\gamma(\sigma)$ is a simple and closed curve in $\mathbb{R}^2$, then so is $\gamma(\sigma^v)$, for all $v \in \mathcal{N}_u(\sigma)$.

The third claim essentially follows from the 2nd one. First note that $\mathcal{M}_u^+(\sigma) \subseteq \Big\{v_r^{(b)}, v_r^{(t)}, v_t^{(r)}, v_t^{(l)}, v_l^{(t)}, v_l^{(b)}, v_b^{(l)}, v_b^{(l)}\Big\}$, whence $\big|\mathcal{M}_u^+(\sigma)\big| \leq 8$. If $\mathcal{M}_u^+(\sigma) \neq \emptyset$, the fact that once $v \in \mathcal{M}_u^+(\sigma)$ is flipped we have $\big|\mathcal{C}\big(\mathsf{E}(\sigma^v)\big)\big| = \big|\mathcal{C}(\sigma)\big| - 2$ was already observed in the proof of Claim 2).

The last item to establish is that $\mathcal{N}_1^+(\sigma) = \emptyset$ (see (114a)). In other words, we need to demonstrate that there are no unstable plus-labeled sites with disagreeing neighbors located only horizontally (to the right and left) or only vertically (above and below). This follows by combining Claims 1) and 2) proven above. Fix $t \geq 0$ and let $\tilde{X}_t = \sigma$. By Claim 2), $\gamma(\sigma)$ is simple and closed, and, in particular, its interior $\mathcal{D}(\sigma)$ is a connected set. Assume in contradiction that there exists $v \in \mathcal{N}_1^+(\sigma)$. Without loss of generality we assume that $v \in \mathcal{C}_{r\to t}(\sigma) \uplus \mathcal{T}(\sigma)$, which implies that $\sigma\big(v+(1,0)\big) = \sigma\big(v-(1,0)\big) = -1$. As argued in the proof of Claim 2, $\sigma\big(v+(1,0)\big) = -1$ implies that $\sigma\big(v+(1,0)+(i,j)\big) = -1$, for all $i,j \geq 0$. Similarly, because $\sigma\big(v-(1,0)\big) = -1$, we have

[10] Recall that $\mathcal{T}_m = \big\{(i,j)\big\}_{\substack{i\in[-m:m]\\ j\in[-1:1]}} \cup \big\{(i,j)\big\}_{\substack{i\in[-1:1]\\ j\in[-m:m]}}$, with $m = \frac{3}{4}k$.

$\sigma\big(v-(1,0)+(i,j)\big)=-1$, for all $i\le 0$ and $j\ge 0$. Denote $v=(i_v,j_v)$ and let $j_t=\max\big\{j\ge j_v\big|(i_v,j)\in\mathcal{C}(\sigma)\big\}$ (note that this set is non-empty because it always contains $v+(0,1)$). The site $(i_v,j_t)\in\mathcal{C}(\sigma)$, but it has three neighbors with negative spin (to its right, to its left and above it). Therefore, $(i_v,j_t)\in\mathcal{N}_d^+(\sigma)$, which contradicts $\mathcal{N}_d^+(\sigma)=\emptyset$.

REFERENCES

[1] P. A. Franaszek. Sequence-state methods for run-length-limited coding. *IBM J. Res. Dev.*, 14(4):376–383, Jul. 1970.

[2] K. A. S. Immink. Runlength-limited sequences. *Proc. of IEEE*, 78(11):1745–1759, Nov. 1990.

[3] I. S. Reed and G. Solomon. Polynomial codes over certain finite fields. *J. Soc. Ind. Appl. Math.*, 8(2):300–304, 1960.

[4] R. J. Glauber. Time-dependent statistics of the Ising model. *J. Mathematical Phys.*, 4(2):294–307, Feb. 1963.

[5] E. Ising. Beitrag zur theorie des ferromagnetismus. *Zeitschrift für Physik A Hadrons and Nuclei*, 31(1):253–258, Feb. 1925.

[6] S. Friedli and Y. Velenik. *Statistical mechanics of lattice systems: a concrete mathematical introduction*. Cambridge University Press, 2017.

[7] D. A. Levin and Y. Peres; with contributions by E. L. Wilmer. *Markov Chains and Mixing Times*. Princeton University Press, Providence, RI, 2nd edition edition, 2017.

[8] E. Lubetzky, F. Martinelli, A. Sly, and F. L. Toninelli. Quasi-polynomial mixing of the 2D stochastic Ising model with "plus" boundary up to criticality. *J. Eur. Math. Soc.*, 15(2):339–386, 2013.

[9] P. Diaconis. The mathematics of mixing things up. *J. Stat. Phys.*, 144(3):445–458, Aug. 2011.

[10] L. Onsager. Crystal statistics. i. a two-dimensional model with an order-disorder transition. *Phys. Rev.*, 65(3-4):117, Feb. 1944.

[11] S. G. Brush. History of the Lenz-Ising model. *Rev. Mod. Phys.*, 39(4):883–893, Oct. 1967.

[12] D. Capocaccia, M. Cassandro, and E. Olivieri. A study of metastability in the Ising model. *Commun. Math. Phys.*, 39(3):185–205, Sep. 1974.

[13] E. B. Davies. Metastability and the Ising model. *J. Stat. Phys.*, 27(4):657–675, Apr. 1982.

[14] E. J. Neves and R. H. Schonmann. Critical droplets and metastability for a Glauber dynamics at very low temperatures. *Commun. Math. Phys.*, 137(2):209–230, Apr. 1991.

[15] E. N. M. Cirillo and J. L. Lebowitz. Metastability in the two-dimensional Ising model with free boundary conditions. *J. Stat. Phys.*, 90(1):211–226, Jan. 1998.

[16] V. Spirin, P. L. Krapivsky, and S. Redner. Fate of zero-temperature Ising ferromagnets. *Phys. Rev. E*, 63(3):036118, Feb. 2001.

[17] H. Lacoin, F. Simenhaus, and F. Toninelli. Zero-temperature 2D stochastic Ising model and anisotropic curve-shortening flow. *J. Eur. Math. Soc. (JEMS) 6*, 16(12):incoonu, Dec. 2014.

[18] M. Aizenman and R. Holley. Rapid convergence to equilibrium of stochastic Ising models in the Dobrushin Shlosman regime. In *Percolation theory and ergodic theory of infinite particle systems*, pages 1–11. Springer, New-York, NY, USA, 1987.

[19] R. B. Griffiths, C.-Y. Weng, and J. S. Langer. Relaxation times for metastable states in the mean-field model of a ferromagnet. *Phys. Rev.*, 149(1):301, Sep. 1966.

[20] D. A. Levin, M. J. Luczak, and Y. Peres. Glauber dynamics for the mean-field Ising model: cut-off, critical power law, and metastability. *Probab. Theory Relat. Fields*, 146(1):223–265, Jan. 2010.

[21] F. Martinelli. Lectures on Glauber dynamics for discrete spin models. In *Lectures on prob. theory and stat.*, pages 93–191. Springer, 1999.

[22] F. Martinelli. On the two-dimensional dynamical Ising model in the phase coexistence region. *J. Stat. Phys.*, 76(5-6):1179–1246, Sep. 1994.

[23] C. M. Grinstead and J. L. Snell. *Introduction to probability*. American Math. Soc., 2012.

[24] R. H. Schonmann. Second order large deviation estimates for ferromagnetic systems in the phase coexistence region. *Commun. Math. Phys.*, 112(3):409–422, Sep. 1987.

[25] L. E. Thomas. Bound on the mass gap for finite volume stochastic Ising models at low temperature. *Comm. Math. Phys.*, 126(1):1–11, Nov. 1989.

[26] N. Sugimine. Extension of thomas' result and upper bound on the spectral gap of $d(\ge 3)$-dimensional stochastic Ising models. *J. Math. Kyoto Univ.*, 42(1):141–160, 2002.

[27] H. V. Poor Y. Polyanskiy and S. Verdú. Channel coding rate in the finite blocklength regime. *IEEE Trans. Inf. Theory*, 56(5):2307–2359, May 2010.

[28] S. Janson. Tail bounds for sums of geometric and exponential variables. *Stat. Prob. Letters*, 135:1–6, Apr. 2018.