Information Storage in the Stochastic Ising Model at Low Temperature

Ziv Goldfeld
MIT
zivg@mit.edu

Guy Bresler
MIT
guy@mit.edu

Yury Polyanskiy
MIT
yp@mit.edu

***Abstract*—Motivated by questions of data stabilization in emerging magnetic storage technologies, we study the retention of information in interacting particle systems. The interactions between particles adhere to the stochastic Ising model (SIM) on the two-dimensional (2D) $\sqrt{n}\times\sqrt{n}$ grid. The measure of interest is the information capacity $I_n(t)\triangleq\max_{p_{X_0}} I(X_0;X_t)$, where the initial spin configuration X_0 is a user-controlled input and the output configuration X_t is produced by running t steps of Glauber dynamics. After the results on the zero-temperature regime reported last year, this work focuses on the positive but low temperature regime. We first show that storing more than a single bit for an exponential time is impossible when the initial configuration is drawn from the equilibrium distribution. Specifically, if X_0 is drawn according to the Gibbs measure, then $I(X_0;X_t)\leq 1+o(1)$ for $t\geq\exp\left(cn^{\frac{1}{4}+\epsilon}\right)$. On the other hand, when scaling time with β, we propose a stripe-based coding scheme that stores order of $\sqrt{n}$ bits for $\exp(\beta)$ time. Key to the analysis of the scheme is a new result on the survival time of a single plus-labeled stripe in a sea of minuses. Together, the 1-bit upper bound and the striped-based storage scheme constitute initial steps towards a general analysis of $I_n(t)$ for $\beta>0$.**

I. Introduction

The increasing demand for high-capacity storage devices inspired new methods for magnetic storage, such as shingled magnetic recording (SMR) [1], heat assisted magnetic recording (HAMR) [2] and bit-patterned media (BPM) [3]. The latter two, for instance, drastically reduce the area magnetic mediums allocate for storing each bit: from roughly 20-30 magnetic grains per bit in today's off-the-shelf hard-drives to a *single* magnetic grain per bit via BPM or HAMR. While shrinking of magnetization domains increases storage capacity, it also entails new challenges in stabilizing the written data long enough to allow later recovery. Specifically, in the time between writing and reading, the stored data (a configuration of states at which the particles comprising the medium are initiated) dissipates due to interparticle interactions driven by quantum/thermal fluctuations. The significance of this physical phenomenon for future hard-drive designs is the motivation for this work.

To address the stabilization of stored data we adhere to the perspective of [4], [5] and study information retention in locally-interacting particle systems. Specifically, a planar grid comprising n particles models the magnetic hard-drive. The encoder chooses an initial spin ($+1$ or -1) configuration X_0, in which to store the data. The read configuration X_t is produced by running t steps of the Glauber dynamics [6] for the Ising model [7] at inverse temperature $\beta\geq 0$. Fig. 1 visualizes the model in a block diagram. The amount of bits that can be stored for t time inside a $\sqrt{n}\times\sqrt{n}$ grid is measured by the information capacity $I_n(t)\triangleq\max_{p_{X_0}} I(X_0;X_t)$.

While [5] focused on the zero-temperature Glauber dynamics, here we consider the more practically-relevant positive temperature regime. At the extreme case of 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. Cooling the system to finite temperatures ($\beta>0$) introduces interactions between particles and changes the behavior of $I_n(t)$. Classical results on the 2D Ising model phase transition and mixing times [8] imply the following: for $\beta<\beta_c\triangleq\frac{1}{2}\log\left(1+\sqrt{2}\right)$, we have $I_n\big(\mathrm{poly}(n)\big)=0$, while for $\beta>\beta_c$, $I_n\big(\exp(\sqrt{n})\big)\geq 1$.[1]

Our first interest is whether anything beyond a single bit can be stored in the SIM on the 2D grid at $\beta>\beta_c$ for exponential time. To gain understanding consider two (stochastic) trajectories with one started from an all-plus configuration and another one from a fixed configuration 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. [9] 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 [10] 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 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 [9, 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

This work was supported in part by the National Science Foundation CAREER award under grant agreement CCF-12-53205 and by the Center for Science of Information (CSoI), an NSF Science and Technology Center, under grant agreement CCF-09-39370

[1] 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 $\beta>\beta_c$ [8].

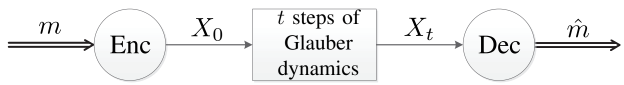


Fig. 1: Data storage inside interacting particle systems: The encoder maps the data m into the initial configuration X_0. The dynamics evolves X_0 by t time steps. The decoder recovers m from X_t.

proven in Section III.

The positive temperature regime also enables scaling time with β (instead of n). It turns out that much more that a single bit can stored in the SIM at low temperature β for $\exp(\beta)$ time. Specifically, 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 IV). Key to the analysis of the scheme is a novel result showing that a single stripe of pluses surrounded by minuses at the bottom of an $\sqrt{n} \times \sqrt{n}$ retains at least half of its original $\sqrt{n}$ pluses for at least $e^{c\beta}$ time, where c is a numerical constant. The storage results provided here constitute first steps towards a comprehensive analysis of the positive-but-low temperature regime.

II. The Stochastic Ising Model

For $k \in \mathbb{N}$, we set $[k] \triangleq \{1, \ldots, k\}$. Let $G_n = (\mathcal{V}_n, \mathcal{E}_n)$ be a square grid of side $\sqrt{n} \in \mathbb{N}$, where $\mathcal{V}_n = \big\{(i,j)\big\}_{i,j \in [\sqrt{n}]}$.[2] The neighborhood of v is $\mathcal{N}_v \triangleq \{w \in \mathcal{V}_n | \{v,w\} \in \mathcal{E}_n\}$.

Fix $\sqrt{n} \in \mathbb{N}$ and let $\Omega_n \triangleq \{-1,+1\}^{\mathcal{V}_n}$. Denote by $\boxplus, \boxminus \in \Omega_n$ the all-plus and all-minus configurations, respectively. For every $\sigma \in \Omega_n$ and $v \in \mathcal{V}_n$, $\sigma(v)$ is the value of σ at v. The Hamiltonian for the Ising model on G_n is given by

$$\mathcal{H}(\sigma) \triangleq -\sum_{\{u,v\} \in \mathcal{E}_n} \sigma(u)\sigma(v), \quad \sigma \in \Omega_n. \tag{1}$$

The Gibbs measure over G_n at inverse temperature $\beta > 0$ and free boundary conditions is

$$\pi(\sigma) = \frac{1}{Z(\beta)} e^{-\beta \mathcal{H}(\sigma)}, \quad \sigma \in \Omega_n \tag{2}$$

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

The SIM is a discrete-time Markov chain (MC) on the state space Ω_n that is reversible with respect to π, for any $\beta > 0$. At each time step, a vertex $v \in \mathcal{V}_n$ is chosen uniformly at random; the spin at v is refreshed to a new value $s \in \{-1,+1\}$ by sampling the conditional Gibbs measure

$$\pi_v(s) \triangleq \pi\big(s \big| \{\sigma(u)\}_{u \neq v}\big) = \pi\big(s \big| \{\sigma(u)\}_{u \in \mathcal{N}_v}\big). \tag{3}$$

Let P be the induced transition kernel and $(X_t)_{t \in \mathbb{N}_0}$ be the MC. If $X_0 \sim p_{X_0}$, then $(X_0, X_t) \sim p_{X_0,X_t}(\sigma, \eta) \triangleq p_{X_0}(\sigma) P^t(\sigma, \eta)$, where P^t is the t-step kernel. The mutual information $I(X_0; X_t)$ is taken with respect to p_{X_0,X_t}.

Remark 1 (Continuous-Time Dynamics) *The SIM can be also set up in continuous-time by assigning with each $v \in \mathcal{V}_n$ an independent Possion clock of rate 1. When the clock at v rings, its spin is redrawn via* (3). *The continuous- and discrete-time versions are equivalent in terms of their information capacity when the Poisson rate is thinned to $\frac{1}{n}$ [11, Proposition 4]. Nonetheless, in Section IV we consider the vanilla rate 1 version when studying $I_n(t)$ for $\exp(\beta)$ time scales.*

[2]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 of large n, and the assumption that $\sqrt{n} \in \mathbb{N}$ has no affect on the asymptotic behavior.

III. 1-Bit Upper Bound under Gibbs Initialization

We show that for sufficiently large $\beta > \beta_c$, it is impossible to store more than a single bit in the SIM for $t \geq e^{cn^{\frac{1}{4}+\epsilon}}$ time by drawing the initial configuration from the Gibbs distribution. This result leverages Proposition 5.2 from [9] which we restate in the following.[3] Some preliminary definition are due.

Fix $\beta \in (0, \infty)$ and let $(X_t)_{t \geq 0}$ be the discrete-time dynamics with $X_0 \sim \pi$. We use $(X_t^\sigma)_{t \geq 0}$ for the dynamics initiated at $X_0 = \sigma$. The case when staring from the all-plus state $\boxplus$ is distinguished by denoting $Y_t \triangleq X_t^{\boxplus}$. Finally, denote the magnetization of $\sigma \in \Omega_n$ by $m(\sigma) \triangleq \frac{1}{n} \sum_{v \in \mathcal{V}_n} \sigma(v)$.

To state the result of [9, Proposition 5.2] we first need to set up the Markov chains $\Big\{(X_t^\sigma)_{t \geq 0}\Big\}_{\sigma \in \Omega_n}$ over the same probability space. Let $\big\{V_t\big\}_{t \in \mathbb{N}}$ be an i.i.d. process of random variables uniformly distributed over $\mathcal{V}_n$. Also let $\big\{U_t\big\}_{t \in \mathbb{N}}$ 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

$$X_t^\sigma(V_t) = \begin{cases} +1, & \text{if } U_t \leq \pi\big(+1 \big| X_t(\mathcal{V}_N \setminus \{V_t\})\big) \\ -1, & \text{otherwise} \end{cases}, \tag{4}$$

where $X_t(\mathcal{A})$ is the restriction of X_t to $\mathcal{A} \subseteq \mathcal{V}_n$. Let $\mathbb{P}$ be the probability measure associated with this probability space.

Proposition 1 (Coupling with All-Plus Phase [9]) *Fix $\sqrt{n} \in \mathbb{N}$, $\epsilon \in \big(0, \frac{1}{2}\big)$ 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*

$$\sum_{\sigma:\ m(\sigma) > 0} \pi(\sigma) \mathbb{P}\Big(X_t^\sigma \neq Y_t\Big) \leq e^{-\gamma\sqrt{n}}. \tag{5}$$

In words, the proposition states that if the initial configuration is distributed according to the restriction of the Gibbs measure to $\big\{\sigma \in \Omega_n \big| m(\sigma) > 0\big\}$, 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 1, we show that for $X_0 \sim \pi$, $I(X_0; X_t)$ 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 1. For any $\beta > \beta_0$ there exist $c(\beta) > 0$ such that*

$$I(X_0; X_t) \leq \log 2 + \epsilon_n(\beta), \quad \forall t \geq n \cdot e^{C\beta n^{\frac{1}{4}+\epsilon}}, \tag{6}$$

[3]The restatement adapts the original continuous-time claim from [9] to the discrete-time dynamics.

where $\lim_{n\to\infty}\epsilon_n(\beta)=0$ *for all* β *values as above.*

Proof: We recall that in the phase coexistence regime, zero magnetization is highly improbable under π [12]. Namely, for each $\beta>\beta_c$ there exists $c(\beta)>0$ such that

$$\pi\Big(\big\{\sigma\big|m(\sigma)=0\big\}\Big)\leq e^{-c(\beta)\sqrt{n}}. \tag{7}$$

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

$$I(X_0;X_t)\leq h\big(e^{-c(\beta)\sqrt{n}}\big)+ne^{-c(\beta)\sqrt{n}}+I(X_0;X_t|E=0) \tag{8}$$

where the last step uses (7), with h as the binary entropy function. Let $S=\mathsf{sign}(m_0)$, $m_0\triangleq m(X_0)$, and further consider

$$\begin{aligned}
&I(X_0;X_t|E=0)\\
&\quad\leq H(S|E=0)+I(X_0;X_t|S,E=0)\\
&\quad\overset{(a)}{\leq}\log 2+2\mathbb{P}\big(S=1\big|E=0\big)I(X_0;X_t|S=1,E=0)\\
&\quad\overset{(b)}{\leq}\log 2+I(X_0;X_t|m_0>0)
\end{aligned} \tag{9}$$

where (a) uses $\mathbb{P}\big(S=0\big|E=0\big)=0$ and the symmetry with respect to a global spin flip of the Gibbs measure and the SIM with free boundary conditions, while (b) is because $\{S=1\}\cap\{E=0\}=\{m_0>0\}$ and $\mathbb{P}\big(S=1\big|E=0\big)\leq\frac{1}{2}$.

Next, define $G=\mathbb{1}_{\{X_t=Y_t\}}$ and use the shorthand $\mathbb{P}_+(\mathcal{A})\triangleq\mathbb{P}(\mathcal{A}|m_0>0)$, for $\mathcal{A}\subseteq\Omega_n$. We have

$$\begin{aligned}
&I\big(X_0;X_t\big|m(X_0)>0\big)\\
&\leq H(G|m_0>0)+\mathbb{P}_+(G=0)I(X_0;X_t|G=0,m_0>0)\\
&\qquad\qquad+\mathbb{P}_+(G=1)I(X_0;X_t|G=1,m_0>0)\\
&\leq h\big(\mathbb{P}_+(X_t=Y_t)\big)+\mathbb{P}_+(X_t\neq Y_t)I\big(X_0;X_t\big|X_t\neq Y_t,m_0>0\big)\\
&\qquad+\mathbb{P}_+(X_t=Y_t)I(X_0;X_t|X_t=Y_t,m_0>0).
\end{aligned} \tag{10}$$

We control $\mathbb{P}_+(X_t\neq Y_t)$ using (5) and (7) as follows:

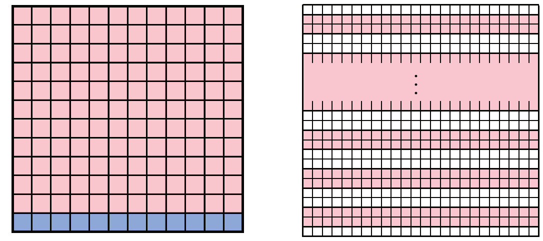
$$\begin{aligned}
\mathbb{P}_+(X_t\neq Y_t)&=\sum_{\sigma\in\Omega_n}\mathbb{P}_+(X_0=\sigma,X_t\neq Y_t)\\
&\overset{(a)}{=}\big(\mathbb{P}(m_0>0)\big)^{-1}\sum_{\sigma:\ m(\sigma)>0}\pi(\sigma)\mathbb{P}\big(X_t^\sigma\neq Y_t\big)\\
&\overset{(b)}{\leq}2\big(1-e^{-c(\beta)\sqrt{n}}\big)^{-1}\sum_{\sigma:\ m(\sigma)>0}\pi(\sigma)\mathbb{P}\big(X_t^\sigma\neq Y_t\big)\\
&\overset{(c)}{\leq}\frac{2e^{-\gamma\sqrt{n}}}{1-e^{-c(\beta)\sqrt{n}}}
\end{aligned} \tag{11}$$

where (a) uses the Markov property, (b) is due to $\pi(m>0)=\frac{1}{2}\pi(m\neq 0)$ and (7), and (c) uses (5). Denoting the RHS of (11) by p_n and inserting back into (10) gives

$$\begin{aligned}
I(X_0;X_t|m_0>0)&\overset{(a)}{\leq}h(p_n)+np_n+I\big(X_0;Y_t\big|G,m_0>0\big)\\
&\leq 2h(p_n)+np_n+I(X_0;Y_t|m_0>0)
\end{aligned}$$

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

We conclude the proof by establishing $I(X_0;Y_t|m_0>0)=0$. 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, $I(X_0;X_t|m_0>0)\leq 2h(p_n)+np_n$, which together with (8) and (9) concludes the proof. ■



(a) (b)

Fig. 2: (a) Initial configuration, with blue and red squares marking plus- and minus-labeled sites, respectively; (b) A partitioning of the 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.

IV. Storing $\sqrt{n}$ bits for Exponential in β Time

We propose analyze a scheme for storing order of $\sqrt{n}$ bits for $\exp(\beta)$ time. We first argue that a single stripe of plus-labeled sites (in a sea of minuses) at the bottom of the grid retains at least half of its original $\sqrt{n}$ pluses for at least $e^{c\beta}$ time, for a numerical constant c. This result generalizes to stripes of width 2 inside the grid (i.e., not at the borders). Together, the two claims enable encoding information into sufficiently separated monochromatic stripes and extracting the written data after $e^{c\beta}$ time via majority decoding.

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 $v\in\mathcal{V}_n$. Calculating the update distribution from 3, if the current configuration is σ and vertex v is selected, then the chance the spin at v is updated to -1 is $p(\sigma,v)=\phi\big(S(\sigma,v)\big)$, where $\phi(a)\triangleq\frac{e^{a\beta}}{e^{a\beta}+e^{-a\beta}}$ and $S(\sigma,v)\triangleq\sum_{w:w\sim v}\sigma(w)=\big|\big\{w\sim v\big|\sigma(w)=-1\big\}\big|-\big|\big\{w\sim v\big|\sigma(w)=+1\big\}\big|$.

Abusing notation, we reuse $(X_t)_{t\geq 0}$ for the continuous-time MC. Let $X_0=\sigma$, 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. 2(a). Define $\mathcal{B}\triangleq\big\{(i,1)\big\}_{i\in[\sqrt{n}]}$, $N_1^{(+)}(\sigma)\triangleq\big|\big\{v\in\mathcal{B}\big|\sigma(v)=+1\big\}\big|$ and $N_1^{(+)}(t)\triangleq N_1^{(+)}(X_t)$, 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 next present a bound on $\mathbb{E}N_1^{(+)}(e^{c\beta})$.

Theorem 1 (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*

$$\mathbb{E}N_1^{(+)}(e^{c\beta})\geq c'\sqrt{n}. \tag{12}$$

A full proof of Theorem 1 is too cumbersome for this short paper, and is available in [4, Appendix G]. Here we only outline the main challenges and our proof strategy.

Challenges & Strategy: The proof comprises two parts: the *phase separation part* and the *analysis part*. To explain each, we first stress the challenges in analyzing the evolution of $N_1^{(+)}(t)$. 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). To circumvent this first complication we simply restrict minus-labeled vertices from flipping. Doing so can only speed up the shrinking of $N_1^{(+)}(t)$, and thus, establishing (12) under this restricted dynamics suffices.

Even with this simplification, the main difficulty in analyzing $N_1^{(+)}(t)$ 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 minus-labeled neighbor. 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 as the dynamics evolves. However, noting that sprinkles have exponentially small probability (in β), while erosion updates have probability exponentially close to 1, one would expect that during some 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 contigs are eaten up.

B. Width-2 Stripe in the Bulk

Our coding scheme also needs a result similar to Theorem 1, but for one monochromatic stripe of width 2 in the bulk of the grid. Specifically, consider an initial state $X_0 = \sigma$, 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*

$$\mathbb{E}N_{j_0}^{(+)}(e^{c\beta}) \geq c'\sqrt{n}. \tag{13}$$

This result follows by the same analysis as in the proof of Theorem [4, Appendix G]), but with an additional preliminary step, as described next. 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. 2(a)). Like before, we first forbid minus-labeled sites to flip. To avoid dealing 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 manually flipped sites 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.
- **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.

With these modification one may repeat the arguments from the proof of Theorem 1 to produce the result of Corollary 2.

C. Stripe-Based Achievability Scheme

To analyze our achievability scheme, we first convert the expected value results from Theorem 1 and Corollary 2 into claims on the probability of a successful majority decoding.

Corollary 3 (Exponential Survival Time Probability) *For sufficiently large β and n (taken from Theorem 1 and Corollary 2), we have*

$$\mathbb{P}\left(N_{j_0}^{(+)}(t_0) > \frac{\sqrt{n}}{2}\right) \geq \frac{2}{3}, \quad \forall j_0 \in [1 : \sqrt{n} - 2]. \tag{14}$$

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

The proof follows by Chebyshev's inequality and is omitted.

We now show that $\Omega(\sqrt{n})$ bits can be stored in the SIM at low 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 2 (Storing $\sqrt{n}$ Bits) *There is $\beta^\star > 0$ such that for any $\beta \geq \beta^\star$ there is $n_0 \in \mathbb{N}$ such that for all $n > n_0$, we have*

$$I_n^{(\beta)}(t) = \Omega(\sqrt{n}), \quad \forall t \leq e^{c\beta}, \tag{15}$$

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.[4]

The partitioning is illustrated in Fig. 2(b). 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. 2(b)). Further set $\mathcal{D}_j \triangleq \mathcal{S}_j \cup \big\{u \in \mathcal{V}_n \big| d(u, \mathcal{S}_j) = 1\big\}$.

Let $\mathcal{C}_n$ be the collection of all configurations whose topology corresponds to Fig. 2(b) 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} assigns an independent and symmetric probability for each stripe to have $+1$ or -1 as its collective spin. 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 $\left(\sigma^{(j)}\right)_{j\in\mathcal{J}}$. 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 $\mathcal{S}_j$. 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.,

$$\psi_j(\sigma) = \begin{cases} +1. & N_j^{(+)}(\sigma) \geq \frac{\sqrt{n}}{2} \\ -1, & N_j^{(+)}(\sigma) < \frac{\sqrt{n}}{2} \end{cases}. \tag{16}$$

Note that X_0 and $\psi_j(X_t)$ are related through a binary channel that inputs a monochromatic stripe $X_0^{(j)}$ (equiprobably all-minus or all-plus), 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 $p_-^{(j)}(\sigma',t) \triangleq \mathbb{P}_{\sigma'}\left(N_j^{(+)}(t) \geq \frac{\sqrt{n}}{2}\right)$. 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

$$\big(\mathsf{T}_j\sigma^{(j)}\big)(v) = \begin{cases} \sigma^{(j)}(v), & v \in \mathcal{S}_j, \\ -\sigma^{(j)}(u), & v \notin \mathcal{S}_j \end{cases}. \tag{17}$$

where u is the bottom left vertex in $\mathcal{S}_j$.[5] By monotonicity of the SIC (see [4, 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 $\mathsf{Ber}\left(\frac{1}{2}\right)$ achieving capacity.

Combining these pieces we have

$$\begin{aligned} I_n^{(\beta)}(t) &\geq I(X_0; X_{t_0}) \\ &\overset{(a)}{=} \sum_{j\in[K]} I\Big(\bar{X}_0^{(j)}; X_{t_0}\Big|\bar{X}_0^{[j-1]}\Big) \\ &\overset{(b)}{\geq} \sum_{j\in[K]} I\Big(\bar{X}_0^{(j)}; \psi_j(X_{t_0})\Big|\bar{X}_0^{[j-1]}\Big) \\ &\overset{(c)}{\geq} K \cdot \mathsf{C}_{\mathsf{BSC}}\left(\frac{1}{3}\right) \end{aligned} \tag{18}$$

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. We conclude by recalling that $K = \Theta(\sqrt{n})$. ■

References

[1] R. Wood, M. Williams, A. Kavcic, and J. Miles. The feasibility of magnetic recording at 10 terabits per square inch on conventional media. *IEEE Trans. Magn.*, 45(2):917–923, February 2009.

[2] M. A. Seigler, W. A. Challener, E. Gage, N. Gokemeijer, G. Ju, B. Lu, K. Pelhos, C. Peng, R. E. Rottmayer, X. Yang, H. Zhou, and T. Rausch. Integrated heat assisted magnetic recording head: Design and recording demonstration. *IEEE Trans. Magn.*, 44(1):119–124, January 2008.

[3] B. D. Terris and T. Thomson. Nanofabricated and self-assembled magnetic structures as data storage media. *Journal of Physics D: Applied Physics*, 38(12):R199–R222, June 2005.

[4] Z. Goldfeld, G. Bresler, and Y. Polyanskiy. Information storage in the stochastic Ising model. *Submitted to IEEE Transactions on Information Theory*, May 2018. Arxiv link: https://arxiv.org/abs/1805.03027.

[5] Z. Goldfeld, H. H. Permuter, and B. M. Zaidel. Capacity region of the finite state MAC with cooperative encoders and delayed CSI. In *Proceedings of the IEEE International Symposium on Information Theory (ISIT-2012)*, Cambridge, Massachusetts, US, July 2012.

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

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

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

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

[10] 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.

[11] Z. Goldfeld, Y. Polyanskiy, and G. Bresler. Information storage in the stochastic Ising model at zero temperature. *Submitted to IEEE Trans. Inf. Theory*, 2018. Available at https://arxiv.org/abs/1805.03027.

[12] 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.

[4] Technically, we allow a wider separation between the two top stripes (to get a minimal distance of at least 2); for simplicity of notation, we henceforth assume all these widths are exactly 2

[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.