Physica A 705 (2027) 132218

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Controlled stochastic growth of structures in the zero-temperature Ising model

Maike C. de Jongh [a,*], Richard J. Boucherie [a], M.N.M. van Lieshout [b,a]

[a] Department of Applied Mathematics, University of Twente, P.O. Box 217, Enschede, NL-7500 AE, The Netherlands
[b] Stochastics Department, Centrum Wiskunde & Informatica, P.O. Box 94079, Amsterdam, NL-1090, The Netherlands

ARTICLE INFO

Keywords:
Bellman optimality equations
Ising model
Metastability
Sequential decision making
Markov decision process

ABSTRACT

Controlled stochastic growth of structures in spatial domains is a fundamental challenge across many scientific disciplines, arising for example in biological pattern formation and solution-based crystallization. In this work, we consider controlled growth in a zero-temperature, two-dimensional Ising model on a finite square lattice evolving under Metropolis dynamics. We investigate how the system can be driven efficiently toward the all-plus configuration through external perturbations applied at predetermined times, starting from a configuration containing a small positive droplet. To analyze this control problem, we formulate it as a Markov decision process, a well-established framework for sequential decision-making under uncertainty. To find an approximately optimal strategy, we propose a reduction of the configuration space to the local minima of the Hamiltonian. Exploiting structural properties of this reduced model, we characterize the optimal policy by recursively solving the associated Bellman optimality equations. The resulting growth mechanism differs fundamentally from the typical evolution observed in the uncontrolled metastable regime. A simulation study demonstrates the effectiveness of this policy and shows that its performance persists at low but positive temperatures.

1. Introduction

The *Ising model*, originally proposed as a model of ferromagnetism [1,2], has since become a paradigmatic model for the study of systems governed by spatially stochastic dynamics with local interactions. Beyond its original physical context, it has found a wide range of applications, including social dynamics [3,4], biological systems [5,6], and image processing [7,8].

In many such settings, it is of interest to understand how the model can be influenced through external interventions. Questions of this type arise, for example, in problems concerning the controlled growth of spatial structures, where local perturbations are used to accelerate or reshape the evolution toward a desired macroscopic configuration, as in crystallization processes [9] or microbial biofilms grown on microfluidic grids [10]. Several works have examined the behavior of the Ising model under external control. For example, control mechanisms based on global or spatially varying parameters, such as temperature, pressure, or external magnetic fields, have been studied with the aim of driving the system between configurations or maintaining it in the vicinity of its critical point [11–13]. Other studies have investigated the effect of various enforcement mechanisms on collective behavior in interacting systems on networks modeled by Ising-type dynamics [14–16].

* Corresponding author.
E-mail address: m.c.dejongh@utwente.nl (M.C. de Jongh).

https://doi.org/10.1016/j.physa.2026.132218
Received 4 March 2026; Received in revised form 3 October 2026
Available online 3 November 2026
0378-4371/© 2026 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

In this work, we focus on a different class of interventions to control the Ising model, namely local perturbations acting on individual spins. Specifically, we consider an external control mechanism that selects lattice sites at predetermined times and flips the corresponding spins. Physically, such an intervention may be realized by applying a strong local magnetic field.

Our objective is to drive the system from a configuration containing a small droplet of +-spins toward the all-plus configuration as efficiently as possible in the zero-temperature regime. In the absence of external control, the behavior of the two-dimensional Ising model at low but positive temperature is well understood [17–24]. In particular, when the system is initialized in the all-minus configuration and subjected to a small positive magnetic field, it typically takes an exponentially long time to reach the stable all-plus configuration. This phenomenon is known as *metastability* [25]. The transition to the stable state generally occurs through the formation of a critical droplet of +-spins, which eventually nucleates the lattice [24]. In this work, we investigate how local intervention mechanisms can optimally accelerate this nucleation process at zero temperature by selecting appropriate sites at which spins are flipped. To assess the robustness of the resulting control strategy, we additionally perform simulations at small positive temperatures. Our results show that the evolution of the system under the optimal control strategy differs structurally from the typical transition observed under uncontrolled dynamics in the metastable regime. Whereas in the latter setting droplets typically grow orthogonally to the sides of a rectangular droplet, the optimal policy favors diagonal growth.

To formalize the control problem, we cast it as a *Markov decision process* (MDP), a well-established framework for modeling and solving problems that involve sequential decision-making under uncertainty [26]. Our formulation belongs to the class of MDPs with a *reachability objective*, where the aim is to reach a specified target state. Since the configuration space of the Ising model grows exponentially with the lattice size, solving this MDP exactly is intractable. A variety of approximate dynamic programming techniques have been developed to address this challenge; see, for example, [27]. One common approach is to reduce the size of the state space through aggregation, or a mapping to a smaller set of representative states. In this work, we propose a reduction of the configuration space to the local minima of the Hamiltonian and show that the resulting simplified model provides an accurate approximation when the time between successive control actions is sufficiently large. Exploiting convenient properties of this reduced model, we recursively solve the associated so-called Bellman optimality equations to deduce the structure of the optimal policy for accelerating nucleation [26].

The main contributions of our work are as follows. First, we introduce a local control mechanism for the zero-temperature Ising model aimed at efficiently driving the system toward the all-plus configuration, starting from an initial small seeded droplet. Second, we formulate this problem as an MDP and relate the first hitting time to the target configuration to the value function of the MDP. Third, we construct an auxiliary MDP by reducing the configuration space to the local minima of the Hamiltonian and deduce the exact structure of the optimal policy within this simplified model. Finally, we present a simulation study that evaluates the performance of this policy at low temperatures and compares it with two heuristic alternatives, inspired by the uncontrolled setting. The results reveal a clear structural difference between the optimal growth mechanism in the controlled setting and the typical evolution of a droplet observed under the uncontrolled metastable dynamics. Moreover, the performance of the optimal strategy is preserved at low but positive temperature.

The remainder of this paper is organized as follows. Section 2 provides the necessary preliminaries on the Ising model and formulates the control problem as an MDP. In Section 3, we study the control problem at zero temperature, construct the auxiliary MDP, and state the exact structure of its optimal policy. Some of the transition probabilities for the auxiliary MDP are deferred to Appendix B. The mathematical justification of the auxiliary MDP and the proof of the structure of the optimal policy are available in the Supplementary Material. Section 4 presents a simulation study, and Section 5 concludes with a discussion of our results and directions for future research.

2. Model definition

2.1. The Ising model

We consider an Ising model defined on a finite, square, two-dimensional lattice $\Lambda = \{0, \dots, N-1\}^2$, equipped with periodic boundary conditions and evolving according to the discrete-time Metropolis dynamics. Each site $i \in \Lambda$ is associated with a spin $\sigma(i) \in \{-1, +1\}$. Let $S = \{-1, +1\}^\Lambda$ denote the set of all configurations of spins on Λ. We denote by $\sigma^+ \in S$ the all-plus configuration, in which all spins are in state +1. The Hamiltonian of a configuration $\sigma \in S$ (see, e.g., [28]) is given by

$$H(\sigma) = -\sum_{\langle i,j \rangle} \sigma(i)\sigma(j) - h \sum_{i \in \Lambda} \sigma(i), \tag{1}$$

where $\langle i, j \rangle$ denotes an unordered nearest neighbor pair in Λ and $h \in (0, 1)$ represents an external magnetic field. For $h \in (0, 1)$, the configuration σ^+ is the unique global minimizer of the Hamiltonian. Given a configuration $\sigma \in S$ and a site $i \in \Lambda$, let σ^i denote the configuration that is obtained after flipping the spin at site i from σ, and given a set $W \subseteq \Lambda$, let σ^W denote the configuration obtained after flipping the spins at all sites in W.

The Metropolis dynamics is defined as a discrete-time Markov chain $(Y_t)_{t \geq 0}$ on S that, given inverse temperature $\beta > 0$, evolves according to the following transition probabilities (see, e.g., [29]) :

$$p_\beta(\sigma, \sigma^i) = \begin{cases} 1/N^2, & \text{if } H(\sigma^i) \leq H(\sigma), \\ (1/N^2)\exp(-\beta(H(\sigma^i) - H(\sigma))), & \text{otherwise,} \end{cases} \tag{2}$$

and $p_\beta(\sigma,\sigma) = 1 - \sum_{i\in\Lambda} p_\beta(\sigma,\sigma^i)$, for $\sigma \in S$. The Metropolis dynamics is reversible with respect to the Gibbs measure associated with H:

$$\mu_\beta(\sigma) = Z_\beta^{-1} \exp(-\beta H(\sigma)), \quad \sigma \in S, \tag{3}$$

where Z_β is the partition function, i.e., $Z_\beta = \sum_{\sigma\in S} \exp(-\beta H(\sigma))$. We are particularly interested in the zero-temperature Metropolis dynamics:

$$p_\infty(\sigma,\sigma^i) = \begin{cases} 1/N^2, & \text{if } H(\sigma^i) \le H(\sigma), \\ 0, & \text{otherwise,} \end{cases} \tag{4}$$

and $p_\infty(\sigma,\sigma) = 1 - \sum_{i\in\Lambda} p_\infty(\sigma,\sigma^i)$, for $\sigma \in S$.

Given configurations $\sigma,\sigma' \in S$, a positive integer k, and inverse temperature $\beta > 0$, we write $p_\beta^{(k)}(\sigma,\sigma')$ for the probability of moving from configuration σ to σ' in k steps under Metropolis dynamics.

2.2. Controlling the Ising model

We consider a setting in which the dynamics of the Ising model can be controlled by means of forced local spin flips. Every $k \in \mathbb{N}^+$ time steps, an external controller may either flip the spin at a chosen site or leave the configuration unchanged. After this control action is applied, the process evolves according to the Metropolis dynamics for k time steps, before the controller can intervene again. We refer to k as the *adjustment time*. The objective is to reach the global minimizer σ^+ as quickly as possible.

A natural framework for analyzing this sequential decision problem is that of a Markov decision process [26]. The state space of this MDP is the set S of spin configurations. At each decision epoch $n = 0, 1, \ldots$, the controller can select an action from the action space

$$A = \Lambda \cup \{0\},$$

where an action $a \in \Lambda$ represents flipping the spin at site a and the action $a = 0$ corresponds to leaving the configuration unchanged. Given a configuration $\sigma \in S$ and an action $a \in A$, let σ^a denote the configuration obtained immediately after applying action a to σ. In particular, $\sigma^0 = \sigma$. We refer to σ^a as the *post-decision configuration*. For inverse temperature $\beta > 0$ and adjustment time $k \in \mathbb{N}^+$, the dynamics of the MDP are described by the transition probability kernel $P_{\beta,k} : S \times A \times S \to [0,1]$ given by

$$P_{\beta,k}(\sigma'|\sigma,a) = p_\beta^{(k)}(\sigma^a,\sigma'), \qquad \sigma,\sigma' \in S,$$

where $p_\beta^{(k)}(\sigma^a,\sigma')$ denotes the probability of a transition from σ^a to σ' in k steps under the Metropolis dynamics defined in (2). In other words, a transition of the MDP consists of two stages: the controller first applies the selected action, after which the system evolves for k time steps according to the Metropolis dynamics. At each decision epoch, the controller receives an immediate reward

$$r(\sigma) = \mathbb{1}\{\sigma = \sigma^+\}, \tag{5}$$

where $\mathbb{1}\{E\}$ denotes the indicator of event E. Since the objective of the controller is to drive the system toward the all-plus configuration σ^+, this reward equals one when the system is in the target configuration and zero otherwise.

The behavior of the controller is governed by a *policy*. We restrict our attention to policies that are *stationary* and *deterministic*. Such a policy is specified by a decision rule $d : S \to A$, which prescribes an action $d(\sigma)$ for each configuration $\sigma \in S$. The same decision rule is applied at every decision epoch, and we denote the resulting policy by $\pi = (d,d,d,\ldots) = d^\infty$. Let Π denote the set of stationary deterministic policies. Each policy $\pi = d^\infty \in \Pi$ induces a Markov chain $(X_n^\pi)_{n\ge0}$ with transition probabilities $P_{\beta,k}(\sigma'|\sigma,d(\sigma))$, $\sigma,\sigma' \in S$. Here, n indexes the decision epochs of the MDP, rather than the individual time steps of the underlying Metropolis dynamics.

A simple example is the policy specified by $d(\sigma) = 0$ for all $\sigma \in S$. Since this policy never forces a spin flip, the induced Markov chain coincides with the uncontrolled Ising model observed every k time steps. As another example, consider a policy that flips the spin at a fixed site $i \in \Lambda$ to $+1$ whenever it is negative. This policy is given by $\pi = d^\infty$, where $d(\sigma) = i$ if $\sigma(i) = -1$ and $d(\sigma) = 0$ otherwise.

The objective of the controller is to reach the all-plus configuration as quickly as possible. We therefore consider a discounted reward criterion that favors small first hitting times to the target configuration σ^+. Starting from a configuration $\sigma \in S$, the expected total discounted reward, or the *value*, of a policy $\pi = d^\infty$ is defined as

$$v_{\lambda,\beta,k}^\pi(\sigma) = \mathbb{E}_\sigma^\pi\Big[\sum_{n=0}^{\infty} \lambda^n r(X_n^\pi)\Big], \tag{6}$$

where $(X_n^\pi)_{n\ge0}$ denotes the Markov chain induced by policy π and $\lambda \in (0,1)$ is a discount factor. The factor λ^n determines the weight assigned to a reward received at decision epoch n. Hence, rewards obtained at later decision epochs receive less weight and this decrease is more rapid for smaller values of λ. The value is obtained by summing these discounted rewards over the entire time horizon. For example, if $\lambda = 1/2$ and the process first reaches σ^+ at decision epoch n^*, the reward obtained at that decision epoch contributes $(1/2)^{n^*}$ to the value function. Thus, if σ^+ is reached one decision epoch earlier, the contribution of this reward is doubled. This illustrates how discounting favors policies that reach the target configuration quickly.

By conditioning (6) on the configuration reached at the next decision epoch, we obtain the following set of equations for the value function $v^{\pi}_{\lambda,\beta,k}$:

$$v^{\pi}_{\lambda,\beta,k}(\sigma) = r(\sigma) + \lambda \mathbb{E}^{\pi}_{\sigma}[v^{\pi}_{\lambda,\beta,k}(X^{\pi}_1)], \tag{7}$$

$$= r(\sigma) + \lambda \sum_{\sigma' \in S} P_{\beta,k}(\sigma'|\sigma, d(\sigma)) v^{\pi}_{\lambda,\beta,k}(\sigma'), \quad \sigma \in S. \tag{8}$$

Eq. (7) expresses the value of a policy π as the sum of the immediate reward at decision epoch 0 and the expected discounted value from the next decision epoch onward. Starting from configuration σ, the controller applies the action prescribed by policy π, after which the system evolves according to the Metropolis dynamics for k time steps and arrives in a random configuration X^{π}_1. The second term in (7) is the expected value of the policy starting from this configuration. Eq. (8) is obtained from (7) by explicitly evaluating the expectation in terms of the transition probabilities to all possible configurations $\sigma' \in S$.

We now establish the precise relation between the value function and the first hitting time to the target configuration. The latter is a key quantity in studies of the metastable behavior of the Ising model, see, e.g., [17–24]. Let $\tau^{\sigma,\pi}_B$ denote the *first hitting time* of a set $B \subseteq S$ by the Markov chain $(X^{\pi}_n)_{n\geq 0}$ induced by policy $\pi \in \Pi$, with initial configuration $X_0 = \sigma \in S$. That is,

$$\tau^{\sigma,\pi}_B = \inf\{n \in \mathbb{N}^+ : X^{\pi}_n \in B\},$$

where $X^{\pi}_0 = \sigma$ and the process evolves under policy π. Thus, $\tau^{\sigma,\pi}_B$ measures the hitting time to B in decision epochs. Consequently, $\lambda^{\tau^{\sigma,\pi}_B}$ represents the discounting associated with the time until B is first reached. For example, if B is reached after three decision epochs, then $\tau^{\sigma,\pi}_B = 3$ and $\lambda^{\tau^{\sigma,\pi}_B} = \lambda^3$. Simplifying notation, we write $\tau^{\sigma,\pi}_{\{\sigma'\}}$ as $\tau^{\sigma,\pi}_{\sigma'}$ for $\sigma, \sigma' \in S$. Theorem 1 expresses the value function in terms of these first hitting times. The proof of this statement is provided in Appendix A. For notational convenience, we write the expectation $\mathbb{E}^{\pi}_{\sigma}\left[\lambda^{\tau^{\sigma,\pi}_{\sigma'}}\right]$ simply as $\mathbb{E}\left[\lambda^{\tau^{\sigma,\pi}_{\sigma'}}\right]$, since the initial configuration and policy are already specified by the notation for the first hitting time. Hence, $\mathbb{E}\left[\lambda^{\tau^{\sigma,\pi}_{\sigma'}}\right]$ quantifies how strongly a reward obtained upon the first visit to σ' would be discounted on average, starting from σ and following policy π.

Theorem 1. *For each stationary, deterministic policy $\pi = d^{\infty}$, we have*

$$v^{\pi}_{\lambda,\beta,k}(\sigma) = r(\sigma) + \sum_{\sigma' \in S} \frac{r(\sigma')\mathbb{E}\left[\lambda^{\tau^{\sigma,\pi}_{\sigma'}}\right]}{1 - \mathbb{E}\left[\lambda^{\tau^{\sigma',\pi}_{\sigma'}}\right]}. \tag{9}$$

Inserting the reward function (5) yields Corollary 2.

Corollary 2. *Given the reward function (5), the value function of a stationary deterministic policy $\pi = d^{\infty}$ is given by*

$$v^{\pi}_{\lambda,\beta,k}(\sigma) = \mathbb{1}\{\sigma = \sigma^+\} + \frac{\mathbb{E}\left[\lambda^{\tau^{\sigma,\pi}_{\sigma^+}}\right]}{1 - \mathbb{E}\left[\lambda^{\tau^{\sigma^+,\pi}_{\sigma^+}}\right]}. \tag{10}$$

Expression (10) clarifies how the value of a policy depends on the first hitting time to the all-plus configuration. More precisely, it indicates that a policy has a high value if the controlled system tends to reach the all-plus configuration quickly and tends to return to it soon after moving away. When $\sigma = \sigma^+$, the first term 1 is the immediate reward obtained at decision epoch 0. Since λ^n decreases with n, trajectories that reach σ^+ at an earlier decision epoch make a larger contribution to the value function. The denominator accounts for the discounted rewards obtained from repeated visits to σ^+ after it has first been reached. By the strong Markov property [30, Section 1.4], the successive return cycles to σ^+ are independent and identically distributed. Hence, the discounted contributions of successive returns to σ^+ form a geometric series with common ratio $\mathbb{E}\left[\lambda^{\tau^{\sigma^+,\pi}_{\sigma^+}}\right]$. The relative contributions of shorter and longer trajectories depend on the value of the discount factor. As we will see in Section 3.2, this dependence gives rise to structurally different optimal policies for different values of λ.

A policy π^* is *optimal* if its value function is at least as large as that of any other policy for every initial configuration. Since in our control problem the state and action spaces are finite and the reward function is bounded, there exists an optimal policy that is stationary and deterministic [26, p. 154, Thm. 6.2.10]. Hence, it suffices to consider policies in Π. In particular, $\pi^* \in \Pi$ is optimal if

$$v^{\pi^*}_{\lambda,\beta,k}(\sigma) \geq v^{\pi}_{\lambda,\beta,k}(\sigma), \quad \sigma \in S, \ \pi \in \Pi.$$

Each optimal policy has the same value function, which we denote by $v^*_{\lambda,\beta,k}$.

Eq. (8) determines the value of a fixed policy. The aim of the controller is to choose a policy that maximizes this value. We proceed to describe this optimization step. Suppose that the system is in a configuration σ. For each possible action a, the controller evaluates the possible configurations σ' that may be reached at the next decision epoch after applying action a and letting the system evolve for k Metropolis steps. The quantity $v^*_{\lambda,\beta,k}(\sigma')$ represents the optimal future value starting from each such configuration σ' at epoch 1. The expected future value associated with action a is therefore obtained by averaging these values with respect to the transition probabilities $P_{\beta,k}(\sigma'|\sigma, a)$. This expected future value is discounted by a factor λ, leading to the contribution of the future

reward $\lambda \sum_{\sigma' \in S} P_{\beta,k}(\sigma'|\sigma,a) v^*_{\lambda,\beta,k}(\sigma')$. Together with the immediate reward $r(\sigma)$, this discounted expected future value gives the value associated with choosing action a now and acting optimally thereafter, i.e.,

$$r(\sigma) + \lambda \sum_{\sigma' \in S} P_{\beta,k}(\sigma'|\sigma,a) v^*_{\lambda,\beta,k}(\sigma').$$

The optimal value is obtained by maximizing this quantity over all available actions. This leads to the *Bellman optimality equations*:

$$v^*_{\lambda,\beta,k}(\sigma) = \max_{a \in A} \Big\{ r(\sigma) + \lambda \sum_{\sigma' \in S} P_{\beta,k}(\sigma'|\sigma,a) v^*_{\lambda,\beta,k}(\sigma') \Big\}, \quad \sigma \in S. \tag{11}$$

For a full derivation of the Bellman optimality equations see [26, Sections 4.3 and 6.2.1]. A policy $\pi^* = (d^*)^\infty \in \Pi$ is optimal if and only if $v^{\pi^*}_{\lambda,\beta,k}$ is a solution to the Bellman optimality equations [26, p. 152]. In particular, an optimal decision rule $d^*(\sigma)$ selects, for each configuration $\sigma \in S$, an action that attains the maximum in (11).

3. Analysis of the two-dimensional zero-temperature Ising model

In this section, we analyze the control problem defined in Section 2 at zero temperature. A major challenge in solving the Bellman optimality equations is the size of the state space. For an $N \times N$ Ising model, the state space contains 2^{N^2} configurations, making direct solution of the Bellman optimality equations computationally intractable except for small values of N. At zero temperature, however, the dynamics simplify considerably, as the definition of the Metropolis dynamics (2) implies that $p_\infty(\sigma,\sigma') = 0$ for all $\sigma' \in S$ satisfying $H(\sigma') > H(\sigma)$, $\sigma \in S$. Hence, at zero temperature, the probability that the process reaches a configuration of higher energy during the k Metropolis steps between two successive decision epochs is zero. This greatly restricts the set of configurations that can be reached at the next decision epoch and thus simplifies the Bellman optimality equations. We therefore analyze the Ising MDP in the zero-temperature regime, that is, for $\beta = \infty$. In addition, we impose a small positive external magnetic field $h \in (0,1)$, which favors the +1 phase and makes σ^+ the unique global minimizer of the Hamiltonian.

The key step in our analysis is the construction of an auxiliary MDP with a reduced state space. To this end, we characterize the configurations in which the Ising model gets trapped under the zero-temperature dynamics. These are the local minima of the Hamiltonian, for which every spin flip increases the energy. For sufficiently large k, the system reaches such a local minimum before the next decision epoch with high probability. Therefore, the transition probabilities of the MDP become increasingly concentrated on these configurations. This motivates the construction of an auxiliary MDP whose state space consists only of local minima of the Hamiltonian. For sufficiently large adjustment times, the resulting reduced model provides an accurate approximation of the original MDP.

The construction of the auxiliary MDP is outlined in Section 3.1. In Section 3.2, we state our main result, Theorem 4, concerning the structure of the optimal policy for the auxiliary MDP. Further details on the construction of the auxiliary MDP and the derivation of the optimal policy are provided in Supplementary Materials I, II, and III.

3.1. The auxiliary MDP

In order to define the auxiliary MDP, we provide some additional notation and terminology specific to the zero-temperature regime. First, we call a site $i \in \Lambda$ *susceptible* in a configuration $\sigma \in S$ if $H(\sigma^i) \leq H(\sigma)$. Equivalently, site i is susceptible if $p_\infty(\sigma,\sigma^i) > 0$. Thus, at zero temperature, only susceptible sites can flip under the Metropolis dynamics. Recall that we impose a small external magnetic field $h \in (0,1)$. This implies that a site with spin +1 is susceptible if and only if it has at least three neighboring sites with spin −1, whereas a site with spin −1 is susceptible if and only if it has at least two neighboring sites with spin +1. Fig. 1 shows an example of a susceptible site with spin +1. The set of susceptible sites in a configuration $\sigma \in S$ is denoted by $\Delta(\sigma)$. We call a configuration $\sigma \in S$ *robust* if no site is susceptible in σ, that is, if $\Delta(\sigma) = \emptyset$. Hence, once the system reaches a robust configuration, no further spin flips are possible under the zero-temperature Metropolis dynamics. Observe that the robust configurations coincide with the local minima of the Hamiltonian. We define $U^{(1)} \subseteq S$ to be the set of robust configurations, excluding the all-minus configuration σ^-, in which the sites with spin +1 form a single connected component, which we refer to as a droplet. Theorem 3 characterizes these configurations geometrically, showing that such a droplet necessarily forms a rectangle.

Theorem 3. *A configuration $\sigma \in S$ belongs to $U^{(1)}$ if and only if $\sigma \neq \sigma^-$ and the sites with spin +1 form a rectangle of size $i \times j$, for some $i,j \in \{2,3,\dots,N-3,N-2,N\}$.*

Proof. A configuration $\sigma \neq \sigma^-$ is robust if and only if each site with spin −1 has at most one neighbor with spin +1 and each site with spin +1 has at least two neighbors with spin +1. A configuration in which the sites with spin +1 form one droplet satisfies these requirements if and only if these sites form a rectangle of size $i \times j$, where $i,j \in \{2,3,\dots,N-3,N-2,N\}$. □

Motivated by this geometrical characterization, we construct the auxiliary MDP by representing all robust configurations with a rectangle of size $i \times j$ by a single state (i,j). Hence, we define the state space $\hat{S}$ of the auxiliary MDP as

$$\hat{S} = \{(i,j) | i,j = 2,3,\dots,N-3,N-2,N\} \cup \{(0,0)\}.$$

A state $(i,j) \in \hat{S}$, $i,j > 0$, serves as a representative of all configurations in which the sites with spin +1 form a single rectangle of size $i \times j$. The state $(0,0)$ represents the all-minus configuration. Due to translation invariance, the position of the rectangle on

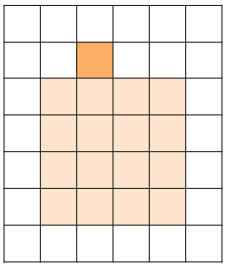


Fig. 1. Example of a susceptible site with spin +1. Orange and white squares represent sites with spin +1 and −1, respectively. The dark orange site, which has spin +1, is susceptible because three of its neighbors have spin −1, so flipping the spin at this site decreases the energy of the configuration.

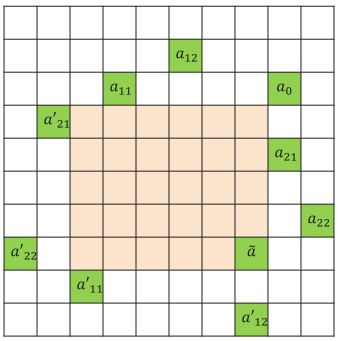


Fig. 2. Visualization of the action space of the auxiliary MDP for a rectangular droplet of +1 spins. Each labeled green site is a representative of the set of sites associated with the corresponding action in the auxiliary MDP. For example, action a_{11} represents sites at distance one from a horizontal side for which the nearest site of the rectangle is not a corner, action a'_{11} represents the corresponding sites for which the nearest site is a corner, and $\tilde{a}$ represents the corners of the rectangle. The complete definition of the action space is provided in the main text.

the lattice does not affect the transition probabilities, and its dimensions contain all information required to describe the transition probabilities of the auxiliary MDP.

Similarly, we construct the action space $\hat{A}$ by grouping sites according to their geometric position relative to a rectangular droplet. The resulting action space, illustrated in Fig. 2, is given by

$$\hat{A} = \{a_{11}, a_{12}, a_{21}, a_{22}, a'_{11}, a'_{12}, a'_{21}, a'_{22}, a_0, \tilde{a}, 0\}.$$

Here $a_{1\ell}$ and $a_{2\ell}$, $\ell = 1, 2$, represent the sets of sites at distance ℓ from the horizontal and vertical side of the rectangle, respectively, for which the nearest site belonging to the rectangle is not a corner. Actions $a'_{1\ell}$, $a'_{2\ell}$, $\ell = 1, 2$, represent the corresponding sets of sites for which the nearest site belonging to the rectangle is a corner. Action a_0 corresponds to the set of sites that are diagonally adjacent to the rectangle, action $\tilde{a}$ represents the set of sites that are corners of the rectangle, and action 0 represents all remaining sites, for which a spin flip is equivalent to taking no action in the auxiliary MDP. By executing an action $a \in \hat{A}$, we mean flipping the spin at one arbitrary site in its associated set. For each action $a \in \hat{A}$, Fig. 2 shows a representative site from the set associated with that action. As will become apparent from the derivation of the dynamics of the auxiliary MDP, different choices of sites within the same set have the same effect for sufficiently large values of k. Therefore, sites associated with the same action are treated as equivalent in the auxiliary MDP.

To illustrate how the auxiliary MDP describes the controlled evolution of a droplet in the original model, consider a droplet represented by the state (i, j). At a decision epoch, the controller selects one of the auxiliary actions shown in Fig. 2, which corresponds to flipping the spin at a site belonging to the associated set. The system then evolves according to the zero-temperature Metropolis dynamics and, for sufficiently large adjustment time, reaches another robust configuration. This new configuration is again represented by the dimensions of its rectangular +1 droplet, after which a new control action is selected. In this way, the auxiliary MDP describes the evolution of the system from one robust configuration to another.

We now give an informal outline of one such transition for action a_{12}, as a representative example. Technical details are deferred to Supplementary Material I. Fig. 3 depicts a configuration obtained immediately after taking action a_{12} from a state (i, j), where $i = 3, \ldots, N-2$, $j = 2, \ldots, N-4$. For sufficiently large k, the evolution of the system now depends entirely on the order in which the sites labeled $1, 2, \ldots, 6$ are selected by the Metropolis dynamics. Observe that in the configuration shown in Fig. 3, there are two susceptible sites, namely those labeled 1 and 2. Spin flips at any other sites are impossible at zero temperature. If site 1 is selected before site 2, the spin at site 1 flips to −1 and the system will be in state (i, j) at the start of the next decision epoch. On the other hand, if site 2 is selected first, it flips to +1, leading to a configuration in which sites 1, 3 and 4 are susceptible. Suppose that, from

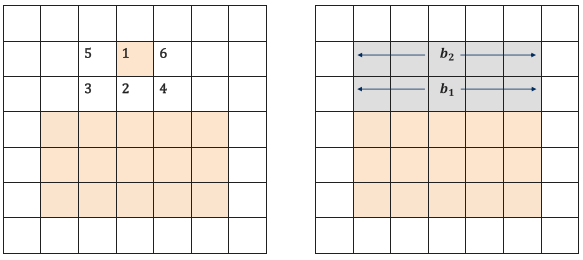


Fig. 3. Post-decision configuration obtained after applying action a_{12} to state (i, j), where the numbered sites indicate the sites whose order of selection determines the resulting transition (left), and illustration of the two strips b_1 and b_2, shown in gray, that may be added to the rectangular droplet during the transition (right).

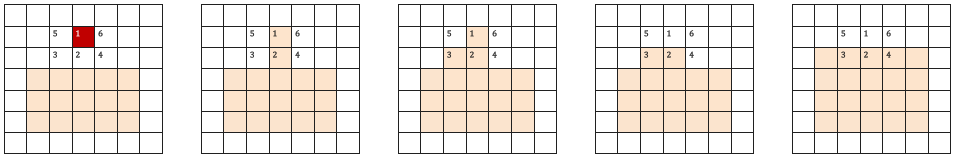


Fig. 4. Schematic illustration of one possible transition after taking action a_{12} from state (i, j), corresponding to the sequence $(2, 3, 1)$ in Table 1. The site selected by the controller is shown in red. In the first configuration, the controller flips the spin at a site corresponding to action a_{12} to $+1$. Under the subsequent zero-temperature Metropolis dynamics, the spins at sites 2 and 3 flip to $+1$, after which the spin at site 1 flips back to -1, as shown in the next three configurations. Thereafter, the remaining sites in strip b_1 successively flip to $+1$, resulting in the state $(i, j+1)$ shown in the final configuration.

this configuration, site 3 is the first selected among these three options. Note that at zero temperature, it is impossible for sites 2 and 3 to become susceptible again and all sites in the strip labeled b_1 in the right panel of Fig. 3 will flip to $+1$ if k is sufficiently large. Similarly, whether the spins in the strip labeled b_2 also flip to $+1$ depends on the order in which the sites labeled 1, 5 and 6 are selected. Hence, starting from state (i, j), the process can either remain in state (i, j), transition to state $(i, j+1)$ by adding one strip to the droplet, or transition to state $(i, j+2)$ by adding two strips, depending on the order in which sites $1, \dots, 6$ are selected. All possible sequences of susceptible sites starting from the post-decision configuration in the left panel of Fig. 3 are listed in Table 1. This argument yields

$$\hat{P}((i', j')|(i, j), a_{12}) = \begin{cases} 5/9, & \text{if } i' = i, j' = j, \\ 7/27, & \text{if } i' = i, j' = j+1, \\ 5/27, & \text{if } i' = i, j' = j+2, \\ 0, & \text{otherwise.} \end{cases} \tag{12}$$

To further illustrate this transition mechanism, Fig. 4 shows one possible evolution of the system after taking action a_{12} from state (i, j). This evolution corresponds to the sequence $(2, 3, 1)$ in Table 1 and leads to state $(i, j+1)$. The local configuration of the sites labeled $1, \dots, 6$ is the same up to translation for each original control action $a \in A$ represented by the auxiliary action a_{12}. Hence, all sites associated with a_{12} induce the same transition probabilities and are treated as equivalent in the auxiliary MDP. The remaining transition probabilities can be derived in a similar way and are provided in Appendix B.

Finally, we define the reward function of the auxiliary process as

$$\hat{r}(s) = \mathbb{1}\{s = (N, N)\}, \qquad s \in \hat{S},$$

reflecting the reward function (5) of the original MDP, as a configuration with a rectangle of size $N \times N$ is equivalent to the all-plus configuration.

For sufficiently large values of k, the value of a policy in the auxiliary MDP provides an accurate approximation of its value in the original zero-temperature Ising MDP. More precisely, for any stationary deterministic policy, the difference between the corresponding value functions converges to zero as $k \to \infty$. This result is established in Theorem S1 in Supplementary Material I. We refer to Remark 1 for a discussion of the relation between our model and the limiting case $\beta, k \to \infty$.

Remark 1 (*The Limiting Case* $\beta, k \to \infty$)**.** There are some subtleties that arise when taking the limits $\beta, k \to \infty$. For any fixed $k \in \mathbb{N}$, note that

$$\lim_{\beta \to \infty} p_\beta^{(k)}(\sigma, \sigma') = p_\infty^{(k)}(\sigma, \sigma').$$

Table 1
Possible sequences of susceptible sites following action a_{12}, corresponding to the labels in Fig. 3, together with their probabilities and the resulting state after the transition. The transition probabilities in the final column are obtained by summing the probabilities of all sequences leading to the same state and yield expression (12).

Sequences of susceptible sites	Probability of selecting sequence	Next state	Transition probability
(1)	1/2	(i, j)	5/9
(2, 1, 2)	1/18		
(2, 1, 3)	1/18	$(i, j+1)$	7/27
(2, 1, 4)	1/18		
(2, 3, 1)	1/18		
(2, 3, 4, 1)	1/54		
(2, 4, 1)	1/18		
(2, 4, 3, 1)	1/54		
(2, 3, 4, 5)	1/54	$(i, j+2)$	5/27
(2, 3, 4, 6)	1/54		
(2, 4, 3, 5)	1/54		
(2, 4, 3, 6)	1/54		
(2, 3, 5)	1/18		
(2, 4, 6)	1/18		

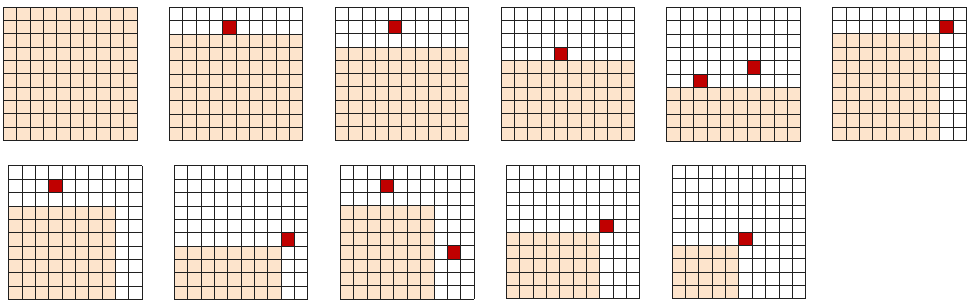


Fig. 5. Visualization of the optimal policy of the auxiliary MDP (Theorem 4). In each state, the red square indicates a representative site associated with the optimal action specified in Theorem 4.
Top row: States $(N, N), (N, N-2), (N, N-3), (N, N-4)$, (N, j) for $j = 2, \dots, N-5$, and $(N-2, N-2)$.
Bottom row: States $(N-2, N-3)$, $(N-2, j)$ for $j = 2, \dots, N-4$, $(N-3, N-3)$, $(N-3, j)$ for $j = 2, \dots, N-4$, and (i, j) for $i, j = 2, \dots, N-4$.

Hence, taking $\beta \to \infty$ first yields the zero-temperature Ising MDP, and subsequently $k \to \infty$ yields the auxiliary MDP. However, this is not the case if the order of the limits is reversed. Indeed, since the state space is finite and the Metropolis dynamics are irreducible, there is a unique stationary Gibbs distribution μ_β, as defined in expression (3), and, since $h > 0$,

$$\lim_{\beta\to\infty} \lim_{k\to\infty} p_\beta^{(k)}(\sigma, \sigma') = \lim_{\beta\to\infty} \mu_\beta(\sigma') = \mathbb{1}\{\sigma' = \sigma^+\}. \quad \square$$

3.2. Optimal control of the zero-temperature Ising model

A convenient feature of the auxiliary MDP is the fact that it is not possible for a rectangle of size at least 3×3 to shrink, as by definition of the transition probability kernel, the probability that a state variable that is greater than 3 decreases in the auxiliary MDP is zero. This fact allows us to solve the Bellman optimality equation recursively, starting from state (N, N), which corresponds to the all-plus configuration. This leads to our main result, which specifies the structure of the optimal policy in the auxiliary MDP and is stated in Theorem 4. A visualization of this result is provided in Fig. 5. The result shows that the optimal policy is to grow the rectangle along its diagonal, as long as the exterior gap between its boundaries is sufficiently large. When this gap becomes too small, the optimal action may instead be to flip a spin at a site between two opposing boundaries. The proof of Theorem 4 is deferred to Supplementary Materials I and III.

Theorem 4. *A stationary, deterministic policy $\pi^* = (d^*)^\infty$ is optimal in the auxiliary MDP if and only if*

$$d^*(i, j) \in \begin{cases} A_1(i, j), & \text{if } \lambda \in (\lambda_c, 1), \\ A_1(i, j) \cup A_2(i, j), & \text{if } \lambda = \lambda_c, \\ A_2(i, j), & \text{if } \lambda \in (0, \lambda_c), \end{cases}$$

for all $(i,j) \in \hat{S}$, *where* $\lambda_c = 15/17$. *Here,* $A_k : \hat{S} \to \mathcal{P}(\hat{A})$, $k = 1,2$, *where* $\mathcal{P}(\hat{A})$ *denotes the power set of the action space* $\hat{A}$, *is defined as*

$$
\begin{aligned}
&A_k(N,N) = \{0\}, && A_k(N-2,N) = \{a_{21}\},\\
&A_k(N,N-2) = \{a_{11}\}, && A_k(N-3,N) = \{a_{22}\},\\
&A_k(N,N-3) = \{a_{12}\}, && A_k(N-4,N) = \{a_{21}\},\\
&A_k(N,N-4) = \{a_{11}\}, && A_k(N-2,N-2) = \{a_0\},
\end{aligned}
$$

$$
\begin{aligned}
&A_k(N-2,N-3) = \{a_{12}\}, && A_k(N-2,j) = A_k(j,N-2) = \{a_0\},\\
&A_k(N-3,N-2) = \{a_{22}\}, && A_k(N-3,j) = A_k(j,N-3) = \{a_0\},\\
&A_k(N-3,N-3) = \{a_{12},a_{22}\}, && A_k(i,j) = \{a_0\},
\end{aligned}
$$

for $i,j = 2,\dots,N-4$, $k = 1,2$, *and*

$$
\begin{aligned}
&A_1(N,j) = \{a_{11}\} \quad \textit{and} \quad A_2(N,j) = \{a_{12}\}, \quad \textit{for } j = 2,\dots,N-5,\\
&A_1(i,N) = \{a_{21}\} \quad \textit{and} \quad A_2(i,N) = \{a_{22}\}, \quad \textit{for } i = 2,\dots,N-5.
\end{aligned}
$$

Remark 2 (*Comparison with the Uncontrolled Low-temperature Setting*)**.** Observe that under the optimal policy, the rectangle predominantly follows a diagonal growth mechanism. This mechanism is structurally different from the growth observed in the uncontrolled low-temperature metastable regime. In the latter case, droplets typically grow through protuberances forming at the sides of the rectangle, corresponding to sites associated with actions a_{11} or a_{21} (e.g., [24]). □

Remark 3 (*Phase Transition in Discount Factor*)**.** An interesting aspect of the result of Theorem 4 is the phase transition with respect to the discount factor for rectangles of size (N,j) or (i,N), $i,j = 2,\dots,N-5$. This phenomenon admits a very intuitive explanation. From the transition probability kernel provided in Appendix B, we observe that after taking action a_{12} or a_{22}, it is possible for the rectangle to grow by two strips at once, whereas choosing action a_{11} or a_{21} results in a higher probability of producing any growth at all, but the gain is limited to a single strip. If the discount factor is high, a reward obtained in the future retains a fair portion of its value in the present. In this regime, it is therefore optimal to choose the safer options a_{11} or a_{21}. By contrast, if the discount factor is low, a reward obtained in the future holds less value in the present, and only trajectories that reach the all-plus configuration as quickly as possible make a significant contribution to the value function. In this case, it is more advantageous to opt for a riskier strategy that has the potential to reach the target sooner. □

4. Controlling the Ising model in the low-temperature regime

In this section, we conduct a simulation study on the performance of the optimal policy for the auxiliary MDP in the original Ising MDP at low temperature. We compare this policy with the following two heuristic growth strategies: policy $\hat{\pi}_1$ always flips the spin at a site immediately adjacent to the current rectangle, alternating between its right and lower sides. Policy $\hat{\pi}_2$ selects a site at distance two from the rectangle, again alternating between these two sides. We let $\hat{\pi}^* = (\hat{d}^*)^\infty$ denote an optimal policy in the auxiliary MDP. The heuristic policies $\hat{\pi}_1 = (\hat{d}_t^1)_{t\in\mathbb{N}}$ and $\hat{\pi}_2 = (\hat{d}_t^2)_{t\in\mathbb{N}}$ in the auxiliary MDP are defined by

$$
\hat{d}_t^1(i,j) = \begin{cases} a_{11}, & \text{if } t \text{ even},\\ a_{21}, & \text{if } t \text{ odd}, \end{cases} \qquad \text{for } i,j = 2,\dots,N-2,
$$

$$
\hat{d}_t^1(i,j) = \begin{cases} a_{11}, & \text{if } i = N, \quad j = 2,\dots,N-2,\\ a_{21}, & \text{if } i = 2,\dots,N-2, \quad j = N,\\ 0, & \text{if } i = j = N, \end{cases}
$$

$$
\hat{d}_t^2(i,j) = \begin{cases} a_{12}, & \text{if } t \text{ even and } j = 2,\dots,N-3,\\ a_{11}, & \text{if } t \text{ even and } j = N-2,\\ a_{22}, & \text{if } t \text{ odd and } i = 2,\dots,N-3,\\ a_{21}, & \text{if } t \text{ odd and } i = N-2, \end{cases} \qquad \text{for } i,j = 2,\dots,N-2,
$$

$$
\hat{d}_t^2(i,j) = \begin{cases} a_{12}, & \text{if } i = N, \quad j = 2,\dots,N-3,\\ a_{11}, & \text{if } i = N, \quad j = N-2,\\ a_{22}, & \text{if } i = 2,\dots,N-3, \quad j = N,\\ a_{21}, & \text{if } i = N-2, \quad j = N,\\ 0, & \text{if } i = j = N. \end{cases}
$$

Flipping the spin at any site further away from the rectangle is ineffective, since it flips back to −1 at zero temperature (and with high probability at low temperature). Hence, $\hat{\pi}_1$ and $\hat{\pi}_2$ are natural benchmark policies. Note that policy $\hat{\pi}_1$ represents the growth mechanism observed in the uncontrolled metastable regime (e.g. [24]). Fig. 6 depicts the values of these policies in state $(3,3)$ for $N = 20$ as a function of the discount factor λ. It can be seen that $\hat{\pi}^*$ indeed achieves higher expected total discounted reward than the policies $\hat{\pi}_1$ and $\hat{\pi}_2$ throughout the range of discount factors considered.

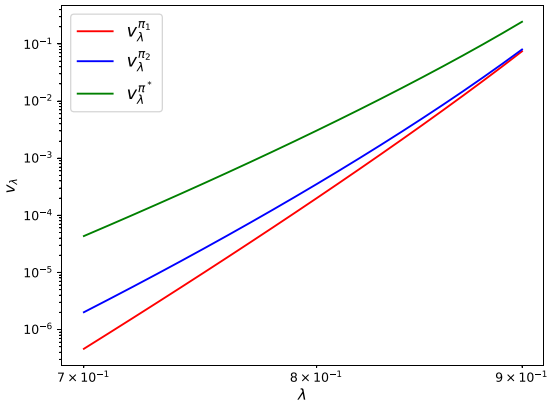


Fig. 6. Values of policies $\hat{\pi}_1$, $\hat{\pi}_2$ and $\hat{\pi}^*$ in state $(3,3)$ in the auxiliary MDP as a function of the discount factor.

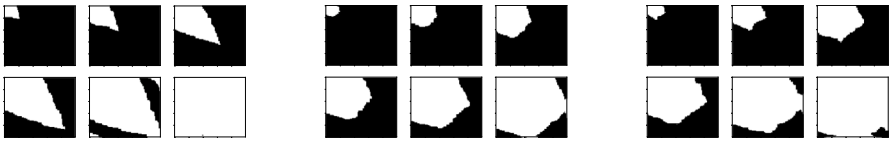


Fig. 7. Illustration of the policies $\tilde{\pi}^*$ (left group), π_1 (middle group), and π_2 (right group) for $\beta = 10$, $N = 100$, and $k = 5000$. Each group shows configurations at times $t = 100, 200, 300, 400, 500$ and 600.



Fig. 8. Illustration of policy $\tilde{\pi}^*$ for various adjustment times: $k = 5000$ (left group), $k = 10\,000$ (middle group), and $k = 50\,000$ (right group), with $\beta = 10$ and $N = 100$. Each group shows configurations at times $t = 50, 100, 150, 200, 250$ and 300.

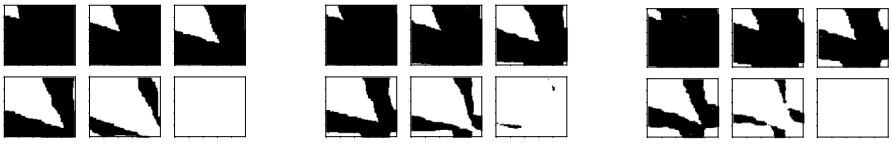


Fig. 9. Illustration of policy $\tilde{\pi}^*$ for various temperatures: $\beta = 2.5$ (left group), $\beta = 1.67$ (middle group), and $\beta = 1.25$ (right group), with $N = 100$ and $k = 5000$. Each group shows configurations at times $t = 100, 200, 300, 400, 500$ and 600.

We now simulate analogues of policies $\tilde{\pi}^*$, $\hat{\pi}_1$ and $\hat{\pi}_2$ in the original Ising MDP at low temperature. In each simulation, the following procedure is applied: at each decision epoch, we first determine the circumscribed rectangle of the current +1 droplet and use the corresponding auxiliary policy to select a control action. The corresponding spin in the original configuration is then forced to flip. Starting from the resulting configuration, we let the system evolve according to the Metropolis dynamics for k time steps. At the next decision epoch, we repeat the same procedure for the resulting configuration. Note that the adjustment time k may not be sufficiently large for the droplet to fill its circumscribed rectangle before the start of the next decision epoch.

More specifically, we define policies $\tilde{\pi}^* = (\tilde{d}^*)^\infty$, $\pi_1 = (d_t^1)_{t\in\mathbb{N}}$ and $\pi_2 = (d_t^2)_{t\in\mathbb{N}}$. For a configuration $\sigma \in S$ in which the circumscribed rectangle has size $i \times j$, the decision rules $\tilde{d}^*(\sigma)$, $d_t^1(\sigma)$, $d_t^2(\sigma)$ select sites corresponding to the auxiliary actions $\hat{d}^*(i,j)$, $\hat{d}_t^1(i,j)$ and $\hat{d}_t^2(i,j)$. Among the sites corresponding to a given action, we choose the one closest to the midpoint of the relevant side of the circumscribed rectangle (or to one of the two central sites if the side length is even). Moreover, whenever a site on a horizontal side of the circumscribed rectangle is selected, we always choose the same horizontal side; the same convention applies to sites on the vertical side.

Fig. 7 shows simulation results for the three different policies for $\beta = 10$, $N = 100$ and an adjustment time of $k = 5000$. The policy $\tilde{\pi}^*$ indeed seems to reach the target configuration faster than the other two policies.

Note that the adjustment time in Fig. 7 is clearly not long enough for a droplet with spin +1 to fully grow into a rectangle before the decision maker again decides to flip a spin. Fig. 8 illustrates the behavior of the policy $\tilde{\pi}^*$ for various different adjustment times. For larger adjustment times, the droplets have smoother sides and more closely resemble rectangles.

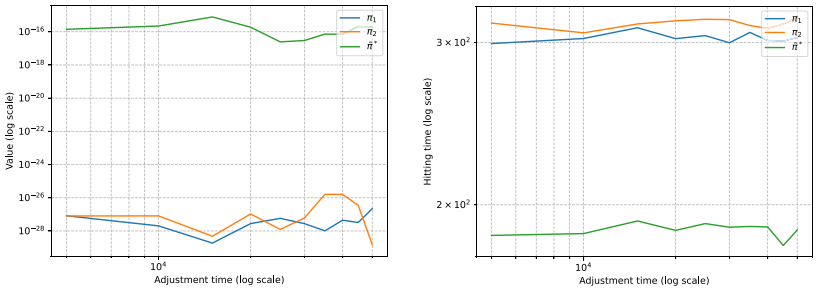


Fig. 10. Comparison of average value (left) and hitting times (right) for policies π_1, π_2, and $\tilde{\pi}^*$ with $\beta = 10$ and $N = 100$, given a starting configuration with a 3×3 square droplet with spin +1. Results are based on 20 experiments.

Fig. 9 explores the behavior of the policy $\tilde{\pi}^*$ for various temperatures. It shows that, at higher temperatures, the droplet with spin +1 evolves more chaotically, with protuberances emerging along different parts of the boundary, thereby rendering the torus edge correction visible. Nevertheless, the diagonal growth mechanism remains apparent, demonstrating robustness with respect to temperature increases.

To study the performance of the three policies more closely, we measure the value and the first hitting times $\tau_{\sigma^*}^{\sigma,\pi}$ to the target configuration across 20 experiments for different values of the adjustment time. The left panel of Fig. 10 shows the average values of each policy as a function of the adjustment time, for a starting configuration with a single 3×3 rectangle with spin +1 and discount factor $\lambda = 0.8$. The right panel of Fig. 10 shows the average hitting times from this starting configuration to the all-plus configuration. All experiments were conducted with $\beta = 10$ and $N = 100$. Recall that the relation between the first hitting time and the value function, given a discount factor λ, has been established in Corollary 2. The results indicate that the policy $\tilde{\pi}^*$ derived from the optimal policy in the auxiliary MDP indeed outperforms the policies π_1 and π_2 at small positive temperature and over a wide range of adjustment times, both in terms of the expected total discounted reward and the mean hitting time.

5. Discussion

We proposed and analyzed a control mechanism for the zero-temperature Ising model on a finite, two-dimensional, square lattice under asynchronous Metropolis dynamics, with the objective of driving the system from a configuration with a small droplet of +-spins to the all-plus configuration most efficiently. We formulated this control problem as an MDP with a reachability objective, for which the optimal policy is characterized by the Bellman optimality equations.

Our analysis relies on a reduction of the state space to local minima of the Hamiltonian, yielding an auxiliary MDP. For sufficiently long adjustment times, these configurations make the largest contribution to the Bellman optimality equations. For the auxiliary MDP, we determined the exact optimal policy by solving these equations recursively. Finally, simulations at low temperatures indicate that the resulting policy outperforms the two alternative policies in terms of expected discounted reward.

This work opens several avenues for future research. First, one could analyze the optimal control policy when the starting configuration contains multiple droplets with spin +1. A natural question would be to what extent the optimal growth mechanism of a droplet would be affected by the presence of other droplets with spin +1. Second, a potential direction for future research would be to analyze the control problem for different target configurations. If the target configuration is a local minimum of the Hamiltonian, the state-space reduction underlying our auxiliary MDP could be adapted to such a control problem. However, if the target configuration is not a local minimum, a reduction of the state space to local minima would be too coarse, so that the methods developed in this article cannot be applied directly. One possible approach would be to use our methods to first drive the system to a nearby local minimum, before driving it to the desired target configuration. Third, a more rigorous analysis of the Ising MDP for small values of the adjustment time could be performed, where the policy obtained from the auxiliary MDP may no longer be optimal. Fourth, the problem of controlling the Ising model at higher temperatures would be a worthwhile object of study, although much more difficult to handle. Fifth, it would be of interest to study the control of the Ising model on the infinite lattice, in different dimensions or on different types of graphs. Finally, the idea of reducing the state space to local minima of the Hamiltonian may yield insights for a more general class of MDPs for interacting particle systems, based on dynamics that are reversible with respect to a Gibbs measure.

CRediT authorship contribution statement

Maike C. de Jongh: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Formal analysis, Conceptualization. **Richard J. Boucherie:** Writing – review & editing, Supervision, Methodology, Conceptualization. **M.N.M. van Lieshout:** Writing – review & editing, Supervision, Methodology, Conceptualization.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial or not-for-profit sectors.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Proof of Theorem 1

Proof. Let $(X_n^\pi)_{n\geq 0}$ denote the Markov chain induced by policy π and let $r(X_n)$ denote the reward obtained in configuration X_n. First, note that for any pair of configurations $\sigma, \sigma' \in S$, conditioning on the first hitting time from σ to σ' and applying the strong Markov property [30, Section 1.4] yields

$$\begin{aligned}\mathbb{E}_\sigma^\pi\left[\sum_{n=1}^{\infty}\lambda^n r(\sigma')\mathbb{1}\{X_n^\pi=\sigma'\}\right] &= \sum_{n'=1}^{\infty}\mathbb{E}_\sigma^\pi\left[\sum_{n=1}^{\infty}\lambda^n r(\sigma')\mathbb{1}\{X_n^\pi=\sigma'\}\Big|\tau_{\sigma'}^{\sigma,\pi}=n'\right]\mathbb{P}(\tau_{\sigma'}^{\sigma,\pi}=n')\\ &= \mathbb{E}\left[\lambda^{\tau_{\sigma'}^{\sigma,\pi}}\right]\mathbb{E}_{\sigma'}^\pi\left[\sum_{n=0}^{\infty}\lambda^n r(\sigma')\mathbb{1}\{X_n^\pi=\sigma'\}\right]. \end{aligned} \tag{A.1}$$

For $\sigma=\sigma'$, it follows from expression (A.1) that

$$\mathbb{E}_\sigma^\pi\left[\sum_{n=0}^{\infty}\lambda^n r(\sigma)\mathbb{1}\{X_n^\pi=\sigma\}\right] = r(\sigma)+\mathbb{E}\left[\lambda^{\tau_\sigma^{\sigma,\pi}}\right]\mathbb{E}_\sigma^\pi\left[\sum_{n=0}^{\infty}\lambda^n r(\sigma)\mathbb{1}\{X_n^\pi=\sigma\}\right].$$

This yields, for each $\sigma\in S$,

$$\mathbb{E}_\sigma^\pi\left[\sum_{n=0}^{\infty}\lambda^n r(\sigma)\mathbb{1}\{X_n^\pi=\sigma\}\right] = \frac{r(\sigma)}{1-\mathbb{E}[\lambda^{\tau_\sigma^{\sigma,\pi}}]}. \tag{A.2}$$

We now obtain that $v_{\lambda,\beta,k}^\pi(\sigma)$ equals

$$\mathbb{E}_\sigma^\pi\left[\sum_{n=0}^{\infty}\lambda^n r(X_n^\pi)\right] = r(\sigma)+\sum_{\sigma'\in S}\mathbb{E}_\sigma^\pi\left[\sum_{n=1}^{\infty}\lambda^n r(\sigma')\mathbb{1}\{X_n^\pi=\sigma'\}\right].$$

Invoking expression (A.1) yields

$$v_{\lambda,\beta,k}^\pi(\sigma) = r(\sigma)+\sum_{\sigma'\in S}\mathbb{E}\left[\lambda^{\tau_{\sigma'}^{\sigma,\pi}}\right]\mathbb{E}_{\sigma'}^\pi\left[\sum_{n=0}^{\infty}\lambda^n r(\sigma')\mathbb{1}\{X_n^\pi=\sigma'\}\right].$$

Inserting (A.2) now completes the proof. □

Remark 4. Theorem 1 in fact holds more generally for finite MDPs (S, A, P, r) with bounded rewards, where the reward function r may depend on both the state and the action. See [26, Section 2.1] for the general framework. The proof for this general case follows exactly the same steps as the proof for the Ising model provided above. □

Appendix B. Transition probabilities

$$\hat{P}((i',j')|(i,j),a_{11}) = \begin{cases}1/3, & \text{if } i'=i, j'=j,\\ 2/3, & \text{if } i'=i, j'=j+1,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for} \quad \begin{matrix} i=3,\dots,N-2,N,\\ j=2,\dots,N-3,\end{matrix}$$

$$\hat{P}((i',j')|(i,j),a_{21}) = \begin{cases}1/3, & \text{if } i'=i, j'=j,\\ 2/3, & \text{if } i'=i+1, j'=j,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-3,\\ j=3,\dots,N-2,N,\end{matrix}$$

$$\hat{P}((i',j')|(i,j),a_{12}) = \begin{cases}5/9, & \text{if } i'=i, j'=j,\\ 7/27, & \text{if } i'=i, j'=j+1,\\ 5/27, & \text{if } i'=i, j'=j+2,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for} \quad \begin{matrix} i=3,\dots,N-2,N,\\ j=2,\dots,N-4,\end{matrix}$$

$$\hat{P}((i',j')|(i,j),a_{22}) = \begin{cases}5/9, & \text{if } i'=i, j'=j,\\ 7/27, & \text{if } i'=i+1, j'=j,\\ 5/27, & \text{if } i'=i+2, j'=j,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-4,\\ j=3,\dots,N-2,N,\end{matrix}$$

$$\hat{P}((i',j')|(i,j),a_0) = \begin{cases} 4/9, & \text{if } i'=i, j'=j, \\ 1/9, & \text{if } i'=i+1, j'=j \\ & \text{or } i'=i, j'=j+1, \\ 1/3, & \text{if } i'=i+1, j'=j+1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-3, \\ j=2,\dots,N-3, \end{matrix}$$

$$\hat{P}((i',j')|(i,j),a'_{11}) = \begin{cases} 1/2, & \text{if } i'=i, j'=j, \\ 1/2, & \text{if } i'=i, j'=j+1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-2, \\ j=2,\dots,N-3, \end{matrix}$$

$$\hat{P}((i',j')|(i,j),a'_{21}) = \begin{cases} 1/2, & \text{if } i'=i, j'=j, \\ 1/2, & \text{if } i'=i+1, j'=j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-3, \\ j=2,\dots,N-2, \end{matrix}$$

$$\hat{P}((i',j')|(i,j),a'_{12}) = \begin{cases} 5/8, & \text{if } i'=i, j'=j, \\ 1/4, & \text{if } i'=i, j'=j+1, \\ 1/8, & \text{if } i'=i, j'=j+2, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-2, \\ j=2,\dots,N-4, \end{matrix}$$

$$\hat{P}((i',j')|(i,j),a'_{22}) = \begin{cases} 5/8, & \text{if } i'=i, j'=j, \\ 1/4, & \text{if } i'=i+1, j'=j, \\ 1/8, & \text{if } i'=i+2, j'=j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for} \quad \begin{matrix} i=2,\dots,N-4, \\ j=2,\dots,N-2, \end{matrix}$$

$$\hat{P}((i',j')|(i,N-2),a_{11}) = \begin{cases} 1/4, & \text{if } i'=i, j'=N-2, \\ 3/4, & \text{if } i'=i, j'=N, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i=3,\dots,N-2,N,$$

$$\hat{P}((i',j')|(N-2,j),a_{21}) = \begin{cases} 1/4, & \text{if } i'=N-2, j'=j, \\ 3/4, & \text{if } i'=N, j'=j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } j=3,\dots,N-2,N,$$

$$\hat{P}((i',j')|(i,N-3),a_{12}) = \begin{cases} 7/18, & \text{if } i'=i, j'=N-3, \\ 31/144, & \text{if } i'=i, j'=N-2 \\ 19/48, & \text{if } i'=i, j'=N, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i=3,\dots,N-2,N,$$

$$\hat{P}((i',j')|(N-3,j),a_{22}) = \begin{cases} 7/18, & \text{if } i'=N-3, j'=j, \\ 31/144, & \text{if } i'=N-2, j'=j, \\ 19/48, & \text{if } i'=N, j'=j, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } j=3,\dots,N-2,N,$$

$$\hat{P}((i',j')|(i,N-2),a_0) = \begin{cases} 5/12, & \text{if } i'=i, j'=j, \\ 1/8, & \text{if } i'=i, j'=N, \\ 1/9, & \text{if } i'=i+1, j'=j, \\ 25/72, & \text{if } i'=i+1, j'=N, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } i=2,\dots,N-3,$$

$$\hat{P}((i',j')|(N-2,j),a_0) = \begin{cases} 5/12, & \text{if } i'=i, j'=j, \\ 1/8, & \text{if } i'=N, j'=j, \\ 1/9, & \text{if } i'=i, j'=j+1, \\ 25/72, & \text{if } i'=N, j'=j+1, \\ 0, & \text{otherwise,} \end{cases} \quad \text{for } j=2,\dots,N-3,$$

$$\hat{P}((i',j')|(N-2,N-2),a_0)=\begin{cases}7/18, & \text{if } i'=j'=N-2,\\ 1/8, & \text{if } i'=N,j'=N-2 \text{ or } i'=N-2,j'=N,\\ 13/36, & \text{if } i'=j'=N,\\ 0, & \text{otherwise,}\end{cases}$$

$$\hat{P}((i',j')|(i,N-2),a'_{11})=\begin{cases}1/3, & \text{if } i'=i,j'=N-2,\\ 2/3, & \text{if } i'=i,j'=N,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } i=2,\dots,N-2,$$

$$\hat{P}((i',j')|(N-2,j),a'_{21})=\begin{cases}1/3, & \text{if } i'=N-2,j'=j,\\ 2/3, & \text{if } i'=N,j'=j,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } j=2,\dots,N-2,$$

$$\hat{P}((i',j')|(i,N-3),a'_{12})=\begin{cases}4/9, & \text{if } i'=i,j'=N-3,\\ 5/27, & \text{if } i'=i,j'=N-2,\\ 10/27, & \text{if } i'=i,j'=N,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } i=2,\dots,N-2,$$

$$\hat{P}((i',j')|(N-3,j),a'_{22})=\begin{cases}4/9, & \text{if } i'=N-3,j'=j,\\ 5/27, & \text{if } i'=N-2,j'=j,\\ 10/27, & \text{if } i'=N,j'=j,\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } j=2,\dots,N-2,$$

$$\hat{P}((i',j')|(i,j),\tilde{a})=\begin{cases}1, & \text{if } (i',j')=(i,j),\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } i,j=3,\dots,N-2,$$

$$\hat{P}((i',j')|(2,j),\tilde{a})=\begin{cases}1/2, & \text{if } (i',j')=(2,j),\\ 1/2, & \text{if } (i',j')=(2,j-1),\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } j=3,\dots,N-2,$$

$$\hat{P}((i',j')|(i,2),\tilde{a})=\begin{cases}1/2, & \text{if } (i',j')=(i,2),\\ 1/2, & \text{if } (i',j')=(i-1,2),\\ 0, & \text{otherwise,}\end{cases} \quad \text{for } i=3,\dots,N-2,$$

$$\hat{P}((i',j')|(2,2),\tilde{a})=\begin{cases}1/3, & \text{if } (i',j')=(2,2),\\ 2/3, & \text{if } (i',j')=(0,0), \quad \text{and}\\ 0, & \text{otherwise,}\end{cases}$$

$$\hat{P}((i,j)|(i,j),0)=1, \quad \text{for all } (i,j)\in\hat{S}.$$

Appendix C. Supplementary data

Additional supporting material is available in the online version of the article at the publisher's website. The Supplementary Material consists of three parts:

1. **Detailed analysis of the zero-temperature Ising MDP**: technical details on the construction of the auxiliary MDP and its relation to the original control problem.
2. **Supporting tables and figures for computation of transition probabilities.**: supporting figures and tables used in deriving the transition probability kernel presented in Appendix B.
3. **Details of the proof of** Theorem 4: detailed computations for the proof of Theorem 4.

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.physa.2026.132218.

Data availability

No data was used for the research described in the article.

References

[1] E. Ising, Beitrag zur Theorie des Ferromagnetismus, Z. Phys. 31 (1925) 253–258.
[2] W. Lenz, Beiträge zum Verständnis der magnetischen Eigenschaften in festen Körpern, Phys. Z. 21 (1920) 613–615.
[3] K.M. Frahm, D.L. Shepelyansky, Ising-PageRank model of opinion formation on social networks, Phys. A 526 (2019) 121069.
[4] V.A. Hurtado-Marín, J.D. Agudelo-Giraldo, S. Robledo, E. Restrepo-Parra, Analysis of dynamic networks based on the Ising model for the case of study of co-authorship of scientific articles, Sci. Rep. 11 (5721) (2021).
[5] M. Ibrahimi, S. Zakany, M.C. Milinkovitch, Cell-biology effective interpretation of the Ising model describing skin color patterning, Phys. Rev. Res. 7 (2025) 023093.
[6] K. Simpson, A. L'Homme, J. Keymer, F. Federici, Spatial biology of Ising-like synthetic genetic networks, BMC Biol. 21 (185) (2023).
[7] E. Cohen, R. Heiman, M. Carmi, O. Hadar, A. Cohen, When physics meets signal processing: Image and video denoising based on Ising theory, Signal Process.: Image Commun. 34 (2015) 14–21.
[8] H. Oliaei-Moghadam, M. Hosseini-Farzad, N. TabeBordbar, Single-pixel quantum ghost imaging using generalized Ising model, Sci. Rep. 15 (22677) (2025).
[9] M.H. Bistervels, B. Antalicz, M. Kamp, W.L. Noorduin, Light-driven nucleation, growth, and patterning of biorelevant crystals using resonant near-infrared laser heating, Nat. Commun. 14 (2023) 6350.
[10] X. Jin, I.H. Riedel-Kruse, Optogenetic patterning generates multi-strain biofilms with spatially distributed antibiotic resistance, Nat. Commun. 15 (2024) 9443.
[11] D.R. Chialvo, S.A. Cannas, T.S. Gigera, D.A. Martin, D. Plenz, Controlling a complex system near its critical point via temporal correlations, Sci. Rep. 10 (12145) (2020).
[12] M.D. Louwerse, D.A. Sivak, Multidimensional minimum-work control of a 2D Ising model, J. Chem. Phys. 156 (2022).
[13] M. Naskar, M. Acharyya, Effects of random fields on the reversal of magnetisation of Ising ferromagnet, Phys. A 551 (2020) 124583.
[14] D.O. Cajueiro, Enforcing social behavior in an Ising model with complex neighborhoods, Phys. A 390 (9) (2011) 1695–1703.
[15] T.S.A.N. Simões, A. Coniglio, H.J. Herrmann, L. de Arcangelis, Modeling public opinion control by a charismatic leader, Phys. A 648 (2024) 129921.
[16] G. Zaklan, F. Westerhoff, D. Stauffer, Analysing tax evastion dynamics via the Ising model, J. Econ. Interact. Coord. 4 (2009) 1–14.
[17] G.B. Arous, R. Cerf, Metastability of the three dimensional Ising model on a torus at very low temperatures, Electron. J. Probab. 1 (1996) 1–55.
[18] R. Cerf, F. Manzo, Nucleation and growth for the Ising model in d dimensions at very low temperatures, Ann. Probab. 41 (6) (2013) 3697–3785.
[19] E.N.M. Cirillo, J.L. Lebowitz, Metastability in the two-dimensional Ising model with free boundary conditions, J. Stat. Phys. 90 (1–2) (1998) 211–226.
[20] E.N.M. Cirillo, V. Jacquier, C. Spitoni, Metastability of synchronous and asynchronous dynamics, Entropy 24 (4) (2022) 1–18.
[21] R. Kotecký, E. Olivieri, Droplet dynamics for asymmetric Ising model, J. Stat. Phys. 70 (5–6) (1993) 1121–1148.
[22] F. Nardi, E. Olivieri, Low temperature stochastic dynamics for an Ising model with alternating field, Markov Process. Relat. Fields 2 (1) (1996) 117–166.
[23] F. Nardi, A. Zocca, Tunneling behavior of Ising and Potts models in the low-temperature regime, Stoch. Process. Appl. 129 (11) (2019) 4556–4575.
[24] E.J. Neves, R.H. Schonmann, Critical droplets and metastability for a Glauber dynamics at very low temperature, Comm. Math. Phys. 137 (1991) 209–230.
[25] O. Penrose, J.L Lebowitz, Rigorous treatment of metastable states in the van der Waals–Maxwell theory, J. Stat. Phys. 3 (2) (1971).
[26] M.L. Puterman, Markov Decision Processes: Discrete Stochastic Dynamic Programming, John Wiley & Sons, New York, 1994.
[27] W.B. Powell, Reinforcement Learning and Stochastic Optimization: A Unified Framework for Sequential Decisions, John Wiley & Sons, Hoboken, 2022.
[28] T.M. Liggett, Interacting Particle Systems, Springer Berlin Heidelberg, New York, 1985.
[29] K. Binder, D.W. Heermann, Monte Carlo Simulation in Statistical Physics, fourth enl. ed., Number 80, Springer, New York, 2002.
[30] J.R. Norris, Markov Chains, in: Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 1997.