Physica A 649 (2024) 129956

Contents lists available at ScienceDirect

Physica A

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

The Bellman equation and optimal local flipping strategies for kinetic Ising models

F. Caravelli

Theoretical Division (T4), Los Alamos National Laboratory, Los Alamos, NM 87545, USA

ARTICLE INFO

Keywords:
Bellman equation
Kinetic Ising model
Work extraction
Kobolds

ABSTRACT

There is a deep connection between thermodynamics, information and work extraction. Ever since the birth of thermodynamics, various types of Maxwell demons have been introduced in order to deepen our understanding of the second law. Thanks to them, it has been shown that there is a deep connection between thermodynamics and information, and between information and work in a thermal system. In this paper we study the problem of energy extraction from a thermodynamic system satisfying detailed balance by an agent with perfect information, e.g. one that has an optimal strategy, given by the solution of the Bellman equation, in the context of Ising models. We call these agents kobolds, in contrast to Maxwell's demons which do not necessarily need to satisfy detailed balance. This is in stark contrast with typical Monte Carlo algorithms, which choose an action at random at each time step. It is thus natural to compare the behavior of these kobolds to a Metropolis algorithm. For various Ising models, we study numerically and analytically the properties of the optimal strategies, showing that there is a transition in the behavior of the kobold as a function of the parameter characterizing its strategy.

1. Introduction

In recent years, there has been a lot of interest in the connection between thermodynamics, information, and the role that thought experiments involving beings such as Maxwell demons play [1–6]. An example of a Maxwell demon [7] in a thermodynamic system is a being able to observe the speed of particles approaching a gate separating two gases, and open it if only the particle speed is such that the temperature of one of the gases can be raised, or acts in order to raise the pressure of one of the gases (temperature vs pressure demons). These thought experiments are useful to understand if the second law of thermodynamics can be locally violated. This line of thought has led later many researchers to actually propose various versions of the Maxwell demon, including notoriously Szilard [8,9], Brillouin [10], Landauer [11] and Bennett [12]. A typical Maxwell demon is able, quintessentially but with some restriction in certain cases, to do anything on the system. Depending on the point of view one takes, these thought experiments often can say something both about the system and the demon itself. If one assumes that thermodynamics is valid, then one can use these arguments to infer what demons cannot do [13]. On the other hand, if one assumes that the ability of the demons is valid, how thermodynamics can be violated? Either way, these thought experiments present a significant challenge for physicists, forcing them to carefully consider the nature of thermodynamic laws. As in the case of Landauer's work, and later Bekenstein and Hawking [14–16], these gedankexperiments with demons can lead to new discoveries about extreme regimes, ranging from nanoscale devices to black holes. A resolution of the violation of the second law can be obtained by assigning information to molecules or atoms and introducing logical irreversibility on the operation of the operation performed by the demon, leading to an entropy increase due to information erasure. If however no information is erased and the operation is reversible, the second law is not violated. In

E-mail address: caravelli@lanl.gov.

https://doi.org/10.1016/j.physa.2024.129956
Received 29 February 2024
Available online 20 July 2024

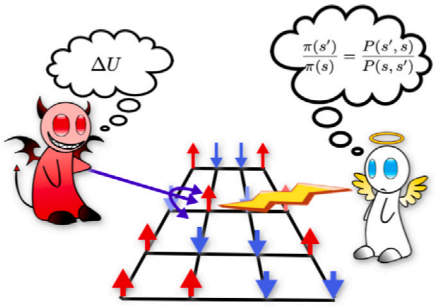

Fig. 1. Representation of the game: the kobolds (minor demons) observe the phase space of the Ising model and decide to attempt to flip a spin according to an optimal strategy in order to gain energy and locally reduce the energy. However, detailed balance is enforced, and the spin-flip can be rejected by the requirement of time reversibility.

fact, physical implementations of a Maxwell demon have been proposed [17] and realized [18], confirming such a picture. These ideas can also be extended from the classical to the quantum realm, proving their extreme generality [19–21], and has also have important applications to quantum thermodynamics [22,23] and non-equilibrium statistical mechanics [24].

Maxwell demons are often very far from real beings, in the sense that they stretch and bend slightly the roles of the games, and time is not as important. Technology innovation is however a constant struggle between satisfying physical laws while gaining as much as possible from the system of interest, whether this is energy or computing power, or something else. In this spirit, we consider here more mundane types of Maxwell demons, those that play by physical laws but with optimal strategies in complex landscapes. An optimal strategy is one which, given any location of the agent in a state space of a system, performs an action that maximizes a certain function. These optimal strategies however require perfect "information", e.g., the demon knows everything about the phase space of the system beforehand, and in particular, has full phase space observability. We will use, for this purpose, the complex landscape of an Ising model. We note that this is not information as it is commonly stated in the theory of communication, but in the game theoretic sense. This simply means that the Kobold has perfect knowledge of the energy landscape in the phase space, but has to satisfy detailed balance when performing a move; it thus follows a Markov decision process.

In particular, this paper addresses the following problem. Consider a thermodynamic demon that knows everything about the dynamics of an Ising model, and desires to reach the ground state of the system by acting locally on a certain spin at a fixed temperature. We choose the demon to act locally to closely resemble the behavior of a physical device, or a Monte Carlo algorithm. For this reason, we introduce kinetics for the Ising system which preserves detailed balance and thus microscopic reversibility. The question we then ask is: how hard would it be to act optimally, and to navigate through the phase space of the model in order to reach one of the states with minimum energy while satisfying the detailed balance? Such an entity is not quite as powerful as a Maxwell demon, as it still has to obey the laws of detailed balance and can only act locally on a single spin. Thus it resembles more of a kobold of German folklore, or goblin in English culture rather than a full-fledged demon. However, the kobold has a perfect knowledge of the state space of the model (perfect information). A cartoon representation of the game is shown in Fig. 1. In addition, the kobold values time, and has an associated parameter with which discounting occurs; discounting is associated to a loss of energy as a function of time, and is a parameter that can be changed by the kobold. While microscopic reversibility is preserved at any intermediate step, there is a particular state at which the kobold decides to not act anymore and saves the energy stored. In this sense, eventually, microscopic reversibility is broken by the kobold's decision to take no action.

To answer this question, we need to specify some further details. First, we assume that we have a statistical physics system, described by spin variables $s_i = \pm 1$, and an energy written in the form

$$H = \sum_{ij} J_{ij} s_i s_j. \tag{1}$$

By navigating, we mean that we are able to observe our current phase space state $\phi = [s_1, \dots, s_n]$, and that we need to act *locally* via a map $a : \phi \to \phi$, where the action a can act on only one spin or none to gain energy from the system.

This setup is not different from the typical Markov Chain Monte Carlo, used to thermalize or anneal a statistical model. For instance, the Metropolis-Hastings [25,26] or Glauber [27] algorithm for the time evolution of a spin model. In the case of the latter, the transition rates are determined by a spin flip. The model dynamics are given by the spin-flip probability

$$W_{ij} = p(\phi_i \to \phi_j) = \frac{e^{-\beta \Delta U(\phi_i,\phi_j)}}{1 + e^{-\beta \Delta U(\phi_i,\phi_j)}}. \tag{2}$$

This implies that the system satisfies detailed balance and can thus approach thermodynamic equilibrium [28]. Then, we have an external clock with a discrete time that labels the operations and discounting. Here, we choose $\Delta U(\phi_i, \phi_j) = H(\phi_i) - H(\phi_j)$, so that the dynamics attempt to find the minimum of the Hamiltonian H of Eq. (1). Note that Glauber dynamics preserves the entropy of the model. This is because the probability distributions and transition rates satisfy $P_i W_{ij} = P_j W_{ji}$, from which we obtain that the Schnakenberg entropy production [29] is zero, e.g.

$$\Pi(t) = \frac{k}{2} \sum_{ij} (P_i(t) W_{ij} - P_j(t) W_{ji}) \log \frac{P_i(t) W_{ij}}{P_j(t) W_{ji}} = 0. \tag{3}$$

Thus, the change in free energy of the model reduces to the change in potential energy of the system. Interestingly, this also means that a kobold acts as a reversible computer [30]. Another point of view is that the system represents the battery, and the kobold is an operation by which we wish to extract energy from the system without wasting it in entropy production.

One question one might ask is why is a kobold different from a standard Monte Carlo "demon". We could in fact consider a system out of equilibrium and let it thermalize, and extract the energy as a result of thermalization. However, if $\bar{E}$ is the expected thermalization energy if $E(0) < \bar{E}$, the demon would actually have to supply energy to the system. A kobold, on the other hand, has an optimal strategy for any initial condition and can adjust its discounting factor to try to maximize it. In doing so, it can also reduce fluctuations over the gained energy.

The paper is organized as follows. We will first discuss the theoretical underpinning of kobolds, envisioned as a stochastic Bellman equation for the optimal strategy [31–33]. We will derive an analytical formulation for the Bellman equation restricted to the case of Glauber dynamics, and show that this can be written in terms of projector operators on the phase space. We will then provide numerical results both for the optimal strategies and discounted gained energy. In particular, we will show that depending on the discounting factor, a kobold can take either a greedy or a wise strategy approach. A greedy strategy is one in which the local spin flip action is such that it always minimizes the energy, while a wise strategy allows for local increases in energy. In this respect, a greedy approach can be compared to a low-temperature Metropolis algorithm, while a wise approach to a higher-temperature one. However, the analogy ends there, as the kobold tries independently to maximize the energy extracted from the system. This is the reason why we compare the behavior of such an agent to a "Metropolis" agent.

Conclusions follow.

2. Optimal strategies and the bellman equation

2.1. The discounted reward

We assume the following rules for the abilities and limitations of the kobold. First, the kobold knows everything about the system: the couplings, has a complete picture of the phase space, and thus also knows the location of the ground state. In the language of economics and game theory, the kobold has perfect information (rather than imperfect, which would occur if the kobold only knew a certain part of the phase space). However, it has some restrictions, most importantly, it has to respect constraints such as the detailed balance. This implies that while it might decide to attempt a spin flip to gain energy, whether this occurs or not is determined by the acceptance rule based on the temperature of the system. The second restriction is that the kobold can only operate locally in time and space, meaning that it has to perform actions sequentially, one spin at a time, but can choose any spin. The third restriction is that the kobold does not get to choose the initial state from which it starts to operate on the system. What the kobold gains by playing this game is energy. If at each time step the kobold gains ΔU_t of energy via a certain action (spin flip), then after τ steps, the kobold will have gained

$$R_\gamma = \sum_{t=1}^{\tau} R_\gamma(t) = \sum_{t=1}^{\tau} \gamma^t \Delta U(\phi_t, \phi_{t-1}). \tag{4}$$

In the equation above, $\gamma \in (0,1]$ is a discounting factor. In a certain sense, the kobold knows the rules of finance and knows that energy today is better than energy tomorrow.

In practice, if $\gamma < \gamma' < 1$, this means that even if the actions are such that the energy gain is positive at each time step, one has $R_\gamma < R_{\gamma'}$.. When $\gamma = 1$, we have

$$R_1 = \sum_{t=1}^{\tau} R_1(t) = \sum_{t=1}^{\tau} (U_t - U_{t-1}) = U_\tau - U_0. \tag{5}$$

In the equation above, and in the context of the Bellman equation, the parameter γ characterizes the kobold strategy, as it defines how important time is to the kobold.

2.2. Bellman equation

In general, a kobold agent wants to maximize the reward over time (possibly over an infinite horizon) via local actions $a_\phi = \pi(\phi)$, meaning that by observing the local state of the system, the kobold can attempt to flip one spin at time t with the intention of gaining energy. Thus, the parameter γ represents how important the speed at which the kobold attempts to reach the ground state is. It is important to stress that the form of the discounted energy in Eq. (4) can be obtained by assumptions on the time invariance of the optimal strategy, and is thus natural.

In the following, we use a notation in which $\pi(\phi)$ determines both the action and the state, as the two in this context are the same. If the action $\pi(\phi)$ is to flip the kth spin of ϕ, then this determines the new state ϕ'. Of course, if the kobold reaches the ground state, it will want to perform no action. This implies that if the system contains N spins, there can be $N+1$ possible actions. We say attempt, because of course the acceptance probability of such action is determined by the temperature of the system, and thus if currently the kobold is in the state ϕ, the action $\pi(\phi)$ of the kobold can lead to a state ϕ' with probability $P_\beta(\phi; \pi(\phi), \phi')$, where $\beta = 1/T$ (we assume $\kappa_B = 1$) being a temperature-like parameter. In fact, sequential Monte Carlo schemes, the detailed balance condition leads to the Gibbs state only with random updates.. The average energy reward is then given by

$$\langle R_\gamma \rangle \equiv \langle U \rangle_{t,\pi(\phi)} = \sum_{\phi_t} P_\beta(\phi_t; \pi(\phi_t), \phi_{t-1}) \gamma^t \Delta U(\phi_t, \phi_{t-1}).$$

For the kobold, an ideal world would be such that $P_\tau(\phi;\pi(\phi),\phi') = \delta_{\phi,\pi(\phi)}$. The setup is now such that the kobold is playing a (reversible) Markov Decision Process in discrete time. A Monte Carlo algorithm works typically by the same sequential rules, with the key difference that the action $\pi(\phi)$ is random in some form, e.g.random in some form, such as a random single (Metropolis or Glauber) or multiple (Kawasaki) spin flip [34].. This is because random sequential MC schemes lead to asymptotically thermal states. instead uses the rules of an economist: which is dynamic programming and the notion of discounting, it knows everything about the system and applies an optimal strategy $\pi(\phi)$.. Under these assumptions, an infinite-horizon decision problem takes the form of a maximization, e.g.,

$$V^*_{\tau,\gamma}(\phi) = \max_{a_0,\dots,a_\tau} \langle R_\gamma(a_0,\dots,a_\tau)\rangle, \tag{6}$$

where ϕ is the starting state and a_i the actions taken at each time step. The optimal reward $V^*_{\tau,\gamma}(\phi)$ is the maximum (discounted) reward that the kobold can obtain starting from a certain state in phase space. For $\gamma = 1$, this corresponds to the maximum energy the kobold can extract. We then see the reason for such a gedankexperiment. The kobold represents the best possible algorithm designed to reach the ground state of the system and represents the best possible line of action an algorithm can take at a fixed temperature. Here, the assumption is that the system is *not* annealed, probabilities are time-independent, and the kobold can act with an infinite time horizon. For this type of problem, the optimal solution $\pi^*(\phi)$, and the optimal reward $V^*_{\tau,\gamma}(\phi)$ can be obtained by solving the stochastic Bellman equation. Above, $\pi^*(\phi)$ is the best possible action that the kobold takes if it finds itself in state ϕ. The Bellman equation is given by the following linear relationship

$$V^*_{\tau,\gamma}(\phi) = \sum_{\phi'} P_\beta(\phi;\pi^*(\phi),\phi')\Big(-\Delta U(\phi,\phi') + \gamma V^*_{\tau,\gamma}(\phi')\Big). \tag{7}$$

The equation above is written implicitly: to solve it, one would already need to have π^*, which we do not have.. There are many ways to solve it; however, we use the strategy iteration method, which starts from a random initial state $V_0(\phi)$ and then uses the iteration scheme

$$\begin{aligned} V_{\tau,\gamma;k+1}(\phi) &= max_a \sum_{\phi'} P_\beta(\phi;a,\phi')\Big(-\Delta U(\phi,\phi') \\ &\qquad + \gamma V_{\tau,\gamma;k}(\phi')\Big), \\ \pi^*(\phi) &= arg\ max_a \sum_{\phi'} P_\beta(\phi;a,\phi')\Big(-\Delta U(\phi,\phi') \\ &\qquad + \gamma V^*_\tau(\phi')\Big). \end{aligned}$$

which gives, at convergence, the optimal strategy $\pi^*(\phi)$.

Note that here we immediately face a computational problem. While the number of actions is $N+1$, the number of states ϕ scales exponentially with the size of the system. Thus, unlike the kobold, we will have to use small systems by Monte Carlo standards. Despite such a curse of dimensionality, the Bellman solution for optimal sequential Markov decision processes is regarded as the most feasible one [33]. Nonetheless, this will be sufficient to obtain a picture of the complexity of Ising models from the point of view of the kobold, as we will discuss in a moment.

2.3. Application in kinetic statistical mechanics

Let us provide an immediate comment on why such a technique is useful. Since we are essentially finding the functions V^* and π^* for every state of the model if one is interested only in the energy of the model, exhaustive search (or brute force) works much better than solving the Bellman equation. However, Bellman's optimal strategy and discounted energy give a more complete picture of how the state space of an Ising model is tangled, and how one could untangle it and navigate through the states, or make less blind moves in Monte Carlo algorithms.

First, we note that Eq. (7) can be written in an explicit form using the Glauber transition of Eq. (2), given a certain action π on the phase space:

$$P_\beta(\phi;\pi(\phi),\phi') = \frac{\delta_{\phi\phi'} + \delta_{\phi'\pi(\phi)} e^{-\beta\Delta U(\phi,\pi(\phi))}}{1+e^{-\beta\Delta U(\phi,\pi(\phi))}}. \tag{8}$$

From this, it follows that we can write the Bellman equation in the form

$$\sum_{\phi'} \mathcal{O}(\phi,\phi') V^*_{\tau,\gamma}(\phi') = -\frac{\Delta U\left(\phi,\pi^*(\phi)\right)}{1+e^{-\beta\Delta U(\phi,\pi^*(\phi))}} \tag{9}$$

where

$$\mathcal{O}(\phi,\phi') = \delta_{\phi\phi'} \frac{1+e^{-\beta\Delta U(\phi,\pi^*(\phi))} - \gamma}{1+e^{-\beta\Delta U(\phi,\pi^*(\phi))}} \tag{10}$$

$$- P_{\pi^*}\left(\phi,\phi'\right) \frac{\gamma e^{-\beta\Delta U(\phi,\pi^*(\phi))}}{1+e^{-\beta\Delta U(\phi,\pi^*(\phi))}} \tag{11}$$

Above, $P_{\pi^*}(\phi, \phi')$ is a state transition matrix, e.g. $P_{\pi^*}(\phi, \phi') = \delta_{\phi',\pi^*(\phi)}$. After a brief calculation, it follows that we can write the exact solution for V^* in the form:

$$V^*_{\tau,\gamma}(\phi) = \sum_{\phi'} (I - \gamma D P_{\pi^*})^{-1}_{\phi,\phi'} \tilde{U}(\phi'). \tag{12}$$

where

$$\tilde{U}(\phi) = -\Delta U(\phi, \pi^*(\phi)) R(\phi), \quad D(\phi, \phi') = \delta_{\phi,\phi'} R(\phi)$$
$$\text{and where } R(\phi) = \frac{e^{-\beta \Delta U(\phi,\pi^*(\phi))}}{1 - \gamma + e^{-\beta \Delta U(\phi,\pi^*(\phi))}}. \tag{13}$$

Assuming a random initial condition, the average maximum discounted energy obtained from the kobold is given by:

$$\langle V_{\tau,\gamma} \rangle = \frac{1}{2^N} \sum_{\phi} V^*_{\tau,\gamma}(\phi), \tag{14}$$

where N is the number of spins. This is the average (discounted) energy that an optimal player (a kobold) can achieve by acting locally on the system, and against the temperature, by reaching the ground state. In general, the Bellman equation selects one possible action for each phase space state, but there might be some other stochastic policies, which we do not consider here, such that π is a stochastic function as well. This helps however in evaluating the amount of information that a policy contains. If we have M possible actions per phase state and K states in phase space, then we have M^K possible policies. We can then evaluate the information in the policy using the entropy (in a base 2) $S_\pi = K \log_2 M$, which is an estimate for the amount of information associated with a kobold, including an optimal one.

2.4. Example: classical zener hamiltonian

Let us consider first a simple application of the equations above. We take as a model the classical Zener Hamiltonian for an Ising spin in an external field, which is given by $H = \mu_0 h s$, with $s = \pm 1$ representing a single spin. In this case, we have only two states, and the optimal strategy is easy to guess: it is a spin flip if $sign(s) = -sign(h)$, and do nothing otherwise. Let us assume $h > 0$. Then,

$$\begin{aligned}
\begin{pmatrix} V^*_{\tau,\gamma}(+) \\ V^*_{\tau,\gamma}(-) \end{pmatrix} &= \begin{pmatrix} \frac{2-\gamma}{2} - \frac{\gamma}{1+\gamma} & 0 \\ \frac{\gamma e^{\beta\mu_0 h}}{1+e^{\beta\mu_0 h}} & \frac{1+e^{\beta\mu_0 h}-\gamma}{1+e^{\beta\mu_0 h}} \end{pmatrix}^{-1} \begin{pmatrix} 0 \\ \mu_0 h \end{pmatrix} \\
&= \frac{1}{d} \begin{pmatrix} \frac{1+e^{\beta\mu_0 h}-\gamma}{1+e^{\beta\mu_0 h}} & -\frac{\gamma e^{\beta\mu_0 h}}{1+e^{\beta\mu_0 h}} \\ 0 & \frac{2-\gamma-\gamma^2}{2+2\gamma} \end{pmatrix} \begin{pmatrix} 0 \\ \mu_0 h \end{pmatrix} \\
&= \frac{\mu_0 h}{1 + e^{\beta\mu_0 h} - \gamma} \begin{pmatrix} -(\gamma e^{\beta\mu_0 h}) \frac{2+2\gamma}{2-\gamma-\gamma^2} \\ 1 + e^{\beta\mu_0 h} \end{pmatrix}
\end{aligned}$$

where $d = \frac{2-\gamma-\gamma^2}{2+2\gamma} \frac{1+e^{\beta\mu_0 h}-\gamma}{1+e^{\beta\mu_0 h}}$. It follows that the average energy the kobold can gain is

$$\langle V^*_{\tau,\gamma} \rangle = \frac{\mu_0 h}{2(1 + e^{\beta\mu_0 h} - \gamma)} \left(1 + e^{\beta\mu_0 h}(1 - 2\frac{\gamma + \gamma^2}{2 - \gamma - \gamma^2})\right)$$

It follows that with a proper discounting strategy, e.g. $\gamma \in [0, \frac{1}{2}(\sqrt{3} - 1)]$, the kobold can always extract energy from the system, at any temperature.

2.5. Properties of P_{π^*}

In this section, we discuss the properties of P_{π^*}, which are useful to understand the behavior of a kobold in the case of an Ising system via local spin flipping.

In fact, as it turns out, the matrix $P_{\pi^*}(\phi, \phi')$ is a projector operator, as its spectrum is only composed of 0's and 1's. Intuitively, the proof follows from the fact that we have a directed graph of outdegree equal to one, with an absorbing state, which is what we discuss below.

The important quantity in the Bellman equation for kinetic Ising models is the transition matrix P_{π^*}, which could be in principle a permutation matrix. However, as it turns out, the optimal strategy is always such that P_{π^*} is a projector operator, e.g. $P^2_{\pi^*} = P_{\pi^*}$, as we will prove in a moment. The kobold's strategy is to lower the energy, by transitioning between states and eventually to one of the absorbing states. Then, the basins of attraction of a particular absorbing state G as a sink in a directed tree, similar to Fig. 5, which follows from the fact that the outdegree of every node is always one. If we have G absorbing states, then the matrix P_{π^*} can be written in block diagonal form. Let us call these sub-blocks $P^G_{\pi*}$; then, simply one has $P_{\pi^*} = \bigoplus_G P^G_{\pi*}$, and we can focus on each sub-block. Now, we can label the nodes such that if i is the numerical value of a certain node and i' down the tree, then $l(i) > l(i')$. Such labeling is always possible because we have a directed acyclic graph, and one example is shown in Fig. 5. The fact that it is acyclic follows from the presence of an absorbing state in a graph with outdegree one. In fact, assume by absurd that one has a cycle in such a graph with an outdegree one. If it were a cycle, there could not be an absorbing state; thus, there would have to be at least one node with degree three and outdegree two, as one directed edge would go into the cycle and another toward the absorbing state. However, this is incompatible with the fact that the outdegree must always be one. With the labeling l we see that these energy state transitions correspond to an upper triangular matrix elements of $P^G_{\pi^*}$. Since the absorbing state is the only state that has null action (the kobold will want to remain in that state), this means the absorbing state is the only element with a one on the diagonal. Each sub-block can always be written, via state relabeling, as

$$P^G_{\pi^*} = \begin{pmatrix} 0 & \times & \times & \times & \times \\ & 0 & \times & \times & \times \\ & & 0 & \times & \times \\ & \mathbf{0} & & 0 & \times \\ & & & & 1 \end{pmatrix}. \tag{15}$$

From this, since the matrix is triangular, it follows that if such a sub-block is D-dimensional, then $D-1$ eigenvalues are 0, and only one is 1. This is enough to show that it is a projector operator, and that it satisfies the condition $(P^G_{\pi^*})^2 = P^G_{\pi^*}$. Thus, the spectrum of P_{π^*} is simply determined by the number of absorbing states. In fact, the number of 1s corresponds to the number of absorbing states of the system, and the rest of the spectrum contains only zeros. Since every diagonal sub-block is a projector operator, P_{π^*} is also a projector operator.

This result is useful for the following reasons. First, in the limit $\gamma \to 1$, we have $D \to I$ and $R(\phi) \to 1$. Having $\gamma = 1$ essentially means no discounting, which can be interpreted as the kobold exhibiting no haste. We can write explicitly the inverse, in the neighborhood of $\gamma = 1^-$ and using the property that $P^2_{\pi^*} = P_{\pi^*}$, as

$$(I - \gamma P_{\pi^*})^{-1} = I + \frac{\gamma}{1-\gamma} P_{\pi^*}, \tag{16}$$

and then the discounted energy values are simply given by

$$V^*_{T,\gamma\approx 1}(\phi) = -\sum_{\phi'} \left(I + \frac{\gamma}{1-\gamma} P_{\pi^*} \right)_{\phi,\phi'} \Delta U(\phi', \pi^*(\phi')) \tag{17}$$

We can thus use the exact inverse to regularize the limit $\gamma \to 1$, which is otherwise ill-defined in the general case. We use the following assumption. Typically, $\partial_\gamma \pi^* = 0$ almost everywhere. Then, assuming that the optimal strategy can be analytically extended from $\gamma = 1-\epsilon$ to $\gamma = 1^-$, i.e. that it is constant, we can obtain a discounted effective energy of the form

$$\begin{aligned} \tilde{V}^*_{\tau,\gamma}(\phi) &= \lim_{\gamma\to 1^-} (1-\gamma) V^*_{\tau,\gamma}(\phi) \\ &= \sum_{\phi'} (P_{\pi^*})_{\phi,\phi'} \Delta U(\phi', \pi^*(\phi')), \end{aligned}$$

which is a regularized $\gamma = 1$ limit for the optimal strategy. In this regime, all absorbing states are ground states.

We have been careful to refer to G as the absorbing state and not the ground state. If $\gamma = 1$, of course, all absorbing states are ground states, but at $\gamma < 1$ this is not guaranteed, although likely for $\gamma \approx 1$. Intuitively, this is because the kobold might find it more rewarding to stop at a certain state rather than attempting to reach the ground state, given that there is a cost in how long the game takes, and in the intermediate steps that might actually increase rather than lowering the energy.

3. Numerical results

We consider, for the purpose of this paper, four types of Ising models, of which three are ferromagnetic and one frustrated. The testbed of our analysis is the Ising models given below

$$H = \begin{cases} \frac{1}{2}\sum_{i=1}^{N-1} s_i s_{i+1} & \text{Ising 1D}, \\ \frac{1}{2}\sum_{<ij>=1}^{N_1 N_2} s_i s_j & \text{Ising 2D}, \\ \frac{1}{2N}\sum_{ij=1}^{N} s_i s_j & \text{Curie-Weiß}, \\ \frac{1}{2\sqrt{N}}\sum_{ij=1}^{N} J_{ij} s_i s_j & \text{Spin Glass} \quad J_{ij} = \pm 1. \end{cases} \tag{18}$$

where in the latter case we consider $P(J_{ij} = \pm) = 1/2$. Since solving the Bellman equation requires solving iteratively a vectorial equation of the size of the phase space, we are forced to study relatively small systems, with 12 spins.

We iterate Eq. until $\frac{1}{2N}\|V_{k+1} - V_{k+1}\|^2 < \epsilon$, with $\epsilon = 10^{-5}$, read the strategy out and sort the spin states with descending discounted potential $V^*(\phi)_{\tau,\gamma}$. The plots are shown in Fig. 2 for $\gamma = 0.95$. In the inset of Fig. 2, we plot $\langle V^*_{\tau,\gamma}\rangle$ for each model at $\gamma = 0.95$. The quantity $\langle V^*_{\tau,\gamma}\rangle$ is the average utility (discounted energy) that the kobold can obtain from the Ising model, assuming that we let him start from any spin state at infinite time. This average shows, in arbitrary units of energy and assuming equal conditions, which model provides more gains via a Monte Carlo method, given the assumption of only local moves. Intuitively, we see that the spin glass led to more gains than the others, with the Ising 2D being harder than Ising 1D model, and the mean-field ferromagnetic model being the easiest in comparison. Such hierarchy seemingly makes sense from a computational perspective, as we would expect the

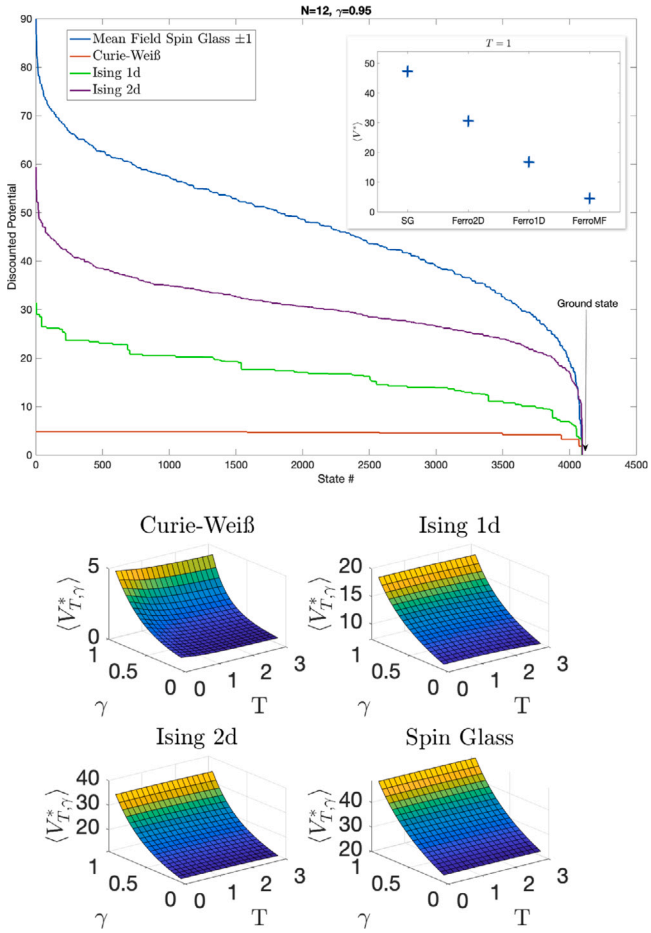

Fig. 2. *Top:* maximum discounted energy $V^*_{\tau,\gamma}$ as a function of the states ϕ for $N = 12$, ordered from high to low, and $\langle V^*_{\tau,\gamma}\rangle$ shown in the inset. We see the difference between the spin glass, Ising 2D and Ising 1D, and the Curie-Weiß model. Since $\gamma = 0.95$, the low energy states are associated with the ground states of the model. *Bottom:* The profile of the function $\langle V^*_{\tau,\gamma}\rangle$ as a function of T and γ. While the dependence on T is negligible for these values of T, the dependence on γ is strong, due to the fact that high discounting rates imply less value to the energy obtained in the future.

spin glass to have a longer way down to the ground state. However, we note that the two key parameters we need to analyze are $\langle V^*_{\tau,\gamma}\rangle$ as a function of temperature, of course, and as a function of the parameter γ. If $\gamma \approx 1$, time for the kobold is not an issue. At strong discounting, however, longer chains to reach the ground state imply losses. This can be seen in Fig. 2 (bottom), in which we see that for all models at stronger discounting ($\gamma \to 0$) the effective value is reduced.

An interesting question is how the optimal strategy of the kobold changes as a function of $\beta^{-1} = T$ and γ. We explore this in two ways. First, we analyze whether the only allowed strategies are those that decrease energy, which we call 'greedy' kobold strategies ($S = +1$). On the other hand, if the kobold uses strategies that momentarily increase the energy, these are termed 'wise' ($S = -1$), as they allow the kobold to reach a lower state more quickly. We find that for both the 1D Ising model and the Curie-Weiss model, the strategy is always greedy. However, we plot S as a function of γ and T in Fig. 3. We see that for the 2D Ising model, the parameter S does not depend on $\beta^{-1} = T$, but it depends on γ and for $\gamma > 0.8$, the strategy becomes *wise*, while it is *greedy* for $\gamma < 0.8$. In the case of the spin glass, we also find a dependence on temperature, with a transition between greedy and wise strategies occurring at approximately $\gamma = 0.62$ for $T = 3$, and $\gamma \approx 0.55$ for $T = 0.1$.

The maximum number of steps required (on average) can be estimated by the depth of the optimal policy tree. The cases of the Ising 2D and Spin Glasses are shown in Fig. 4, showing the difference between high and low discounting rates for $T = 3$. The change in the strategy trees, depicted in different colors, illustrates that kobolds use different strategies at various discounting rates. Notably, at high discounting rates, kobolds develop two additional absorbing states on top of the ground states in the Ising 2D case.

"The difference between the kobold strategy and the Monte Carlo thermalization process is shown in Fig. 6 for the Ising 2D and spin glass cases. The random strategy fluctuates significantly in energy, whereas the kobold's strategy reaches the ground state in just a few steps.

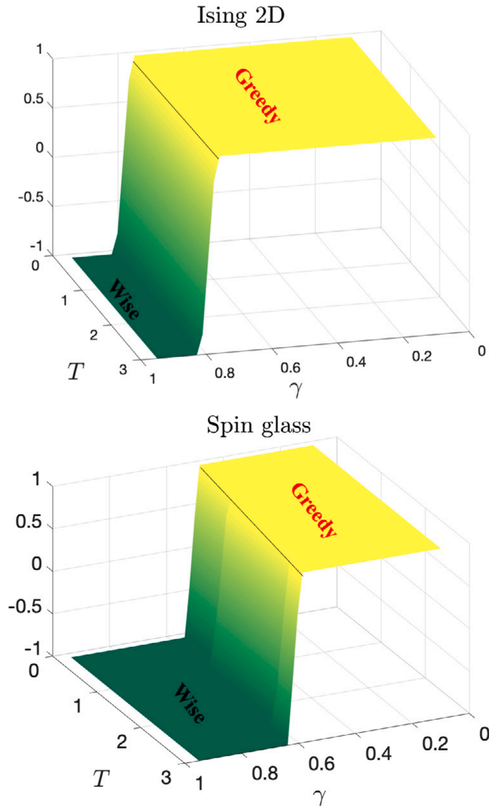

Fig. 3. Change in kobold strategy as a function of temperature and discounting rate. A greedy strategy implies only lowering the energy at every step, while a wise one can also accept intermediate energy increases.

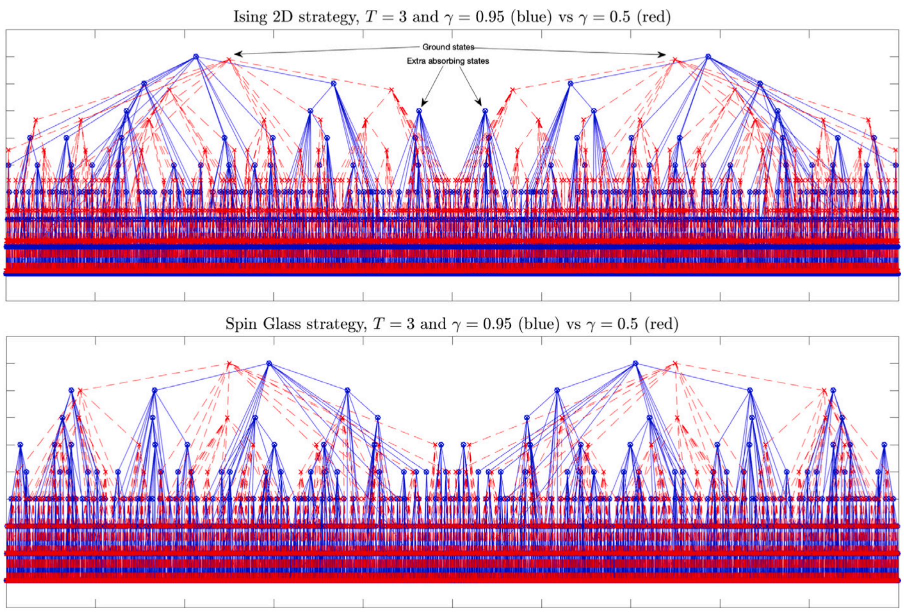

Fig. 4. Kobold strategy tree (top are absorbing states) as a function of the γ for $T = 3$ for the Ising 2D and spin glass cases. Red curves are low discounting, while blue is high discounting rates. We see the emergence of two extra absorbing states on top of the Ising 2D ground states.

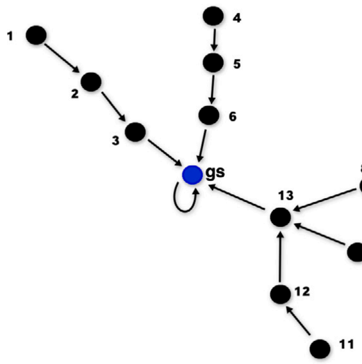

Fig. 5. The graph representation of the optimal strategy π^* is restricted to the basin of attraction of the absorbing state. Since the graph is directed and acyclic, there is a node labeling $l(i)$ such that $l(i) > l(i')$ if i' can be reached by a directed path from i.

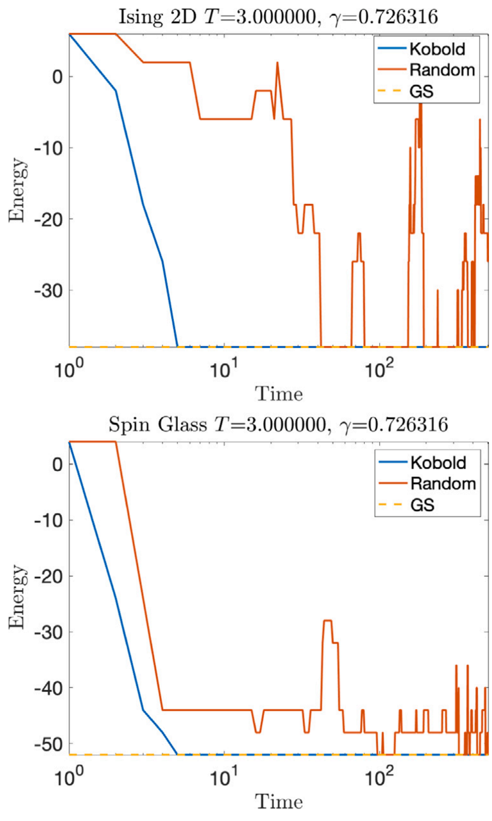

Fig. 6. Example of Markov Chain Monte Carlo for the Ising and Spin Glass examples, for $N = 12$. We compare the dynamics to the Ising ground state of the associated model.

4. Conclusions

The present paper introduced optimal flipping policies for kinetic Ising models, using the stochastic Bellman equation as a prototypical Maxwell demon eager to extract energy from a thermal system. As demonstrated in this paper, these demons have less power than a traditional Maxwell demon, as they can only operate locally on a single spin and must satisfy detailed balance. They thus parallel the typical strategy of a Monte Carlo algorithm, with the key difference being that, instead of thermalizing the model, their intent is to lower the energy given the Markov chain transition probabilities. As shown, their strategy and approach strongly depend on how quickly they want to extract energy from the system, shifting from a greedy to a wise approach depending on whether discounting is high or low, respectively. Unfortunately, the Bellman equation is still plagued by the curse of dimensionality, limiting analysis to small systems. Nonetheless, we have demonstrated that these strategies exhibit interesting changes even in small systems.

This approach can also be interpreted as the optimal sequential strategy to optimize Ising models in an uncertain but time-invariant environment. To overcome the curse of dimensionality, we note that this problem aligns with the goals of reinforcement learning, which is a natural extension of this work [35]. In reinforcement learning, agents operate with “imperfect information" and do not necessarily fully know the state space, analogous to our case where the energy landscape is not fully known. Thus, a natural extension of our work involves "training" a kobold using a reinforcement learning algorithm. The kobold would learn the optimal strategy by testing actions within the phase space and iteratively updating its strategy based on the results. This will be the focus of future work.

CRediT authorship contribution statement

F. Caravelli: Conceptualization, Investigation, Methodology, Writing – original draft.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Francesco Caravelli reports was provided by Los Alamos National Laboratory. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

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

Acknowledgments

This work was carried out under the auspices of the NNSA of the U.S. DoE at LANL under Contract No. DE-AC52-06NA25396, and in particular grant PRD20190195 from the LDRD, United States. This paper was inspired by “Meister Eder und sein Pumuckl”.

References

[1] K. Muruyama, F. Nori, V. Vedral, Rev. Modern Phys. 81 (2009) 1.
[2] G.N. Price, S.T. Bannerman, K. Viering, E. Narevicius, M.G. Raizen, Phys. Rev. Lett. 100 (2008) 093004.
[3] J.V. Koski, V.F. Maisi, T. Sagawa, J.P. Pekola, Phys. Rev. Lett. 113 (2014) 030601.
[4] R. Kosloff, L. Amikam, Annu. Rev. Phys. Chem. 65 (2014) 365.
[5] N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, Phys. Rev. E 85 (2012) 051117.
[6] M. Plesch, O. Dahlsten, J. Goold, Vlatko Vedral, Sci. Rep. 4 (2014) 6995.
[7] J.C. Maxwell, Theory of Heat, Longman, London, 1871.
[8] L. Szilard, Z. Phys. 53 (1929) 840.
[9] E. Lubkin, Internat. J. Theoret. Phys. 26 (1987) 523.
[10] L. Brillouin, J. Appl. Phys. 22 (1951) 334.
[11] R. Landauer, IBM J. Res. Dev. 5 (1961) 183.
[12] C.H. Bennett, Internat. J. Theoret. Phys. 21 (1982) 905.
[13] P.A. Skordos, W. H. Zurek, Am. J. Phys. 60 (1992) 876.
[14] J.D. Bekenstein, Phys. Rev. D 7 (1973) 2333.
[15] J.D. Bekenstein, Phys. Rev. D 9 (1974) 3292.
[16] S.W. Hawking, Nature 248 (1974) 30.
[17] S. Lloyd, Phys. Rev. A 56 (1997) 3374.
[18] M.D. Vidrighin, O. Dahlsten, M. Barbieri, M.S. Kim, V. Vedral, I.A. Walmsley, Phys. Rev. Lett. 116 (2015) 050401.
[19] A. Lande, Phys. Rev. 87 (1952) 267.
[20] A. Peres, Phys. Rev. Lett. 63 (1989) 1114.
[21] W.H. Zurek, Phys. Rev. A 67 (2003) 012320.
[22] L. Del Rio, et al., Nature 474 (2011) 61–63.
[23] M. Horodecki, J. Oppenheim, Nat. Comm. 4 (2013) 2059.
[24] C. Jarzynski, Eur. Phys. J. B 64 (2008) 331340.
[25] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, J. Chem. Phys. 21 (1953) 1087–1092.
[26] W.K. Hastings, Biometrika, Vol. 57, 1970, pp. 97–109.
[27] R.J. Glauber, J. Math. Phys. 4 (1963) 294.
[28] L. Onsager, Phys. Rev. 37 (1931) 405–426, II 38, 2265–2279.
[29] J. Schnakenberg, Rev. Modern Phys. 48 (1976) 571.
[30] M.P. Frank, in: I. Phillips, R. Hafizur (Eds.), RC 2017, in: LNCS, vol. 10301, Springer, Cham, 2017, pp. 19–34.
[31] R.E. Bellman, Proc. Natl. Acad. Sci. USA 38 (8) (1952) 716–719.
[32] R.E. Bellman, Dynamic Programming, Princeton University Press, 2003, Dover paperback edition.
[33] R.S. Sutton, A.G. Barto, Reinforcement Learning: An Introduction, Second edition, MIT University Press, Boston, 2015.
[34] P.L. Krapivsky, S. Redner, E. Ben-Naim, A Kinetic View of Statistical Physics, Cambridge University Press, Boston, 2010, ISBN:9780511780516.
[35] A. Barto, R.S. Sutton, Reinforcement Learning: An Introduction, Adaptive Computation and Machine Learning Series, 2016.