The Ising Model and The Statistical Mechanics of Learning

compiled by D.Gueorguiev 12/30/2024

# The Ising Model

## What the Ising Model is about

The model consists of discrete variables that represent magnetic dipole moments of atomic “spins” that can be in one of two states (+1 or -1). The spins are arranged in a graph, usually a lattice allowing each spin to interact with its neighbors. Neighboring spins that agree have a lower energy than those that disagree; obviously the system attempts to occupy the lowest possible energy state, but heat would disturb this equilibrium, thus creating the possibility of different structural phases. The model allows identification of phase transitions.

The Ising problem without an external field can be equivalently formulated as a graph maximum cut (Max-Cut) problem (ref. [2]) that can be solved via combinatorial optimization.

## More formal discussion

Consider a set of lattice sites, each with a set of adjacent sites (e.g. a graph) forming a -dimensional lattice. For each lattice site there is a discrete variable such that , representing the site’s spin. A *spin configuration*, is an assignment of spin value to each lattice site.

For any two adjacent sites there is an *interaction* . Also a site has an *external magnetic field* interacting with it. The *energy* of a configuration is given by the Hamiltonian function

where the first sum is over pairs of adjacent spins (every pair is counted once). The notation indicates that sites and are nearest neighbors. The magnetic moment is given by .

The configuration probability is given by the Boltzmann distribution with inverse temperature :

where , and the normalization constant

# References

[1] <https://en.wikipedia.org/wiki/Ising_model>

[2] <https://en.wikipedia.org/wiki/Maximum_cut>

[3] <https://en.wikipedia.org/wiki/Partition_function_(statistical_mechanics)>

[4] [Information Theory and Statistical Mechanics, E.T. Jaynes, Department of Physics, Stanford University, 1957](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/articles/Information_theory_and_statistical_mechanics_part1_Jaynes_1957.pdf)

[5] [Information Theory and Statistical Mechanics, E.T. Jaynes, Department of Physics, Stanford University II, 1957](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/articles/Information_theory_and_statistical_mechanics_part2_Jaynes_1957.pdf)

[6] [Exactly Solved Models in Statistical Mechanics, R.J. Baxter, The Australian National University, 1989](https://github.com/dimitarpg13/information_theory_and_statistical_mechanics/blob/main/literature/books/Exactly_Solved_Models_In_Statistical_Mechanics_Baxter_1989.pdf)

[7] [Statistical Mechanics of Deep Learning Neural Networks - The Back-Propagating Kernel Renormalization, Q. Li, H. Sompolinsky, 2021](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Statistical_Mechanics_of_Deep_Linear_Neural_Networks-The_Back-Propagating_Kernel_Renormalization_Li_Sompolinsky_2021.pdf)

[8] [Statistical Mechanics of Learning from Examples, HS Seung, H. Sompolinsky, 1994](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Statistical_mechanics_of_learning_from_examples_1992-seung-somplinsky.pdf)

[9] [Visualizing the Loss Landscape of Neural Nets, H. Li et al, 2018](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Visualizing_the_Loss_Landscape_of_Neural_Nets_Li_2018.pdf)

[10] [Qualitatively Characterizing Neural Network Optimization Problems, Ian Goodfellow et al, 2015](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Qualitatively_Characterizing_Neural_Network_Optimization_Problems_Goodfellow_2015.pdf)

[11] [The Loss Surfaces of Multilayer Networks, A. Choromanska et al, 2015](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/The_Loss_Surfaces_of_Multilayer_Networks_Choromanska_2015.pdf)

[12] [Spin Glasses and the Statistical Mechanics of Protein Folding, JD Bryngelson, PG Wolynes, 1987](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/bryngelson-wolynes-1987-spin-glasses-and-the-statistical-mechanics-of-protein-folding.pdf)

[13] [Mimicking The Folding Pathway to Improve Homology-Free Protein Structure Prediction, J. DeBartolo, 2008](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/debartolo-et-al-2009-mimicking-the-folding-pathway-to-improve-homology-free-protein-structure-prediction.pdf)

[14] [Funnels in Energy Landscapes, K. Klemm et al, 2007](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Funnels_in_Energy_Landscapes_Klemm_2007.pdf)

[15] [Understanding Protein Folding with Energy Landscape Theory, K. Plotkin, JN Onuchic, 2002](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Understanding_protein_folding_with_energy_landscape_theory_PlotkinOnuchic_part1.pdf)

[16] [Landscape Statistics of low autocorrelated binary string problem, F. Ferreira et al, 2000](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Landscape_statistics_of_the_low_autocorrelated_binary_string_problem_Ferreira_2000.pdf)

[17] [Energy Landscapes, Supergraphs, and "Folding Funnels" in Spin Systems, Piotr Garstecki et al, 1999](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/Energy_Landscapes_Supergraphs_and_Folding_Funnels_in_Spin_Systems_Garstecki_1999.pdf)

[18] [From Levinthal to pathways to funnels, K. Dill, HS Chan, 1997](https://github.com/dimitarpg13/deep_learning_and_neural_networks/blob/main/literature/articles/statistical_mechanics_of_DL/From_Levintal_to_pathways_to_funnels_Dill_1997.pdf)

[19] [Towards a new Theory of Learning: Statistical Mechanics of Deep Neural Networks, 2019, Charles H Martin](https://calculatedcontent.com/2019/12/03/towards-a-new-theory-of-learning-statistical-mechanics-of-deep-neural-networks/)

[20] [Why does Deep Learning work?, Charles H Martin, 2015](https://calculatedcontent.com/2015/03/25/why-does-deep-learning-work/)

# Appendix

## Microscopic Definitions of Thermodynamic Concepts

### Ensembles, Microstates, Macrostates, Partition Functions

**Definition** *Statistical* and *Thermodynamic ensembles –*

*Statistical ensemble* is an idealization consisting of a large number of virtual copies (countable set) of a system, considered all at once, each of which represents a possible state that the real system might be in.

*Thermodynamic ensemble* is a specific variety of statistical ensemble that, among other properties, is in statistical equilibrium and is used to study the properties of thermodynamic systems.

#### Physical considerations

The ensemble formalizes the notion that an experimenter repeating an experiment again and again under the same macroscopic conditions, but unable to control the microscopic details, may expect to observe a range of different outcomes. The size of the ensemble can be very large, including every possible microscopic state the system could be in, consistent with its observed macroscopic properties. In many cases it is possible to calculate averages directly over the whole of the thermodynamic ensemble, to obtain explicit formulas for many of the thermodynamic quantities of interest, often in terms of appropriate partition function.

The concept of an *equilibrium* (or *stationary*) *ensemble* implies that although the system certainly evolves over time, the ensemble does not necessarily have to evolve. In fact, the ensemble will not evolve if it contains all past and future phases of the system.

The Thermodynamics is concerned with systems that appear to human perception to be “static” (despite the motion of their internal parts), and which can be described simply by a set of macroscopically observable variables. These systems can be described by statistical ensembles that depend on a few observable parameters, which are in statistical equilibrium.

#### Types of ensembles

*Microcanonical ensemble (NVE ensemble)-*

A statistical ensemble where the total energy of the system and the number of particles in the system are each fixed for particular values; each of the members of the ensemble are required to have the same total energy and particle number. The system must remain totally isolated (unable to exchange energy or particles with its environment) in order to stay in statistical equilibrium

*Canonical ensemble (NVT ensemble)-*

A statistical ensemble where the energy is not known exactly but the number of particles is fixed. In place of the energy, the temperature is specified. The canonical ensemble is appropriate for describing a closed system which is in, or has been in, weak thermal contact with outside body. In order to be in statistical equilibrium, the system must remain totally closed (unable to exchange particles with its environment) and may come into weak thermal contact with other systems that are described by ensembles with the same temperature.

*Grand canonical ensemble* (VT *ensemble*)-

A statistical ensemble where neither the energy nor particle number are fixed. In their place, the temperature and chemical potential are specified. The grand canonical ensemble is appropriate for describing an open system: one which is in, or has been in, weak contact (thermal, chemical, radiative, electrical) with a reservoir. The ensemble remains in statistical equilibrium if the system comes into weak contact with other systems that are described by ensembles with the same temperature and chemical potential.

**Definition** *Microstate* –

A specific configuration of a system that describes the precise positions and momenta of all the individual particles or components that make up the system. Each microstate has a certain probability of occurring during the course of the system’s thermal fluctuations.

**Definition** *Macrostate* -

Macrostate of a system refers to its macroscopic properties, such as its temperature, pressure, volume, and density. More precisely, a microstate is a particular set of values of energy, the number of particles, and the volume of an isolated thermodynamic system. Microstates represent specific possible ways in which the system can achieve a particular microstate. Thus, a microstate is characterized by a probability distribution of possible states across a certain statistical ensemble of all microstates. This distribution describes the probability of finding the system in a certain microstate. In the *thermodynamic limit*, the microstates visited by a macroscopic system during its fluctuations all have the same macroscopic properties.

**Definition** *Thermodynamic (macroscopic) limit* -

The *Thermodynamic (macroscopic) limit* of system is the limit for a large number of particles where the volume is taken to grow in proportion with the number of particles. Thus, the thermodynamic limit is defined as the limit of a system with a large volume, where the particle density is held fixed:

In this limit, macroscopic thermodynamics is valid. This implies that the thermal fluctuations in global quantities are negligible, and all thermodynamic quantities, such as pressure and energy, are simply functions of the thermodynamic variables, such as temperature and density.

The thermodynamic limit is essentially a consequence of the central limit theorem. The internal energy of a gas of molecules is the sum of order contributions, each of which is approximately independent, and so the central limit theorem predicts that the ratio of the size of the fluctuations to the mean is of order . Thus for a macroscopic volume with number of molecules comparable to the Avogadro number those fluctuations are negligible.

Statistical Mechanics links the empirical properties of a system to the statistical distribution of an ensemble of microstates. All macroscopic thermodynamic properties of a system may be calculated from the partition function that sums of all its microstates.

At any moment a system is distributed across an ensemble of microstates, each labeled by , and having a probability of occupation , and an energy .

**Definition** *Internal energy*-

The internal energy of the microstate is the mean over all microstates of the system’s energy

This is a microscopic statement of the notion of energy associated with the first law of thermodynamics.

**Definition** *Entropy-*

For the more general case of the canonical ensemble, the absolute entropy depends exclusively on the probabilities of the microstates and is defined as

where is the Boltzmann constant. For the microcanonical ensemble, consisting of only those microstates equal to the energy of the microstate this simplifies to

with the number of microstates .

#### A Note on Heat and Work

For a closed system (no transfer of matter), *heat* is the energy transfer associated with a disordered, microscopic action on the system, associated with (small) changes in occupation numbers of the quantum energy levels of the system, without change in the values of the energy levels themselves.

Work is the energy transfer associated with an ordered, macroscopic action on the system. If this action acts very slowly, then the adiabatic theorem of quantum mechanics implies that this will not cause jumps between energy levels of the system. In this case, the internal energy of the system only changes due to a change of the system’s energy levels.

The microscopic definitions of heat and work for the quantum case are:

so that

.

#### The microstate in phase space for the classical case

The description of a classical system of DOF may be stated in terms of a dimensional phase space, whose coordinate axes consist of the generalized coordinates of the system, and its generalized momenta . The microstate of such system will be specified by a single point in the phase space. For a system with very large number of DOFs its exact microstate usually is not important. So the phase space can be divided into cells of the size each treated as a microstate. Now the microstates are discrete and countable and the internal energy has no longer an exact value but is between and with .

The number of microstates that a closed system can occupy is proportional to its phase space volume:

where is an indicator function. It is if the Hamilton function at the point in phase space is between and and if not. The constant makes dimensionless. For an ideal gas is .

In this description, the particles are distinguishable. If the position and momentum of two particles are exchanged, the new state will be represented by a different point in phase space. In this case a single point will represent a microstate. If a subset of M particles are indistinguishable from each other, then

#### On Partition Functions

The *Partition function* describes the statistical properties of a system in thermodynamic equilibrium. Partition functions are functions of the thermodynamic state variables such as temperature and volume. Most of the aggregate thermodynamic variables of the system, such as the total energy, free energy, entropy, and pressure, can be expressed in terms of the partition function or its derivatives. The partition function is dimensionless.

Each partition function represents a particular statistical ensemble which in turn corresponds to a particular free energy. The most common statistical ensembles have named partition functions-

*Canonical partition function* : applies to the canonical ensemble, in which the system is allowed to exchange heat with the environment at fixed temperature, volume and number of particles.

*Grand canonical partition function* : applies to grand canonical ensemble , in which the system can exchange both heat and particles with the environment, at fixed temperature, volume and chemical potential.

Classical discrete system

For a canonical ensemble that is classical and discrete, the canonical partition function is defined as

where

* is index for the microstates of the system;
* is the thermodynamic beta coefficient given with where is the Boltzmann constant;
* is the total energy of the system in the respective microstate.

is also known as the Boltzmann factor

Derivation of canonical partition function for classical discrete system

According to the second law of thermodynamics, a system assumes a configuration of maximum entropy at thermodynamic equilibrium. We seek a probability distribution of states that maximizes the discrete Gibbs entropy

subject to two physical constraints:

1. the probabilities of all states add to unity (second axiom of probability):

2. in the canonical ensemble, the system is in thermal equilibrium, so the average energy does not change over time;