Top Ten Terms

That You Need to Know:

Statistical Mechanics for Artificial Intelligence

Alianna J. Maren, Ph.D.
Themesis, Inc.

Top Ten Terms

That You Need to Know:

Statistical Mechanics for Artificial Intelligence

by

Alianna J. Maren, Ph.D.,Themesis, Inc.

Contents

Abstract

The big challenge, for many AI students and researchers, is that ***the important papers in deep learning are written in the language of statistical mechanics***. In order for someone without a strong, graduate-level physics background to read these papers, that person needs to at least know the *vocabulary of statistical mechanics*, also called *statistical physics* or even *statistical thermodyamics*.

This paper functions as a short phrase book for those who are venturing into the realm of energy-based neural networks, which are the networks on which all of ***deep learning architectures and GANs*** (generative adversarial networks) are based. It presents ***ten common statistical mechanics terms***, and orients the reader to their meaning and to how these terms are used in neural networks applications.

We translate *statistical mechanics* terms to common language.

1 Introduction: Why You Need the "Top Ten Terms"

If you are an artificial intelligence (AI) student or practitioner, then you probably already know that the AI is an *integrative discipline*. This means that AI is more than multi-disciplinary. In order to understand the AI fundamentals, you need to master *enough* of certain underlying disciplines so that you can understand how the leading AI systems are constructed and how they work.

The following Figure 1 identifies *seven key papers* in the evolution of *energy-based neural networks*. These papers - and their corresponding neural networks - all work with the same *fundamental statistical mechanics equation*.

This means - if you can master a *very small set of statistical mechanics concepts and equations*, then you can *understand these neural networks*.

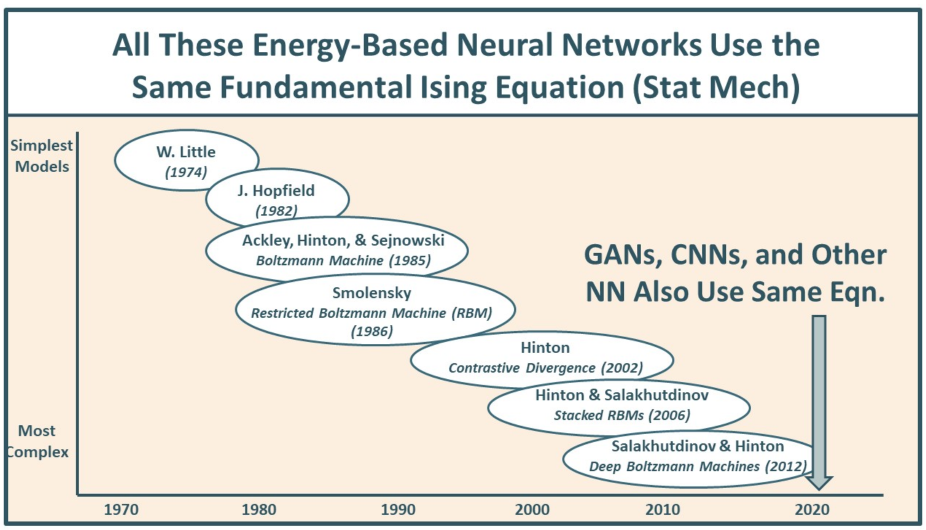


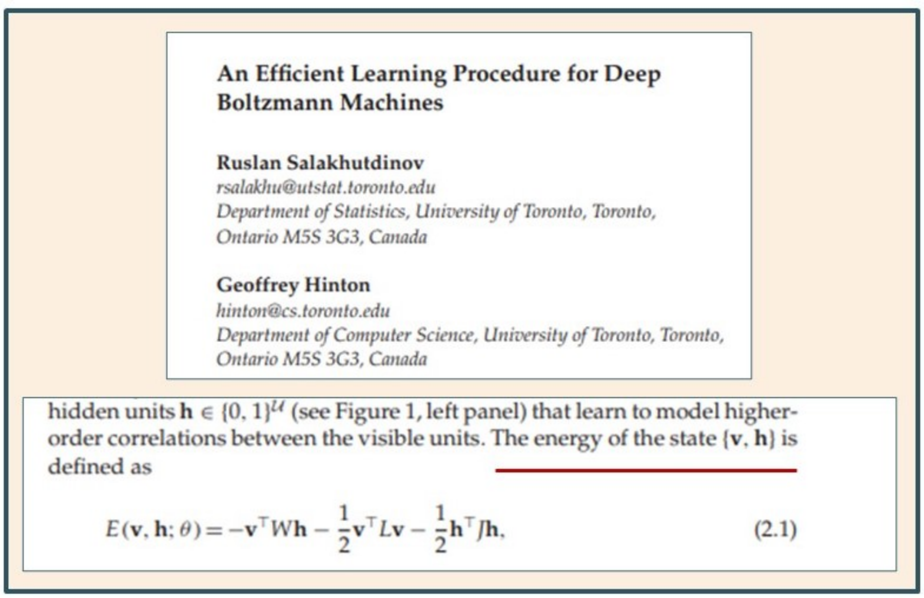
Figure 1: Seven key papers in the evolution of energy-based neural networks.

2 Top Ten Terms that You Need to Know

The classic papers of *energy-based neural networks*, including the Hopfield neural network, the (restricted) Boltzmann machine, and ***all*** deep learning methods use the language of statistical mechanics, also known as statistical physics or even statistical thermodynamics.

This is illustrated in the following paper (Salakhutdinov and Hinton 2012). In this paper, Salakhutdinov and Hinton presented a learning method that enabled very powerful and effective learning in "deep" architectures. These deep architectures used layers of (restricted) Boltzmann machines, and enabled the current massive evolution in AI performance.

Salakhutdinov and Hinton describe the Boltzmann machine in terms of *the energy of the state* of a given set of neurons or nodes. This notion of the *energy of a system* draws on the language and concepts of statistical mechanics.



An Efficient Learning Procedure for Deep Boltzmann Machines

Ruslan Salakhutdinov
rsalakhu@utstat.toronto.edu
Department of Statistics, University of Toronto, Toronto, Ontario M5S 3G3, Canada

Geoffrey Hinton
hinton@cs.toronto.edu
Department of Computer Science, University of Toronto, Toronto, Ontario M5S 3G3, Canada

hidden units $\mathbf{h} \in \{0, 1\}^{\mathcal{U}}$ (see Figure 1, left panel) that learn to model higher-order correlations between the visible units. The energy of the state $\{\mathbf{v}, \mathbf{h}\}$ is defined as

$$E(\mathbf{v}, \mathbf{h}; \theta) = -\mathbf{v}^{\top} W\mathbf{h} - \frac{1}{2}\mathbf{v}^{\top} L\mathbf{v} - \frac{1}{2}\mathbf{h}^{\top} J\mathbf{h}, \qquad (2.1)$$

Figure 2: *Statistical mechanics* terms, such as *energy* or *partition function*, are prevalent in the deep learning literature. An example is this paper by Salakhutdinov and Hinton (2012).

If we can learn just *ten terms from statistical mechanics*, we'll be able to read the important papers in neural networks and deep learning.

3 Term 1: *Thermodynamics*

Thermodynamics is the science behind heat engines. For example, when we say that an engine has a certain efficiency, the notions behind having "efficiency" in an internal combustion engine all come from thermodynamics.

In a somewhat larger scale, thermodynamics forms the theoretical basis behind all power-generation systems that rely on heat production, such as any power production plant that burns fossil fuels.

The *Encyclopedia Brittanica* gives us a useful definition:

> "Thermodynamics is the study of the relations between heat, work, temperature, and energy. The laws of thermodynamics describe how the energy in a system changes and whether the system can perform useful work on its surroundings." (https://www.britannica.com/science/thermodynamics, accessed Feb. 9, 2022.)

Figure 3: A power plant measures its efficiency using the laws of thermodynamics.

Thermodynamics is a ***macroscopic concept.*** In contrast, ***statistical mechanics*** (for our purposes, the same as statistical physics/ thermodynamics), ***works at a very, VERY microscopic scale***.

The true beauty of this is that the very microscopic notions advanced in statistical mechanics (statistical physics/thermodynamics) work out, when the equations are applied to large-scale systems, to yield *exactly the same results* as the macroscopic notions from thermodynamics.

4 Term 2: *Free Energy*

There's nothing "free" about *free energy.* Instead, the notion of free energy refers to the amount of energy that is available to do work - such as drive a piston in an engine.

The free energy notion is common to both regular (macroscopic) thermodynamics and to its microscopic corollary; statistical mechanics (or statistical physics/thermodynamics).

If you read the literature, you might come across terms such as "Gibbs free energy" or "Helmholtz free energy." The distinction between these two is important in macroscopic thermodynamics, and also for some realms of statistical mechanics. Each term refers to a relationship between free energy and changes in the pressure, volume, or temperature of a system - typically considered macroscopic variables.

For our purposes, those distinctions don't matter, because we're ultimately going to use a simple form of a statistical mechanics-based free energy equation as a framework, or shell, into which we can insert neural network concepts. This framework takes the notion of free energy into a new direction - away from pure physics (or physical chemistry) and into a new discipline altogether.

Figure 4: "Free energy" - is not really "free;" it is fundamental in both macroscopic thermodynamics and microscopic statistical thermodynamics.

The free energy function can be expressed in two different ways. They are equivalent, but they look very different.

One way is that the free energy is the difference between two terms; *enthalpy* and *entropy*. (We'll address these in subsequent sections.) The other is as a function of the *partition function*.

The first way of expressing free energy is macroscopic, and is used in classic thermodynamics. The second way is microscopic, or based on statistical mechanics.

The formulation of free energy as a function of enthalpy and entropy is shown in the following figure, taken from a Themesis YouTube on statistical mechanics and neural networks (Maren 2021).

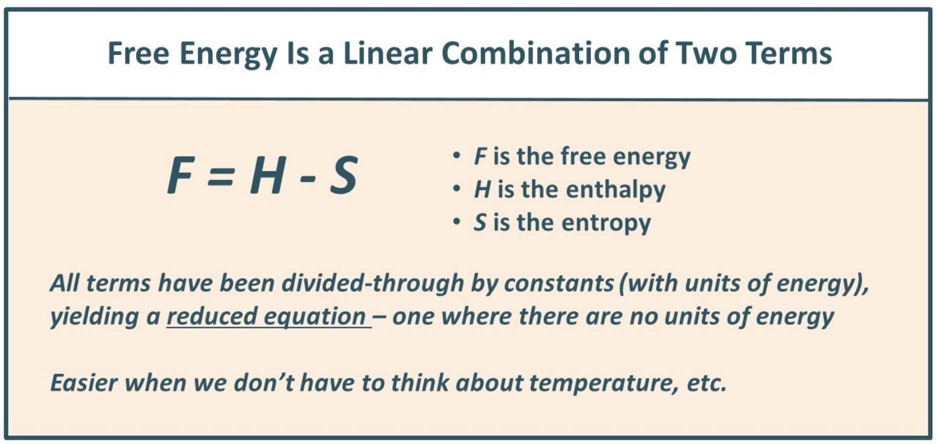


Figure 5: *Free energy* can be expressed as the difference between enthalpy and entropy. This difference is the amount of energy that is "free," or available to do work - such as move a piston in an internal combustion engine.

5 Term 3: *Statistical Mechanics*

For our purposes, statistical physics, statistical mechanics, and statistical thermodynamics *all mean the same thing*.

Statistical mechanics, invented by Ludwig Boltzmann, considers a system in which the only things present are small particles, and we pretend that these particles have no mass and no size, e.g., they are "point particles."

Each particle DOES, however, reside in a specific energy state. ***The properties of the system as a whole are dependent on how many particles are in each different energy state***.

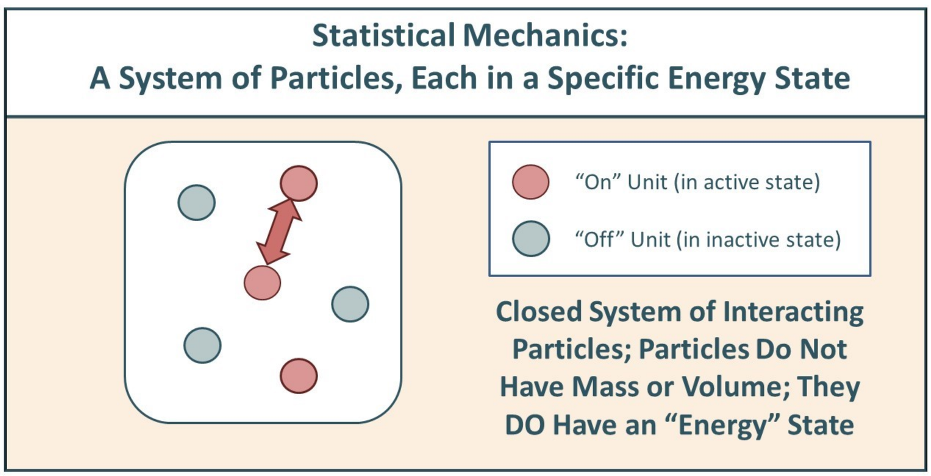


Figure 6: *Statistical mechanics* is based on the behaviors of a large number of interacting particles, each of which is in a distinct *energy state*.

Figure 6 illustrates a very simplified statistical mechanics system. The "particles" shown are blown up in size; in statisical mechanics thinking, they actually take up no volume at all. The system shown has particles in only ***two energy states***, i.e., it is a *bistate system*.

Now, we think about how energy-based neural networks work within a statistical mechanics framework. The important thing is that in (almost all) energy-based neural networks, ***the neural network nodes are also bistate units***. This means that we can ***readily apply the statistical mechanics ideas and methods to energy-based neural networks***.

6 Term 4: *Enthalpy*

Enthalpy is an energy concept. In statistical mechanics, there are (in a simplified sense) *two kinds of enthalpy (energy)*:

- ***Activation enthalpy***, which is the enthalpy (energy) associated with each individual unit in the system; this is directly related to the ***energy state*** in which each unit is residing, and
- ***Interaction enthalpy***, which is the energy associated with the ***interaction between any two particles in the system***.

Both the notions of activation enthalpy and interaction enthalpy are used in energy-based neural networks. The authors of energy-based neural networks papers (e.g., (Hinton and Salakhutdinov 2006)), refer to the combination of these terms as the *energy* of the system, as illustrated in Figure 7.

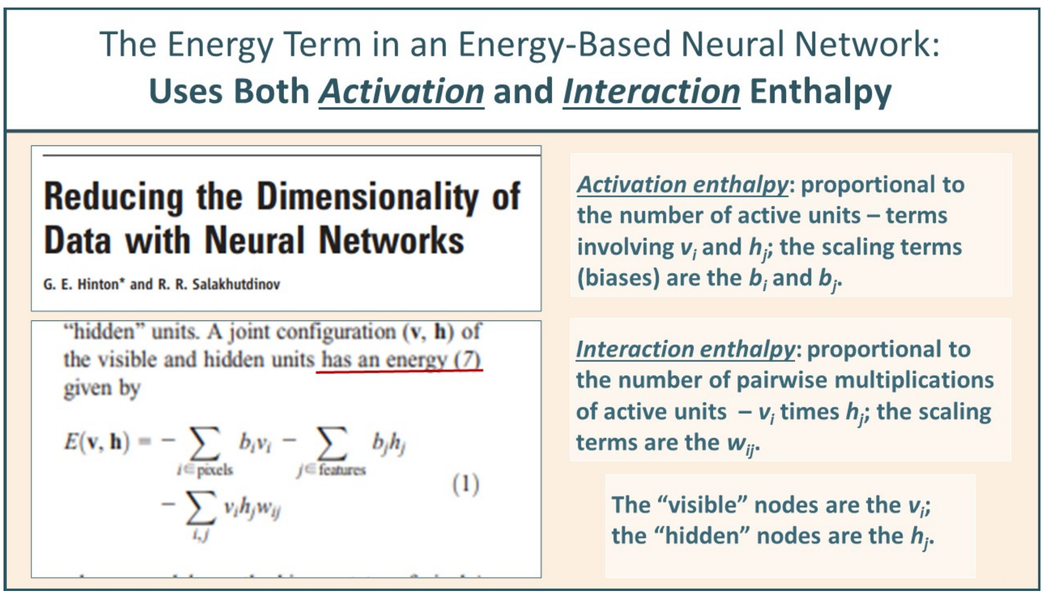


Figure 7: *Statistical mechanics* uses the notion of enthalpy (energy), which includes both ***activation energy*** and ***interaction energy***. These same two types of energy are ***used in all energy-based neural networks***.

The equation used by Geoffrey Hinton and Ruslan Salakhutdinov, extracted in the above Figure 7, is consistent across all uses of Boltzmann machines. Boltzmann machines are essential for deep learning.

7 Term 5: *Entropy*

If you're going to learn more about any of the concepts presented here, ***your best investment would be to focus on entropy***.

The reason is that the ***notion of entropy***, completely apart from statistical mechanics, has taken on a life of its own and ***become the basis for information theory***. From there, it has also become the basis for many practices within the AI and ML community; we won't take the time here to identify all of them.

That said, not many people coming into the AI arena have a good "gut-feel" for what entropy really is.

Figure 8: Entropy - it's more than just "disorder" or "randomness."

Many scientists describe entropy as a measure of the disorder or randomness in a system.

It may be more useful to think about entropy as a ***measure of the distribution of units in a system among all possible (energy) states***.

These two statements are very close; they're really saying the same thing - just from slightly different perspectives.

You may have seen the basic ***entropy equation*** in any one of many different disciplines - from statistical mechanics to information theory.

This equation is given as

The Entropy Equation:

$$S = -\sum_j p_j ln(p_j). \tag{1}$$

The j in this case is the number of allowed energy states, and p_j is the probability of units being in that state; it is really the total fraction of units in that state.

When we have a bistate system, that is, there are only two allowed energy levels, then we have $p_1 + p_2 = 1$. We can let $x = p_1$, and then $1 - x = p_2$.

This gives us

The Entropy Equation:

$$S = -[xln(x) + (1-x)ln(1-x)]. \tag{2}$$

A few minutes of playing around with a calculator or the simplest program will easily reveal that the entropy is at a maximum when $x_1 = x_2 = 0.5$, meaning that the entropy is at a maximum when we have distributed the available units as broadly as possible among the (two) available states.

In fact, if we graph the entropy as a function of x, we'll see that the entropy forms a bowl-shaped (symmetrical) curve, with the bottom of the bowl facing up, and the maximal entropy value is when $x = 0.5$.

The free energy, as we saw in Figure 5, is the enthalpy minus the entropy. Thus, if the enthalpy were zero, then the free energy would simply be that "entropy bowl," turned upside down - so that it shows up as a regular (concave) bowl.

In nature, systems tend towards equilibrium. That is, they tend towards a free energy minimum.

If there were no competing factors, that is, if the enthalpy were zero, then the free energy minimum of a system would always be that “bottom of the bowl” point. We would live in a universe where everything tends to be maximally distributed among all possible states.

The introduction of the enthalpy term, though, shifts the location of the free energy minimum away from the center point of the entropy bowl. It makes other real-world configurations possible.

8 Term 6: *Equilibrium*

There are really only a few powerful *force-combinations* that hold the universe together. Interestingly - it's not the *singular forces* that are important; it's how *two of them - working together* - provide the "glue" that makes the universe work as it should.

Take gravity, for example. We all know about the "Law of Gravity." Babies spend hours experimenting with gravity, dropping their toys on the floor. Toys always fall down; they never fall up.

Yet, the important thing is *not* just gravity. It is gravity working together with one of Newton's "Laws of Motion" - the one that says, "an object in motion remains in motion at constant speed and in a straight line ..."

If gravity were the only important thing, we would expect that the moon would fall into the earth, and the earth would fall into the sun. But, they don't.

Figure 9: The moon stays in orbit around the earth because its orbit balances the *Law of Gravity* and the *First Law of Motion.* Photo courtesy Ken G. Kosada, *Lethal Lens Photography* on Instagram. (See Acknowledgements at the end of this paper.)

Instead, the moon travels in an orbit around the earth, and the earth travels in an orbit around the sun. When a spacecraft or rocket hits a certain point at which it can go into orbit, we say that it has "achieved orbital velocity." That means that, for that particular distance from the earth, the rocket has achieved enough velocity for it to stay in orbit (at that distance) without falling back to earth.

What's important is not just Law of Gravity, or the Law of Motion. It's how the two work together to give us *stable systems*.

The "Principle of Equilibrium" is similar. It is the combination - the *balancing* - of two different things.

Figure 10: Equilibrium - the balance between entropy and enthalpy - when the free energy of the system is at a minimum.

In nature, this "Principle of Equilibrium" says that a system will *tend towards a free energy minimum*. "Free energy" is an important notion - perhaps *the most important notion* in statistical mechanics (or statistical physics/thermodynamics), and in regular thermodynamics as well.

We introduced free energy previously in Section 4, and noted that it could be expressed two different ways; it's a *dimorphic equation*. The version that we'll use here expresses the free energy very simply; free energy is the difference between two terms - enthalpy and entropy. (Recall Figure 4.)

So while we might think of equilibrium as being a balance between two or more things - such as a "work/life balance," for our purposes, the best way to think about equilibrium is that it is *finding the minimum in the free energy.*

This notion of "finding the miniimum" works so well that it not only underlies much of statistical mechanics - and hence, models a lot of nature - but it also works in creating a model for neural network training.

So here we have it. ***This is the essence of all energy-based neural networks***; the Hopfield (Little-Hopfield) neural network, the Boltzmann machine (whether simple or restricted), and all forms of "deep" architectures.

All of these networks work as well as they do because they each are trained to find a free energy minimum. The free energy minimum that is used by these neural networks is *NOT* the same in as is found in models describing nature; there are no "particles" that have mass (or not), etc.:

1. Instead of "particles in a volume," we have nodes in a neural network,
2. Instead of having some sort of interactions between these particles, we have connection weights between the nodes, and
3. Instead of entropy, we have building a training and testing data set.

(We'll discuss these points in the *Bonus Chapters* included at the end of the *Top Ten Terms* short course.

With all of these changes - in applying statistical mechanics to neural networks - we'd think it is surprising (perhaps even shocking) that the free energy notion works as well as it does in neural networks. In fact, for neural networks, "free energy" is almost a poetic metaphor, and less a model of an actual, real, observable and measurable physical system.

However, as a poetic metaphor, the free energy equation functions remarkably well. That's why these neural networks (Hopfield, Boltzmann, and "deep") function as well as they do.

It is just because these neural networks are indeed so effective that we often want to learn about them by reading the original literature - and also, the emerging new work that will keep us up-to-date in our field.

The problem that most of us encounter is that the authors of these very important papers bring in a lot of concepts and terms, and they don't often explain the context. This means that we are often left feeling frustrated, and at a loss, when it comes to reading (or *deciphering*) these papers.

What may help us, when we read these important works, is to realize that the authors of these papers are often physicists. The notion of free energy is as real to them as the ocean is to a fish swimming among the coral reefs.

If a fish was communicating with another fish, it would mention specific things about the coral reef. It would NOT have to remind the other fish, "Oh, by the way, we're swimming in an ocean." That part would be assumed as common knowledge.

Thus, when we read the important, foundational neural networks papers, the authors don't mention free energy. They might, in fact, discuss something that is very much about one *aspect* of free energy - for example, *entropy* (which we discussed in the previous Section 7). But they will assume that this concept is so well understood by their reader that they don't even mention it as a common reference frame.

Figure 11: When a fish communicates with another fish, that fish does not need to reference the ocean - it is already a common reference frame. Similarly, when physicists talk or write about energy-based neural networks, they will often assume that we (the readers) understand statistical mechanics as a common reference frame. Photo courtesy Ken G. Kosada, *Lethal Lens Photography* on Instagram. (See Acknowledgements at the end of this paper.)

Instead, these (physicist) authors will write about things that are important as they attempt to implement the notion of free energy into their neural network. Instead of saying, "Let's talk about entropy," they'll jump right into something that is very *relevant* to implementing entropy in the neural network training - such as Gibbs sampling or Markov chains.

They will assume that their reader understands the context in which these new topics are being mentioned. They do this because, in their minds, they are writing for other physicists.

It is up to us, who are entering the field without this kind of background, to decode their reference frame and figure out the context in which they are bringing in certain topics.

That is why we need this phrasebook covering the *Top Ten Terms* in statistical mechanics. Like a decoder ring, it helps us decode the messages that these researchers were sending to their colleagues. With key phrases under our belt, we can also decode their messages.

9 Term 7: *Microstates*

The notion of *microstates* is the most foundational to statistical mechanics. It is also the most abstract.

The insight into microstates is due to Ludwig Boltzmann, who invented the field of statistical mechanics. His ideas were not accepted right away - despite his ability to show how when we looked at the behavior of *microscopic systems* (those described by statistical mechanics) at a large enough scale, we came to exactly the results of the classical thermodynamics. (This was an amazing intellectual *tour de force* at its time, and still ranks as one of the most significant discoveries of all time.)

The notion of microstates is that we can describe a system in terms of how many units (or particles with zero volume and zero mass) are at distinctly different *energy levels*. Figure 12 shows a system with ten particles or units, and three energy levels. This particular figure illustrates the case where there is one unit (#10) at the highest energy level ($e_j = 2$), two units at the middle energy level ($e_j = 1$), and the remaining seven units are at the lowest energy level ($e_j = 0$). Here, the subscript j indices the three different energy levels. The "energies" are assigned arbitrary values of 0, 1, and 2.

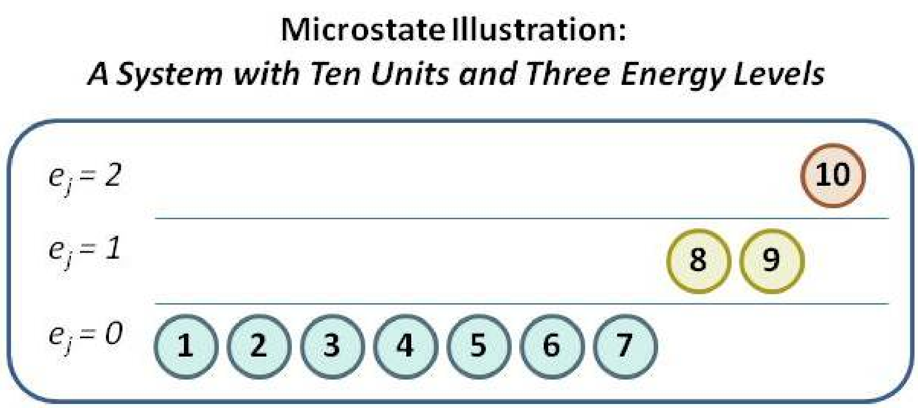


Figure 12: Microstates - the ways in which a system with a specific energy distribution can be configured.

The key notion about microstates is that we can *count the number of microstates* associated with a particular distribution of units into energy states. For example, in the energy state-distribution shown in Figure 12, we can create a total of ten different ways of putting a unit into the top-most energy state. (This is because we have ten different units, and so there are ten ways to do this.) Then, we have $9 * 8$ distinct ways of putting two of the remaining units into the middle-most energy state. OF course, we then have to divide by two - because it doesn't matter *where* we put each of those units; we can interchange them and still have the same two units at that energy level.

In short, it doesn't matter what *order* we arrange the units in at a particular energy level. It just matters which units are where.

Imagine going to a party at a house where there are three stories, and ten guests. If only one person is allowed onto the top floor at a time, and only two persons are allowed onto the middle floor, then we have a similar situation. The host doesn't care where the guests are on a particular floor; for example, guests #8 and #9 can move about and change places on the middle floor. As far as the host is concerned, that's still just one of the potentially allowable patterns for guest-configurations. However, if guest #1 changes places with guest #8 on the second floor, that is a new configuration. Or, if guest #8 changes places with guest #10, so that guest #8 is now on the top floor, and guest #10 is on the middle floor, that is also a new configuration.

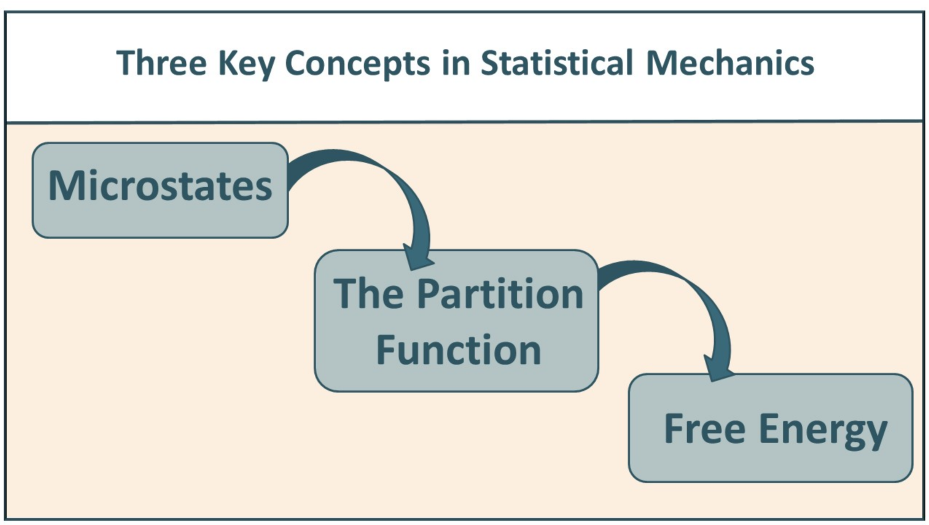


Figure 13: Three key concepts in statistical mechanics will get you established well enough to read - and understand - the important papers in neural networks and deep learning.

When we allow progressively more units at the higher energy levels (or more guests are allowed to the upper floors at one time), the total number of possible arrangements increases.

If we did these calculations, we'd find that we get the maximal number of possible arrangements when we allow an equal distribution of units-per-level. (In our case, we'd have three units on one level, three on another, and four on the third - it wouldn't matter how we allotted those limits.)

However we did this, we'd get the *maximal number of arrangements* when we allow the units to spread out, as much as possible, between all energy levels.

This "spreading out between levels" leads us directly to the notion of *entropy*, which we discussed previously.

To be more precise, the notion of *microstates* leads us to something called the *partition function*, which leads directly to *entropy*. This is shown in Figure 13.

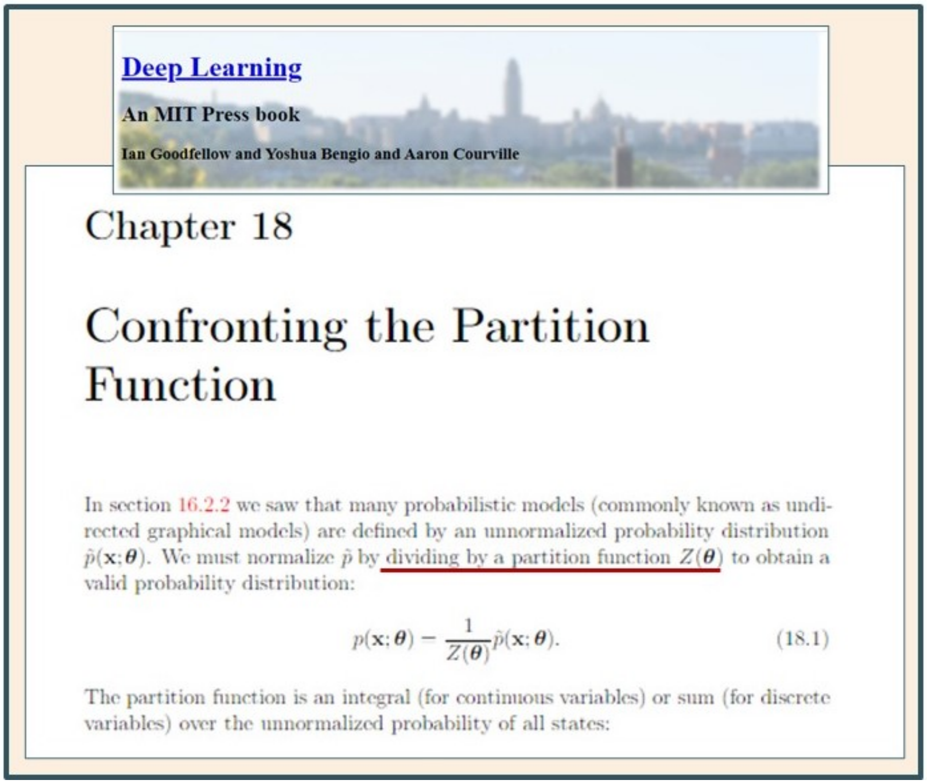
The bottom line is that the notion of *microstates* is foundational to statistical mechanics. However, we almost never see this term mentioned in the AI or machine learning (ML) or deep learning or variational methods literature. This is because when we start using statistical mechanics as a model in AI/ML, we use it in a metaphorical sense. In energy-based neural networks, for example, we have nodes that can be "on" or "off," and we adapt the connection weights between certain kinds of nodes. However, we don't typically do a microstates-type calculation using those nodes.

Instead, when we have to think about the entropy for an energy-based neural network, we think about creating a well-balanced data set, instead of putting nodes into different energy states. It leads to the same place (when done right); it's just very different ways of conceptualizing the notion of "arranging nodes in the allowable energy states."

We *DO* occasionally see the term "partition function" in the literature. Therefore, we discuss that in the next section.

10 Term 8: *Partition Function*

The *partition function* crops up in many books and papers on deep learning.



Deep Learning

An MIT Press book

Ian Goodfellow and Yoshua Bengio and Aaron Courville

Chapter 18

Confronting the Partition Function

In section 16.2.2 we saw that many probabilistic models (commonly known as undirected graphical models) are defined by an unnormalized probability distribution $\tilde{p}(\mathbf{x};\boldsymbol{\theta})$. We must normalize $\tilde{p}$ by dividing by a partition function $Z(\boldsymbol{\theta})$ to obtain a valid probability distribution:

$$p(\mathbf{x};\boldsymbol{\theta}) = \frac{1}{Z(\boldsymbol{\theta})}\tilde{p}(\mathbf{x};\boldsymbol{\theta}). \tag{18.1}$$

The partition function is an integral (for continuous variables) or sum (for discrete variables) over the unnormalized probability of all states:

Figure 14: Extract from *Deep Learning* by Ian Goodfellow, Yoshua Bengio, and Aaron Courville. Published in 2016 by The MIT Press, Cambridge, MA.

The partition function, Z, for the German word *zusammanfügen* (literally “put together’), is a summation over all the possible microstates, j, that a system can find itself in, and is given as

The Partition Function:

$$Z = \sum_j \exp(-\beta E_j). \tag{3}$$

The very weird, strange, and sort-of-sneaky thing about Eqn. 3 is the summation index j is *not* over the total number of particles in the system, and it is *not* over the total number of energy states.

Instead, the summation is over the *total number of microstates*.

This was the radical, explosively-innovative notion that Ludwig Boltzmann conceptualized and published in a series of articles in the 1870's.

Boltzmann's ideas were ridiculed by other physicists. His insights were not validated until shortly before the 1900's, when new discoveries in atomic physics gave support to his work.

Even today, statistical mechanics is still regarded as one of the most intellectually challenging fields. If we can "wrap our heads" around the concepts of microstates and the partition function, we have made substantial gains - and ascended into very lofty elevations of abstract conceptualization!

However, to *understand the AI/ML literature*, we typically *do not need* to understand the notions of *microstates* and the *partition function* in depth. (It does help, though, to be familiar with how researchers formulate the *entropy equation*.)

What is most important is to remember that, in AI and ML, statistical physics is used as a model. It's not to be taken literally. Instead, it's a conceptual framework that - when applied to energy-based neural networks, or variational inference, or other tasks - gives us very good results.

Thus, statistical mechanics validates itself (for our purposes) by being very useful. It is a *pragmatically useful tool*.

For what we need to do, the terms that we've covered so far get us very far in terms of using statistical mechanics as a model.

Most of us, even if we're studying the most subtle and arcane papers in AI and ML, *do not* need a full graduate-level course in statistical mechanics in order to read (and understand) the AI/ML papers. We just need to identify how the AI/ML work is positioned within the statistical mechanics reference frame.

In the two remaining sections, we take on just two more terms. Both of them have to do with how statistical mechanics is specifically used in application to AI/ML. These two terms are:

- ***The Ising equation***, which is how we write the free energy equation (enthalpy minus entropy) in specific terms, and
- ***Interaction enthalpy***, which is the energy associated with the ***interaction between any two particles in the system***.

11 Term 9: *Ising Equation*

The Ising equation, or Ising model, is one of the most common free energy equations in statistical mechanics. It has two major parts; one dealing with the *enthalpy* term, and the other dealing with *entropy*. We've discussed each of these two terms, in Sections 6 and 7 respectively.

The important thing about the Ising equation is that is is the *simplest possible* useful free energy model from statistical mechanics. It's also the basis for all *energy-based neural networks*; this includes the Little-Hopfield neural network, the Boltzmann machine (both original and restricted), and all "deep" architectures, as well as GANs (generative adversarial networks).

In short, the Ising equation is a profoundly useful tool - both in statistical mechanics and in energy-based neural networks.

The following Figure 15 is taken from a Themesis YouTube (Maren 2021).

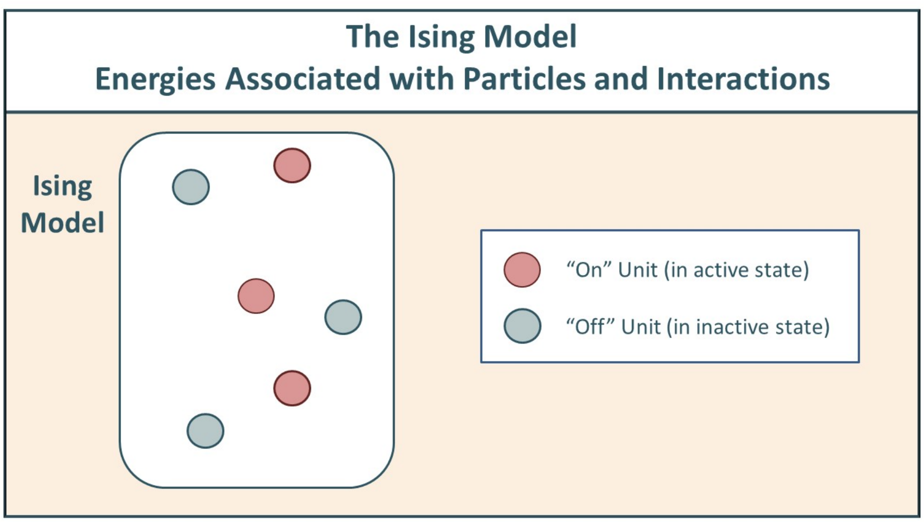


Figure 15: The Ising equation is a statistical mechanics equation that expresses the free energy of a system in terms of the *enthalpy* (energies) of individual particles, together with the overall system *entropy*.

Three key points help us understand why the Ising equation is so useful:

1. ***Bistate system*** - there are only ***two energy states*** allowed for the units in the Ising model, so we can think of these units as being either "on" or "off."

2. ***Interaction enthalpy can be easily modeled*** - there are several different ways in which the *interaction enthalpy* can be expressed in the Ising equation; we'll discuss these more in the next section as our final "term" in this study of *Top Ten Terms.*

3. ***Powerful model*** - from the statistical mechanics perspective, even though the Ising equation is simple, it can be used to model a great deal of naturally-occuring phenomena. From the neural networks perspective, the Ising equation has made possible an entire class of neural networks, and is the foundation for all manner of "deep" architectures.

When we limit the system being modeled by the Ising equation to having just two allowable energy states, then the mathematics is *much easier.*

This is why, in almost all energy-based neural networks (as well as a great deal of the statistical mechanics literature), we see that the Ising model is applied to a system where there are only two allowable energy states. These translate to specific nodes (or neurons) being either "on" or "off."

We will continue this discussion, making a stronger connection on how the Ising model is used in an energy-based neural network, in the *Bonus Chapters* that are provided in the last week of the *Top Ten Terms* short course.

12 Term 10: *Energy Function: The Interaction Energy*

In classic statistical mechanics, the interactions between nodes are just a single scalar value. This is shown in the following Figure 16 (taken from a Themesis YouTube presented by A.J. Maren (Maren 2021)).

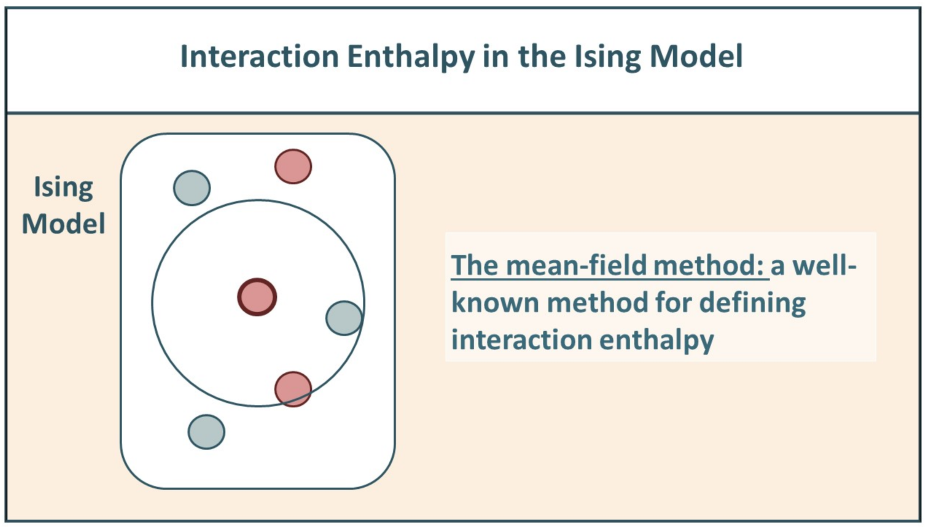


Figure 16: The *interaction energy (enthalpy)* in a typical Ising equation usually represents pairwise interactions, and will (in many Ising formulations) be a function of how many other particles are within a certain radius of the particle for which you're computing the interaction enthalpy.

The breakthrough notion for neural networks - both in the backpropagation method for Multilayer Perceptrons (MLPs) and for Boltzmann machines (and all forms of deep learning architectures) is that the neural network "learns" an individual and specific interaction energy between each different pairwise combination of "visible" and "hidden" nodes.

When a MLP neural network uses backpropagation for learning, it minimizes a squared error function across all the output nodes. When an energy-based neural network (e.g., a Boltzmann machine) does learning, it minimizes an energy function.

In a simple Boltzmann machine, there are connections between all nodes; *visible-to-hidden*, *visible-to-visible*, and *hidden-to-hidden*.

In a restricted Boltzmann machine, the connections are *restricted*; they are between visible-to-hidden nodes only. (And have the same values going the other way; hidden-to-visible.)

The energy function used in Boltzmann machines is very similar to the Ising energy equation in statistical mechanics. The big difference is that, in statistical mechanics, there is a single value for the interaction energy between nodes (or "units") in the system. In an energy-based neural network (e.g., a Boltzmann machine), each connection between two nodes - that is, the interaction energy for each node-to-node connection - is determined uniquely for that network.

We will go into the relation between the Ising equation and the Boltzmann machine (and all energy-based neural networks) in the accompanying *Bonus Chapters*.

For now, we complete our study of statistical mechanics terms by looking at how the interaction energy is formulated for a simple Ising equation.

If we refer to Figure 16, we see that there is a big, red circle drawn around an "on" node (shown in the near-center of the figure). The total *interaction enthalpy* associated with this "on" node depends on the fraction of other "on" nodes within a radius around this central "on" node. (This is a very simplified explanation; there are other, much more complex and elaborate ways to compute interaction enthalpy.)

Still referring to Figure 16, we get the total interaction enthalpy for a given active node (a scalar term) by multiplying the interaction enthalpy-per-active-node by the total fraction of active nodes. (This scales the overall fraction of active nodes to those that can be found within the specified radius around that individual active node.) If we use x to refer to the fraction of nodes that are active, this gives us the interaction enthalpy per active node as a function linear in x.

Then, to get the total interaction enthalpy for the system, we multiply that value by the fraction of nodes that are active across the system. This gives us a total interaction enthalpy that is proportional to x^2.

When we get to the *Bonus Chapters*, we'll learn how the energy of the node-to-node connections is similar to that of the interaction energy in the Ising equation. The real difference is that in the Ising model, there is just one value for a node-to-node interaction energy, and in a Boltzmann machine (or any related neural network), the interaction energy is computed specifically for each different connection.

Once we study this correspondence, we'll realize that *all* energy-based neural networks share this fundamental reliance on the same Ising model. This means that the Little-Hopfield neural network, the Boltzmann machine (both simple and restricted), and then all forms of deep learning and GANs are based on the same fundamental statistical mechanics metaphor.

After we've studied the *Bonus Chapters*, we will know enough to go to the great classics of energy-based neural networks, such as the articles identified in Figure 1, and understand them with much greater insight.

13 Summary and Next Steps

In studying this short article, you've gained a useful and solid understanding of the *top ten terms* used in statistical mechanics (also known as statistical physics, and sometimes as statistical thermodynamics).

You are now in a position to apply what you've learned to understanding the classic works in *energy-based neural networks.*

This next step will take you beyond a simple understanding of the *top ten terms.*

Actually, there are ***three components to your next step***:

1. ***First pass - identify what terms you now understand*** - take a paper that you may have previously attempted to read, and read it again - at least the portion where you were figuratively "blown out of the water" earlier; this time, identify the terms that you ***now understand***; this means that (with sufficient attention) you can now work through that paper,
2. ***Second pass - reading for clarity and understanding*** - go through the same paper (more realistically; the specific section of that paper), and this time, work through the details; at this point, you'll want to correlate the equations (which you now understand) with the neural network structure and operation; in short, ***make the important connections***, and then
3. ***Extend your understanding across multiple works*** - there are many important papers, and you'll want to read more than one. Each time, the equations will seem to be subtly different. The question that you'll want to answer is: "Is this the ***same neural network***, or is it ***just different enough*** so that you can ***confidently discern*** whether or not they are the same - or if you need to understand a different neural network.

You've just taken the ***first and most important step*** - you've accessed this PDF by enrolling in the Themesis short course in the *Top Ten Terms in Statistical Mechanics.*

Your ***next necessary task*** is to get to the end of the course sequence, where you'll have a link to a ***Bonus PDF***. These *Bonus Chapters* will get you started on these three steps that we just identified.

This bonus material will give you an overview. ***It will get you started.***

Then, ***follow through***.

Look for your follow-on Themesis emails. Look for opportunities to ***take the next short course*** in this sequnce. Identify a few more papers that you'd like to have under your belt. (The papers identified in Figure 1 are some good starting points.)

You're now in a different place in your mastery of neural networks fundamentals. You can phrase a different set of questions. You can have different objectives in your next round of study. This puts you in a more powerful position as you progress in mastery.

You are now more mentally prepared to investigate in greater depth, with more penetrating and insightful questions.

As always, ***Themesis welcomes your input and feedback***.

Please feel free to email us at themesis1@themesis.com

We look forward to hearing from you, and to connecting with you soon in the bonus materials that you already can access, and in the next short course.

Very best wishes from all of us at Themesis, Inc.

Alianna J. Maren, Ph.D.
Founder and Chief Scientist

Acknowledgments

Cover photography by Ken G. Kosada, extreme nature and adventure photographer/videographer. See Ken G.'s Instagram at *Lethal Lens Photo* (lethallensphoto at https://www.instagram.com/lethallensphoto/).

References

Hinton, Geofrrey E., and Ruslan R. Salakhutdinov. 2006. "Reducing the Dimensionality of Data with Neural Networks." *Science* 313 (July): 504–507. Accessed on Feb. 12, 2022. https://www.cs.toronto.edu/ hinton/science.pdf.

Maren, Alianna J. 2021, August. Statistical Physics Underlying Energy-Based Neural Networks. Accessed on Feb. 11, 2022. https://www.youtube.com/watch?v=ZazyMS-IDg8&t=251s.

Salakhutdinov, Ruslan R., and Geofrrey E. Hinton. 2012. "An Efficient Learning Procedure for Deep Boltzmann Machines." *Neural Computation* 24 (8): 1967–2006 (August). Accessed on Jan. 5, 2022. https://www.cs.cmu.edu/r̃salakhu/papers/neco_DBM.pdf.

Related YouTubes

All YouTube videos referenced here are on the Themesis, Inc. YouTube channel.

- Maren, Alianna J. 2021. "Statistical Physics Underlying Neural Networks." Themesis YouTube Channel (August 26, 2021). (Accessed August 31, 2022; https://www.youtube.com/watch?v=ZazyMS-IDg8&t=322s.)