

MATH327: More Is Different

Statistical Mechanics, Thermodynamics, and All That

David Schaich

Spring Term 2025

LECTURE NOTES

Last modified 6 May 2025

Module information and logistics	2
Unit 1: Central limit theorem and diffusion	8
Unit 2: Micro-canonical ensemble	26
Unit 3: Canonical ensemble	40
Unit 4: Ideal gases	56
Unit 5: Thermodynamic cycles	68
Unit 6: Grand-canonical ensemble	80
Unit 7: Quantum statistics	90
Unit 8: Quantum gases	101
Unit 9: Interacting systems	127
Unit 10: Synthesis and broader applications	151

Module information and logistics

Coordinator

David Schaich
Theoretical Physics Wing Room 123
david.schaich@liverpool.ac.uk
calendly.com/daschaich
www.davidschaich.net

Overview

“[More Is Different](#)” is the title of a famous 1972 essay by [Philip Anderson](#), which established the concept of emergent phenomena — the idea that large, complex physical systems generally can’t be understood by extrapolating the properties of small, simple systems. For example, consider the $\sim 10^{22}$ H₂O molecules in a cubic centimetre of water. While we can use Newton’s laws (or the laws of quantum mechanics) to analyse a few of these molecules, this does not allow us to predict processes such as phase transitions of this water into steam or ice.

Instead, we have to apply the stochastic (i.e., probabilistic) techniques of *statistical mechanics* — one of the central pillars of modern physics, along with quantum mechanics and relativity. While statistical mechanics was originally developed in the context of thermodynamics in the nineteenth century, it is more generally applicable to any large-scale (macroscopic) behaviour that emerges from the microscopic dynamics of *many* underlying objects. It is intimately connected to quantum field theory, and has been applied to topics from nuclear physics and cosmology to climate science and biophysics, often with outstanding success (recently recognized by the [2021](#) and [2024](#) Nobel Prizes in Physics, to name just two).

The module outline on the previous page is organized around the concept of *statistical ensembles* introduced in the early 1900s. In essence, a statistical ensemble is a mathematical framework for concisely describing the properties of idealized physical systems subject to certain constraints. After studying the probability foundations underlying these frameworks, we meet the *micro-canonical ensemble* in unit 2 and the *canonical ensemble* in unit 3. The following two units 4–5 apply the canonical ensemble to investigate non-interacting (“ideal”) gases and thermodynamic cycles. Unit 6 introduces a third statistical ensemble, the *grand-canonical ensemble*, which units 7–8 apply to several types of non-interacting quantum gases. (No prior exposure to quantum mechanics is required — see below for more information.) Finally, in unit 9 we begin to explore the effects of interactions, which open up a much broader landscape of applications that we will survey for the remainder of the term.

Schedule

Most weeks we will have lectures at 13:00–14:00 on Monday and 11:00–13:00 on Wednesday, with tutorials at 10:00–11:00 on Thursday, all in Room 210. The tutorials

in weeks 2, 3 and 4 (on 6, 13 and 20 February) will be computer lab sessions in Hub 502 PC Teaching Centre B, to provide opportunities for you to work on the computer assignment summarized below.

I will use Panopto to record lectures (and lecturey bits of tutorials and computer labs). Although Panopto can produce a live webcast as it records, it does not do this very well. Therefore I will use Zoom to provide a way for anyone off campus or out of town to connect remotely through [this link](#) (meeting ID 992 7015 4627, passcode Math327!). Since Panopto and Zoom sometimes fight over the microphone and camera, I encourage you to attend in person if you are able. The Panopto recordings will appear along with all other resources on [our Canvas site](#), canvas.liverpool.ac.uk/courses/76365

Office hours will take place at 14:00 on Mondays and 11:00 on Thursdays, after the corresponding lecture and tutorial. They will be held in Room 123 of the Theoretical Physics Wing, and will also connect to the [Zoom link](#) above. If these times do not work with your schedule, you can also make an appointment through calendly.com/daschaich, use the [Canvas discussion board](#) (where anonymous posting is enabled), or reach me by email at david.schaich@liverpool.ac.uk. I will aim to respond to emails and discussion board queries within 48 hours.

Assessment and academic integrity

There will be three in-term assignments. Each accounts for 15% of the module mark, with the remaining 55% coming from the final exam. Although the deadlines listed below are not ideal, they have been centrally coordinated within the Department to minimize pile-up across different modules. The Department also sets each deadline to be at 17:00.

15% A computer assignment due **Friday, 21 February**

30% Two equally weighted homework assignments, the first due **Friday, 7 March** and the second due **Friday, 4 April**

55% A two-hour in-person final examination to be centrally scheduled within the May exam period

According to the University's [Code of Practice on Assessment](#) (COPA), late submissions completed within 120 hours after the submission deadline will have 5% of the total marks deducted each 24-hour period after the deadline. Submissions more than 120 hours late will be awarded zero marks, though I will still endeavour to provide feedback on them. I will aim to return feedback and share model solutions within two or three weeks of the deadline for the homeworks or computer assignment, respectively.

For your other modules you already should have read and understood the Department's current [academic integrity guidance](#) as well as the [Academic Integrity Policy](#) detailed in COPA Appendix L. If you have any questions about what is or is not acceptable, please ask me or our Academic Integrity Officer Alena Haddley. In all cases, the spirit of the Academic Integrity Policy should take precedence over legalistic convolutions of the text.

In particular, I encourage you to discuss the in-term assignments with each other, since discussing and debating your work is a very effective way to learn. Note that I say *your work* — your submissions for all assignments must be your own work representing your own understanding, and the examination must be done on your own. It is unacceptable to copy solutions in part or in whole from other students (current or prior) or from other sources (commercial or otherwise). Should you make use of resources beyond the module materials — including generative AI tools such as ChatGPT — these must be explicitly referenced in your submissions. Clear and neat presentations of your workings and the logic behind them will contribute to your mark.

Main resources and materials

The main materials we will use are the lecture notes you are currently reading. As you read further, you will encounter gaps in the notes, which provide bite-sized exercises to help you check your understanding. While we will fill most gaps during lectures, I encourage you to use them as opportunities to practice.

The ten units into which the content is organized won't neatly match up with the twelve weeks of the term. Some units will require more time than others. Regular Canvas announcements will summarize what we cover each week.

We will use 'natural units' in which the Boltzmann constant $k = 1$, and logarithms have base e unless otherwise specified (i.e., $\log x = \ln x$). There is no need to memorize any equations. Many equations are numbered so that they can be referenced later on, not necessarily because they are important. Key results, definitions and concepts are highlighted by coloured boxes, and you should aim to be confident in your understanding of these.

These lecture notes were first written 'live' during the 2021 and 2022 editions of this module. While they are now much more stable, they continue to be improved, refined and sometimes corrected. The "Last modified" date at the bottom of each page will flag any changes that occur during the term. You can track the changes themselves through the version control repository at github.com/daschaich/MATH327_2025. The [Software Carpentry](#) project provides an introduction to [Version Control with Git](#).

Expected background

No prior exposure to quantum mechanics or computer programming is required — all necessary information on these topics will be provided. I do anticipate that you have previously seen the [standard deviation](#), the [binomial coefficient](#)

$$\binom{N}{k} = \frac{N!}{k!(N-k)!} = \binom{N}{N-k}$$

that counts the number of possible ways to choose k objects out of a set of $N \geq k$ total objects, and [gaussian integrals](#),

$$\int_{-\infty}^{\infty} e^{-a(x+b)^2} dx = \sqrt{\frac{\pi}{a}} \quad a > 0.$$

Programming

You are welcome to complete the computer assignment using the programming language of your choice. I recommend [Python](#), which is free, user-friendly, and very widely used around the world. During the first two weeks of the term we will review [this demo](#) that explains all the Python programming tools you'll need. Python is available on University computers and should work on personal computers. You can also write and run Python code using many cloud services, of which I have had the best experiences with Google's [Colab](#) and [CoCalc](#).¹ You may need to create a free account, and you should make sure to save a local copy to reduce the risk of losing your work. Alternative languages could include [C](#), [Fortran](#), [R](#), or even [MATLAB](#) (through the University's site license). I advise against using Maple, which may struggle to handle parts of the assignment.

How to get the most out of this module

At this point in your studies, this advice should be familiar, but it's worth repeating.

Come to class — ideally in person, if necessary by Zoom. This will ensure regular contact with the material, and help you check that you understand it. If the module is moving slower than you'd prefer, coming to class will give you opportunities to ask about more interesting extensions, applications or complications.

Before class, take a quick look at the upcoming pages in the lecture notes, and think about how any gaps could be filled. Look for the big ideas rather than digging in to every detail, and see if you have any questions (or objections) to raise in class. After class, take a closer look at the details, and make sure the gaps have been filled to your satisfaction. Even though the lecture notes reflect my plans for the module, they may not exactly match what happens in class, especially when questions arise. We may gloss over some topics that are explained clearly in the notes, and we may delve deeper into other topics that merit further consideration.

Work on the homework problems, computer assignment and tutorial exercises. The best way to learn mathematics is by doing mathematics, and these assignments and activities are designed to make you think and help develop your mathematical muscles. In particular, the homework problems will be harder than exam questions, since you'll have much more time to work on them — so make sure you start thinking about them well in advance of the deadline. Afterwards, review the model solutions and feedback, to make sure any confusing points are resolved.

Ask questions. Ask questions you think you're supposed to know the answer to. Ask questions you think everyone else knows the answer to. (They don't.) Ask questions about the big ideas, the specific details, and the connections between them. The opportunity to ask questions is the main benefit of taking a module. You can ask me; you can ask your classmates; you can ask the additional resources below.

¹I have had worse experiences with [replit.com](#), [onlinegdb.com](#), [mybinder.org](#) and [trinket.io](#) — use these at your own risk.

Additional resources

The optional additional resources listed below may be helpful. You can use the module [Reading List](#) on Canvas to reach our library's records for the books.

Resources at roughly the level of this module:

1. David Tong, [Lectures on Statistical Physics](#) (2012), www.damtp.cam.ac.uk/user/tong/statphys.html
2. MIT OpenCourseWare for undergraduate [Statistical Physics I](#) (2013) and [Statistical Physics II](#) (2005), ocw.mit.edu/courses/8-044-statistical-physics-i-spring-2013/ ocw.mit.edu/courses/8-08-statistical-physics-ii-spring-2005/
3. Daniel V. Schroeder, *An Introduction to Thermal Physics* (2021)
4. J. Allday and S. Hands, *Introduction to Entropy: The Way of the World* (2024)
5. C. Kittel and H. Kroemer, *Thermal Physics* (1980)
6. F. Reif, *Fundamentals of Statistical and Thermal Physics* (1965)

More advanced and more specialized resources, which may be useful to consult concerning specific questions or topics:

7. MIT OpenCourseWare for postgraduate [Statistical Mechanics I](#) (2013) and [Statistical Mechanics II](#) (2014), ocw.mit.edu/courses/8-333-statistical-mechanics-i-statistical-mechanics-of-particles-fall-2013/ ocw.mit.edu/courses/8-334-statistical-mechanics-ii-statistical-physics-of-fields-spring-2014/
8. R. K. Pathria and P. D. Beale, *Statistical Mechanics* (2021)
9. Sidney Redner, *A Guide to First-Passage Processes* (2001)
10. Pavel L. Krapivsky, Sidney Redner and Eli Ben-Naim, *A Kinetic View of Statistical Physics* (2010)
11. Kerson Huang, *Statistical Mechanics* (1987)
12. Andreas Wipf, *Statistical Approach to Quantum Field Theory* (2013)
13. Weinan E, Tiejun Li and Eric Vanden-Eijnden, *Applied Stochastic Analysis* (2019)
14. Michael Plischke & Birger Bergersen, *Equilibrium Statistical Physics* (2006)
15. Sacha Friedli and Yvan Velenik, *Statistical Mechanics of Lattice Systems* (2018)
16. L. D. Landau and E. M. Lifshitz, *Statistical Physics, Part 1* (1969)

A general book about learning, emphasizing (among other things) the value of retrieval practice compared to re-reading lecture notes or re-watching videos:

17. Peter C. Brown, Henry L. Roediger III and Mark A. McDaniel, *Make it Stick: The Science of Successful Learning* (2014)
A [short summary video](#) is also available

Programming resources:

18. MATH327 [Python programming demo](#) (2025)
19. [Beginner's Guide to Python](#) (2024)
20. [W3Schools Python Tutorial](#) (2024)
21. [Software Carpentry](#) tutorials:
 - [Version Control with Git](#) (2024)
 - [Programming with Python](#) (2024)
 - [Plotting and Programming in Python](#) (2024)
22. Stormy Attaway, *MATLAB: A Practical Introduction to Programming and Problem Solving* (2013)
23. B. Barnes and G. R. Fulford, *Mathematical Modelling with Case Studies: Using Maple and MATLAB* (2014)

In addition, there is a vast constellation of other online resources such as [Stack Exchange](#) and [Wikipedia](#). These can be great places to *start* learning about a topic, but are often terrible places to *stop*.

Unit 1: Central limit theorem and diffusion

Introductory remarks: More Is Different

What is most exciting about our work is that it illuminates the chain of connections between, on the one hand, the simple underlying laws that govern the behavior of all matter in the universe and, on the other hand, the complex fabric that we see around us, exhibiting diversity, individuality, and evolution. The interplay between simplicity and complexity is the heart of our subject. —[Murray Gell-Mann](#) (1969 Nobel laureate), 1996

Mathematical sciences such as physics aim to determine the laws of nature and understand how these govern experimental observations — both in everyday circumstances and under extreme conditions. This mathematical understanding is typically guided by reproducing a set of observations, with the resulting framework then used to make predictions for other “observables”.

Over the past few centuries this process has been tremendously successful, with theoretical physics accurately predicting experimental and observational results from sub-atomic through to cosmological scales. Modern physics labs can create a vacuum better than in outer space and the coldest temperatures in the known universe, as well as going to the other extreme to reach temperatures of millions of degrees and pressures millions of times atmospheric pressure at sea level. Amazingly, many aspects of these realms of physics can be described by mathematics developed centuries ago.²

A crucial aspect of this success is the *emergence* of complicated physical phenomena from simple underlying mathematical laws. This module’s name borrows from a famous 1972 essay by [Philip Anderson](#) (1977 Nobel laureate), “[More Is Different](#)”, which helped to establish modern perspectives on emergent phenomena. We will focus on one particular domain in which simple mathematical principles enable amazing predictive capabilities: statistical mechanics (sometimes called probabilistic mechanics, statistical physics or statistical thermodynamics). Initially developed in the context of thermodynamics in the nineteenth century, statistical mechanics remains a central pillar of modern physics, and will retain this position in years to come. The foundations of statistical mechanics lie in the use of probability theory to mathematically describe experimental observations and corresponding laws of nature that involve stochastic randomness rather than being perfectly predictable.

The lack of perfect predictability in statistical mechanics is a matter of practicality rather than one of principle. It arises due to working with a large number of degrees of freedom — that is, a large number of independent objects such as atoms. For illustration, Avogadro’s number $N_A \approx 6.022 \times 10^{23}$ is the large number of molecules in everyday amounts of familiar substances — about 18 grams of water or about 22 litres of air at sea-level atmospheric pressure. Specifying the positions and velocities of $\sim 10^{23}$

²A famous 1960 essay by [Eugene Wigner](#) (1963 Nobel laureate), “[The Unreasonable Effectiveness of Mathematics in the Natural Sciences](#)”, and subsequent work in the philosophy of physics, elaborates on why this may be considered ‘amazing’. While this module will not comment extensively on philosophy, Paolo Beltrame will lead [a short course](#) on this subject during February and March.

objects would require far more information than could be stored by even the biggest existing supercomputers. Statistical mechanics instead produces simple mathematical descriptions of large-scale properties such as temperature, pressure and diffusion, which are generally of such outstanding quality that the underlying ‘randomness’ is effectively invisible.

Historically, the difficulty of detecting the stochastic processes underlying such *thermodynamic* properties made it challenging to convince skeptics that atoms and molecules really exist. [Ludwig Boltzmann](#), a prominent early developer of statistical mechanics, endured a constant struggle to defend his ideas, which likely contributed to his deteriorating mental health and eventual suicide in 1906. A significant step to convincingly establish the existence of atoms was Albert Einstein’s use of statistical mechanics to explain the observed “[Brownian motion](#)” of particles suspended in fluids — this work was part of Einstein’s “miracle year” in 1905, along with special relativity and early contributions to quantum physics. [Jean Perrin](#) soon verified Einstein’s predictions and used them to determine Avogadro’s number; he was awarded the 1926 Nobel Prize in Physics for helping to demonstrate “the discontinuous structure of matter”.

Applications of statistical mechanics continue to advance “our understanding of complex physical systems” — quoting the [2021 Nobel Prize](#) shared between Syukuro Manabe, Klaus Hasselmann and Giorgio Parisi. Even more recently, the [2024 Nobel Prize](#) was awarded to John Hopfield and Geoffrey Hinton for “using ideas from statistical physics” to lay the foundations for machine learning and artificial neural networks. Other modern topics we will encounter in this module include explaining why stars don’t collapse under the ‘weight’ of their own gravity, and identifying effects of dark matter in temperature fluctuations observable in the cosmic microwave background lingering from the early years of the universe.

In this unit we focus on some of the foundational mathematics that will underlie our later development and application of statistical mechanics. Looking back to Boltzmann’s times, we can consider the following question some of his critics might have asked: *If the pressure of a gas in a container results from molecules stochastically colliding with the walls of that container, then how can the pressure be so stable, rather than itself fluctuating stochastically?* The mathematical answer lies in the **law of large numbers** and the **central limit theorem**, which we will review and apply to the physics of diffusion in one dimension.

1.1 Probability foundations

We begin by building a more formal mathematical framework around the concept of probability, through a sequence of definitions. First, a random **experiment** \mathcal{E} involves setting up, manipulating and/or observing some (physical or hypothetical) system that involves some element of randomness. Flipping a coin is a simple random experiment. For the statistical ensembles we will focus on in later units, a typical experiment will be to allow a collection of particles to evolve in time, subject to certain constraints.

Each time an experiment is performed, the world is observed to be in some **state** ω . The specification of the experiment and the state must include all objects of interest,

and may include more besides. When flipping a coin, for example, the full state could contain information not only about the final orientation of the coin, but also about its position — where exactly did it land?

The **set of all states** Ω collects all possible states ω that the given experiment \mathcal{E} can produce, and is therefore intricately tied to \mathcal{E} itself.

We are generally not interested in all aspects of the full state ω . For example, we won't care where a flipped coin lands. Instead we're typically only interested in whether it lands heads up or tails up — and we may want to set aside any state that doesn't cleanly reflect those options. The **measurement** $X(\omega)$ extracts and quantifies this information, acting as a function that maps the state ω to a number that we can mathematically manipulate. If we repeat the given experiment \mathcal{E} many times and carry out the measurement X on each resulting state ω_i , we will obtain a sequence of numbers $X(\omega_i)$ that behave as a *random variable*.

Acting with the measurement X on all of the possible states in the set Ω defines the **set of all outcomes** (or **outcome space**) A :

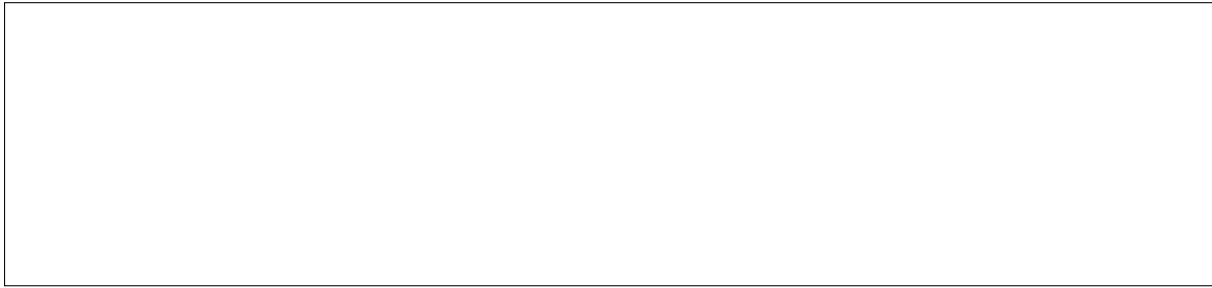
$$X : \Omega \rightarrow A.$$

That is, A collects all possible measurement results that the given experiment \mathcal{E} and measurement X can produce. A can be finite, countably infinite, or uncountably infinite (i.e., continuous).

Let's consider some examples to clarify these definitions. With an experiment of rolling a six-sided die and measuring the number (1–6) that comes out on top, what is the set of all outcomes A ? What additional information could be included in a corresponding state ω ?

What is the outcome space A if we toss a coin four times and measure whether it lands heads up (H) or tails up (T) each time?

What information could characterize a state ω for a gas of 10^{23} argon atoms in a container? What might be interesting to measure?



For convenience, we can introduce a unique number as a *label* to characterize each state ω in the set Ω . Generalizing the concept of measurement, this provides a label function $L(\omega)$ that also behaves as a random variable. Our condition of uniqueness makes $L(\omega)$ isomorphic, so that the label can be used interchangeably with the full state, $\omega \longleftrightarrow L(\omega)$.

While the measurements $X(\omega)$ we consider generally will not produce a unique number for each ω , we will design them precisely to remove irrelevant information that doesn't interest us. Ignoring that irrelevant information leaves us free to interchange the set of outcomes A for the set of states Ω . (Some textbooks may never distinguish between A vs Ω in the first place, though this can be a source of confusion.)

Only a couple of definitions remain. The next is to define an **event** to be any subset of the set of all outcomes A . For example, events resulting from rolling a die could include (i) rolling a 6, (ii) rolling anything but a 6, (iii) rolling any even number, and many more. Collecting all events of interest defines the **set of events** (or **event space**) \mathcal{F} .

We are now prepared for the final foundational definition in this section — the **probability** P of an event in the set \mathcal{F} . Mathematically, P is a *measure function*,

$$P : \mathcal{F} \rightarrow [0, 1],$$

which must satisfy the following two requirements:

1. The probability of a countable union of mutually exclusive events must equal the sum of the probabilities of each of these events.
2. The probability of the outcome space ($\mathcal{F} = A$) must equal 1, even if A is uncountable. This simply means that the experiment \mathcal{E} must produce a measurable outcome. We discard any experiment that doesn't produce such an outcome.

Combining the outcome space, event space and probability measure gives us a *probability space* (A, \mathcal{F}, P) .

For example, consider an experiment that can only produce N possible states, so that

$$\Omega = \{\omega_1, \omega_2, \dots, \omega_N\}.$$

As described above, the subscript is a label, and it is possible for two different states $\omega_i \neq \omega_j$ to produce the same measurement outcome $X(\omega_i) = X(\omega_j)$. This means that the size n of the outcome space A may be smaller than the size of Ω , $n \leq N$. We can write

$$A = \{X_1, X_2, \dots, X_n\},$$

where each X_α is distinct and its label does not necessarily match the one on ω_i . We can take the individual X_α themselves to be the events we're interested in, choosing the event space

$$\mathcal{F} = \{X_1, X_2, \dots, X_n\} = A. \quad (1)$$

These events are all mutually exclusive, so if we assign them probabilities

$$P(X_\alpha) \equiv p_\alpha \quad \text{for } \alpha = 1, \dots, n,$$

then the above requirements on probabilities demand that for any $\alpha \neq \beta$ we have

$$P(X_\alpha \text{ or } X_\beta) = p_\alpha + p_\beta$$

$$P(A) = P(X_1 \text{ or } X_2 \text{ or } \dots \text{ or } X_n) = \sum_{\alpha=1}^n p_\alpha = 1.$$

Similarly choosing $\mathcal{F} = A$ for the six-sided die considered in an earlier gap, what are the probabilities p_1 through p_6 that result from assuming the die is fair?

Again taking $\mathcal{F} = A$ for the case of tossing a coin four times, what are the probabilities p_α that result from assuming the coin is fair? If we instead consider the event space

$$\mathcal{F} = \{\text{equal number of } H \text{ and } T, \text{ different numbers of } H \text{ and } T\},$$

what are the probabilities p_{equal} and p_{diff} for the two events in this \mathcal{F} ?



The standard European roulette wheel shown to the left ([source](#)) has 37 pockets labelled “0” through “36”. 18 of these pockets are coloured red, 18 are coloured black and 1 (pocket “0”) is coloured green. Let an experiment be a spin of the roulette wheel, measuring the label of the pocket where the ball comes to rest (which also provides the pocket’s colour).

What is the outcome space A for a spin of the roulette wheel? With $\mathcal{F} = A$, what are the probabilities p_α for a fair wheel? With

$$\mathcal{F} = \{\text{ball in a red pocket, ball in a black pocket, ball in the green pocket}\},$$

what are the corresponding probabilities p_{red} , p_{black} and p_{green} ?

The process of assigning probabilities to events is called *modelling*. The gaps above demonstrate that *symmetries* are a powerful way to constrain probabilities. The symmetry between the six sides of a fair die, the two sides of a fair coin, and the 37 pockets of a fair roulette wheel each sufficed to completely determine the corresponding probabilities p_α .

Modelling can also be guided by empirical data obtained by repeating an experiment many times. For example, if we don’t know whether a set of dice are fair, we will be able to infer their probabilities p_α (with a certain confidence level) by rolling them enough times. The need to repeat the experiment many times comes from the law of large numbers, to which we now turn.

1.2 Law of large numbers

Let’s return to the setup leading to Eq. 1 above, with

$$\mathcal{F} = A = \{X_1, X_2, \dots, X_n\}$$

for finite n , and probabilities $p_\alpha = P(X_\alpha)$ that obey

$$p_\alpha \in [0, 1] \qquad \sum_{\alpha=1}^n p_\alpha = 1.$$

We can generalize this notation by writing instead

$$\sum_{X \in A} P(X) = 1,$$

and introducing similar expressions for the **mean** μ and **variance** σ^2 of the probability space,

$$\mu = \langle X \rangle = \sum_{X \in A} X P(X) \quad (2)$$

$$\sigma^2 = \langle (X - \mu)^2 \rangle = \sum_{X \in A} (X - \mu)^2 P(X). \quad (3)$$

The angle bracket notation indicates the **expected** (or **expectation**) **value** with general definition

$$\langle f(X) \rangle = \sum_{X \in A} f(X) P(X), \quad (4)$$

which is a linear operation,

$$\langle c \cdot f(X) + g(X) \rangle = c \langle f(X) \rangle + \langle g(X) \rangle.$$

The square root of the variance, $\sqrt{\sigma^2} = \sigma$, is the **standard deviation**. What is σ expressed in terms of $\langle X^2 \rangle$ and $\langle X \rangle^2$?

We now define a new experiment that consists of *repeating* the original experiment R times, with each repetition independent of all the others. Using the same measurement as before for each repetition, we obtain a new outcome space that we can call B . For $R = 4$, what are some representative outcomes in the set B ? What is the total size of B ?

Each outcome in B contains R different $X^{(r)} \in A$, one for each repetition $r = 1, \dots, R$, and each with mean $\langle X^{(r)} \rangle = \mu$ and variance $\langle (X^{(r)} - \mu)^2 \rangle = \sigma^2$. Considering the case $R = 4$ for simplicity, any element of B can be written as $X_i^{(1)} X_j^{(2)} X_k^{(3)} X_l^{(4)} \in B$ with corresponding probability

$$P_B \left(X_i^{(1)} X_j^{(2)} X_k^{(3)} X_l^{(4)} \right) = P_A \left(X_i^{(1)} \right) P_A \left(X_j^{(2)} \right) P_A \left(X_k^{(3)} \right) P_A \left(X_l^{(4)} \right),$$

using subscripts to distinguish between the probability spaces for the single experiment (A) and repeated experiment (B).

Averaging over all R repetitions defines the *arithmetic mean*

$$\bar{X}_R = \frac{1}{R} \sum_{r=1}^R X^{(r)}. \quad (5)$$

Unlike the true mean μ , the arithmetic mean \bar{X}_R is a random variable — a number that may be different for each element of B . That said, \bar{X}_R and μ are certainly related, and so long as the standard deviation exists — that is, so long as σ^2 is finite — this relation can be proved rigorously in the limit $R \rightarrow \infty$.³

Here we will not be fully rigorous, and take it as given that

$$\langle (X^{(i)} - \mu) (X^{(j)} - \mu) \rangle = \sigma^2 \delta_{ij} = \begin{cases} \sigma^2 & \text{for } i = j \\ 0 & \text{for } i \neq j \end{cases},$$

where the *Kronecker delta* $\delta_{ij} = 1$ for $i = j$ and vanishes for $i \neq j$. This is a consequence of our requirement that each repetition is independent of all the others. Using this result and the relation $(\sum_i a_i)(\sum_j b_j) = \sum_{i,j} (a_i b_j)$, express the following quantity in terms of σ and R :

$$\left\langle \left(\frac{1}{R} \sum_{r=1}^R X^{(r)} - \mu \right)^2 \right\rangle =$$

³In the computer project we will numerically investigate a situation where σ^2 diverges.

You should find that your result vanishes in the limit $R \rightarrow \infty$, so long as σ^2 is finite. Since the square makes this expectation value a sum of non-negative terms, it can vanish only if every one of those terms is individually zero.

This establishes the **law of large numbers**:

$$\lim_{R \rightarrow \infty} \frac{1}{R} \sum_{r=1}^R X^{(r)} = \mu, \quad (6)$$

where we have assumed finite $\langle X^{(r)} \rangle = \mu$ and $\langle (X^{(r)} - \mu)^2 \rangle = \sigma^2$.

1.3 Probability distributions

It is not necessary to make the assumption (Eq. 1) that our outcome space contains only a countable number of possible outcomes. The considerations above continue to hold even if the random variable X is a continuous real number. In this case, however, the identification of probabilities with outcomes is slightly more complicated, which will be relevant when we consider the central limit theorem in the next section.

When the outcome can be any number on the real line, the fundamental object is a **probability distribution** (or **density function**) $p(x)$ defined for all $x \in \mathbb{R}$. Starting from this density, a probability is determined by integrating over a given interval. Calling this interval $[a, b]$, the integration produces the probability that the outcome X lies within the interval,

$$P(a \leq X \leq b) = \int_a^b p(x) dx.$$

We similarly generalize the definition of an expectation value (Eq. 4) to an integral over the entire domain of the probability distribution,

$$\langle f(x) \rangle = \int f(x) p(x) dx.$$

We will omit the limits on integrals over the entire domain, so for $x \in \mathbb{R}$ we implicitly have $\int dx = \int_{-\infty}^{\infty} dx$, with $\int p(x) dx = 1$. An important set of expectation values is

$$\langle x^\ell \rangle = \int x^\ell p(x) dx, \quad (7)$$

which provides the mean and variance of the probability distribution $p(x)$, through generalizations of Eqs. 2–3:

$$\mu = \langle x \rangle = \int x p(x) dx \quad \sigma^2 = \langle x^2 \rangle - \langle x \rangle^2. \quad (8)$$

The expression for the variance should be familiar from your determination of the standard deviation in an earlier gap. Unless stated otherwise, we will assume the mean and variance are both finite for the probability distributions we consider.

1.4 Central limit theorem

The central limit theorem is a central result of probability theory (hence its name). Over the years it has been expressed in several equivalent ways, and there are also many distinct variants of the theorem accommodating different conditions and assumptions. Here we are interested in applying rather than proving the central limit theorem; you can find [proofs](#) in many textbooks.

The version of the theorem we use in this module assumes we have N independent random variables x_1, \dots, x_N , each of which has the same (finite) mean μ and variance σ^2 . Such random variables are said to be *identically distributed*, and a common way to obtain them is to repeat an experiment N times, as we considered in Section 1.2. Just as in Eq. 5, the sum

$$s = \sum_{i=1}^N x_i \quad (9)$$

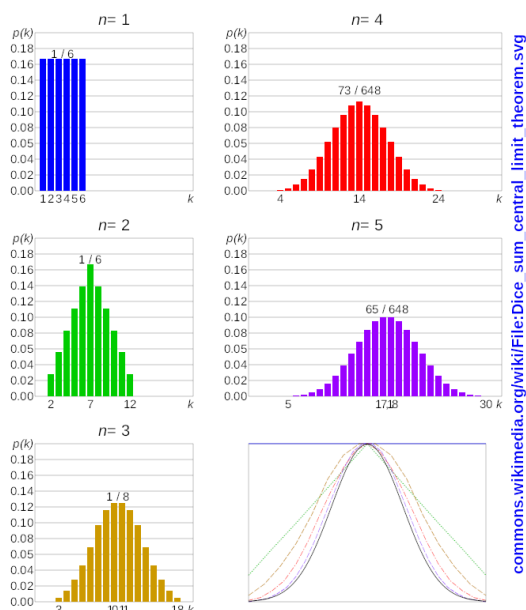
is itself a random variable.

The **central limit theorem** states that for large $N \gg 1$ the probability distribution for s is

$$p(s) \approx \frac{1}{\sqrt{2\pi N\sigma^2}} \exp\left[-\frac{(s - N\mu)^2}{2N\sigma^2}\right], \quad (10)$$

with the approximation becoming exact in the $N \rightarrow \infty$ limit.

In addition to asserting that the collective behaviour of many independent and identically distributed random variables x_i is governed by a **normal** (or **gaussian**) **distribution**, the central limit theorem further specifies the precise form of this distribution in terms of the mean and variance of *each individual* x_i .



In practice, N often doesn't need to be very large in order for the central limit theorem to provide a reasonable approximation. This is illustrated by the figure to the left, from [Wikimedia Commons](#), which shows the probabilities for the sum of $n \leq 5$ rolls of a six-sided die rapidly approaching a gaussian distribution.

1.5 Diffusion and the central limit theorem

1.5.1 Random walk on a line

As a more powerful application and illustration of the central limit theorem, let's consider the behaviour of a randomly moving object. Such **random walks** appear frequently in mathematical modelling of stochastic phenomena (including Brownian motion), and can be applied to movement through either physical space or more abstract vector spaces. They are examples of [Markov processes](#), in which the state of the system (in this case the position of the 'walker') at any time probabilistically depends only on the system's prior state at the previous point in time — there is no 'memory' of any earlier states. The resulting sequence of system states is known as a *Markov chain*, since each state is produced from the one before, like links in a chain.

Let's consider the simple example of a random walker that moves only in a single spatial dimension — to the left or to the right on a line — and can only take 'steps' of a fixed length, which we can set to $\ell = 1$ without loss of generality. At each point in time, the walker takes either a step to the right (*R*) with probability p or a step to the left (*L*) with probability $q = 1 - p$. We will further assume that each step takes a constant amount of time δt , so a walk of N steps will last for total time

$$t = N\delta t. \tag{11}$$

As an example, for $N = 6$ a representative walk can be written as *LRLRRR*, which leaves the walker $x = 2$ steps to the right of its starting point ($x = 0$). The opposite walk *RLRLLL* would leave the walker at $x = -2$, with negative numbers indicating positions to the left of the starting point. How many possible walks are there for $N = 6$, and what is the probability (in terms of p and q) for the walks *LRLRRR* and *RLRLLL* to occur? How many possible walks are there for general N , and what is the probability for any particular walk involving r steps to the right to occur?



We are interested in the walker's final position x at time t after it has taken N steps. Just as for the sum of n rolls of a die considered in Section 1.4, there are a range of possible final positions x , each of which has some probability $P(x)$ of being realized. The key pieces of information we want to determine are the expectation value $\langle x \rangle$ and the standard deviation $\Delta x \equiv \sqrt{\langle x^2 \rangle - \langle x \rangle^2}$ that indicates the scale of fluctuations we can expect around $\langle x \rangle$ as the N -step walk is repeated many times from the same starting point. (We reserve the variables μ and σ^2 for the mean and variance of the single-step process, which will play an important role when we apply the central limit theorem in Section 1.5.3.)

Suppose the N total steps involve r steps to the right. What is the final position x of the walker in terms of N and r ? Check your general answer for the cases $N = 6$ and $r = 4, 2$ considered above.

This relation makes it equivalent to consider either the probability P_r of taking r steps to the right, or the probability $P(x)$ of ending up at final position x . This equivalence **will not hold** for more general random walks in which the step length is no longer fixed and ℓ_i can vary from one step to the next.

Because the order in which steps are taken does not affect the final position x , to determine the probability $P(x)$ we have to count all possible ways of walking to x . For $N = 6$, what are all the possible walks that produce $x = 4$, and what is the corresponding probability $P(4)$?

Your answer should have a factor of 6 that corresponds to the binomial coefficient $\binom{N}{r} = \binom{6}{5} = 6$. In terms of this binomial coefficient, what is the general probability P_r that an N -step walk will include r steps to the right in any order?

Given this probability P_r , we can apply Eqs. 2–3 to find the expectation value $\langle x \rangle$ and the standard deviation Δx . As a first step, what are $\langle x \rangle$ and $\langle x^2 \rangle$ in terms of the expectation values $\langle r^n \rangle = \sum_{r=0}^N r^n P_r$?

$\langle x \rangle =$

$\langle x^2 \rangle =$

Now we need to calculate the necessary $\langle r^n \rangle$. A powerful way to do this is to define the **generating function**

$$G(\theta) = \sum_{r=0}^N e^{r\theta} P_r. \quad (12)$$

This approach introduces a parameter θ that we subsequently remove by setting $\theta = 0$. For example, $G(0) = \sum_{r=0}^N P_r = 1$. What do you obtain upon taking derivatives of the generating function and then setting $\theta = 0$?

$\left. \frac{d}{d\theta} G(\theta) \right|_{\theta=0} =$

$\left. \frac{d^n}{d\theta^n} G(\theta) \right|_{\theta=0} =$

For the current case of a fixed-step-length random walk in one dimension, the probabilities P_r produce a simple closed-form expression for the generating functional:

$$G(\theta) = \sum_{r=0}^N e^{r\theta} P_r = \sum_{r=0}^N e^{r\theta} \binom{N}{r} p^r q^{N-r} = (e^\theta p + q)^N, \quad (13)$$

making use of the binomial formula $(a + b)^N = \sum_{i=0}^N \binom{N}{i} a^i b^{N-i}$.

It's straightforward to take the necessary derivatives of Eq. 13, which simplify pleasantly since $(e^\theta p + q)|_{\theta=0} = p + q = 1$:

$$\left. \frac{d}{d\theta} (e^\theta p + q)^N \right|_{\theta=0} =$$

$$\left. \frac{d^2}{d\theta^2} (e^\theta p + q)^N \right|_{\theta=0} =$$

Insert the resulting $\langle x \rangle$ and $\langle x^2 \rangle$ into the relations derived above:

$$\langle x \rangle =$$

$$(\Delta x)^2 = \langle x^2 \rangle - \langle x \rangle^2 =$$

In the end, you should obtain

$$\langle x \rangle = N(2p - 1) \qquad \Delta x = 2\sqrt{Npq}. \qquad (14)$$

We can check that this $\langle x \rangle$ produces the expected results in the special cases $p = 0$, $1/2$ and 1 , while the standard deviation Δx also behaves appropriately by vanishing for both $p = 0$ and 1 .

1.5.2 Law of diffusion

It's possible to interpret the results in Eq. 14 in a more intuitive way by expressing them in terms of the total time t taken by the random walk (Eq. 11). Inserting $N = t/\delta t$ into Eq. 14,

$$\langle x \rangle = \frac{t}{\delta t}(2p - 1) = \frac{2p - 1}{\delta t}t \equiv v_{\text{dr}}t,$$

we see that the *expected* final position of the walker depends linearly on time, with **drift velocity**

$$v_{\text{dr}} = \frac{2p - 1}{\delta t} = \frac{N(2p - 1)}{t} = \frac{\langle x \rangle}{t}. \qquad (15)$$

The sign of v_{dr} indicates whether the walker is drifting to the right ($p > \frac{1}{2}$) or to the left ($p < \frac{1}{2}$). The typical scale of fluctuations (or the ‘uncertainty’) around the expected final position $\langle x \rangle$ is given by the standard deviation Δx , and also depends on time:

$$\Delta x = \sqrt{\langle x^2 \rangle - \langle x \rangle^2} = 2\sqrt{Npq} = 2\sqrt{\frac{pq}{\delta t}}\sqrt{t},$$

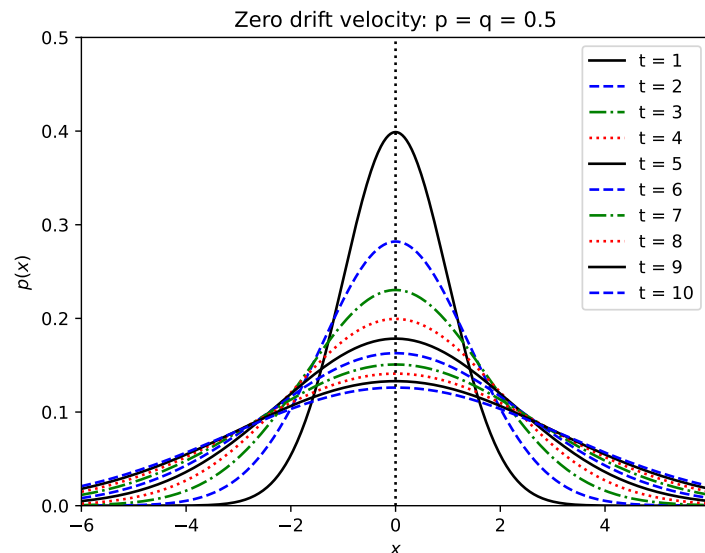
where the constant factor is non-negative. This \sqrt{t} dependence is a very general result.

The **law of diffusion** states that

$$\Delta x = D\sqrt{t}, \tag{16}$$

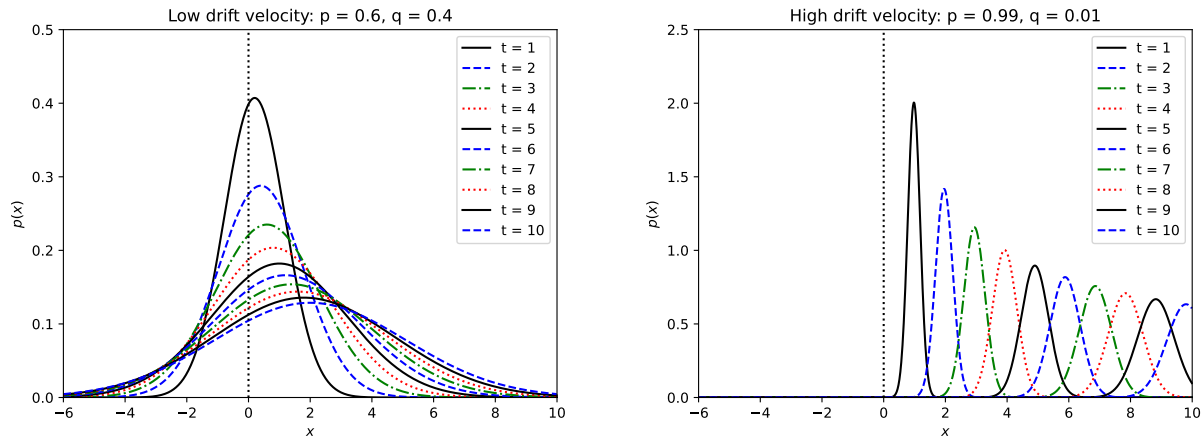
where $D > 0$ is the **diffusion constant**. The standard deviation Δx is sometimes also called the *diffusion length*.

The result $D = 2\sqrt{\frac{pq}{\delta t}}$ that we computed above is specific to the simple example of a fixed-step-length random walk in one dimension and **will not hold** for more general random walks. The behaviour it describes is illustrated by the figure below, which plots the t -dependent probability distribution $p(x)$ that we’ll soon derive using the central limit theorem (Eq. 17). What we can see already, even before completing that derivation, is that the probability distribution steadily spreads out — or *diffuses* — as time passes:



Here we are considering the special case $p = q = \frac{1}{2}$, for which the drift velocity $v_{\text{dr}} = 0$ and the expectation value is always $\langle x \rangle = 0$ for any walk time t . However, as time goes on, there is a steady decrease in the probability that the walker will end up around its starting point $x \approx 0$. (As described in Section 1.3, we can determine this probability by integrating the distribution $p(x)$ over an appropriate interval.) Instead, the interval $-D\sqrt{t} \leq x \leq D\sqrt{t}$ within which we can expect to find the walker (with a constant ‘one-sigma’ or $\sim 68\%$ probability) steadily grows with characteristic dependence on the square root of the time the diffusive process lasts.

Except in the trivial cases $p = 0$ or $q = 0$, diffusion also occurs when the drift velocity is non-zero. This is shown in the two figures below, considering a low but non-zero drift velocity $v_{\text{dr}} = 0.2$ on the left, and a high $v_{\text{dr}} = 0.98$ on the right. All three figures were produced by [this Python script](#).



In the figure on the left, the large diffusion constant $D \approx 0.980$ produces a plot that looks similar to the $v_{\text{dr}} = 0$ case on the previous page, but now the central peaks (and expectation values $\langle x \rangle$) for each t drift steadily to the right. The lower $D \approx 0.199$ for the figure on the right leads to less overlap among the distributions, but they still diffuse to exhibit lower and broader peaks as time passes.

When $p \neq \frac{1}{2}$ so that $\langle x \rangle \neq 0$, it is interesting to compare the drift in the expectation value against the growth in fluctuations around $\langle x \rangle$ due to diffusion. We can do this by considering the following *relative* uncertainty:

$$\frac{\Delta x}{\langle x \rangle} =$$

You should find that at large times this ratio vanishes proportionally to $1/\sqrt{t} \propto 1/\sqrt{N}$. Although the absolute uncertainty $\Delta x = D\sqrt{t}$ grows by diffusion, for $v_{\text{dr}} \neq 0$ the linear drift in the expectation value becomes increasingly dominant as time goes on.

1.5.3 Applying the central limit theorem

Based on our work in Section 1.4, we can see how to apply the central limit theorem to analyse this fixed-step-length random walk in one dimension, for large numbers of steps N or equivalently large times $t = N\delta t$. Each step in the random walk is an independent and identically distributed random variable x_i . The corresponding probability space involves only two possible outcomes: a step of length $\ell = 1$ to the right or to the left with probability p or q , respectively. From this we can easily compute the mean and variance of the single-step process:

$$\begin{aligned}\mu &= \langle x_i \rangle = \\ \langle x_i^2 \rangle &= \\ \sigma^2 &= \langle x_i^2 \rangle - \langle x_i \rangle^2 =\end{aligned}$$

The final position x of the walker after N steps is exactly the sum over these x_i given in Eq. 9. Its probability distribution $p(x)$ from the central limit theorem is therefore obtained directly from these single-step μ and σ^2 , which we can also express in terms of the drift velocity and diffusion constant:

$$\begin{aligned}p(x) &= \frac{1}{\sqrt{2\pi(4Npq)}} \exp\left[-\frac{(x - N(2p - 1))^2}{8Npq}\right] \\ &= \frac{1}{\sqrt{2\pi D^2 t}} \exp\left[-\frac{(x - v_{\text{dr}} t)^2}{2D^2 t}\right].\end{aligned}\tag{17}$$

This expression was used to produce the three figures above. We could have jumped straight to the final line by considering Eq. 10 and noting

$$v_{\text{dr}} t = N(2p - 1) = N\mu \qquad D^2 t = 4pq \frac{t}{\delta t} = N\sigma^2.\tag{18}$$

While this dependence on p and q is specific to the particular fixed-step-length random walk we're currently considering, the generic results $v_{\text{dr}} t = N\mu$ and $D^2 t = N\sigma^2$ in Eq. 18 **do hold** for any random walk with finite μ and σ^2 (so that the central limit theorem can be applied). This is remarkable, because it means that the diffusive process as a whole is determined entirely by the single-step mean and variance. So long as μ and σ^2 are finite, we end up with Eq. 17 as the large- t probability distribution for any markovian random walk in a single variable x .

This result is related to the generality of the law of diffusion (Eq. 16), which we can recognize in the structure of Eq. 17. Since $t > 0$, the exponential in the gaussian distribution $p(x)$ peaks at the drifting expectation value $x = v_{\text{dr}} t = \langle x \rangle$. The factor

$(x - v_{dr}t)^2$ simply quantifies the distance from this peak. As t increases, so does the factor $2D^2t$ dividing this $(x - v_{dr}t)^2$, meaning that a larger distance from the peak is needed for the overall argument of the exponential to reach a given value — in other words, the peak becomes broader. This in turn requires a lower peak, reflected in the $\frac{1}{\sqrt{t}}$ in the overall coefficient, which is set by requiring $\int p(x) dx = 1$. In other words, the law of diffusion holds whenever the central limit theorem is applicable. This requires that the mean and variance of the single-step process are finite, and in the computer project we will numerically investigate the *anomalous diffusion* that occurs when this condition is not satisfied.

Unit 2: Micro-canonical ensemble

2.1 Statistical ensembles and thermodynamic equilibrium

We begin this unit by establishing the concept of *statistical ensembles*, which was formalized by [J. Willard Gibbs](#) in 1902. (Gibbs also introduced the term ‘statistical mechanics’, in 1884.) Building on the probability foundations laid above, we will be interested in ‘experiments’ that simply allow a collection of degrees of freedom to evolve in time, subject to certain constraints. At a given time t_1 , the arrangement of these degrees of freedom defines the state ω_1 of the system.

As a concrete example, consider a system of *spins* — arrows that can point either ‘up’ or ‘down’ — arranged in a line. Such [spin systems](#) will appear several times in the remainder of this module, since in addition to obeying simple mathematics analogous to flipping coins, spins also serve as good models of physical systems such as magnetic molecules. What would be a representative state (or *configuration*) for a system of $N = 8$ spins? How many distinct states are there for this system?



At a different time t_2 , the system’s state ω_2 is generally different from ω_1 . However, there are some measurements we can perform that always produce the same outcome even as the system’s state changes over time. These measurements define *conserved quantities*, such as the number of spins considered in the example above.

Another important conserved quantity is the *internal energy* E of an isolated (or ‘closed’) system,

$$E(\omega_1) = E(\omega_2).$$

The conservation of energy is presumably a familiar concept, and you may also know that it can be rigorously proven through [Emmy Noether’s first theorem](#).⁴ Because conservation of energy was empirically observed long before Noether’s theorem was proven, it also has a more grandiose name: the **first law of thermodynamics**. Another way of stating the first law is that any change in the internal energy of one particular system Ω must be matched by an equal and opposite change in the energy of some other system with which Ω is in contact. This will be important when we consider thermodynamic cycles later in the term.

For now, let’s return to the example above, and endow the spin system with an internal energy by placing it in a ‘magnetic field’ of strength H . That is, if a spin is

⁴There are [complications](#) when considering the dynamical space-time of general relativity, but that’s beyond the scope of this module.

parallel to the field, it contributes energy $-H$ to the total energy E of the system. If a spin is anti-parallel to the field, it instead contributes energy H . For later convenience, we define a positive magnetic field $H > 0$ to point upward, and also define n_+ to be the number of spins pointing upward — parallel to the field and therefore contributing *negative* energy. Similarly, the remaining $n_- = N - n_+$ downward-pointing spins are anti-parallel to the field and contribute positive energy. What is the total energy E of the system in terms of n_+ and n_- ? What is E for the representative 8-spin state you wrote down above? What fraction of the states of the spin system have this energy?

If instead we consider $N \sim 10^{23}$ hydrogen (H_2) molecules in a container, we can write a simple expression for the internal energy E by treating each molecule as a *point-like particle*, with no size or structure. In this case each molecule contributes only its kinetic energy, and

$$E = \frac{m}{2} \sum_{n=1}^N \vec{v}_n^2 = \frac{1}{2m} \sum_{n=1}^N \vec{p}_n^2,$$

where \vec{v}_n is the velocity of the n th molecule, $\vec{p}_n = m\vec{v}_n$ is its momentum, and all molecules have the same mass m .

As forecast at the start of the module, we treat the time evolution of any given system as a stochastic process in which the system probabilistically adopts a sequence of states $\omega_i \in \Omega$:

$$\omega_1 \longrightarrow \omega_2 \longrightarrow \omega_3 \longrightarrow \omega_4 \longrightarrow \dots$$

This approach is a matter of practicality rather than one of principle. In principle, Newton's laws would allow us to predict the exact time evolution of (say) 10^{23} hydrogen molecules, but only by specifying 10^{23} initial conditions and solving 10^{23} differential equations. Since we cannot hope to record so much information or carry out so many computations, we instead apply probability theory in order to analyse these systems.

This leads us to the following core definition: A **statistical ensemble** is the set of all states $\Omega = \{\omega_1, \omega_2, \dots\}$ that a system can possibly adopt through its time evolution. Each state ω_i has some probability p_i of being adopted by the system, so we can recognize a statistical ensemble as a probability space.

Because these states ω_i depend on the 'microscopic' degrees of freedom that compose the overall system, we will refer to them as **micro-states** from now on. From the definition of probability in Section 1.1, we have the requirement $\sum_i p_i = 1$, which

simply means that the system must be in *some* micro-state at any point in time. The fact that time evolution cannot change any conserved quantities, as discussed above, means that such conserved quantities characterize statistical ensembles. Throughout the next seven units we will consider different statistical ensembles with different sets of conserved quantities.

First we define the **micro-canonical ensemble** to be a statistical ensemble characterized by conserved internal energy E and conserved number of degrees of freedom N — which we will call **particle number** for short.

According to the discussion above, this means that a system governed by the micro-canonical ensemble is *isolated* in the sense that it cannot exchange energy or particles with any other system.

Now that the micro-canonical ensemble is defined, we can connect it to our intuition from everyday physical systems. Let's consider a collection of particles moving around and bouncing (or '*scattering*') off each other in a sealed container. To a first approximation, this should describe the behaviour of air in a room, which our lived experience indicates is spread quite uniformly throughout the room in a way that is stable as time passes. We do not expect all the air in a room to be concentrated in any one corner, nor do we expect strong collective gusts of wind without some clear external influence.

These qualitative expectations illustrate the idea of **thermodynamic equilibrium**, an axiomatic concept in statistical mechanics.⁵ We can mathematically define thermodynamic equilibrium through the probabilities p_i that appear in the micro-canonical ensemble.

A micro-canonical system Ω with M micro-states ω_i is in thermodynamic equilibrium if and only if all probabilities p_i are equal. If M is finite, the requirement $\sum_i p_i = 1$ implies

$$p_i = \frac{1}{M}. \quad (19)$$

The full meaning and significance of this definition are not immediately obvious, and we will continue exploring them through consideration of derived quantities such as entropy and temperature. First, it's important to emphasize that this equilibrium is *dynamic*: There is not a single 'equilibrium micro-state' that the system sits in. Instead, the equilibrium system continues probabilistically adopting all possible micro-states as it evolves in time.

⁵Our expectation that physical systems generically evolve towards thermodynamic equilibrium as time passes is more formally expressed as the [ergodic hypothesis](#).

2.2 Entropy and its properties

2.2.1 Definition of entropy

We can gain further insight into thermodynamic equilibrium by considering a famous derived quantity.

The **entropy** of a statistical ensemble Ω with a countable number of micro-states M is defined to be

$$S = - \sum_{i=1}^M p_i \log p_i, \quad (20)$$

where p_i is the probability for micro-state ω_i to occur. Unless otherwise specified, “log” indicates the natural logarithm with base e .

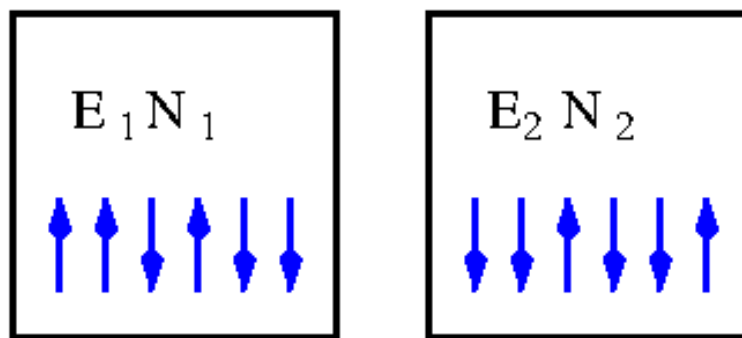
When the system under consideration is in thermodynamic equilibrium, we expect derived quantities such as the entropy to be stable over time, even as different micro-states are probabilistically adopted. This implies that such derived quantities are functions of the conserved quantities that are the same for all micro-states. Therefore, for the micro-canonical ensemble, the equilibrium entropy $S(E, N)$ is a function of the conserved energy and particle number.

By inserting Eq. 19 into Eq. 20 you can quickly compute a simple expression for the entropy of a micro-canonical ensemble in thermodynamic equilibrium:

Your result should depend only on the number of micro-states M , diverging as $M \rightarrow \infty$. While the energy E and particle number N are not explicit in this expression, $\{E, N, M\}$ are inter-related and can be expressed in terms of each other given the details of any specific situation under consideration. For example, what is the equilibrium entropy of the system of N spins considered above, if the magnetic field is turned off, $H = 0$? What is the entropy if $E = 0$ with $H > 0$ (which requires $n_+ = n_-$)?

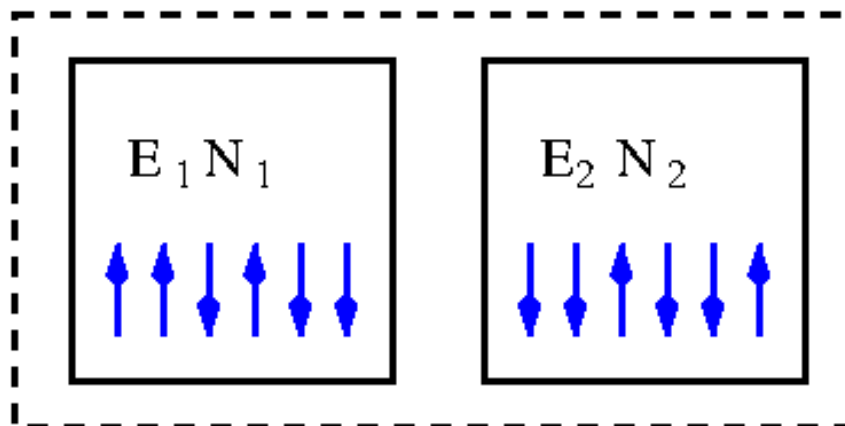
2.2.2 Extensivity

The increase in entropy for an increasing number of micro-states M is a reflection of entropy being an *extensive* quantity. Extensive quantities are formally defined by considering how they behave if two isolated subsystems are *analysed* as a single system — while still remaining isolated from each other, exchanging neither energy nor particles. This is clearest to consider through the specific example shown below of two isolated spin subsystems, Ω_1 & Ω_2 , characterized by the energies E_1 & E_2 and particle numbers N_1 & N_2 , respectively. To simplify the subsequent analysis, we can assume that both subsystems are placed in magnetic fields with the same H , so that $E_S = -H(n_+^{(S)} - n_-^{(S)})$ for $S \in \{1, 2\}$.



We can take system Ω_1 to have M_1 micro-states with probabilities p_i while system Ω_2 has M_2 micro-states with probabilities q_k . As discussed above, each M_S is determined by E_S and N_S . The entropies of the two systems are

$$S_1 = - \sum_{i=1}^{M_1} p_i \log p_i \qquad S_2 = - \sum_{k=1}^{M_2} q_k \log q_k.$$



Now we keep these two subsystems isolated from each other, but consider them as a combined system Ω_{1+2} , as illustrated above. In order to compute the entropy S_{1+2} ,

we need to figure out the number of micro-states M_{1+2} the combined system could possibly adopt, and then determine the corresponding probability for each micro-state. Both steps are simplified by the subsystems being isolated from each other, so that they are statistically independent. Specifically, with subsystem Ω_1 in any one of its M_1 micro-states $\omega_i^{(1)}$, subsystem Ω_2 could independently adopt any of its M_2 micro-states, implying $M_{1+2} = M_1 M_2$.

Similarly, statistical independence means that the combined probability of subsystem Ω_1 adopting micro-state $\omega_i^{(1)}$ while subsystem Ω_2 adopts $\omega_k^{(2)}$ is the product of the individual probabilities, $p_i q_k$. We can check this is a well-defined probability, with

$$\sum_{M_{1+2}} p_i q_k = \sum_{i=1}^{M_1} \sum_{k=1}^{M_2} p_i q_k = \left[\sum_{i=1}^{M_1} p_i \right] \cdot \left[\sum_{k=1}^{M_2} q_k \right] = 1 \cdot 1 = 1.$$

Inserting the probability $p_i q_k$ into Eq. 20, and recalling $\log(a \cdot b) = \log a + \log b$, what is the combined entropy S_{1+2} of these two independent subsystems?

$S_{1+2} =$

You should find that the total entropy is the sum of the entropies of the two isolated subsystems, which is also how the energies and particle numbers behave,

$$E_{1+2} = E_1 + E_2 \qquad N_{1+2} = N_1 + N_2.$$

This behaviour identifies the energy, particle number and entropy as **extensive** quantities, which are **defined** to be those that add up across independent subsystems. This can be contrasted with **intensive** quantities, which are **defined** to be independent of the extent of the system, and hence the same (on average) for subsystems as for the combined system. Temperature and density are everyday examples of intensive quantities, though we will see below that the micro-canonical approach introduces some subtleties. It is possible for quantities to be neither extensive nor intensive, for example the number of micro-states $M_{1+2} = M_1 M_2$.

Finally, suppose that each subsystem is independently in thermodynamic equilibrium, with finite M_1 and M_2 , implying

$$p_i = \frac{1}{M_1} \qquad q_k = \frac{1}{M_2}$$

$$S_1 = \log M_1 \qquad S_2 = \log M_2.$$

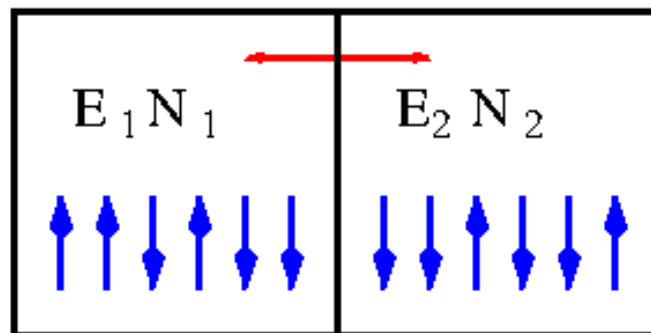
As a consequence we can establish that Ω_{1+2} is also in thermodynamic equilibrium, since the probabilities

$$p_i q_k = \frac{1}{M_1 M_2} = \frac{1}{M_{1+2}}$$

are identical all of its micro-states. In this situation it's even easier to see $S_{1+2} = \log(M_1 M_2) = \log M_1 + \log M_2 = S_1 + S_2$.

2.2.3 Second law of thermodynamics

Let's continue considering two subsystems, with one significant change: Suppose the subsystems are now able to exchange energy (but not particles) with each other. We'll say they are in *thermal contact* with each other, rather than being fully isolated. We'll also wait long enough after establishing thermal contact for the combined system Ω to reach equilibrium. This is illustrated below:



The total energy $E = E_1 + E_2$ remains conserved, so the overall system Ω is still governed by the micro-canonical ensemble. However, the individual energies E_1 and E_2 can now change over time, meaning that *each subsystem is no longer micro-canonical*.

The overall system Ω is *not* the same as the combined Ω_{1+2} considered above. We need to reconsider the total number of micro-states M that Ω could adopt, which is much more difficult than before because we can no longer apply statistical independence. Our key remaining tool is the conservation of the total energy E .

Considering a micro-state in which the N_1 spins contribute energy e_1 to the total, we know that the N_2 spins must contribute the remaining $E - e_1$. Our work above implies there are $M_{e_1} = M_{e_1}^{(1)} M_{E-e_1}^{(2)}$ micro-states providing this particular distribution of energies, where $M_{e_1}^{(1)}$ is the number of micro-states of the formerly isolated subsystem Ω_1 with energy e_1 , and $M_{E-e_1}^{(2)}$ similarly corresponds to Ω_2 with energy $E - e_1$. We also know that it's possible to have $e_1 = E_1$, since that's the initial energy of Ω_1 before it was brought into thermal contact with Ω_2 . When $e_1 = E_1$, we have $M_{E_1} = M_1 M_2$, covering all the micro-states of the combined Ω_{1+2} when the two subsystems were isolated. *In addition*, we also have to count any other micro-states for which $e_1 \neq E_1$:

$$M = \sum_{e_1} M_{e_1}^{(1)} M_{E-e_1}^{(2)} = M_1 M_2 + \sum_{e_1 \neq E_1} M_{e_1}^{(1)} M_{E-e_1}^{(2)} \geq M_1 M_2. \quad (21)$$

Equality holds when $e_1 = E_1$ is the only possibility — this is an extremely special case, in which the two subsystems remain individually micro-canonical, with fixed E_1 and E_2 . This is all we can say in full generality, without specifying more details of a particular example, but it allows us to obtain a famous result for the total entropy S of Ω in *thermodynamic equilibrium*:

$$S = \log M \geq \log(M_1 M_2) = S_{1+2}.$$

This is a form of the **second law of thermodynamics**,

$$S \geq S_{1+2} = S_1 + S_2.$$

In words, whenever initially isolated (sub)systems in thermodynamic equilibrium are brought into thermal contact with each other and allowed to exchange energy, the total entropy of the overall system can never decrease. Indeed, it generically increases except in extremely special cases.

Though we won't go through a more general derivation here, it turns out that the total entropy never decreases (and generically increases) as time passes, under *any* circumstances. This has many far-reaching consequences, the first of which is a more general definition of thermodynamic equilibrium that (unlike Eq. 19) will also apply when we consider statistical ensembles other than the micro-canonical ensemble. For simplicity we assume that any system under consideration has a finite number of micro-states, which means that its entropy is bounded from above. To motivate the definition below, note that the overall system Ω may have undergone an equilibration process to reach its thermodynamic equilibrium after any independently equilibrated subsystems were brought into thermal contact — and in this process the entropy was non-decreasing.

A system is defined to be in **thermodynamic equilibrium** if its entropy is maximal.

We can *derive* Eq. 19 from this definition. All we need to do is maximize the entropy $S = -\sum_i p_i \log p_i$ subject to the three micro-canonical constraints of conserved energy, conserved particle number, and well-defined probabilities $\sum_i p_i = 1$. It turns out that only the final constraint needs to be incorporated into the maximization, through the method of **Lagrange multipliers**. As a reminder, this method involves maximizing the modified entropy

$$\bar{S}(\lambda) = S + \lambda \left(\sum_{i=1}^M p_i - 1 \right) = -\sum_{i=1}^M p_i \log p_i + \lambda \left(\sum_{i=1}^M p_i - 1 \right),$$

and subsequently imposing $\sum_i p_i = 1$. Here λ is a parameter called the 'multiplier'. In short, this procedure is valid because $\frac{\partial \bar{S}}{\partial \lambda} = 0$ once we impose $\sum_i p_i = 1$, so that any extremum of \bar{S} corresponds to an extremum of $S = \bar{S}(\lambda = 0)$.

Recalling $\frac{\partial}{\partial x_k} \sum_i f(x_i) = \frac{\partial f(x_k)}{\partial x_k}$, what is the probability p_k that maximizes the modified entropy \bar{S} ?

$$0 = \frac{\partial \bar{S}}{\partial p_k} =$$

You should find that p_k is some constant that depends on λ . We don't care about λ ; so long as we know p_k is constant, then we must have $p_k = \frac{1}{M}$ in order to satisfy $\sum_k p_k = 1$. As advertised, we recover Eq. 19 from our new definition of thermodynamic equilibrium based on the second law.

2.3 Temperature

In the micro-canonical ensemble, the conserved internal energy and particle number are fundamental, while the temperature (like the entropy) is a derived quantity. As discussed below Eq. 20, in thermodynamic equilibrium such derived quantities are functions of the conserved $\{E, N\}$. In this section we will state the definition of temperature for the micro-canonical ensemble and apply this to a spin system. In the next section we will check that this definition reproduces our expectations from everyday experiences.

In thermodynamic equilibrium, the **temperature** $T(E, N)$ in the micro-canonical ensemble is defined by

$$\frac{1}{T} = \left. \frac{\partial S}{\partial E} \right|_N. \quad (22)$$

In words, the (inverse) temperature is set by the dependence of the entropy on the internal energy for a fixed number of degrees of freedom.

Since this definition is not terribly intuitive, we will again gain insight by considering N spins in a line, in a magnetic field of strength H . We saw above that $E = -H(n_+ - n_-)$ for n_+ and $n_- = N - n_+$ spins respectively pointing up and down. With N fixed, each (conserved) value of E defines a *different* micro-canonical system, which we can expect to have a different number of micro-states $M(E)$, different entropy $S(E)$ and different temperature $T(E)$. We will compute the functional forms of each of these three quantities, starting with $M(E)$.

Even though the total energy E remains fixed as time passes, individual spins can ‘flip’ between pointing up or down. Such spin flips simply have to come in pairs so that the overall n_{\pm} both remain the same. As illustration, what are representative spin configurations that produce the minimal energy $E_{\min} \equiv E_0$ and the next-to-minimal E_1 ? What are E_0 and E_1 in terms of $\{N, H\}$, and how many distinct micro-states are there for each of E_0 and E_1 ?

Your results should generalize to

$$M(E_{n_-}) = \binom{N}{n_-} = \frac{N!}{n_-! (N - n_-)!} = \binom{N}{n_+}. \quad (23)$$

To take the derivative in Eq. 22, we need to express n_- in terms of $\{E, N\}$. It will also be useful to avoid the factorial operation, which is inconvenient to differentiate. For $N \gg 1$, we can accomplish both these goals by treating the spin system as a random walk in the space of its possible energies E and applying the central limit theorem:⁶

- Each spin adds to $x \equiv \frac{E}{-H} = 2n_+ - N$ a ‘step’ of fixed ‘length’ ± 1 . Our task therefore coincides with the special case we considered in Section 1.5.
- We don’t impose any preference for positive vs. negative energies, meaning $p = q = \frac{1}{2}$ in the terminology of Section 1.5.
- With $p = q = \frac{1}{2}$, every one of the 2^N possible configurations of N spins is equally probable. Therefore the probability P_{n_+} that our overall ‘walk’ ends up producing a configuration with $n_+ = \frac{1}{2}(x + N)$ is simply the fraction of those 2^N states with this n_+ , in which we can recognize Eq. 23:

$$P_{n_+} = \frac{1}{2^N} \binom{N}{n_+} = \frac{M(E_{n_-})}{2^N} \quad \implies \quad M(E_{n_-}) = 2^N P_{n_+}.$$

- To estimate P_{n_+} for $N \gg 1$, we apply the central limit theorem just as in Section 1.5.3. In particular, we can re-use our computation that $\mu = 2p - 1 = 0$ and $\sigma^2 = 4pq = 1$, to find

$$p(x) \approx \frac{1}{\sqrt{2\pi N}} \exp\left[-\frac{x^2}{2N}\right].$$

This is the probability *distribution* from which we want to extract P_{n_+} .

⁶Applying Stirling’s formula, $\log(N!) \approx N \log N - N$, is another possible approach.

- From a tutorial problem we know that $P_{\text{const}}(n_+) = p(2n_+ - N)\Delta n_+$ is a good approximation. With $\Delta n_+ = 1$ and $2n_+ - N = \frac{E}{-H}$, we therefore find

$$M(E) \approx 2^N p(2n_+ - N) = \frac{2^N}{\sqrt{2\pi N}} \exp\left[-\frac{E^2}{2NH^2}\right]. \quad (24)$$

What is the derivative of the log of Eq. 24 with N fixed?

$$\left. \frac{\partial}{\partial E} \log M \right|_N =$$

You should find the temperature

$$T \approx -\frac{NH^2}{E} \quad N \gg 1, \quad (25)$$

which in several ways does *not* seem to match our expectations from everyday experiences: This T diverges as $E \rightarrow 0$ for $n_+ \approx n_-$, and it is negative whenever $n_+ < n_-$ to produce $E > 0$. You can check that this $T < 0$ corresponds to the number of micro-states decreasing for larger internal energies, $\frac{\partial M}{\partial E} < 0$. In **natural** systems, larger energies make more micro-states accessible, producing $\frac{\partial M}{\partial E} > 0$ and a positive temperature. If $H = 0$, we also have $E = 0$ and T is ill-defined.

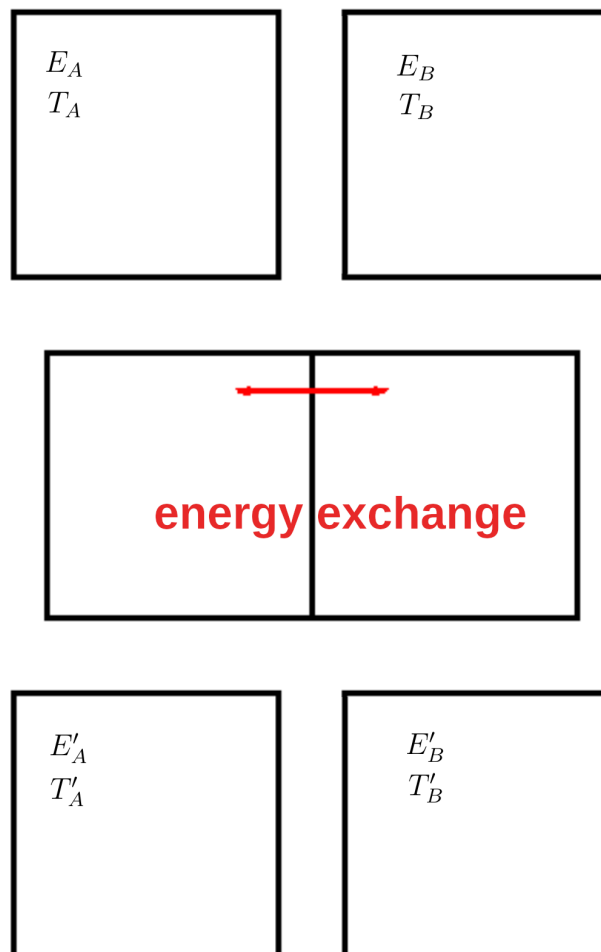
Restricting our attention to $H > 0$ and $n_+ > n_-$, we also see that the resulting non-negative temperature cannot vanish. It is minimized by the most-negative energy you found above, $T_{\text{min}} = H > 0$ for $E_{\text{min}} = -NH$. The non-zero minimum temperature is specific to spin systems, while some of the other oddities result from the micro-canonical approach more generally. This will motivate turning to the canonical ensemble in Unit 3, but first we can check that some aspects of the micro-canonical temperature defined in Eq. 22 do match our everyday expectations, at least in the ‘natural’ positive-temperature regime.

2.4 Heat exchange

From Eq. 25 for the temperature of a micro-canonical spin system, we can see that ‘natural’ positive temperatures correspond to negative energies, and therefore increase as the energy increases by becoming less negative (with a smaller magnitude). Such a direct relation between energy and temperature is very generic, and we will study it in more detail when considering thermodynamic cycles in a few weeks. For now, considering unspecified systems that exhibit this natural behaviour, let’s ask what

would happen if we take two initially isolated micro-canonical systems — Ω_A and Ω_B with temperatures T_A and T_B in thermodynamic equilibrium — and bring them into thermal contact.

In micro-canonical terms, the temperatures T_A and T_B are derived from the corresponding energies E_A and E_B , while thermal contact allows the two systems to exchange energy (but not particles) as non-isolated subsystems of a combined micro-canonical system Ω_C . Once the two subsystems have been in thermal contact long enough for the combined system to have reached thermodynamic equilibrium, it will have temperature T_C . We can then re-isolate the two subsystems, which will remain in thermodynamic equilibrium with energies $\{E'_A, E'_B\}$ and temperatures $\{T'_A, T'_B\}$. This three-step procedure is illustrated below.



From everyday experience, we expect that this energy exchange will result in a net flow of energy from the hotter system to the colder system, cooling the former by heating the latter. We will now check that the micro-canonical definition of temperature in Eq. 22 predicts this expected behaviour. With $S \in \{A, B\}$, we can write

$$E'_S = E_S + \Delta E_S$$

and consider for simplicity the case where the change in energy is relatively small,

$$\left| \frac{\Delta E_S}{E_S} \right| \ll 1.$$

Since we can build up large changes in energy through a series of smaller changes, this assumption doesn't lead to any loss of generality. We also know $\Delta E_B = -\Delta E_A$ thanks to conservation of energy.

Equation 22 tells us that we need to consider the entropies as functions of E_S and E'_S in order to connect the temperatures to any flow of energy. Because we don't change the number of particles in each system, we only need to consider the energy dependence of the entropy. We assume $S(E)$ is continuous and infinitely differentiable,⁷ which allows us to expand each of the final entropies $S(E'_S)$ in a Taylor series,

$$S(E'_S) = S(E_S + \Delta E_S) \approx S(E_S) + \left. \frac{\partial S}{\partial E} \right|_{E_S} \Delta E_S,$$

neglecting all $\mathcal{O}(\Delta E_S^2)$ terms because we consider relatively small changes in energy. What is the expression above in terms of the initial temperatures T_S ?

From the second law of thermodynamics, we know that the total entropy of these systems can never decrease as time passes:

$$S(E_A) + S(E_B) \leq S(E_A + E_B) = S(E'_A) + S(E'_B). \quad (26)$$

The final equality means that re-isolating the two subsystems doesn't change the entropy. This is because E'_A is not fixed and could take any value from zero to $E_A + E_B$ at the moment when the subsystems are re-isolated. Computing the final entropy $S(E'_A) + S(E'_B)$ therefore requires summing over all possible values of E'_A , producing exactly the sum in Eq. 21 for the overall system. We will see something similar when we consider the 'Gibbs paradox' in Unit 4.

What do you find when you insert your linearized Taylor series into Eq. 26?

⁷This assumption breaks down at a *phase transition*, where we would need to be more careful. We will learn about phase transitions towards the end of the term.

Applying conservation of energy should produce

$$\left(\frac{1}{T_A} - \frac{1}{T_B}\right) \Delta E_A \geq 0.$$

Recalling from Section 2.2.3 that equality holds only in extremely special cases, we can identify three possibilities consistent with this result. If $T_A > T_B$, then $\left(\frac{1}{T_A} - \frac{1}{T_B}\right)$ is negative and we will generically have $\Delta E_A < 0$, so that energy flows out of the hotter system Ω_A and into the colder one. Our restriction to natural systems means this flow of energy reduces the higher temperature, and increases the lower temperature, bringing the temperatures of the two subsystems closer to each other. Similarly, if $T_A < T_B$, we will generically have $\Delta E_A > 0$, meaning that energy still flows from the hotter system Ω_B into the colder one, again reducing the difference in their temperatures. We can finally conclude that $T_A = T_B$ is the very special case where there is no energy flow, $\Delta E_S = 0$, keeping the temperatures the same. All of this is exactly what we would expect based on our everyday experience of temperature as an intensive quantity.

Unit 3: Canonical ensemble

3.1 The thermal reservoir

3.1.1 Replicas and occupation numbers

While it is reasonable to forbid particle exchange, for example by sealing gases inside airtight containers, it is not practical to prevent energy exchange as would be needed to fully isolate statistical systems. Any thermal insulator is imperfect, and any observation of the system would require exchanging energy with the external observer. In practice it is more convenient to work with physical systems that are characterized by their (intensive) temperatures rather than their (extensive) internal energies.

This leads us to define the **canonical ensemble** to be a statistical ensemble characterized by its fixed temperature T and conserved particle number N , with the temperature held fixed through contact with a **thermal reservoir**.

The second part of this definition connects the fixed temperature to the fundamental fact of energy conservation (the first law of thermodynamics). This is done by proposing that our system of interest Ω is in thermal contact with a much larger external system Ω_{res} — the thermal reservoir, sometimes called a “heat bath”. The overall combined system $\Omega_{\text{tot}} = \Omega_{\text{res}} \otimes \Omega$ is governed by the micro-canonical ensemble, with conserved total energy $E_{\text{tot}} = E_{\text{res}} + E \approx E_{\text{res}}$, while the energy E of Ω is allowed to fluctuate. The key qualitative idea is that, in thermodynamic equilibrium, Ω has a negligible effect on the overall system. In particular, the temperature of that overall system — and therefore the temperature of Ω , by intensivity — is set by the reservoir and remains fixed even as E fluctuates. This effectively generalizes the setup we used to analyse heat exchange in the previous section, where we saw that thermal contact causes a net flow of energy from hotter systems to colder systems. When these systems are ‘natural’, this cools the hotter one by heating the colder one.

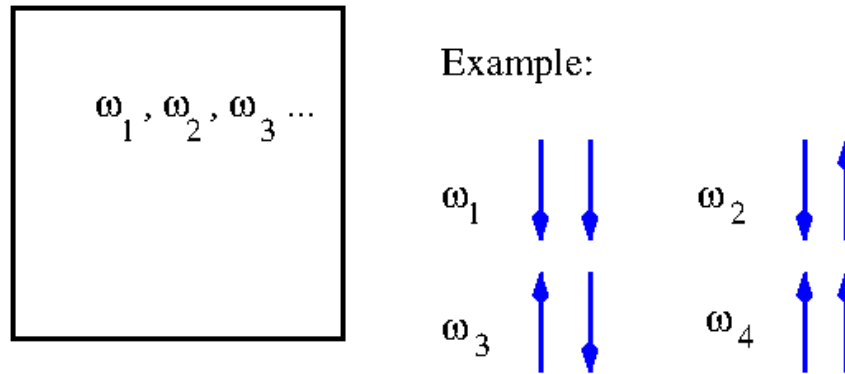
The mathematical implementation of this argument, as developed by Gibbs, proceeds by considering a well-motivated ansatz for the form of the thermal reservoir Ω_{res} . **The goal**, which will be useful to keep in mind as we go through the lengthy analysis, is to show that the specific form of Ω_{res} is ultimately irrelevant. This will allow us to work directly with the system of interest, Ω , independent of the details of the thermal reservoir that fixes its temperature.

Without further ado, we take Ω_{tot} to consist of many ($R \gg 1$) identical **replicas** of the system Ω that we’re interested in. All of these replicas are in thermal contact with each other, and in thermodynamic equilibrium.⁸ Choosing any one of the replicas to be the system of interest, Ω , the other $R - 1 \approx R$ replicas collectively form the thermal

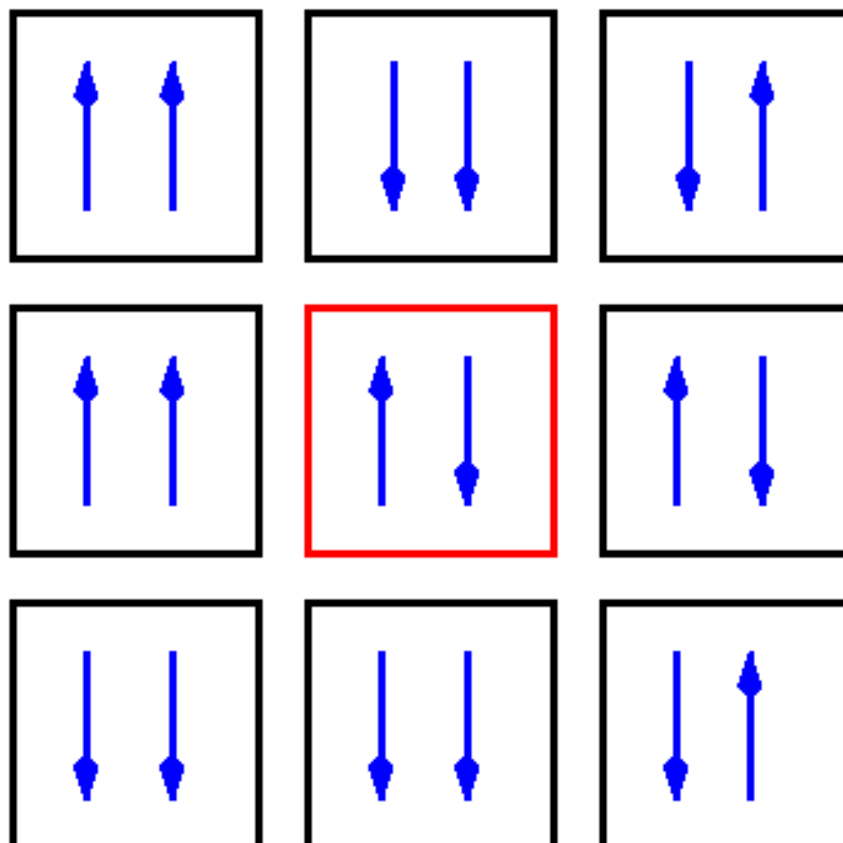
⁸The thermal contact between any two replicas can be indirect, mediated by a sequence of intermediate replicas. This transitivity of thermodynamic equilibrium is sometimes called the **zeroth law of thermodynamics**. It declares that if systems Ω_A & Ω_B are in thermodynamic equilibrium while systems Ω_B & Ω_C are in thermodynamic equilibrium, then Ω_A & Ω_C must also be in thermodynamic equilibrium.

reservoir Ω_{res} . Assuming we want to study reasonable systems Ω , this ansatz ensures that Ω_{res} is also reasonable, simply much larger.

An extremely small example of this setup is illustrated by the figures below, where the system of interest is just $N = 2$ spins. For now we assume the spins are *distinguishable*, so that $\downarrow\uparrow$ and $\uparrow\downarrow$ are both distinct micro-states. This means that each individual replica has the $M = 4$ micro-states ω_i defined below.



To form the overall system Ω_{tot} we now bring together the $R = 9$ replicas shown below. We draw boxes around each replica to remind us that they are allowed to exchange only energy with each other, while the $N = 2$ spins per replica are fixed in place. We pick out one of these replicas (in the red box) to serve as the system Ω we will consider. The other 8 are the thermal reservoir Ω_{res} that fixes the temperature of Ω .



A convenient way to analyse the overall system of R replicas, Ω_{tot} , is to define the **occupation number** n_i to be the number of replicas that adopt the micro-state $\omega_i \in \Omega$ in any given micro-state of Ω_{tot} . The index $i \in \{1, 2, \dots, M\}$ runs over all M micro-states of Ω . In the example above, three of the replicas have the micro-state $\omega_1 = \downarrow\downarrow$, meaning $n_1 = 3$. What are the occupation numbers $\{n_2, n_3, n_4\}$ for the other three ω_i in the figures above? Are all replicas accounted for, $\sum_i n_i = R$?

Normalizing the occupation number by R gives us a well-defined *occupation probability*, $p_i = n_i/R$ with $\sum_i p_i = 1$. At the moment, this p_i is the probability that if we choose a replica at random it will be in micro-state ω_i .

Now let us consider conservation of energy, which continues to apply to the total energy E_{tot} of the overall system Ω_{tot} . We assume that each replica's energy E_r is independent of all the other replicas. This is guaranteed for the non-interacting systems we will focus on until Unit 9, and also holds when interactions are allowed within each replica but not between different replicas. The thermal contact between replicas allows E_r to fluctuate subject to conservation of E_{tot} , but there are at most M possible values E_i it can have, corresponding to the M micro-states $\omega_i \in \Omega$. Some distinct micro-states $\omega_i \neq \omega_j$ may have the same energy $E_i = E_j$, which doesn't affect the analysis. This allows us to rearrange a sum over replicas into a sum over the micro-states of Ω :

$$E_{\text{tot}} = \sum_{r=1}^R E_r = \sum_{i=1}^M n_i E_i, \quad (27)$$

with the occupation number n_i counting how many times micro-state ω_i appears among the R replicas. We can assume that R and M are both finite, so we don't need to worry about rearranging the sums.

3.1.2 Partition function

Following Gibbs, we've taken the thermal reservoir Ω_{res} to consist of $R - 1$ replicas of the system of interest, Ω . The next step is to further simplify the mathematics by assuming that the overall R -replica system Ω_{tot} is fully specified by a fixed set of M occupation numbers $\{n_i\}$. This is equivalent to assuming that the occupation probabilities $\{p_i\}$ are constant in time, as a reflection of thermodynamic equilibrium. From Eq. 27, we see that this ensures conservation of the total energy E_{tot} , and we can apply the micro-canonical tools we developed in the previous unit. Recall our ultimate goal of showing that such details of the thermal reservoir are irrelevant to the system Ω .

Based on the conservation of E_{tot} , we want to determine the (intensive) temperature of Ω_{tot} , which fixes the temperature of the system of interest, Ω . According to our previous work, to do this we first need to compute the overall number of micro-states

M_{tot} as a function of E_{tot} , from which we can derive the micro-canonical entropy and temperature since the system is in thermodynamic equilibrium. From the fixed occupation numbers n_i , we already know how many times each micro-state ω_i appears among the R replicas. To determine M_{tot} we just need to count how many possible ways there are of distributing the $\{n_i\}$ micro-states among the R replicas.

If we consider first the micro-state ω_1 , the number of possible ways of distributing n_1 copies of this micro-states among the R replicas is just the binomial coefficient

$$\binom{R}{n_1} = \frac{R!}{n_1! (R - n_1)!}.$$

Moving on to ω_2 , we need to keep in mind that n_1 replicas have already been assigned micro-state ω_1 , so there are only $R - n_1$ replicas left to choose from. What is the resulting number of possible ways of distributing these n_2 micro-states?

Repeating this process for all micro-states $\{\omega_1, \omega_2, \dots, \omega_M\}$, recalling $(R - \sum_i n_i)! = 0! = 1$, you should obtain a product that ‘telescopes’ to

$$M_{\text{tot}} = \frac{R!}{n_1! n_2! \dots n_M!}. \quad (28)$$

This confirms that the order in which we assign micro-states to replicas is irrelevant, since integer multiplication is commutative.

Thanks to thermodynamic equilibrium, the entropy of the micro-canonical Ω_{tot} is

$$S(E_{\text{tot}}) = \log M_{\text{tot}} = \log(R!) - \sum_{i=1}^M \log(n_i!),$$

where the dependence on E_{tot} enters through the occupation numbers via Eq. 27. With $R \gg 1$ and $n_i \gg 1$ for all $i = 1, \dots, M$, we can approximate each of these logarithms using the first two terms in [Stirling’s formula](#),

$$\log(N!) = N \log N - N + \mathcal{O}(\log N) \approx N \log N - N \quad \text{for } N \gg 1.$$

In order for *every* occupation number to be large, $n_i \gg 1$, the number of replicas must be much larger than the number of micro-states of Ω . As we have discussed before, the number of micro-states M is typically a very large number, so with $R \gg M$ we are formally considering truly enormous thermal reservoirs! This enormity helps ensure that the detailed form of the reservoir will be irrelevant.

Using the approximation above, what do you find for $S(E_{\text{tot}})$ in terms of R and n_i ? What is the entropy in terms of the occupation probabilities $p_i = n_i/R$?

$$S(E_{\text{tot}}) = \log(R!) - \sum_{i=1}^M \log(n_i!) \approx$$

In your result, the dependence on E_{tot} now enters through the occupation probabilities p_i . In order to determine the temperature, we have to express the entropy explicitly in terms of E_{tot} . We do this by applying our knowledge that thermodynamic equilibrium implies maximal entropy.

Following the same steps as in Section 2.2.3, we maximize the entropy, now with two Lagrange multipliers to account for two constraints on the occupation probabilities:

$$\sum_{i=1}^M p_i = 1 \qquad \sum_{i=1}^M n_i E_i = R \sum_{i=1}^M p_i E_i = E_{\text{tot}}.$$

Writing everything in terms of occupation probabilities, we therefore need to maximize the modified entropy

$$\bar{S} = -R \sum_{i=1}^M p_i \log p_i + \alpha \left(\sum_{i=1}^M p_i - 1 \right) - \beta \left(R \sum_{i=1}^M p_i E_i - E_{\text{tot}} \right).$$

Here we've chosen the sign of β for later convenience. What is the occupation probability p_k that maximizes \bar{S} ?

$$0 = \frac{\partial \bar{S}}{\partial p_k} =$$

By defining a new parameter Z in terms of α , you should find

$$p_k = \frac{1}{Z} e^{-\beta E_k}. \quad (29)$$

As before, we need to fix the parameters $\{Z, \beta\}$ by demanding that the two constraints above are satisfied. The first of these constraints is straightforward and produces an important result:

$$1 = \sum_{i=1}^M p_i = \frac{1}{Z} \sum_{i=1}^M e^{-\beta E_i} \implies Z(\beta) = \sum_{i=1}^M e^{-\beta E_i}. \quad (30)$$

Eq. 30 defines the canonical **partition function** $Z(\beta)$, a fundamental quantity in the canonical ensemble, from which many other derived quantities can be obtained.

$Z(\beta)$ still depends on the other as-yet-unknown parameter $\beta(E_{\text{tot}})$. By applying our second constraint, Eq. 27, we can relate β to E_{tot} :

$$E_{\text{tot}} = R \sum_{i=1}^M p_i E_i = \frac{R}{Z(\beta)} \sum_{i=1}^M E_i e^{-\beta E_i} = R \frac{\sum_{i=1}^M E_i e^{-\beta E_i}}{\sum_{j=1}^M e^{-\beta E_j}}. \quad (31)$$

This relation is a bit complicated, but will suffice for our goal of expressing the entropy in terms of E_{tot} . Inserting Eq. 29 for p_i into your earlier result for the entropy, what do you obtain upon applying Eqs. 30 and 31?

$$S(E_{\text{tot}}) = -R \sum_{i=1}^M p_i \log p_i =$$

There is a pleasant simplification when we take the derivative to determine the temperature. Defining $\beta' \equiv \frac{\partial}{\partial E_{\text{tot}}} \beta(E_{\text{tot}})$, we have

$$\frac{1}{T} = \frac{\partial}{\partial E_{\text{tot}}} S(E_{\text{tot}}) = \frac{\partial}{\partial E_{\text{tot}}} [E_{\text{tot}} \beta + R \log Z(\beta)] = \beta + E_{\text{tot}} \beta' + R \frac{1}{Z} \frac{\partial Z(\beta)}{\partial \beta} \beta'.$$

Using Eq. 31 we can compute

$$\frac{1}{Z} \frac{\partial Z(\beta)}{\partial \beta} = \frac{1}{Z} \frac{\partial}{\partial \beta} \sum_{i=1}^M e^{-\beta E_i} = -\frac{1}{Z} \sum_{i=1}^M E_i e^{-\beta E_i} = -\sum_{i=1}^M p_i E_i = -\frac{E_{\text{tot}}}{R},$$

so that we don't need to figure out the explicit form of β' :

$$\frac{1}{T} = \beta + E_{\text{tot}}\beta' - E_{\text{tot}}\beta' = \beta. \quad (32)$$

What's truly remarkable about Eqs. 29, 30 and 32 is that they make no reference to the R replicas or any extensive quantity of the overall system, such as E_{tot} — all information about the thermal reservoir has vanished. This is **the goal** we have been pursuing since the start of this unit! The large thermal reservoir is still present to fix the temperature T characterizing the canonical system Ω , but beyond that nothing about it is relevant (or even knowable) in the canonical approach. Every aspect of Ω can now be specified in terms of its fixed temperature T and conserved particle number N , starting with the parameter $\beta = 1/T$.

In particular, the partition function from Eq. 30 is simply

$$Z(T) = \sum_{i=1}^M e^{-E_i/T}. \quad (33)$$

and together with β specifies the probabilities

$$p_i = \frac{1}{Z} e^{-E_i/T} \quad (34)$$

from Eq. 29. This p_i is now the thermodynamic equilibrium probability that Ω adopts micro-state ω_i with (non-conserved) internal energy E_i . This probability distribution is called either the **Boltzmann distribution** or the **Gibbs distribution**, while $e^{-E_i/T}$ itself is known as a **Boltzmann factor**. All micro-states with the same energy have the same probability in thermodynamic equilibrium, which is consistent with the micro-canonical behaviour we saw in Unit 2.

3.2 Internal energy, heat capacity, and entropy

In addition to fixing the temperature of the system Ω , the thermal reservoir also allows the internal energy of Ω to fluctuate. The system simply exchanges energy with the reservoir, satisfying the first law of thermodynamics. Although the internal energy fluctuates, its expectation value $\langle E \rangle$ is an important derived quantity in thermodynamic equilibrium. Applying the general definition from Eq. 4 to the probability space of the canonical ensemble,

$$\langle E \rangle(T) = \sum_{i=1}^M E_i p_i = \frac{1}{Z} \sum_{i=1}^M E_i e^{-\beta E_i}.$$

Here we highlight the dependence of $\langle E \rangle$ on the temperature, and also freely interchange $\beta = 1/T$.

The expression above may look familiar from our work in the previous section:

$$\frac{\partial}{\partial \beta} \log Z =$$

In this case it is easier to take the derivative with respect to β as opposed to

$$\frac{\partial}{\partial \beta} = \frac{\partial T}{\partial \beta} \frac{\partial}{\partial T} = -\frac{1}{\beta^2} \frac{\partial}{\partial T} = -T^2 \frac{\partial}{\partial T}. \quad (35)$$

In Section 2.3, we saw that ‘natural’ micro-canonical systems exhibit higher (derived) temperatures for larger (conserved) internal energies. Here, in the canonical approach, the average internal energy $\langle E \rangle$ is the derived quantity while the temperature is fixed. From our everyday experience, we expect a similar direct relation between temperature and energy, which the following result confirms.

The **heat capacity** is defined to be

$$c_v = \frac{\partial}{\partial T} \langle E \rangle, \quad (36)$$

and is always non-negative, $c_v \geq 0$.

The subscript indicates that the volume of the system is kept fixed; we’ll consider the role of the volume more carefully starting in Unit 4. In a homework assignment you will confirm $c_v \geq 0$ by deriving a **fluctuation–dissipation** (or **fluctuation–response**) **relation**. That relation will be a special case of a [more general theorem](#), and will connect the fluctuations of the internal energy around its expectation value, $(E_i - \langle E \rangle)^2$, to the energy’s *response* to a change in temperature, $\frac{\partial}{\partial T} \langle E \rangle$. Only extremely special cases will produce $c_v = 0$, meaning that the heat capacity is generically positive, in agreement with our intuition that higher temperatures produce larger internal energies.

Finally, we can compute the entropy of Ω with no reference to the thermal reservoir, apart from its role fixing the temperature in thermodynamic equilibrium. Since the general definition of the entropy in Eq. 20 continues to hold for the canonical ensemble, we just need to insert the probabilities p_i from Eq. 34:

$$S(T) = - \sum_{i=1}^M p_i \log p_i =$$

You should find that the entropy depends on $\log Z$.

3.3 Helmholtz free energy

This dependence of the entropy on $\log Z$ is in accordance with our earlier claim that the partition function is a fundamental quantity in the canonical ensemble. Recalling from Eq. 33 that Z is a sum over all micro-states, we can view this result as the canonical counterpart to the micro-canonical entropy being the logarithm of the number of micro-states. (Thermodynamic equilibrium is required in both cases.) This motivates the following definition of a quantity with the dimensions of energy that is related to $\log Z$, which provides simpler and more elegant expressions for the derived quantities we considered above.

The **Helmholtz free energy** of a system in the canonical ensemble is

$$F(T) = -T \log Z(T) \qquad F(\beta) = -\frac{\log Z(\beta)}{\beta}, \qquad (37)$$

where Z is the partition function of the system. In terms of this free energy, Eqs. 33 and 34 are

$$Z = e^{-F/T} \qquad p_i = e^{(F-E_i)/T}.$$

The Helmholtz free energy is named after [Hermann von Helmholtz](#) and reveals its usefulness when we take its derivative. The derivative involves $\frac{\partial}{\partial T} \log Z$, which is worth collecting in advance based on Eq. 35:

$$-\frac{\partial}{\partial T} \left(\frac{F(T)}{T} \right) = \frac{\partial}{\partial T} \log Z(T) =$$

$$\frac{\partial}{\partial T} F(T) =$$

From these results we can read off the more elegant expressions promised above:

$$S(T) = -\frac{\partial}{\partial T} F(T) \tag{38}$$

$$\langle E \rangle(T) = -T^2 \frac{\partial}{\partial T} \left(\frac{F(T)}{T} \right) = \frac{\partial}{\partial \beta} [\beta F(\beta)] = TS(T) + F(T). \tag{39}$$

3.4 The physics of information

As a first application of the canonical ensemble, we will explore physically observable effects that depend on the information content of a statistical system. A famous scientific question illustrating the importance of information, which you may have heard of, is the [black hole information paradox](#). However, that topic is well beyond the scope of this module since it depends on quantum mechanics and general relativity in addition to statistical mechanics. Here we will consider simple spin systems as introduced in Section 2.1, contrasting the behaviour of their average internal energy $\langle E \rangle$ and entropy S depending on whether or not the spins can (in principle) be distinguished from each other. It's important to appreciate that the "information" discussed here is an intrinsic property of the system — what is *knowable* about it in principle. It does not matter whether or not any observer actually knows this information; so long as it can possibly be known it will have an effect.

3.4.1 Distinguishable spins in a solid

We begin with the setup from Section 2.1: A system of N spins arranged in a line, placed in an external magnetic field of strength H , and in thermodynamic equilibrium. We further specify that the spins are embedded in a solid material that fixes their positions and prevents them from moving. This allows them to be distinguished from one another: An observer can target an appropriate position in the solid to measure the corresponding spin. Each spin measured in this way will be either parallel or

anti-parallel to the magnetic field. The canonical system therefore has $M = 2^N$ distinct micro-states ω_i with energies E_i and probabilities $p_i = \frac{1}{Z} e^{-E_i/T}$, each defined by the orientations of all N spins.

To streamline our notation, we can represent the orientation of the n th spin as $s_n \in \{1, -1\}$, where $s_n = 1$ indicates alignment parallel to the field and $s_n = -1$ indicates alignment anti-parallel to the field. With this notation, the internal energy of the system in micro-state ω_i specified by the N spins $\{s\}$ is simply

$$E_i = -H \sum_{n=1}^N s_n. \quad (40)$$

To compute the canonical partition function Z_D , where the subscript reminds us of the spins' distinguishability, we have to sum over all 2^N possible spin configurations $\{s\}$. In this process we can save some space by defining the dimensionless variable $x = \beta H = \frac{H}{T}$:

$$\begin{aligned} Z_D &= \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} e^{-\beta E_i} = \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} \exp \left[x \sum_{n=1}^N s_n \right] \\ &= \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} e^{x s_1} \cdots e^{x s_N} = \left(\sum_{s_1=\pm 1} e^{x s_1} \right) \cdots \left(\sum_{s_N=\pm 1} e^{x s_N} \right) \\ &= \left(\sum_{s=\pm 1} e^{x s} \right)^N = (e^x + e^{-x})^N = [2 \cosh(\beta H)]^N, \end{aligned} \quad (41)$$

distributing the summations since all the spins are independent of each other.

The corresponding Helmholtz free energy

$$F_D(\beta) = -\frac{\log Z(\beta)}{\beta} = -\frac{N \log [2 \cosh(\beta H)]}{\beta} \quad (42)$$

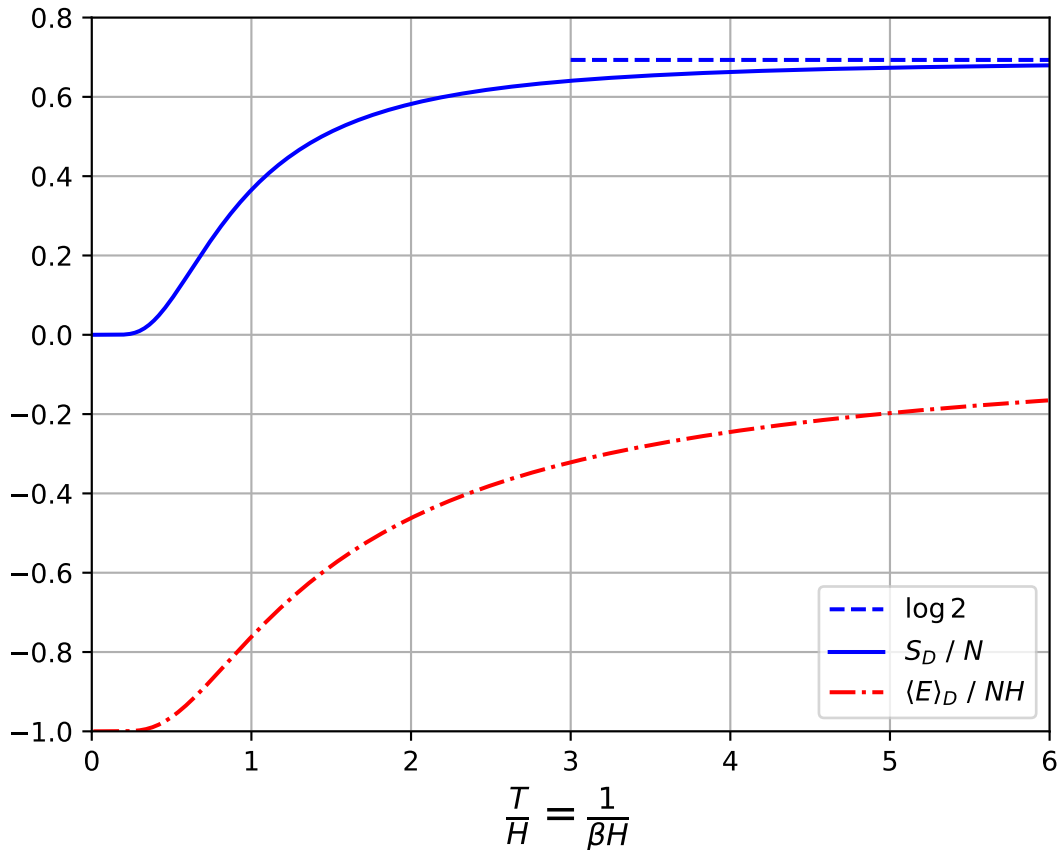
is all we need to compute the average internal energy:

$$\langle E \rangle_D = \frac{\partial}{\partial \beta} [\beta F_D(\beta)] =$$

From this we immediately obtain the entropy

$$S_D = \beta (\langle E \rangle_D - F_D) = -N \beta H \tanh(\beta H) + N \log [2 \cosh(\beta H)]. \quad (43)$$

These results for $\langle E \rangle_D$ and S_D are plotted on the next page as functions of $\frac{T}{H} = \frac{1}{\beta H}$, using [this Python code](#). Since both these quantities are extensive, we normalize them by showing $\frac{\langle E \rangle_D}{NH}$ and $\frac{S_D}{N}$.



Let's analyse the asymptotic behaviour of these functions, starting with **low temperatures**. In contrast to the micro-canonical Eq. 25, in the canonical ensemble there is no issue with taking the independent variable $T \rightarrow 0$. This corresponds to $\beta H \rightarrow \infty$ and $\tanh(\beta H) \rightarrow 1$, approaching the “ground-state” energy $E_{\min} = E_0 = -NH$ you computed in Section 2.3. This energy is only produced by the single micro-state in which all the spins are aligned with the magnetic field, $s_n = 1$ for all n (or $n_+ = N$ and $n_- = 0$ in the notation from Section 2.3). Correspondingly, $\log [2 \cosh(\beta H)] \rightarrow \log e^{\beta H} = \beta H$ and the two terms in Eq. 43 cancel out, so that $S_D \rightarrow 0$. This vanishing entropy is a generic consequence of temperatures approaching *absolute zero*.

For low but non-zero temperatures, $\langle E \rangle_D$ and S_D will be affected by the non-zero probability for the system to adopt micro-states ω_i with higher energies $E_i > E_0$. These higher-energy configurations are often referred to as “excited states”. Note that each energy $E_i > E_0$ may correspond to many different micro-states. For example, in Section 2.3 you also computed the energy $E_1 = -(N - 2)H$ of the first excited state, which is realized by the N distinct micro-states with $n_- = 1$. In the case of spin systems, we can instead refer to *energy levels* that are all separated by a constant *energy gap* $\Delta E \equiv E_{n_-+1} - E_{n_-} = 2H$.

We can compute the effects of the higher energy levels at low temperatures $\beta H \gg 1$ by expanding $\langle E \rangle_D$ in powers of $e^{-\beta H} \ll 1$. What is the first temperature-dependent term in this expansion?

$$\frac{\langle E \rangle_D}{NH} =$$

You should find that the excited-state effects are *exponentially* suppressed by the energy gap ΔE at low temperatures,

$$\frac{\langle E \rangle_D}{NH} = -1 + 2e^{-\beta\Delta E} + \mathcal{O}(e^{-2\beta\Delta E}).$$

This is a generic feature of canonical systems with a non-zero energy gap, and is due to the exponentially suppressed probability for the system to adopt any of the microstates with the higher energy,

$$\frac{\frac{1}{Z}e^{-\beta E_{n_{-}+1}}}{\frac{1}{Z}e^{-\beta E_{n_{-}}}} = e^{-\beta\Delta E}.$$

The low-temperature expansion of Eq. 43 for the entropy in powers of $e^{-\beta H} \ll 1$ is similar:

$$\frac{S_D}{N} =$$

Here the leading term includes a linear factor of $\beta\Delta E \gg 1$, but this can't overcome the now-expected exponential suppression:

$$\frac{S_D}{N} = \beta\Delta E e^{-\beta\Delta E} + e^{-\beta\Delta E} + \mathcal{O}(\beta\Delta E e^{-2\beta\Delta E}).$$

In the limit of **high temperatures** we should instead expand in powers of the small factor $\beta H \ll 1$. This is straightforward for $\langle E \rangle_D$:

$$\frac{\langle E \rangle_D}{NH} = -\tanh(\beta H) = -\beta H + \frac{(\beta H)^3}{3} + \mathcal{O}([\beta H]^5),$$

which vanishes $\sim \frac{1}{T}$ as $T \rightarrow \infty$. This matches the micro-canonical behaviour we saw for this system from Eq. 25, where the derived temperature diverged as the conserved energy approached zero.

For the entropy, there is a similar connection to micro-canonical behaviour at high temperatures:

$$\frac{S_D}{N} =$$

As $\frac{T}{H} \rightarrow \infty$, the result

$$\frac{S_D}{N} = \log 2 - \frac{(\beta H)^2}{2} + \mathcal{O}([\beta H]^4)$$

approaches the asymptotic value $S_D \rightarrow N \log 2 = \log M$ for the $M = 2^N$ micro-states (with different energies). Conceptually, in this limit the energy of each spin is negligible compared to the temperature, and the system approximately behaves as though the energy were zero for all micro-states (and hence conserved).

3.4.2 Indistinguishable spins in a gas

Next, let's consider nearly the same setup, with N spins in thermodynamic equilibrium, in an external magnetic field of strength H . The only difference is that now the spins are allowed to move, like particles in a one-dimensional gas. We demand that they move slowly, so that we can ignore their kinetic energy and the total energy of the system continues to be given by Eq. 40. Since the spins don't interact with each other, they can freely move past each other, and even occupy the same space, making it impossible for them to be distinguished from one another.

To compute the fundamental canonical partition function (Eq. 33), we have to sum over the micro-states of the system. These micro-states are no longer in one-to-one correspondence with the full configurations $\{s\}$ of the N spins. Because the spins are now indistinguishable, certain spin configurations also cannot be distinguished from each other. The simplest example comes from the two-spin system considered in Section 3.1.1, where the configurations $\downarrow\uparrow$ and $\uparrow\downarrow$ now both correspond to a single

micro-state. In this micro-state, we know only that one spin is $s_i = 1$ while the other is $s_k = -1$; it's not possible to distinguish which is which.

Generalizing, we can conclude that a single distinct micro-state corresponds to all possible permutations of spins with fixed $\{n_+, n_-\}$. This means that each micro-state is now in one-to-one correspondence with the energy $E = -H(n_+ - n_-)$, which we can continue to organize as energy levels separated by a constant energy gap $\Delta E = 2H$. As a quick example, enumerate the energy levels when $N = 4$ and list the spin configurations associated with the corresponding micro-states. How many micro-states are there for N spins?

A convenient way to label these micro-states and energy levels is to define

$$E_k = -NH + 2Hk = -H(N - 2k)$$

for micro-state ω_k with $k = n_- = 0, \dots, N$. To compute the partition function Z_I , with the subscript reminding us about the spins' indistinguishability, we now have

$$Z_I = \sum_{k=0}^N e^{-\beta E_k} = \sum_{k=0}^N e^{\beta H(N-2k)} = e^{N\beta H} \sum_{k=0}^N (e^{-2\beta H})^k = e^{N\beta H} \frac{1 - e^{-2(N+1)\beta H}}{1 - e^{-2\beta H}}. \quad (44)$$

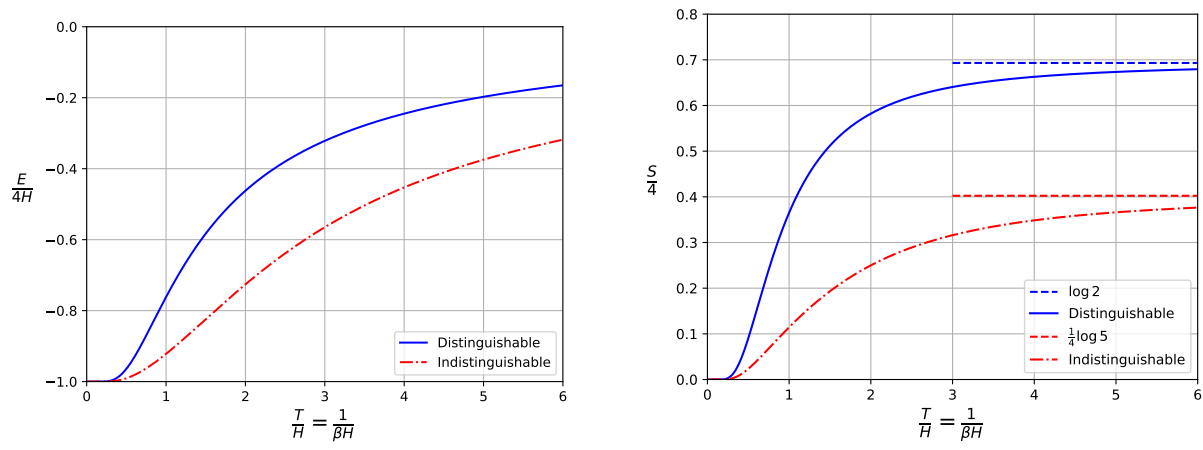
The geometric series in the last step can be reconstructed by considering

$$\sum_{k=0}^N x^k = \sum_{k=0}^{\infty} x^k - \sum_{k=N+1}^{\infty} x^k = \frac{1}{1-x} - x^{N+1} \sum_{\ell=0}^{\infty} x^{\ell} = \frac{1}{1-x} - \frac{x^{N+1}}{1-x}.$$

The corresponding Helmholtz free energy is

$$F_I(\beta) = -\frac{\log Z_I(\beta)}{\beta} = -NH - \frac{\log [1 - e^{-2(N+1)\beta H}]}{\beta} + \frac{\log [1 - e^{-2\beta H}]}{\beta}. \quad (45)$$

In contrast to Eq. 42, $F_I(\beta)$ is no longer proportional to N . In a homework assignment you will use F_I to determine the average internal energy $\langle E \rangle_I$ and entropy S_I shown in the figures on the next page, and also analyse the low- and high-temperature expansions like we did for the distinguishable case above. Unlike our results for the distinguishable case, you will find that $\frac{\langle E \rangle_I}{NH}$ and $\frac{S_I}{N}$ depend on N , which requires us to fix $N = 4$ in the plots on the next page.



The solid blue lines in these figures are exactly the distinguishable-spin results we previously discussed. The red dash-dotted lines are the new results for indistinguishable spins. We see that the same $T \rightarrow 0$ limits are approached in both cases: $E \rightarrow -NH$ and $S \rightarrow 0$. At low temperatures, the indistinguishable results approach these limits more quickly — they still feature exponential suppression of excited-state effects by the energy gap, $\propto e^{-\beta\Delta E}$, but this now comes with additional factors of N .

At high temperatures there is an even more striking difference. While the average internal energy $\langle E \rangle_I$ continues to vanish $\sim \frac{1}{T}$ as $T \rightarrow \infty$ (with different N dependence), the entropy approaches the asymptotic value $S_I \rightarrow \log(N + 1) = \log M$ for the $M = N + 1$ micro-states. This logarithmic dependence on N is very different from the $S_D \rightarrow N \log 2$ limit we found for distinguishable spins, and reflects the exponentially smaller number of micro-states that exist for indistinguishable spins, $N + 1$ vs. 2^N .

Finally, away from those low- and high-temperature limits, the left figure above shows a significant difference in the internal energies of the spin systems, depending only on whether or not the spins can be distinguished from each other in principle. This is a physically measurable effect caused by the intrinsic information content of a statistical system, and a simple illustration of phenomena that remain at the leading edge of ongoing research. As [Rolf Landauer](#) put it in a famous 1991 essay: [Information is physical](#).

Unit 4: Ideal gases

4.1 Volume, energy levels, and partition function

We now apply the canonical ensemble to investigate non-relativistic, classical, ideal gases. Using statistical mechanics we will explore how the large-scale behaviours of such gases emerge from the properties of the particles that compose them. The key particle properties are specified by the adjectives used above:

- **Classical** systems are those for which we can ignore the effects of quantum mechanics. Among other things, this allows us to simultaneously define both the position (x, y, z) and the momentum $\vec{p} = (p_x, p_y, p_z)$ of each particle with arbitrary precision.
- **Non-relativistic** particles move with speeds small compared to the speed of light, which allows us to ignore small effects due to special relativity. The particles are therefore governed by the laws Isaac Newton published all the way back in 1687. In particular, the energy of each particle of mass m is

$$E_n = \frac{1}{2m} p_n^2,$$

where $p_n^2 = \vec{p}_n \cdot \vec{p}_n = (p_x)_n^2 + (p_y)_n^2 + (p_z)_n^2$ is the inner (or 'dot') product of the momentum vector for the n th particle in the micro-state of interest.

- **Ideal** gases are those whose constituent particles don't interact with each other. As a result, the total energy of the gas is simply the sum of the energies of the N individual particles,

$$E = \frac{1}{2m} \sum_{n=1}^N p_n^2. \quad (46)$$

As is now familiar for the canonical ensemble, we consider the gas to be in thermodynamic equilibrium, and in thermal contact with a large external thermal reservoir with which it can exchange energy but not particles. To prevent particle exchange, we can specify that the gas is enclosed in a cubic box with volume $V = L^3$. The thermal reservoir fixes the temperature T of the gas.

The starting point for our analysis is to compute the partition function

$$Z = \sum_i e^{-E_i/T}.$$

Unfortunately there is a challenge confronting this sum over all possible micro-states ω_i of the N -particle system. These micro-states depend on the momenta \vec{p}_n for all N particles, and it's intuitive to suppose that each component of $(p_x, p_y, p_z)_n$ is a continuously varying real number that can (in principle) be distinguished with arbitrary precision. This implies an uncountably infinite set of distinct momenta and hence an uncountably infinite set of micro-states, making the summation above ill-defined.

To proceed, we *regularize* the system so that there are a countable number of micro-states we can sum over to define the partition function. We do this by positing

that the particles' momentum components can take only discrete (or 'quantized') values that depend on the volume of the box. Specifically, we declare that the possible momenta are

$$\vec{p} = (p_x, p_y, p_z) = \hbar \frac{\pi}{L} (k_x, k_y, k_z) \quad k_{x,y,z} \in \mathbb{Z}. \quad (47)$$

The constant factor \hbar ("h-bar") is known as the (reduced) Planck constant, named after [Max Planck](#)). Like the Boltzmann constant k , the Planck constant is another unit conversion factor (relating inverse length $\frac{1}{L}$ and momentum p), which is just $\hbar = 1$ in natural units. Despite this, we will retain explicit factors of \hbar in these lecture notes.

What are the energies that correspond to these discretized momenta?

You should find energies that fall into discrete *energy levels*, somewhat similar to the spin system considered in Section 3.4. Unlike the spin system, in this case the energy gaps between subsequent energy levels are not constant.

Similar discrete energy levels turn out to be realized in nature, thanks to quantum mechanics — if you have previously studied quantum physics, you may spot a resemblance with a [particle in a box](#). For the purposes of this module we can just adopt Eq. 47 as an ansatz. Even with this perspective, it will be useful to maximize the similarity with quantum physics. We can do this by observing that negating any component of the momentum has no effect on the energy $E \propto p^2 \propto k^2$, and hence all non-zero $\pm k_i$ pairs make equal contributions to the partition function. Therefore, restricting the discrete momenta to non-negative $k_{x,y,z} = 0, 1, 2, \dots$ only changes Z by a constant factor C , which cancels out in the expectation value of any observable that depends only on the inner product p^2 :

$$\langle f(p^2) \rangle = \frac{\sum_{k_{x,y,z} \in \mathbb{Z}} f(p^2) e^{-E_i/T}}{Z} \quad \longrightarrow \quad \frac{C \sum_{k_{x,y,z} \in \mathbb{N}} f(p^2) e^{-E_i/T}}{CZ}.$$

For example, the constant factor does not contribute to $\langle E \rangle = \frac{\partial}{\partial \beta} \log Z$.

Although there are still an infinite number of possible momenta and energy levels for each particle in the gas, these are now countable, making our partition function well-defined. Let's start by considering the partition function Z_1 for a single particle. The micro-states for this single-particle system are completely specified by the particle's p^2 ,

$$Z_1 = \sum_i \exp \left[-\frac{E_i}{T} \right] = \sum_{\vec{p}} \exp \left[-\frac{p^2}{2mT} \right] = \sum_{k_{x,y,z}=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2}{2mTL^2} (k_x^2 + k_y^2 + k_z^2) \right].$$

We can separately sum over each of the independent (k_x, k_y, k_z) , and recognize that all three summations are identical:

$$\begin{aligned} Z_1 &= \sum_{k_x=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2}{2mTL^2} k_x^2 \right] \sum_{k_y=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2}{2mTL^2} k_y^2 \right] \sum_{k_z=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2}{2mTL^2} k_z^2 \right] \\ &= \left(\sum_{k_i=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2}{2mTL^2} k_i^2 \right] \right)^3. \end{aligned}$$

Now that the sum over micro-states has been turned into a sum over momenta, our system has been regularized and we are free to switch back from discrete to continuous momenta.⁹ We can start by converting from integer k_i to continuous real \hat{k}_i :

$$\sum_{k_i=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2 k_i^2}{2mTL^2} \right] \rightarrow \int_0^{\infty} \exp \left[-\frac{\hbar^2 \pi^2 \hat{k}_i^2}{2mTL^2} \right] d\hat{k}_i = \frac{1}{2} \int_{-\infty}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2 \hat{k}_i^2}{2mTL^2} \right] d\hat{k}_i.$$

The final equality simply notes that the integrand is an even function of \hat{k}_i , as it depends only on \hat{k}_i^2 . Next we use Eq. 47 to return to the original momenta $p_i = \hbar \frac{\pi}{L} \hat{k}_i$,

$$\sum_{k_i=0}^{\infty} \exp \left[-\frac{\hbar^2 \pi^2}{2mTL^2} k_i^2 \right] \rightarrow \frac{1}{2} \int \exp \left[-\frac{p_i^2}{2mT} \right] \left(\frac{L}{\pi \hbar} dp_i \right).$$

We end up with the single-particle partition function

$$Z_1 = \left(\frac{L}{2\pi \hbar} \right)^3 \int \exp \left[-\frac{p^2}{2mT} \right] d^3 p,$$

where $p^2 = p_x^2 + p_y^2 + p_z^2$ and $d^3 p = dp_x dp_y dp_z$. (Some textbooks may skip the formal regularization and simply introduce this expression as a definition, using dimensional analysis to justify the factors of L and \hbar .) We can now account for all N particles in the ideal gas, which are completely independent and don't interact with each other. Assuming we can distinguish these particles from each other, then each of them simply contributes an independent factor of Z_1 to the overall partition function

$$Z_D = \left(\frac{L}{2\pi \hbar} \right)^{3N} \int \exp \left[-\sum_{n=1}^N \frac{p_n^2}{2mT} \right] d^{3N} p, \quad (48)$$

where the subscript reminds us of the particles' distinguishability. We will consider the indistinguishable case below.

We can recognize that each of the $3N$ independent integrations in Eq. 48 is a gaussian integral,

$$\frac{L}{2\pi \hbar} \int \exp \left[-\frac{p_i^2}{2mT} \right] dp_i = \frac{L}{2\pi \hbar} \sqrt{2\pi mT} = \sqrt{\frac{mTL^2}{2\pi \hbar^2}} \equiv \frac{L}{\lambda_{\text{th}}(T)}.$$

⁹If we were truly doing quantum physics, this switch would be an approximation that is valid when $\hbar^2 \pi^2 \ll 2mTL^2$, which holds unless T or L is *extremely* small. In this regime, the function being summed above varies very smoothly as the integer k_i increases, for any k_i small enough to leave the exponential factor non-negligible. You can find further discussion of this in Section 6.7 of Dan Schroeder's *Introduction to Thermal Physics*.

In the last step we have made the notation more compact by defining the *thermal de Broglie wavelength* (named after [Louis de Broglie](#)),

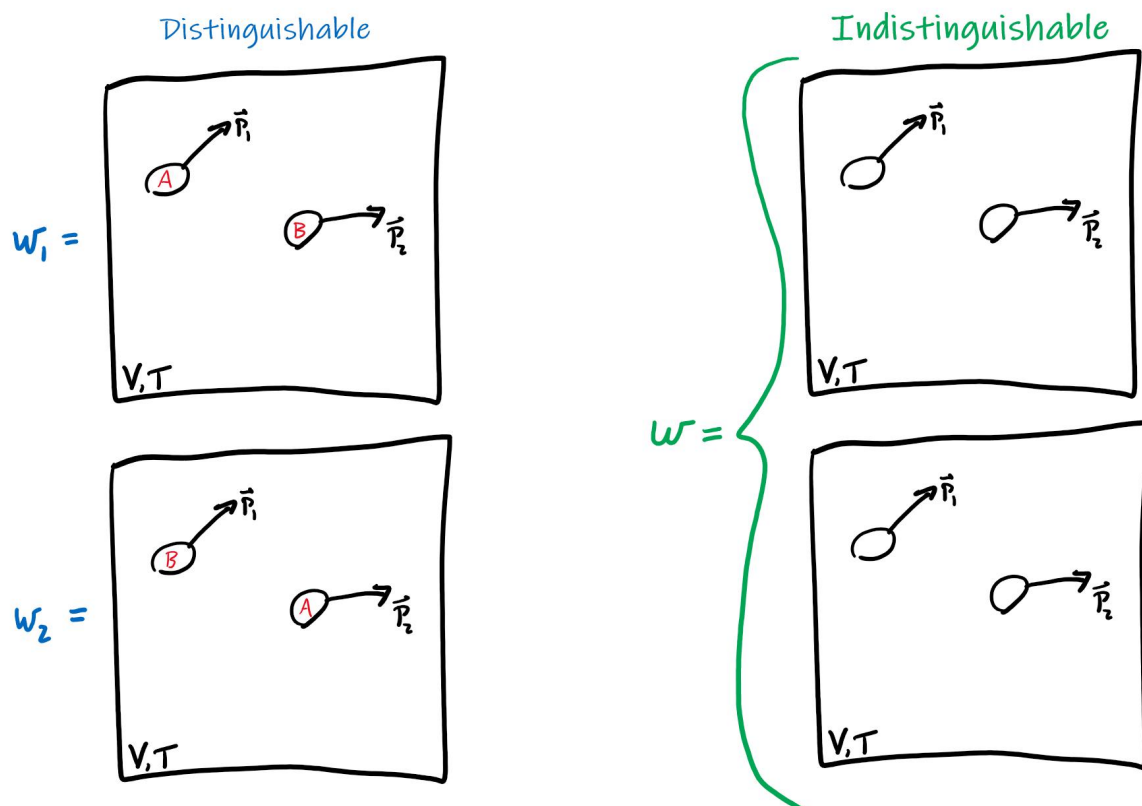
$$\lambda_{\text{th}}(T) = \sqrt{\frac{2\pi\hbar^2}{mT}}. \quad (49)$$

Performing all $3N$ gaussian integrals,

$$Z_D = \left(\frac{mTL^2}{2\pi\hbar^2}\right)^{3N/2} = \left(\frac{L}{\lambda_{\text{th}}}\right)^{3N} = \left(\frac{V}{\lambda_{\text{th}}^3}\right)^N, \quad (50)$$

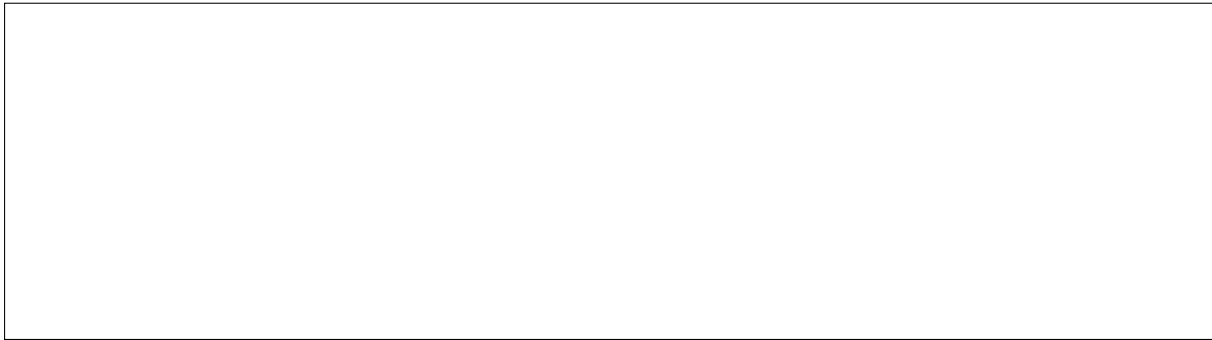
since the volume of the box is $V = L^3$. It is worth emphasizing here that the partition function *depends on the volume of the gas*, in addition to the fixed temperature T and conserved particle number N . This dependence may persist in other quantities derived from the partition function, which we will consider in the next section.

First, let's determine what we would have with indistinguishable particles. For a classical gas, distinguishability means that we can label the particles and use those labels to tell them apart. In the simple two-particle example illustrated below, these labels mean we have a different micro-state ω_1 when particle A has momentum \vec{p}_1 while particle B has momentum \vec{p}_2 , compared to micro-state ω_2 in which particle A has momentum \vec{p}_2 while particle B has momentum \vec{p}_1 .



If the particles are indistinguishable, no such labeling is possible, and there is only one micro-state for these $\{\vec{p}_1, \vec{p}_2\}$, rather than two. This factor of 2 is not accidental, as

you can explore by counting how many micro-states there are for three distinguishable particles with momenta $\{\vec{p}_1, \vec{p}_2, \vec{p}_3\}$, compared to the single micro-state for the indistinguishable case:



Generalizing to N particles, we find that ideal gases with distinguishable particles have $N!$ times more micro-states compared to otherwise-identical ideal gases with indistinguishable particles: There are N possible ways to label the particle with momentum \vec{p}_1 , then $N - 1$ possible labels for \vec{p}_2 , and so on.¹⁰ The partition function sums over these micro-states, but depends only on their energies, which are independent of any labeling. Therefore this factor of $N!$ is the only difference between Eq. 50 and the partition function for indistinguishable particles,

$$Z_I = \frac{1}{N!} \left(\frac{mTL^2}{2\pi\hbar^2} \right)^{3N/2} = \frac{1}{N!} \left(\frac{L}{\lambda_{\text{th}}} \right)^{3N} = \frac{1}{N!} \left(\frac{V}{\lambda_{\text{th}}^3} \right)^N. \quad (51)$$

4.2 Internal energy, and entropy

Now that we have the canonical partition function, let's apply our work from Unit 3 to predict the large-scale behaviour of the ideal gas it describes. Our first targets are the average internal energy $\langle E \rangle$ and entropy S for the gas, as functions of its fixed temperature T , conserved particle number N , and the volume $V = L^3$ of the box in which it is contained. Let's begin with the slightly more complicated case of indistinguishable particles, Eq. 51. Recalling the derivatives in Eqs. 38–39, we should keep the temperature dependence explicit in our workings, rather than hidden inside the thermal de Broglie wavelength $\lambda_{\text{th}}(T)$.

By writing down the Helmholtz free energy,

$$F_I = -T \log Z_I = -\frac{3NT}{2} \log \left(\frac{mTL^2}{2\pi\hbar^2} \right) + T \log(N!),$$

we can quickly extract the internal energy,

$$\langle E \rangle_I = -T^2 \frac{\partial}{\partial T} \left(\frac{F_I}{T} \right) = -T^2 \frac{\partial}{\partial T} \left(-\frac{3N}{2} \log T + T\text{-independent} \right) = \frac{3}{2} NT.$$

¹⁰This argument assumes the momenta themselves are distinguishable, $\vec{p}_i \neq \vec{p}_k$ for any $i \neq k$. This is a reliable assumption for classical gases with $L\sqrt{mT} \gg \hbar$, but will need to be revisited when we consider quantum statistics.

This in turn provides the entropy

$$S_I = \frac{\langle E \rangle_I - F_I}{T} = \frac{3}{2}N + \frac{3N}{2} \log \left(\frac{mTL^2}{2\pi\hbar^2} \right) - \log(N!).$$

We can clean this up by reintroducing the thermal de Broglie wavelength,

$$\frac{3N}{2} \log \left(\frac{mTL^2}{2\pi\hbar^2} \right) = \frac{3N}{2} \log \left(\frac{L^2}{\lambda_{\text{th}}^2} \right) = N \log \left(\frac{V}{\lambda_{\text{th}}^3} \right),$$

and by applying Stirling's formula to find

$$S_I = \frac{3}{2}N + N \log \left(\frac{V}{\lambda_{\text{th}}^3} \right) - N \log N + N = \frac{5}{2}N + N \log \left(\frac{V}{N\lambda_{\text{th}}^3} \right).$$

We can interpret $N\lambda_{\text{th}}^3$ as the volume 'occupied' by the N particles.

What are the corresponding results for the case of distinguishable particles, starting from Eq. 50 for the partition function Z_D ?

$$F_D =$$

$$\langle E \rangle_D =$$

$$S_D =$$

You should find that the energy is the same whether or not we can label the particles:

$$\langle E \rangle_D = \langle E \rangle_I = \frac{3}{2}NT. \quad (52)$$

This agrees with the argument in the previous section that multiplying Z by a constant factor (here $N!$) does not change the internal energy expectation value.¹¹

However, the entropy reflects the extra information that distinguishability provides:

$$S_D = \frac{3}{2}N + N \log \left(\frac{V}{\lambda_{\text{th}}^3} \right) \quad S_I = \frac{5}{2}N + N \log \left(\frac{V}{N\lambda_{\text{th}}^3} \right). \quad (53)$$

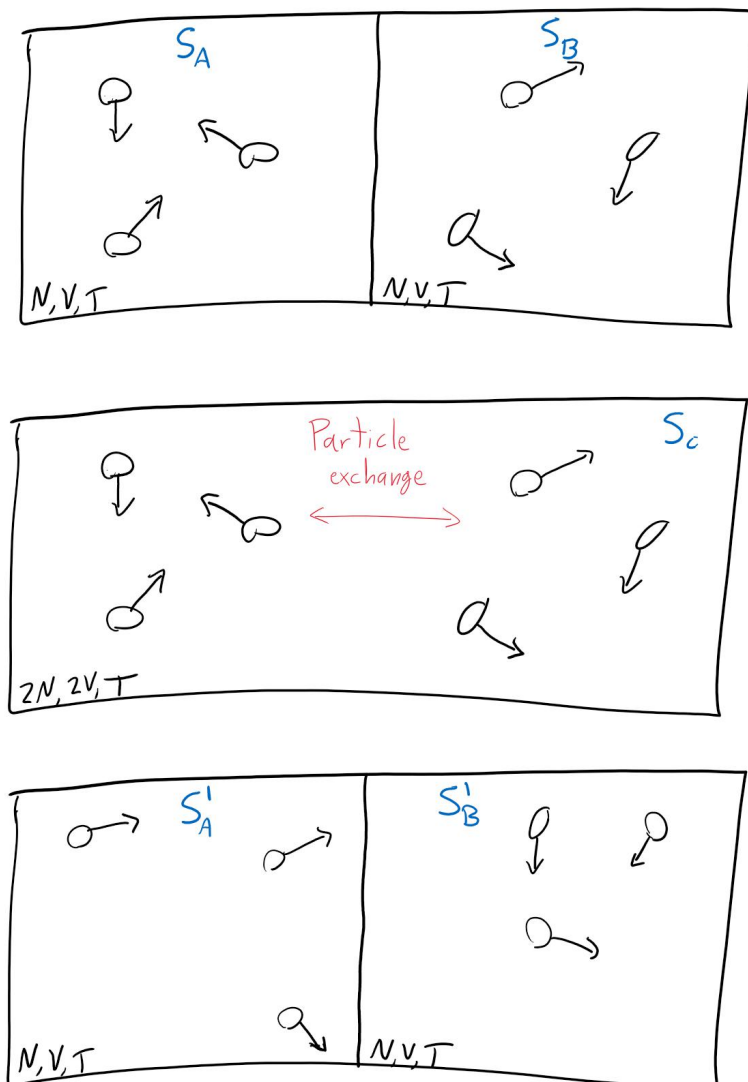
¹¹The spin system we considered in Section 3.4 behaved differently because its number of distinguishable micro-states per indistinguishable micro-state was the energy-dependent binomial coefficient $\binom{N}{n_{\uparrow}}$. The energy dependence caused Z_D vs. Z_I to differ by more than a simple constant factor.

The difference $S_I - S_D = N - N \log N \rightarrow -\log(N!) < 0$, meaning $S_I < S_D$, as expected. We can also note that $\lambda_{\text{th}} \rightarrow \infty$ as the temperature approaches absolute zero, $T \rightarrow 0$, apparently producing negative entropies for fixed V . This is a warning sign that our classical assumptions are breaking down in this regime, and quantum effects would need to be taken into account.

4.3 The mixing entropy and the ‘Gibbs paradox’

In Section 2.4 we analysed what would happen if we allowed two micro-canonical systems to exchange energy, and then re-isolated them. We saw that this procedure obeys the second law of thermodynamics — the entropy never decreases, though we have to be careful to account for all of the entropy after re-isolating the two systems.

We can now carry out a similar thought experiment of allowing two *canonical* systems to exchange *particles*, and then re-separating them. We demand that both canonical ensembles are in thermodynamic equilibrium with each other, for instance by sharing the same thermal reservoir with temperature T . This procedure is illustrated below, where we simplify the setup by taking the two initial systems to have equal volumes, $V_A = V_B = V$, and numbers of particles, $N_A = N_B = N$.



We can represent the process of combining and re-separating these systems as

$$\Omega_A + \Omega_B \longrightarrow \Omega_C \longrightarrow \Omega'_A + \Omega'_B.$$

What is the entropy for each of these three stages? Since the entropies depend on whether or not the particles in the gas are distinguishable from each other, let's first consider the case of *indistinguishable* particles.

The initial entropy is the sum of the contributions from the two canonical systems, $S_A + S_B$, both of which are the same thanks to our simplification above:

$$S_A + S_B =$$

To find the entropy S_C of the combined system, we just need to consider what happens when we double the volume and also double the number of particles:

$$S_C =$$

You should find $S_C = S_A + S_B$, which is consistent with the second law.

Things are more complicated when we re-separate the systems. Analogously to our considerations in Section 2.4, we need to sum over all the possible ways of dividing the $2N$ indistinguishable particles between the two re-separated boxes. In particular, we need to perform this sum at the stage of computing the partition function Z' for $\Omega'_A + \Omega'_B$, since this is the fundamental quantity from which the entropy is then derived as $S' = \frac{\partial}{\partial T} (T \log Z')$. In other words, we have to consider a logarithm of a sum rather than a sum of logarithms.

If ν particles end up in system Ω'_A , then the other system Ω'_B must contain the remaining $2N - \nu$ particles, giving us

$$Z_\nu = \frac{1}{\nu!} \left(\frac{V}{\lambda_{\text{th}}^3} \right)^\nu \times \frac{1}{(2N - \nu)!} \left(\frac{V}{\lambda_{\text{th}}^3} \right)^{2N - \nu} = \frac{1}{\nu! (2N - \nu)!} \left(\frac{V}{\lambda_{\text{th}}^3} \right)^{2N}.$$

Summing over all possible values of $0 \leq \nu \leq 2N$,

$$\begin{aligned} Z' &= \sum_{\nu=0}^{2N} Z_\nu = \left(\frac{V}{\lambda_{\text{th}}^3} \right)^{2N} \sum_{\nu=0}^{2N} \frac{1}{\nu! (2N - \nu)!} = \left(\frac{V}{\lambda_{\text{th}}^3} \right)^{2N} \frac{1}{(2N)!} \sum_{\nu=0}^{2N} \binom{2N}{\nu} \\ \implies S'_A + S'_B &= 2N \frac{\partial}{\partial T} \left(T \log \left[\frac{V}{\lambda_{\text{th}}^3} \right] \right) - \log[(2N)!] + \log \left[\sum_{\nu=0}^{2N} \binom{2N}{\nu} \right]. \end{aligned}$$

This is a complicated expression. In the 1870s, Gibbs introduced the following argument that helps to simplify it, which we will explore further in tutorials: For large $N \gg 1$, the entropy of the two subsystems is nearly saturated by the case in which the particles are divided roughly evenly between them, rather than being mostly in one of them. Equivalently, there are far more micro-states with $N'_A \approx N'_B \approx N$, compared to all other terms in the sum above.

Therefore we can safely set $N'_A = N'_B = N$, which was already incorporated into the illustration above.¹² This means $\Omega'_A = \Omega_A$ and $\Omega'_B = \Omega_B$, producing a final entropy of $S'_A + S'_B = S_A + S_B$ that satisfies the second law:

$$S'_A + S'_B = S_C = S_A + S_B.$$

This is just what we would expect from everyday experience: Opening a door between two identical rooms doesn't produce any observable effects, nor does reversing that process by closing the door.

Something interesting happens when we repeat this analysis for the case of *distinguishable* particles, using our result for $S_D(N, V)$ in Eq. 53. If we consider the difference between the combined entropy S_C and the initial entropy $S_A + S_B$,

$$\begin{aligned} \Delta S_{\text{mix}} &= S_C - (S_A + S_B) = S_D(2N, 2V) - 2S_D(N, V) \\ &= 3N + 2N \log \left(\frac{2V}{\lambda_{\text{th}}^3} \right) - \left[3N + 2N \log \left(\frac{V}{\lambda_{\text{th}}^3} \right) \right] = 2N \log 2 > 0, \end{aligned} \quad (54)$$

we find the entropy increases upon combining the two initial systems. This $\Delta S_{\text{mix}} > 0$ is known as the **mixing entropy**.

This result $S_C > S_A + S_B$ is to be expected from the second law of thermodynamics. However, repeating the argument above — that we should have $N'_A \approx N'_B \approx N$ leading to $S'_A + S'_B = S_A + S_B$ after re-separating the systems — would produce the prediction $S'_A + S'_B < S_C$, indicating a *decrease* in the entropy by ΔS_{mix} and an apparent violation of the second law. This is known as the 'Gibbs paradox', though Gibbs himself explained how a paradox is avoided.

The explanation is that because the particles are now distinguishable, $N'_A = N_A$ no longer suffices to establish $\Omega'_A = \Omega_A$ and $S'_A = S_A$. Recovering Ω_A would additionally require that the N'_A particles in the re-separated system are the *same* distinguishable particles that were initially in Ω_A . While we can still expect $N'_A \approx N'_B \approx N$, the vast majority of the resulting micro-states will not correspond to micro-states of Ω_A and Ω_B . Summing over these additional possibilities ensures $S'_A + S'_B > S_A + S_B$, and it turns out $S'_A + S'_B \geq S_C$ as well, obeying the second law of thermodynamics.

These thought experiments provide another example of behaviour that depends on the intrinsic information content of the system — whether or not the particles in an ideal gas can be distinguished from each other in principle. Mixing gases of distinguishable particles introduces a positive mixing entropy, Eq. 54, but for gases of indistinguishable particles there is no change in entropy when we let two subsystems mix, or when we reverse that process and re-separate them. Due to the second law, processes that produce an increase in entropy are **irreversible**.

¹²Formally this is only exact in the *thermodynamic limit* $N \rightarrow \infty$, a concept we will discuss in Unit 9.

4.4 Pressure, ideal gas law, and equations of state

Below Eq. 50 we emphasized that the ideal gas partition function depends on the volume of the gas, V , in addition to the fixed temperature T and conserved particle number N that always characterize systems governed by the canonical ensemble. Parameters like V that appear in the partition function are called **control parameters**, with the idea that they can (in principle) be controlled in experiments. Control parameters generally enter the partition function through the definition of the energies E_i for the micro-states ω_i . Another example is the magnetic field strength H for the spin systems we considered earlier.

Focusing on ideal gases for now, we see that all dependence on V drops out in our results for the average internal energy, Eq. 52. On the other hand, the entropies in Eq. 53 do depend on the volume. For both cases of distinguishable and indistinguishable particles, the entropy S depends on the same combination of volume and temperature: $V\lambda_{\text{th}}^{-3} \propto VT^{3/2}$. If we keep N fixed and consider using our experimental control to change the volume and the temperature of the system, the entropy will typically change as a consequence, unless the following relation is satisfied:

$$VT^{3/2} = \text{constant} \quad \implies \quad S = \text{constant}.$$

Such **constant-entropy (or isentropic) processes** will be important in our upcoming analyses of thermodynamic cycles.¹³ These cycles will involve making changes to control parameters, which is a topic we have already started to consider through the micro-canonical temperature (Eq. 22) and the canonical heat capacity (Eq. 36). The pressure of an ideal gas is similarly connected to a change in its volume, which we can motivate by thinking about squeezing an inflated balloon into a small box.

The **pressure** is defined to be

$$P = - \left. \frac{\partial}{\partial V} \langle E \rangle \right|_S, \quad (55)$$

with constant entropy S . In words, the pressure is the isentropic response of the system's internal energy to a change in its volume.

In Unit 5 we will look in detail at processes that change some or all of the pressure, volume, temperature, or internal energy of an ideal gas, with N fixed. Although changing the temperature departs from the assumptions of the canonical ensemble, we will be able to analyse such a process as a change from one canonical system (in thermodynamic equilibrium with a thermal reservoir that fixes the initial temperature T_0) to another (in thermodynamic equilibrium with a different thermal reservoir that fixes the final temperature T_f).

If we consider an isentropic process with N fixed, then the temperature and volume are related,

$$VT^{3/2} = c^{3/2} \quad \longrightarrow \quad T = cV^{-2/3},$$

¹³The term *isentropic* is based on the Greek word *ισος* ("isos"), meaning "equal".

with c a constant. By inserting this into Eq. 52, we can relate the average internal energy to the volume,

$$\langle E \rangle = \frac{3}{2}NT = \frac{3c}{2}NV^{-2/3} \quad \text{for constant entropy.}$$

Using this constant-entropy expression, what is the pressure for the ideal gas?

$$P = - \left. \frac{\partial}{\partial V} \langle E \rangle \right|_S =$$

You should find the **ideal gas law**,

$$PV = NT, \tag{56}$$

which is an example of an **equation of state**.

The “state” referred to by this terminology is different from the micro-states that we have mostly discussed up until now. Whereas each micro-state is defined by detailed information about the microscopic degrees of freedom that constitute the system, this **macro-state** concerns only the large-scale (*macroscopic*) properties of the system, such as its pressure, volume, temperature, entropy, or internal energy. (Macro-states are sometimes called “system states” or “thermodynamic states”.) Equations of state are relations between these large-scale properties.

Historically, equations of state were observed empirically and studied experimentally well before the mathematical development of statistical mechanics. In the 1660s, for instance, [Robert Boyle](#) experimented with changing the pressure of a gas while holding its temperature fixed, finding a special case of the ideal gas law,

$$PV = \text{constant} \quad \text{for constant } N \text{ and } T,$$

which became known as “[Boyle’s law](#)”. (I include the quotation marks to emphasize the limitations of assigning an individual sole credit for advances arising from the work of broad scientific communities.)

Other equations of state reflecting different aspects of the ideal gas law were uncovered during the Industrial Revolution:

- $\frac{V}{T} = \text{constant}$ for constant N and P (1787, “[Charles’s law](#)”)

- $\frac{P}{T} = \text{constant}$ for constant N and V (1802, “[Gay-Lussac’s law](#)”)
- $\frac{V}{N} = \text{constant}$ for constant P and T (1812, “[Avogadro’s law](#)”)

In the 1830s [Émile Clapeyron](#) combined these empirical results into the ideal gas law itself, which [August Krönig](#) and [Rudolf Clausius](#) independently derived on the basis of statistical mechanics in the 1850s. These historical considerations are useful to illustrate how progress in scientific and mathematical understanding went hand-in-hand with industrial developments, including the design of engines and related machines, which are connected to our next topic of thermodynamic cycles.

Unit 5: Thermodynamic cycles

5.1 Work, pressure and force

In the previous section we defined the pressure of an ideal gas in the canonical ensemble as the thermodynamic response of the internal energy to an isentropic change in the volume (Eq. 55). We motivated this definition by thinking about ‘squeezing’ the system — exerting a force on it — which suggests a connection between pressure and force. Here we make this connection explicit by considering how the energy of an object changes when a force acts on it.

Let’s begin by considering a single object at position $\vec{r} = (x, y, z)$, which is displaced by a vector $d\vec{r}$ due to a force $\vec{F}(\vec{r})$. The *work* done by this force is defined to be the resulting change in the energy of the object. Infinitesimally, $W = dE = \vec{F} \cdot d\vec{r}$, which generalizes to the line integral $W = \Delta E = \int \vec{F}(\vec{r}) \cdot d\vec{r}$.

A famous example is an object falling due to the force of the Earth’s gravity. That force is $\vec{F} = (0, 0, -mg)$, where m is the mass of the object, $g \approx 9.8 \text{ m/s}^2$ (metres per second per second) is the strength of gravity near the surface of the Earth, and the negative sign indicates that the gravitational force is directed downward. Suppose the object starts from rest, with initial kinetic energy $E_0 = 0$, and falls downward, parallel to \vec{F} , from a height h . Its final energy E_f upon hitting the ground comes from the work done by the Earth’s gravity:

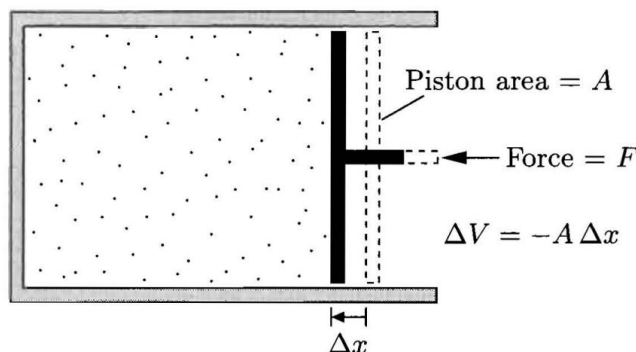
$$W = \int \vec{F}(\vec{r}) \cdot d\vec{r} = -mg \int_h^0 dz = mgh > 0$$

$$E_f = E_0 + \Delta E = 0 + W = mgh = \frac{p_z^2}{2m} \quad \longrightarrow \quad p_z = -m\sqrt{2gh},$$

where Eq. 46 relates the energy to the momentum $\vec{p} = (p_x, p_y, p_z)$.

Generalizing to $N \gg 1$ objects in a statistical system governed by the canonical ensemble, we define the **work** done by a force to be the resulting change in the system’s average internal energy *due to that force*, $W = \Delta \langle E \rangle_{\text{force}}$. In practice, the volume is the control parameter that such a force will change.

This change in $\langle E \rangle$ due to a change in volume suggests that the work is related to the pressure defined by Eq. 55. We can formalize this relation by considering the setup shown below (from Schroeder’s *Introduction to Thermal Physics*).



Here we have an ideal gas in a container of volume V , with one wall of that container being a piston that we can move by applying a force F . Let's demand that this process leaves the entropy of the gas constant. The displacement $\Delta x > 0$ shown in the figure reduces the volume of the gas, by $\Delta V = -A\Delta x < 0$ where A is the surface area of the piston. Since the force F is parallel to the piston's displacement Δx , it does positive work $W = F\Delta x > 0$. Therefore the internal energy of the gas increases by $\Delta\langle E \rangle = W$, at the same time as its volume decreases isentropically, so from Eq. 55

$$P = - \left. \frac{\partial}{\partial V} \langle E \rangle \right|_S = - \frac{W}{\Delta V} = \frac{F\Delta x}{A\Delta x} = \frac{F}{A}. \quad (57)$$

This identifies the pressure of an ideal gas in a container as the force per unit area that the gas exerts on the container wall, in agreement with our everyday experiences.

Rearranging the expressions above, we can obtain an expression for the work *done on the gas by its surroundings* — that is, by the external force applied to move the piston and change the volume. Still assuming a constant-entropy (isentropic) process, this input work must match the increase in the gas's average internal energy,

$$W = \Delta\langle E \rangle = -P\Delta V \quad \text{for constant entropy.}$$

If the entropy is allowed to change, this relation between work and pressure will still hold. However, as we will see in the next section, the non-constant entropy will introduce an additional change in the average internal energy unrelated to a force, leading to $W \neq \Delta\langle E \rangle$ and leaving only the relation

$$W = -P\Delta V \quad \text{more generally.} \quad (58)$$

Later we will be interested in thermodynamic engines where work is *done by the gas on its surroundings*. This removes energy from the gas, corresponding to a negative $W < 0$, and we will need to be careful to keep track of the negative signs and their physical meaning.

Of course, as we change the volume of the gas, the pressure itself may change as described by the gas's equation of state — such as the ideal gas law, Eq. 56. In all these considerations we will keep the particle number N fixed, though in principle it could change in the same way as discussed below Eq. 55 for the temperature. With the equation of state providing an expression $P(V)$ for the pressure as a function of the volume, Eq. 58 generalizes to

$$W = - \int_{V_0}^{V_f} P(V) dV. \quad (59)$$

5.2 Heat and entropy

Now let's switch things up by changing the temperature T of an ideal gas while keeping its volume V and particle number N constant. Since the volume is constant, Eq. 59 indicates that no work is done, $W = 0$. Even so, from Eq. 52 we have $\langle E \rangle = \frac{3}{2}NT$ and can see that the average internal energy still changes,

$$d\langle E \rangle = \frac{3}{2}N dT. \quad (60)$$

In order to remain consistent with our discussion in the previous section, we should expect a change in the entropy to accompany this change in the internal energy that occurs with no work done. Indeed, both distinguishable and indistinguishable particles lead to the same the temperature dependence in Eq. 53 for the entropy:

$$S = N \log (\lambda_{\text{th}}^{-3}) + T\text{-independent} = N \log (T^{3/2}) + T\text{-independent}.$$

What is the change in entropy that results from changing the temperature by dT ?

$$dS =$$

Looking back to Eq. 60, you should find $d\langle E \rangle = T dS$, which leads us to another important definition.

The **heat** added to or removed from a statistical system is defined to be

$$Q = T dS, \tag{61}$$

and corresponds to the change in the average internal energy of the system when the volume and particle number are kept constant.

In the same way as our considerations of the work W in the previous section, we can generalize this infinitesimal definition to

$$Q = \int_{S_0}^{S_f} T(S) dS, \tag{62}$$

with $Q = \Delta\langle E \rangle$ when the volume is constant. Here we assume it is possible to invert the usual canonical relation that expresses the entropy as a function of the temperature, $S(T)$. Some textbooks may refer to infinitesimal heat and work as “ dQ ” and “ dW ”, but this notation can easily be misinterpreted as describing a ‘change’ in heat or work. Instead, heat and work are themselves changes in the internal energy.

Like the work W , the heat Q is positive when energy is added to the system to increase $\langle E \rangle$, and negative when energy is removed. Recalling that the canonical ensemble involves placing the system in thermal contact with a large external thermal reservoir, we can recognize that this energy is not being created or destroyed, but is instead flowing back and forth between the system and the reservoir. When considering heat, we will also demand that no *entropy* is created or destroyed — a positive dS will indicate entropy flowing into the system from the reservoir, while a negative value reflects entropy moving from the system to the reservoir. Because the total entropy of the system plus its reservoir is constant, these processes are reversible, making it possible for the system to return to its starting macro-state.¹⁴

¹⁴In the case of irreversible processes, there must be sources of entropy creation, which change Eq. 61 to $Q < T dS$. The **Clausius inequality** $Q \leq T dS$ covers both reversible and irreversible cases.

We have already considered isentropic processes with $dS = 0$, for example in the definition of pressure in Eq. 55. With our assumption of reversibility, the definition of heat provides a new perspective on such processes:

We define an **adiabatic** process to be a change in the control parameters of a system that occurs without transferring heat, $Q = 0$. When this process is reversible, Eq. 61 guarantees that it also does not change the system's entropy.

Since the canonical ensemble requires thermal contact between the system and its surroundings, the practical way to avoid heat exchange is to change the control parameters quickly. That is, **adiabatic processes are fast** enough that the system does not have time to exchange heat (and hence entropy) with its surroundings. The opposite extreme would be a process slow enough that any and all possible heat exchange can be completed while it is underway. Based on our work in Section 2.4, we can see that such heat exchange will keep the system's temperature equal to the temperature of its surroundings. Taking that surrounding temperature to be constant, we reach the conclusion that **constant-temperature (or isothermal) processes are slow**. Real processes generally exist in between these two extremes, usually closer to the adiabatic limit.

5.3 Thermodynamic cycles

Now we can generalize our considerations in the previous two sections to address simultaneous changes in the temperature T and the volume V of an ideal gas, still with fixed particle number N . We are used to working with the internal energy $\langle E \rangle(T, V)$ and entropy $S(T, V)$ as functions of the temperature and volume. Inverting the latter relation allows us to instead express the temperature $T(S, V)$ as a function of the entropy and volume, which carries through to the internal energy $\langle E \rangle = \frac{3}{2}NT$,

$$\langle E \rangle(T, V) \rightarrow \langle E \rangle(S, V).$$

Let's expand this to first order in a multi-variable Taylor expansion:

$$\langle E \rangle(S, V) \approx \langle E \rangle(S_0, V_0) + (S - S_0) \left. \frac{\partial \langle E \rangle}{\partial S} \right|_V + (V - V_0) \left. \frac{\partial \langle E \rangle}{\partial V} \right|_S.$$

This approximation becomes exact in the limit of infinitesimal changes

$$\langle E \rangle(S, V) - \langle E \rangle(S_0, V_0) \rightarrow d\langle E \rangle \quad S - S_0 \rightarrow dS \quad V - V_0 \rightarrow dV.$$

At the same time, we can recognize the temperature from Eq. 22 and the (negative) pressure from Eq. 55, to obtain

$$d\langle E \rangle = T dS - P dV = Q + W. \tag{63}$$

This is a generalized form of the **first law of thermodynamics**: Any change in the internal energy of a canonical system must be matched by (either or both) heat exchange with its surroundings or work done by or on those surroundings.

We now have all the concepts and **key equations** needed to consider a variety of ways to manipulate a canonical, classical, non-relativistic ideal gas in a container:

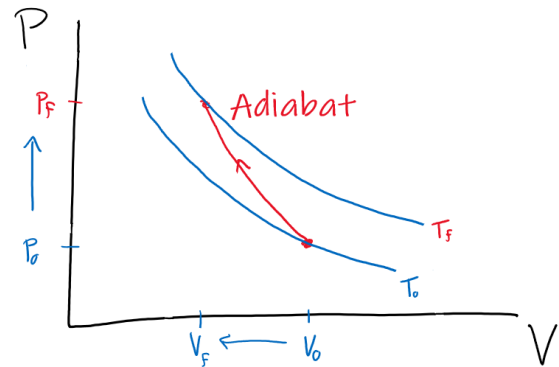
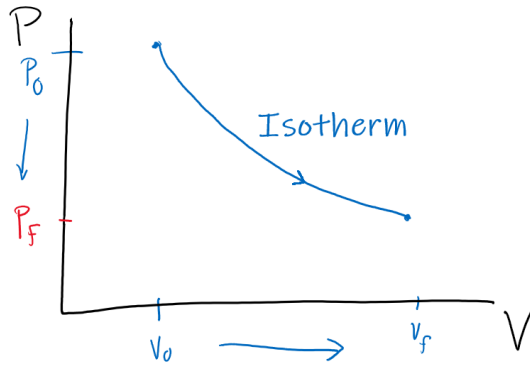
- Eq. 52 for the internal energy: $\langle E \rangle = \frac{3}{2}NT$
- Eq. 53 for the condition of constant entropy: $VT^{3/2} = \text{constant}$
- Eq. 56 for the equation of state (ideal gas law): $PV = NT$
- Eq. 63 for the first law of thermodynamics: $d\langle E \rangle = T dS - P dV = Q + W$

As examples of manipulations we can carry out by changing the system's control parameters, the piston we considered in Section 5.1 allows us to compress or expand the gas. This change in volume could be fast to keep the entropy constant (adiabatic), or slow to keep the temperature constant (isothermal). Alternatively, we can clamp the piston in place to keep the volume constant, and add heat to the gas to increase its temperature — according to the ideal gas law, this will also increase the pressure of the gas. Or we can add heat while keeping the pressure constant by applying a constant force to the piston. The ideal gas law then implies the volume will increase, pushing out the piston and potentially doing work on the surroundings.

It's possible to carry out a sequence of such manipulations that cause the system to end up in the same thermodynamic (macro-)state in which it started, with the same pressure, volume, temperature and internal energy. This sequence can then be repeated over and over again, always returning to the same starting point. Such a repeatable process is known as a **thermodynamic cycle**. As we will see in the next section, these cycles can make use of heat to have the system do work on its surroundings (providing an *engine*), or make use of work to remove heat from the system (providing a *refrigerator*), among other applications.

With N fixed, the key equations above allow us to specify the full macro-state for an ideal gas solely in terms of the pressure P and the volume V . The ideal gas law provides the temperature $T = \frac{PV}{N}$, which then determines the internal energy $\langle E \rangle \propto NT$. This makes it convenient to represent the system's macro-state as a point in a **pressure–volume (or PV) diagram** — a graph with the volume on the horizontal axis and the pressure on the vertical axis. The manipulations discussed above correspond to lines in PV diagrams. In the case of a thermodynamic cycle, the lines must meet up to form a closed path for the system to go around as the cycle is repeated.

For example, the left figure on the next page shows the PV diagram for isothermal expansion of the gas, which is slow enough for heat to enter the system to keep the temperature fixed despite the expansion.



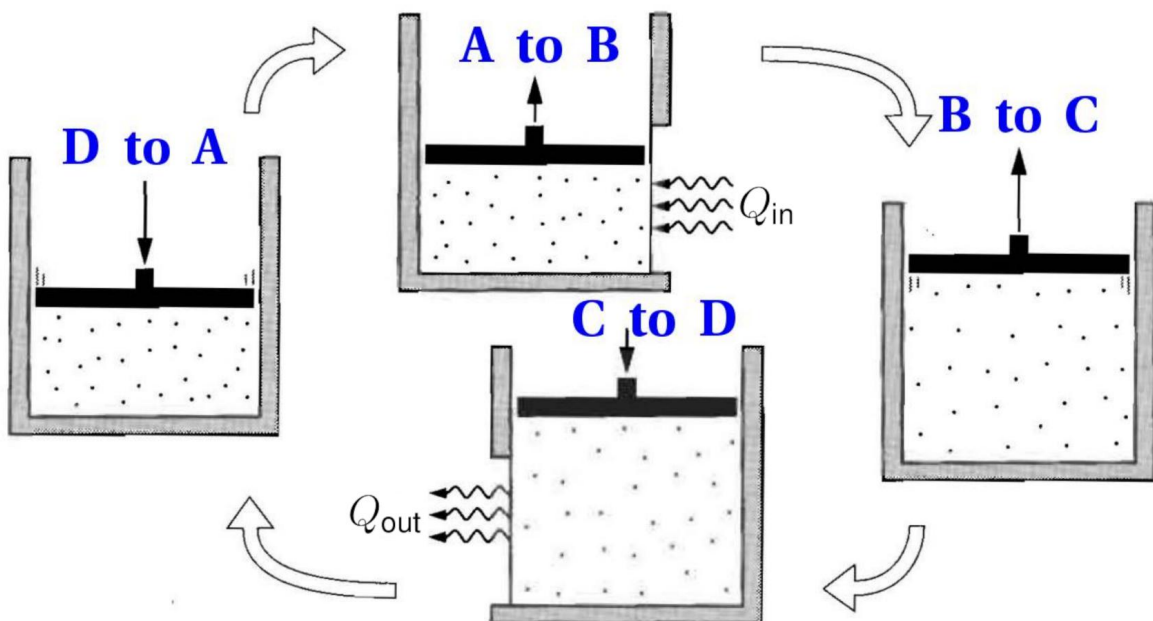
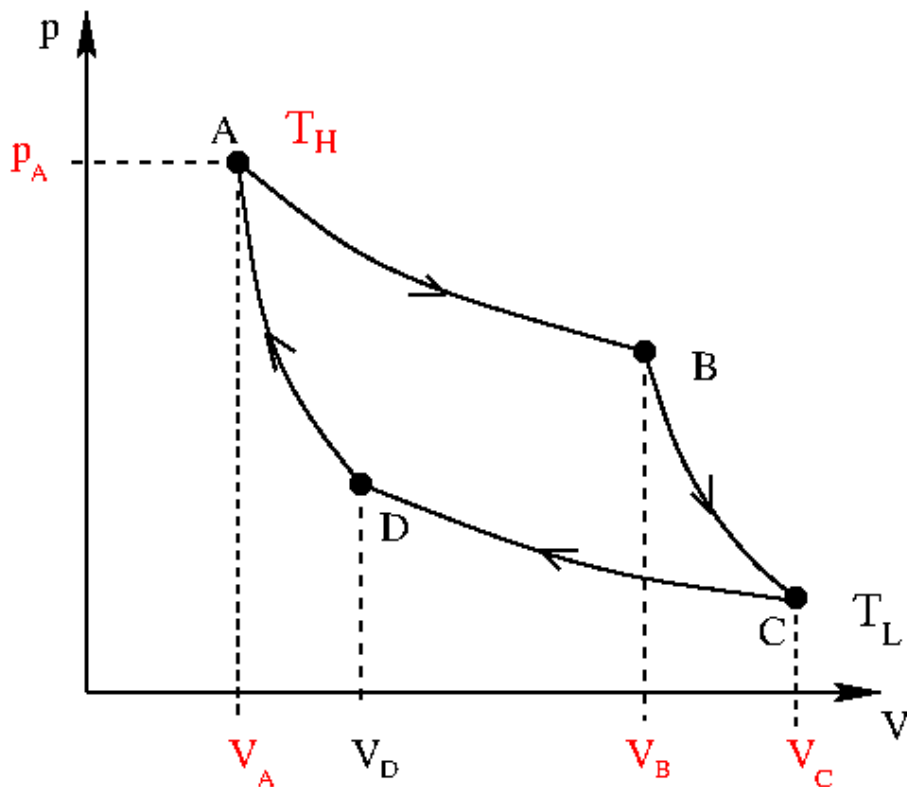
The line in a PV diagram for an isothermal process is known as an *isotherm*. As the volume expands from V_0 to V_f , the temperature (and therefore PV) is constant. What is the change in pressure $\Delta P = P_f - P_0$ in terms of P_0 , V_0 and V_f ? What would it mean if two isotherms were to cross in a PV diagram?

Similarly, we can consider the PV diagram on the right for adiabatic compression of the gas, which occurs too quickly for heat to be exchanged. In this case both the pressure and temperature change, while the entropy (and therefore $VT^{3/2}$) is constant. What are ΔP and the change in the temperature $\Delta T = T_f - T_0$ in terms of P_0 , V_0 , V_f and the fixed number of particles N ? Can you convince yourself that two distinct reversible adiabats never cross?

5.4 The Carnot cycle

A famous thermodynamic cycle was proposed by [Sadi Carnot](#) in 1824, and laid the groundwork for subsequent development of engines and refrigerators later in the

nineteenth century. The key idea is to propose that the ideal gas in its container can exchange energy with either of *two* different thermal reservoirs: a hot reservoir with a higher temperature T_H and a cold reservoir with a lower temperature T_L . The Carnot cycle consists of four stages, which are first shown below in the form of a PV diagram, then illustrated in a sketch (adapted from Schroeder's *Introduction to Thermal Physics*) that provides a more concrete picture of the physical processes, and finally summarized in words.



The illustration above supposes that the hot reservoir is located to the right of the system, while the cold reservoir is on its left. In words, the four stages are the following:

- From point A to point B the system undergoes slow isothermal expansion, bringing in heat Q_{in} from the hot reservoir in order to keep its temperature T_H fixed.
- From point B to point C the system undergoes fast adiabatic expansion, with no heat exchange, until its temperature falls from T_H down to T_L .
- From point C to point D the system undergoes slow isothermal compression, expelling heat Q_{out} into the cold reservoir to keep its temperature T_L fixed.
- From point D to point A the system undergoes fast adiabatic compression, with no heat exchange, until its temperature rises from T_L back up to T_H .

We need to make sure that these processes really do produce a self-consistent closed cycle that our system could repeatedly follow. In a real experiment, we would have full control over the four input variables $\{P_A, V_A, V_B, V_C\}$ coloured red in the PV diagram above. Specifically, we can prepare our N -particle system in initial macro-state A with our choice of pressure P_A and volume V_A , which sets the higher temperature $T_H = \frac{P_A V_A}{N}$ of the hot reservoir. We can then freely choose the volume $V_B > V_A$ at which to switch from isothermal expansion to adiabatic expansion, and similarly choose the volume $V_C > V_B$ at which we stop expanding and start compressing. These choices of V_B and V_C set the lower temperature T_L of the cold reservoir.

At this point, however, we are no longer free to choose an arbitrary volume $V_D < V_C$ at which to switch from isothermal compression to adiabatic compression — this switch needs to happen at precisely the correct point in order for the final stage to bring the system back to its initial macro-state A . While we can expect that this will be possible for the Carnot cycle, a priori there is no guarantee that a given sequence of processes will close to form a self-consistent thermodynamic cycle.

In order to confirm the self-consistency of the Carnot cycle, we need to express the unknown quantities $\{P_B, P_C, T_L, P_D, V_D\}$ in terms of the four inputs described above, along with the fixed number of particles N . At point B , we know the system's temperature remains $T_H = P_A V_A / N$. What is the pressure P_B in terms of $\{P_A, V_A, V_B, V_C, N\}$?

At point C , we know the system's entropy is the same as at point B . What are the temperature T_L and pressure P_C in terms of $\{P_A, V_A, V_B, V_C, N\}$?

At point D , we know the system's temperature remains T_L . We have to demand that its entropy is the same as at point A , in order for the final adiabatic stage to connect points D and A . What are the resulting pressure P_D and volume V_D in terms of $\{P_A, V_A, V_B, V_C, N\}$?

You should find that all of $\{P_B, P_C, T_L, P_D, V_D\}$ can be consistently specified by the inputs under our control, which confirms that the Carnot cycle is a valid thermodynamic cycle (as expected). We now also have the ingredients needed to investigate how much net work (if any) this cycle can do on its surroundings, compared to the amount of heat it would transfer from the hot reservoir to the cold reservoir. It will simplify this calculation to use the following positive quantities, with subscripts (rather than negative signs) indicating whether energy is flowing into or out of the gas:

- When work is done on the system by its surroundings, $W_{\text{in}} = W > 0$ from Eq. 59
- When work is done by the system on its surroundings, $W_{\text{out}} = -W > 0$
- When heat enters the system, $Q_{\text{in}} = Q > 0$ from Eq. 62
- When heat leaves the system, $Q_{\text{out}} = -Q > 0$

We can now define a convenient combination of heat and work to consider.

The **efficiency** η of a thermodynamic engine is defined to be

$$\eta = \frac{W_{\text{done}}}{Q_{\text{in}}} = \frac{W_{\text{out}} - W_{\text{in}}}{Q_{\text{in}}}, \quad (64)$$

where $W_{\text{done}} = W_{\text{out}} - W_{\text{in}}$ is the *net* amount of work done by each repetition of the cycle, while Q_{in} is the *total* amount of heat that enters the system in each repetition.

By specifying a thermodynamic *engine*, we assume $W_{\text{out}} > W_{\text{in}}$, so that the overall cycle does more work on its surroundings than it requires to operate. This corresponds to $\eta > 0$, and we can also put an upper bound on the efficiency, due to the first law of thermodynamics, Eq. 63. Because the system returns to its initial macro-state after each repetition of the cycle, we have

$$\begin{aligned}\Delta\langle E\rangle = 0 &= Q_{\text{in}} - Q_{\text{out}} + W_{\text{in}} - W_{\text{out}} \\ \implies W_{\text{out}} - W_{\text{in}} &= Q_{\text{in}} - Q_{\text{out}} \leq Q_{\text{in}},\end{aligned}\tag{65}$$

or $\eta \leq 1$, with equality occurring when no ‘waste’ heat is expelled during the cycle, so that $Q_{\text{out}} = 0$. All together, $0 < \eta \leq 1$ lets us interpret the efficiency as the fraction of the incoming heat that the engine is able to use to do work on its surroundings.

We can illustrate these ideas by computing the efficiency of the Carnot cycle. It is useful to divide this calculation into smaller pieces by considering the contributions to W_{done} and Q_{in} from each of the cycle’s four stages.

First, in the isothermal expansion from point A to point B , the ideal gas law provides $P(V)$ to insert into Eq. 59:

$$W_{AB} = - \int_{V_A}^{V_B} P(V) dV =$$

You should find $W_{AB} < 0$, meaning the system does work on its surroundings during this stage. At the same time, the constant temperature means $\Delta\langle E\rangle \propto \Delta T = 0$ from Eq. 52, so that $Q_{AB} = -W_{AB} > 0$, in agreement with our earlier observation of heat flowing into the system during this stage.

Next, in the adiabatic expansion from point B to point C , we know $Q_{BC} = 0$, which lets us use the first law of thermodynamics to compute the work:

$$W_{BC} =$$

You should find that the system continues doing work on its surroundings during this stage, $W_{BC} < 0$.

Finally, the computations for the two compression stages are directly analogous to those for the two expansion stages considered above. For the isothermal compression from point C to point D , the ideal gas law again provides $P(V)$:

$$W_{CD} = - \int_{V_C}^{V_D} P(V) dV =$$

Now you should find $W_{CD} > 0$, meaning this compression requires work to be done on the system by its surroundings, while $Q_{CD} = -W_{CD} < 0$ means heat flows out of the system. For the adiabatic compression from point D to point A , we know $Q_{DA} = 0$ while the change in temperature is exactly opposite the ΔT of the $B \rightarrow C$ adiabatic expansion. Therefore $W_{DA} = -W_{BC} > 0$ and more work has to be done on the system to complete the cycle.

Putting everything together,

$$\begin{aligned} W_{\text{out}} &= -W_{AB} - W_{BC} \\ W_{\text{in}} &= W_{CD} + W_{DA} = W_{CD} - W_{BC} \\ Q_{\text{in}} &= Q_{AB} = -W_{AB} \\ \eta &= \frac{-W_{AB} - W_{BC} - W_{CD} + W_{BC}}{-W_{AB}} = 1 + \frac{W_{CD}}{W_{AB}} = 1 - \frac{T_L}{T_H}. \end{aligned} \quad (66)$$

We can check that our result $\eta = 1 - \frac{T_L}{T_H}$ for the efficiency of the Carnot cycle makes sense. Since $T_L < T_H$, we have $\eta > 0$. If the temperatures of the hot and cold reservoirs approach each other, $\frac{T_L}{T_H} \rightarrow 1$, then the cycle would collapse to a single isotherm with $W_{\text{out}} = W_{\text{in}}$ and vanishing efficiency $\eta \rightarrow 0$. In the opposite limit of a large difference in the temperatures $T_L \ll T_H$, the efficiency would improve, with $\eta \rightarrow 1$ as $\frac{T_L}{T_H} \rightarrow 0$.

It turns out to be generic for heat engines to operate more efficiently as the temperature difference between their hot and cold reservoirs increases, and they always cease performing net work as $\frac{T_L}{T_H} \rightarrow 1$. The Carnot cycle is special because its efficiency $\eta = 1 - \frac{T_L}{T_H}$ is the theoretical maximum allowed by the second law of thermodynamics. We can show this by using Eq. 65 to rewrite

$$\eta = \frac{Q_{\text{in}} - Q_{\text{out}}}{Q_{\text{in}}} = 1 - \frac{Q_{\text{out}}}{Q_{\text{in}}} = 1 - \frac{T_L \Delta S_{\text{out}}}{T_H \Delta S_{\text{in}}},$$

where the last equality uses Eq. 62 and the fact that the input heat $Q_{\text{in}} = T_H \Delta S_{\text{in}}$ enters the engine from the hot reservoir with temperature T_H , while the waste heat $Q_{\text{out}} = T_L \Delta S_{\text{out}}$ is expelled to the cold reservoir with temperature T_L . After each repetition of the cycle, the gas returns to its original macro-state, with its original entropy, after absorbing entropy ΔS_{in} from its surroundings and expelling ΔS_{out} back out again. The

second law therefore demands $\Delta S_{\text{out}} \geq \Delta S_{\text{in}}$, so that

$$\eta = 1 - \frac{T_L}{T_H} \frac{\Delta S_{\text{out}}}{\Delta S_{\text{in}}} \leq 1 - \frac{T_L}{T_H}$$

in principle, for any thermodynamic engine.

Finally, if we were to operate the Carnot cycle in reverse, with isothermal expansion at temperature T_L and compression at T_H , we would do work on the system in order to bring heat in from the cold reservoir (i.e., $Q_{\text{in}} = T_L \Delta S_{\text{in}}$) and expel it into the hot reservoir ($Q_{\text{out}} = T_H \Delta S_{\text{out}}$). In other words, we would have a refrigerator rather than an engine. The 'efficiency' of a refrigerator is called its *coefficient of performance*, and defined as

$$\text{COP} = \frac{Q_{\text{in}}}{W_{\text{in}} - W_{\text{out}}} = \frac{Q_{\text{in}}}{Q_{\text{out}} - Q_{\text{in}}} = \frac{1}{Q_{\text{out}}/Q_{\text{in}} - 1} \leq \frac{1}{T_H/T_L - 1} = \frac{T_L}{T_H - T_L},$$

which can be greater than one. The reversed Carnot cycle provides the best possible COP for a refrigerator. Despite its efficiency, the Carnot cycle does not provide a practical engine or refrigerator, simply because its slow isothermal stages take too long! Real engines and refrigerators sacrifice efficiency for functionality.

Unit 6: Grand-canonical ensemble

6.1 The particle reservoir and chemical potential

Now that we have had some fun with applications of the canonical ensemble, we will complete our more formal development of statistical ensembles by considering the grand-canonical ensemble. Recall that statistical ensembles are probability spaces describing the micro-states that a system can adopt as it evolves in time, subject to certain constraints. Back in Unit 2 we first considered the micro-canonical ensemble, for which these constraints are conservation of the internal energy E and particle number N . We then introduced the canonical ensemble in Unit 3 by allowing the system's internal energy to fluctuate, while keeping its temperature T fixed through thermal contact with a large external thermal reservoir.

Building on this pattern, the next step is to allow *both* the system's energy and its particle number to fluctuate. Generalizing our earlier work on the canonical ensemble, these fluctuations occur through contact between the system and a large external reservoir. This is now a **particle reservoir**, with which the system can exchange both energy and particles.

In the same way that energy exchange leads to a fixed temperature, we expect there to be some quantity that will be fixed due to particle exchange. Recall that we initially defined the temperature in the context of the micro-canonical ensemble in thermodynamic equilibrium (Eq. 22), as the dependence of the entropy on the internal energy for a fixed number of degrees of freedom:

$$\frac{1}{T} = \left. \frac{\partial S}{\partial E} \right|_N.$$

The quantity we are now interested in comes from interchanging the roles of E and N .

In thermodynamic equilibrium, the **chemical potential** in the micro-canonical ensemble is defined by

$$\mu = -T \left. \frac{\partial S}{\partial N} \right|_E. \quad (67)$$

This definition is not very intuitive, and unlike the temperature the chemical potential is not a familiar concept from everyday experiences. To gain some insight into the chemical potential, we can first note that μ has dimensions of energy. It is also an intensive quantity, like the temperature — it is independent of the extent of the system, and remains the same if we consider only a part of a larger system. Finally, we can expect the chemical potential to be negative, at least for 'natural' systems with $T > 0$. This is because the partial derivative $\frac{\partial S}{\partial N}$ is generally positive — systems with more degrees of freedom typically have more entropy, reflecting the greater amount of information they can contain even with the energy fixed. This can be checked explicitly from Eq. 24 for the micro-canonical spin system we considered in Section 2.3.

The presence of the negative sign in Eq. 67 is really a choice we have made. The motivation for this choice comes from considering a net flow of particles between two systems Ω_A and Ω_B with the same temperature $T > 0$ but different

$$\frac{\partial S_A}{\partial N_A} > \frac{\partial S_B}{\partial N_B} > 0 \quad \implies \quad \mu_A < \mu_B < 0.$$

Due to the negative sign in Eq. 67, the system with the larger partial derivative has the smaller (more-negative) chemical potential. Particle exchange between these two systems, $\Delta N_A = -\Delta N_B$, causes the (extensive) entropy to change by

$$\Delta S = \Delta S_A + \Delta S_B = \left(\frac{\partial S_A}{\partial N_A} \right) \Delta N_A + \left(\frac{\partial S_B}{\partial N_B} \right) \Delta N_B = \left[\frac{\partial S_A}{\partial N_A} - \frac{\partial S_B}{\partial N_B} \right] \Delta N_A.$$

According to the second law of thermodynamics, $\Delta S \geq 0$, and since the term in square brackets is positive we must also have $\Delta N_A \geq 0$. Equivalently, we can note that more entropy is gained by adding each particle to system Ω_A than is lost by removing it from system Ω_B , ensuring that the total entropy of the universe doesn't decrease.

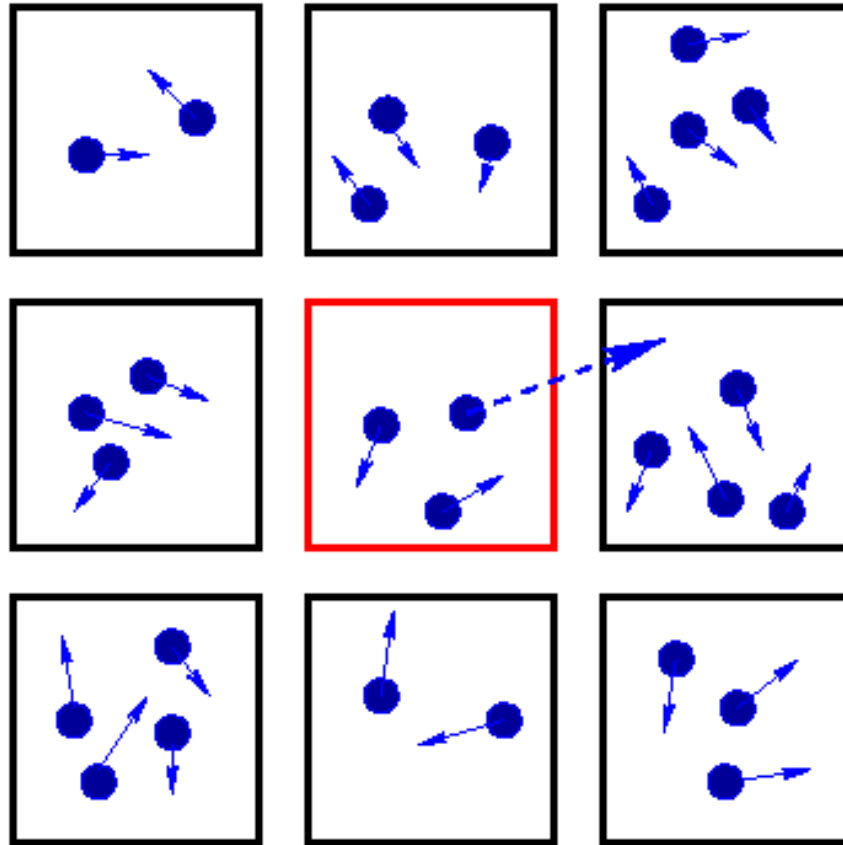
In other words, the choice of sign in Eq. 67 ensures that particles flow *from* systems with larger chemical potentials *to* systems with smaller chemical potential. This is similar to heat flowing from hotter systems with larger temperatures to colder systems with smaller temperatures, allowing us to reuse our intuition based on the temperature. Had we instead chosen to make μ positive for natural systems, we would have ended up with counter-intuitive flow of particles from small to large chemical potential. This connection to particle flow also implies that we should consider only systems of indistinguishable particles when working with a chemical potential.

We are now able to define the **grand-canonical ensemble** to be a statistical ensemble characterized by its fixed temperature T and fixed chemical potential μ , with the temperature and chemical potential held fixed through contact with a particle reservoir.

6.2 The grand-canonical partition function

Let's now place the grand-canonical ensemble on a more concrete mathematical foundation, by following the same procedure we used for the canonical ensemble. That is, we introduce a well-motivated ansatz for the form of the particle reservoir Ω_{res} , then show that the form of the reservoir is ultimately irrelevant. This will allow us to work directly with the system of interest, Ω , independent of the details of the particle reservoir that fixes its temperature and chemical potential.

As before, our ansatz is to take $\Omega_{\text{tot}} = \Omega_{\text{res}} \otimes \Omega$ to consist of many ($R \gg 1$) identical replicas of the system Ω that we're interested in. All of these replicas are in thermodynamic equilibrium, and can exchange both energy and particles with each other. The overall system Ω_{tot} is governed by the micro-canonical ensemble, with conserved total energy E_{tot} and conserved total particle number N_{tot} . An extremely small example of this setup is illustrated by the figure on the next page, where the system of interest is an ideal gas in a volume V .



Although we draw a box around each replica (and colour one red to highlight the system Ω we will consider), these boxes are now merely mental constructions, and don't interfere with particles moving from one replica to another. For example, we could take our system to be a cubic centimetre of air in a room, with the rest of the room forming its reservoir. As in Section 3.1.1, we assume that this system $\Omega = \{\omega_1, \omega_2, \dots, \omega_M\}$ has a finite number of M possible micro-states, where now different micro-states may involve different numbers of particles.

This again allows us to analyse the overall system of R replicas in terms of occupation numbers n_i and the corresponding occupation probabilities p_i . Recall that n_i is the number of replicas that adopt the micro-state $\omega_i \in \Omega$ in any given micro-state of the overall system Ω_{tot} , so that $\sum_i n_i = R$. Similarly, $p_i = n_i/R$ is the probability that a randomly chosen replica will be in micro-state ω_i , with $\sum_i p_i = 1$ as usual. In terms of n_i and p_i , the total number of micro-states of Ω_{tot} , and the corresponding entropy, are the same as we derived in Section 3.1.2,

$$M_{\text{tot}} = \frac{R!}{n_1! n_2! \dots n_M!} \quad \longrightarrow \quad S(E_{\text{tot}}, N_{\text{tot}}) = -R \sum_{i=1}^M p_i \log p_i,$$

assuming $R \gg 1$ and $n_i \gg 1$ for all $i = 1, \dots, M$. In this expression, the dependence on both E_{tot} and N_{tot} now enters through the occupation probabilities p_i , since the micro-states ω_i may involve different numbers of particles in addition to different energies.

As before, we want to determine the (intensive) temperature and chemical potential of the micro-canonical system Ω_{tot} through Eqs. 22 and 67, which requires expressing S directly in terms of E_{tot} and N_{tot} . We again proceed by maximizing the entropy

subject to the constraints on the conserved quantities of Ω_{tot} . Labelling the energy and particle number of each replica E_r and N_r , respectively, as in Eq. 27 we can rearrange sums over replicas into sums over the micro-states of Ω :

$$\begin{aligned}
 1 &= \sum_{i=1}^M p_i & E_{\text{tot}} &= \sum_{r=1}^R E_r = \sum_{i=1}^M n_i E_i = R \sum_{i=1}^M p_i E_i \\
 N_{\text{tot}} &= \sum_{r=1}^R N_r = \sum_{i=1}^M n_i N_i = R \sum_{i=1}^M p_i N_i, & & (68)
 \end{aligned}$$

where E_i and N_i are the energies and particle numbers of the M micro-states $\omega_i \in \Omega$. The first two constraints, on the occupation probabilities and the total energy, are the same as we had in Section 3.1.2. The third constraint, on the total particle number, is the new ingredient for us to incorporate.

Writing everything in terms of occupation probabilities, we see that we need to maximize the modified entropy

$$\begin{aligned}
 \bar{S} &= -R \sum_{i=1}^M p_i \log p_i + \alpha \left(\sum_{i=1}^M p_i - 1 \right) \\
 &\quad - \beta \left(R \sum_{i=1}^M p_i E_i - E_{\text{tot}} \right) + \gamma \left(R \sum_{i=1}^M p_i N_i - N_{\text{tot}} \right),
 \end{aligned}$$

adding the Lagrange multiplier γ to the α and (negative) β we previously had in Section 3.1.2. What is the occupation probability p_k that maximizes \bar{S} ?

$$0 = \frac{\partial \bar{S}}{\partial p_k} =$$

You should find a probability of the form

$$p_k = \frac{1}{Z_g} e^{-\beta E_k + \gamma N_k}, \quad (69)$$

defining $Z_g = \exp \left[1 - \frac{\alpha}{R} \right]$ to work in terms of the parameters $\{Z_g, \beta, \gamma\}$. As usual, we fix these three parameters by demanding that the three constraints above are satisfied.

Using the first constraint, what is Z_g in terms of β and γ ?

$$1 = \sum_{i=1}^M p_i =$$

Analogously to Eq. 31 in Section 3.1.2, the other two constraints

$$E_{\text{tot}} = R \sum_i p_i E_i \qquad N_{\text{tot}} = R \sum_i p_i N_i$$

now produce complicated relations between $\{\beta, \gamma\}$ and $\{E_{\text{tot}}, N_{\text{tot}}\}$:

$$E_{\text{tot}} = R \frac{\sum_{i=1}^M E_i e^{-\beta E_i + \gamma N_i}}{\sum_{j=1}^M e^{-\beta E_j + \gamma N_j}} = -\frac{R}{Z_g} \frac{\partial}{\partial \beta} \sum_{i=1}^M e^{-\beta E_i + \gamma N_i} = -R \frac{\partial}{\partial \beta} \log Z_g(\beta, \gamma) \quad (70)$$

$$N_{\text{tot}} = R \frac{\sum_{i=1}^M N_i e^{-\beta E_i + \gamma N_i}}{\sum_{j=1}^M e^{-\beta E_j + \gamma N_j}} = \frac{R}{Z_g} \frac{\partial}{\partial \gamma} \sum_{i=1}^M e^{-\beta E_i + \gamma N_i} = R \frac{\partial}{\partial \gamma} \log Z_g(\beta, \gamma). \quad (71)$$

These relations between $\{E_{\text{tot}}, N_{\text{tot}}\}$ and partial derivatives of $\log Z_g$ will be useful when we consider the partial derivatives of the entropy that define the temperature and chemical potential. Of course, in order to consider those partial derivatives, we need to express the entropy itself in terms of E_{tot} , N_{tot} and the parameters

$$\{Z_g(\beta, \gamma), \quad \beta(E_{\text{tot}}, N_{\text{tot}}), \quad \gamma(E_{\text{tot}}, N_{\text{tot}})\}$$

that we have now related to E_{tot} and N_{tot} . What do you obtain upon inserting Eq. 69 for p_i into the formula for the entropy?

$$S(E_{\text{tot}}, N_{\text{tot}}) = -R \sum_{i=1}^M p_i \log p_i =$$

Taking the derivative of the resulting entropy with respect to E_{tot} , keeping N_{tot} fixed, gives the temperature from Eq. 22. Thanks to Eqs. 70 and 71, the result should simplify in a pleasant way:

$$\frac{1}{T} = \left. \frac{\partial S}{\partial E_{\text{tot}}} \right|_{N_{\text{tot}}} =$$

In the same way, the derivative with respect to N_{tot} , keeping E_{tot} fixed, gives the chemical potential with similar simplifications:

$$\mu = -T \left. \frac{\partial S}{\partial N_{\text{tot}}} \right|_{E_{\text{tot}}} =$$

In the end you should find

$$\beta = \frac{1}{T} \qquad \gamma = \beta\mu = \frac{\mu}{T}, \qquad (72)$$

meaning that (as desired) all information about the particle reservoir has dropped out, with no remaining reference to R , E_{tot} or N_{tot} . This large external reservoir is still present to fix the temperature T and chemical potential μ that characterize the grand-canonical system Ω , but beyond that nothing about it is relevant (or even knowable) in the grand-canonical approach.

Every aspect of Ω can now be specified in terms of its fixed temperature T and chemical potential μ , starting with the parameters $\beta = 1/T$ and $\gamma = \mu/T$. In particular, the thermodynamic equilibrium probability that Ω adopts micro-state ω_i with (non-conserved) internal energy E_i and particle number N_i is

$$p_i = \frac{1}{Z_g} e^{-\beta(E_i - \mu N_i)} = \frac{1}{Z_g} e^{-(E_i - \mu N_i)/T}. \qquad (73)$$

Since the particle number N_i is dimensionless, the combination $E_i - \mu N_i$ that appears here reflects our observation below Eq. 67 that the chemical potential μ has dimensions of energy.

These micro-state probabilities are normalized by the **grand-canonical partition function**

$$Z_g(T, \mu) = \sum_{i=1}^M e^{-\beta(E_i - \mu N_i)} = \sum_{i=1}^M e^{-(E_i - \mu N_i)/T}. \quad (74)$$

Analogously to the canonical partition function, this Z_g is a fundamental quantity in the grand-canonical ensemble, from which many other derived quantities can be obtained.

6.3 The grand-canonical potential, internal energy, entropy, and particle number

The development of the grand-canonical ensemble we have seen so far closely resembles our earlier work setting up the canonical ensemble. We have generalized the thermal reservoir to a particle reservoir that fixes both the temperature T and chemical potential μ , while allowing the system's internal energy and particle number to vary between different micro-states ω_i . By adapting the replica ansatz to this setup, we determined the micro-state probabilities p_i and the grand-canonical partition function Z_g , confirming that they are independent of the details of the particle reservoir.

We now continue by considering a similar set of derived quantities for the grand-canonical ensemble in thermodynamic equilibrium. In addition to the expectation value of the internal energy introduced in Section 3.2, the fluctuations of the particle number mean that we also need to consider its expectation value,

$$\begin{aligned} \langle E \rangle(T, \mu) &= \sum_{i=1}^M E_i p_i = \frac{1}{Z_g} \sum_{i=1}^M E_i e^{-\beta(E_i - \mu N_i)} \\ \langle N \rangle(T, \mu) &= \sum_{i=1}^M N_i p_i = \frac{1}{Z_g} \sum_{i=1}^M N_i e^{-\beta(E_i - \mu N_i)}. \end{aligned}$$

Looking back to Eqs. 70 and 71, we can expect both of these derived quantities to be related to derivatives of the logarithm of the grand-canonical partition function. In Section 3.3, similar relations led us to define the Helmholtz free energy for the canonical ensemble, which we can also generalize to the grand-canonical case.

We define the **grand-canonical potential** of a grand-canonical ensemble to be

$$\Phi(T, \mu) = -T \log Z_g(T, \mu) = -\frac{\log Z_g(\beta, \mu)}{\beta}, \quad (75)$$

where Z_g is the grand-canonical partition function of the ensemble. In terms of this quantity, Eqs. 73 and 74 are

$$Z_g = e^{-\Phi/T} \qquad p_i = e^{(\Phi - E_i + \mu N_i)/T}.$$

The grand-canonical potential is similar to the Helmholtz free energy, and to highlight this similarity it is sometimes called the *Landau free energy* (named after [Lev Landau](#)). As mentioned above, we want to consider derivatives of the grand-canonical potential, the simplest of which is with respect to the chemical potential,

$$\frac{\partial}{\partial \mu} \Phi(\beta, \mu) =$$

The derivative with respect to the temperature is a little messier, but can be simplified by recalling $\frac{\partial}{\partial T} = -\beta^2 \frac{\partial}{\partial \beta}$ from Eq. 35. As in Section 3.3, it involves $\frac{\partial}{\partial T} \log Z_g$, which is again worth collecting in advance,

$$-\frac{\partial}{\partial T} \left[\frac{\Phi(T, \mu)}{T} \right] = \frac{\partial}{\partial T} \log Z_g(T, \mu) =$$

$$\frac{\partial}{\partial T} \Phi(T, \mu) =$$

You should find

$$\frac{\partial \Phi}{\partial T} = \frac{\Phi - \langle E \rangle + \mu \langle N \rangle}{T} = -\log Z_g - \beta \langle E \rangle + \beta \mu \langle N \rangle,$$

which we can connect to the entropy by inserting the probabilities p_i from Eq. 73 into the general definition of the entropy from Eq. 20:

$$S(T, \mu) = - \sum_{i=1}^M p_i \log p_i =$$

From this work we can read off the following relations involving the grand-canonical potential $\Phi(T, \mu)$:

$$\langle N \rangle(T, \mu) = - \frac{\partial}{\partial \mu} \Phi \quad (76)$$

$$S(T, \mu) = - \frac{\partial}{\partial T} \Phi \quad (77)$$

$$\langle E \rangle(T, \mu) = -T^2 \frac{\partial}{\partial T} \left[\frac{\Phi}{T} \right] + \mu \langle N \rangle \quad (78)$$

$$\Phi(T, \mu) = -T S + \langle E \rangle - \mu \langle N \rangle \quad (79)$$

Finally, the connections between the energy, entropy and particle number provided by these relations motivate a further extension of the generalized first law of thermodynamics we derived in Eq. 63. To make the notation less cumbersome here, we write $\langle E \rangle$ and $\langle N \rangle$ as E and N , keeping in mind that these are properties of the system's thermodynamic macro-state rather than its fluctuating micro-state. In this notation, Eq. 63 reads $dE = T dS - P dV = Q + W$, and relates any changes in the internal energy of a canonical system to changes in its entropy (heat) or volume (work).

Extending this to the grand-canonical ensemble, we can express the entropy as a function of the internal energy, volume and particle number, $S(E, V, N)$, and consider the change in entropy due to changes in each of these three parameters,

$$dS = \left. \frac{\partial S}{\partial E} \right|_{V,N} dE + \left. \frac{\partial S}{\partial V} \right|_{E,N} dV + \left. \frac{\partial S}{\partial N} \right|_{V,E} dN = \frac{1}{T} dE + \left. \frac{\partial S}{\partial V} \right|_{E,N} dV - \frac{\mu}{T} dN.$$

We can interpret the remaining partial derivative by considering Eq. 63 in the case of fixed internal energy E . This equation already incorporates the fixed particle number N , since it was derived in the framework of the canonical ensemble:

$$dE = 0 = T dS - P dV \quad \implies \quad \left. \frac{\partial S}{\partial V} \right|_{E,N} = \frac{P}{T}.$$

Putting things together, we obtain the generalized thermodynamic identity

$$dE = T dS - P dV + \mu dN . \quad (80)$$

Due to this result, the term μdN is sometimes referred to as “chemical work”, in analogy to the mechanical work $W = -P dV$ done on a system by changing its volume. This thermodynamic identity provides a convenient way to remember (or derive) relations between the internal energy, entropy, volume and particle number in thermodynamic equilibrium, by considering processes in which any two of these are fixed. For example, fixing N and V gets us back to Eq. 22 for the temperature,

$$dE = T dS \quad \Longrightarrow \quad \frac{1}{T} = \left. \frac{\partial S}{\partial E} \right|_{N,V} ,$$

while fixing N and S gives Eq. 55 for the pressure,

$$dE = -P dV \quad \Longrightarrow \quad P = - \left. \frac{\partial E}{\partial V} \right|_{N,S} .$$

If we fix the entropy S and volume V , we end up with another way of understanding the chemical potential,

$$dE = \mu dN \quad \Longrightarrow \quad \mu = \left. \frac{\partial E}{\partial N} \right|_{S,V} . \quad (81)$$

That is, the chemical potential is the change in the internal energy when we change the number of particles in the system, without changing its entropy or volume. If we consider adding particles to the system, $\Delta N > 0$, we argued below Eq. 67 that we should generically expect an increase in the entropy. In order to keep the entropy fixed in this process, we therefore need the change in the energy to decrease the entropy by the corresponding amount. For natural systems with positive temperatures, this requires decreasing the energy, $\Delta E < 0$. Similarly, keeping the entropy fixed as we decrease N would require increasing E , so that Eq. 81 confirms our earlier finding that for natural systems the chemical potential is negative in general.

Unit 7: Quantum statistics

7.1 Quantized energy levels and their micro-states

Now that we have defined the grand-canonical ensemble, we will apply it to investigate quantum statistical systems. The first step is to introduce quantum statistics itself — it is worth reiterating that no prior knowledge of quantum physics is assumed, nor will this module attempt to teach quantum mechanics. We will simply consider quantum behaviour as an ansatz (that turns out to be realized in nature), and analyse the resulting systems by making use of the statistical mechanics tools we have developed.

Quantum physics was mentioned in passing during our derivation of the canonical partition function for a classical (non-quantum) ideal gas in Section 4.1. You may have felt that this derivation involved seemingly circular reasoning. First, because the partition function is defined as a sum over micro-states ω_i ,

$$Z = \sum_i e^{-E(\vec{p}_i)/T},$$

we regularized the system by supposing that the gas particles' momenta \vec{p}_i are *quantized* and can take only particular discrete values, rather than varying continuously. These quantized momenta produce a countable number of discrete *energy levels*, leading to a countable number of micro-states and hence a well-defined partition function that involves sums over all possible discrete momenta for each particle. Second, after the system was regularized, we switched back to continuously varying momenta, converting sums to integrals while retaining the dimensionful factors we had derived along the way.

Two changes are required to define quantum statistics. First, not surprisingly, we need to retain discrete energy levels rather than approximating these as continuous. This will allow our calculations to remain valid even in the quantum regime where $L\sqrt{mT} \sim \hbar$ (or equivalently $\lambda_{\text{th}}^3 \sim V$). The second change is more subtle, and is connected to the fundamental indistinguishability of identical particles governed by quantum mechanics — a fact about nature that we will take as given. The issue is how to handle micro-states in which multiple indistinguishable particles occupy the same energy level.

To build up to this issue, we will first see what happens when we ignore it and apply our usual classical approach to compute the grand-canonical partition function for a system with discrete energy levels. Despite the quantized energy levels, this calculation will still produce a non-quantum result known as **Maxwell–Boltzmann** (MB) statistics, named after [James Clerk Maxwell](#) and Ludwig Boltzmann. We will then consider how this approach can break down, and use this insight to develop true quantum statistics in the following sections. Finally, we will wrap up this unit by confirming that Maxwell–Boltzmann statistics remains an excellent approximation to quantum statistics in the classical limit.

7.1.1 Maxwell–Boltzmann statistics

Let's begin our classical derivation of the grand-canonical partition function for a system with discrete energy levels by defining some necessary notation. We label the discrete energy levels \mathcal{E}_ℓ for $\ell = 0, 1, \dots, \mathcal{L}$, where \mathcal{L} can be taken to infinity while retaining a countable number of micro-states and hence well-defined partition functions. The energy level \mathcal{E}_ℓ may be characterized by extra information in addition to the actual value of its energy, E_ℓ . As we saw in Section 4.1, it is therefore possible for distinct energy levels $\{\mathcal{E}_m, \mathcal{E}_n\}$ to have the same energy $E_m = E_n$ for $m \neq n$. Such energy levels with the same value of the energy are said to be *degenerate*. We will label energy levels so that $E_m \leq E_n$ for $m < n$. Without loss of generality, we can take $E_\ell \geq E_0 \geq 0$.

Now starting from the general expression for the grand-canonical partition function, Eq. 74,

$$Z_g(\beta, \mu) = \sum_i e^{-\beta(E_i - \mu N_i)},$$

we just need to define the micro-states ω_i with energy E_i and particle number N_i . In the classical Maxwell–Boltzmann approach, we arrange this into a series of sums with fixed particle numbers:

$$Z_g^{\text{MB}}(\beta, \mu) = \sum_{i, N_i=0} e^{-\beta E_i} + \sum_{j, N_j=1} e^{-\beta(E_j - \mu)} + \sum_{k, N_k=2} e^{-\beta(E_k - 2\mu)} + \dots,$$

where the micro-states labelled $\{\omega_i, \omega_j, \omega_k, \dots\}$ are those that have $N = 0, 1, 2, \dots$ particles, respectively. We can recognize N -particle canonical partition functions $Z_N(\beta)$ in the expression above:

$$Z_g^{\text{MB}}(\beta, \mu) = Z_0(\beta) + e^{\beta\mu} Z_1(\beta) + e^{2\beta\mu} Z_2(\beta) + \dots = \sum_{N=0}^{\infty} [e^{\beta\mu}]^N Z_N(\beta). \quad (82)$$

This is a general result known as the *fugacity expansion*, where $e^{\beta\mu}$ is called the fugacity. Organizing the calculation in this way allows us to take advantage of our experience with the canonical ensemble.

In particular, because we continue to consider only ‘ideal’ systems in which the particles don't interact with each other, each $Z_N(\beta)$ is simply the product of the single-particle partition functions $Z_1(\beta)$ for all N independent particles,

$$Z_N(\beta) = \frac{1}{N!} [Z_1(\beta)]^N,$$

with the factor of $N!$ included to correct for over-counting indistinguishable particles. This is exactly the derivation we performed in Section 4.1, to obtain Eq. 51 for the classical ideal gas. Inserting this Z_N into Eq. 82, we have

$$Z_g^{\text{MB}}(\beta, \mu) = \sum_{N=0}^{\infty} [e^{\beta\mu}]^N \frac{1}{N!} [Z_1(\beta)]^N = \sum_{N=0}^{\infty} \frac{1}{N!} [e^{\beta\mu} Z_1(\beta)]^N = \exp [e^{\beta\mu} Z_1(\beta)].$$

In the case of a system with discrete energy levels \mathcal{E}_ℓ , the single-particle partition function is simply

$$Z_1(\beta) = \sum_{\ell=0}^{\mathcal{L}} e^{-\beta E_\ell},$$

where each micro-state has the particle in a different energy level. This gives us the Maxwell–Boltzmann grand-canonical partition function

$$Z_g^{\text{MB}}(\beta, \mu) = \exp \left[e^{\beta\mu} \sum_{\ell=0}^{\mathcal{L}} e^{-\beta E_{\ell}} \right] = \exp \left[\sum_{\ell=0}^{\mathcal{L}} e^{-\beta(E_{\ell}-\mu)} \right]. \quad (83)$$

7.1.2 Over-counting and quantum statistics

The problem with the derivation above was already mentioned in the footnote accompanying Eq. 51. Recall that the factor of $\frac{1}{N!}$ comes from the $N!$ different ways the N particles could be labeled — since indistinguishable particles cannot be labeled, we have only a single micro-state rather than the $N!$ micro-states arising from Z_1^N itself. We will see that this analysis is only correct when each of the particles is in a different energy level. When the particles' energies can vary continuously it is practically impossible for two of them to have exactly the same energy, making it safe to assume a different energy level for each particle. More generally, this assumption can remain an excellent approximation whenever there are many more accessible energy levels than there are particles to occupy them. Here we are interested in situations where there is a non-negligible chance of multiple particles occupying the same energy level, which we can call the 'quantum regime').

Let's illustrate the problem by considering a simple system with $N = 2$ particles that can each occupy any of five energy levels, \mathcal{E}_0 through \mathcal{E}_4 . The canonical partition function for this system sums over all possible ways these particles can occupy the available energy levels — equivalent to the ways $N = 2$ balls can be placed into $\mathcal{L} + 1 = 5$ boxes. We can represent possible micro-states by drawing these balls and boxes, for example $\square \bullet \square \square \square$ and $\square \square \bullet \bullet \square$. How many terms are there in the sum for the distinguishable partition function, Z_D ?

For indistinguishable particles, our earlier derivation predicts there are half as many ($\frac{1}{2!}$) terms in the sum for the indistinguishable partition function, Z_I , but this would be a non-integer number of terms and therefore cannot be correct.

We can see where the derivation goes wrong by explicitly writing down all the micro-states in both cases of distinguishable and indistinguishable particles. In the distinguishable case, we can suppose that the balls are red (\bullet) and blue (\circ), and compactly represent micro-states by recording whether each box is empty ("0") or contains the red ball ("R"), the blue ball ("B") or both balls ("2"):

$$\square \square \bullet \square \circ = 00R0B \qquad \square \square \bullet \circ \square = 00200.$$

The full set of micro-states is then

$RB000$	$0R0B0$	$BR000$	$0B0R0$	20000
$R0B00$	$0R00B$	$B0R00$	$0B00R$	02000
$R00B0$	$00RB0$	$B00R0$	$00BR0$	00200
$R000B$	$00R0B$	$B000R$	$00B0R$	00020
$0RB00$	$000RB$	$0BR00$	$000BR$	00002

If we now consider indistinguishable particles where we can only know the number of particles in each box ($R, B \rightarrow 1$), we see that the third and fourth columns above duplicate the first two columns. This is exactly the over-counting that the usual factor of $\frac{1}{N!} = \frac{1}{2}$ corrects. On the other hand, the micro-states in the final column, with both particles in the same energy level, were not over-counted, and must not be divided by $N!$. This leaves us with 15 micro-states, rather than $25/2$:

11000	01010	20000
10100	01001	02000
10010	00110	00200
10001	00101	00020
01100	00011	00002

We can generalize this simple exercise by systematically labeling the micro-states for indistinguishable particles by *occupation numbers* n_ℓ , similar to those that we encountered when using replicas to set up the canonical ensemble in Section 3.1 and the grand-canonical ensemble in Section 6.2. In this case the occupation number n_ℓ is simply the number of particles in energy level \mathcal{E}_ℓ . This new approach is the second change promised above, and the final ingredient needed for the following definition.

In **quantum statistics**, the micro-states are defined by considering each energy level \mathcal{E}_ℓ in turn, and summing over the possible occupation numbers n_ℓ that it could have. This differs from the classical approach in which the micro-states were defined by considering each particle in turn, and summing over the possible energies (etc.) that it could have.

Because quantum mechanics requires all particles of the same type to be indistinguishable, the classical approach forces us to correct for over-counting, and we have now seen how this becomes non-trivial whenever multiple particles can occupy the same energy level. The quantum approach of summing over the occupation numbers of the quantized energy levels avoids this issue, and doesn't encounter any over-counting that would require correction.

7.2 Bosons and fermions

In Sections 7.3 and 7.4 we will carry out explicit computations to show how the quantum statistics defined above work in practice. In preparation, we will take into account some additional information about nature, specifically concerning the occupation

numbers n_ℓ that are possible for each energy level \mathcal{E}_ℓ . As before, for the purposes of this module we can simply consider this information to be part of our ansatz.

Using quantum mechanics and special relativity, it is possible to prove that all particles in three spatial dimensions fall into two distinct classes. ([More exotic behaviour](#) is possible for particles confined to two-dimensional surfaces.) This result is known as the [spin–statistics theorem](#), while the two types of particles it describes are called **bosons** (named after [Satyendra Nath Bose](#)) and **fermions** (named after [Enrico Fermi](#)). These two classes of particles obey different rules for their possible occupation numbers, and therefore give rise to distinct quantum statistics.

Any non-negative number of identical bosons can simultaneously occupy the same energy level, corresponding to occupation numbers $n_\ell = 0, 1, 2, \dots$. Physical examples of bosons include photons (particles of light), pions, helium-4 atoms and the famous Higgs particle.

In contrast, it is impossible for multiple identical fermions to occupy the same energy level, meaning that their only possible occupation numbers are $n_\ell = 0$ and 1 . This behaviour is known as the *Pauli exclusion principle* (named after [Wolfgang Pauli](#)) and has extremely important consequences, including the existence of chemistry and life. Physical examples of fermions include electrons, protons, neutrons, neutrinos and helium-3 atoms.

The reason multiple identical fermions cannot occupy the same energy level is due to a feature of quantum mechanics, and not because they physically repel each other. This paragraph will imprecisely describe how this works, for those who may be curious, and can be skipped without any problem. Consider a system of identical quantum particles occupying various energy levels. Loosely speaking, all observable properties of this system depend on the *square* of the quantum function that defines it. Interchanging any pair of indistinguishable particles must leave all these observable properties unchanged. Just as $\sqrt{1} = \pm 1$, there are two ways the underlying quantum function can behave to leave its square unchanged: it can be completely symmetric or completely antisymmetric under all possible interchanges. Bosons correspond to the symmetric case, while fermions correspond to the antisymmetric case. At the same time, if two identical particles occupy the same energy level, then the quantum function itself is unchanged (i.e., symmetric) when they are interchanged. In the fermionic case, the resulting quantum function must therefore be simultaneously symmetric and antisymmetric, which is only possible if it is exactly zero. In other words no systems with multiple identical fermions in the same energy level can possibly exist.

Looking back at the example system of $N = 2$ particles with five energy levels in the previous section, all 15 micro-states we wrote down are possible if the particles are bosons. How many micro-states are allowed if the particles are fermions?

This difference in the possible micro-states ensures that bosons and fermions exhibit different quantum statistics. We will now consider each case in turn.

7.3 Bose–Einstein statistics

The quantum statistics of bosons is known as **Bose–Einstein** (BE) statistics, named after Satyendra Nath Bose and Albert Einstein. As described above, to carry out the sum over all micro-states in the grand-canonical partition function

$$Z_g(\beta, \mu) = \sum_i e^{-\beta(E_i - \mu N_i)},$$

we first sum over all energy levels \mathcal{E}_ℓ , and then over all possible occupation numbers $n_\ell \in \mathbb{N}_0$ for each energy level.

Consider first the simple case of a system that only has a single energy level \mathcal{E}_0 , with energy E_0 . In this case, each micro-state ω_i is uniquely described by its particle number N_i , which is just the occupation number of \mathcal{E}_0 . What is the energy E_i of micro-state ω_i with occupation number $n_0 = N_i$?

$E_i =$

Summing over all possible occupation numbers for this single energy level, the Bose–Einstein grand-canonical partition function for this system is

$$Z_g^{\text{BE}}(\beta, \mu) = \sum_{n_0=0}^{\infty} e^{-\beta(E_0 - \mu)n_0} = \sum_{n_0=0}^{\infty} [e^{-\beta(E_0 - \mu)}]^{n_0} = \frac{1}{1 - e^{-\beta(E_0 - \mu)}}. \quad (84)$$

In the last step we recognized the geometric series

$$\frac{1}{1 - x} = 1 + x + x^2 + \dots,$$

which only converges for $x = e^{-\beta(E_0 - \mu)} < 1$. For natural systems with $\beta = 1/T > 0$, this condition requires $E_0 - \mu > 0$ or equivalently $\mu < E_0$. Since we can take all $E_\ell \geq 0$ without loss of generality, this constraint is consistent with the negative chemical potential $\mu < 0$ that we discussed in Unit 6.

At this point, it is straightforward to generalize to multiple energy levels \mathcal{E}_ℓ with $\ell = 0, 1, \dots, \mathcal{L}$. For systems of non-interacting particles, the micro-state ω_i defined by the set of occupation numbers $\{n_\ell\}$ has energy and particle number

$$E_i = \sum_{\ell=0}^{\mathcal{L}} E_\ell n_\ell \qquad N_i = \sum_{\ell=0}^{\mathcal{L}} n_\ell. \quad (85)$$

The general Bose–Einstein grand-canonical partition function is therefore

$$Z_g^{\text{BE}}(\beta, \mu) = \sum_{n_0=0}^{\infty} \sum_{n_1=0}^{\infty} \cdots \sum_{n_{\mathcal{L}}=0}^{\infty} \exp \left[-\beta \sum_{\ell=0}^{\mathcal{L}} (E_\ell - \mu) n_\ell \right].$$

We can apply the identity $e^{\sum_i x_i} = \prod_i e^{x_i}$ to rewrite

$$\exp \left[-\beta \sum_{\ell=0}^{\mathcal{L}} (E_{\ell} - \mu) n_{\ell} \right] = (e^{-\beta(E_0 - \mu)n_0}) (e^{-\beta(E_1 - \mu)n_1}) \dots (e^{-\beta(E_{\mathcal{L}} - \mu)n_{\mathcal{L}}}).$$

Recalling $\mu < E_{\ell}$ for all ℓ , by rearranging the terms we find

$$\begin{aligned} Z_g^{\text{BE}}(\beta, \mu) &= \left(\sum_{n_0=0}^{\infty} e^{-\beta(E_0 - \mu)n_0} \right) \left(\sum_{n_1=0}^{\infty} e^{-\beta(E_1 - \mu)n_1} \right) \dots \left(\sum_{n_{\mathcal{L}}=0}^{\infty} e^{-\beta(E_{\mathcal{L}} - \mu)n_{\mathcal{L}}} \right) \\ &= \prod_{\ell=0}^{\mathcal{L}} \frac{1}{1 - e^{-\beta(E_{\ell} - \mu)}}. \end{aligned} \quad (86)$$

This grand-canonical partition function defines the quantum Bose–Einstein statistics of bosons. Its structure as the product of an independent contribution for each energy level is reminiscent of the result $Z_N \propto Z_1^N$ for the classical N -particle canonical partition function discussed above. In such situations we say that the calculation *factorizes* into a product of many simpler terms, allowing us to build up the full result on the basis of much easier computations. Looking back to Eq. 83, we can also observe factorization in the classical Maxwell–Boltzmann grand-canonical partition function,

$$Z_g^{\text{MB}}(\beta, \mu) = \exp \left[\sum_{\ell=0}^{\mathcal{L}} e^{-\beta(E_{\ell} - \mu)} \right] = \prod_{\ell=0}^{\mathcal{L}} \exp [e^{-\beta(E_{\ell} - \mu)}]. \quad (87)$$

In both cases factorization results from working with non-interacting particles. Starting in Unit 9 we will consider non-ideal systems in which the particles can interact with each other, where the absence of factorization will make analyses much more difficult.

7.4 Fermi–Dirac statistics

Next, the quantum statistics of fermions is known as **Fermi–Dirac** (FD) statistics, named after Enrico Fermi and [Paul Dirac](#). The derivation of the Fermi–Dirac grand-canonical partition function is very similar to the Bose–Einstein case above. We again proceed by summing over all energy levels \mathcal{E}_{ℓ} , and just have to account for the more limited possible occupation numbers $n_{\ell} \in \{0, 1\}$ for each energy level.

Taking the same approach of first considering a simple system with only a single energy level, Eq. 84 would just change to

$$Z_g^{\text{FD}}(\beta, \mu) = \sum_{n_0=0}^1 e^{-\beta(E_0 - \mu)n_0} = 1 + e^{-\beta(E_0 - \mu)}.$$

Generalizing to multiple energy levels \mathcal{E}_{ℓ} with $\ell = 0, 1, \dots, \mathcal{L}$, the micro-state energies $E_i = \sum_{\ell} E_{\ell} n_{\ell}$ and particle numbers $N_i = \sum_{\ell} n_{\ell}$ remain the same as in Eq. 85, and the

computation again factorizes,

$$\begin{aligned}
 Z_g^{\text{FD}}(\beta, \mu) &= \sum_{n_0=0}^1 \sum_{n_1=0}^1 \cdots \sum_{n_{\mathcal{L}}=0}^1 \exp \left[-\beta \sum_{\ell=0}^{\mathcal{L}} (E_{\ell} - \mu) n_{\ell} \right] \\
 &= \left(\sum_{n_0=0}^1 e^{-\beta(E_0 - \mu)n_0} \right) \left(\sum_{n_1=0}^1 e^{-\beta(E_1 - \mu)n_1} \right) \cdots \left(\sum_{n_{\mathcal{L}}=0}^1 e^{-\beta(E_{\mathcal{L}} - \mu)n_{\mathcal{L}}} \right) \\
 &= \prod_{\ell=0}^{\mathcal{L}} [1 + e^{-\beta(E_{\ell} - \mu)}].
 \end{aligned} \tag{88}$$

This grand-canonical partition function defines the quantum Fermi–Dirac statistics of fermions. In both this case and the case of classical Maxwell–Boltzmann statistics there is no constraint on $\beta(E_{\ell} - \mu)$.

In Unit 8 we will take Z_g^{BE} and Z_g^{FD} as starting points to analyse quantum gases of bosons and fermions, respectively. Before beginning those more detailed analyses, let’s quickly compare the three types of statistics that we have derived in this unit, while they are all close to hand.

7.5 The classical limit

In Section 7.1 we claimed that if the probability of multiple particles occupying the same energy level is negligible, then the classical Maxwell–Boltzmann statistics can be an excellent approximation to quantum statistics — both bosonic and fermionic. We will wrap up this unit by demonstrating this result and clarifying the conditions that correspond to this ‘classical limit’ of quantum statistics.

It is useful to start by asking when we should expect classical statistics to predict a non-negligible probability for multiple particles to occupy the same energy level, leading to the over-counting problems that are solved by quantum statistics. This is actually a question we have already considered, back in Section 3.4. There we used the canonical ensemble to analyse classical spin systems with discrete energy levels, finding that at low temperatures the systems are dominated by their lowest-energy micro-states, with only exponentially suppressed corrections coming from any higher-energy micro-states. In the present context, this corresponds to a classical prediction of exponentially small probabilities for particles to occupy any energy levels with $E_{\ell} > E_0$ — effectively guaranteeing that the lowest energy level \mathcal{E}_0 will be occupied by multiple particles and classical statistics will break down.

In short, the low-temperature regime is where quantum and classical statistics disagree, while **high temperatures correspond to the classical limit** of quantum statistics. If you are not convinced by the argument leading to this conclusion, you can find a more detailed derivation based on the equation of state and thermal de Broglie wavelength in Section 3.5 of David Tong’s [Lectures on Statistical Physics](#).

For now, we want to consider the grand-canonical ensemble at high temperatures, to see whether the quantum and classical statistics we derived in the previous sections become equivalent in this regime. However, it can be subtle to work with the

grand-canonical ensemble at high temperatures, due to the dependence of the average number of particles on the temperature. To demonstrate this subtlety, let's compute the average particle number $\langle N \rangle(T, \mu)$ starting from the grand-canonical partition function, for both classical and quantum statistics.

For convenience, let's collect our earlier results for the grand-canonical partition functions corresponding to classical Maxwell–Boltzmann statistics (Eq. 87), the quantum Bose–Einstein statistics of bosons (Eq. 86) and the quantum Fermi–Dirac statistics of fermions (Eq. 88):

$$Z_g^{\text{MB}} = \prod_{\ell=0}^{\mathcal{L}} \exp [e^{-\beta(E_\ell - \mu)}]$$

$$Z_g^{\text{BE}} = \prod_{\ell=0}^{\mathcal{L}} \frac{1}{1 - e^{-\beta(E_\ell - \mu)}} \qquad Z_g^{\text{FD}} = \prod_{\ell=0}^{\mathcal{L}} [1 + e^{-\beta(E_\ell - \mu)}].$$

Recalling $\log [\prod_i x_i] = \sum_i \log x_i$, the three corresponding grand-canonical potentials $\Phi = -T \log Z_g$ are

$$\Phi_{\text{MB}} = -T \sum_{\ell=0}^{\mathcal{L}} e^{-\beta(E_\ell - \mu)}$$

$$\Phi_{\text{BE}} = T \sum_{\ell=0}^{\mathcal{L}} \log [1 - e^{-\beta(E_\ell - \mu)}] \qquad \Phi_{\text{FD}} = -T \sum_{\ell=0}^{\mathcal{L}} \log [1 + e^{-\beta(E_\ell - \mu)}].$$

We are now ready to compute the average particle numbers. Using the result we derived in Section 6.3,

$$\langle N \rangle = -\frac{\partial \Phi}{\partial \mu},$$

what are the average particle numbers resulting from the three grand-canonical potentials above?

$$\langle N \rangle_{\text{MB}} =$$

$$\langle N \rangle_{\text{BE}} =$$

$$\langle N \rangle_{\text{FD}} =$$

You should find that the average particle number in all three cases can be expressed as a sum over the average occupation numbers,

$$\langle N \rangle = \sum_{\ell=0}^{\mathcal{L}} \langle n_{\ell} \rangle,$$

where the corresponding average occupation numbers are

$$\langle n_{\ell} \rangle_{\text{MB}} = \frac{1}{e^{\beta(E_{\ell}-\mu)}}$$

$$\langle n_{\ell} \rangle_{\text{BE}} = \frac{1}{e^{\beta(E_{\ell}-\mu)} - 1} \qquad \langle n_{\ell} \rangle_{\text{FD}} = \frac{1}{e^{\beta(E_{\ell}-\mu)} + 1}.$$

Note that $\langle n_{\ell} \rangle \geq 0$ in all cases, and also $\langle n_{\ell} \rangle_{\text{FD}} \leq 1$, as required for fermions. From these results we see that the classical limit $\langle n_{\ell} \rangle_{\text{BE}} \approx \langle n_{\ell} \rangle_{\text{FD}} \approx \langle n_{\ell} \rangle_{\text{MB}}$ corresponds to

$$e^{\beta(E_{\ell}-\mu)} \pm 1 \approx e^{\beta(E_{\ell}-\mu)} \implies e^{\beta(E_{\ell}-\mu)} \gg 1.$$

We can also confirm that this limit corresponds to $\langle n_{\ell} \rangle \ll 1$ for all energy levels \mathcal{E}_{ℓ} and all three types of statistics, confirming our starting point of very small probabilities for multiple particles to occupy the same energy level.

Now we can appreciate the subtlety promised above, because

$$\beta(E_{\ell} - \mu) = \frac{E_{\ell} - \mu}{T} \gg 1. \tag{89}$$

does not look like a high-temperature limit! Indeed, if we consider the naive high-temperature limit $\beta = 1/T \rightarrow 0$ with fixed $(E_{\ell} - \mu)$, we would find large average occupation numbers,

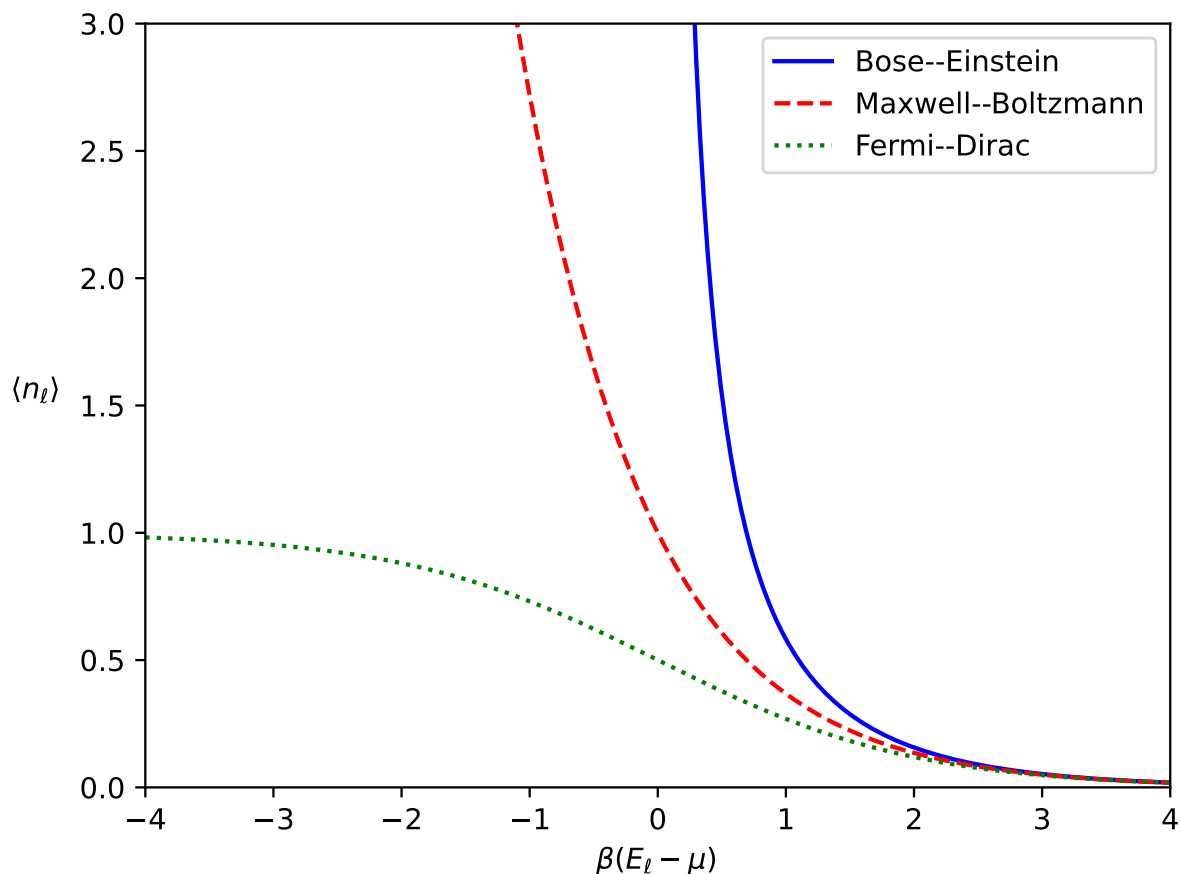
$$\langle n_{\ell} \rangle_{\text{MB}} \approx 1 \qquad \langle n_{\ell} \rangle_{\text{BE}} \rightarrow \infty \qquad \langle n_{\ell} \rangle_{\text{FD}} \approx \frac{1}{2}.$$

In addition to implying non-negligible probabilities for multiple particles to occupy the same energy level, this result indicates that higher temperatures in the grand-canonical ensemble lead to larger particle numbers in total — at least when $(E_{\ell} - \mu)$ is fixed.

In order to balance this effect, we need to adjust the other parameter offered by the grand-canonical ensemble: the chemical potential μ . Specifically, in order to satisfy Eq. 89 in the high-temperature limit, we need $E_{\ell} - \mu \gg T$, which requires that $\mu \rightarrow -\infty$ as $T \rightarrow \infty$. Making the chemical potential more negative reduces the probability of having large numbers of particles in the system, at the same time as the smaller β increases the number of energy levels that these particles can occupy with non-negligible probability. Taken together, these two effects guarantee that there are many more accessible energy levels than there are particles, allowing us to conclude that the true high-temperature limit in which quantum statistics becomes classical is

$$-\mu \gg T \gg E_{\ell} \implies \frac{E_{\ell} - \mu}{T} \gg 1. \tag{90}$$

This corresponds to the right edge of the plot on the next page, where we can confirm that all three sets of statistics predict essentially the same average occupation number $\langle n_{\ell} \rangle$ for any energy level \mathcal{E}_{ℓ} .



In the low-temperature regime $\frac{E_\ell - \mu}{T} \ll 1$ corresponding to the left portion of the plot, we see dramatically different behaviour for the three cases. The classical Maxwell–Boltzmann prediction for the average occupation number grows exponentially, while the quantum Bose–Einstein prediction diverges as either $E_\ell \rightarrow \mu$ or $\beta \rightarrow 0$ with $(E_\ell - \mu)$ fixed, and the Fermi–Dirac prediction slowly approaches its maximum possible value $\langle n_\ell \rangle_{\text{FD}} \rightarrow 1$. In the next unit we will explore these results in more detail by considering specific examples of quantum gases of bosons and fermions.

Unit 8: Quantum gases

8.1 The photon gas

8.1.1 Massive bosons in a box

In Section 7.3 we derived the grand-canonical partition function (Eq. 86) that defines quantum Bose–Einstein statistics for systems of non-interacting bosons,

$$Z_g^{\text{BE}}(\beta, \mu) = \prod_{\ell=0}^{\mathcal{L}} \frac{1}{1 - e^{-\beta(E_\ell - \mu)}}.$$

Following the quantum approach, we obtained this result by considering in turn each energy level \mathcal{E}_ℓ with energy E_ℓ , and summing over all possible occupation numbers that it could have. For bosons, $n_\ell \in \mathbb{N}_0$ produces sums that only converge if $\mu < E_\ell$ for all ℓ . The corresponding grand-canonical potential is

$$\Phi_{\text{BE}} = -T \log Z_g^{\text{BE}} = T \sum_{\ell=0}^{\mathcal{L}} \log [1 - e^{-\beta(E_\ell - \mu)}],$$

from which we can determine the large-scale properties of the system, including its average internal energy $\langle E \rangle$, average particle number $\langle N \rangle$, entropy S , and pressure P .

To do so, we have to specify the energy levels of the particles that compose the system of interest, including any degenerate energy levels $\{\mathcal{E}_m, \mathcal{E}_n\}$ with the same energy $E_m = E_n$ for $m \neq n$. One example of this that we have already considered is the analysis of non-relativistic ideal gas particles in Section 4.1. For a single particle with mass m in a volume $V = L^3$, we adopted an ansatz for the quantized energies,

$$E(k_x, k_y, k_z) = \frac{\hbar^2 \pi^2}{2mL^2} (k_x^2 + k_y^2 + k_z^2) = \varepsilon k^2 \quad \varepsilon \equiv \frac{\hbar^2 \pi^2}{2mL^2}, \quad (91)$$

where the positive integers $k_{x,y,z} = 1, 2, \dots$ specify the possible momentum magnitude of the particle, $p = \hbar \frac{\pi}{L} k$.

Compared to Section 4.1, where $k_{x,y,z} = 0$ was allowed, here we have adjusted our ansatz to require strictly positive values. This adjustment is required by another feature of quantum mechanics, which this paragraph will imprecisely describe for the curious. This description can be skipped without any problem, with the adjusted ansatz simply taken as input. The feature at play here is known as [Heisenberg's uncertainty principle](#) (named after [Werner Heisenberg](#)), which relates the precision with which the position and momentum of each particle can simultaneously be *defined*:

$$(\Delta x) (\Delta p_x) \gtrsim \hbar$$

and similarly for y and z . The ' \gtrsim ' sign here hints that we're ignoring irrelevant factors of 2 and π , while ' Δ ' refers to the precision (or uncertainty) with which x and p_x are defined. Since the particle is within a volume $V = L^3$, we know $\Delta x \lesssim L$. Therefore the uncertainty principle requires $\Delta p_x \gtrsim \hbar/L$, which is only possible if p_x is non-zero, corresponding to $k_x \geq 1$. Note that smaller lengths L imply larger momenta and therefore larger energies.

With this adjusted ansatz, $k_{x,y,z} \geq 1$, we can adapt an exercise from Section 4.1 and ask: What are the lowest energies and the degeneracies of the corresponding energy levels?

8.1.2 Massless photons

Now we will adapt these considerations to analyse a gas of *photons*, massless bosonic quantum particles of light. For our purposes, with no prior knowledge of particle physics, we can define photons simply by specifying their energy levels. Clearly $E \propto 1/m$ from Eq. 91 is problematic for massless particles with $m = 0$. Instead, a photon's energy is proportional to the magnitude of its momentum,

$$E_{\text{ph}}(p) = cp = c\sqrt{p_x^2 + p_y^2 + p_z^2}.$$

Here the speed of light c is really just a unit conversion factor (like the Boltzmann constant) that we could set to $c = 1$ by working in appropriate units.

This relation is connected to the non-relativistic energy $E = \frac{p^2}{2m}$ that we considered in Section 4.1 through the general expression

$$E^2 = (mc^2)^2 + (pc)^2,$$

which is sometimes called Einstein's triangle. When $m = 0$, or $m \ll p/c$ more generally, this reproduces the *ultra-relativistic* relation above. For stationary particles with $p = 0$ it reduces to the famous 'mass-energy' $E = mc^2$, while the non-relativistic kinetic energy is recovered for $m \gg p/c$:

For photons in a volume $V = L^3$, we again have quantized energies,

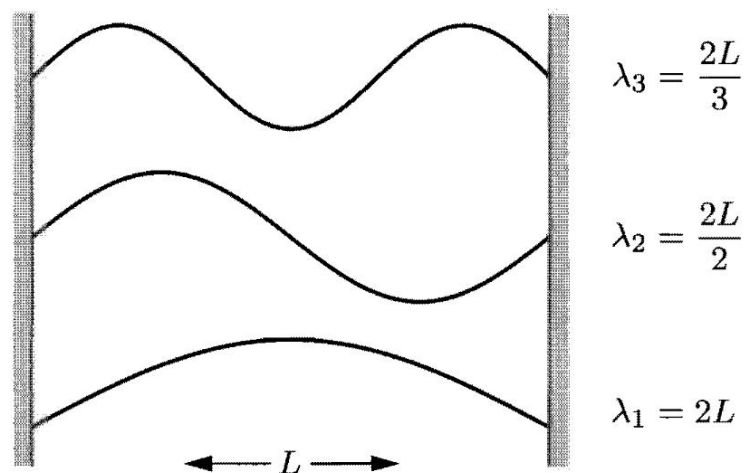
$$E_{\text{ph}}(k) = \hbar c \frac{\pi}{L} k = \hbar c \frac{\pi}{L} \sqrt{k_x^2 + k_y^2 + k_z^2} \quad k_{x,y,z} = 1, 2, \dots, \quad (92)$$

with the same $p = \hbar \frac{\pi}{L} k$ as in Eq. 91. Another feature of photons is that each ‘wavenumber’ (k_x, k_y, k_z) corresponds to two degenerate energy levels with the same $E_{\text{ph}}(k)$. This arises from photons’ connection to electric and magnetic fields, which allows each photon to be polarized in two different ways. If there is interest we can discuss this further in a tutorial, but it is not relevant to the statistical mechanics of photons, for which we can take this double degeneracy as input. Note that this factor of two multiplies all other degeneracies, for instance from permutations of (k_x, k_y, k_z) .

Because light is an electromagnetic wave, it is convenient to work in terms of photons’ wavelength λ and angular frequency $\omega = 2\pi f$ (not to be confused with generic micro-states ω_i). Together, the wavelength and frequency determine the speed of the wave’s propagation, in this case the speed of light

$$c = \frac{\lambda \omega}{2\pi}.$$

In quantum physics, a particle’s momentum is related to its de Broglie wavelength, implying that in a volume $V = L^3$ the wavelengths are also quantized as illustrated in the picture below (from Schroeder’s *Introduction to Thermal Physics*).



Specifically, the length L must be an integer multiple of half a wavelength,

$$L = k \frac{\lambda}{2} \quad \implies \quad c = \frac{L \omega}{k \pi} = \frac{\omega}{\frac{\pi}{L} k},$$

and we can rewrite Eq. 92 as

$$E_{\text{ph}}(\omega) = \hbar \omega. \tag{93}$$

Since $\lambda \propto c/\omega$, we can observe the relation between length and energy scales mentioned above, in the discussion of Heisenberg’s uncertainty principle. Low (*infrared*) frequencies correspond to small energies and long wavelengths, while high (*ultraviolet*) frequencies correspond to large energies and short wavelengths.

We are now ready to write down the grand-canonical potential for a photon gas:

$$\Phi_{\text{ph}} = T \sum_{\ell=0}^{\mathcal{L}} \log [1 - e^{-\beta(E_{\ell}-\mu)}] = 2T \sum_{\vec{k}} \log [1 - e^{-\beta(E_{\text{ph}}(k)-\mu)}],$$

where the factor of 2 in the final expression accounts for the doubly degenerate energy levels. We can simplify this expression by appreciating that photons are easy to create and absorb. For example, every time a light is turned on, it starts emitting a constant flood of photons (with wavelengths of several hundred nanometres). Similarly, food in a microwave oven gets hot by absorbing many lower-energy photons (with longer wavelengths around 12 centimetres). In general an enormous number of photons is required to make even a small change in energy, so that Eq. 81 implies the chemical potential of a photon gas is negligible,¹⁵

$$\mu = \left. \frac{\partial E}{\partial N} \right|_S \approx 0 \quad \implies \quad \Phi_{\text{ph}} \approx 2T \sum_{\vec{k}} \log [1 - e^{-\beta E_{\text{ph}}(k)}].$$

Since we have $k_{x,y,z} \geq 1$, the strictly positive energies $E_{\text{ph}}(k) \propto k/L > 0$ ensure Bose–Einstein statistics is still convergent even with $\mu = 0$.

Another simplification comes from considering the photon gas in a volume that is large compared to the typical wavelengths of the photons. Then the energies $E_{\text{ph}}(k) \propto k/L$ are very closely spaced, and we can approximate the sums over integer $k_{x,y,z}$ as integrals over continuous real $\hat{k}_{x,y,z}$,¹⁶

$$\Phi_{\text{ph}} \approx 2T \int \log [1 - e^{-\beta E_{\text{ph}}(\hat{k})}] d\hat{k}_x d\hat{k}_y d\hat{k}_z.$$

Since the energy $E_{\text{ph}}(\hat{k})$ depends only on the magnitude \hat{k} , we can profit from converting to spherical coordinates. When we do so, we have to keep in mind that $k_{x,y,z} > 0$ corresponds to the positive octant of the sphere,

$$\int_0^\infty d\hat{k}_x \int_0^\infty d\hat{k}_y \int_0^\infty d\hat{k}_z = \int_0^\infty \hat{k}^2 d\hat{k} \int_0^{\pi/2} \sin \theta d\theta \int_0^{\pi/2} d\phi = \frac{\pi}{2} \int_0^\infty \hat{k}^2 d\hat{k},$$

so that

$$\Phi_{\text{ph}} \approx \pi T \int_0^\infty \hat{k}^2 \log [1 - e^{-\beta E_{\text{ph}}(\hat{k})}] d\hat{k}.$$

We can finally change variables to integrate over the angular frequency $\omega = c\frac{\pi}{L}k$, with $E_{\text{ph}} = \hbar\omega$, to find

$$\Phi_{\text{ph}} \approx \pi T \left(\frac{L}{c\pi} \right)^3 \int_0^\infty \omega^2 \log [1 - e^{-\beta\hbar\omega}] d\omega = \frac{VT}{\pi^2 c^3} \int_0^\infty \omega^2 \log [1 - e^{-\beta\hbar\omega}] d\omega, \quad (94)$$

recognizing $L^3 = V$. With this grand-canonical potential derived, we just need to take the appropriate derivatives to determine the thermodynamics and equation of state for the photon gas.

¹⁵If you are not convinced by the argument leading to this conclusion, you can find alternative analyses in Section VI.C of “[The elusive chemical potential](#)” by Ralph Baierlein.

¹⁶We mentioned this approximation in a footnote in Section 4.1, and you can find further discussion in Section 6.7 of Dan Schroeder’s *Introduction to Thermal Physics*.

8.2 The sun and the void

It will be very interesting to use Φ_{ph} in Eq. 94 to analyse the average internal energy for a photon gas. With $\mu = 0$, Eq. 78 from Section 6.3 becomes

$$\langle E \rangle_{\text{ph}} = -T^2 \frac{\partial}{\partial T} \left[\frac{\Phi_{\text{ph}}}{T} \right] = \frac{\partial}{\partial \beta} [\beta \Phi_{\text{ph}}].$$

To begin, let's consider the energy density, as an integral over photon frequencies:

$$\frac{\langle E \rangle_{\text{ph}}}{V} = \int_0^\infty P(\omega) d\omega,$$

where the function $P(\omega)$ is known as the *spectral density*, or simply the *spectrum*. (It's not the pressure!) What is the spectrum for a photon gas?

You should find

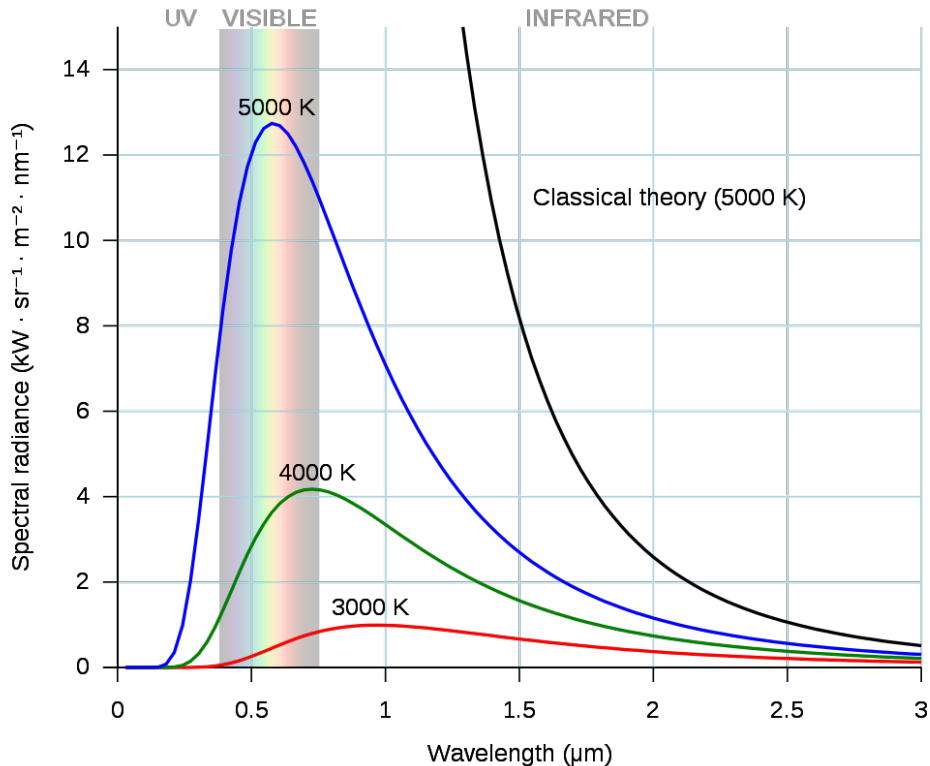
$$P(\omega) = \left(\frac{\hbar}{\pi^2 c^3} \right) \frac{\omega^3}{e^{\beta \hbar \omega} - 1}, \quad (95)$$

which is known as the Planck spectrum, named after Max Planck. We can equally well consider the Planck spectrum $P(\lambda)$ as a function of the wavelength $\lambda = 2\pi c/\omega$, by changing variables in the expression above:

You should find

$$P(\lambda) = \left(\frac{16\pi^2\hbar c}{\lambda^5} \right) \frac{1}{e^{2\pi\beta\hbar c/\lambda} - 1}, \quad (96)$$

which is plotted for three temperatures $T = 1/\beta$ in the figure below, which comes from [Wikimedia Commons](#). (The plot divides our $P(\lambda)$ by 4π steradian and multiplies it by c to convert from a spectral density to a spectral power per unit area per unit of solid angle. For our purposes only the functional form is significant.)



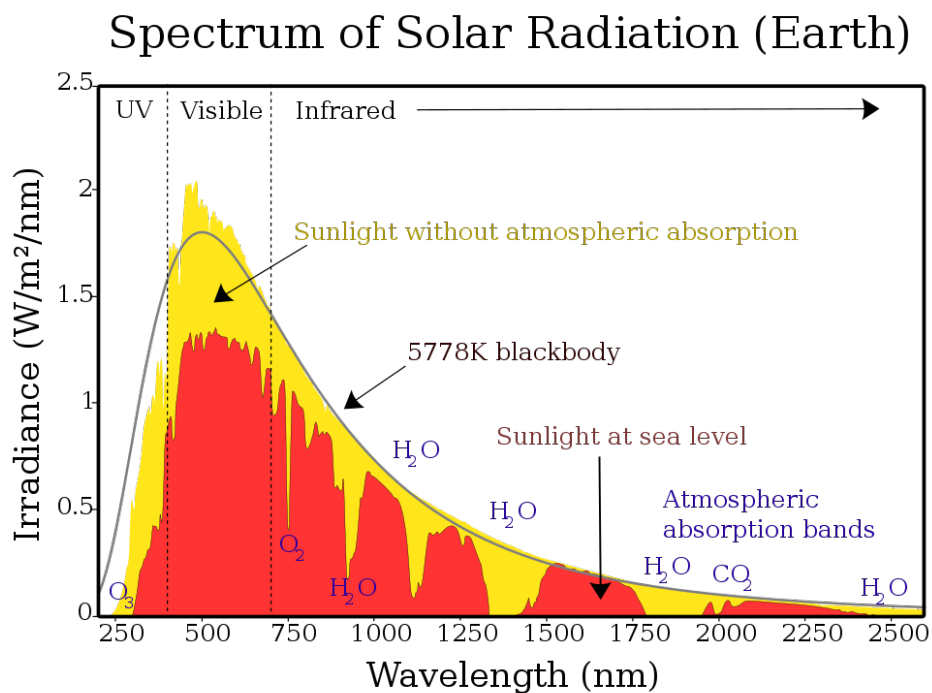
Considering the high-energy ultraviolet (UV) limit of short wavelengths λ , we can see from Eq. 96 that $P(\lambda)$ is exponentially suppressed, which overwhelms the diverging factor $\propto 1/\lambda^5$ in parentheses. In the low-energy infrared (IR) limit, the large λ has the same effect that a large temperature ($\beta \ll 1$) would have: $e^{2\pi\beta\hbar c/\lambda} - 1 \approx 2\pi\beta\hbar c/\lambda$ and

$$P(\lambda) \approx \left(\frac{16\pi^2\hbar c}{\lambda^5} \right) \frac{\lambda}{2\pi\beta\hbar c} = \frac{8\pi T}{\lambda^4}.$$

The connection to large temperatures indicates that this is what classical statistical mechanics would predict for the spectrum of light. It is known as the Rayleigh–Jeans spectrum, named after [the third Baron Rayleigh](#) and [James Jeans](#). Recall that the classical approach sums over all possible energies for each degree of freedom, corresponding to a light-emitting object (historically known as a *black body*) emitting light of all wavelengths λ . According to the classical Rayleigh–Jeans spectrum, in the limit $\lambda \rightarrow 0$ this light would carry an infinite amount of energy, a problem that became known as the *ultraviolet catastrophe*. Planck described his 1900 derivation of the UV-suppressed $P(\lambda)$ as “an act of desperation” to avoid this problem; it turned out to be one of the first steps towards quantum physics.

Another noteworthy feature of the Planck spectrum shown above is that as the temperature increases, the maximum of $P(\lambda)$ moves to shorter wavelengths and correspondingly larger energies. It is not a coincidence that the peak of the spectrum for $T \approx 5000$ K falls within the wavelengths of visible light, roughly 400–700 nm. As shown in the figure below, also from [Wikimedia Commons](#), the amount of sunlight that reaches the surface of the earth is also largest for visible wavelengths, which are visible to us because we have evolved to use this sunlight.

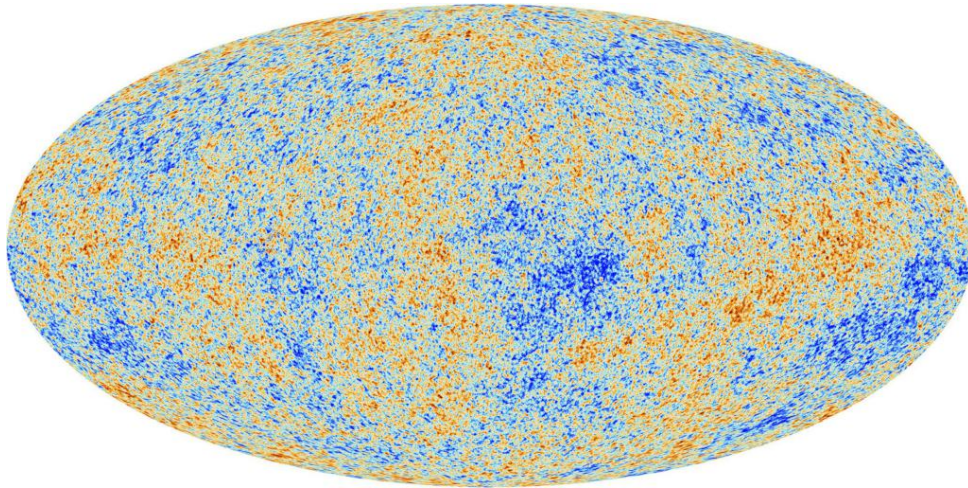
Taking into account the absorption of some sunlight by molecules in the atmosphere, we can see from the figure below that the energy spectrum of the sunlight reaching the top of the atmosphere is quite close to a Planck (or ‘blackbody’) spectrum with temperature $T \approx 5778$ K. The agreement isn’t perfect, which is to be expected since the Planck spectrum relies on the non-trivial assumption of an ideal gas of non-interacting particles. Despite that caveat, numerically fitting the measured sunlight to the Planck spectrum is how this ‘effective’ surface temperature of the sun is determined. This same fitting procedure can even be done for distant stars, with red stars corresponding to relatively low temperatures $T \lesssim 3500$ K and blue stars corresponding to relatively high temperatures $T \gtrsim 10\,000$ K.



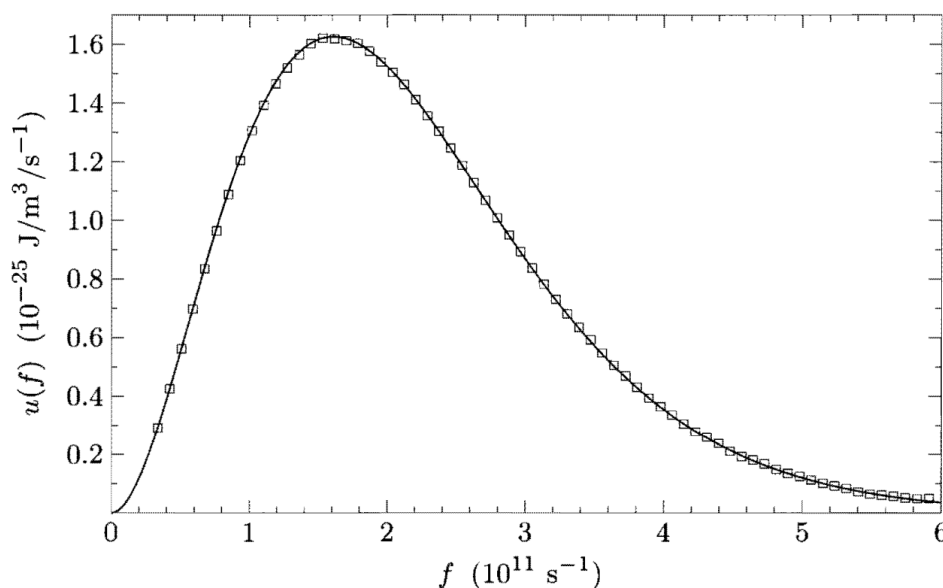
Even more remarkably, we can use the Planck spectrum to determine the temperature of intergalactic space. Rather than being empty, these voids are actually permeated by a very low-temperature photon gas left over from the Big Bang roughly 14 billion years ago. This photon gas is known as the [cosmic microwave background](#) (CMB), and carries information about the early evolution of the universe, including some of the strongest evidence for the existence of dark matter.

The picture below is a famous visualization of the CMB, provided by the [European Space Agency](#) and produced from measurements taken by their ‘Planck’ satellite. For each point in the sky the satellite measures the photon spectrum reaching it from that

direction. The contributions from stars and galaxies are subtracted, and the remaining data are fit to the Planck spectrum to find the temperature of the intergalactic CMB photon gas at that point. From point to point, there are only small temperature fluctuations around the average $T_{\text{CMB}} \approx 2.725$ K. That average temperature is subtracted and the fluctuations themselves are shown below, with warmer red-coloured regions only $\Delta T \approx 0.0002$ K hotter than the cooler blue-coloured regions.



The final figure below illustrates such a fit of CMB data to the Planck spectrum, using measurements taken by the Cosmic Background Explorer (COBE) satellite and [published in 1990](#). (This version of the figure is adapted from that publication, and copied from Schroeder's *Introduction to Thermal Physics*.) The squares are the measured data, and their size represents a cautious estimate of uncertainties. They are plotted with the frequency $f = \omega/(2\pi)$ on the horizontal axis, where $f \approx 3 \times 10^{11} \text{ s}^{-1}$ corresponds to a low-energy wavelength $\lambda = c/f \approx 1 \text{ mm}$, roughly 1000 times longer than the wavelengths of visible light. The solid line is a fit to the data, which produces $T_{\text{CMB}} = 2.735 \pm 0.060$ K. While more recent satellites have increased the precision with which we know T_{CMB} , this first result was awarded the 2006 Nobel Prize in Physics.



Even though we derived the Planck spectrum by assuming an ideal gas of non-interacting photons, we see that it provides an excellent mathematical model for real physical systems, stretching from the hottest to the coldest places in the universe.

8.3 Equation of state for the photon gas

Having looked in some detail at the integrand for the photon gas energy density, Eq. 95, let's complete the integration, which is related to the Riemann zeta function:

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = \Gamma(4)\zeta(4) = \frac{\pi^4}{15}.$$

Using this result, what is the average energy density for an ideal photon gas?

You should find a result proportional to T^4 , which appears significantly more complicated than Eq. 52 for the energy of an N -particle non-relativistic ideal gas in the canonical ensemble. This is related to the fluctuating particle number now that we are working in the grand-canonical ensemble. It's possible to simplify the current situation by computing the average photon number from Eq. 76,

$$\langle N \rangle_{\text{ph}} = - \left. \frac{\partial}{\partial \mu} \Phi_{\text{ph}} \right|_{\mu=0} = - \frac{VT}{\pi^2 c^3} \int_0^\infty \omega^2 \frac{\partial}{\partial \mu} \log [1 - e^{-\beta \hbar \omega} e^{\beta \mu}] d\omega \Big|_{\mu=0},$$

recalling $\mu = 0$ for photon gases. The calculation is quite similar to that for the average internal energy density, now involving the integral

$$\int_0^\infty \frac{x^2}{e^x - 1} dx = \Gamma(3)\zeta(3) = 2\zeta(3).$$

Using this result, what is the average particle number density for an ideal photon gas?

You should find a result proportional to $VT^3 \propto \langle E \rangle_{\text{ph}}/T$, so that

$$\langle E \rangle_{\text{ph}} = \frac{\pi^2}{15\hbar^3 c^3} VT^4 = \frac{\pi^4}{30\zeta(3)} \langle N \rangle_{\text{ph}} T. \quad (97)$$

The functional form is the same as Eq. 52, with a larger numerical factor

$$\frac{\pi^4}{30\zeta(3)} = \frac{\Gamma(4)\zeta(4)}{\Gamma(3)\zeta(3)} \approx 2.7$$

compared to $\frac{3}{2}$ for the classical non-relativistic case.

To get the rest of the way to the equation of state for the photon gas, we need to compute the *radiation pressure*

$$P_{\text{ph}} = - \left. \frac{\partial}{\partial V} \langle E \rangle_{\text{ph}} \right|_{S_{\text{ph}}},$$

which requires first figuring out the condition of constant entropy S_{ph} for a photon gas. From Eq. 79 with $\mu = 0$, we have

$$S_{\text{ph}} = \frac{\langle E \rangle_{\text{ph}} - \Phi_{\text{ph}}}{T}.$$

Looking back to Eq. 94 for the grand-canonical potential, we see

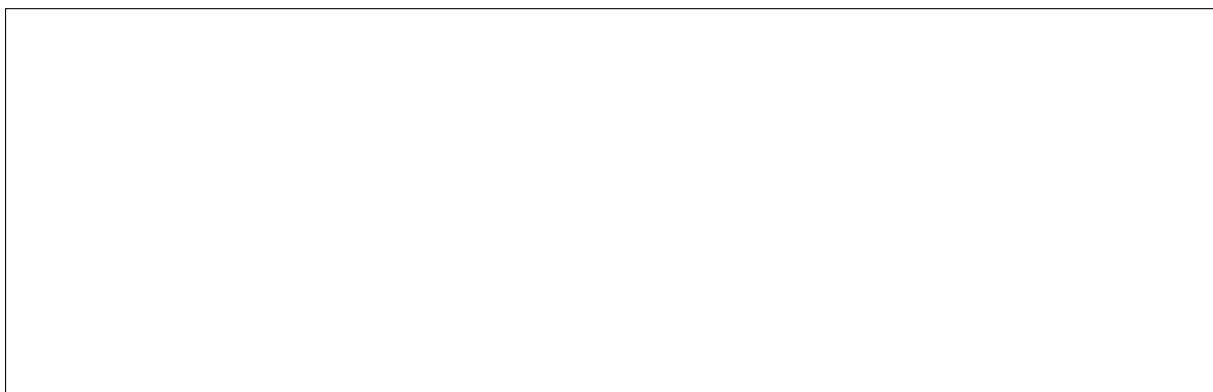
$$\frac{\Phi_{\text{ph}}}{T} = \frac{V}{\pi^2 c^3} \int_0^\infty \omega^2 \log [1 - e^{-\beta\hbar\omega}] d\omega = \frac{VT^3}{\pi^2 \hbar^3 c^3} \int_0^\infty x^2 \log [1 - e^{-x}] dx,$$

changing variables to $x = \beta\hbar\omega = \hbar\omega/T$. The final factor in this expression is yet another delightful integral,

$$\int_0^\infty x^2 \log [1 - e^{-x}] dx = -2\zeta(4) = -\frac{\pi^4}{45}.$$

Since this gives us $S \propto VT^3$, we can conclude that the condition of constant entropy for a photon gas is $VT^3 = \text{constant}$, in contrast to the $VT^{3/2}$ dependence of Eq. 53 for classical non-relativistic particles.

At this point it is straightforward to take the derivative of the average internal energy if we express the constant-entropy condition as $T = bV^{-1/3}$, with b a constant:



For the resulting equation of state for the photon gas, you should find

$$P_{\text{ph}}V = \frac{1}{3} \langle E \rangle_{\text{ph}} = \frac{\pi^4}{90\zeta(3)} \langle N \rangle_{\text{ph}} T. \quad (98)$$

The functional form is the same as the (classical, non-relativistic) ideal gas law, with just an additional numerical factor of

$$\frac{\pi^4}{90\zeta(3)} = \frac{\zeta(4)}{\zeta(3)} \approx 0.9004.$$

8.4 Non-relativistic ideal fermion gas

For the remainder of this unit we will apply the grand-canonical ensemble to investigate ideal gases of non-interacting fermions. We again take the approach of quantum statistics, defining micro-states by summing over the possible occupation numbers n_ℓ for each energy level \mathcal{E}_ℓ with (possibly not unique) energy E_ℓ . In contrast to the bosonic case considered above, the only possible occupation numbers are now $n_\ell = 0$ and 1, since the Pauli exclusion principle prevents multiple identical fermions from occupying the same energy level.

In Section 7.4 we derived the grand-canonical partition function (Eq. 88) that defines quantum Fermi–Dirac statistics for systems of non-interacting fermions,

$$Z_g^{\text{FD}}(\beta, \mu) = \prod_{\ell=0}^{\mathcal{L}} [1 + e^{-\beta(E_\ell - \mu)}],$$

in terms of the inverse temperature $\beta = 1/T$ and chemical potential μ . Recall that it is possible for systems of fermions to have any value for the chemical potential, either positive or negative, in contrast to the systems of bosons we considered above. From the corresponding grand-canonical potential,

$$\Phi_{\text{FD}} = -T \log Z_g^{\text{FD}} = -T \sum_{\ell=0}^{\mathcal{L}} \log [1 + e^{-\beta(E_\ell - \mu)}]$$

we can determine the large-scale properties of the system, including its average internal energy $\langle E \rangle$, average particle number $\langle N \rangle$, entropy S , and pressure P , along with the equation of state relating these quantities.

Concrete calculations require specifying the energy levels of the system, including the degeneracies of any distinct energy levels $\{\mathcal{E}_m, \mathcal{E}_n\}$, $m \neq n$, with the same energy $E_m = E_n$. In this section we'll consider non-relativistic particles, expanding on our review of such systems in Section 8.1. In a volume $V = L^3$, the energy levels are defined by the non-zero quantized energies

$$E(k) = \varepsilon (k_x^2 + k_y^2 + k_z^2) \quad \varepsilon \equiv \frac{\hbar^2 \pi^2}{2mL^2} \quad k_{x,y,z} = 1, 2, \dots$$

In addition to the usual degeneracies coming from permutations of (k_x, k_y, k_z) that we have already analysed, for each distinct \vec{k} typical fermions such as electrons have two

degenerate energy levels with the same energy $E(k)$. This arises from a quantum property called *spin*, rather than the two polarizations for photons discussed in Section 8.1: ‘spin-up’ and ‘spin-down’ electrons with everything else the same (including their energies) occupy distinct, degenerate energy levels. This property of spin is related to the spin–statistics theorem mentioned in Section 7.2, and is another topic we can discuss further in a tutorial if there is interest. For our statistical mechanics purposes it will suffice simply to incorporate this information into our ansatz as input.

Thus the grand-canonical potential for an ideal gas of non-relativistic fermions is

$$\Phi_f = -T \sum_{\ell=0}^{\mathcal{L}} \log [1 + e^{-\beta(E_\ell - \mu)}] = -2T \sum_{\vec{k}} \log \left[1 + \exp \left(-\frac{\hbar^2 \pi^2 k^2}{2mL^2 T} + \frac{\mu}{T} \right) \right].$$

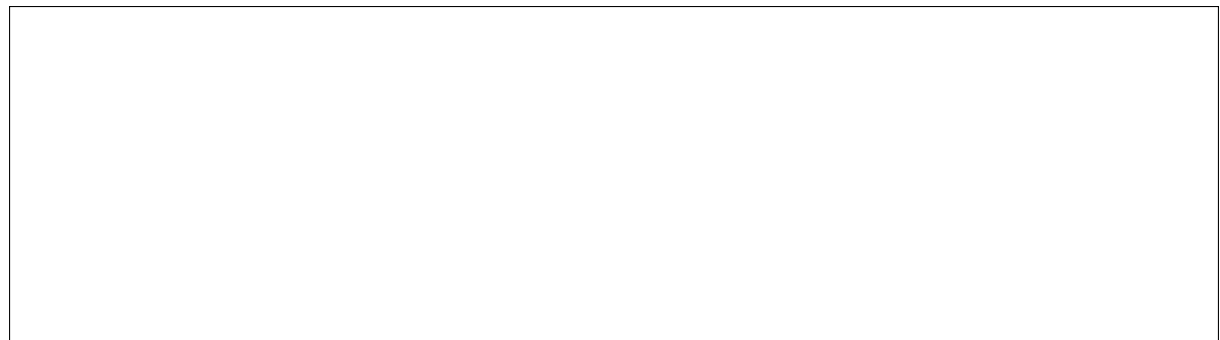
We can again proceed by considering the gas in a large volume and approximating the sum over discrete integer $k_{x,y,z}$ by integrals over continuous real $\hat{k}_{x,y,z}$:

$$\Phi_f \approx -2T \int \log \left[1 + \exp \left(-\frac{\hbar^2 \pi^2 \hat{k}^2}{2mL^2 T} + \frac{\mu}{T} \right) \right] d\hat{k}_x d\hat{k}_y d\hat{k}_z.$$

Converting to spherical coordinates and carrying out the angular integrations over the $\frac{\pi}{2}$ solid angle of the octant of the sphere with $k_{x,y,z} > 0$, we have

$$\Phi_f \approx -\pi T \int_0^\infty \hat{k}^2 \log \left[1 + \exp \left(-\frac{\hbar^2 \pi^2 \hat{k}^2}{2mL^2 T} + \frac{\mu}{T} \right) \right] d\hat{k}.$$

In the same spirit as the change of variables we carried out to integrate over photon frequencies ω , let’s now work in terms of the fermion energy $E = \frac{\hbar^2 \pi^2}{2mL^2} \hat{k}^2$:



As for the case of a photon gas, Eq. 94, you should find $\Phi_f \propto VT$. It will be convenient to keep this grand-canonical potential in the form of an integral over the energy E , which we will evaluate after taking appropriate derivatives to determine the thermodynamics and equation of state for non-relativistic fermions.

8.5 Low-temperature equation of state

In contrast to the photon gas, we need to retain the chemical potential in our analyses of non-relativistic fermions, which makes these calculations more complicated. To

achieve a different simplification, we can focus on the low-temperature regime where we expect quantum Fermi–Dirac statistics to differ significantly from the classical case we considered back in Section 4.1. As we saw in Section 7.5, it is only at high temperatures, with large negative chemical potential, that the classical approach provides a good approximation to the true quantum physics.

To see how low temperatures simplify the analysis of the non-relativistic fermion gas, it will prove profitable to first consider the average particle number

$$\langle N \rangle_f = -\frac{\partial}{\partial \mu} \Phi_f,$$

using the grand-canonical potential we computed above. In analogy to the Planck spectrum we derived for the photon gas in Section 8.2, we first express the average particle number density as an integral over energies,

$$\frac{\langle N \rangle_f}{V} = \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \int_0^\infty F(E) \sqrt{E} dE, \quad (99)$$

where the function $F(E)$ is known as the Fermi function. In contrast to the Planck spectrum, some constant factors are kept separate from $F(E)$, so that it more closely resembles the average occupation numbers $\langle n_\ell \rangle$ we computed in Section 7.5:



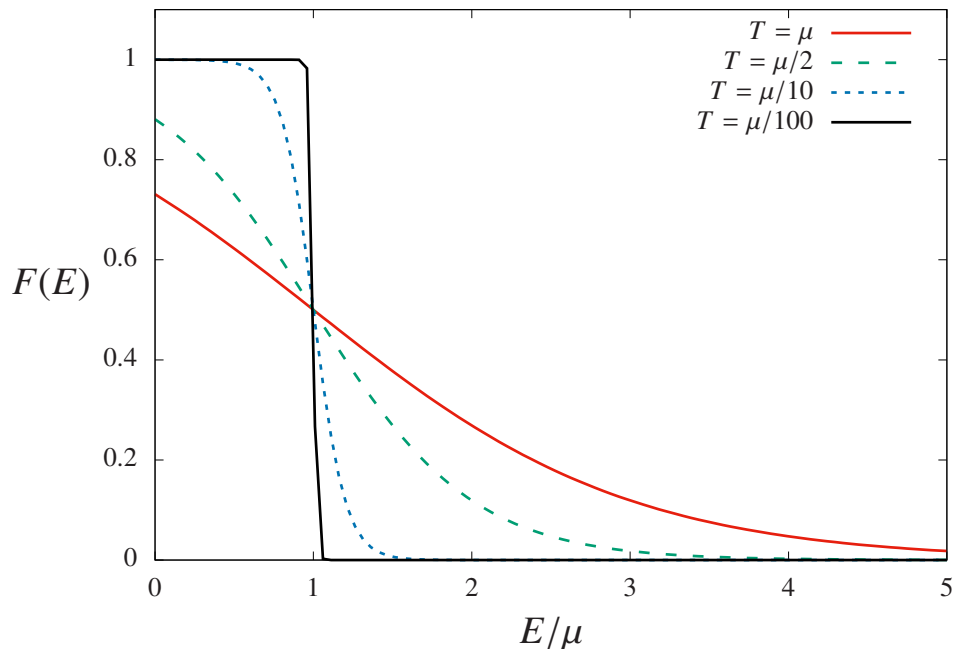
As usual in the grand-canonical approach, the average particle number density and Fermi function depend on the inverse temperature β and the chemical potential μ . Expressing $F(E)$ in terms of the dimensionless ratios E/μ and T/μ ,

$$F(E) = \frac{1}{\exp\left[\frac{E-\mu}{T}\right] + 1} = \frac{1}{\exp\left[\frac{\mu}{T}\left(\frac{E}{\mu} - 1\right)\right] + 1} = \frac{1}{\left(\exp\left[\frac{E}{\mu} - 1\right]\right)^{\mu/T} + 1},$$

we can highlight the two main features of the figure on the next page, which plots the Fermi function against E/μ for various temperatures T/μ . Here we assume a positive chemical potential, $\mu > 0$, which we will soon show is required for low-temperature non-relativistic fermion gases.

First, we can see that the point $E = \mu$, where $F(E) = 1/2$ for any temperature, is a threshold at which the behaviour of the Fermi function changes. For larger energies $E > \mu$, the factor $\exp\left[\frac{E}{\mu} - 1\right] > 1$ and drives $F(E) \rightarrow 0$ as the energy increases. For smaller energies $E < \mu$, the factor $\exp\left[\frac{E}{\mu} - 1\right] < 1$ and becomes negligible if raised to

a sufficiently large power $\frac{\mu}{T}$, leaving $F(E) \rightarrow 1$. These two asymptotic limits reflect the possible energy level occupation numbers for fermions, $n_\ell = 0$ and 1. Second, smaller temperatures cause much more rapid approach to these two limits, with the exponential factor either enhanced (if $E > \mu$) or suppressed (if $E < \mu$) by a power $\mu/T \gg 1$.



Therefore, for small temperatures $T \ll \mu$, we can simplify our calculations by approximating the Fermi function as a step function,

$$F(E) \approx \begin{cases} 1 & \text{for } 0 \leq E < \mu \\ 0 & \text{otherwise} \end{cases} . \quad (100)$$

Using this approximation, what is the resulting average particle number density?

You should find a result proportional to $\mu^{3/2}$ but independent of T . The temperature independence turns out to be the leading-order behaviour of a more general result that can be organized in powers of $T/\mu \ll 1$, through a method known as the *Sommerfeld expansion* (named after [Arnold Sommerfeld](#)). The $\mu^{3/2}$ dependence on the chemical potential is something we could have predicted before doing the explicit calculation. This is because the step function in Eq. 100 corresponds to a single fermion occupying each and every energy level with $E_\ell < \mu$, while all energy levels with $E_\ell > \mu$

are unoccupied. Since $E(k) \propto k^2$, summing over all $k_{x,y,z}$ for which $E(k) < \mu$ corresponds to computing (a portion of) the volume of a sphere of radius $r_k = \sqrt{\mu}$. This volume is proportional to $r_k^3 = \mu^{3/2}$, in agreement with our result above. Inverting that result defines the **Fermi energy**, the maximum fermion energy at zero temperature:

$$E_F = \mu = \frac{\hbar^2}{2m} \left(\frac{3\pi^2 \langle N \rangle_f}{V} \right)^{2/3}. \quad (101)$$

Like the step-function approximation to the Fermi function, this equality between the Fermi energy and the chemical potential is only exact at zero temperature, with $0 < T \ll E_F$ introducing small corrections discussed in Section 8.8.

Let's compute the average energy density of the non-relativistic fermion gas at low temperatures. Rather than taking another derivative of the grand-canonical potential, we can note from Eq. 85 and from our work on the photon gas in Section 8.3 that

$$\frac{\langle E \rangle_f}{V} = \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \int_0^\infty E F(E) \sqrt{E} dE. \quad (102)$$

That is, instead of simply counting the number of fermions in the system, we need to add up their energies, introducing an extra factor of E compared to Eq. 99. Still using the low-temperature step-function approximation for the Fermi function in Eq. 100, what is the average energy density?

You should find

$$\langle E \rangle_f = \frac{3}{5} \mu \langle N \rangle_f, \quad (103)$$

which means that the average energy of each fermion in the low-temperature ideal gas, $\langle E \rangle_f / \langle N \rangle_f$, is three-fifths of the Fermi energy $E_F = \mu$.

In particular, we find that ideal gases of non-relativistic fermions have positive internal energy even as the temperature approaches absolute zero, $T \rightarrow 0$.

This can be understood by recalling that the lowest-energy pair of degenerate energy levels can each hold only a single fermion, forcing any additional fermions to 'fill' energy levels with larger energies $E_\ell > 0$, up to the Fermi energy. It is a stark contrast to the $\langle E \rangle = \frac{3}{2} NT$ we found for classical ideal gases in the canonical ensemble in Eq. 52, as well as the $\langle E \rangle_{\text{ph}} \approx 2.7 \langle N \rangle_{\text{ph}} T \propto T^4$ we more recently computed in

Eq. 97 for a grand-canonical quantum gas of photons. In both of those cases the average energy vanishes in the zero-temperature limit. This is because all the particles in those classical and bosonic systems are able to occupy the lowest energy level at low temperatures, with exponentially small probabilities $\propto e^{-E_\ell/T}$ for particles to occupy any energy levels with $E_\ell > E_0$.

This picture of fermions filling energy levels up to the Fermi energy also clarifies why the chemical potential for a fermion gas must be positive at low temperatures. Recall Eq. 81 for the chemical potential,

$$\mu = \left. \frac{\partial E}{\partial N} \right|_{S,V},$$

which we derived from the generalized thermodynamic identity in Section 6.3. Let's use this to consider what happens when we increase the number of particles in a zero-temperature fermion gas. In this limit $T \rightarrow 0$, there is only the single quantum micro-state described above, with all energy levels filled below the Fermi energy E_F and empty above E_F . Adding particles, $\Delta N > 0$, doesn't increase the number of accessible micro-states, and therefore doesn't increase the entropy $S_f = -\sum_{i=1}^M p_i \log p_i = 0$, satisfying the constant-entropy condition required by this equation. However, it does increase the energy, because the added particles must fill the first available energy levels, just above the Fermi energy. That is, $\Delta E = E_F \Delta N > 0$, and we find $\mu = E_F > E_0 \geq 0$ as claimed earlier in this section. It is an interesting but lengthy exercise (discussed in Section 8.8) to show that the chemical potential becomes negative as the temperature increases and we approach the classical limit.

To get the rest of the way to the low-temperature equation of state for ideal gases of non-relativistic fermions, we need to compute the pressure

$$P_f = - \left. \frac{\partial}{\partial V} \langle E \rangle_f \right|_{N,S_f}.$$

As we just discussed, the single accessible micro-state for $T \rightarrow 0$ automatically satisfies the condition of constant entropy, $S_f = 0$. Applying Eq. 101 that relates the chemical potential to the average particle number density, we have

$$\langle E \rangle_f = \frac{3}{5} \mu \langle N \rangle_f = \frac{3}{5} \left(\frac{\hbar^2}{2m} \right) \left(\frac{3\pi^2}{V} \right)^{2/3} \langle N \rangle_f^{5/3}.$$

This is all we need to determine the pressure, which we can relate to the energy density, the Fermi energy $E_F = \mu$ and the particle number density:



In particular, we can see that the pressure (like the energy) remains positive even as the temperature approaches absolute zero, with

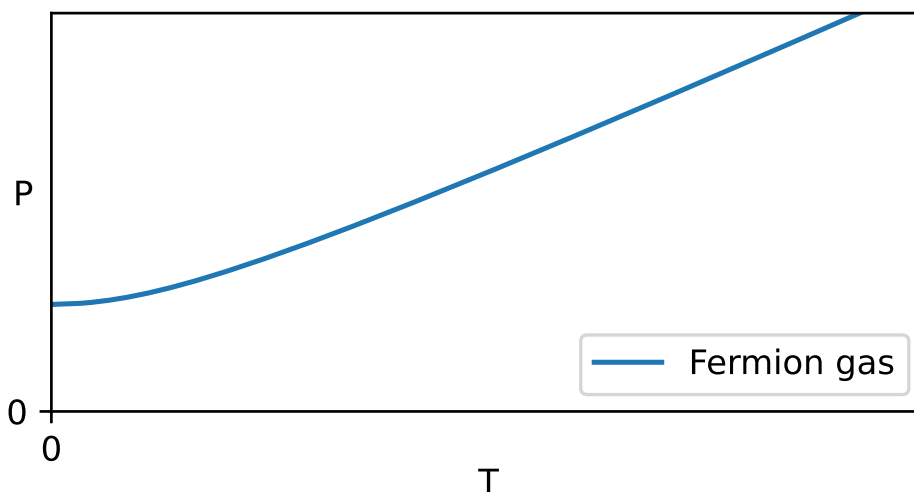
$$P_f = (3\pi^2)^{2/3} \frac{\hbar^2}{5m} \rho_f^{5/3}, \quad (104)$$

where we define the density $\rho_f = \langle N \rangle_f / V$. This positive pressure in the $T \rightarrow 0$ limit is not due to any direct force between the fermions, which remain non-interacting in this ideal gas. It is a purely quantum effect resulting from the Pauli exclusion principle.

As we saw earlier in this section, the temperature independence of the pressure P_f is due to approximating the low-temperature Fermi function as a step function in Eq. 100, and systematic corrections to this approximation can be computed through a Sommerfeld expansion. Even without getting into such detailed calculations, we know that in the high-temperature classical regime the quantum ideal gas of massive fermions will be well approximated by the classical ideal gas we considered in Section 4.4, with equation of state

$$PV = NT \quad \implies \quad P = \frac{N}{V}T = \rho T. \quad (105)$$

In words, at high temperatures the pressure depends linearly on the temperature, with the slope corresponding to the density ρ . The plot below (produced by [this Python code](#)) shows how the pressure changes from a positive constant as $T \rightarrow 0$ to this linear behaviour at higher temperatures.



8.6 Type-Ia supernovas

The positive pressure that remains for a fermion gas even at zero temperature, Eq. 104, is known as the *degeneracy pressure*. (This use of the word ‘degeneracy’ is unrelated to its other use describing multiple energy levels with the same value of the energy.) The degeneracy pressure plays a key role in a certain type of supernova explosions of stars — a famous astrophysical phenomenon.

To begin exploring this, note that the temperature doesn’t need to be exactly zero in order for the degeneracy pressure to be significant. The temperature just needs to

be small compared to the Fermi energy, $T \ll E_F$. From Eq. 101 we see that $E_F \propto \rho_f^{2/3}$ increases for larger densities $\rho_f = \langle N \rangle_f / V$. While the densities of stars can be very large, due to the enormous amount of matter that is squeezed together by gravitational attraction, these extreme conditions can also create very high temperatures.

For reference, everyday solids have densities around 10^{28} – 10^{29} atoms per cubic metre (roughly Avogadro's number per cubic centimetre). In the case of metals, the density of conducting electrons (a gas of fermions that are not bound to particular atoms) is similar, leading to Fermi energies $E_F \sim 10^4$ K, very large compared to everyday temperatures. Perhaps surprisingly, the average density of the sun is not much different, around 10^{30} atoms per cubic metre, corresponding to $E_F \sim 10^5$ K. This is because the sun's much larger gravitational forces are counter-balanced by the radiation pressure coming from the fusion of hydrogen and helium nuclei. That fusion also heats up the core of the sun to $T \sim 10^7$ K $\gg E_F$, making our low-temperature derivations above inapplicable.

Something interesting happens when this hydrogen and helium 'fuel' is exhausted and the radiation pressure decreases precipitously. This causes stars to be gravitationally compressed into much denser and more compact objects. The precise chain of events depends on the mass of the star. Stars with masses comparable to our sun turn into *white dwarfs* with radii comparable to the radius of the earth, roughly a hundred times smaller than that of the sun. In other words, the density of a white dwarf is $\rho \sim (100)^3 \rho_{\text{sun}} \sim 10^{36}$ atoms per cubic meter, equivalent to a mass density around one tonne per cubic centimetre.

The corresponding white dwarf Fermi energy is

$$E_F \sim \left(\frac{\rho}{\rho_{\text{sun}}} \right)^{2/3} E_F^{(\text{sun})} \sim (10^6)^{2/3} 10^5 \sim 10^9 \text{ K.}$$

Even for a young white dwarf that retains its initial core temperature of roughly ten million kelvin, we have $T \sim 10^7$ K $\ll E_F$ and can accurately describe the star using the low-temperature ideal (non-interacting) fermion gas we analysed above. In particular, the degeneracy pressure, Eq. 104, coming from the electrons in the white dwarf is what stabilizes these stars and prevents them from collapsing further into even denser objects such as neutron stars or black holes.

That gravitational collapse could trigger a supernova explosion, but won't happen for a white dwarf in isolation — these stars will happily cool for trillions of years, supported by their degeneracy pressure, until they reach thermal equilibrium with the ~ 2.725 K cosmic microwave background radiation we discussed in Section 8.2. (The coldest known white dwarfs have cooled to temperatures of thousands of kelvin over the past ~ 14 billion years.) Things become more interesting for a white dwarf in a binary system with a companion star. If this companion star is still burning hydrogen or helium through nuclear fusion, it will emit matter that is then captured by the white dwarf, slowly increasing the white dwarf's mass. Such a binary system is pictured on the next page, in an artist's illustration provided by the [European Space Agency](#).



As the white dwarf accumulates the matter emitted by its companion, its mass and its density steadily increase. As the mass of the white dwarf approaches a value roughly 40% larger than the mass of our sun, known as the *Chandrasekhar limit* (named after [Subrahmanyan Chandrasekhar](#)), its density becomes large enough for new types of nuclear fusion reactions to occur. Instead of hydrogen or helium, which the white dwarf has already burned, these new fusion reactions involve carbon and oxygen, which remain present in abundance. In just a few seconds, these fusion reactions run away, increase the temperature of the white dwarf to billions of kelvin, and blast it apart in a supernova explosion about five billion times brighter than the sun.

For obscure historical reasons, these particular stellar explosions are known as *type-Ia* (“one-A”) *supernovas*. The degeneracy pressure of low-temperature fermion gases plays a crucial role in these supernovas, by ensuring that a specific amount of mass has to build up in order to trigger the explosion. The resulting regularity of type-Ia supernovas allows them to be used as “standard candles” providing a consistent measure of astronomical distances. This in turn enabled the discovery that the expansion of the universe is accelerating (a phenomenon popularly called ‘dark energy’), which was awarded the 2011 Nobel Prize in Physics.

8.7 Relativistic ideal fermion gas

Gases of relativistic fermions also play important roles in nature. In fact, by changing units we can appreciate that the white dwarf Fermi energy discussed above, $E_F \sim 10^9 \text{ K} \sim 0.3 \text{ MeV}$ is comparable to the 0.511 MeV mass-energy of electrons, suggesting that relativistic effects may be non-negligible in white dwarfs. Such effects are indeed crucial to the computation of the Chandrasekhar limit mentioned above.

While full calculations for massive relativistic particles are beyond the scope of this module, we can take advantage of our earlier analyses of gases of massless photons to quickly obtain results for similar gases of massless fermions. *Neutrinos* (denoted ‘ ν ’) are physical examples of fermions whose masses are so small that they can be very well approximated as massless. In fact, for many years neutrinos were

thought to be exactly massless — the discovery that neutrinos have non-zero masses was awarded the 2015 Nobel Prize in Physics.

In the same way as photons, massless fermions would travel at the speed of light, c , and have energies determined by their angular frequencies, $E_\nu = cp = \hbar\omega$. In a volume $V = L^3$, these energies are quantized as usual,

$$\omega = \frac{2\pi c}{\lambda} = c \frac{\pi}{L} k,$$

where $k^2 = k_x^2 + k_y^2 + k_z^2$ and $k_{x,y,z} > 0$ are positive integers. Just as for the non-relativistic case considered in Section 8.4, typical massless fermions, including neutrinos, have two degenerate energy levels for each distinct \vec{k} , with opposite spin.

The computations required to analyse a gas of massless fermions are very similar to the work we recently did for photon gases. In particular, massless fermions are also easy to create and absorb, implying a vanishing chemical potential, $\mu \approx 0$. Again approximating sums over discrete integer $k_{x,y,z}$ by integrals over continuous real $\hat{k}_{x,y,z}$, and changing variables to integrate over the angular frequency, we end up with the grand-canonical potential

$$\Phi_\nu = -\frac{VT}{\pi^2 c^3} \int_0^\infty \omega^2 \log [1 + e^{-\beta\hbar\omega}] d\omega. \quad (106)$$

The only changes here compared to Eq. 94 for the photon Φ_{ph} are a couple of negative signs, precisely as we would expect from comparing the Bose–Einstein and Fermi–Dirac grand-canonical potentials in Section 7.5.

Due to these negative signs, when we take derivatives of the potential to compute quantities like $\langle E \rangle_\nu = \frac{\partial}{\partial \beta} [\beta\Phi_\nu]$ and $\langle N \rangle_\nu = -\frac{\partial}{\partial \mu} \Phi_\nu \Big|_{\mu=0}$, we will encounter slightly different but equally enjoyable integrals:

$$\int_0^\infty \frac{x^3}{e^x + 1} dx = \left(1 - \frac{1}{2^3}\right) \Gamma(4)\zeta(4) = \frac{7\pi^4}{120},$$

$$\int_0^\infty \frac{x^2}{e^x + 1} dx = \left(1 - \frac{1}{2^2}\right) \Gamma(3)\zeta(3) = \frac{3}{2}\zeta(3).$$

Using these results, what are the average particle number and the average internal energy for a gas of massless fermions?

You should again find $\langle E \rangle_\nu \propto \langle N \rangle_\nu T \propto VT^4$, and by noting that

$$\frac{\Phi_\nu}{T} = -\frac{V}{\pi^2 c^3} \left(\frac{T}{\hbar}\right)^3 \int_0^\infty x^2 \log [1 + e^{-x}] dx \propto VT^3,$$

we can see that the entropy $S_\nu = (\langle E \rangle_\nu - \Phi_\nu)/T$ is constant when $VT^3 = \text{const.}$ Applying this, what is the pressure for a gas of massless fermions?

You should find an equation of state with the usual functional form and just a new numerical factor:

$$P_\nu V = \frac{1}{3} \langle E \rangle_\nu = \frac{7\pi^4}{540\zeta(3)} \langle N \rangle_\nu T = \frac{(7/8)\zeta(4)}{(3/4)\zeta(3)} \langle N \rangle_\nu T \approx 1.05 \langle N \rangle_\nu T.$$

8.8 Density of states & Sommerfeld expansion

In Eq. 103 we found that the average internal energy of a non-relativistic fermion gas becomes independent of the temperature in the limit $T \rightarrow 0$. This means that its heat capacity,

$$c_v = \left. \frac{\partial}{\partial T} \langle E \rangle \right|_{N,V},$$

vanishes in this limit, which we could also see by considering the fluctuation–dissipation relation $c_v \propto \langle (E - \langle E \rangle)^2 \rangle$ with only a single micro-state.

To derive the non-trivial heat capacity for a gas with a small but non-zero temperature, we need to move beyond approximating the Fermi function

$$F(E) = \frac{1}{e^{\beta(E-\mu)} + 1}$$

as a step function, and return to the full Eq. 99 for the average particle number,

$$\langle N \rangle_f = V \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \int_0^\infty F(E) \sqrt{E} dE \equiv \int_0^\infty g(E) F(E) dE. \quad (107)$$

Here we have defined the **density of states**

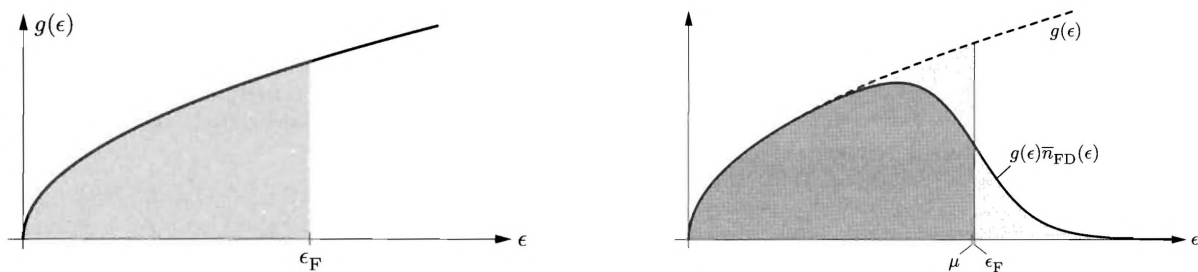
$$g(E) \equiv V \frac{\sqrt{2m^3}}{\pi^2 \hbar^3} \sqrt{E} \equiv g_0 \sqrt{E}$$

as the number of energy levels per unit energy, each of which can hold at most a single fermion. We can read Eq. 107 as saying that the total number of particles is given by integrating over the single-particle energy levels, $g(E)$, times the probability $F(E)$ that each of these energy levels is occupied.

The figures below, from Schroeder's *Introduction to Thermal Physics*, illustrate this integral in the case of $T = 0$ (left) and $T > 0$ (right). As we have already seen, when $T \rightarrow 0$ all energy levels with $E < E_F$ are occupied while all those with $E > E_F$ are unoccupied. With $T > 0$, there is an exponentially suppressed probability for some energy levels with $E > E_F$ to be occupied. Because the Fermi energy is set by the number of particles,

$$E_F = \frac{\hbar^2}{2m} \left(\frac{3\pi^2 \langle N \rangle_f}{V} \right)^{2/3},$$

having some of these particles occupy energy levels with $E > E_F$ requires that an equal number of energy levels with $E < E_F$ be unoccupied.

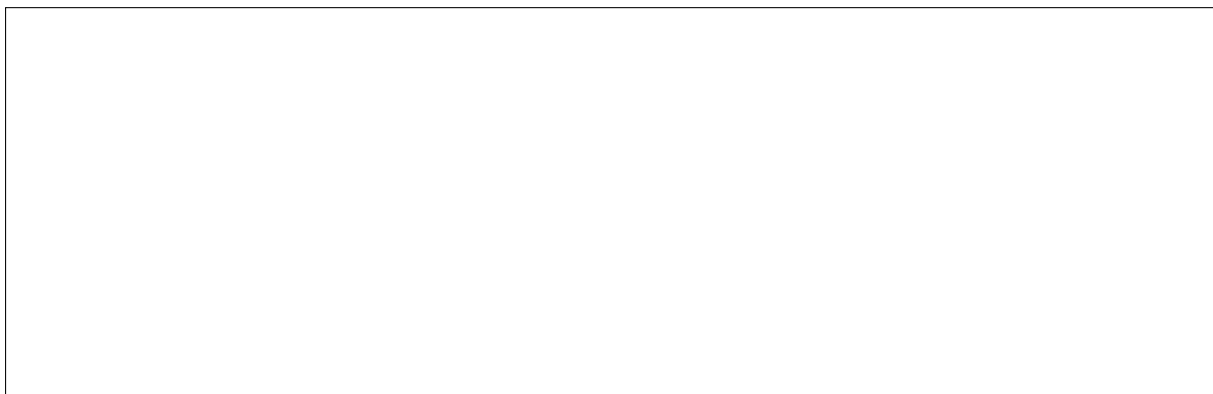


Note that $E = \mu$ is the point where the Fermi function $F(E) = \frac{1}{2}$. We will see below that $\mu \leq E_F$, as shown in the right figure above, with equality when $T = 0$.

In order to determine the particle number and internal energy, we need to evaluate the integrals

$$\langle N \rangle_f = g_0 \int_0^\infty F(E) \sqrt{E} dE \qquad \langle E \rangle_f = g_0 \int_0^\infty E F(E) \sqrt{E} dE \qquad (108)$$

without approximating the Fermi function as a step function. For $T \ll E_F$, we can do this through a **Sommerfeld expansion**. Let's begin by considering the particle number. The first step in the Sommerfeld expansion is integrating by parts:



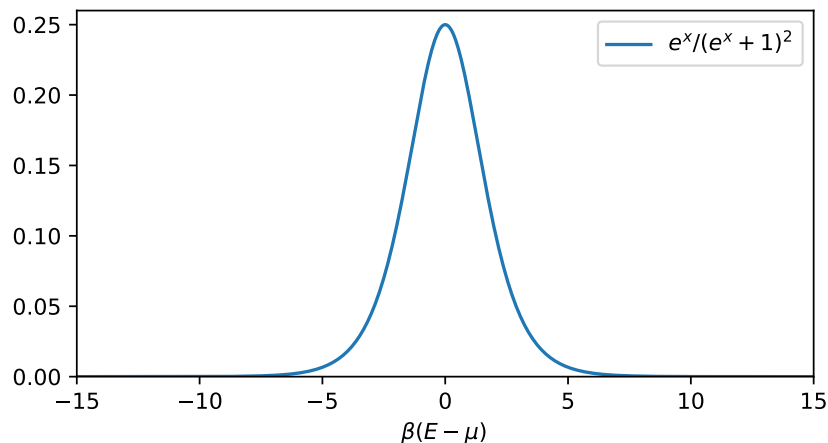
Changing variables to $x \equiv \beta(E - \mu)$, you should find

$$\langle N \rangle_f = \frac{2}{3} g_0 \int_{-\beta\mu}^{\infty} \frac{e^x}{(e^x + 1)^2} E^{3/2} dx.$$

This is not obviously simpler than the expression we started with (especially because E depends on x), but has the benefit of being exponentially suppressed for *both*

$$\begin{aligned} x \gg 1 &\implies \frac{e^x}{(e^x + 1)^2} \approx \frac{e^x}{e^{2x}} = \frac{1}{e^x} \\ \text{and } x \ll -1 &\implies \frac{e^x}{(e^x + 1)^2} \approx \frac{e^x}{1} = \frac{1}{e^{-x}}. \end{aligned}$$

The additional $E^{3/2}$ factor is far too mild to overcome this exponential suppression. In other words, non-negligible contributions to the integral as a whole come only from a region centered at $E = \mu$, which becomes narrower in E as the temperature decreases (corresponding to larger $\beta = 1/T$). This is illustrated by the plot below, which shows the exponential suppression setting in when $|E - \mu|$ is larger than a few times the temperature, and certainly for $|E - \mu| \gtrsim 5/\beta = 5T$.



These considerations justify two low-temperature approximations. First, recalling $\mu > 0$ at low temperatures, for large β we are free to extend the lower limit of the integral to obtain a more convenient domain of integration,

$$\langle N \rangle_f = \frac{2}{3} g_0 \int_{-\beta\mu}^{\infty} \frac{e^x}{(e^x + 1)^2} E^{3/2} dx \approx \frac{2}{3} g_0 \int_{-\infty}^{\infty} \frac{e^x}{(e^x + 1)^2} E^{3/2} dx.$$

Second, we can expand $E^{3/2}$ in a Taylor series around $E = \mu$, and truncate after the first few terms:

$$\begin{aligned} E^{3/2} &\approx \mu^{3/2} + (E - \mu) \left. \frac{\partial}{\partial E} E^{3/2} \right|_{E=\mu} + \frac{1}{2} (E - \mu)^2 \left. \frac{\partial^2}{\partial E^2} E^{3/2} \right|_{E=\mu} \\ &= \mu^{3/2} + \frac{3}{2} (E - \mu) \mu^{1/2} + \frac{3}{8} (E - \mu)^2 \mu^{-1/2} = \mu^{3/2} + \frac{3}{2} \frac{x \mu^{1/2}}{\beta} + \frac{3}{8} \frac{x^2 \mu^{-1/2}}{\beta^2}. \end{aligned}$$

Switching to work with $T = 1/\beta$, the Sommerfeld expansion has given us a series of manageable integrals we can consider one by one:

$$\langle N \rangle_f \approx \frac{2}{3} g_0 \mu^{3/2} \mathcal{I}_0 + g_0 T \mu^{1/2} \mathcal{I}_1 + \frac{g_0 T^2}{4 \mu^{1/2}} \mathcal{I}_2.$$

$$\mathcal{I}_0 = \int_{-\infty}^{\infty} \frac{e^x}{(e^x + 1)^2} dx =$$

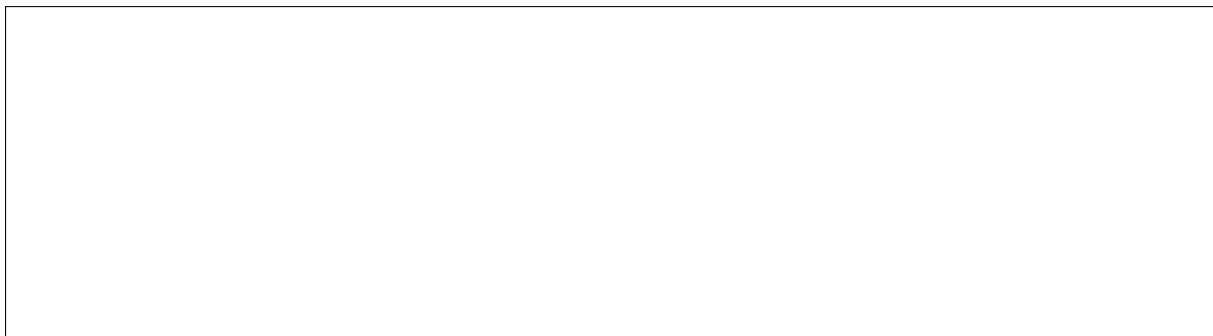
$$\mathcal{I}_1 = \int_{-\infty}^{\infty} \frac{x e^x}{(e^x + 1)^2} dx =$$

$$\mathcal{I}_2 = \int_{-\infty}^{\infty} \frac{x^2 e^x}{(e^x + 1)^2} dx =$$

Collecting the results and restoring $g_0 = V \frac{\sqrt{2m^3}}{\pi^2 \hbar^3}$, you should find

$$\langle N \rangle_f \approx \frac{2}{3} g_0 \mu^{3/2} + g_0 \frac{\pi^2 T^2}{12 \mu^{1/2}} = V \frac{(2m\mu)^{3/2}}{3\pi^2 \hbar^3} + V \frac{\sqrt{2m^3}}{12 \hbar^3 \mu^{1/2}} T^2.$$

The first term reproduces what we found with the step-function approximation in Section 8.5, while the second term provides the promised leading-order temperature dependence in the Sommerfeld expansion. This becomes more interesting if we rearrange Eq. 101 to work in terms of the Fermi energy:



The result

$$\frac{\mu}{E_F} \approx \left[1 - \frac{\pi^2 T^2}{8 E_F^{3/2} \mu^{1/2}} \right]^{2/3}$$

can be simplified through one final low-temperature approximation. Because T is small, just in the second term above we can set $E_F \approx \mu$ (the zero-temperature relation). Then

$$\frac{\mu}{E_F} \approx 1 - \frac{\pi^2 T^2}{12 E_F^2}, \quad (109)$$

which confirms our earlier claim $\mu \leq E_F$, and reveals that the leading correction to the zero-temperature relation is quadratic in $\frac{T}{E_F} \ll 1$.

The calculation is essentially the same for the internal energy from Eq. 108. With $E^{3/2}$ in place of $E^{1/2}$, integrating by parts just gives

$$\langle E \rangle_f = g_0 \int_0^\infty \frac{E^{3/2}}{e^{\beta(E-\mu)} + 1} dE \approx \frac{2}{5} g_0 \int_{-\infty}^\infty \frac{e^x}{(e^x + 1)^2} E^{5/2} dx$$

with the same $x = \beta(E - \mu)$ and extended limit of integration. The Taylor expansion

$$E^{5/2} \approx \mu^{5/2} + \frac{5}{2}(E - \mu)\mu^{3/2} + \frac{15}{8}(E - \mu)^2\mu^{1/2} = \mu^{5/2} + \frac{5}{2} \frac{x\mu^{3/2}}{\beta} + \frac{15}{8} \frac{x^2\mu^{1/2}}{\beta^2}$$

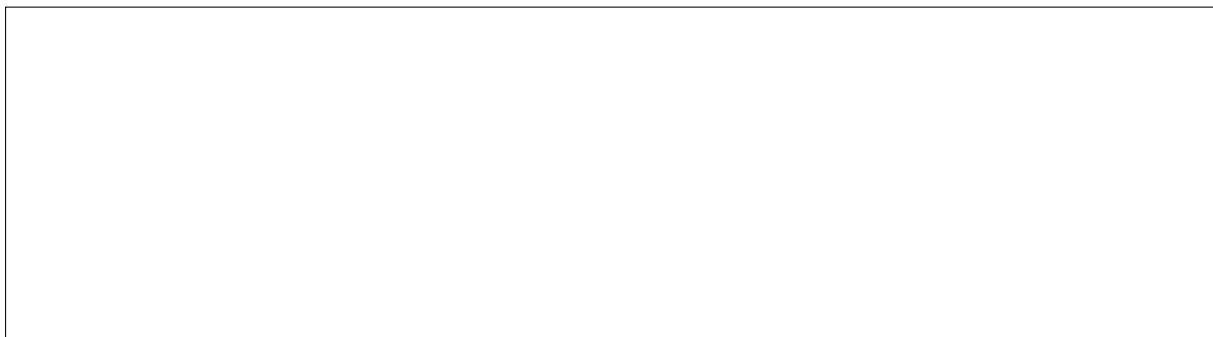
also produces the same integrals, with different coefficients:

$$\langle E \rangle_f \approx \frac{2}{5} g_0 \mu^{5/2} \mathcal{I}_0 + g_0 T \mu^{3/2} \mathcal{I}_1 + \frac{3}{4} g_0 T^2 \mu^{1/2} \mathcal{I}_2 = \frac{2}{5} g_0 \mu^{5/2} + \frac{1}{4} g_0 \pi^2 T^2 \mu^{1/2}.$$

Inserting $g_0 = \frac{3\langle N \rangle_f}{2E_F^{3/2}}$, we have

$$\langle E \rangle_f \approx \frac{3}{5} \langle N \rangle_f \frac{\mu^{5/2}}{E_F^{3/2}} + \frac{3}{8} \langle N \rangle_f \pi^2 T^2 \frac{\mu^{1/2}}{E_F^{3/2}},$$

which we can simplify by applying Eq. 109 and dropping $\mathcal{O}(T^3)$ terms:



From your result you should obtain the heat capacity

$$c_v = \left. \frac{\partial}{\partial T} \langle E \rangle \right|_{N,V} \approx \frac{\pi^2}{2} \frac{\langle N \rangle_f}{E_F} T.$$

This low-temperature linear dependence on T agrees with experimental heat capacity measurements, as we have seen in tutorials.

As a final comment in this section, let's consider what would happen to $\langle N \rangle_f$ at higher temperatures $T^2 \sim E_F^2$, for which the two terms in Eq. 109 would cancel out, leaving $\mathcal{O}(T^4/E_F^4)$ effects non-negligible. In this regime, there's no guarantee that the low-temperature Sommerfeld expansion would even converge, so we need to work with the full integral from Eq. 108. Fortunately, it is not hard to numerically evaluate this integral, which is done by [this Python code](#). For the purpose of numerical analysis, it's best to express everything in terms of dimensionless ratios, such as

$$t \equiv \frac{T}{E_F} \qquad c \equiv \frac{\mu}{E_F} \qquad x \equiv \frac{E}{T} = \beta E.$$

Also inserting $g_0 = \frac{3\langle N \rangle_f}{2E_F^{3/2}}$, we have

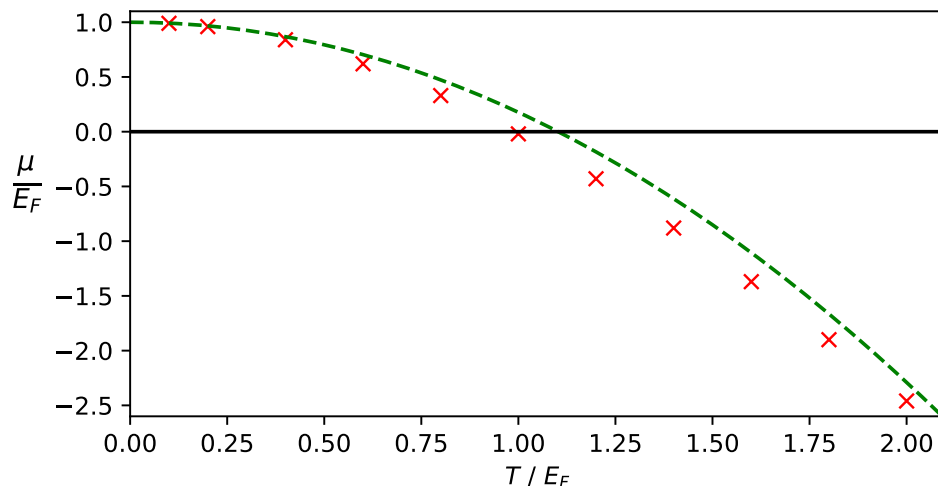
$$\langle N \rangle_f = \frac{3\langle N \rangle_f}{2E_F^{3/2}} \int_0^\infty \frac{\sqrt{E}}{e^{\beta(E-\mu)} + 1} dE = \frac{3\langle N \rangle_f}{2} t^{3/2} \int_0^\infty \frac{\sqrt{x}e^{-x}}{e^{-c/t} + e^{-x}} dx,$$

working with small $e^{-\beta E} = e^{-x}$ rather than large e^x to avoid numerical overflow.

As in Eq. 109, $\langle N \rangle_f$ drops out, and we end up with the consistency condition

$$1 = \frac{3}{2} t^{3/2} \int_0^\infty \frac{\sqrt{x}e^{-x}}{e^{-c/t} + e^{-x}} dx. \tag{110}$$

If we fix the temperature $t = T/E_F$ in units of the Fermi energy, by numerically evaluating this integral with different values of $c = \mu/E_F$ we can determine the self-consistent value of the chemical potential, also in units of the Fermi energy. The red \times 's in the figure below are results of such work for eleven temperatures $0.1 \leq t \leq 2$, compared to the $\mathcal{O}(t^2)$ result from the Sommerfeld expansion, Eq. 109. This leading-order Sommerfeld expansion clearly deviates from the full results by the time $T \sim E_F$. The more interesting result is that the chemical potential continues to decrease as the temperature increases, becoming negative for $T \gtrsim E_F$ and approaching the expected high-temperature classical limit.



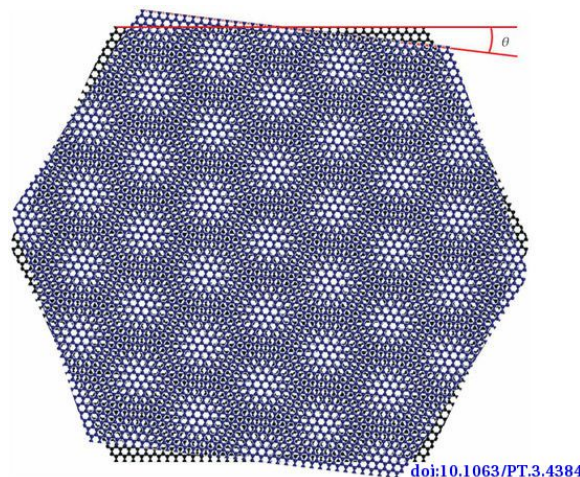
Unit 9: Interacting systems

9.1 The Ising model

So far in this module we have considered ‘ideal’ systems whose degrees of freedom do not interact with each other. While we have seen that statistical mechanics employing this approximation of non-interacting particles can provide excellent descriptions of real physical systems ranging from the low-temperature heat capacities of solids to solar radiation and the cosmic microwave background, there are crucial emergent phenomena that this approach fails to predict.

An important class of examples, which we investigate in this unit, are *phase transitions*. These occur when interactions allow extremely different large-scale behaviours to arise from the same set of degrees of freedom, depending on control parameters such as the temperature or pressure. Phase transitions occur in both everyday and extreme situations. Everyday examples include the liquid–gas transition of H_2O molecules from water to steam upon boiling a kettle, as well as the transition from liquid water to solid ice as the temperature decreases. In the extreme conditions following the big bang, matter in the universe existed as a charged plasma of quarks and gluons. Once the universe was a few micro-seconds old, it cooled enough for this matter to transition into the protons and neutrons we are made out of today.

An intermediate example in between the everyday and the extreme involves two layers of graphene as illustrated in the figure below.¹⁷ Graphene is an amazing material (recognized by the 2010 Nobel Prize in Physics) that consists of a single-atom-thick sheet of carbon atoms arranged in a hexagonal ‘honeycomb’ lattice. Under most conditions, graphene is an electrical insulator. However, if two graphene sheets are stacked and rotated with respect to each other by a small “magic angle” $\theta \approx 1.1^\circ$, the system transitions into a superconducting phase at low temperatures $T \lesssim 1.7$ K. Superconductivity allows electrical current to flow with no resistance, meaning that no energy is lost to the production of waste heat. If we could discover or design materials that exhibit superconductivity at everyday temperatures $T \sim 300$ K rather than low $T \sim \mathcal{O}(1)$ K, it would revolutionize the energy efficiency of electronics and the power grid.



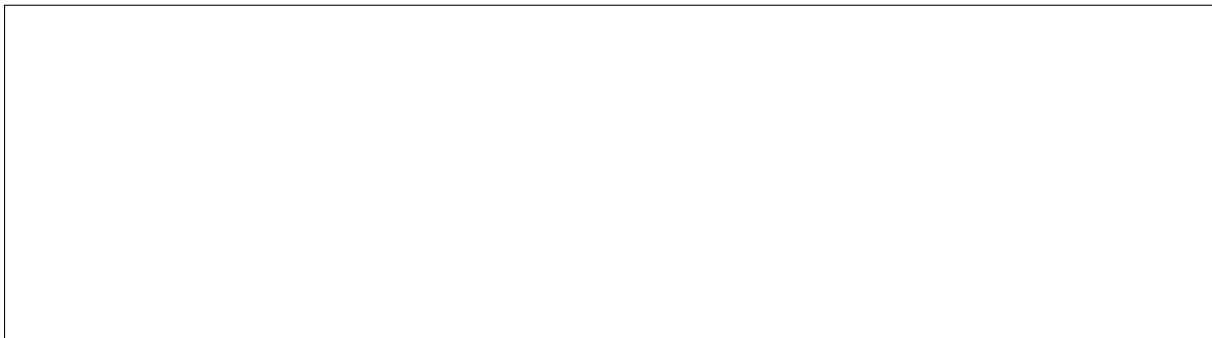
¹⁷Heather M. Hill, “[Twisted bilayer graphene enters a new phase](#)”, *Physics Today* **73**:18, 2020.

With this motivation for investigating phase transitions, let's step back to introduce interactions and explore their effects using simple spin systems of the sort we considered in units 2 and 3. In the non-interacting case we previously analysed (Eq. 40), the internal energy of the system is

$$E_i = -H \sum_{n=1}^N s_n \quad (\text{non-interacting})$$

for micro-state ω_i specified by the N spins $\{s_n\}$ (and, as always for this module, in thermodynamic equilibrium). Here $H > 0$ is the constant strength of an external magnetic field and the orientation of the n th spin, s_n , takes one of only two possible values: $s_n = 1$ if the spin is aligned parallel to the field and $s_n = -1$ if the spin is aligned anti-parallel to the field. The ground state of the system features all N spins aligned parallel to the magnetic field, with minimal energy $E_0 = -NH$.

In this unit we will only consider systems of distinguishable spins that we label by their fixed position in a d -dimensional simple cubic **lattice**. The $d = 1$ case of a one-dimensional lattice is precisely the system of spins arranged in a line that we analysed in Section 3.4.1. This and the case of $d = 2$ are both easy to visualize and draw on a sheet of paper:



While we can only have physical lattices with $d = 1, 2$ or 3 in nature, the mathematical construction works just as well for any integer $d \geq 1$.

We can see that the total internal energy of the non-interacting system can easily be written as a sum over energies ε_n for each individual spin,

$$\varepsilon_n = -H s_n \quad E_i = \sum_{n=1}^N \varepsilon_n \quad (\text{non-interacting}).$$

This is a generic feature of non-interacting systems, and an aspect of the **factorization** that enormously simplifies calculations — in this case by causing the N -particle partition function (Eq. 41) to take the form of a product of N identical terms, $Z_N = [2 \cosh(\beta H)]^N = Z_1^N$. However, it is possible to have non-factorizable systems in which the internal energy can be expressed as a sum of this sort. A stronger condition needs to be satisfied in order to guarantee factorization, and this conditions rigorously defines what it means for a system to be non-interacting.

Let ΔE_j be the change in the system's internal energy caused by changing its j th degree of freedom. Then the system is defined to be **non-interacting** if and only if ΔE_j is independent of all other degrees of freedom $k \neq j$.

For our system of N distinguishable spins, the only possible change we can make to a degree of freedom is to negate it, $s_j \rightarrow -s_j$, which corresponds to flipping its alignment relative to the external magnetic field. It is easy to check that the change in the internal energy resulting from such a spin flip satisfies our definition of a non-interacting system:



Now let's make things more interesting by considering a different spin system that includes a simple two-spin contribution to the internal energy:

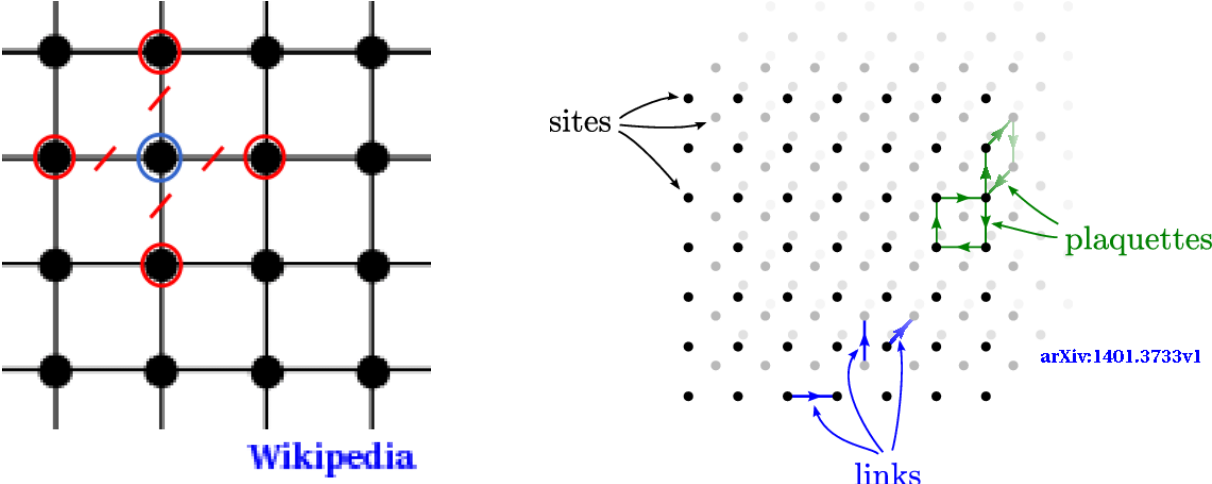
$$E_i = - \sum_{(jk)} s_j s_k - H \sum_{n=1}^N s_n. \quad (111)$$

The first sum runs over all pairs of nearest-neighbour spins in the lattice, denoted (jk) . What is the change in energy ΔE_j from Eq. 111 upon negating $s_j \rightarrow -s_j$? Does this indicate an interacting or non-interacting system?

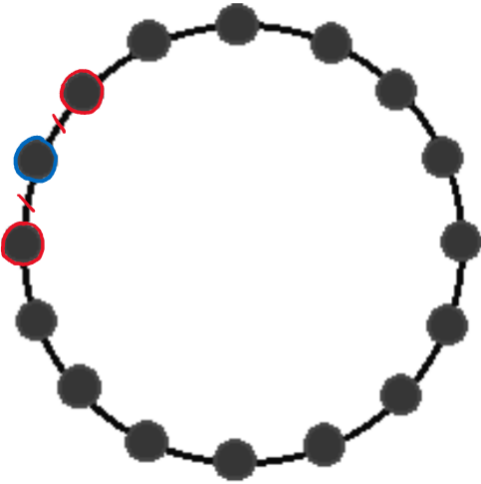


The pictures on the next page illustrate nearest-neighbour pairs for simple cubic lattices in $d = 2$ and 3 dimensions, while also introducing some additional lattice terminology. Instead of drawing up- and down-pointing arrows, these pictures identify the spins with *sites* in the lattice represented as points (or larger filled circles). In simple cubic lattices, all sites are positioned in a regular grid, separated by a constant distance along each basis vector. We can also draw *links* as solid lines connecting these nearest-neighbour sites, with each link corresponding to a term in $\sum_{(jk)}$. The picture of

a two-dimensional lattice on the left highlights the four links (with red hatch marks) that correspond to the four nearest neighbours (circled in red) of a particular site (circled in blue). For $d \geq 2$, the elementary unit of surface area is called a *plaquette*, while for $d \geq 3$ the elementary unit of volume is called a *cube*.

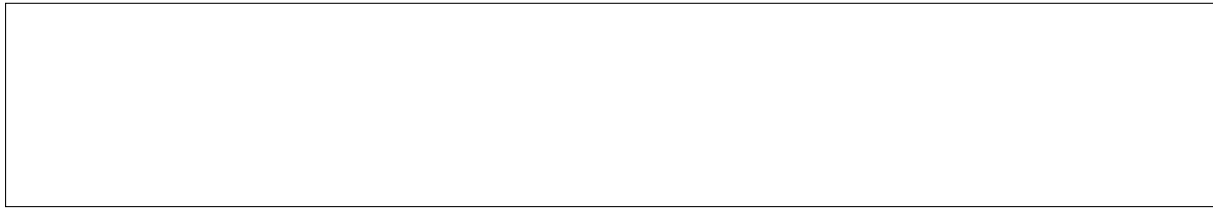


The nearest-neighbour pairs appearing in the energy, Eq. 111, correspond to the links in the lattice, $\ell = (jk)$. Assuming a finite lattice, the presence of any edges will make analyses more complicated, since sites on the edges or in the corners will have a different number of nearest neighbours than other sites. We can avoid this complication by imposing **periodic boundary conditions**, which remove these edges by adding links between each site on the left edge of the lattice and the corresponding site on the right edge, and similarly in all other dimensions. This is illustrated below for the simple case of the one-dimensional lattice, drawn as a circle to emphasize that all N sites remain separated by a constant distance. In higher dimensions, periodic boundary conditions produce flat (zero-curvature) d -dimensional tori that preserve the simple cubic lattice structure.



Periodic boundary conditions cause the one-dimensional lattice drawn above to have the same number of links as sites: Each site has two links connecting it to its two nearest neighbours, and each of those links is shared between two sites, so that

$\#l = 2N/2 = N$. Looking back to the two-dimensional lattice drawn farther above, the four links per site produce $\#l = 4N/2 = 2N$. How many terms are there in the sum $\sum_{(jk)}$ in Eq. 111 for N -site lattices with periodic boundary conditions in d dimensions?



The energy of interacting spins given by Eq. 111, with a lattice structure specifying which spins form the nearest-neighbour pairs (jk) , defines a famous system known as the d -dimensional **Ising model**. Since the 1960s, the Ising model has been the basis of thousands of scientific studies analysing everything from ferromagnetism to neural networks to urban segregation.¹⁸ The model was proposed in 1920 by [Wilhelm Lenz](#), whose PhD student [Ernst Ising](#) solved the one-dimensional system for his thesis in 1924. Exactly solving the two-dimensional case (with $H = 0$) took another twenty years, culminating in renowned work by [Lars Onsager](#) in 1944. The three-dimensional Ising model remains an open mathematical question, with no known exact solution.

In this context, ‘solving’ the Ising model means deriving a closed-form expression for its canonical partition function,

$$Z(\beta, N, H) = \sum_{\{s_n\}} \exp[-\beta E(s_n)] = \sum_{\{s_n\}} \exp \left[\beta \sum_{(jk)} s_j s_k + \beta H \sum_n s_n \right].$$

As in Section 3.4, the partition function sums over all possible spin configurations $\{s_n\}$, which amounts to a sum of 2^N exponential factors for N spins, with $\mathcal{O}(N)$ terms within each exponential. Now that the system is interacting, the partition function no longer factorizes into the N identical \cosh factors of Eq. 41, making it extremely difficult to evaluate. This is why there is no known exact solution for the three-dimensional Ising model, and it also makes ‘brute-force’ numerical computations impractical. Even for a system of $N = 1023$ spins, twenty orders of magnitude smaller than our typical $N \sim 10^{23}$, there are roughly $2^{1023} \sim 10^{310}$ terms in the partition function, far beyond the capabilities of existing or foreseeable supercomputers.

9.2 Ising model phases and phase transition

Even without exactly solving the Ising model, we can still make robust predictions for its large-scale behaviour by considering the simplified limits of high and low temperature, much as we did for non-interacting spin systems in Section 3.4. We can also simplify the system by setting $H = 0$ in this section, and considering just

$$E_i = - \sum_{(jk)} s_j s_k \qquad Z(\beta, N) = \sum_{\{s_n\}} \exp \left[\beta \sum_{(jk)} s_j s_k \right]. \qquad (112)$$

¹⁸For a brief discussion, see Charlie Wood, “[The Cartoon Picture of Magnets That Has Transformed Science](#)”, *Quanta Magazine*, 2020.

We will see that the behaviour of this zero-field Ising model is qualitatively different at high temperatures compared to low temperatures. In other words, the system exhibits at least two distinct phases for different temperatures. This is a necessary but not sufficient condition for there to be a true phase transition — it leaves open the possibility of a gradual *crossover* between these two phases, as opposed to a rapid transition. In this section we will use the Ising model to more rigorously define what exactly constitutes a phase transition, and how this can be distinguished from a crossover.

First, though, let's consider the **high-temperature** limit $\beta \rightarrow 0$, where the Ising model partition function becomes extremely simple:



In this limit, all 2^N spin configurations are adopted with the same probability $p_i = 1/2^N$, regardless of their internal energy from Eq. 112. In effect, that energy has become negligible compared to the temperature.

There is a simple observable we can use to characterize this high-temperature phase. This is the **magnetization** $M = n_+ - n_-$, retaining our definition of n_{\pm} as the number of spins with value ± 1 , even without an external field for them to align with or align against. It is convenient to normalize the magnetization by the number of spins,

$$m \equiv \frac{M}{N} = \frac{n_+ - n_-}{n_+ + n_-}, \quad (113)$$

so that $-1 \leq m \leq 1$ for any value of N . In addition, without an external field to distinguish between ± 1 spins, it is also convenient to consider the absolute magnitude $0 \leq |m| \leq 1$.

Our task is now to determine the expectation value of the magnetization at high temperatures. Above we found that all spin configurations are equally probable in this regime, so $\langle |m| \rangle$ will be determined by how many of these equally probable micro-states have a particular magnetization. For example, there are only two micro-states with $|m| = 1$, corresponding to $(n_+, n_-) = (N, 0)$ and $(0, N)$. In general, just as we saw in Eq. 23, there are

$$\binom{N}{n_+} = \binom{N}{n_-} = \frac{N!}{n_+! n_-!}$$

equally probable micro-states with a given $n_+ = N - n_-$. For large $N \gg 1$ this binomial coefficient has a factorially narrow peak around

$$n_+ = n_- = \frac{1}{2}N \quad \longrightarrow \quad |m| = 0.$$

This characterizes a **disordered phase** with similar numbers of up- and down-pointing spins producing a small magnetization. In the *thermodynamic limit* $N \rightarrow \infty$, the expectation value of the magnetization in the disordered phase vanishes exactly, $\langle |m| \rangle \rightarrow 0$.

We next need to determine $\langle |m| \rangle$ in the **low-temperature** limit $\beta \rightarrow \infty$. In this regime, as we saw in Section 3.4.1, the Boltzmann factor $\exp \left[\beta \sum_{\langle jk \rangle} s_j s_k \right]$ makes it exponentially more likely for the system to adopt micro-states with lower energies. In particular, we can expect the ground state to dominate the expectation value of the magnetization, $\langle |m| \rangle$, up to exponentially suppressed corrections from excited states. With $H = 0$, the Ising model has two degenerate ground states corresponding to the two ways all the spins can be aligned with each other: $(n_+, n_-) = (N, 0)$ and $(0, N)$. What is the ground-state energy of the N -site Ising model in d dimensions?

As mentioned above, both of these degenerate ground states have the maximal magnetization $|m| = 1$. Let's check what effect the first excited state would have on the overall magnetization of the system. For $d > 1$, the first excited energy level is obtained by flipping a single spin, negating its value. Starting from the two degenerate ground states, this produces all possible micro-states with $(n_+, n_-) = (N - 1, 1)$ and $(1, N - 1)$. Because any one of the N spins in the lattice could be flipped, the degeneracy of this first excited energy level grows with N :

$$\binom{N}{1} + \binom{N}{N-1} = 2N.$$

At the same time, as N increases the magnetization of each of these micro-states gets closer to that of the ground state,

$$|m| = \frac{N-2}{N} = 1 - \frac{2}{N}.$$

The key factor is the probability for the system to be in one of these micro-states, which depends on the value of the energy E_1 for the first excited energy level. What is this E_1 for the N -site Ising model in d dimensions?

Let's bring everything together by computing the relative probability for the d -dimensional Ising model to be in its ground state with $|m| = 1$ compared to its first excited state with $|m| = 1 - \frac{2}{N}$. We just need to multiply the degeneracy of each energy

level times the corresponding Boltzmann factor, while the $\frac{1}{Z}$ normalization cancels out in the ratio

$$\frac{P(E_0)}{P(E_1)} = \frac{2 \cdot \exp[\beta d \cdot N]}{2N \cdot \exp[\beta(d \cdot N - 4d)]} = \frac{\exp[4\beta d]}{N}.$$

For any fixed N , a sufficiently low temperature will cause the ground state to dominate, with exponentially suppressed contributions from higher energy levels, just as we previously found for simpler non-interacting systems. This characterizes an **ordered phase** with essentially all spins aligned in the same direction, producing a large expectation value for the magnetization, $\langle |m| \rangle \rightarrow 1$.

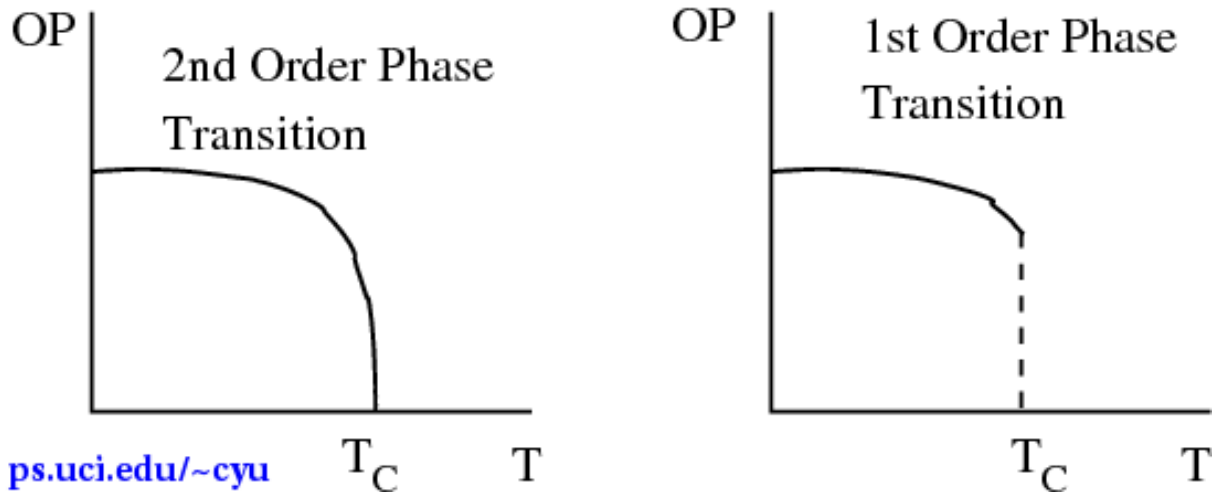
We have now seen how the behaviour of the magnetization $\langle |m| \rangle$ distinguishes the high- and low-temperature phases of the zero-field Ising model in $d > 1$ dimensions. In the high-temperature disordered phase, the magnetization is small and $\langle |m| \rangle \rightarrow 0$ in the thermodynamic limit $N \rightarrow \infty$. In the low-temperature ordered phase, the magnetization is large and $\langle |m| \rangle \rightarrow 1$ as $T \rightarrow 0$.

This contrast between ordered and disordered phases is typical behaviour for interacting statistical systems. These two phases are distinguished by an **order parameter** — an observable (related to a derivative of the free energy) that is zero in the disordered phase but non-zero in the ordered phase.¹⁹ The magnetization is the order parameter for the Ising model, and we will relate it to the free energy in the next section. Note that the order parameter need not reach its maximum value in the ordered phase — it only needs to be non-zero. The details of how the order parameter changes between zero and non-zero values distinguish crossovers from true transitions.

A phase transition is defined by a discontinuity or divergence in the order parameter or its derivative(s), in the $N \rightarrow \infty$ thermodynamic limit. The value(s) of the control parameter(s) at which the discontinuity occurs define the *critical point* corresponding to the transition.

For the zero-field Ising model, since we have set $H = 0$, the only remaining control parameter is the temperature T . Any phase transition would therefore occur at a **critical temperature** T_C . The sketches on the next page illustrate the most common types of phase transitions. When the order parameter (OP) itself is discontinuous (shown by a dashed line), the transition is said to be a *first-order* phase transition. When the order parameter is continuous at T_C but its first derivative with respect to the control parameter is discontinuous (typically divergent), the transition is said to be a *second-order* phase transition. Historically, this naming scheme was extended to n th-order phase transitions for which discontinuities don't occur until the $(n-1)$ th derivative of the order parameter (related to an n th derivative of the free energy). At present it is more common for any phase transition with a continuous order parameter to be called simply a second-order transition.

¹⁹There are atypical (but interesting and important) *topological phase transitions* that are not characterized by such an order parameter. The most famous example is the BKT phase transition named after [Vadim Berezinskii](#), [J. Michael Kosterlitz](#) and [David Thouless](#), which was recognized by the 2016 Nobel Prize in Physics. It is also possible for a single system to have multiple distinct phase transitions, each characterized by a different order parameter.



In practice, any system with a finite number of degrees of freedom will not exhibit a true discontinuity or divergence in any observable. As a result, it is sometimes said that true phase transitions only occur in the $N \rightarrow \infty$ thermodynamic limit, but I consider this excessively pedantic, especially given the finite number of atoms in the observable universe. We are still able to distinguish crossovers from true phase transitions when considering a finite number of degrees of freedom, by analysing the way in which the system approaches the thermodynamic limit. If there is time and interest, we can explore such *finite-size scaling*, but first we will develop a useful approximation technique, and apply it to the Ising model to investigate its (dimensionality-dependent) phase transition in more detail.

9.3 The mean-field approximation

Having identified the low-temperature ordered phase and high-temperature disordered phase of the zero-field Ising model, let's now restore a non-zero external magnetic field, $H > 0$. This will allow us to gain a deeper appreciation of the magnetization (now with no absolute value) by noting that Eq. 113 means it is just the average spin:

$$m = \frac{M}{N} = \frac{1}{N} (n_+ - n_-) = \frac{1}{N} \sum_{n=1}^N s_n.$$

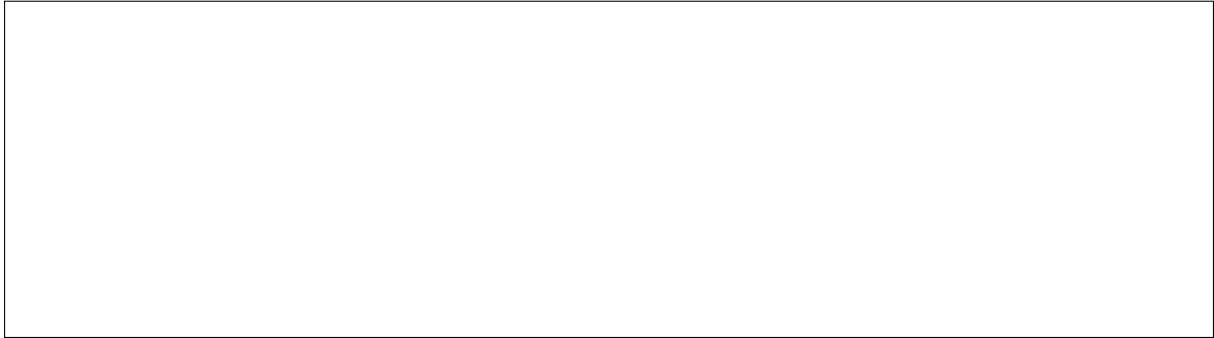
We can benefit from this observation in two ways. First, we can recognize the magnetization in the internal energy of the full Ising model with $H > 0$:

$$E_i = - \sum_{(jk)} s_j s_k - H \sum_{n=1}^N s_n = - \sum_{(jk)} s_j s_k - H N m = - \sum_{(jk)} s_j s_k - H M.$$

The corresponding canonical partition function is

$$Z = \sum_{\{s_n\}} \exp \left[\beta \sum_{(jk)} s_j s_k + \beta H M \right].$$

Based on this expression, and our earlier experience with the entropy and internal energy, we can anticipate that $\langle m \rangle = \langle M \rangle / N$ is related to the derivative of the Helmholtz free energy $F = -T \log Z$ with respect to the field strength H :



As promised in the previous section, this relation ensures that the magnetization is an appropriate order parameter for the Ising model phase transition.

The second way we can benefit from relating the magnetization to the average spin is to express the Ising model in terms of the expectation value

$$\langle m \rangle = \frac{1}{Z} \sum_{\{s_n\}} m e^{-\beta E(s_n)} = \frac{1}{N} \sum_{n=1}^N \langle s_n \rangle.$$

The expectation value of the average spin, $\frac{1}{N} \sum_{n=1}^N \langle s_n \rangle$, is independent of the spin configuration $\{s_n\}$ and is simply a function of the inverse temperature β and magnetic field strength H . By adding and subtracting factors of $\langle m \rangle$, we can exactly rewrite each nearest-neighbour term in the Ising model energy, Eq. 111, as

$$\begin{aligned} s_j s_k &= [(s_j - \langle m \rangle) + \langle m \rangle] \times [(s_k - \langle m \rangle) + \langle m \rangle] \\ &= (s_j - \langle m \rangle)(s_k - \langle m \rangle) + (s_j + s_k)\langle m \rangle - \langle m \rangle^2. \end{aligned} \quad (114)$$

This is beneficial because we can note that the factors of $(s_j - \langle m \rangle)$ correspond to the spins' fluctuations around their mean value $\langle m \rangle$. By conjecturing that these fluctuations are small *on average*, we can approximate the Ising model energy by neglecting the first term in Eq. 114 when summing over all links:

$$E_i = - \sum_{(jk)} s_j s_k - H \sum_{n=1}^N s_n \quad \longrightarrow \quad E_{\text{MF}} = - \sum_{(jk)} [(s_j + s_k)\langle m \rangle - \langle m \rangle^2] - H \sum_{n=1}^N s_n.$$

The sum over the links $\ell = (jk)$ in d dimensions simply counts $d \cdot N$ factors of the constant $\langle m \rangle^2$. Similarly, since the first term includes both spins $(s_j + s_k)$ on each end of the link, every individual spin appears $2d$ times in the sum over links. Therefore this term just gives us $2d \langle m \rangle$ times another sum over the spins s_n , which we can combine with the final term above:

$$E_{\text{MF}} = d \cdot N \langle m \rangle^2 - (2d \langle m \rangle + H) \sum_{n=1}^N s_n \equiv d \cdot N \langle m \rangle^2 - H_{\text{eff}} \sum_{n=1}^N s_n. \quad (115)$$

In this expression we define an effective magnetic field $H_{\text{eff}} = 2d \langle m \rangle + H$ that depends on the mean spin. This approach of neglecting the squared fluctuations $(s_j - \langle m \rangle)(s_k - \langle m \rangle)$ is therefore known as the **mean-field approximation**. It can be applied to generic lattice systems, not just the Ising model, and essentially supposes that we can average over all $2d$ nearest neighbours of each lattice site and end up with an approximately constant factor that behaves like a modification of the magnetic field. Given the resulting mean-field energy E_{MF} from Eq. 115, let's check the change in this energy, ΔE_j , upon negating any $s_j \rightarrow -s_j$:



Your result should suggest that the mean-field approximation producing Eq. 115 makes it very easy to compute the corresponding canonical partition function

$$\begin{aligned} Z_{\text{MF}} &= \sum_{\{s_n\}} \exp[-\beta E_{\text{MF}}] = \exp[-\beta d \cdot N \langle m \rangle^2] \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} \exp\left[-x \sum_{n=1}^N s_n\right] \\ &= \exp[-\beta d \cdot N \langle m \rangle^2] (2 \cosh[\beta H_{\text{eff}}])^N \\ &= \exp[-\beta d \cdot N \langle m \rangle^2] (2 \cosh[\beta (2d \langle m \rangle + H)])^N, \end{aligned} \quad (116)$$

where we defined $x \equiv -\beta H_{\text{eff}}$ to put the sums into the same form as in Eq. 41. Although this factorized result is far simpler than the partition function for the full Ising model, it does involve some complicated dependence on $\langle m \rangle$ — especially when we recall that $\langle m \rangle$ itself is related to a derivative of $\log Z_{\text{MF}}$. With

$$\log Z_{\text{MF}} = N \log \cosh[\beta (2d \langle m \rangle + H)] + \{H\text{-independent terms}\},$$

the relation we derived above gives us

$$\langle m \rangle = \frac{1}{N\beta} \frac{\partial}{\partial H} \log Z_{\text{MF}} = \frac{1}{\beta} \frac{1}{\cosh[\beta (2d \langle m \rangle + H)]} \frac{\partial}{\partial H} \cosh[\beta (2d \langle m \rangle + H)].$$

Simplifying, we obtain a **self-consistency condition** for the Ising model magnetization in the mean-field approximation:

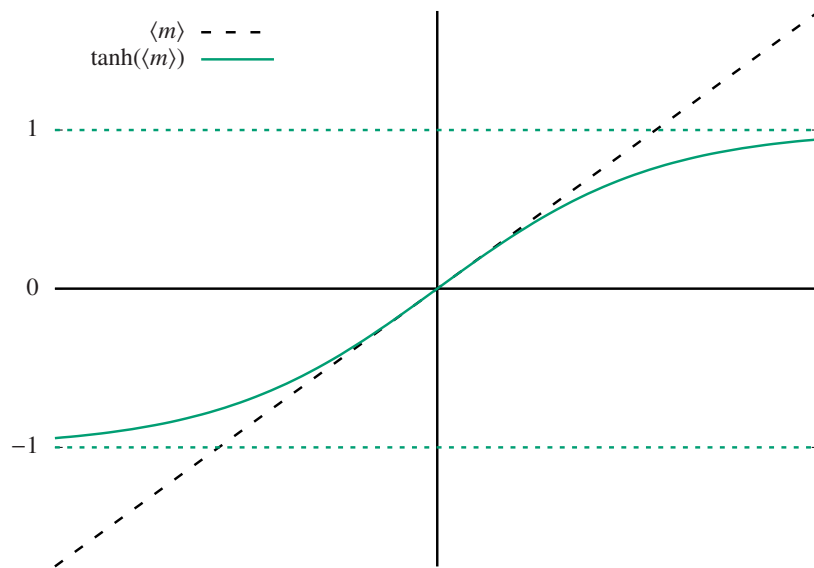
$$\langle m \rangle = \tanh[\beta (2d \langle m \rangle + H)]. \quad (117)$$

Solving this equation for $\langle m \rangle$ is equivalent to finding the roots of the transcendental equation $\tanh[\beta(2d \cdot x + H)] - x = 0$.

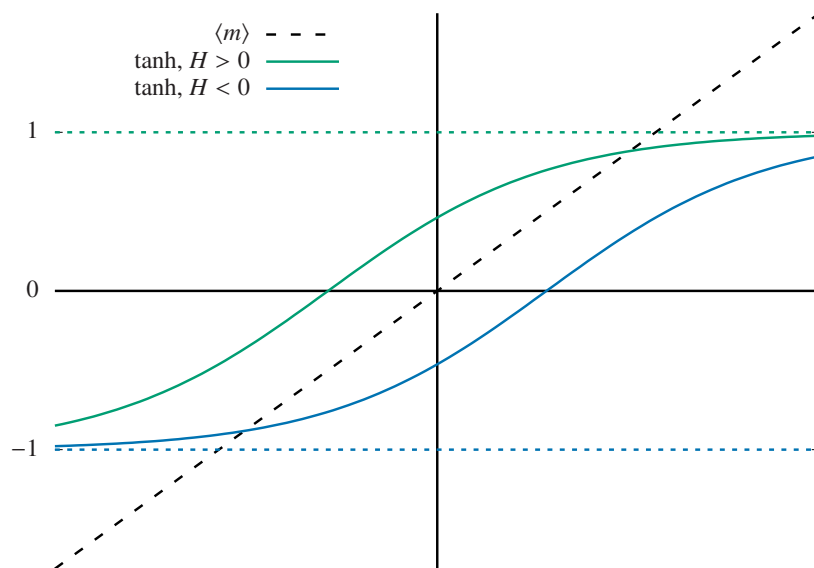
A straightforward way to inspect such solutions is by plotting both

$$f(\langle m \rangle) = \langle m \rangle \qquad g(\langle m \rangle) = \tanh[\beta(2d \langle m \rangle + H)]$$

and monitoring the intersections of these two functions. Fixing $d = 2$, the plot on the next page considers the simplest case $\beta = \frac{1}{4}$ and $H = 0$ for which $g(\langle m \rangle) = \tanh[\langle m \rangle]$ (the solid line). There is only a single intersection between this function and $f(\langle m \rangle)$ (the dashed line), at $\langle m \rangle = 0$, which suggests the disordered phase.

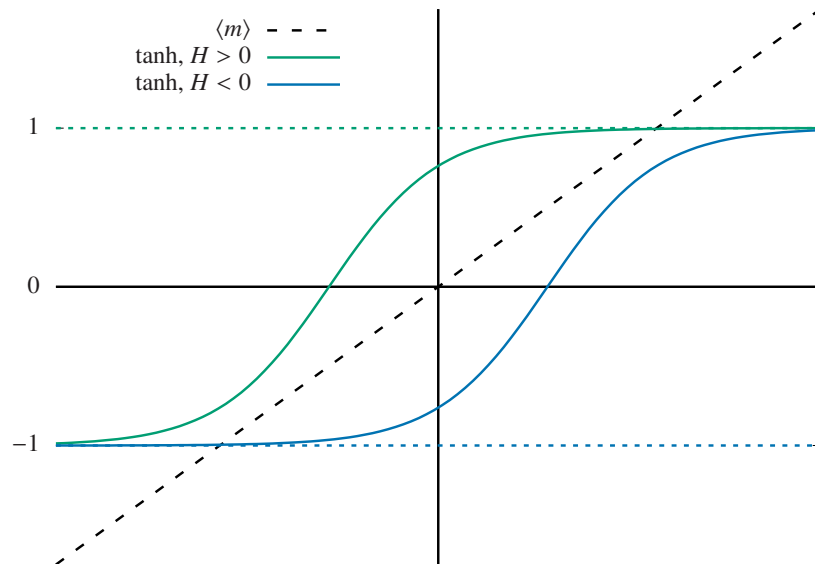


To confirm our interpretation of this result, let's check how the intersections depend on β and H . In the next plot below we keep the same temperature $T = 1/\beta = 4$ while comparing two non-zero values for the external magnetic field. A positive $H = 2$ simply shifts $g(\langle m \rangle)$ to the left (the green line), while a negative $H = -2$ shifts it to the right (the blue line). In both cases there is still only a single intersection, at $\langle m \rangle \approx \pm 0.88$ for $H = \pm 2$. We can interpret this non-zero result as an indication that the system is in an ordered phase where the spins tend to align with the external field.

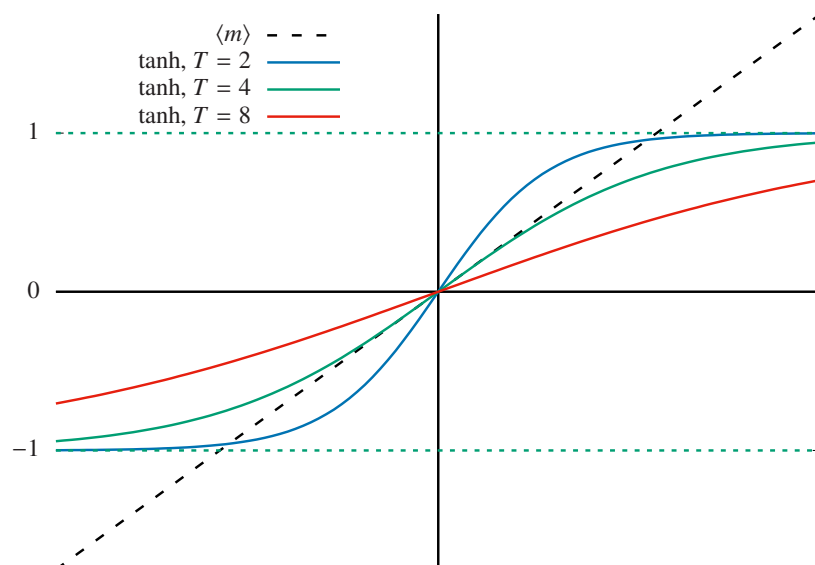


From our work in the previous section, we can expect that the spins' alignment will increase — reflecting the minimum-energy ground states — as the temperature decreases. Decreasing the temperature increases β , which causes the argument of the

\tanh to vary more rapidly with $\langle m \rangle$, making $g(\langle m \rangle)$ a steeper function that more rapidly approaches its limiting values ± 1 . The plot below illustrates this for $T = 1/\beta = 2$. For this temperature with magnetic field $H = \pm 2$, the intersection is $\langle m \rangle \approx \pm 1$ to a very good approximation. We can also appreciate that $-1 \leq \tanh x \leq 1$ ensures that the mean-field self-consistency condition can only ever be satisfied for $-1 \leq \langle m \rangle \leq 1$, reassuringly consistent with the definition of the magnetization.



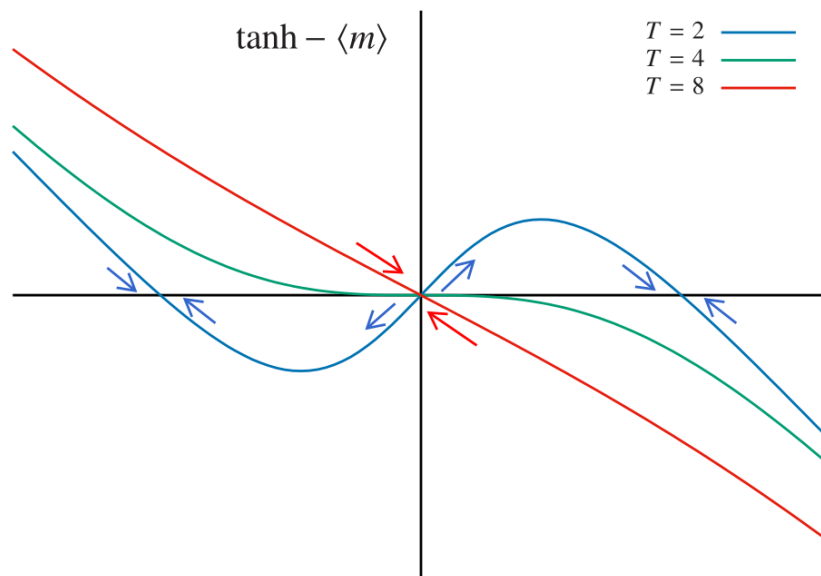
Also in the previous section, we saw that the Ising model spins should align at low temperatures, even without an external field to promote one direction over the other. To check whether this behaviour is preserved by the mean-field approximation, we can consider the self-consistency condition for various temperatures with $H = 0$. The plot below shows the results, considering a low temperature $T = 2$ (the blue line), the same $T = 4$ green curve shown in the first plot above, and a high temperature $T = 8$ (the red line). While the $\langle m \rangle = 0$ expected in the disordered phase is always a possible solution, something interesting happens at lower temperatures, where the steeper \tanh function introduces two additional solutions at non-zero $\langle m \rangle = \pm m_0$ corresponding to the ordered phase. As $T \rightarrow 0$, this magnetization approaches its maximum value $m_0 \rightarrow 1$.



When there are three solutions $\langle m \rangle = \{-m_0, 0, m_0\}$ at low temperatures, we can determine that the $\langle m \rangle = 0$ solution is actually *unstable*. Here we are venturing briefly into non-equilibrium territory, and thinking of the mean-field system as a ‘blind’ process that attempts to satisfy the self-consistency condition $\langle m \rangle = \tanh [2\beta d \langle m \rangle]$, based only on whether the expectation value of the magnetization is too small or too large compared to the \tanh . Once the magnetization is self-consistent, the system can happily settle into thermodynamic equilibrium.

From the figure above we can see that with $H = 0$ we can have three solutions only when the slope of the \tanh at $\langle m \rangle = 0$ is greater than 1. Any positive value $\langle m \rangle = \varepsilon > 0$ would then produce $\tanh [2\beta d \langle m \rangle] > \langle m \rangle$, which the system ‘feels’ as a magnetization that is too small to be self-consistent. This drives the system to continue increasing its magnetization, until it eventually settles at the non-zero solution $\langle m \rangle = m_0$. Similarly, any negative magnetization $\langle m \rangle = -\varepsilon < 0$ would drive $\langle m \rangle$ away from zero and to the $\langle m \rangle = -m_0$ solution.

This argument can be visualized more easily by plotting $\tanh [2\beta d \langle m \rangle] - \langle m \rangle$ vs. $\langle m \rangle$ as shown in the final plot below. When this difference is negative, $\langle m \rangle$ is larger than the self-consistency condition allows, driving the system to smaller $\langle m \rangle$ as shown by the arrows pointing to the left. Conversely, when the difference is positive, the system ‘seeks’ self-consistency by increasing $\langle m \rangle$ as shown by the arrows pointing to the right. For the low temperature $T = 2$, we see that the arrows move the system away from the unstable solution $\langle m \rangle = 0$ and to the stable solutions $\langle m \rangle = \pm m_0$.



At this point we can conclude that the non-interacting mean-field approximation successfully captures at least the high- and low-temperature limits of the interacting zero-field Ising model that we determined in the previous section. For high temperatures the mean-field self-consistency condition demands $\langle m \rangle = 0$ in the disordered phase, while for low temperatures it produces $\langle m \rangle = \pm m_0 \neq 0$ in the ordered phase.

Going further, now that we have a more tractable non-interacting system we can determine the value of the temperature at which its $\langle m \rangle = \pm m_0$ solutions appear and

the $\langle m \rangle = 0$ solution becomes unstable. As described above, this occurs whenever the slope of the \tanh function at $\langle m \rangle = 0$ is greater than 1. Let's call the corresponding temperature T_c , though it remains to be determined whether it is really the critical temperature of a true phase transition. Expanding $\tanh(x)$ for $x \approx 0$, what is T_c ?

You should find that the change from the high-temperature disordered phase to the low-temperature ordered phase occurs at $T_c = 2d$ in d dimensions, or equivalently $\beta_c = \frac{1}{2d}$ — corresponding to the green lines in the two figures above with $d = 2$. In order to determine whether or not this is a true critical temperature, we need to check whether the order parameter $\langle m \rangle$ or its T -derivatives are discontinuous at T_c . We can do this by considering the self-consistency condition for a temperature T lower than but very near to $T_c = 2d$, which would produce $0 < |\langle m \rangle| \ll 1$ and allow us to expand $\tanh(x) = x - \frac{x^3}{3} + \mathcal{O}(x^5)$. What is the resulting prediction for $\langle m \rangle$?

Making the approximation $\left(\frac{T}{T_c}\right)^2 \approx 1$, your result should resemble

$$\langle m \rangle = \pm \sqrt{3} \left(\frac{T_c - T}{T_c} \right)^{1/2} \quad \text{for } T \lesssim T_c.$$

From this, we can see that the order parameter $\langle m \rangle$ is continuous at T_c :

$$\langle m \rangle \propto \begin{cases} (T_c - T)^{1/2} & \text{for } T \lesssim T_c \\ 0 & \text{for } T \gtrsim T_c \end{cases}. \quad (118)$$

However, its first derivative

$$\frac{d\langle m \rangle}{dT} \propto \frac{1}{(T_c - T)^{1/2}}$$

diverges as $T \rightarrow T_c$ from below. This is the situation we discussed at the end of the previous section, which predicts a second-order phase transition with critical temperature $T_c = 2d$ in d dimensions. The power-law dependence $\langle m \rangle \propto (T_c - T)^b$ with $0 < b < 1$ is a generic feature of second-order phase transitions. The power b is known as a **critical exponent**, in this case $b = 1/2$.

At this point we have invested some effort to find that the mean-field approximation of the d -dimensional Ising model, with $H = 0$, predicts a second-order phase transition at $T_c = 2d$ with critical exponent $1/2$. Let's wrap up this section with some quick comments on the reliability of the mean-field approximation and the accuracy of these results it has given us.

The accuracy of the mean-field results turns out to depend on the number of dimensions. For the one-dimensional Ising model that Ising himself solved, there is no phase transition at all, as we will derive in the next section. In other words, the mean-field approximation simply fails for $d = 1$.

The situation improves for the two-dimensional Ising model. Onsager's exact $H = 0$ solution features a second-order phase transition, at an inverse critical temperature $\beta_c = \frac{1}{2} \log(1 + \sqrt{2}) \approx 0.44$ that had been exactly determined a few years before his work. For $T \lesssim T_c$, the magnetization vanishes as $\langle m \rangle \propto (T_c - T)^{1/8}$, corresponding to a critical exponent $1/8$. While the mean-field prediction of a second-order phase transition is now qualitatively correct, at a quantitative level its predicted $\beta_c = \frac{1}{2d} = 0.25$ is off by almost a factor of 2, while the mean-field critical exponent $b = 1/2$ is four times larger than the true $b = 1/8$.

For higher dimensions $d \geq 3$ there is no known exact solution for the Ising model, but the existence of a second-order phase transition can be established and the corresponding critical temperature and critical exponents can be computed numerically, as we will discuss in Unit 10. In three dimensions the mean-field $T_c = 2d = 6$ and $b = 1/2$ are still significantly different from the true $T_c \approx 4.5$ and $b \approx 0.32$. The mean-field prediction for the critical exponent $b = 1/2$ turns out to be correct for $d \geq 4$, while the critical temperature $T_c = 2d$ gradually approaches the true value as the number of dimensions increases. Numerical computations find $T_c \approx 6.7, 8.8, 10.8, 12.9$ and 14.9 for $d = 4, 5, 6, 7$ and 8 , respectively, so that the mean-field result improves from being $\sim 19\%$ too high for $d = 4$ to only $\sim 7\%$ too high for $d = 8$. Formally, the mean-field approximation exactly reproduces the Ising model in the abstract limit of infinite dimensions, $d \rightarrow \infty$. Roughly speaking, the greater reliability of the mean-field approach in higher dimensions is due to the larger number of nearest neighbours for each site, $2d$. The larger number of nearest-neighbour spins produces a more reliable approximation of the mean spin in the effective field seen by each site in the mean-field approach.

9.4 Exact results for the Ising model

If time permits, it is not too hard to prove some of the exact results mentioned above, for the Ising model in one and two dimensions where the mean-field approximation is least reliable.

9.4.1 One-dimensional partition function and magnetization

The special property of the one-dimensional Ising model that helps us derive a closed-form expression for its partition function is the fact that it has exactly as many links as it has sites. Looking back to the illustration on page 130, we can rewrite the

nearest-neighbour interaction term as

$$\sum_{(jk)} s_j s_k = \sum_{n=1}^N s_n s_{n+1},$$

where the periodic boundary conditions identify $s_{N+1} = s_1$. If we also rewrite the magnetic-field term as $H \sum_{n=1}^N s_n = \frac{H}{2} \sum_{n=1}^N (s_n + s_{n+1})$, then the full internal energy is

$$E_i = - \sum_{n=1}^N \left[s_n s_{n+1} + \frac{H}{2} (s_n + s_{n+1}) \right].$$

Inserting this into the partition function $Z(\beta, N, H) = \sum_{\{s_n\}} \exp[-\beta E(s_n)]$, we can convert the exponential of the sum into a product of exponentials,

$$Z = \sum_{s_1=\pm 1} \cdots \sum_{s_N=\pm 1} \prod_{n=1}^N \exp \left[\beta s_n s_{n+1} + \frac{\beta H}{2} (s_n + s_{n+1}) \right].$$

Similarly to Eq. 41 for the non-interacting case, we are going to distribute the summations. Now, however, we have to keep track of the fact that a given spin s_j will appear both when $n = j$ and when $n + 1 = j$, and for each spin configuration it must have the same value both times it appears. An elegant way to account for all allowed possibilities is through matrix multiplication. Use the 2×2 matrix

$$T_n = \begin{pmatrix} e^{\beta+\beta H} & e^{-\beta} \\ e^{-\beta} & e^{\beta-\beta H} \end{pmatrix} \quad (119)$$

to collect the exponential factors for the four possibilities

$$\{s_n, s_{n+1}\} = \begin{pmatrix} \{1, 1\} & \{1, -1\} \\ \{-1, 1\} & \{-1, -1\} \end{pmatrix}.$$

The matrix product $T_n \cdot T_{n+1}$ then provides the sum over all contributions with consistent values for s_{n+1} . Repeating this for all terms in $\prod_{n=1}^N$, the periodic boundary conditions produce the (cyclic) trace, making the exact partition function simply

$$Z = \text{Tr} \left[\prod_{n=1}^N T_n \right].$$

What's more, since $T_n \equiv T$ is actually independent of n , this simplifies further to

$$Z = \text{Tr} [T^N]. \quad (120)$$

T is known as the **transfer matrix** — roughly speaking, it ‘transfers’ information about the values of the spins from one link to the next. At this point we can appreciate that our earlier rewriting of the magnetic-field term in the energy just helped to make T more symmetric. If we now diagonalize

$$T = \begin{pmatrix} e^{\beta} e^{\beta H} & e^{-\beta} \\ e^{-\beta} & e^{\beta} e^{-\beta H} \end{pmatrix} \longrightarrow \begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix},$$

the partition function will become simply

$$Z = \text{Tr} \left[\begin{pmatrix} \lambda_+ & 0 \\ 0 & \lambda_- \end{pmatrix}^N \right] = \text{Tr} \begin{pmatrix} \lambda_+^N & 0 \\ 0 & \lambda_-^N \end{pmatrix} = \lambda_+^N + \lambda_-^N.$$

What are the two eigenvalues λ_{\pm} of T ?

With $\beta \geq 0$ and $H \geq 0$, we can check that both eigenvalues are real and $\lambda_+ > \lambda_-$. For asymptotically high temperatures, $\beta \rightarrow 0$, the eigenvalues reduce to $\lambda_+ = 2$ and $\lambda_- = 0$ independent of H . In the special case $H = 0$, the eigenvalues are $\lambda_+ = 2 \cosh \beta$ and $\lambda_- = 2 \sinh \beta$, while $H > 0$ typically produces $\lambda_+ \gg \lambda_-$. Because $\lambda_-/\lambda_+ < 1$, for sufficiently large $N \gg 1$ we can further simplify

$$Z = \lambda_+^N \left[1 + \left(\frac{\lambda_-}{\lambda_+} \right)^N \right] \approx \lambda_+^N = e^{N\beta} \cosh^N(\beta H) \left[1 + \sqrt{1 - \frac{2 \sinh(2\beta)}{e^{2\beta} \cosh^2(\beta H)}} \right]^N.$$

So there we have it — the solution of the Ising model in one dimension. As usual, the partition function Z is not so revelatory in and of itself. Its principal value lies in enabling us to predict observables like the magnetization — so let's do that, returning to the zero-field case for which we computed the mean-field critical temperature and critical exponent in the previous section:

Note that we set $H = 0$ only *after* computing the derivative of the free energy, which allows us to consider $\langle m \rangle$ rather than $\langle |m| \rangle$. Upon setting $H = 0$, something remarkable happens: $\langle m \rangle = 0$ for all temperatures!

As claimed at the end of the previous section, the zero-field one-dimensional Ising model has no phase transition at all. It is always in the disordered phase, even in the limit of absolute zero $T \rightarrow 0$. In addition to showing that the mean-field approximation fails in one dimension, this result also reveals how the balance of energy level degeneracy vs. Boltzmann factor considered in Section 9.2 depends on the lattice structure.

Specifically, for the case of a one-dimensional lattice, the first excited energy level with energy E_1 includes more micro-states than the $2N$ we get by flipping a single spin

to oppose the full alignment of the ground state. Suppose we start from the ground state and flip spin s_j to reach the first excited energy level. Relative to the ground-state energy $E_0 = -N$, the energy of this micro-state is increased to $E_1 = -N + 4$ due to the positive contributions from the $s_{j-1}s_j$ and $s_j s_{j+1}$ links. But if we now consider *also* flipping spin s_{j+1} , the $s_j s_{j+1}$ link goes back to providing a negative contribution while the positive contribution shifts to the $s_{j+1}s_{j+2}$ link. This gives us an additional $2N$ micro-states featuring a flipped nearest-neighbour *pair* of spins, with the same energy E_1 but a smaller magnetization $|m| = 1 - \frac{4}{N}$. And we can continue this process, finding more degenerate micro-states with a flipped block of *any* number of neighbouring spins up to $N - 1$, and hence any magnetization, including $m = 0$.

The non-existence of an ordered phase in one dimension also holds for more general interacting systems beyond just the Ising model. A useful way to analyse this sort of behaviour is to describe all of the micro-states in the first excited energy level as consisting of two *domains* separated by two **domain walls** (recalling the periodic boundary conditions). For the Ising model, one domain will contain only up spins while the other will contain only down spins. The two domain walls are able to move freely through the lattice without changing the energy, but as the domain walls move the magnetization samples the full range of values $-(1 - \frac{2}{N}) \leq m \leq 1 - \frac{2}{N}$.

9.4.2 Two-dimensional critical temperature

While the zero-field Ising model on the $d = 2$ square lattice has also been exactly solved, both Onsager's original calculation and subsequent re-derivations using simpler techniques are much too complicated to cover here. However, a few years before Onsager published his famous result, [Hans Kramers](#) and [Gregory Wannier](#) were able to determine the exact $d = 2$ critical temperature in 1941. They did this by identifying a relation between two Ising model partition functions, without actually evaluating either sum over micro-states:

$$\frac{Z(\beta)}{2^N \cosh^{2N} \beta} = \frac{Z(\tilde{\beta})}{2e^{2N\tilde{\beta}}}, \quad (121)$$

where the two inverse temperatures β and $\tilde{\beta}$ are related by

$$\sinh(2\beta) = \frac{1}{\sinh(2\tilde{\beta})}. \quad (122)$$

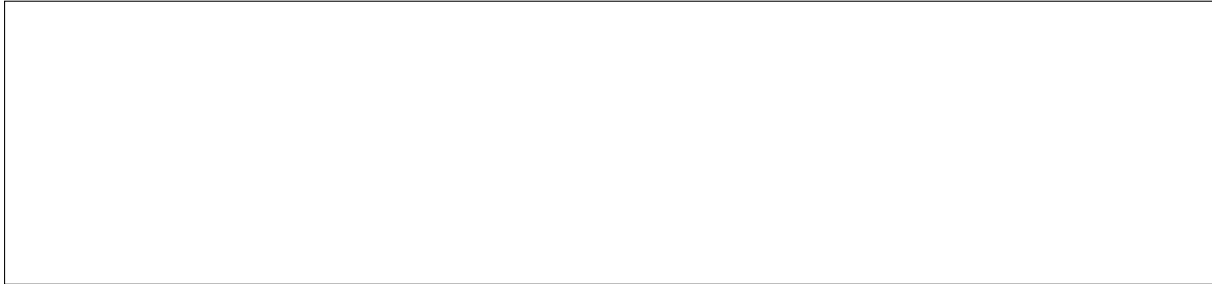
This relation is now known as [Kramers–Wannier duality](#), and the general concept of **duality** has become a powerful tool in modern theoretical physics. Note that small β implies large $\tilde{\beta}$ and vice versa — the duality relates one $d = 2$ Ising model at a high temperature to another one at a low temperature.

Although it can be useful to explicitly compare such high- and low-temperature partition functions, by computing series expansions as we did for the non-interacting spin system in Section 3.4 and for the Einstein solid in tutorial exercises, we'll skip that exercise, to keep this section from becoming too long. Those who are interested can find related discussions in Sections 5.3.2 and 5.3.3 of David Tong's [Lectures on Statistical Physics](#). Some of the manipulations below, which may seem to come out of thin air, can be motivated by considering these expansions.

The first manipulation is to express the zero-field partition function as

$$Z = \sum_{\{s_n\}} \exp \left[\beta \sum_{(jk)} s_j s_k \right] = \sum_{\{s_n\}} \prod_{(jk)} \exp [\beta s_j s_k] = \sum_{\{s_n\}} \prod_{(jk)} [\cosh \beta + s_j s_k \sinh \beta],$$

which relies on the fact $s_j s_k = \pm 1$ for the Ising model. It's easy to check the relation $e^{\beta s_j s_k} = \cosh \beta + s_j s_k \sinh \beta$ for both $s_j s_k = 1$ and $s_j s_k = -1$:



Next, we write the sum over the cosh and sinh as a summation,

$$Z = \sum_{\{s_n\}} \prod_{(jk)} [\cosh \beta + s_j s_k \sinh \beta] = \sum_{\{s_n\}} \prod_{(jk)} \sum_{p_{jk}=0,1} C_{p_{jk}}(\beta) (s_j s_k)^{p_{jk}},$$

by defining $C_0(\beta) \equiv \cosh \beta$ and $C_1(\beta) \equiv \sinh \beta$ while raising $s_j s_k$ to the corresponding power p_{jk} . Recall that the nearest-neighbour pairs (jk) correspond to all $2N$ links in the $d = 2$ lattice. To make the language a little less awkward, we can say that $p_{jk} = 1$ corresponds to link jk being 'on' while $p_{jk} = 0$ when it is turned 'off'. The product and sum above account for all possible configurations of links that are turned on and off, which we can more conveniently represent as another configuration sum,

$$\sum_{\{p\}} \equiv \sum_{p_1=0,1} \cdots \sum_{p_{2N}=0,1} .$$

Our partition function is now written in terms of two configuration sums, but we can expect only one to be necessary. We will aim to eliminate the sum over all spin configurations, by pulling out everything independent of the spins:

$$Z = \sum_{\{s_n\}} \sum_{\{p\}} \prod_{(jk)} C_{p_{jk}}(\beta) (s_j s_k)^{p_{jk}} = \sum_{\{p\}} \left[\prod_{(jk)} C_{p_{jk}}(\beta) \right] \sum_{\{s_n\}} \left[\prod_{(jk)} s_j^{p_{jk}} s_k^{p_{jk}} \right].$$

The final factor can be converted from a product over links to a product over sites. For $d = 2$, any spin s_n will appear four times in the product, once for each of the four links connected to it. The product of these four factors can be rewritten as

$$\prod_{(nk)} s_n^{p_{nk}} = s_n^{P_n}, \quad \text{defining } P_n \equiv \sum_{(nk)} p_{nk}.$$

We are now left with a product over individual s_n :

$$Z = \sum_{\{p\}} \left[\prod_{(jk)} C_{p_{jk}}(\beta) \right] \sum_{\{s_n\}} \left[\prod_{n=1}^N s_n^{P_n} \right].$$

Although each s_n is raised to a power that depends on p_{nk} for all four of its links to its nearest neighbours, consider what happens when we sum over the two values $s_n = \pm 1$. There are two possibilities: If P_n is odd, then the $s_n = 1$ and $s_n = -1$ contributions cancel — making the entire product over sites vanish! Otherwise, if P_n is even, they add up to give twice the product over the other sites. Repeating for all sites, the only non-zero contribution must have N factors of 2:

$$Z = \sum_{\{p\}} \prod_{\{jk\}} C_{p_{jk}}(\beta) \prod_{n=1}^N 2\delta_2(P_n) = 2^N \sum_{\{p\}} \prod_{\{jk\}} C_{p_{jk}}(\beta) \prod_{n=1}^N \delta_2(\sum_{(nk)} p_{nk}), \quad (123)$$

where the ‘mod-2’ Kronecker delta $\delta_2(P_n)$ vanishes if P_n is odd and equals one if P_n is even.

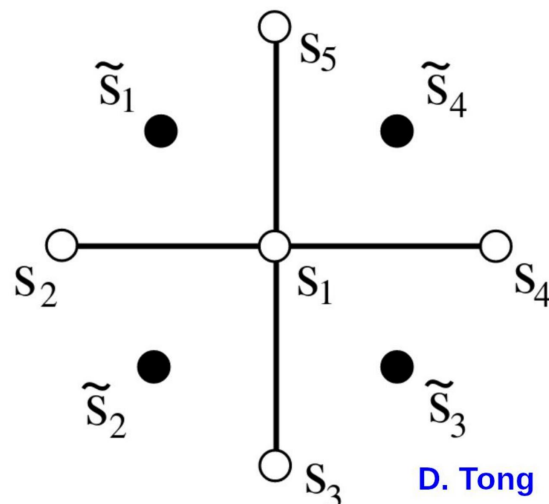
We have successfully eliminated the spin configuration sum $\{s_n\}$ in Eq. 123, which no longer explicitly refers to our original degrees of freedom at all! In its place we now have a configuration sum over all on-or-off links — $p_{jk} = 1$ or 0, respectively. And there is a very tricky set of N inter-dependent constraints coming from the product of δ_2 factors, which require that an even number of links be turned on at every lattice site, in order to get a non-zero contribution to the partition function. This is a sign that our new variables p_{jk} aren’t entirely independent of each other — which makes sense, since we have $2N$ of them, but started off with only N degrees of freedom.

What we can do about these constraints? Well, the constraints arose as a result of eliminating a configuration sum over $s_n = \pm 1$, so it is at least plausible that we can unwind them by introducing a *different* set of spin variables, $\tilde{s}_n = \pm 1$. We want the ± 1 values of \tilde{s}_n to be related to the $\{0, 1\}$ values of p_{jk} , which we can achieve by implicitly defining \tilde{s}_n via

$$p_{12} = \frac{1 - \tilde{s}_1 \tilde{s}_2}{2} \qquad p_{13} = \frac{1 - \tilde{s}_2 \tilde{s}_3}{2}$$

$$p_{14} = \frac{1 - \tilde{s}_3 \tilde{s}_4}{2} \qquad p_{15} = \frac{1 - \tilde{s}_1 \tilde{s}_4}{2}$$

and so on for all $2N$ links. A convenient way to keep track of the subscripts above is to identify these \tilde{s}_n with the **dual lattice** drawn on the next page. Each \tilde{s}_n is identified with one of the N plaquettes of the original lattice, and pairs \tilde{s}_a and \tilde{s}_b determine p_{jk} for the link passing between them.



This provides another way to see that p_{jk} aren't completely independent variables: Both p_{12} and p_{15} depend on \tilde{s}_1 , both p_{12} and p_{13} depend on \tilde{s}_2 , and so on. Delightfully, these patterns of dependence are precisely what we need to handle the δ_2 factors that are still in our partition function. If we consider

$$\begin{aligned} P_1 &= p_{12} + p_{13} + p_{14} + p_{15} = 2 - \frac{\tilde{s}_1\tilde{s}_2 + \tilde{s}_2\tilde{s}_3 + \tilde{s}_3\tilde{s}_4 + \tilde{s}_1\tilde{s}_4}{2} \\ &= 2 - \frac{(\tilde{s}_1 + \tilde{s}_3)(\tilde{s}_2 + \tilde{s}_4)}{2} \in \{0, 2, 4\}, \end{aligned}$$

we see that working in terms of \tilde{s}_n automatically turns on an even number of links at every site, producing $\prod_{n=1}^N \delta_2(P_n) = 1!$

In order to replace $\sum_{\{p\}}$ by $\sum_{\{\tilde{s}\}}$, we need to confirm that the latter accounts for all possible $P_n = \{0, 2, 4\}$. It turns out that the dual-spin configuration sum provides all values of P_n twice, which we can check for the simple case $P_1 = 4$:



This generalizes to the full N -spin system: Negating all $\tilde{s}_n \rightarrow -\tilde{s}_n$ leaves all P_n unchanged.²⁰ Therefore converting the partition function from Eq. 123 into a configuration sum over $\{\tilde{s}\}$ gives us

$$Z = \frac{1}{2} 2^N \sum_{\{\tilde{s}\}} \prod_{(jk)} C_{(1-\tilde{s}_j\tilde{s}_k)/2}(\beta),$$

where (jk) now refers to the dual lattice.

The final step is to express $C_0(\beta) = \cosh \beta$ and $C_1(\beta) = \sinh \beta$ in terms of the dual variables \tilde{s}_n that we're now working with. The trick here is to write

$$C_p(\beta) = (\cosh \beta) \exp [p \log \tanh \beta] = (\cosh \beta) \exp \left[\frac{1 - \tilde{s}_j\tilde{s}_k}{2} \log \tanh \beta \right],$$

substituting $p = (1 - \tilde{s}_j\tilde{s}_k)/2$. Breaking up the exponential gives us

$$C_p(\beta) = (\cosh \beta \sinh \beta)^{1/2} \exp \left[-\frac{1}{2} \tilde{s}_j\tilde{s}_k \log \tanh \beta \right].$$

Inserting this into the partition function, the product over all links just provides $2N$ factors of the \tilde{s}_n -independent first term, and we can then convert the product of exponentials into an exponential of the sum, producing

$$Z = \frac{1}{2} (2 \cosh \beta \sinh \beta)^N \sum_{\{\tilde{s}\}} \exp \left[-\frac{\log \tanh \beta}{2} \sum_{(jk)} \tilde{s}_j\tilde{s}_k \right].$$

²⁰This is discussed in more detail by Robert Savit, "Duality in field theory and statistical systems", *Reviews of Modern Physics* 52 (1980), 453.

By defining $\tilde{\beta} \equiv -\frac{1}{2} \log \tanh \beta$, we can recognize the sum over $\{\tilde{s}\}$ configurations as simply a zero-field Ising model partition function $Z(\tilde{\beta})$, Eq. 112. We can also confirm that this definition of $\tilde{\beta}$ is equivalent to Eq. 122:



Using similar manipulations, we can express the spin-independent prefactor in terms of either β or $\tilde{\beta}$,

$$2 \cosh \beta \sinh \beta = \sinh(2\beta) = \frac{1}{\sinh(2\tilde{\beta})},$$

or in the mixed form that reproduces Eq. 121:

$$2 \cosh \beta \sinh \beta = 2 \cosh^2 \beta \tanh \beta = \frac{2 \cosh^2 \beta}{e^{2\tilde{\beta}}} \implies \frac{Z(\beta)}{(2 \cosh^2 \beta)^N} = \frac{Z(\tilde{\beta})}{2e^{2N\tilde{\beta}}}.$$

We have successfully derived Kramers–Wannier duality! Now let’s briefly interpret what it means. It’s a worthwhile **exercise** to show that multiplying a partition function by an overall spin-independent factor, $Z(\beta) \rightarrow c(\beta)Z(\beta)$, has no effect on expectation values. (Try it!) Therefore the relation above identifies a $d = 2$ Ising system at temperature $1/\beta$ with another such system at temperature $1/\tilde{\beta}$, where small β corresponds to large $\tilde{\beta}$ and vice versa.

This does *not* mean that $d = 2$ Ising model behaves the same at low and high temperatures. Indeed, we saw already in Section 9.2 that it changes between qualitatively different ordered and disordered phases in these two regimes. What Kramers–Wannier duality is telling us is that the ordered phase of Ising spins in two dimensions, characterized by their order parameter, is secretly equivalent to the disordered phase of the *dual* spins — a different set of degrees of freedom, which can be characterized by a ‘disorder parameter’. Similarly, the disordered phase of the original system maps onto the ordered phase of the dual system.

If we assume there is a single phase transition where the ordered and disordered phases coincide (which [Rudolf Peierls](#) had established in 1936), then Kramers–Wannier duality implies this must occur when $\beta = \tilde{\beta}$. In other words,

$$\sinh^2(2\beta_c) = 1 \implies \beta_c = \frac{1}{2} \operatorname{arcsinh}(1) = \frac{\log(1 + \sqrt{2})}{2} = 0.440686 \dots,$$

recalling (or looking up) that $\operatorname{arcsinh}(x) = \log(x + \sqrt{x^2 + 1})$. Because this exact critical temperature $T_c = 2/\log(1 + \sqrt{2}) = 2.269185 \dots$ was predicted three years before Onsager analytically solved the $d = 2$ Ising model, the fact that his solution reproduced it was an important check of correctness.

As mentioned at the start of this subsection, dualities of this sort are central to theoretical physics in the 21st century. In general, these dualities are much more complicated than Kramers–Wannier duality, in two main ways. First, the dual degrees of freedom are typically different — and have different interactions — than the original degrees of freedom. For example, if we were to carry out a similar analysis of the three-dimensional Ising model, we would find that the dual system is *not* just another Ising model. The $d = 2$ Ising model is a special case of a **self-dual** system, which we exploited to determine T_c .

Second, the Ising model is special in that we were able to explicitly derive the duality it exhibits, which is typically not (yet) possible. Instead, most dualities have to be conjectured and then checked by subjecting them to as many tests as possible. For example, this is the case for *holographic dualities* that are conjectured to relate certain theories of quantum gravity to non-gravitational quantum systems that can be much easier to analyse mathematically. As an aside, the fully connected lattice encountered in a tutorial also makes an appearance in holography — the behaviour of interacting fermions on this fully connected lattice (known as the [SYK model](#), named after Subir Sachdev, Jinwu Ye and Alexei Kitaev) is conjectured to be dual to the gravitational dynamics of quantum black holes. [More than 1500](#) scientific studies related to the SYK model and its conjectured holographic duality have been published since 2016!

Unit 10: Synthesis and broader applications

10.1 Monte Carlo importance sampling

Although we have now seen how to derive some exact results for the Ising model in one and two dimensions, it's worth recalling that for $3 \leq d < \infty$ no exact solution is known even for this simple system. In general, interacting statistical systems are not exactly solvable. In order to explore their broad applications throughout the mathematical sciences and beyond, we therefore need to analyse them either through systematic approximation schemes (such as *perturbation theory*) or by numerical computations. Numerical approaches have become increasingly important over the past fifty years, and in this section we'll outline the general methods they employ.

Our goal is to compute expectation values that may be experimentally observed, which are formally defined by sums over all micro-states. Considering the canonical ensemble for simplicity,

$$\langle \mathcal{O} \rangle = \sum_{i=1}^M \mathcal{O}_i p_i = \frac{1}{Z} \sum_{i=1}^M \mathcal{O}_i e^{-\beta E_i} = \frac{\sum_{i=1}^M \mathcal{O}_i e^{-\beta E_i}}{\sum_{i=1}^M e^{-\beta E_i}}.$$

We already saw, at the end of Section 9.1, that enormous computational resources would be required to explicitly carry out such sums over micro-states. Even for tiny Ising systems with $N \sim 1000$ spins, the largest existing or foreseeable supercomputers would have to run for far longer than the age of the universe in order to evaluate the roughly $2^{1000} \sim 10^{300}$ terms in the partition function. To quantify 'tiny', consider that $N \sim 1000$ would correspond to a $10 \times 10 \times 10$ lattice in three dimensions or a $6 \times 6 \times 6 \times 6$ lattice in four dimensions, both very far from the $N \rightarrow \infty$ thermodynamic limit of interest for phase transitions.

And yet, at the end of Section 9.3 we were able to quote numerical results for the Ising model critical temperature for $3 \leq d \leq 8$, along with a $d = 3$ critical exponent. These results are obtainable because practical numerical computations do not perform a 'brute-force' evaluation of every single micro-state. Instead, they proceed by (pseudo-)randomly **sampling** a very small subset of those micro-states, and using this subset to compute results for the average energy, magnetization, and other thermodynamic quantities. As long as this sampling is done appropriately, without *bias*, the law of large numbers allows us to treat these averages as controlled approximations to the true ensemble expectation values.

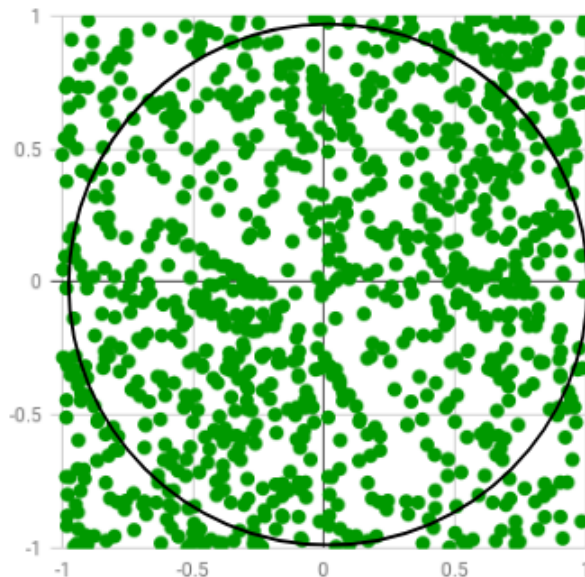
As we saw in the computer assignment, such numerical calculations employ pseudo-random numbers rather than complete randomness, which allows them to be reproducible up to very high precision by different people using different computers. Due to the role of randomness, these numerical approaches have become known as **Monte Carlo** methods — a whimsical reference to the famous gambling centre in Monaco. Monte Carlo methods are crucial in statistical mechanics, and related disciplines, because they are very broadly applicable to interacting systems that no longer benefit from dramatic simplifications through factorization.

We can gain some intuition about how Monte Carlo methods work by using such pseudo-random sampling to numerically evaluate a simple integral. The idea is that

the integral can be numerically approximated by evaluating its integrand at randomly sampled points in the integration domain, and normalizing by the number of samples. An amusing example is to compute

$$\pi = \int_{-1}^1 dx \int_{-1}^1 dy H(1 - \{x^2 + y^2\}) \quad H(r) = \begin{cases} 1 & \text{for } r \geq 0 \\ 0 & \text{for } r < 0 \end{cases},$$

where the Heaviside step function $H(r)$ picks out a disk with radius $R = 1$ in a square integration domain with area 4, as shown below. Since the integrand is either 0 or 1 for each randomly sampled point in that domain, simply counting the fraction of the S samples that lie in the disk provides an unbiased numerical determination of π , with a *statistical uncertainty* that vanishes $\propto 1/\sqrt{S}$. In just a few minutes, [this Python code](#) predicts $\pi = 3.14163 \pm 0.00022$ purely from sampling pseudo-random numbers.



Of course, numerically computing π can be done far more efficiently with other, more specialized, techniques. Monte Carlo integration is most useful when we need to consider very high-dimensional integrals — such as canonical partition functions of interacting statistical systems, interpreted as N -dimensional integrals over the system's N degrees of freedom. To illustrate the scale of computations that can currently be carried out, ongoing theoretical physics research here in Liverpool routinely uses Monte Carlo methods to numerically evaluate roughly billion-dimensional integrals.

At this point, you should be concerned that such sampling can account for only an extremely small fraction of the possible micro-states for the systems under consideration, suggesting a risk of inaccurate results from insufficient sampling. This is a new manifestation of the obstacle we encountered when considering brute-force computations above. If the brute-force evaluation of every single micro-state takes far longer than the age of the universe, then the fraction we could sample in days — or even years — is almost vanishingly small.

As a concrete example, if we generously suppose our computer only needs a few nanoseconds to sample a micro-state of a tiny $N \sim 1000$ Ising system, over the course of a day it would sample roughly ten trillion (10^{13}) spin configurations — only about one part in 10^{287} of the total $2^N \sim 10^{300}$ micro-states. To make the situation even worse, as N increases the number of possible Ising model micro-states grows *exponentially* quickly, $\sim 2^N$, in addition to the more modest growth in the amount of computing required to sample each micro-state. For illustration, [this 2015 publication](#) numerically predicting T_c for the Ising model in $d = 5, 6$ and 7 dimensions includes calculations up to $N = 64^5 \approx 10^9$. Out of the roughly $2^{10^9} \sim 10^{323,000,000}$ micro-states for this systems, only $\sim 10^4$ could be sampled in a reasonable amount of time. How much trust should we place in results from such numerical work?

Thinking back to our consideration of the ordered and disordered phases of the Ising model in Section 9.2, we can argue that this should be all right in the high-temperature disordered phase. In the infinite-temperature limit, all the micro-states become equally probable, and observable expectation values are determined by the degeneracies of the different energy levels. Random sampling is more likely to account for the dominant energy levels with large degeneracies, making it plausible that reasonable results could be obtained even by averaging over such a tiny fraction of the total number of micro-states.

In the low-temperature ordered phase, however, the opposite occurs. As the temperature decreases, the large-scale behaviour of the system in this phase is dominated by a very small number of micro-states. For sufficiently low temperatures, observable expectation values are effectively determined by the two degenerate minimum-energy micro-states with all spins aligned either up or down. Only exponentially suppressed corrections would then be introduced by higher-energy excited states. As there is essentially no chance of randomly sampling either of those two minimum-energy micro-states, the approach described above seems doomed to fail.

A key breakthrough that made numerical results truly reliable was the invention of stochastic procedures to sample any micro-state ω_i with a probability proportional to its Boltzmann factor, $p_i \propto e^{-\beta E_i}$. Such automated procedures are known as *algorithms* (a term that evolved from the name of [Muhammad ibn Musa al-Khwarizmi](#)), and the overall approach is called **importance sampling**, since it preferentially samples the important micro-states that make the most significant contributions to the partition function and expectation values. Assuming we have such an algorithm, applying it to the $\beta \rightarrow \infty$ low-temperature phase considered above would produce an exponentially enhanced probability of sampling low-energy micro-states, as desired. As $\beta \rightarrow 0$ in the high-temperature phase, there would be little change compared to the naive random sampling considered above, since all micro-states would become equally probable.

The challenge is to design importance sampling algorithms in the first place. In particular, these algorithms can't rely on knowing the full set of micro-state energies E_i , and the corresponding probabilities p_i , since enumerating this information would be equivalent to brute-force computation of the full partition function. In essence, the algorithm has to exploit its stochastic nature — its use of pseudo-random numbers — to guide the sampling to important, high-probability micro-states. And this guiding needs to be done without introducing any other bias that might cause distorted results

to be obtained from averaging over the tiny fraction of micro-states that can be sampled in a reasonable amount of time.

A famous solution to this challenge was developed in 1953 by Nick Metropolis, Arianna Rosenbluth, Marshall Rosenbluth, Mici Teller and Edward Teller. I will call this solution the MRRTT algorithm.²¹ It relies on the concept of Markov chains that we briefly discussed when considering random walks all the way back in Section 1.5. To reiterate the essence of this concept, a Markov chain is a process in which the next micro-state to be sampled is pseudo-randomly chosen based on the micro-state currently under consideration, with no ‘memory’ of any other micro-states that may previously have been sampled.

We can use the Ising model to illustrate how the MRRTT algorithm employs Markov chains to sample micro-states proportionally to their importance. Any spin configuration can serve equally well as the initial micro-state at the start of the chain. Starting from this initial configuration, we pseudo-randomly select one spin, s_j , and compute ΔE_j , the change in the system’s energy that would be caused by flipping $s_j \rightarrow -s_j$. We then update the spin configuration by ‘accepting’ this flip with probability

$$P_{\text{accept}} = \min \{1, e^{-\beta \Delta E_j}\}, \quad (124)$$

which defines the next micro-state in the Markov chain. Importantly, this new micro-state may be identical to the previous micro-state — this occurs with probability $P_{\text{reject}} = 1 - P_{\text{accept}}$. The final step of the algorithm is to repeat this single-spin update procedure as many times as our computers can handle.

We can appreciate why micro-states may need to be repeated in the Markov chain by considering what should happen if we were to sample the ground state at low temperatures. In this regime, the ground state should dominate, so the algorithm should sample it repeatedly, proportionally to its probability.

Digging into Eq. 124, we can see that any spin flip that lowers the energy will always be accepted, since $\Delta E < 0 \implies e^{-\beta \Delta E} > 1$. The MRRTT algorithm is therefore free to approach the minimum-energy ground state of the system. If it is in the ground state, then any spin flip will increase the energy, $\Delta E > 0$, and will only be accepted with an exponentially suppressed probability $e^{-\beta \Delta E} \rightarrow 0$ as $T = 1/\beta \rightarrow 0$, as desired. More generally, if we consider two micro-states ω_A and ω_B , then the relative probabilities of moving between these two micro-states are

$$\frac{P(A \rightarrow B)}{P(B \rightarrow A)} = \frac{\min \{1, e^{-\beta(E_B - E_A)}\}}{\min \{1, e^{-\beta(E_A - E_B)}\}} = e^{-\beta(E_B - E_A)} = \frac{e^{-\beta E_B}}{e^{-\beta E_A}}, \quad (125)$$

regardless of whether $E_A \leq E_B$ or $E_B \leq E_A$.

So long as every micro-state can be generated from any other micro-state by making a series of pseudo-random changes to individual degrees of freedom, Eq. 125

²¹In an infamous misfiring of alphabetical ordering, this remains widely known as the “Metropolis algorithm” even though Metropolis’s role was providing specialized computing equipment rather than developing the algorithm itself. In addition, the key contributions of Arianna Rosenbluth and Mici Teller were widely under-appreciated for many years.

ensures that the micro-states ω_i will indeed be sampled with probabilities proportional to the Boltzmann factors $e^{-\beta E_i}$ that quantify their importance. The necessary condition that the Markov chain can reach every micro-state (at least in principle) is called **ergodicity**. Because the micro-state probabilities p_i are effectively ‘felt out’ through the accept/reject test described above, in non-ergodic situations the algorithm will fail to account for the probabilities of micro-states that can’t be reached by the Markov chain. This can easily lead to incorrect results.

You can read more about the MRRTT algorithm in Section 8.2 of Dan Schroeder’s *Introduction to Thermal Physics*, which provides a single-page annotated code implementing it for the two-dimensional Ising model. While the conceptually simple MRRTT algorithm is the most famous means to carry out Markov-chain Monte Carlo importance sampling, it is far from the only option, and often far from the best. If time permits, we may discuss some of the challenges that make it advantageous to go beyond the MRRTT algorithm, in particular the issue of **auto-correlations**.

For now, suffice it to say that there is an enormous amount of ongoing research developing, optimizing and applying more elaborate Monte Carlo methods to investigate topics throughout the mathematical sciences and beyond. In Section 10.3 we will briefly look at some of these broader applications. First, there is another important concept to introduce, called universality, which helps to reveal why interacting statistical systems are so useful to apply to such a diverse range of scientific investigations.

10.2 Universality

In Section 9.3 we defined the critical exponent b as the non-integer power governing the behaviour of the order parameter $\langle m \rangle \propto (T_c - T)^b$ for temperatures slightly lower than the critical temperature T_c of the second-order phase transition. In addition to being an important characteristic any specific phase transition, it was discovered during the twentieth century that precisely the same critical exponents turn out to govern the behaviour of phase transitions that we initially would not have expected to resemble each other at all.

A famous example of universality involves the **liquid–gas phase transition**. We no longer have time to analyse this phase transition in full detail. It occurs only for interacting (non-ideal) gases, and after our studies of simple interacting spin systems in Unit 9 we can appreciate that such interacting gases are far more complicated than the ideal gases we considered in Units 4 and 8. For example, Section 8.2 of Schroeder’s *Introduction to Thermal Physics* considers the equation of state of a low-density interacting gas, organized as an expansion around the ideal gas law,

$$PV = NT \left[1 + B(T)\rho_{\text{gas}} + \mathcal{O}(\rho_{\text{gas}}^2) \right].$$

Here $\rho_{\text{gas}} = N/V$ is the density of the atoms in the gas, which needs to be small in order for this approach to work. The temperature-dependent function that governs the leading correction to the ideal gas law,

$$B(T) = -2\pi \int_0^\infty r^2 \left(e^{-u(r)/T} - 1 \right) dr,$$

in turn depends on the interaction energy $u(r)$ between pairs of particles separated by distance r . This function $u(r)$ can change significantly for different types of particles. You should confirm that $B(T) = 0$ in the non-interacting case $u(r) = 0$, and consider the consequences of both an *attractive* interaction with $u(r) < 0$ and a *repulsive* interaction with $u(r) > 0$. (Try it!)

Unfortunately, the restriction to low densities in this analysis prevents its application to the liquid–gas phase transition. From everyday experience we can appreciate that the liquid phase of a given set of particles (for example, water) has a higher density than the gas phase of those same particles (in this case, steam). The liquid–gas phase transition occurs at a critical temperature T_c that produces equal densities for the two phases — this is the largest density the gas can reach while still remaining a gas, rather than the low density assumed above.

An alternative approach, pioneered by [Johannes van der Waals](#) in the late 1800s (and recognized by the 1910 Nobel Prize), is to propose modifications to the equation of state based on qualitative physical arguments. The van der Waals equation of state,

$$(P + a\rho^2)(V - Nb) = NT,$$

aims to model two aspects of interacting fluids and gases, through two positive parameters a and b . First, Nb represents the volume occupied by the N particles, which is subtracted from the total available volume V . Second, the $a\rho^2$ dependence on the density represents repulsive forces that the particles are assumed to exert on each other when they are packed more closely together.

These are obviously rough, imprecise arguments, and you may not be surprised to hear that the van der Waals equation of state does not provide a very accurate description of real fluids and gases. However, as discussed in Section 5.1 of David Tong’s [Lectures on Statistical Physics](#), it does successfully predict a phase transition between the liquid and gas phases, along with a set of critical exponents characterizing this transition. One of these critical exponents concerns the behaviour of the densities,

$$\frac{1}{\rho_{\text{gas}}} - \frac{1}{\rho_{\text{liquid}}} \propto (T_c - T)^b.$$

The van der Waals equation of state predicts $b = 1/2$, while the true value is $b \approx 0.32$, for *any* liquid–gas transition (not just H₂O water/steam).

It may seem surprising that all liquids involve the same critical exponent. If we look back to the end of Section 9.3, we can be even more surprised that this $b \approx 0.32$ is the same as the critical exponent of the magnetization in the three-dimensional Ising model! This is not merely a numerical coincidence, but an example of an amazing phenomenon known as **universality**. In essence, universality states that the specific details of interacting statistical systems become irrelevant close to critical points at which phase transitions occur. It doesn’t matter whether we are considering a three-dimensional lattice of Ising spins or a liquid such as water — the behaviour of both systems is governed by the same set of critical exponents, completely specified by their *universality class*.

The detailed mathematics underlying universality is beyond the scope of this module. Important contributions to its development were made by [Michael Fisher](#), [Leo](#)

Kadanoff and Ken Wilson, among many others (with Wilson’s contributions recognized by the 1982 Nobel Prize). Suffice it to say that universality causes the same large-scale behaviour to appear in the vicinity of critical points even for systems that appear completely different. This insensitivity to the details of the system helps to explain the power of even simple interacting statistical systems such as the Ising model, and their applicability in so many different domains.

10.3 Broader applications

In this section we will quickly highlight a few ways in which the concepts and tools of statistical mechanics find application within and beyond the mathematical sciences. These discussions are just brief high-level summaries, without detailed derivations. The familiarity you have now gained with statistical mechanics can help you to uncover and understand further examples in addition to the five below.

Quantum field theory

First considering an example within the mathematical sciences, let’s elaborate on the Liverpool theoretical physics research mentioned in Section 10.1 as a major use of Monte Carlo importance sampling. This research investigates the behaviours of various quantum field theories (QFTs), with applications ranging from the strong nuclear force (explaining how protons, neutrons and other composite particles arise from the fundamental interactions of quarks and gluons), to the Higgs boson, dark matter, and even holographic gauge/gravity duality.

Qualitatively, any particular QFT involves a set of fields, $\Phi(\vec{x}, t)$, that fill all of space and time. These fields interact with each other (and with themselves), producing behaviour that is governed by the **Feynman path integral**

$$\mathcal{Z}[\Phi(\vec{x}, t)] = \int \mathcal{D}\Phi \exp \left[\frac{i}{\hbar} \int \mathcal{L}[\Phi(\vec{x}, t)] d^3x dt \right],$$

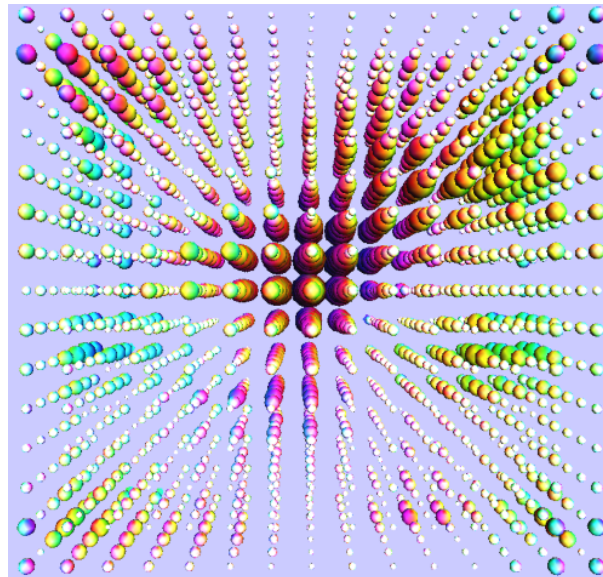
where the lagrangian $\mathcal{L}[\Phi(\vec{x}, t)]$ specifies the interactions between the degrees of freedom. Both the lagrangian and the path integral are *functionals* of the fields, which are in turn functions of space and time. The measure $\mathcal{D}\Phi$ represents integration over all configurations of all the fields at every (\vec{x}, t) .

The path integral should look tantalizingly similar to the partition function in the canonical ensemble, with \mathcal{L} appearing in place of the energy of each field configuration, and Planck’s constant \hbar appearing in place of the temperature — which we can interpret as quantum fluctuations playing a role analogous to thermal fluctuations in statistical mechanics. The main difference is the imaginary unit i in the exponential, which we can address through the trick of working in “imaginary time” $i\tau = t$. This is known as a *Wick rotation*, and gives us

$$\begin{aligned} \mathcal{Z}[\Phi(\vec{x}, \tau)] &= \int \mathcal{D}\Phi \exp \left[-\frac{1}{\hbar} \int \mathcal{L}[\Phi(\vec{x}, \tau)] d^3x d\tau \right] \\ \langle \mathcal{O} \rangle &= \frac{1}{\mathcal{Z}} \int \mathcal{D}\Phi \mathcal{O}[\Phi(\vec{x}, \tau)] \exp \left[-\frac{1}{\hbar} \int \mathcal{L}[\Phi(\vec{x}, \tau)] d^3x d\tau \right], \end{aligned}$$

which we can analyse through Monte Carlo importance sampling in much the same way as we would for an Ising model. The catch is that Wick rotation is only valid in equilibrium, limiting the phenomena that can be analysed through this approach.

Finally, there is one more complication we need to deal with. Treating space and time as continuous implies that there are infinitely many degrees of freedom for each field, making \mathcal{Z} formally an infinite-dimensional integral. We need to *regularize* this integral by replacing space and (imaginary) time with a discrete lattice of space-time points, as illustrated below.



(Image credit: Claudio Rebbi)

This approach of Wick rotating and then discretizing space and time is known as **lattice field theory**. The image above comes from a lattice field theory computation that predicted properties of protons. It visualizes a snapshot (at a single point in imaginary time) of the spatial configuration of a quark field within that proton. Each quark degree of freedom has a “colour charge” — a linear combination of {red, blue, green} — whose orientation in colour space is indicated by the colour of each point in the illustration above, while the size of each point represents the magnitude of the charge. By repeating these numerical calculations for discrete space-times with more lattice sites placed closer together, we can recover the original, continuous system by extrapolating to the *continuum limit* where there are infinitely many lattice sites infinitesimally close together. Lattice field theory is an extremely successful and broadly applicable approach, which has proven able to predict many aspects of complicated interacting QFTs with high precision and controlled uncertainties that are systematically improvable through additional computing resources or algorithmic innovations.

Machine learning and artificial intelligence

In tutorials we are exploring generalizations of the Ising model both to different lattice structures (hyper-diamond, A_d^* , kagome, etc.) as well as to different interactions

between the $s_j = \pm 1$ spins, including the general

$$E_{\text{SG}} = - \sum_{j \neq k} J_{jk} s_j s_k$$

with zero external magnetic field. From this expression we can recover the usual Ising model by setting the interaction strength $J_{jk} = \frac{1}{2}$ for nearest-neighbour pairs, otherwise $J_{jk} = 0$. Alternatively, a great deal of interesting physics and mathematics can be investigated by considering symmetric non-zero J_{jk} that connect all pairs of lattice sites (more generally called *nodes*). In tutorials, for example, we focus on the exactly solvable Curie–Weiss model defined by constant $J_{jk} = \frac{J}{2N}$ for all $j \neq k$.

As another example, if the values of J_{jk} are randomly sampled from a gaussian distribution, the resulting model is also solvable, and the 2021 Nobel Prize recognized [Giorgio Parisi](#) in part for finding this solution. This is often called the Sherrington–Kirkpatrick model, to recognize [David Sherrington](#) and [Scott Kirkpatrick](#) for developing it in 1975, and beginning to work on the solution that Parisi eventually found in 1979. More generally, this is called a [spin glass](#), in analogy to everyday glass — the latter is formally in a liquid phase without solid order, but disorder sets in so slowly that remnant solidity persists for long periods (centuries). Similarly, when a spin glass transitions from an ordered phase to a disordered phase, disorder sets in so slowly that remnant magnetization persists for long periods (days).

A third name given to this same system is to call it a [Hopfield network](#), named after [John Hopfield](#), which is an example of a recurrent neural network in which the J_{jk} are interpreted as *weights* that can be used to store and access (or “learn”) information. Similarly, adding a node-dependent external magnetic field term,

$$E_{\text{BM}} = - \sum_{j \neq k} J_{jk} s_j s_k - \sum_n H_n s_n,$$

defines another type of neural network known as a [Boltzmann machine](#), co-invented by [Geoffrey Hinton](#). Boltzmann machines have proven most useful if *restricted* so that the nodes are divided into *layers*, with non-zero weights connecting only pairs of nodes in different, neighbouring layers. Such restricted Boltzmann machines with many layers are one of the most common approaches to the type of artificial intelligence (AI) known as *deep learning*. Other forms of AI, such as the large language models that form the foundation for generative AI systems in the 2020s, can also be [analysed and interpreted](#) in terms of interacting statistical mechanics systems. Hopfield and Hinton were recognized for their work on neural networks by the 2024 Nobel Prize, and many other researchers are making major contributions in these areas. The annual [Elizabeth Gardner Lectures](#) are a good place to hear about some of these advances; they are named after [Elizabeth Gardner](#), an early leader in the field who died at the age of 30.

Voter models

Moving beyond the mathematical sciences, we can consider sociology as a domain where the applicability of statistical mechanics is less obvious. Despite this, there is a diverse and active field of “sociophysics”, which uses methods and perspectives

of statistical mechanics to describe various aspects of social and political behaviour. Serge Galam's 2012 book [Sociophysics](#) provides a comprehensive introduction. In the years since it was published, the continued growth of enormous online social networks has provided a flood of data that can be modeled through frameworks based on interacting statistical systems.

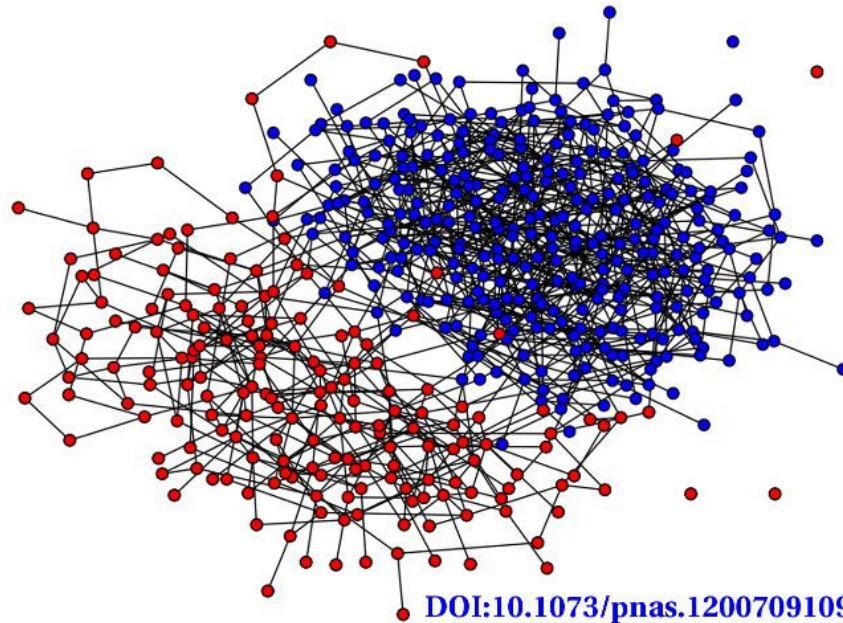
A particular branch of sociology where connections to statistical mechanics may be more apparent is the field of opinion formation, where “[voter models](#)” have been widely used since the 1970s to model elections, and more general political debates, with varying degrees of success. Voter models are interacting statistical systems not too different from the Ising model, where the spin degrees of freedom that we have been working with are reinterpreted as voters' opinions on a certain topic. For example, support for a particular proposition, candidate or political party can be represented as $s_n = +1$, with $s_n = -1$ indicating opposition. Just like the interactions in the Ising model encourage spins to align with each other, voter models incorporate a tendency for voters to align (i.e., agree) with the majority of the other voters they interact with.

There are many generalizations that can then be added to better describe the outcomes of surveys and elections. A simple extension is to allow voters to be neutral or indifferent, represented as $s_n = 0$. Similarly, the strength of a voter's commitment to their opinion can be modelled by extending the range of possible spin values,

$$s_n \in \{\dots, -2, -1, 0, +1, +2, \dots\}.$$

Voter models should also be defined on flexible and evolving *graphs*, rather than the regular lattices we considered for the Ising model, in order to capture the possibility of people interacting with different sets of individuals over time. The figure on the next page, from Richard Durrett et al., “[Graph fission in an evolving voter model](#)” (2012), illustrates how such a graph can look for a two-state voter model with $s_n \in \{-1, +1\}$ coloured red and blue. As in everyday experience (and inspired by social networks), different voters have different numbers of connections with each other. In this particular investigation, network connections between disagreeing voters are probabilistically severed, which enables a transition between two phases. In one of these phases a consensus develops among the vast majority of voters. In the other phase the population becomes increasingly polarized, as shown below. Voters with a particular opinion increasingly interact only with others who share that opinion, and eventually nearly all connections are severed between the two opposing groups.

Simpler studies, that don't allow connections to be severed, can also observe a transition to consensus once an opinion reaches a critical concentration. For example, Alexander Balankin et al., “[Ising percolation in a three-state majority vote model](#)” (2016) considers a case with a second-order phase transition in the universality class of the two-dimensional Ising model. This work also observed the possibility of a “stable non-consensus” phase, with long-term polarization between ‘clusters’ of aligned voters who interact mainly with each other rather than with voters holding different opinions. These stable clusters eventually formed even without severing connections, as an emergent consequence of the seemingly simple interactions voters had with each other.

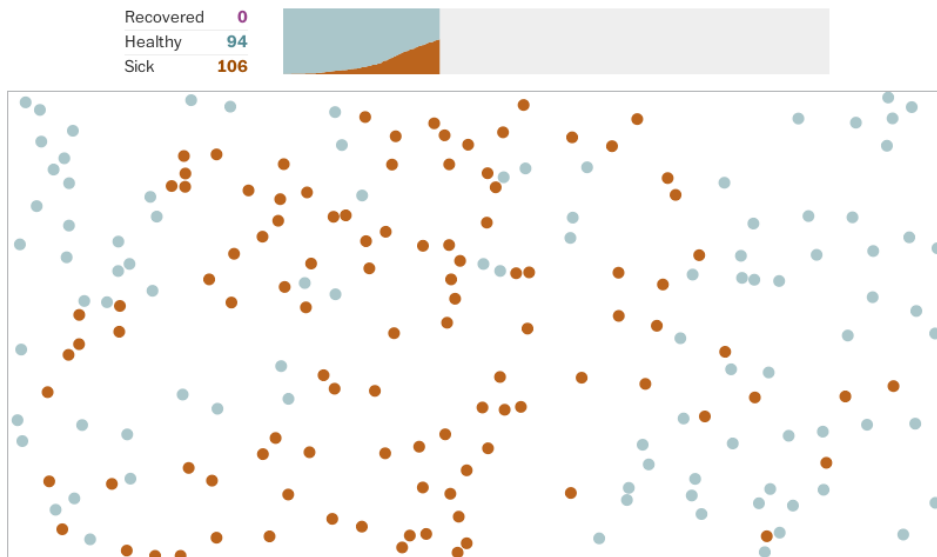


Epidemiology

Another application of interacting statistical systems, which attracted a lot of attention in the early 2020s, is to model the spread of diseases in populations. As in the case of voter models, the degrees of freedom under consideration are again individuals, whose interactions with each other can allow the infection to spread to those who are susceptible. Numerical Monte Carlo calculations can then be used to model how many people are likely to be infected as time passes, guided by data on typical movement and contacts.

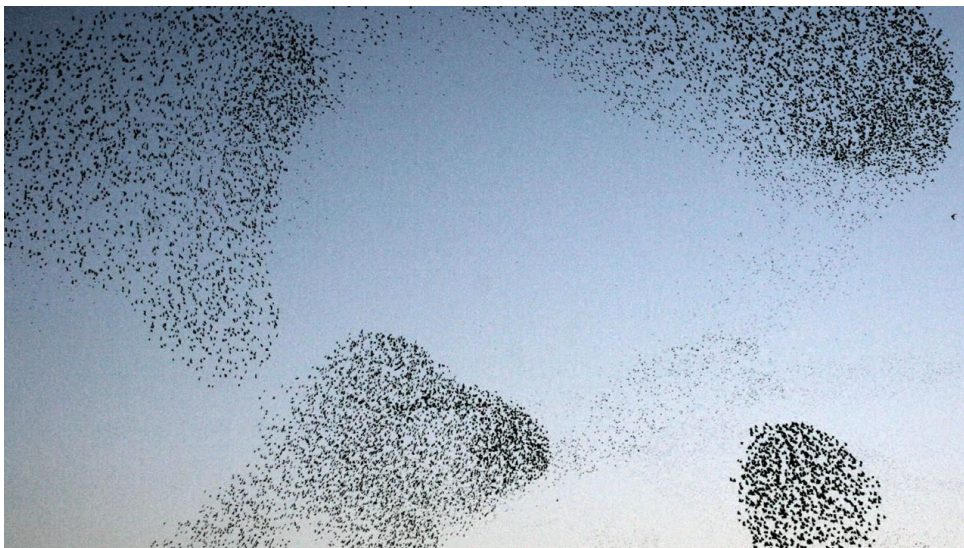
The picture on the next page is a snapshot from [an over-simplified simulation](#) provided by *The Washington Post* to illustrate these concepts. Here individuals are modeled simply as a gas of interacting particles in two dimensions, and various ways of restricting their motion are used to explore the likely effects of measures such as quarantines and social distancing. By coincidence, the scale of this simulation, which considers a population of just 200 people, is comparable to the first importance sampling Monte Carlo calculation carried out in 1953, which computed the pressure (i.e., equation of state) for 224 interacting particles in a two-dimensional ‘volume’. At the time, this required several days of computing on a state-of-the-art machine; now these sorts of calculations are easily done on a smartphone.

Larger-scale and more realistic versions of these epidemiological simulations provide important input into government deliberations regarding what restrictions (such as lockdowns) would be most beneficial to reduce the spread of disease, and how long it would be best to maintain such restrictions. Rather than investigating a phase transition, the goal is to quantify likely effects of various potential restrictions. As described in Section 1.2, the numerical experiments are therefore repeated many times with different sequences of pseudo-random numbers, to produce an ensemble of possibilities from which the likely outcomes of interventions can be inferred.



Flocking

Let's conclude these brief highlights by considering another biological application, in which interacting statistical systems are used to model the large-scale collective motion of certain groups of animals. The image below, from Marcus Woo, "[How bird flocks are like liquid helium](#)" (2014) illustrates the "flocking" behaviour of groups of starlings, which fly through the sky in surprisingly tight coordination, producing shape-shifting clouds known as [murmurations](#). Qualitatively similar behaviour is also seen in schools of fish, swarms of insects, and even [crowds of humans](#).



Many models based on interacting statistical systems have been — [and continue to be](#) — developed to describe this emergent collective behaviour. Much of this work builds upon a particularly simple "[Vicsek model](#)" introduced in 1995, which hypothesizes that each particle (i.e., bird) will interact with others near to it. These interactions encourage the particle to move in the same direction as its neighbours, similar to how the nearest-neighbour interactions in the Ising model encourage its spins to align.

Models of this sort exhibit a transition between two distinct phases. When there is a low density of particles, there are relatively few interactions and the particles' motion is disordered, with no murmurations or swarms. At high densities, in contrast, large-scale collective motion appears, based solely on the interactions between the individual particles in the system. An order parameter sensitive to this behaviour is the average angle of the particles' motion. Different model variations can predict either first-order or second-order transitions between these two phases, and in the second-order case the universality class of the critical exponents also depends on the model details. [In some cases](#), as the critical density is approached from the disordered phase, the particles exhibit anomalous super-diffusion like that investigated in the computer assignment.

10.4 Wrap-up recap and synthesis

In this module we have built on foundations from probability theory to develop and apply the core concepts of statistical mechanics. In **Unit 1** we defined probability spaces, expectation values and variances, and used these to establish the law of large numbers through which stable large-scale behaviour occurs for stochastic systems that involve a large number of degrees of freedom, $N \gg 1$. We also saw how the central limit theorem relates large- N probability distributions to the underlying mean and variance of the microscopic degrees of freedom, and practiced extracting probabilities from such distributions. The law of diffusion results from the central limit theorem, and the computer assignment applied inverse transform sampling to study how anomalous diffusion can arise when the central limit theorem's assumption of a finite mean and variance breaks down.

Starting in **Unit 2** we focused on statistical ensembles in thermodynamic equilibrium as more specialized realizations of probability spaces. These statistical ensembles consist of the set of micro-states ω_i that a system can possibly adopt through its evolution in time, each with the corresponding probability p_i that it will actually be adopted. The laws of nature — such as the first law of thermodynamics that requires conservation of energy — impose constraints on statistical ensembles. The micro-canonical ensemble directly implements such constraints, by requiring the system's internal energy and particle number be constant as it evolves in time. This implies that the system must be completely isolated from the rest of the universe. Important derived quantities in the micro-canonical ensemble are the entropy (Eq. 20) and temperature (Eq. 22), which we explored through the simple example of non-interacting spin systems. In particular, we derived a form of the second law of thermodynamics, which states that the total entropy never decreases as time passes and therefore implies that maximal entropy corresponds to thermodynamic equilibrium.

Motivated by the impracticality of demanding that a system be completely isolated, in **Unit 3** we turned to the canonical ensemble, which allows systems to exchange energy with a large thermal reservoir that imposes a constant temperature. The particle number is still fixed, while the entropy, internal energy and heat capacity are now important derived quantities, with a fluctuation–dissipation relation between the last two. Maximizing the entropy to consider systems in thermodynamic equilibrium provides the partition function (Eq. 33) and Boltzmann distribution (Eq. 34) as a fundamental mathematical definition of the system. Derived quantities are determined

from the partition function, or equivalently from the Helmholtz free energy (Eq. 37).

By analysing non-interacting systems of both distinguishable and indistinguishable spins, we demonstrated that the intrinsic information content of statistical systems has physically measurable effects. These physical measurements are possible for canonical systems that don't have to be completely isolated, and in tutorial activities we saw that experimentally measured heat capacities for many materials can be modeled by the Einstein solid, with improved results obtainable through the Debye theory of solids. These analyses highlighted the key strategy of considering high- and low-temperature limits in which statistical systems can simplify dramatically.

In **Unit 4** we continued working with the canonical ensemble, applying it to study non-interacting (ideal) gases of non-relativistic, classical particles. At the start of this analysis we had to regulate the system to ensure that its partition function was well-defined. We did this by assuming that only discrete momenta are possible in a finite volume V , and then taking the limit of continuously varying momenta that allowed us to replace discrete sums with continuous integrals. Considering ideal gases of both distinguishable and indistinguishable particles, we determined their partition functions and used these to derive the internal energy (Eq. 52) and entropy (Eq. 53). We saw that mixing two gases of distinguishable particles produces a positive mixing entropy, meaning this process is irreversible, in contrast to combining and reseparating gases of indistinguishable particles. Finally we defined the pressure as the change in the system's internal energy upon changing its volume while keeping its entropy constant (Eq. 55), and derived the ideal gas law (Eq. 56) as a famous equation of state.

Building on these analyses of ideal gases in the canonical ensemble, in **Unit 5** we considered thermodynamic cycles as systems that perform a repeatable sequence of expansions, compressions and heat transfers in order to act as heat engines or refrigerators. This required first considering the work done on the system through these processes (Eq. 59), the heat added to or removed from it (Eq. 61), and the first law of thermodynamics expressed in terms of these quantities (Eq. 63). As two limiting cases, we contrasted slow isothermal processes that feature sufficient heat exchange to keep the temperature constant, versus more realistic adiabatic processes that occur too quickly for any heat to be exchanged. After introducing PV diagrams as a convenient way to visualize thermodynamic cycles and the individual processes that comprise them, we analysed the Carnot heat engine and computed how much work it can do on its surroundings as heat flows through it from a hot thermal reservoir to a cold reservoir. This balance of work and heat defines the engine's efficiency (Eq. 64). We showed that the Carnot cycle achieves the maximum possible efficiency allowed by the second law of thermodynamics, and in tutorials we confirmed the lower efficiency of the Otto cycle that describes a petrol engine.

Returning to the formulation of statistical ensembles, in **Unit 6** we further generalized our perspective to consider systems that can exchange both particles and energy with a large particle reservoir. Introducing the chemical potential (Eq. 67) as the new quantity that the particle reservoir keeps constant — along with the temperature — defined the grand-canonical ensemble. A final round of entropy maximization gave us the grand-canonical partition function (Eq. 74) and the corresponding grand-canonical potential (Eq. 75) that determine the derived quantities, which now include the particle

number in addition to the internal energy and the entropy. We were also able to derive a generalized thermodynamic identity (Eq. 80) that relates the chemical potential to the change in energy upon adiabatically adding a particle to the system.

Our main applications of the grand-canonical ensemble involved analysing quantum gases, and as preparation for that we introduced quantum statistics in **Unit 7** — simply considering this as an ansatz rather than relying on prior knowledge of quantum mechanics. After demonstrating how our earlier classical (non-quantum) approach to ideal gases breaks down when there is a non-negligible probability for multiple identical particles to occupy the same energy level, we got around this problem by organizing the micro-states in terms of the possible occupation numbers of the energy levels. These possible occupation numbers distinguish the two types of particles that appear in nature: bosons can have any non-negative occupation numbers $n_\ell \in \mathbb{N}_0$, while fermions obey the Pauli exclusion principle and can have only $n_\ell \in \{0, 1\}$. We derived (and factorized) the respective Bose–Einstein (Eq. 86) and Fermi–Dirac (Eq. 88) statistics for these two types of quantum particles, and checked that both approach classical Maxwell–Boltzmann statistics (Eq. 87) in the limit of high temperature with large negative chemical potential $-\mu \gg T$.

With quantum statistics in hand, we applied the grand-canonical ensemble to analyse quantum gases in **Unit 8**, first considering a non-interacting ideal gas of ultra-relativistic photons with energies defined in terms of frequencies (Eq. 93). Based on the grand-canonical potential, we derived the Planck spectrum (Eq. 95) governing the frequency dependence of the energy density for the photon gas, and learned how it solves the ultraviolet catastrophe of the classical Rayleigh–Jeans spectrum. We also saw how the Planck spectrum provides an excellent mathematical model for both solar radiation from stars — among the hottest environments in the universe — as well as the cosmic microwave background that fills frigid inter-galactic space and provides strong evidence for the existence of dark matter. We finally derived the radiation pressure of photon gases, and the corresponding equation of state (Eq. 98), which has the same form as the ideal gas law, just with a different numerical factor.

We continued with a similar application of the grand-canonical ensemble to ideal quantum gases of non-interacting fermions, focusing mainly on non-relativistic particles and considering the low-temperature regime where quantum Fermi–Dirac statistics differs the most from the classical case. Again based on the grand-canonical potential, we derived the Fermi function $F(E)$ and saw how it approaches a step function at low temperatures. This corresponds to all the low-lying energy levels being filled — each level occupied by the single fermion allowed by the Pauli exclusion principle. The transition from filled to empty energy levels defines the Fermi energy, which in the zero-temperature limit is simply the chemical potential (Eq. 101). The resulting internal energy (Eq. 103) and pressure (Eq. 104) remain positive even as the temperature approaches absolute zero. This non-zero degeneracy pressure helps to explain the regularity of type-Ia supernovas, which play a key role in establishing the existence of so-called dark energy. After also deriving the equation of state for a relativistic ideal fermion gas, we wrapped up the non-interacting portion of the module by using the Sommerfeld expansion to predict the leading low-temperature corrections to the zero-temperature limits of $\frac{\mu}{E_F}$ (Eq. 109) and the heat capacity ($c_v \propto T$).

Our final major topic was to explore interacting statistical systems, making further use of the tools we had previously developed. Statistical systems in which the microscopic degrees of freedom interact with each other can exhibit a broader array of phenomena, including phase transitions. At the same time, they become enormously more difficult to analyse, because their partition functions and derived quantities no longer factorize into independent single-particle contributions. In **Unit 9** we focused on the famous Ising model, which is simple to write down as a system of spins interacting with their nearest neighbours on a d -dimensional lattice (Eq. 111), but extremely difficult to solve exactly in two or more dimensions. For $d \geq 2$ dimensions, the Ising model exhibits a second-order phase transition between its high-temperature disordered phase and its low-temperature ordered phase, with its magnetization serving as the order parameter. We analysed the Ising model through the mean-field approximation that produced a self-consistency condition (Eq. 117) for this order parameter. We then exactly solved the one-dimensional Ising model and used Kramers–Wannier duality to predict the exact critical temperature in two dimensions, showing that the mean-field approximation is not very reliable in a low number of dimensions.

Although the mean-field approximation to the Ising model does capture a second-order phase transition, the critical temperature and critical exponents it predicts for this transition only slowly converge towards the correct values as the dimensionality of the system increases. Determining the correct critical behaviour in $d > 2$ dimensions requires numerical analyses that we discussed in this final **Unit 10**. Monte Carlo importance sampling algorithms are the standard way to carry out reliable numerical computations with controlled uncertainties in a realistic amount of time, and are very broadly applicable to more general interacting statistical systems. Amazingly, extremely different physical systems can exhibit phase transitions characterized by exactly the same critical exponents, a phenomenon known as universality, which helps to explain the broad applicability and far-reaching power of statistical mechanics.

In summary, we have learned foundations including statistical ensembles, entropy, and the laws of thermodynamics. We have studied applications including diffusion, ideal gases, thermodynamic cycles, and phase transitions. And we have explored advanced topics including numerical methods and universality. All together, our new knowledge of statistical mechanics and thermodynamics enables us to observe and appreciate many further uses of these concepts and tools across the mathematical sciences and beyond.