

Statistical Mechanics

Entropy, Information, and Quantum Systems

Marcelo Santos Guimarães
Department of Theoretical Physics
Institute of Physics, UERJ

Graduate course, 2026.2

Preface

These notes accompany the graduate course in Statistical Mechanics offered at UERJ in the second semester of 2026. They grew out of an earlier undergraduate text, but the organizing question has changed:

What does entropy measure, why do several inequivalent formulas carry the same name, and under which assumptions do they describe the same physics?

Standard statistical mechanics remains the subject. We will derive the equilibrium ensembles, calculate classical and quantum gases, study phase transitions and the Ising model, and finish with the renormalization group. Information theory enters because it supplies exact tools for discussing inference, correlations, distinguishability, and loss of information. It doesn't replace mechanics or thermodynamics.

Entropy conventions

Units cause avoidable confusion, so we fix them now. Information-theoretic entropies are dimensionless and use natural logarithms. For a probability distribution p and a density operator ρ ,

$$h(p) = - \sum_i p_i \ln p_i, \quad s(\rho) = - \text{Tr}(\rho \ln \rho). \quad (0.1)$$

Their unit is the nat. Thermodynamic entropies carry units of energy divided by temperature:

$$S_G[p] = k_B h(p), \quad S_{VN}(\rho) = k_B s(\rho). \quad (0.2)$$

Relative entropy and mutual information stay dimensionless. Thus

$$I(A : B) = s(\rho_A) + s(\rho_B) - s(\rho_{AB}) = D(\rho_{AB} \| \rho_A \otimes \rho_B) \quad (0.3)$$

has consistent units, while the nonequilibrium free-energy identity reads

$$F_T(\rho) - F_T(\rho_\beta) = k_B T D(\rho \| \rho_\beta) \quad (T \neq 0). \quad (0.4)$$

How claims are marked

Statistical mechanics mixes exact identities with limiting arguments and approximations. Each matters, but they are not interchangeable. Boxes in the margin of the argument use the following labels:

- **Exact:** an identity or theorem under the assumptions stated nearby.
- **Thermodynamic limit:** a statement obtained as $N, V \rightarrow \infty$ with densities fixed.

- **Controlled approximation:** an expansion with a stated small parameter and an estimate of what has been dropped.
- **Mean field:** fluctuations have been replaced by an average field.
- **Model-specific exact result:** exact for the named model, not a universal theorem.
- **Heuristic:** a physical argument that guides calculation but does not prove the general claim.

The labels become particularly important when we discuss the second law. Fine-grained Gibbs entropy is conserved by Hamiltonian flow, and von Neumann entropy is conserved by unitary evolution. An increasing entropy requires another ingredient: coarse graining, coupling to an environment, stochastic dynamics, an assumption about correlations, or a restriction to macroscopic observables.

Course map

The first ten lectures cover equilibrium thermodynamics, probability, and ensembles. Lectures 11–18 study classical systems, irreversibility, and information thermodynamics; lectures 19–20 introduce quantum entropy before the second exam. Four further lectures develop ideal quantum gases. The final five introduce phase transitions, the Ising model, and renormalization.

Chapter 1 previews the semester-long entropy map. We return to that map after each part of the course, and Chapter 12 places the several definitions and approximation statements side by side.

Each chapter ends with exercises. Problems marked with an asterisk require a longer calculation or an idea not demonstrated in the main text. The hints appendix gives starting points, not full solutions.

Notation at a glance

Symbol	Meaning
$h(p)$, $s(\rho)$	Dimensionless Shannon and von Neumann entropies.
$S_G = k_B h$, $S_{vN} = k_B s$	The corresponding entropies with thermodynamic units.
$S_B = k_B \ln \Omega$	Boltzmann entropy of a specified macrostate or energy shell.
$D(\cdot \parallel \cdot)$, $I(A : B)$	Dimensionless relative entropy and mutual information.
Σ	Dimensionless entropy production, measured in units of k_B , unless units are restored explicitly.
β , z , λ_T	$\beta = (k_B T)^{-1}$, fugacity $z = e^{\beta \mu}$, and thermal wavelength.
$\mathcal{F}_T[p]$, $F_T(\rho)$	The same nonequilibrium-free-energy construction for classical probabilities and quantum states.

Prerequisites

We assume undergraduate thermodynamics, classical mechanics, and quantum mechanics. Students should know linear algebra, density operators, tensor products, and partial traces. Chapter 7 reviews these notions before using them.

Contents

Preface	ii
I Equilibrium and inference	1
1 Thermodynamics and the macroscopic entropy	2
1.1 Why begin with thermodynamics?	2
1.2 The laws as constraints on processes	3
1.2.1 Zeroth law and temperature	3
1.2.2 First law	3
1.2.3 Second law and Carnot's theorem	3
1.3 From Carnot's theorem to entropy	5
1.3.1 Signed heat and a cycle with several reservoirs	5
1.3.2 What the auxiliary engines do	6
1.3.3 The Clausius inequality for a cycle	8
1.3.4 Reversible cycles define entropy	8
1.3.5 Irreversible processes and entropy production	9
1.4 The fundamental relation	10
1.4.1 Equilibrium between subsystems	10
1.4.2 Extensivity and Euler's relation	11
1.5 Thermodynamic potentials	11
1.6 Stability, concavity, and response	13
1.7 Third law and residual entropy	14
1.8 Exercises	14
1.9 Further reading	14
2 Probability, information, and typicality	15
2.1 Probability enters through a description	15
2.1.1 Frequency, inference, and physical ensembles	16
2.1.2 Generating functions and cumulants	17
2.2 Sample averages and concentration inequalities	19
2.3 Binomial model: multiplicity becomes entropy	20
2.4 Shannon entropy	23
2.4.1 Joint and conditional entropy	24
2.4.2 Typical sequences	24
2.5 Relative entropy and distinguishability	25
2.6 Mutual information and data processing	26
2.6.1 Stochastic maps and the log-sum inequality	26
2.6.2 Markov chains and information loss	27
2.7 Maximum entropy and exponential families	27

2.8	Large deviations	29
2.9	Continuous variables and reference measures	30
2.10	Exercises	31
2.11	Further reading	32
3	Equilibrium ensembles	33
3.1	Classical state counting	33
3.2	Microcanonical ensemble	33
3.3	Canonical ensemble from a reservoir	34
3.3.1	Thermodynamics from the partition function	35
3.3.2	Fluctuations and response	36
3.4	Grand-canonical ensemble	37
3.5	Canonical variational principle	38
3.6	Microcanonical to canonical: Laplace principle	39
3.7	When are ensembles equivalent?	40
3.8	Worked example: independent two-level systems	41
3.9	Exercises	42
3.10	Further reading	42
II	Classical systems and irreversibility	43
4	Classical gases and interacting particles	44
4.1	Monatomic ideal gas	44
4.1.1	Gibbs paradox	45
4.2	Maxwell distribution	46
4.2.1	Distribution of kinetic energy	47
4.3	Equipartition theorem	47
4.3.1	Classical harmonic solid	48
4.4	Diatomic molecules	48
4.5	No classical orbital magnetism	50
4.6	Interacting gases	50
4.6.1	Mayer function and the second virial coefficient	50
4.6.2	van der Waals model	52
4.7	Collisions and the dilute-gas length scale	53
4.8	Exercises	54
4.9	Further reading	54
5	Microscopic dynamics and irreversibility	55
5.1	The apparent contradiction	55
5.2	Liouville's theorem	56
5.2.1	Volume preservation	56
5.2.2	The Liouville equation	57
5.2.3	Conservation of the fine-grained entropy	58
5.3	Coarse graining	58
5.3.1	Cells and the smoothed density	59
5.3.2	Averaging never lowers the entropy	59
5.3.3	Where the arrow of time enters	60
5.3.4	A closed dynamics for the cell probabilities	61
5.3.5	Ergodicity, mixing, and chaos	64
5.4	Master equations	66

5.4.1	From transition probabilities to rates	66
5.4.2	Stationary states and detailed balance	67
5.4.3	Relaxation and relative entropy	68
5.5	Entropy production in jump processes	71
5.5.1	Entropy of the system	71
5.5.2	Entropy of the environment: local detailed balance	71
5.5.3	Total entropy production	72
5.6	The Boltzmann equation	73
5.6.1	The one-particle distribution and free streaming	73
5.6.2	Counting collisions	74
5.6.3	The H -theorem	76
5.6.4	Loschmidt, Zermelo, and the arrow of time	78
5.7	Nonequilibrium work	79
5.7.1	Work on a fluctuating system	79
5.7.2	The Jarzynski equality	80
5.7.3	Crooks' relation and the arrow of time	81
5.8	Summary: which entropy increases?	82
5.9	Exercises	83
5.10	Further reading	84
6	Information is physical	85
6.1	Maxwell's demon	85
6.2	A memory is a physical system	86
6.2.1	Logical states and microstates	86
6.2.2	Logical and thermodynamic reversibility	88
6.2.3	Where the erased information goes	88
6.3	Szilard's engine	89
6.4	Work and free energy out of equilibrium	91
6.4.1	An exact identity	91
6.4.2	Three readings of the bound	93
6.4.3	Reaching the bound	93
6.5	Landauer's bound	94
6.5.1	Erasing a random bit	94
6.5.2	A reversible erasure	94
6.5.3	Biased bits and asymmetric memories	96
6.6	Measurement and feedback	97
6.6.1	Measurement creates correlation	97
6.6.2	The feedback bound	98
6.6.3	Processing the record	99
6.7	Closing the cycle	100
6.7.1	The cost of measurement and erasure	100
6.7.2	The entropy ledger of the Szilard cycle	101
6.7.3	Maxwell's demon revisited	102
6.8	Summary	103
6.9	Exercises	104
6.10	Further reading	105

III	Quantum statistics	106
7	Quantum states, entropy, and equilibrium	107
7.1	From probabilities to operators	107
7.2	Density operators	108
7.2.1	Qubit review	109
7.3	Composite systems and partial trace	110
7.3.1	Purification	111
7.4	von Neumann entropy	111
7.5	Quantum relative entropy	113
7.6	Quantum Gibbs state and free energy	114
7.7	Quantum mutual information	115
7.8	Entanglement entropy	116
7.9	Rényi entropies	117
7.10	Dynamics: closed and open	117
7.11	Typicality and thermalization	119
7.11.1	Canonical typicality: a statement about most vectors	119
7.11.2	Time evolution and the diagonal ensemble	119
7.11.3	The eigenstate thermalization hypothesis	120
7.12	Further reading	120
7.13	Exercises	120
8	Identical particles and ideal quantum gases	122
8.1	Identical particles	122
8.2	Grand partition function	123
8.3	Density of states	124
8.4	Degenerate Fermi gas	125
8.4.1	Sommerfeld expansion	127
8.5	Ideal Bose gas and condensation	128
8.6	Photon gas and blackbody radiation	131
8.7	Classical limit and exchange corrections	133
8.8	Exercises	134
8.9	Further reading	134
IV	Collective behavior and scale	135
9	Phase transitions, order parameters, and scaling	136
9.1	Nonanalyticity requires infinitely many degrees of freedom	136
9.2	First-order transitions and coexistence	137
9.2.1	van der Waals loop	138
9.3	Continuous transitions and order parameters	138
9.4	Landau theory	139
9.4.1	First-order behavior from the same expansion	140
9.5	Spatial fluctuations and correlation length	141
9.5.1	When is mean field self-consistent?	142
9.6	Scaling hypothesis	143
9.7	Exercises	144
9.8	Further reading	144

10 The Ising model	145
10.1 Definition and symmetry	145
10.2 One-dimensional transfer matrix	145
10.2.1 Correlations	147
10.2.2 Domain-wall picture	148
10.3 Mean-field theory	149
10.4 Two-dimensional square lattice	151
10.5 Metropolis sampling	152
10.6 Finite-size scaling	153
10.7 Exercises	154
10.8 Further reading	154
11 Renormalization and universality	155
11.1 Why scale enters	155
11.2 Exact definition of a coarse-grained Hamiltonian	156
11.3 One-dimensional Ising decimation	156
11.3.1 Fixed points and the free-energy contribution	157
11.4 Fixed points and scaling fields	158
11.5 Critical exponents from RG eigenvalues	159
11.6 Power counting in the Landau–Ginzburg theory	160
11.6.1 What the first epsilon-expansion step says	160
11.7 Universality	161
11.8 Entropy and coarse graining	162
11.9 Entanglement near a one-dimensional quantum critical point	162
11.10 Core results and further reading	163
11.11 Exercises	163
11.12 Further reading	164
12 One word, several entropies	165
12.1 The map	165
12.2 Exact coincidences	166
12.2.1 Uniform microcanonical shell	166
12.2.2 Equilibrium Gibbs distribution	167
12.2.3 Free-energy excess	167
12.2.4 Correlation as distinguishability from a product	167
12.3 Statements that require a limit or approximation	168
12.3.1 Boltzmann entropy and thermodynamic entropy	168
12.3.2 Equivalence of ensembles	168
12.3.3 Entropy increase	168
12.3.4 An exact system–bath entropy balance	168
12.4 Three false shortcuts and their counterexamples	169
12.4.1 “Entropy is missing information”	169
12.4.2 “Entanglement is thermal entropy”	169
12.4.3 “Entropy always increases after discarding information”	169
12.5 Four diagnostic examples	169
12.5.1 Free expansion of an ideal gas	169
12.5.2 A thermal qubit	170
12.5.3 A Bell pair	170
12.5.4 A critical chain	170
12.6 A decision procedure	170
12.7 Synthesis exercises	170

V	Appendices	172
A	Mathematical tools	173
A.1	Stirling's formula	173
A.2	Gaussian integrals	173
A.3	Gamma, zeta, and Bose–Fermi integrals	174
A.4	Laplace's method	175
A.5	Legendre and Legendre–Fenchel transforms	176
A.6	Delta functions and density of states	177
A.7	Fourier transforms and Gaussian correlations	177
A.8	Cumulants, sources, and response	178
A.9	Operator derivatives and traces	178
A.10	Schmidt decomposition	179
A.11	Useful inequalities	179
A.11.1	Log-sum inequality and classical data processing	179
A.11.2	Norms and distinguishability	180
A.12	Partial traces and operator logarithms	180
A.13	Constants and scale checks	181
B	Hints for selected exercises	182
C	Formula and assumption summary	186
C.1	Thermodynamics	186
C.2	Equilibrium ensembles	187
C.3	Classical ideal and dilute gases	187
C.4	Information and nonequilibrium relations	188
C.5	Quantum entropy and correlations	189
C.6	Ideal quantum gases	190
C.7	Critical phenomena and renormalization	191
	References	192

Part I

Equilibrium and inference

Chapter 1

Thermodynamics and the macroscopic entropy

Lecture 1–4. Macroscopic states, the laws of thermodynamics, Clausius entropy, thermodynamic potentials, and stability.

What you should be able to do

- Read the fundamental relation as a geometric statement about equilibrium states.
- Derive the conditions for thermal, mechanical, and diffusive equilibrium.
- Explain what the Clausius entropy does and does not say about microscopic dynamics.
- Use Legendre transforms and response functions without losing track of the controlled variables.

1.1 Why begin with thermodynamics?

A gas contains too many microscopic variables to monitor. Thermodynamics starts elsewhere: it describes equilibrium states by a small set of quantities such as energy E , volume V , particle number N , pressure P , temperature T , and entropy S . The reduction is drastic. It is also empirical. Statistical mechanics must explain when such a reduced description works and why its predictions become sharp for large systems.

Thermodynamics therefore comes before microscopic counting in these notes. It tells us which relations statistical mechanics must reproduce. It also fixes the operational meaning of entropy before we reinterpret entropy as multiplicity or missing information.

An *equilibrium state* remains macroscopically unchanged while the external constraints remain fixed. For a simple one-component system, we may use (E, V, N) as independent extensive variables. The entropy representation assigns a number

$$S = S(E, V, N) \tag{1.1}$$

to each equilibrium state. Equivalently, the energy representation uses

$$E = E(S, V, N). \tag{1.2}$$

The two descriptions contain the same information when the relevant derivatives exist and $T > 0$.

The entropy map for the semester

Quantity	Role
Thermodynamic S	A state function defined by reversible heat and used to compare equilibrium endpoints.
Boltzmann $S_B = k_B \ln \Omega$	The logarithm of the multiplicity assigned to a macrostate.
Gibbs–Shannon $S_G = k_B h(p)$	Entropy of a probability distribution; on a uniform shell it equals S_B .
von Neumann $S_{vN} = k_B s(\rho)$	Spectral entropy of a quantum state; conserved globally under unitary evolution.
Entropy production Σ	A process-dependent balance term whose definition depends on the chosen open, stochastic, or coarse-grained description.

Each part of the course returns to this list and states which equalities are exact and which need a limit or approximation. Chapter 12 develops the full map and its counterexamples.

1.2 The laws as constraints on processes

1.2.1 Zeroth law and temperature

Suppose systems A and B can exchange energy but not volume or particles. If they no longer show a net energy current, they are in thermal equilibrium. The zeroth law states that thermal equilibrium is transitive: if A is in equilibrium with B , and B with C , then A is in equilibrium with C . This equivalence relation allows us to label an equilibrium class by a scalar temperature.

The zeroth law does not fix the numerical scale of temperature. A monotone function of an empirical temperature labels the same equilibrium classes. Carnot’s theorem will select an absolute scale up to a multiplicative constant.

1.2.2 First law

Energy is a state function. Heat and work describe transfers during a process and need not be state functions. With work defined as work done by the system,

$$dE = \delta Q - \delta W. \quad (1.3)$$

For quasistatic compression and particle exchange,

$$\delta W = P dV - \mu dN, \quad (1.4)$$

where the sign of the chemical term follows from treating particle addition as work done on the system. Combining terms gives

$$dE = \delta Q - P dV + \mu dN. \quad (1.5)$$

The notation matters. The symbol dE is an exact differential; its integral depends only on the endpoints. The symbols δQ and δW remind us that heat and work depend on the path.

1.2.3 Second law and Carnot’s theorem

No cyclic engine can convert heat drawn from a single reservoir entirely into work. An equivalent Clausius statement says that heat does not flow spontaneously from a colder body to a hotter one.

Consider an engine operating in a cycle between a hot reservoir and a cold reservoir, at empirical temperatures θ_h and θ_c . During one cycle it absorbs heat $Q_h > 0$ from the hot reservoir, releases heat $Q_c > 0$ to the cold reservoir, and delivers work $W > 0$. Because the engine returns to its initial state, its energy change over the cycle vanishes. The first law therefore gives

$$W = Q_h - Q_c.$$

The *efficiency* is the fraction of the absorbed heat converted into work:

$$\eta \equiv \frac{W}{Q_h} = 1 - \frac{Q_c}{Q_h}.$$

Thus $0 < \eta < 1$ for an ordinary heat engine. In the proof of Carnot's theorem, reversibility supplies one precise operation: the same cycle can run backward with every heat and work transfer reversed. In that direction the engine consumes work W , removes Q_c from the cold reservoir, and deposits $Q_h = Q_c + W$ in the hot reservoir.

Carnot's theorem states that all reversible engines operating between the same two reservoirs have the same efficiency, independent of their working substances. Suppose, to the contrary, that two such engines A and B satisfy $\eta_A > \eta_B$. Run each engine for a suitable number of cycles so that, in the forward direction, each transfers the same total heat Q from the hot reservoir. Heat and work both scale with the number of cycles. Now run A forward and B backward. In the following ledger, ΔE_h and ΔE_c are changes in the reservoir energies, while $W_{\text{out}} > 0$ denotes work delivered to the surroundings:

device	ΔE_h	ΔE_c	W_{out}
A forward	$-Q$	$+(1 - \eta_A)Q$	$+\eta_A Q$
B backward	$+Q$	$-(1 - \eta_B)Q$	$-\eta_B Q$
net	0	$-(\eta_A - \eta_B)Q$	$+(\eta_A - \eta_B)Q$

The hot reservoir has no net change. Both engines complete cycles and also have no net change. The cold reservoir alone loses the heat $(\eta_A - \eta_B)Q$, and the combined device converts exactly that amount into work. This is the one-reservoir engine forbidden by the Kelvin–Planck statement of the second law. Hence two reversible engines between the same reservoirs cannot have different efficiencies.

The argument used only the identities of the two reservoirs, not the working substance or the engine design. The common reversible efficiency therefore depends only on θ_c and θ_h . Since $Q_c/Q_h = 1 - \eta$, the heat ratio also depends only on the two reservoir temperatures:

$$\frac{Q_c}{Q_h} = f(\theta_c, \theta_h). \quad (1.6)$$

The function f obeys a composition rule. Take three reservoirs with $\theta_1 < \theta_2 < \theta_3$. Run a reversible engine from θ_3 to θ_2 , and use the heat Q_2 that it releases as the input to a second reversible engine from θ_2 to θ_1 . The intermediate reservoir receives and loses the same Q_2 , so its net change is zero. The two devices together form a reversible engine from θ_3 to θ_1 . Consequently,

$$\frac{Q_2}{Q_3} = f(\theta_2, \theta_3), \quad \frac{Q_1}{Q_2} = f(\theta_1, \theta_2),$$

and multiplication gives

$$f(\theta_1, \theta_3) = f(\theta_1, \theta_2)f(\theta_2, \theta_3). \quad (1.7)$$

This relation constructs the absolute temperature scale. Running a reversible engine backward exchanges the two arguments and inverts the heat ratio, so define $f(\theta_2, \theta_1) =$

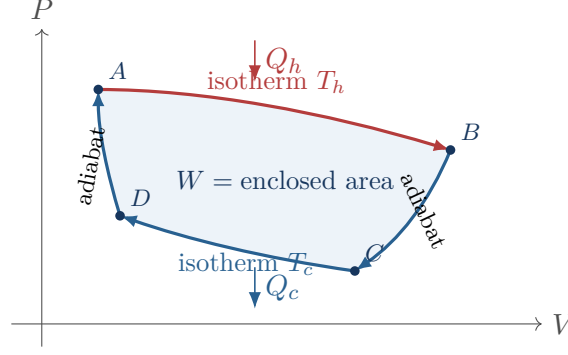


Figure 1.1: A reversible Carnot cycle in the P - V plane. Heat enters on the hot isotherm, leaves on the cold isotherm, and the enclosed area is the work delivered per cycle. The precise curve shapes depend on the working substance; the heat ratio does not.

$1/f(\theta_1, \theta_2)$. The same two-engine construction works for any ordering of three temperatures: reverse whichever engine is needed, and the intermediate heat still cancels. We may therefore choose an arbitrary reference reservoir θ_0 and define

$$T(\theta) = T_0 f(\theta, \theta_0),$$

where $T_0 > 0$ fixes the unit. Applying the composition rule with the reference reservoir gives

$$f(\theta_1, \theta_2) = \frac{f(\theta_1, \theta_0)}{f(\theta_2, \theta_0)} = \frac{T(\theta_1)}{T(\theta_2)}.$$

Changing T_0 multiplies every temperature by the same constant and leaves all heat ratios unchanged. For a reversible engine, $Q_c/Q_h = T_c/T_h$, so

$$\frac{Q_h}{T_h} - \frac{Q_c}{T_c} = 0, \quad \eta_{\text{rev}} = 1 - \frac{T_c}{T_h}. \quad (1.8)$$

1.3 From Carnot's theorem to entropy

Equation (1.8) concerns a reversible engine coupled to two reservoirs. A general cycle is less restrictive: during different parts of the cycle, a system may exchange heat at different temperatures. The purpose of this section is to find the restriction imposed by the second law on all those heat exchanges. For a reversible cycle, that restriction will become an equality and will define entropy.

1.3.1 Signed heat and a cycle with several reservoirs

First rewrite the two-reservoir Carnot result using signed heats. The quantities Q_h and Q_c in Eq. (1.8) were positive magnitudes. The engine absorbed Q_h from the hot reservoir and released Q_c to the cold reservoir. If heat is positive when it enters the engine, then

$$q_h = Q_h, \quad q_c = -Q_c.$$

Carnot's heat ratio reads

$$\frac{q_h}{T_h} + \frac{q_c}{T_c} = \frac{Q_h}{T_h} - \frac{Q_c}{T_c} = 0.$$

The same statement holds for any reversible cyclic engine operating between two reservoirs. If its signed heat inputs are q_0 and q_i , and the reservoir temperatures are T_0 and T_i , then

$$\frac{q_0}{T_0} + \frac{q_i}{T_i} = 0. \quad (1.9)$$

Running the engine backward changes (q_0, q_i) into $(-q_0, -q_i)$, so the identity remains true.

Now consider one system, denoted by K , that completes a cycle. The system K need not be reversible and need not deliver positive work. During the cycle it is placed, in succession, in thermal contact with reservoirs

$$R_1, R_2, \dots, R_n, \quad \text{at temperatures} \quad T_1, T_2, \dots, T_n.$$

There are therefore n reservoirs in this construction. They need not touch K simultaneously. For example, a gas may absorb heat while in contact with one reservoir during one leg of its cycle and release heat to another reservoir during a later leg. A continuously varying boundary temperature will be obtained below as the limit of a finer list of reservoirs.

Let Q_i be the total signed heat that enters K from reservoir R_i during the complete cycle:

- $Q_i > 0$: K absorbs heat Q_i from R_i ;
- $Q_i < 0$: K releases heat $|Q_i|$ to R_i .

Because K returns to its initial state, $\Delta E_K = 0$. The first law therefore gives

$$W_K = \sum_{i=1}^n Q_i, \quad (1.10)$$

where W_K is the net work delivered by K . A negative W_K means that net work was done on it.

1.3.2 What the auxiliary engines do

The heats Q_i come from reservoirs at different temperatures, so their unweighted sum contains only the first-law information. Carnot's relation lets us express every one of these exchanges as an equivalent heat exchange at one common temperature.

Choose an additional reservoir R_0 at an arbitrary temperature $T_0 > 0$. The system K never interacts with R_0 . For each reservoir R_i , introduce a reversible auxiliary engine C_i operating between R_i and R_0 . Thus there are n auxiliaries,

$$C_1, C_2, \dots, C_n,$$

one for each of the n reservoirs used by K . These auxiliary engines are hypothetical comparison devices. They do not describe the internal operation of K .

Operate each C_i in the direction that cancels what K did to R_i :

- If $Q_i > 0$, then K removed heat Q_i from R_i . Run C_i so that it returns the same heat Q_i to R_i .
- If $Q_i < 0$, then K deposited heat $|Q_i|$ in R_i . Run C_i backward so that it removes the same heat $|Q_i|$ from R_i .

In both cases, the signed heat entering C_i from R_i is $-Q_i$. The heat exchanges of K and C_i with R_i add to zero:

$$\underbrace{Q_i}_{\text{into } K} + \underbrace{(-Q_i)}_{\text{into } C_i} = 0.$$

Both terms measure heat drawn from the same reservoir R_i , so this cancellation means that R_i returns to its initial energy. The size of the auxiliary, or the number of times it is operated, is chosen so that these two amounts agree.

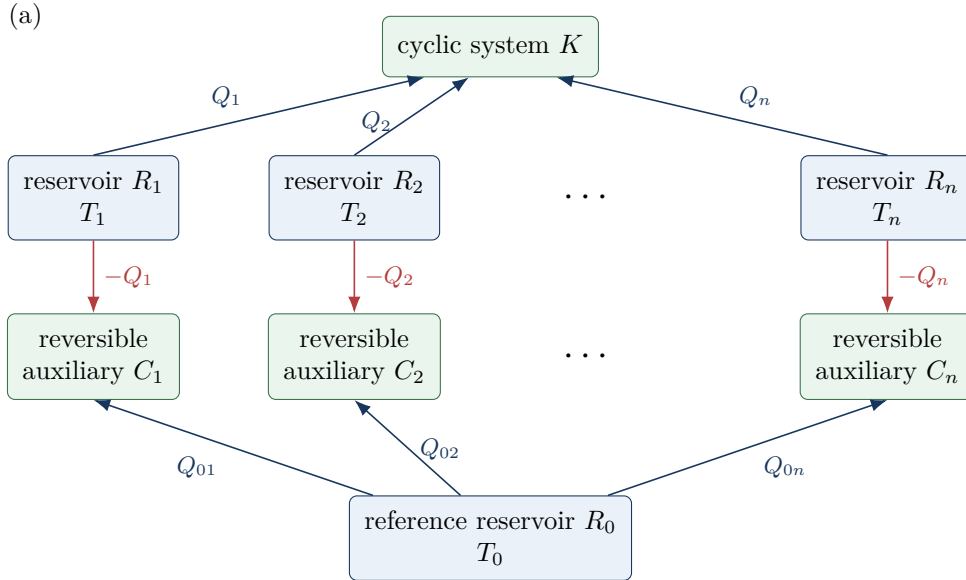
Let Q_{0i} denote the signed heat entering C_i from the reference reservoir R_0 . Since C_i is reversible, Eq. (1.9) gives

$$\frac{Q_{0i}}{T_0} + \frac{-Q_i}{T_i} = 0, \quad \text{hence} \quad Q_{0i} = T_0 \frac{Q_i}{T_i}. \quad (1.11)$$

This equation determines the heat exchanged with R_0 . Work may enter or leave C_i as required by the first law; its direction does not affect the cancellation at R_i .

What is being combined?

The system K is one cyclic system with n thermal contacts. The devices $C_1, \dots, C_n$ are separate reversible engines added only for this proof. They are chosen so that every reservoir R_i has zero net heat exchange with $K + C_i$. The composite system used below is $K + C_1 + \dots + C_n$.



Each C_i cancels the exchange of K with R_i .

A negative signed heat flows opposite to the arrow shown.

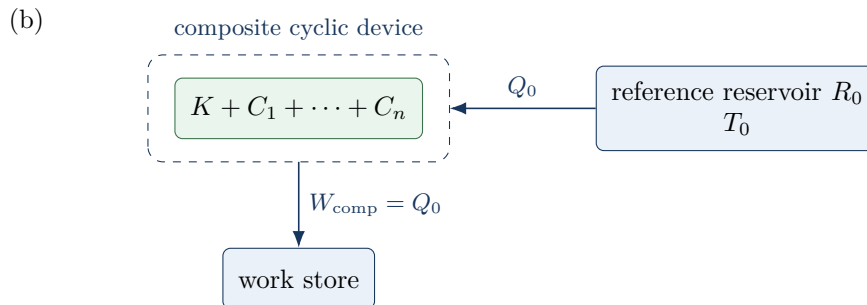


Figure 1.2: The construction for a cycle with n reservoirs. (a) The single system K exchanges heats $Q_1, \dots, Q_n$ with the reservoirs $R_1, \dots, R_n$. Each reservoir has its own reversible auxiliary C_i , which exchanges $-Q_i$ with that reservoir and cancels the action of K on it. (b) After all n cancellations, the composite $K + C_1 + \dots + C_n$ exchanges heat only with the reference reservoir R_0 .

1.3.3 The Clausius inequality for a cycle

Combine K with all the auxiliaries $C_1, \dots, C_n$. Every component is cyclic. For each i , the heat Q_i entering K from R_i cancels the heat $-Q_i$ entering C_i from the same reservoir. The complete heat accounting is

cyclic devices	heat received from each R_i for $i = 1, \dots, n$	heat received from the reference R_0
K	Q_i	0
all auxiliaries $C_1, \dots, C_n$	$-Q_i$	$Q_0 \equiv \sum_i Q_{0i}$
composite	0 for every i	Q_0

The composite therefore has no net heat exchange with any of the reservoirs $R_1, \dots, R_n$. Its only heat exchange is with R_0 , and Eq. (1.11) gives

$$Q_0 = \sum_{i=1}^n Q_{0i} = T_0 \sum_{i=1}^n \frac{Q_i}{T_i}. \quad (1.12)$$

Since the composite is cyclic, its total energy change is zero. The first law says that its net work output equals its only net heat input:

$$W_{\text{comp}} = Q_0.$$

If Q_0 were positive, the composite would take heat Q_0 from the single reservoir R_0 and convert all of it into work. Kelvin–Planck forbids such a cyclic device. Hence $Q_0 \leq 0$. Since $T_0 > 0$, Eq. (1.12) implies

$$\boxed{\sum_{i=1}^n \frac{Q_i}{T_i} \leq 0.} \quad (1.13)$$

This is the Clausius inequality for a cyclic process. The reference temperature T_0 has disappeared, as it must: R_0 was only used to compare heat exchanges occurring at different temperatures.

1.3.4 Reversible cycles define entropy

Suppose now that K itself executes a reversible cycle. It can then be run backward. The reversed cycle exchanges $-Q_i$ with every reservoir, so the Clausius inequality applied to the two directions gives

$$\sum_{i=1}^n \frac{Q_i}{T_i} \leq 0, \quad \sum_{i=1}^n \frac{-Q_i}{T_i} \leq 0.$$

The first inequality says that the sum is nonpositive; the second says that the same sum is nonnegative. Therefore

$$\sum_{i=1}^n \frac{Q_i}{T_i} = 0 \quad \text{for every reversible cycle.} \quad (1.14)$$

A general process may exchange infinitesimal heats through a boundary whose temperature changes continuously. Divide the process into short segments, approximate the boundary temperature on segment i by T_i , and then refine the partition. In this limit the finite sums become

$$\oint \frac{\delta Q}{T_b} \leq 0 \quad \text{for any cycle,} \quad (1.15)$$

and

$$\oint \frac{\delta Q_{\text{rev}}}{T} = 0 \quad \text{for a reversible cycle.} \quad (1.16)$$

Here T_b is the temperature of the reservoir or boundary through which δQ crosses. During a reversible heat exchange, the system temperature equals the boundary temperature, which is why Eq. (1.16) uses T .

Let $\mathcal{P}_1$ and $\mathcal{P}_2$ be two reversible paths from equilibrium state A to equilibrium state B . Follow $\mathcal{P}_1$ from A to B , then follow $\mathcal{P}_2$ backward from B to A . This is a reversible cycle, so

$$0 = \oint \frac{\delta Q_{\text{rev}}}{T} = \int_{\mathcal{P}_1} \frac{\delta Q_{\text{rev}}}{T} - \int_{\mathcal{P}_2} \frac{\delta Q_{\text{rev}}}{T}.$$

The two integrals are equal. Within any connected region of equilibrium state space where such reversible paths exist, the integral depends only on its endpoints. It therefore defines a state function S :

$$S(B) - S(A) = \int_A^B \frac{\delta Q_{\text{rev}}}{T}, \quad dS = \frac{\delta Q_{\text{rev}}}{T}. \quad (1.17)$$

Changing the chosen reference state adds the same constant to every value of S and leaves all entropy differences unchanged.

1.3.5 Irreversible processes and entropy production

Let an actual, possibly irreversible process take the system from equilibrium state A to equilibrium state B . Add a reversible return path from B to A . Together they form a cycle, so Eq. (1.15) gives

$$\int_A^B \frac{\delta Q}{T_b} + \int_B^A \frac{\delta Q_{\text{rev}}}{T} \leq 0.$$

The return path is reversible, and the entropy definition gives

$$\int_B^A \frac{\delta Q_{\text{rev}}}{T} = S(A) - S(B).$$

Substitution yields the Clausius inequality between equilibrium endpoints:

$$S(B) - S(A) \geq \int_A^B \frac{\delta Q}{T_b}. \quad (1.18)$$

The temperature in this integral belongs to the boundary or reservoir. The system need not possess a well-defined temperature while it follows the irreversible path.

A local differential statement needs the system to remain in internal equilibrium, so that an entropy can be assigned at each instant. If heat crosses one boundary at temperature T_b , then

$$dS \geq \frac{\delta Q}{T_b}, \quad (1.19)$$

or, equivalently,

$$dS = \frac{\delta Q}{T_b} + d_i S, \quad d_i S \geq 0. \quad (1.20)$$

The term $d_i S$ is entropy produced inside the system. With several boundaries, the exchange term is $\sum_i \delta Q_i / T_i$. If the system is too far from equilibrium for entropy to be assigned along the path, only the endpoint statement (1.18) is justified.

For an isolated process, no heat crosses the boundary. Equation (1.18) then reduces to

$$\Delta S \geq 0. \quad (1.21)$$

Equality holds for a reversible adiabatic process.

Status: exact

The proof of the cyclic inequality uses Carnot's heat ratio and the Kelvin–Planck statement. It does not use a microscopic definition of entropy. The local entropy-production formula requires internal equilibrium; without it, the endpoint inequality is the valid statement.

Worked example: entropy produced by thermal contact

Take two identical bodies with constant heat capacity C . Initially they have temperatures T_h and $T_c < T_h$; the pair is isolated. Energy conservation fixes the final temperature,

$$T_f = \frac{T_h + T_c}{2}. \quad (1.22)$$

The entropy change cannot be found by integrating the actual heat flow, because the bodies do not pass through equilibrium states while the temperature difference is finite. Entropy is a state function, however, so we may connect the same endpoints by reversible paths:

$$\Delta S_{\text{tot}} = C \ln \frac{T_f}{T_h} + C \ln \frac{T_f}{T_c} = C \ln \frac{(T_h + T_c)^2}{4T_h T_c} \geq 0. \quad (1.23)$$

The last inequality follows from the arithmetic–geometric mean inequality. For $T_h = T + \Delta T/2$ and $T_c = T - \Delta T/2$,

$$\Delta S_{\text{tot}} = \frac{C}{4} \left(\frac{\Delta T}{T} \right)^2 + \mathcal{O}((\Delta T/T)^4). \quad (1.24)$$

Entropy production is therefore second order in a small temperature imbalance even though the heat current is first order.

1.4 The fundamental relation

For a reversible change of a simple system, combine the first law with $\delta Q_{\text{rev}} = T dS$:

$$dE = T dS - P dV + \mu dN. \quad (1.25)$$

Thus

$$T = \left(\frac{\partial E}{\partial S} \right)_{V,N}, \quad P = - \left(\frac{\partial E}{\partial V} \right)_{S,N}, \quad \mu = \left(\frac{\partial E}{\partial N} \right)_{S,V}. \quad (1.26)$$

The entropy representation follows by solving for dS :

$$dS = \frac{1}{T} dE + \frac{P}{T} dV - \frac{\mu}{T} dN. \quad (1.27)$$

1.4.1 Equilibrium between subsystems

Let an isolated composite system contain two weakly coupled parts. Its totals $E = E_A + E_B$, $V = V_A + V_B$, and $N = N_A + N_B$ are fixed, while the parts exchange these quantities. Additivity gives

$$S_{\text{tot}} = S_A(E_A, V_A, N_A) + S_B(E - E_A, V - V_A, N - N_A). \quad (1.28)$$

Equilibrium maximizes S_{tot} . Differentiating with respect to the exchange variables yields

$$\frac{\partial S_{\text{tot}}}{\partial E_A} = \left(\frac{\partial S_A}{\partial E_A} \right)_{V_A, N_A} - \left(\frac{\partial S_B}{\partial E_B} \right)_{V_B, N_B} = \frac{1}{T_A} - \frac{1}{T_B} = 0. \quad (1.29)$$

The volume and particle derivatives have the same subtraction because $V_B = V - V_A$ and $N_B = N - N_A$. Using Eq. (1.27) in all three directions gives

$$\frac{1}{T_A} = \frac{1}{T_B}, \quad \frac{P_A}{T_A} = \frac{P_B}{T_B}, \quad \frac{\mu_A}{T_A} = \frac{\mu_B}{T_B}. \quad (1.30)$$

Therefore $T_A = T_B$, $P_A = P_B$, and $\mu_A = \mu_B$. Temperature, pressure, and chemical potential appear as the variables conjugate to conserved exchanges.

1.4.2 Extensivity and Euler's relation

For a homogeneous bulk system with short-range interactions, doubling (S, V, N) doubles the energy:

$$E(\lambda S, \lambda V, \lambda N) = \lambda E(S, V, N). \quad (1.31)$$

Differentiate the homogeneity relation with respect to λ and set $\lambda = 1$. The chain rule gives

$$S \frac{\partial E}{\partial(\lambda S)} + V \frac{\partial E}{\partial(\lambda V)} + N \frac{\partial E}{\partial(\lambda N)} = E, \quad (\lambda = 1), \quad (1.32)$$

where Eq. (1.25) identifies the three derivatives as T , $-P$, and μ . Hence

$$E = TS - PV + \mu N. \quad (1.33)$$

Differentiating Eq. (1.33) gives

$$dE = T dS + S dT - P dV - V dP + \mu dN + N d\mu. \quad (1.34)$$

Subtracting Eq. (1.25) cancels the $T dS$, $-P dV$, and μdN terms and leaves the Gibbs–Duhem relation

$$S dT - V dP + N d\mu = 0. \quad (1.35)$$

Only two of the three intensive variables (T, P, μ) are independent for a one-component phase.

Limits of extensivity

Surface terms, long-range interactions, gravity, and nonadditive constraints can spoil the simple homogeneity assumption. In such cases Euler's relation and ordinary ensemble equivalence need not hold, and a thermodynamic-limit statement must specify how energy and system size scale.

1.5 Thermodynamic potentials

A laboratory often fixes T rather than S , or P rather than V . Legendre transforms replace inconvenient natural variables by their conjugates.

The minimum principles in this section assume an ordinary reservoir at $T > 0$. This sign matters: the entropy accounting below contains the factor $-1/T$. For a normalizable negative-temperature Gibbs state, possible only for a suitably bounded spectrum, the corresponding free-energy extremum is a maximum rather than a minimum.

The Helmholtz free energy is

$$F(T, V, N) = E - TS, \quad dF = -S dT - P dV + \mu dN. \quad (1.36)$$

At fixed (T, V, N) and $T > 0$, equilibrium minimizes F .

This minimum principle follows directly from the entropy maximum of system plus reservoir. Let the reservoir have energy E_R , temperature T , heat capacity $C_R > 0$, and exchange energy with the system while (V, N) stay fixed. If the system energy changes by ΔE , the reservoir energy changes by $-\Delta E$. First note the curvature identity

$$\frac{\partial}{\partial E} \left(\frac{1}{T} \right) = -\frac{1}{T^2} \frac{\partial T}{\partial E} = -\frac{1}{T^2 C}, \quad C = \left(\frac{\partial E}{\partial T} \right)_{V,N}. \quad (1.37)$$

The second equality uses $\partial T / \partial E = 1/C$ at fixed (V, N) . Taylor expansion of the reservoir entropy now gives

$$\begin{aligned} \Delta S_R &= S_R(E_R - \Delta E) - S_R(E_R) \\ &= -\left(\frac{\partial S_R}{\partial E_R} \right) \Delta E + \frac{1}{2} \left(\frac{\partial^2 S_R}{\partial E_R^2} \right) (\Delta E)^2 + \dots \\ &= -\frac{\Delta E}{T} - \frac{(\Delta E)^2}{2T^2 C_R} + \dots, \end{aligned} \quad (1.38)$$

where $\partial S_R / \partial E_R = 1/T$ follows from Eq. (1.27), and Eq. (1.37) supplies the second derivative. Therefore

$$\Delta S_{\text{tot}} = \Delta S + \Delta S_R = \Delta S - \frac{\Delta E}{T} - \frac{(\Delta E)^2}{2T^2 C_R} + \dots = -\frac{\Delta F}{T} - \frac{(\Delta E)^2}{2T^2 C_R} + \dots. \quad (1.39)$$

In the second equality, $\Delta F \equiv \Delta E - T\Delta S$ uses the fixed bath temperature. The fluctuating system need not possess a temperature away from equilibrium; F means $E - TS$ evaluated with the bath's T . Chapter 3 makes the same point for an arbitrary probability distribution through $\mathcal{F}_T[p]$ and Eq. (3.44). In the ideal-reservoir limit $C_R \rightarrow \infty$, maximizing total entropy is therefore equivalent to minimizing F .

The pressure-controlled principle follows from the same reservoir accounting. If the system gains $(\Delta E, \Delta V)$ while the reservoir loses both, the first-order reservoir entropy change is

$$\Delta S_R = -\frac{\Delta E}{T} - \frac{P_{\text{ext}} \Delta V}{T} + \dots. \quad (1.40)$$

Consequently,

$$\Delta S_{\text{tot}} = -\frac{1}{T} \Delta(F + P_{\text{ext}} V) + \dots. \quad (1.41)$$

Thus, for $T > 0$, equilibrium minimizes $F + P_{\text{ext}} V$, equivalently G , at fixed (T, P_{ext}, N) . Allowing particle exchange in the same way gives the Φ minimum principle at fixed (T, V, μ) .

The enthalpy and Gibbs free energy are

$$H(S, P, N) = E + PV, \quad dH = T dS + V dP + \mu dN, \quad (1.42)$$

$$G(T, P, N) = E - TS + PV, \quad dG = -S dT + V dP + \mu dN. \quad (1.43)$$

For a homogeneous one-component phase, $G = \mu N$.

The grand potential is

$$\Phi(T, V, \mu) = E - TS - \mu N, \quad d\Phi = -S dT - P dV - N d\mu. \quad (1.44)$$

Using (1.33), $\Phi = -PV$ for a homogeneous phase.

Equality of mixed derivatives produces Maxwell relations. For example,

$$\left(\frac{\partial S}{\partial V} \right)_{T,N} = \left(\frac{\partial P}{\partial T} \right)_{V,N} \quad (1.45)$$

follows from $F(T, V, N)$.

1.6 Stability, concavity, and response

Entropy maximization imposes curvature conditions. For energy exchange at fixed total energy,

$$\frac{\partial^2 S_{\text{tot}}}{\partial E_A^2} < 0 \quad (1.46)$$

at a stable equilibrium. Since

$$\frac{\partial^2 S}{\partial E^2} = -\frac{1}{T^2 C_V}, \quad (1.47)$$

by Eq. (1.37), the full two-system condition is

$$\frac{\partial^2 S_{\text{tot}}}{\partial E_A^2} = -\frac{1}{T^2} \left(\frac{1}{C_A} + \frac{1}{C_B} \right) < 0. \quad (1.48)$$

For a canonical reservoir, $C_B = C_R \rightarrow \infty$, so the subsystem term must have $C_A > 0$. With a finite reservoir only the sum of inverse heat capacities is constrained; Exercise 1.6 treats that more general stability condition.

At fixed (T, P, N) with $T > 0$, Eq. (1.41) shows that a volume fluctuation minimizes $F + P_{\text{ext}}V$. At equilibrium $P = P_{\text{ext}}$, and the second derivative is

$$\left(\frac{\partial^2 F}{\partial V^2} \right)_{T,N} = - \left(\frac{\partial P}{\partial V} \right)_{T,N} = \frac{1}{V \kappa_T}.$$

Local stability therefore requires positive isothermal compressibility,

$$\kappa_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_{T,N} > 0. \quad (1.49)$$

In the canonical ensemble these inequalities will become positivity statements for variances.

The full local stability criterion is a matrix statement. In the entropy representation, the Hessian of $S(E, V, N)$ must be negative semidefinite along directions compatible with the constraints. In the canonical description, the Hessian of $F(T, V, N)$ is mixed: for example,

$$\left(\frac{\partial^2 F}{\partial T^2} \right)_{V,N} = -\frac{C_V}{T}, \quad \left(\frac{\partial^2 F}{\partial V^2} \right)_{T,N} = \frac{1}{V \kappa_T}. \quad (1.50)$$

Hence F is concave in T and convex in V . Saying only that “free energy is convex” loses the information carried by its natural variables.

Worked example: a measurable entropy difference

For one mole of an ideal gas, take a reversible path from (T_1, V_1) to (T_2, V_2) . With constant molar heat capacity C_V ,

$$dS = \frac{C_V}{T} dT + \frac{R}{V} dV, \quad \Delta S = C_V \ln \frac{T_2}{T_1} + R \ln \frac{V_2}{V_1}. \quad (1.51)$$

The first term can be obtained calorimetrically; the second follows from a reversible isothermal expansion. No microscopic model was used. Later the Sackur–Tetrode formula will reproduce this difference and will also supply an absolute entropy once the quantum phase-space cell is specified.

Physical picture

Thermodynamic potentials are not competing definitions of energy. Each answers the same optimization question with different external controls. Entropy is maximized for an isolated composite system. For $T > 0$, F is minimized at fixed (T, V, N) , G at fixed (T, P, N) , and Φ at fixed (T, V, μ) .

1.7 Third law and residual entropy

The third law concerns the low-temperature limit. In its entropy form, it states that the entropy approaches a constant as $T \rightarrow 0$ for systems satisfying suitable equilibrium conditions. If the ground state is unique and the limit is regular, statistical mechanics will give $S \rightarrow 0$. A finite ground-state degeneracy g_0 instead gives

$$S(0) = k_B \ln g_0. \quad (1.52)$$

This is residual entropy. What vanishes more generally in an extensive system with subexponential ground-state degeneracy is the entropy density.

1.8 Exercises

Exercise 1.1. Prove the cyclic identity

$$\left(\frac{\partial x}{\partial y}\right)_z \left(\frac{\partial y}{\partial z}\right)_x \left(\frac{\partial z}{\partial x}\right)_y = -1, \quad (1.53)$$

and use it to derive $C_P - C_V = TV\alpha^2/\kappa_T$, where $\alpha = V^{-1}(\partial V/\partial T)_P$.

Exercise 1.2. Starting from $dF = -S dT - P dV + \mu dN$, derive all three Maxwell relations associated with F and state the variables held fixed in each derivative.

Exercise 1.3. Two bodies with constant heat capacities C_A and C_B start at temperatures T_A and T_B and are placed in thermal contact inside an isolated container. Find the final temperature and the total entropy change. Show directly that the change is nonnegative.

Exercise 1.4. Assume $E(S, V, N)$ is homogeneous of degree one. Derive Euler's relation and the Gibbs–Duhem relation. Then explain which step fails for a self-gravitating gas.

Exercise 1.5. For a paramagnet with work term $-M dB$, write the fundamental relation, identify the magnetic conjugate variables, and construct the potential whose natural variables include (T, B) .

Exercise 1.6 (*). Show that concavity of $S(E, V, N)$ implies positivity of C_V and κ_T for an ordinary additive system. Carefully identify where additivity enters.

Exercise 1.7. A system has g_0 exactly degenerate ground states and a gap Δ to the first excited level of degeneracy g_1 . Find the leading low-temperature correction to $S(T)$.

1.9 Further reading

Callen gives the cleanest axiomatic account of the fundamental relation and stability [8]. Kardar emphasizes the bridge to microscopic counting [23]. Pathria and Reichl provide broader collections of examples [42, 47].

Chapter 2

Probability, information, and typicality

Lecture 5, 10, and the first part of 17. Probability for many-body systems; Shannon entropy, relative entropy, maximum entropy, correlations, and a first encounter with large deviations.

What you should be able to do

- Distinguish an arithmetic sample average from a probability expectation.
- Use moments, cumulants, Markov's inequality, and Chebyshev's inequality without treating them as black boxes.
- Quantify the concentration of extensive observables around their mean values.
- Separate the probability calculus from frequentist and Bayesian interpretations, and state which physical assumptions enter an ensemble.
- Prove positivity of relative entropy and use it to solve constrained inference problems.
- Explain typicality without identifying “typical” with “certain.”
- Read a rate function as the cost of a macroscopic fluctuation.

2.1 Probability enters through a description

Statistical mechanics starts with a physical description: a state space, a Hamiltonian, conserved quantities, a preparation procedure, and any relevant couplings to reservoirs. Within that description, a probability distribution assigns weights to the microscopic alternatives. We first set up the common probability calculus and then ask what those weights represent.

A discrete random variable X assigns a numerical value to each possible outcome of an experiment. If its possible values are $x_1, x_2, \dots$ and

$$\mathbb{P}(X = x_i) = p_i,$$

then $p_i \geq 0$ and $\sum_i p_i = 1$. The expectation of a function of X is

$$\langle f(X) \rangle = \sum_i p_i f(x_i). \tag{2.1}$$

Angle brackets always denote a probability expectation. In particular,

$$\langle X \rangle = \sum_i p_i x_i, \quad \langle X^2 \rangle = \sum_i p_i x_i^2.$$

The variance measures the mean squared distance from the expectation:

$$\text{Var}(X) \equiv \langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - \langle X \rangle^2. \quad (2.2)$$

The last equality follows by expanding the square and using linearity of the expectation:

$$\langle (X - \langle X \rangle)^2 \rangle = \langle X^2 \rangle - 2\langle X \rangle \langle X \rangle + \langle X \rangle^2.$$

For two random variables X and Y , the covariance is

$$\text{Cov}(X, Y) = \langle XY \rangle - \langle X \rangle \langle Y \rangle. \quad (2.3)$$

It measures the part of their fluctuations that varies together. Direct expansion gives

$$\text{Var}(aX + bY) = a^2 \text{Var}(X) + b^2 \text{Var}(Y) + 2ab \text{Cov}(X, Y). \quad (2.4)$$

The joint distribution is $p_{XY}(x, y) = \mathbb{P}(X = x, Y = y)$. Its marginals are obtained by summing over the other variable, for example $p_X(x) = \sum_y p_{XY}(x, y)$. If $p_X(x) > 0$, the conditional probability of $Y = y$ given $X = x$ is

$$p_{Y|X}(y|x) = \frac{p_{XY}(x, y)}{p_X(x)}, \quad p_{XY}(x, y) = p_X(x)p_{Y|X}(y|x).$$

The variables are independent when

$$p_{XY}(x, y) = p_X(x)p_Y(y).$$

Independence implies $\text{Cov}(X, Y) = 0$ because the expectation of XY factorizes. The converse fails: covariance detects only one kind of dependence.

2.1.1 Frequency, inference, and physical ensembles

The mathematical rules of probability do not determine what a probability represents. In a frequentist interpretation, probabilities describe the behavior of a repeatable experiment or sampling procedure. In Bayesian inference, probability also quantifies uncertainty about an unknown hypothesis or parameter, conditional on a stated model and the available information. Both interpretations use the same rules for conditional probability.

Let H range over mutually exclusive and exhaustive hypotheses, and let D denote observed data with $p(D) > 0$. The two factorizations of their joint probability give

$$p(H | D)p(D) = p(H, D) = p(D | H)p(H). \quad (2.5)$$

Therefore

$$p(H | D) = \frac{p(D | H)p(H)}{p(D)}, \quad p(D) = \sum_H p(D | H)p(H). \quad (2.6)$$

For a continuous parameter, the sum becomes an integral. Equation (2.6) is Bayes' theorem, an identity within the probability calculus. Bayesian inference adds an interpretation: $p(H)$ is a prior, $p(D | H)$ is the likelihood, and $p(H | D)$ is the posterior state of knowledge after observing D .

The frequentist connection with repeated trials can be stated without identifying model probability with an observed frequency. If an event $\mathcal{E}$ occurs $N_{\mathcal{E}}$ times in N independent repetitions, its empirical frequency is

$$f_N(\mathcal{E}) = \frac{1}{N} \sum_{a=1}^N \mathbf{1}_{\mathcal{E}}(X_a) = \frac{N_{\mathcal{E}}}{N}. \quad (2.7)$$

The weak law proved in section 2.2 shows, under the stated independence and finite-variance assumptions, that $f_N(\mathcal{E})$ converges in probability to $p(\mathcal{E})$. In frequentist inference an unknown model parameter remains fixed; the data, estimators, and confidence procedures vary across hypothetical repetitions.

For comparison, suppose independent Bernoulli trials have an unknown success probability θ . If m successes occur in N trials and $x = m/N$, the likelihood is

$$p(m \mid \theta, N) = \binom{N}{m} \theta^m (1 - \theta)^{N-m}. \quad (2.8)$$

A frequentist estimator is $\hat{\theta} = x$. A Bayesian assigns a prior density $\pi(\theta)$ and obtains

$$\pi(\theta \mid m, N) \propto \pi(\theta) \theta^m (1 - \theta)^{N-m}. \quad (2.9)$$

The first construction studies the sampling behavior of x at fixed θ . The second conditions on the observed value of m and assigns a posterior density to θ . The binomial model in section 2.3 will show how the same relative entropy controls their regular large- N behavior.

Statistical mechanics uses probability in more than one of these senses. An ensemble may describe relative frequencies across repeated preparations, incomplete microscopic information under macroscopic constraints, or a measure with respect to which certain behavior is typical. The canonical distribution will be derived from reservoir counting in section 3.3 and from constrained inference in section 2.7. The two routes give the same distribution but answer different questions.

What must accompany a probability distribution?

Whenever probabilities enter, one should name the physical state space, the preparation procedure, the retained constraints, and any dynamical assumptions. These notes do not require a single interpretation in every setting. *Interpretive choices should not replace the physical argument.*

2.1.2 Generating functions and cumulants

The r th moment of X is $\langle X^r \rangle$. Instead of listing all moments separately, package them into the moment-generating function

$$M_X(t) = \langle e^{tX} \rangle, \quad (2.10)$$

provided the expectation is finite for t in a neighborhood of zero. Expanding the exponential gives

$$M_X(t) = \sum_{r=0}^{\infty} \frac{t^r}{r!} \langle X^r \rangle, \quad M_X^{(r)}(0) = \langle X^r \rangle. \quad (2.11)$$

The cumulant-generating function is the logarithm of M_X :

$$K_X(t) = \ln M_X(t) = \sum_{r=1}^{\infty} \frac{t^r}{r!} \kappa_r(X). \quad (2.12)$$

This equation defines the cumulants $\kappa_r(X)$; equivalently, $\kappa_r(X) = K_X^{(r)}(0)$. They are combinations of moments in which contributions that factor into lower moments have been removed. The first four are

$$\begin{aligned}\kappa_1(X) &= \langle X \rangle, \\ \kappa_2(X) &= \text{Var}(X), \\ \kappa_3(X) &= \langle (X - \langle X \rangle)^3 \rangle, \\ \kappa_4(X) &= \langle (X - \langle X \rangle)^4 \rangle - 3 \text{Var}(X)^2.\end{aligned}\tag{2.13}$$

The first two formulas follow directly from differentiating the logarithm. Since $M_X(0) = 1$,

$$K'_X(0) = \left. \frac{M'_X(t)}{M_X(t)} \right|_{t=0} = \langle X \rangle, \tag{2.14}$$

$$K''_X(0) = \left[\frac{M''_X(t)}{M_X(t)} - \left(\frac{M'_X(t)}{M_X(t)} \right)^2 \right] \Big|_{t=0} = \langle X^2 \rangle - \langle X \rangle^2. \tag{2.15}$$

Why take the logarithm? Let X and Y be independent. Then

$$M_{X+Y}(t) = \langle e^{tX} e^{tY} \rangle = M_X(t) M_Y(t),$$

so

$$K_{X+Y}(t) = K_X(t) + K_Y(t), \quad \kappa_r(X+Y) = \kappa_r(X) + \kappa_r(Y). \tag{2.16}$$

Cumulants are therefore additive for independent contributions. This is why they are useful for many-body observables.

If $S_N = \sum_{a=1}^N X_a$ is a sum of independent identically distributed random variables, then

$$M_{S_N}(t) = \langle e^{t \sum_a X_a} \rangle = \prod_{a=1}^N \langle e^{t X_a} \rangle = [M_X(t)]^N, \tag{2.17}$$

because independence factorizes the expectation. Taking a logarithm gives

$$K_{S_N}(t) = N K_X(t), \quad \kappa_r(S_N) = N \kappa_r(X). \tag{2.18}$$

The variance $\kappa_2(S_N)$ grows as N , so the standard deviation grows as $\sqrt{N}$. Dividing a higher cumulant by the matching power of the standard deviation gives

$$\frac{\kappa_r(S_N)}{\kappa_2(S_N)^{r/2}} = N^{1-r/2} \frac{\kappa_r(X)}{\kappa_2(X)^{r/2}}.$$

For $r = 3$ this ratio is the skewness; for $r = 4$ it is the excess kurtosis. Both vanish as $N \rightarrow \infty$ when the underlying cumulants are finite. This scaling is consistent with Gaussian fluctuations, but it is not by itself a proof of convergence in distribution.

For a Bernoulli variable $B \in \{0, 1\}$ with $\mathbb{P}(B = 1) = p$,

$$K_B(t) = \ln(1 - p + pe^t), \quad \kappa_1 = p, \quad \kappa_2 = p(1 - p), \quad \kappa_3 = p(1 - p)(1 - 2p). \tag{2.19}$$

The number of successes in N independent Bernoulli trials is a sum of N such variables, so each of these cumulants is multiplied by N .

2.2 Sample averages and concentration inequalities

Let $X_1, \dots, X_N$ be independent copies of a random variable with

$$\langle X_a \rangle = \mu, \quad \text{Var}(X_a) = \sigma^2.$$

Their sample mean is the random variable

$$\bar{X}_N = \frac{1}{N} \sum_{a=1}^N X_a. \quad (2.20)$$

The bar and the angle brackets are different operations

The bar in $\bar{X}_N$ means an arithmetic average over the N variables in one sample. Before the sample is observed, $\bar{X}_N$ is still random. The brackets $\langle \bar{X}_N \rangle$ mean the expectation of that random sample mean over repeated samples. Thus there is no duplicate notation for the same operation.

Linearity of expectation gives

$$\langle \bar{X}_N \rangle = \frac{1}{N} \sum_{a=1}^N \langle X_a \rangle = \mu. \quad (2.21)$$

For the variance, use Eq. (2.4) repeatedly, or write the full covariance sum:

$$\text{Var}(\bar{X}_N) = \frac{1}{N^2} \sum_{a,b=1}^N \text{Cov}(X_a, X_b) \quad (2.22)$$

$$= \frac{1}{N^2} \sum_{a=1}^N \text{Var}(X_a) = \frac{\sigma^2}{N}. \quad (2.23)$$

The second line uses independence: $\text{Cov}(X_a, X_b) = 0$ when $a \neq b$. This calculation is the origin of the familiar $N^{-1/2}$ scale for fluctuations of an average.

Variance alone does not yet bound the probability of a large fluctuation. The needed step is Markov's inequality.

Proposition 2.1 (Markov inequality). *If $Y \geq 0$ and $a > 0$, then*

$$\mathbb{P}(Y \geq a) \leq \frac{\langle Y \rangle}{a}. \quad (2.24)$$

Proof. Let $\mathbf{1}_{Y \geq a}$ equal 1 when $Y \geq a$ and 0 otherwise. Pointwise,

$$Y \geq a \mathbf{1}_{Y \geq a}.$$

Taking expectations preserves the inequality and gives

$$\langle Y \rangle \geq a \langle \mathbf{1}_{Y \geq a} \rangle = a \mathbb{P}(Y \geq a).$$

Division by a proves Eq. (2.24). □

Chebyshev's inequality follows by applying Markov's inequality to the nonnegative random variable $Y = (X - \langle X \rangle)^2$ with threshold $a = \epsilon^2$:

$$\mathbb{P}(|X - \langle X \rangle| \geq \epsilon) = \mathbb{P}((X - \langle X \rangle)^2 \geq \epsilon^2) \leq \frac{\text{Var}(X)}{\epsilon^2}. \quad (2.25)$$

Applying this result to $X = \bar{X}_N$ and using Eq. (2.23) yields

$$\mathbb{P}(|\bar{X}_N - \mu| \geq \epsilon) \leq \frac{\sigma^2}{N\epsilon^2}. \quad (2.26)$$

For every fixed $\epsilon > 0$, the right-hand side tends to zero. This is convergence in probability of $\bar{X}_N$ to μ , also called the weak law of large numbers under the assumptions used here.

The same scaling can be stated for the extensive sum $S_N = \sum_a X_a$. Its expectation is $N\mu$ and its standard deviation is $\sigma\sqrt{N}$. If $\mu \neq 0$, the relative width is

$$\frac{\sqrt{\text{Var}(S_N)}}{|\langle S_N \rangle|} = \frac{\sigma}{|\mu| \sqrt{N}}. \quad (2.27)$$

When $\mu = 0$, this ratio is undefined; one should compare the width with another physically relevant scale.

For independent identically distributed variables with finite nonzero variance, the Lindeberg–Lévy central limit theorem refines this statement:

$$\frac{\sum_{a=1}^N X_a - N\mu}{\sigma\sqrt{N}} \xrightarrow{\text{distribution}} \mathcal{N}(0, 1). \quad (2.28)$$

Convergence in distribution means that probabilities computed from the rescaled sum approach those of a standard Gaussian at continuity points of the Gaussian distribution. We quote this theorem without proof; Ref. [9] gives a proof and variants. It describes fluctuations of size $\sqrt{N}$ around the mean. Fluctuations in which the average itself differs from μ by an amount of order one require large-deviation theory.

Quoted theorem

The Markov and Chebyshev inequalities and the weak-law bound were proved above. The central limit theorem is quoted. Its finite-variance and independence assumptions matter; heavy-tailed variables and strongly correlated systems can have different scaling.

2.3 Binomial model: multiplicity becomes entropy

Let $B_1, \dots, B_N$ be independent Bernoulli variables:

$$\mathbb{P}(B_a = 1) = p, \quad \mathbb{P}(B_a = 0) = 1 - p.$$

The random variable $M_N = \sum_a B_a$ counts the number of ones. We use m for one possible observed value of M_N and $x = m/N$ for the corresponding observed fraction.

To obtain $M_N = m$, one must choose which m of the N trials produced a one. There are $\binom{N}{m} = N!/[m!(N-m)!]$ such choices. Each chosen sequence has probability $p^m(1-p)^{N-m}$, so

$$P_N(m) \equiv \mathbb{P}(M_N = m) = \binom{N}{m} p^m (1-p)^{N-m}. \quad (2.29)$$

The binomial theorem checks the normalization:

$$\sum_{m=0}^N P_N(m) = [p + (1-p)]^N = 1.$$

The cumulants in Eq. (2.19) are additive, hence

$$\langle M_N \rangle = Np, \quad \text{Var}(M_N) = Np(1-p), \quad \text{Var}(M_N/N) = \frac{p(1-p)}{N}. \quad (2.30)$$

The exact distribution also describes fluctuations much larger than the standard deviation. Let $x = m/N$ and assume that x stays away from 0 and 1 as N grows. Stirling's formula is

$$\ln N! = N \ln N - N + \frac{1}{2} \ln(2\pi N) + \mathcal{O}(N^{-1}), \quad (2.31)$$

and therefore

$$\ln \binom{N}{Nx} = \ln N! - \ln(Nx)! - \ln[N(1-x)]! \quad (2.32)$$

$$= N \ln N - Nx \ln(Nx) - N(1-x) \ln[N(1-x)] - \frac{1}{2} \ln[2\pi Nx(1-x)] + \mathcal{O}(N^{-1}) \quad (2.33)$$

$$= Nh_2(x) - \frac{1}{2} \ln[2\pi Nx(1-x)] + \mathcal{O}(N^{-1}), \quad (2.34)$$

where the terms $N \ln N$ and $-N$ cancel, and

$$h_2(x) = -x \ln x - (1-x) \ln(1-x). \quad (2.35)$$

Thus the multiplicity grows exponentially as $\exp[Nh_2(x)]$. The probability also contains the weight of each individual sequence:

$$\frac{1}{N} \ln P_N(Nx) = h_2(x) + x \ln p + (1-x) \ln(1-p) + o(1) \quad (2.36)$$

$$= -D(x||p) + o(1), \quad (2.37)$$

where

$$D(x||p) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}. \quad (2.38)$$

This function measures the exponential cost of observing a fraction x when the underlying success probability is p . Indeed,

$$\frac{\partial D}{\partial x} = \ln \frac{x(1-p)}{p(1-x)}, \quad \frac{\partial^2 D}{\partial x^2} = \frac{1}{x(1-x)} > 0. \quad (2.39)$$

The unique minimum is at $x = p$, where $D(p||p) = 0$. Expanding about that minimum gives

$$D(x||p) = \frac{(x-p)^2}{2p(1-p)} + \mathcal{O}((x-p)^3), \quad (2.40)$$

which agrees with the Gaussian variance in Eq. (2.30) near the mean.

At leading exponential order,

$$P_N(Nx) \asymp e^{-ND(x||p)}, \quad (2.41)$$

for values of x such that Nx is an integer. The notation $\asymp$ means logarithmic equivalence:

$$\lim_{N \rightarrow \infty} \frac{1}{N} \ln P_N(Nx) = -D(x||p). \quad (2.42)$$

Keeping the subextensive term in Eq. (2.34) gives the sharper estimate

$$P_N(Nx) \simeq \frac{e^{-ND(x||p)}}{\sqrt{2\pi Nx(1-x)}}. \quad (2.43)$$

Status: thermodynamic-limit statement

Equation (2.41) keeps the part of $\ln P_N$ proportional to N . Equation (2.43) also keeps the leading prefactor. Neither formula is an exact finite- N identity; the exact probability is Eq. (2.29).

Bayesian concentration and the binomial rate function

If the Bernoulli parameter is unknown, rename it θ and assign a prior density $\pi(\theta)$. For the observed fraction $x = m/N$, the θ -dependent part of the likelihood in Eq. (2.8) obeys the exact identity

$$\theta^m (1 - \theta)^{N-m} = \exp\{-N[h_2(x) + D(x||\theta)]\}. \quad (2.44)$$

Because $h_2(x)$ does not depend on θ , Bayes' theorem gives

$$\pi(\theta | m, N) \propto \pi(\theta) e^{-ND(x||\theta)}. \quad (2.45)$$

Thus $D(x||\theta)$ has two readings in the same model. At fixed θ , it is the large-deviation cost of observing x . At fixed observed x , it suppresses parameter values θ that poorly fit the data. The same divergence will acquire a hypothesis-testing meaning in [section 2.5](#).

Suppose $0 < x < 1$ and the prior is fixed, continuous, and nonzero near x . Expansion with respect to θ gives

$$D(x||\theta) = \frac{(\theta - x)^2}{2x(1-x)} + \mathcal{O}((\theta - x)^3). \quad (2.46)$$

The posterior then concentrates near x with width of order $N^{-1/2}$, and its mode is $x + \mathcal{O}(N^{-1})$ for a smooth nonuniform prior. Frequentist and Bayesian uncertainty estimates have the same leading Gaussian scale in this regular interior regime, although their definitions remain different. The conclusion can fail at $x = 0$ or $x = 1$, or when the prior vanishes near x or excludes it from its support.

Worked example: how unlikely is a macroscopic imbalance?

For $N = 1000$ independent trials with $p = 0.3$, ask for the probability of observing $x = 0.4$. The rate is

$$D(0.4||0.3) = 0.4 \ln \frac{4}{3} + 0.6 \ln \frac{6}{7} \simeq 2.26 \times 10^{-2}. \quad (2.47)$$

At exponential accuracy, $P_N(400) \asymp \exp(-22.6)$. Including the Stirling prefactor gives

$$P_N(400) \simeq \frac{\exp[-ND(0.4||0.3)]}{\sqrt{2\pi Nx(1-x)}} \simeq 4.0 \times 10^{-12}. \quad (2.48)$$

This is a probability for exactly 400 successes. A tail probability, such as $\mathbb{P}(m \geq 400)$, contains many terms and is larger, but has the same leading exponential rate. The distinction between a point probability and a tail is easy to lose when only the symbol $\asymp$ is retained.

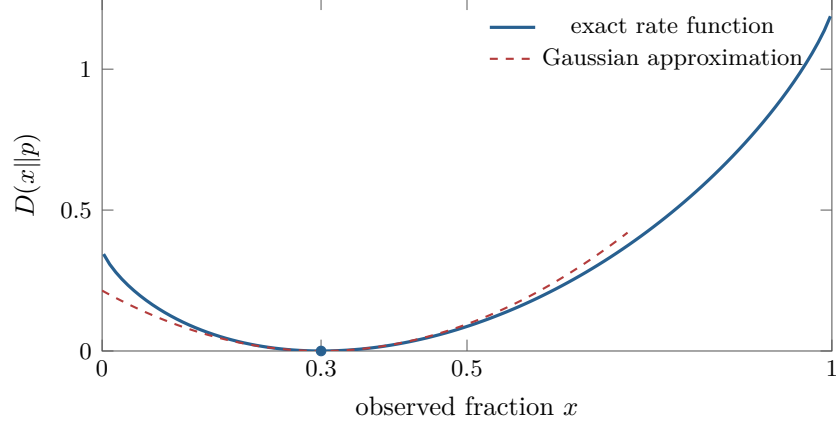


Figure 2.1: The Bernoulli rate function for $p = 0.3$. Its quadratic expansion is accurate near the typical value $x = p$ but underestimates the cost of fluctuations in the tails.

Lecture 5 ends here. The next three sections supply the probability and correlation tools revisited in the first part of Lecture 17.

2.4 Shannon entropy

The binary function $h_2(x)$ arose from counting sequences with two possible symbols. For a distribution $p = (p_1, \dots, p_d)$ on d symbols, the same counting structure leads to the Shannon entropy

$$h(p) = - \sum_{i=1}^d p_i \ln p_i. \quad (2.49)$$

We use the continuous extension $0 \ln 0 = 0$. Natural logarithms measure entropy in *nats*; logarithms to base two measure it in *bits*, with $h_{\text{bits}} = h / \ln 2$.

If outcome i occurs with probability p_i , define its information content, or surprise, as $-\ln p_i$. A rare outcome has greater surprise. Shannon entropy is the expectation of that random quantity:

$$h(p) = \langle -\ln p_X(X) \rangle. \quad (2.50)$$

Here $p_X(X)$ means: evaluate the probability mass function at the value taken by X . This notation does not introduce a second random variable.

Two limiting cases fix the interpretation. If one outcome has probability one, then $h = 0$. If all d outcomes have probability $1/d$, then

$$h = -d \frac{1}{d} \ln \frac{1}{d} = \ln d.$$

These are the minimum and maximum values:

$$0 \leq h(p) \leq \ln d. \quad (2.51)$$

The lower bound follows term by term from $-p_i \ln p_i \geq 0$. For the upper bound, maximize h subject to $\sum_i p_i = 1$. At an interior stationary point,

$$0 = \frac{\partial}{\partial p_i} \left[- \sum_j p_j \ln p_j - \alpha \left(\sum_j p_j - 1 \right) \right] = -\ln p_i - 1 - \alpha,$$

so all p_i are equal and normalization gives $p_i = 1/d$. Moreover,

$$\frac{\partial^2 h}{\partial p_i \partial p_j} = -\frac{\delta_{ij}}{p_i},$$

which is negative definite in the interior. The stationary point is therefore the unique maximum. A distribution supported on fewer than d outcomes has maximum $\ln d'$ with $d' < d$, so it cannot exceed $\ln d$.

Continuity, additivity for independent choices, and a grouping rule characterize Eq. (2.49) up to a positive multiplicative constant [54, 9]. This is a theorem about measures of uncertainty. It is not a thermodynamic law and does not yet contain Boltzmann's constant.

2.4.1 Joint and conditional entropy

For jointly distributed variables (X, Y) , define

$$h(X, Y) = -\sum_{x,y} p_{XY}(x, y) \ln p_{XY}(x, y), \quad (2.52)$$

$$h(Y|X) = -\sum_{x,y} p_{XY}(x, y) \ln p_{Y|X}(y|x). \quad (2.53)$$

The conditional entropy can also be read as an average:

$$h(Y|X) = \sum_x p_X(x) h(p_{Y|X=x}). \quad (2.54)$$

Each entropy inside this average is nonnegative, hence $h(Y|X) \geq 0$ for classical random variables.

Insert $p_{XY}(x, y) = p_X(x)p_{Y|X}(y|x)$ into the joint entropy:

$$h(X, Y) = -\sum_{x,y} p_{XY}(x, y) [\ln p_X(x) + \ln p_{Y|X}(y|x)] \quad (2.55)$$

$$= h(X) + h(Y|X). \quad (2.56)$$

This is the entropy chain rule. If X and Y are independent, then $p_{Y|X}(y|x) = p_Y(y)$ and $h(Y|X) = h(Y)$. Therefore

$$h(X, Y) = h(X) + h(Y) \quad \text{for independent } X, Y. \quad (2.57)$$

2.4.2 Typical sequences

Draw independent symbols $X_1, \dots, X_N$ from p on a finite alphabet. The probability of a particular sequence $x^N = (x_1, \dots, x_N)$ factorizes:

$$p(x^N) = \prod_{a=1}^N p(x_a).$$

Its information per symbol is

$$-\frac{1}{N} \ln p(x^N) = -\frac{1}{N} \sum_{a=1}^N \ln p(x_a). \quad (2.58)$$

The variables $-\ln p(X_a)$ are independent and have expectation $h(p)$ by Eq. (2.50). The weak law of large numbers therefore makes Eq. (2.58) converge in probability to $h(p)$.

For $\epsilon > 0$, the typical set is

$$A_\epsilon^{(N)} = \left\{ x^N : \left| -\frac{1}{N} \ln p(x^N) - h(p) \right| < \epsilon \right\}. \quad (2.59)$$

By the weak law, $\mathbb{P}(A_\epsilon^{(N)}) \rightarrow 1$. Exponentiating the defining inequality shows that every typical sequence obeys

$$e^{-N[h(p)+\epsilon]} < p(x^N) < e^{-N[h(p)-\epsilon]}. \quad (2.60)$$

These probability bounds also determine the number of typical sequences. The lower probability bound and normalization give

$$1 \geq \sum_{x^N \in A_\epsilon^{(N)}} p(x^N) > |A_\epsilon^{(N)}| e^{-N[h(p)+\epsilon]},$$

so $|A_\epsilon^{(N)}| < e^{N[h(p)+\epsilon]}$. For any $\delta > 0$ and sufficiently large N , the typical set has probability greater than $1 - \delta$. Using the upper probability bound gives

$$1 - \delta < \sum_{x^N \in A_\epsilon^{(N)}} p(x^N) < |A_\epsilon^{(N)}| e^{-N[h(p)-\epsilon]},$$

and hence

$$(1 - \delta) e^{N[h(p)-\epsilon]} < |A_\epsilon^{(N)}| < e^{N[h(p)+\epsilon]}. \quad (2.61)$$

This is the precise sense in which most probability lies on about $e^{Nh(p)}$ sequences of probability about $e^{-Nh(p)}$.

What typical does not mean

Typical sequences need not have exactly equal probabilities. Atypical sequences remain possible at every finite N . Typicality says that the total probability of the typical set approaches one, while the probability of its complement approaches zero.

2.5 Relative entropy and distinguishability

For distributions p and q on the same finite set,

$$D(p||q) = \sum_i p_i \ln \frac{p_i}{q_i}. \quad (2.62)$$

We use $0 \ln(0/q_i) = 0$. If $p_i > 0$ while $q_i = 0$ for any i , then $D(p||q) = +\infty$. Relative entropy is the expectation, under p , of the log-likelihood ratio $\ln(p_i/q_i)$.

The proof of its main property starts with an elementary logarithm inequality. For $u > 0$,

$$\ln u \leq u - 1, \quad (2.63)$$

with equality only at $u = 1$. To verify this, define $g(u) = u - 1 - \ln u$. Then $g'(u) = 1 - 1/u$ vanishes only at $u = 1$, while $g''(u) = 1/u^2 > 0$. Thus g has a global minimum $g(1) = 0$.

Proposition 2.2 (Gibbs inequality). *Relative entropy is nonnegative, and $D(p||q) = 0$ exactly when $p = q$.*

Proof. If some $p_i > 0$ and $q_i = 0$, the result is immediate because the relative entropy is infinite. Otherwise apply Eq. (2.63) with $u = q_i/p_i$ wherever $p_i > 0$:

$$-D(p\|q) = \sum_{i:p_i>0} p_i \ln \frac{q_i}{p_i} \quad (2.64)$$

$$\leq \sum_{i:p_i>0} p_i \left(\frac{q_i}{p_i} - 1 \right) \leq \sum_i (q_i - p_i) = 0. \quad (2.65)$$

Equality in both inequalities requires $q_i = p_i$ for every i with $p_i > 0$ and no probability mass of q outside that support. Hence $p = q$. $\square$

The Bernoulli quantity $D(x\|p)$ in Eq. (2.38) is precisely the relative entropy between the binary distributions $(x, 1-x)$ and $(p, 1-p)$. Thus Gibbs' inequality supplies another proof that the binomial rate function is nonnegative and vanishes only at $x = p$.

Relative entropy is asymmetric and does not satisfy the triangle inequality, so it is not a metric. Its order has an operational meaning. The classical Stein lemma states that, at any fixed type-I error below one, the best type-II error exponent for N independent samples is $D(p\|q)$ [9]. We quote that theorem and do not use it in later derivations. Chapter 3 will derive a separate thermodynamic meaning through Eq. (3.44).

2.6 Mutual information and data processing

For jointly distributed variables (X, Y) ,

$$I(X : Y) = D(p_{XY}\|p_X p_Y) = h(X) + h(Y) - h(X, Y). \quad (2.66)$$

The second equality follows by expanding the relative entropy:

$$D(p_{XY}\|p_X p_Y) = \sum_{x,y} p_{XY}(x, y) [\ln p_{XY}(x, y) - \ln p_X(x) - \ln p_Y(y)] \quad (2.67)$$

$$= -h(X, Y) + h(X) + h(Y). \quad (2.68)$$

Gibbs' inequality gives $I(X : Y) \geq 0$. Equality holds exactly when $p_{XY} = p_X p_Y$, which is the condition for independence. Using the chain rule also gives

$$I(X : Y) = h(X) - h(X|Y) = h(Y) - h(Y|X). \quad (2.69)$$

It is the average reduction in uncertainty about either variable after the other one is known.

2.6.1 Stochastic maps and the log-sum inequality

A stochastic map $\mathcal{T}$ sends an input value x to an output value y with conditional probabilities $T_{y|x} \geq 0$ satisfying $\sum_y T_{y|x} = 1$. It maps an input distribution p to

$$(\mathcal{T}p)_y = \sum_x T_{y|x} p_x. \quad (2.70)$$

Discarding detail, adding noise, and recording only a coarse-grained label are all stochastic maps.

For nonnegative numbers a_i and b_i , the log-sum inequality states

$$\sum_i a_i \ln \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \ln \frac{\sum_i a_i}{\sum_i b_i}. \quad (2.71)$$

To prove it, first suppose $A = \sum_i a_i > 0$ and $B = \sum_i b_i > 0$, and set $r_i = a_i/A$ and $s_i = b_i/B$. The difference between the two sides is

$$A \sum_i r_i \ln \frac{r_i}{s_i} = AD(r\|s) \geq 0.$$

If $A = 0$, both sides vanish. If $B = 0 < A$, both sides are infinite under the same support convention used for relative entropy. Thus the remaining boundary cases follow directly or by continuity.

Now let $p' = \mathcal{T}p$ and $q' = \mathcal{T}q$. Insert $\sum_y T_{y|x} = 1$ into the input relative entropy, and apply the log-sum inequality separately for each output y :

$$D(p\|q) = \sum_{x,y} T_{y|x} p_x \ln \frac{T_{y|x} p_x}{T_{y|x} q_x} \quad (2.72)$$

$$\geq \sum_y \left(\sum_x T_{y|x} p_x \right) \ln \frac{\sum_x T_{y|x} p_x}{\sum_x T_{y|x} q_x} \quad (2.73)$$

$$= D(\mathcal{T}p\|\mathcal{T}q). \quad (2.74)$$

Terms with $T_{y|x} = 0$ contribute zero. Equation (2.74) is the data-processing inequality for classical relative entropy.

2.6.2 Markov chains and information loss

A Markov chain $X \rightarrow Y \rightarrow Z$ means that, once Y is known, the rule producing Z does not depend further on X :

$$p_{Z|X,Y}(z|x,y) = p_{Z|Y}(z|y). \quad (2.75)$$

Equivalently, Z is obtained from Y by a stochastic map. The term *Markov chain* is unrelated to *Markov's inequality*; they share a name, not a mathematical role.

Apply the channel $Y \rightarrow Z$ to both distributions in $I(X : Y) = D(p_{XY}\|p_X p_Y)$. It maps them to p_{XZ} and $p_X p_Z$, respectively. Equation (2.74) therefore gives

$$I(X : Z) \leq I(X : Y). \quad (2.76)$$

Processing Y cannot create information about X that was absent before the processing. Appendix A records the same log-sum proof as a reusable mathematical reference.

Status: exact

Data processing is an exact mathematical statement about stochastic maps. Interpreting a particular physical evolution as such a map requires a modeling decision, often involving coarse graining or an open-system description.

Lecture 10 resumes here with constrained inference and large deviations.

2.7 Maximum entropy and exponential families

Suppose a system can occupy states i , but the available information specifies only the expectation values

$$\sum_i p_i A_i^{(a)} = a_a, \quad a = 1, \dots, r, \quad (2.77)$$

along with normalization. The maximum-entropy prescription chooses, among all distributions satisfying these conditions, the one with largest $h(p)$.

Introduce a multiplier α for normalization and multipliers λ_a for the r expectation constraints:

$$\mathcal{L} = - \sum_i p_i \ln p_i - \alpha \left(\sum_i p_i - 1 \right) - \sum_a \lambda_a \left(\sum_i p_i A_i^{(a)} - a_a \right). \quad (2.78)$$

Since $\partial[-p_i \ln p_i]/\partial p_i = -\ln p_i - 1$, stationarity with respect to each p_i requires

$$\frac{\partial \mathcal{L}}{\partial p_i} = -\ln p_i - 1 - \alpha - \sum_a \lambda_a A_i^{(a)} = 0. \quad (2.79)$$

Solving this equation gives $p_i = \exp[-1 - \alpha - \sum_a \lambda_a A_i^{(a)}]$. Normalization determines the factor $\exp(1 + \alpha)$ and yields

$$p_i = \frac{1}{Z(\boldsymbol{\lambda})} \exp \left[- \sum_a \lambda_a A_i^{(a)} \right], \quad Z(\boldsymbol{\lambda}) = \sum_i \exp \left[- \sum_a \lambda_a A_i^{(a)} \right]. \quad (2.80)$$

The multipliers are fixed by demanding that Eq. (2.80) reproduce the specified values a_a .

The partition function generates these averages. Differentiate it with respect to one multiplier:

$$\frac{\partial Z}{\partial \lambda_a} = - \sum_i A_i^{(a)} \exp \left[- \sum_b \lambda_b A_i^{(b)} \right] \quad (2.81)$$

$$= -Z \langle A^{(a)} \rangle. \quad (2.82)$$

Division by Z gives

$$a_a = \langle A^{(a)} \rangle = - \frac{\partial \ln Z}{\partial \lambda_a}. \quad (2.83)$$

For the second derivative, differentiate the normalized average and use the quotient rule:

$$\frac{\partial \langle A^{(a)} \rangle}{\partial \lambda_b} = - \langle A^{(a)} A^{(b)} \rangle + \langle A^{(a)} \rangle \langle A^{(b)} \rangle \quad (2.84)$$

$$= -\text{Cov}(A^{(a)}, A^{(b)}). \quad (2.85)$$

Combining this with Eq. (2.83) gives

$$\frac{\partial^2 \ln Z}{\partial \lambda_a \partial \lambda_b} = \text{Cov}(A^{(a)}, A^{(b)}). \quad (2.86)$$

Thus curvature of $\ln Z$ measures fluctuations of the constrained observables.

The entropy Hessian found below Eq. (2.51) is negative definite in the interior of the probability simplex. Since linear constraints define a convex feasible set, any interior stationary point is the unique global maximum. The following calculation proves the same claim and quantifies the entropy gap.

Let p^* denote the exponential-family solution. From Eq. (2.80),

$$\ln p_i^* = -\ln Z - \sum_a \lambda_a A_i^{(a)}, \quad (2.87)$$

$$h(p^*) = - \sum_i p_i^* \ln p_i^* = \ln Z + \sum_a \lambda_a a_a. \quad (2.88)$$

For any other feasible distribution p , insert the first line into the relative entropy:

$$D(p\|p^*) = \sum_i p_i \ln p_i - \sum_i p_i \ln p_i^* \quad (2.89)$$

$$= -h(p) + \ln Z + \sum_a \lambda_a \sum_i p_i A_i^{(a)} \quad (2.90)$$

$$= -h(p) + \ln Z + \sum_a \lambda_a a_a \quad (2.91)$$

$$= h(p^*) - h(p). \quad (2.92)$$

The third line uses the fact that p satisfies the same constraints as p^* . Gibbs' inequality now gives

$$h(p^*) - h(p) = D(p\|p^*) \geq 0. \quad (2.93)$$

Equality holds only for $p = p^*$. Jaynes proposed this construction as an inference principle for statistical mechanics [19].

Inference is not dynamics

Maximum entropy selects a distribution from stated constraints. It does not prove that a real system approaches that distribution. Thermalization requires a dynamical argument, coupling to a bath, typicality, or an experimentally justified equilibrium assumption.

2.8 Large deviations

The central limit theorem describes fluctuations of the sample average on the scale $\sigma/\sqrt{N}$, so that the extensive sum fluctuates around $N\mu$ on the scale $\sigma\sqrt{N}$. A large-deviation principle asks a different question: how unlikely is it for the average $\bar{X}_N$ itself to differ from μ by an amount that remains finite as N grows?

For a narrow interval around x , the characteristic large- N form is

$$\mathbb{P}(\bar{X}_N \approx x) \asymp e^{-NI(x)}. \quad (2.94)$$

The symbol $\approx$ represents a small interval whose precise treatment belongs to the formal definition of a large-deviation principle. The rate function $I(x)$ records the exponential cost. It is nonnegative and vanishes at the typical value.

The quantity used to calculate I is the scaled cumulant-generating function

$$\lambda(k) = \lim_{N \rightarrow \infty} \lambda_N(k), \quad \lambda_N(k) = \frac{1}{N} \ln \langle e^{kN\bar{X}_N} \rangle. \quad (2.95)$$

For independent identically distributed variables, $N\bar{X}_N = \sum_a X_a$ and factorization gives

$$\lambda_N(k) = \frac{1}{N} \ln \prod_{a=1}^N \langle e^{kX_a} \rangle = \ln M_X(k). \quad (2.96)$$

Thus the scaled cumulant-generating function reduces to the one-variable $K_X(k) = \ln M_X(k)$. In particular,

$$\lambda(0) = 0, \quad \lambda'(0) = \mu, \quad \lambda''(0) = \sigma^2, \quad (2.97)$$

whenever these derivatives exist.

The Gärtner–Ellis theorem states, under differentiability and other regularity hypotheses, that

$$I(x) = \sup_k \{kx - \lambda(k)\}. \quad (2.98)$$

This is the Legendre–Fenchel transform of λ . Choosing $k = 0$ in the supremum shows $I(x) \geq 0$. At a smooth interior optimum,

$$x = \lambda'(k_x), \quad I(x) = k_x x - \lambda(k_x). \quad (2.99)$$

For $x = \mu = \lambda'(0)$, the optimum is $k_x = 0$ and $I(\mu) = 0$.

The binomial model already supplied a concrete large-deviation principle: Eq. (2.41) has rate $I(x) = D(x\|p)$. For a Bernoulli variable,

$$\lambda(k) = \ln(1 - p + pe^k), \quad \lambda'(k) = \frac{pe^k}{1 - p + pe^k}. \quad (2.100)$$

The stationary condition $\lambda'(k_x) = x$ gives

$$e^{k_x} = \frac{x(1-p)}{p(1-x)}, \quad 1 - p + pe^{k_x} = \frac{1-p}{1-x}. \quad (2.101)$$

Substitution into the transform gives

$$I(x) = x \ln \frac{x(1-p)}{p(1-x)} - \ln \frac{1-p}{1-x} \quad (2.102)$$

$$= x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p} \quad (2.103)$$

$$= D(x\|p). \quad (2.104)$$

Near $x = p$, Eq. (2.40) turns $e^{-NI(x)}$ into the Gaussian form predicted by the central limit theorem. Farther into the tails, the full nonquadratic rate function is needed.

Quoted theorem

The Bernoulli large-deviation form was derived directly from the exact binomial distribution. The general Gärtner–Ellis theorem is quoted; its differentiability and regularity conditions cannot be dropped without changing the conclusion. Ref. [58] gives the precise statement and examples.

Exercise 2.9 applies the same construction to a biased random walk. The transform reappears in Chapter 3, where Eq. (3.48) connects microcanonical entropy to canonical free energy.

2.9 Continuous variables and reference measures

For a continuous variable, probabilities belong to intervals:

$$\mathbb{P}(a \leq X \leq b) = \int_a^b p_X(x) \, dx, \quad \int p_X(x) \, dx = 1.$$

The density $p_X(x)$ has units inverse to those of x . A direct analogue of Shannon entropy is the differential entropy

$$h[p_X] = - \int p_X(x) \ln p_X(x) \, dx. \quad (2.105)$$

Unlike discrete Shannon entropy, this expression depends on the coordinate and may be negative. There is also a dimensional problem: p_X generally has units, so its logarithm needs a reference scale.

The coordinate dependence can be calculated explicitly. Let $Y = f(X)$ with f one-to-one and differentiable. Conservation of probability gives

$$p_Y(y) = p_X(x) \left| \frac{dx}{dy} \right|. \quad (2.106)$$

Changing variables in the entropy integral then gives

$$h[p_Y] = - \int dx p_X(x) \ln \left[p_X(x) \left| \frac{dx}{dy} \right| \right] \quad (2.107)$$

$$= h[p_X] + \langle \ln |f'(X)| \rangle. \quad (2.108)$$

Differential entropy is therefore not coordinate invariant.

Choose instead a reference measure $d\mu(x) = m(x) dx$, where $m(x)$ has the same units as $p_X(x)$. Define the entropy relative to that measure by

$$h_\mu[p_X] = - \int dx p_X(x) \ln \frac{p_X(x)}{m(x)}. \quad (2.109)$$

The ratio inside the logarithm is dimensionless. Under $y = f(x)$, both p and m acquire the same Jacobian, so their ratio and the full integral remain unchanged.

If $m(x)$ is itself a normalized probability density $q(x)$, then $h_\mu[p] = -D(p||q) \leq 0$. A reference measure need not be normalized, however. In that case Eq. (2.109) remains coordinate invariant but has no general sign bound. Keeping these cases separate prevents an apparent contradiction with Gibbs' inequality.

Worked example: why differential entropy needs a reference

Let X be uniform on an interval of length L , so its density is $p_X(x) = 1/L$. The differential entropy is $h[p_X] = \ln L$, which is already problematic because the logarithm acts on a dimensional length. Rescale to $Y = X/L$. Then Y is uniform on $[0, 1]$ and $h[p_Y] = 0$. Nothing physical changed; only the coordinate did.

Equation (2.108) shows the change directly: $f'(x) = 1/L$, so $h[p_Y] = h[p_X] - \ln L = 0$. Choose a reference measure with constant density $m(x) = 1/L_0$. Equation (2.109) gives

$$h_\mu[p_X] = - \int_0^L \frac{dx}{L} \ln \frac{1/L}{1/L_0} = \ln \frac{L}{L_0}. \quad (2.110)$$

The result is dimensionless and records the chosen resolution L_0 . The measure $m(x) dx$ is not a normalized probability measure on the whole line, so h_μ may have either sign.

In classical phase space, Liouville measure supplies the coordinate-invariant volume element. The cell h^{3N} makes that volume dimensionless, while the factor $N!$ removes permutations of identical particles. Later chapters will therefore use the reference measure

$$d\mu(\Gamma) = \frac{d^{3N}q d^{3N}p}{N! h^{3N}}. \quad (2.111)$$

The appearance of h here anticipates quantum mechanics; it fixes the phase-space resolution rather than changing the classical equations of motion.

2.10 Exercises

Exercise 2.1. Starting from the pointwise inequality $Y \geq a \mathbf{1}_{Y \geq a}$ for $Y \geq 0$, rederive Markov's inequality. Apply it to $(X - \langle X \rangle)^2$ to obtain Chebyshev's inequality, then apply Chebyshev to the sample mean of N independent variables with variance σ^2 .

Exercise 2.2. For N independent spins with probabilities p and $1 - p$, calculate the mean, variance, and third cumulant of the number of up spins. Determine how the skewness scales with N .

Exercise 2.3. Use Stirling's approximation to derive (2.41), including the leading prefactor. Expand $D(x||p)$ to quadratic order around $x = p$ and recover the Gaussian variance.

Exercise 2.4. Prove $h(p) \leq \ln d$ by writing $D(p||u) \geq 0$, where $u_i = 1/d$.

Exercise 2.5. Let X be a fair bit and let $Y = X \oplus Z$, where Z is an independent bit with $\mathbb{P}(Z = 1) = \epsilon$. Compute $I(X : Y)$ and examine the limits $\epsilon = 0$ and $\epsilon = 1/2$.

Exercise 2.6. Maximize Shannon entropy subject to fixed mean and variance on the real line, using Lebesgue measure as the reference. Show that the result is Gaussian. Which part of the calculation depends on the reference measure?

Exercise 2.7. Derive the canonical distribution by maximizing $h(p)$ at fixed mean energy. Identify the multiplier β by comparing two weakly coupled systems that exchange energy.

Exercise 2.8 (*). Prove the log-sum inequality and use it to establish the data-processing inequality for a finite stochastic matrix.

Exercise 2.9. For a biased random walk, calculate the rate function for the displacement per step. Plot its qualitative shape and compare its quadratic approximation with the exact function.

2.11 Further reading

Cover and Thomas develop the information identities and typical-set arguments [9]. Witten gives a short physics-oriented route through the same ideas [62]. Touchette reviews large deviations and their role in ensemble theory [58].

Chapter 3

Equilibrium ensembles

Lecture 6–10. Microcanonical, canonical, and grand-canonical ensembles; partition functions, fluctuations, variational principles, and ensemble equivalence.

What you should be able to do

- Derive each ensemble from its physical constraints.
- Extract thermodynamics and fluctuations from partition functions.
- State conditions for ensemble equivalence instead of treating it as automatic.
- Use relative entropy to prove the canonical variational principle.

3.1 Classical state counting

For N classical particles in d dimensions, a microstate is a point

$$\Gamma = (\mathbf{q}_1, \dots, \mathbf{q}_N; \mathbf{p}_1, \dots, \mathbf{p}_N), \quad (3.1)$$

in a $2dN$ -dimensional phase space. The Hamiltonian $H_N(\Gamma)$ assigns an energy to each point. The natural volume element on this space is the Liouville measure

$$d\Gamma = \prod_{a=1}^N d^d q_a d^d p_a. \quad (3.2)$$

For identical particles, the semiclassical counting measure is

$$\frac{d\Gamma}{N! h^{dN}}. \quad (3.3)$$

The factor $N!$ removes permutations of identical particles. The factor h^{dN} divides phase space into cells with the dimensions of action and makes the count dimensionless; its value records the quantum origin of the classical limit. Classical mechanics alone fixes neither this cell size nor the additive constant in the resulting entropy.

3.2 Microcanonical ensemble

An isolated system has fixed (E, V, N) up to a small energy resolution ΔE . Define the shell count

$$\Omega(E, V, N; \Delta E) = \frac{1}{N! h^{dN}} \int_V d\Gamma \mathbf{1}_{[E, E+\Delta E]}(H_N(\Gamma)). \quad (3.4)$$

Here $\mathbf{1}_A(x)$ is one when $x \in A$ and zero otherwise. Thus Ω counts the phase-space cells whose energies lie between E and $E + \Delta E$. The width must be large enough to contain many states but small on the macroscopic energy scale. Its precise value changes the entropy only by an additive term $k_B \ln \Delta E$ when the density of states varies slowly across the shell.

The equilibrium postulate for an isolated system assigns equal weight, with respect to the counting measure (3.3), to every microstate in the allowed shell. The resulting density with respect to the bare phase-space element $d\Gamma$ is

$$\rho_{\text{mc}}(\Gamma) = \frac{\mathbf{1}_{[E, E+\Delta E]}(H_N(\Gamma))}{N! h^{dN} \Omega(E, V, N; \Delta E)}. \quad (3.5)$$

Indeed,

$$\int \rho_{\text{mc}}(\Gamma) d\Gamma = \frac{1}{N! h^{dN} \Omega} \int \mathbf{1}_{[E, E+\Delta E]}(H_N) d\Gamma = 1.$$

Boltzmann entropy is

$$S_B(E, V, N) = k_B \ln \Omega(E, V, N; \Delta E). \quad (3.6)$$

An alternative uses the cumulative phase volume

$$\Sigma(E) = \frac{1}{N! h^{dN}} \int_V d\Gamma \Theta(E - H_N(\Gamma)), \quad (3.7)$$

where Θ is the Heaviside step function. Its derivative $\omega(E) = \partial_E \Sigma(E)$ is the density of states, so that $\omega(E)\Delta E$ approximates the shell count. The two entropies $k_B \ln \Sigma$ and $k_B \ln[\omega\Delta E]$ differ by subextensive terms for many ordinary systems. To see why, suppose $\omega(Ne) \sim \exp[Ns(e)/k_B]$ and $s(e)$ increases over the relevant range. In

$$\Sigma(Ne) = N \int^e de' \omega(Ne'),$$

the exponential is largest at the upper endpoint. Laplace's method then gives $\ln \Sigma(Ne) = Ns(e)/k_B + \mathcal{O}(\ln N)$, the same extensive term as $\ln[\omega(Ne)\Delta E]$. The two definitions need not agree for small systems, bounded spectra near an entropy maximum, or any setting where this endpoint argument fails.

Identifying S_B with the macroscopic equilibrium entropy is a physical postulate, not a consequence of state counting alone. Under that identification, Eq. (1.27) gives

$$\frac{1}{T} = \left(\frac{\partial S_B}{\partial E} \right)_{V, N}, \quad \frac{P}{T} = \left(\frac{\partial S_B}{\partial V} \right)_{E, N}, \quad -\frac{\mu}{T} = \left(\frac{\partial S_B}{\partial N} \right)_{E, V}. \quad (3.8)$$

Lecture 6 ends here. Lecture 7 starts from a small system coupled weakly to a large energy reservoir.

3.3 Canonical ensemble from a reservoir

Let a small system S exchange energy with a much larger reservoir R . Assume weak coupling, so the interaction energy is negligible and the total energy is additive. The composite system is isolated with total energy E_{tot} and is assigned the microcanonical distribution. If S occupies a microstate i of energy E_i , the number of compatible composite microstates is the reservoir multiplicity $\Omega_R(E_{\text{tot}} - E_i)$. Therefore

$$p_i = \frac{\Omega_R(E_{\text{tot}} - E_i)}{\sum_j \Omega_R(E_{\text{tot}} - E_j)} \propto \exp \left[\frac{1}{k_B} S_R(E_{\text{tot}} - E_i) \right]. \quad (3.9)$$

The equal-probability postulate applies to the composite system. The subsystem probabilities are unequal because different subsystem microstates leave different numbers of reservoir states available.

Choose a reference reservoir energy E_R^0 near its equilibrium value and write $\delta E_i = E_i - E_S^0$. The reservoir energy is $E_R^0 - \delta E_i$. Taylor expansion gives

$$S_R(E_R^0 - \delta E_i) = S_R(E_R^0) - \frac{\delta E_i}{T} - \frac{(\delta E_i)^2}{2T^2 C_R} + \mathcal{O}((\delta E_i)^3). \quad (3.10)$$

Here the derivatives are evaluated at E_R^0 ,

$$\left(\frac{\partial S_R}{\partial E_R} \right)_{V,N} = \frac{1}{T}, \quad \left(\frac{\partial^2 S_R}{\partial E_R^2} \right)_{V,N} = -\frac{1}{T^2 C_R},$$

as derived in Eq. (1.37). Terms containing the reference energy E_S^0 can be absorbed into the normalization, leaving the Boltzmann factor $e^{-\beta E_i}$. Exercise 3.2 keeps the quadratic finite-reservoir correction. When the reservoir is large enough for that term to be negligible over the subsystem's probable energy range,

$$p_i = \frac{e^{-\beta E_i}}{Z(\beta, V, N)}, \quad Z(\beta, V, N) = \sum_i e^{-\beta E_i}, \quad \beta = \frac{1}{k_B T}. \quad (3.11)$$

The partition function Z is the normalization required by $\sum_i p_i = 1$.

Status: controlled approximation

The canonical form follows from the leading reservoir expansion. The small parameter is set by the system-to-reservoir size ratio. An ideal infinite reservoir makes the approximation exact for the subsystem model.

For a classical system,

$$Z(\beta, V, N) = \frac{1}{N! h^{dN}} \int_V d\Gamma e^{-\beta H_N(\Gamma)}. \quad (3.12)$$

For a quantum system,

$$Z(\beta, V, N) = \text{Tr}_N e^{-\beta H_N}. \quad (3.13)$$

3.3.1 Thermodynamics from the partition function

The mean energy follows directly from differentiating the normalization:

$$\frac{\partial Z}{\partial \beta} = - \sum_i E_i e^{-\beta E_i} = -Z \langle E \rangle, \quad (3.14)$$

$$U \equiv \langle E \rangle = -\frac{1}{Z} \frac{\partial Z}{\partial \beta} = -\frac{\partial \ln Z}{\partial \beta}. \quad (3.15)$$

The Gibbs entropy of the canonical probabilities is

$$S = -k_B \sum_i p_i \ln p_i = -k_B \sum_i p_i (-\beta E_i - \ln Z) = k_B \ln Z + \frac{U}{T}.$$

Substitution in $F = U - TS$ then gives

$$F(T, V, N) = -k_B T \ln Z. \quad (3.16)$$

Energy, entropy, pressure, and chemical potential follow:

$$U = -\frac{\partial \ln Z}{\partial \beta}, \quad (3.17)$$

$$S = k_B(\ln Z + \beta U) = -\left(\frac{\partial F}{\partial T}\right)_{V,N}, \quad (3.18)$$

$$P = k_B T \left(\frac{\partial \ln Z}{\partial V}\right)_{T,N}, \quad (3.19)$$

$$\mu = \left(\frac{\partial F}{\partial N}\right)_{T,V}. \quad (3.20)$$

For example,

$$-\left(\frac{\partial F}{\partial T}\right)_{V,N} = k_B \ln Z + k_B T \frac{\partial \ln Z}{\partial T} = k_B \ln Z + \frac{U}{T} = S,$$

where $\partial_T \ln Z = U/(k_B T^2)$. The pressure formula follows from $P = -(\partial F/\partial V)_{T,N}$. The derivative with respect to V must include both any explicit V dependence of the Hamiltonian and the changing domain of integration. For large N , the chemical potential is the derivative shown above; at finite integer N , the corresponding definition uses a free energy difference such as $F(T, V, N+1) - F(T, V, N)$.

The thermodynamic entropy obtained from $-\partial_T F$ therefore agrees with the Gibbs entropy when the energy levels have no explicit temperature dependence. If an effective Hamiltonian depends on T , its derivative adds an extra term and the preceding identity must be reconsidered.

3.3.2 Fluctuations and response

Lecture 8 begins here. It combines positive-temperature free-energy minimization with fluctuations and linear response.

For a Hamiltonian with no explicit temperature dependence, differentiating the mean energy once more gives

$$\frac{\partial U}{\partial \beta} = \frac{\partial}{\partial \beta} \left(\frac{\sum_i E_i e^{-\beta E_i}}{Z} \right) \quad (3.21)$$

$$= -\langle E^2 \rangle + \langle E \rangle^2 = -\text{Var}(E). \quad (3.22)$$

Since $\partial T/\partial \beta = -k_B T^2$ and $C_V = (\partial U/\partial T)_{V,N}$,

$$\text{Var}(E) = \frac{\partial^2 \ln Z}{\partial \beta^2} = k_B T^2 C_V. \quad (3.23)$$

If $U \propto N$ and $C_V \propto N$, then

$$\frac{\sqrt{\text{Var}(E)}}{U} \sim N^{-1/2}, \quad (3.24)$$

for an ordinary extensive system away from a critical point. This estimate is about relative fluctuations: the absolute width still grows as $\sqrt{N}$. The same derivative structure relates susceptibilities to fluctuations of their conjugate observables.

More generally, perturb a classical Hamiltonian as $H_f = H_0 - fA$. The sign convention makes positive f favor states with larger A . The same formulas apply to a quantum system when A commutes with H_f . In a simultaneous eigenbasis, write $Z_f = \sum_i e^{-\beta(E_i - fA_i)}$. If A

has no explicit f dependence, the quotient rule gives

$$\frac{\partial \langle A \rangle}{\partial f} = \frac{1}{Z_f} \sum_i \beta A_i^2 e^{-\beta(E_i - f A_i)} - \frac{\langle A \rangle}{Z_f} \frac{\partial Z_f}{\partial f}, \quad (3.25)$$

$$\frac{\partial Z_f}{\partial f} = \beta Z_f \langle A \rangle. \quad (3.26)$$

Combining these lines gives

$$\left(\frac{\partial \langle A \rangle}{\partial f} \right)_T = \beta \text{Var}(A), \quad (3.27)$$

when A has no explicit f dependence. At $f = 0$, the left side is the static susceptibility of A to its conjugate field. Under the same classical or commuting assumptions,

$$\frac{\partial \langle B \rangle}{\partial f} = \beta (\langle AB \rangle - \langle A \rangle \langle B \rangle) = \beta \text{Cov}(A, B).$$

For noncommuting quantum observables, differentiating an operator exponential requires the Duhamel formula and gives an imaginary-time Kubo–Mori correlation, stated in Eq. (A.54). This is a static fluctuation–response relation: an equilibrium correlation fixes the linear response to a time-independent weak bias. Dynamical response to a time-dependent perturbation requires additional information.

Status: exact

Equation (3.23) is exact in the canonical ensemble for a temperature-independent Hamiltonian. The $N^{-1/2}$ scaling adds extensivity and finite correlation length. It can fail near criticality or in nonadditive systems.

Lecture 9 begins here with energy and particle exchange.

3.4 Grand-canonical ensemble

Now let the system exchange energy and particles with a reservoir. If the system microstate has energy E_i and particle number N_i , the reservoir has $(E_{\text{tot}} - E_i, N_{\text{tot}} - N_i)$. To first order,

$$S_R(E_{\text{tot}} - E_i, N_{\text{tot}} - N_i) = S_R(E_{\text{tot}}, N_{\text{tot}}) - \frac{E_i}{T} - \left(\frac{\partial S_R}{\partial N_R} \right)_{E_R, V} N_i + \dots \quad (3.28)$$

$$= S_R(E_{\text{tot}}, N_{\text{tot}}) - \frac{E_i - \mu N_i}{T} + \dots, \quad (3.29)$$

where Eq. (1.27) gives $(\partial S_R / \partial N_R)_{E_R, V} = -\mu/T$. Exponentiating the reservoir multiplicity and normalizing gives

$$p_i = \frac{1}{\Xi} e^{-\beta(E_i - \mu N_i)}, \quad \Xi(\beta, V, \mu) = \sum_i e^{-\beta(E_i - \mu N_i)}. \quad (3.30)$$

The index i now runs over microstates in every particle-number sector. An equivalent and often more useful form is

$$\Xi(\beta, V, z) = \sum_{N=0}^{\infty} z^N Z_N(\beta, V), \quad z = e^{\beta\mu}, \quad (3.31)$$

because the states with fixed N sum to the canonical partition function Z_N . The Gibbs entropy of the grand-canonical probabilities is

$$S = -k_B \sum_i p_i \ln p_i \quad (3.32)$$

$$= k_B \ln \Xi + k_B \beta \langle E - \mu N \rangle \quad (3.33)$$

$$= k_B \ln \Xi + \frac{U - \mu \langle N \rangle}{T}. \quad (3.34)$$

Rearranging the definition $\Phi = U - TS - \mu \langle N \rangle$ gives the grand potential

$$\Phi = -k_B T \ln \Xi. \quad (3.35)$$

Its derivatives give

$$U = - \left(\frac{\partial \ln \Xi}{\partial \beta} \right)_{z, V}, \quad \langle N \rangle = z \frac{\partial \ln \Xi}{\partial z}, \quad P = k_B T \left(\frac{\partial \ln \Xi}{\partial V} \right)_{T, \mu}. \quad (3.36)$$

The subscript z in the energy derivative matters. At fixed fugacity, $\partial_\beta z^N = 0$, so $-\partial_\beta \ln \Xi|_z = \langle E \rangle$. At fixed μ the derivative instead produces $\langle E - \mu N \rangle$.

The dimensionless fugacity $z = e^{\beta\mu}$ weights each added particle. Acting once with $z \partial_z$ gives

$$z \frac{\partial \Xi}{\partial z} = \sum_{N=0}^{\infty} N z^N Z_N = \Xi \langle N \rangle.$$

Acting again and differentiating the quotient gives

$$z \frac{\partial \langle N \rangle}{\partial z} = \langle N^2 \rangle - \langle N \rangle^2 = \text{Var}(N), \quad (3.37)$$

$$\frac{\partial}{\partial \mu} = \beta z \frac{\partial}{\partial z} \quad (T \text{ fixed}). \quad (3.38)$$

Therefore number fluctuations obey

$$\text{Var}(N) = \frac{1}{\beta} \left(\frac{\partial \langle N \rangle}{\partial \mu} \right)_{T, V}. \quad (3.39)$$

Exercise 3.5 repeats this calculation and checks it in a concrete model. For a homogeneous extensive phase, the Euler relation $U = TS - PV + \mu N$ gives $\Phi = U - TS - \mu N = -PV$.

3.5 Canonical variational principle

In this section we assume $T > 0$. The algebraic identity below also holds for a normalizable negative-temperature Gibbs distribution, but multiplication by $k_B T < 0$ reverses the variational inequality: at negative temperature the Gibbs distribution maximizes, rather than minimizes, $\mathcal{F}_T$.

Let p be any distribution on energy eigenstates, and let

$$p_i^{\text{eq}} = \frac{e^{-\beta E_i}}{Z}. \quad (3.40)$$

Define the nonequilibrium free-energy functional

$$\mathcal{F}_T[p] = \sum_i p_i E_i - k_B T h(p). \quad (3.41)$$

Here $h(p) = -\sum_i p_i \ln p_i$ is dimensionless, so the physical entropy is $k_B h(p)$. Insert $\ln p_i^{\text{eq}} = -\beta E_i - \ln Z$ into the relative entropy:

$$k_B T D(p \| p^{\text{eq}}) = k_B T \sum_i p_i (\ln p_i + \beta E_i + \ln Z) \quad (3.42)$$

$$= k_B T \sum_i p_i \ln p_i + \sum_i p_i E_i + k_B T \ln Z \quad (3.43)$$

$$= \mathcal{F}_T[p] - F_{\text{eq}}. \quad (3.44)$$

Positivity of relative entropy proves

$$\mathcal{F}_T[p] \geq F_{\text{eq}}, \quad (3.45)$$

with equality only at the Gibbs distribution.

This proof assumes that p has support only where the Gibbs distribution is nonzero; for finite energies at positive temperature that support is the whole state space. The same argument works for a classical phase-space density when both distributions are defined relative to the same reference measure.

Worked example: one bit of nonequilibrium free energy

Take a degenerate two-state memory, $E_0 = E_1 = 0$, in contact with a bath. Its equilibrium distribution is $(1/2, 1/2)$ and $F_{\text{eq}} = -k_B T \ln 2$. If the memory is prepared with certainty in state 0, then $h(p) = 0$ and $\mathcal{F}_T[p] = 0$. Hence

$$\mathcal{F}_T[p] - F_{\text{eq}} = k_B T \ln 2 = k_B T D((1, 0) \| (1/2, 1/2)). \quad (3.46)$$

The ordered bit stores free energy even though the two logical states have equal internal energy. Chapter 6 will turn this observation into a work bound and will state the assumptions under which the bound is attainable.

One identity, two readings

For $T > 0$, Eq. (3.44) says that equilibrium minimizes free energy. It also gives the exact numerical identity $\mathcal{F}_T[p] - F_{\text{eq}} = k_B T D(p \| p^{\text{eq}})$. Interpreting this free-energy excess as extractable work requires the process assumptions stated in Chapter 6.

The final part of Lecture 9 and Lecture 10 connect the ensembles to large deviations and Legendre–Fenchel transforms.

3.6 Microcanonical to canonical: Laplace principle

Group canonical states by energy:

$$Z(\beta) = \int dE \omega(E) e^{-\beta E} \simeq \frac{1}{\Delta E} \int dE \exp \left[\frac{S_B(E; \Delta E)}{k_B} - \beta E \right]. \quad (3.47)$$

The first equality groups together all states in an interval dE . The second form uses $\Omega(E; \Delta E) \simeq \omega(E) \Delta E$ for a narrow shell. The factor $1/\Delta E$ affects a prefactor but not the leading thermodynamic rate.

For a large system, write $E = Ne$ and $S_B(E) = Ns(e)$. Then

$$Z(\beta) \simeq \frac{N}{\Delta E} \int de \exp \left\{ N \left[\frac{s(e)}{k_B} - \beta e \right] \right\}.$$

If the function in square brackets has a unique stable maximum, contributions away from it are exponentially smaller. Taking $N^{-1} \ln$ of the integral gives

$$-\beta f(\beta) = \sup_e \left[\frac{s(e)}{k_B} - \beta e \right], \quad (3.48)$$

where $f = F/N$. If $s(e)$ is differentiable and concave, the maximizing energy satisfies

$$\frac{1}{k_B} \frac{ds}{de} = \beta. \quad (3.49)$$

The equality holds because the derivative of the exponent vanishes at the maximum. Since $ds/de = 1/T$ and $\beta = 1/(k_B T)$, this is the equality of microcanonical and canonical temperatures. Strict negative curvature makes the stationary point a maximum and gives a Gaussian leading approximation to the canonical energy distribution.

Equation (3.48) has the same Legendre–Fenchel structure as the rate-function formula (2.98). Probability theory minimizes a fluctuation cost; thermodynamics maximizes entropy minus the energetic bias. Touchette develops this bridge in detail [58].

3.7 When are ensembles equivalent?

The phrase “ensembles are equivalent in the thermodynamic limit” needs hypotheses and a statement of which observables are compared. For short-range additive systems with a strictly concave entropy density and a single stable phase, the canonical energy distribution

$$P_\beta(E) \propto \exp \left[\frac{S_B(E)}{k_B} - \beta E \right]$$

concentrates around one energy density. Relative energy fluctuations vanish, and canonical averages of local observables approach their microcanonical values at that energy.

If $s(e)$ contains a nonconcave region, no supporting line of slope $k_B \beta$ touches the graph there. The canonical supremum skips those energies, and the inverse transform returns the concave envelope rather than the original entropy. Skipped energies may occur microcanonically but not as canonical equilibrium energies. Long-range interactions can sustain this inequivalence because the usual additivity argument for concavity fails.

First-order coexistence in an additive short-range system works differently. At the transition temperature, the finite-size canonical energy distribution is bimodal: one peak represents each phase. As T crosses the transition temperature, the dominant peak switches and the canonical mean energy approaches a jump in the thermodynamic limit. Microcanonically, energies between the two pure-phase values remain accessible as spatial mixtures of the phases, with an interfacial free-energy correction at finite size. Local intensive observables measured well inside either phase agree between ensembles, but the energy–temperature relation and observables sensitive to the phase fraction need not. This is the meaning of *partial equivalence*: agreement for a stated class of observables, not equality of the probability distributions.

Status: thermodynamic-limit statement

Ensemble equivalence is a thermodynamic-limit statement that depends on additivity, concavity, and the observables being compared. It is not an identity between finite probability distributions.

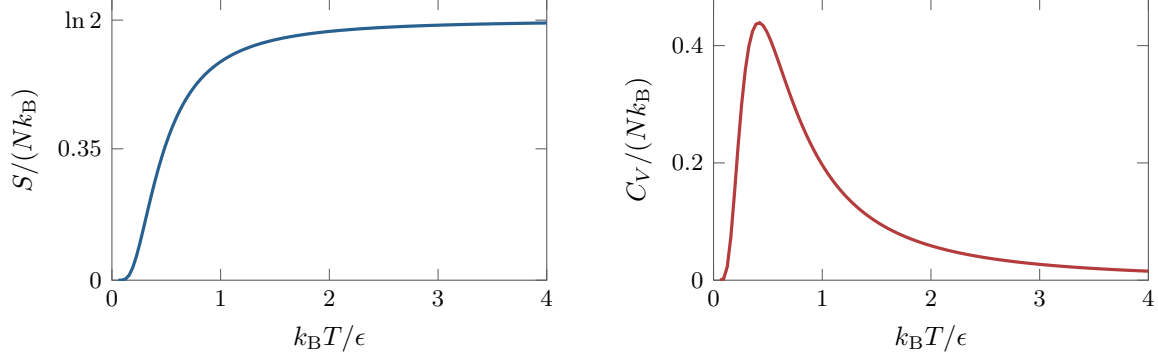


Figure 3.1: Entropy and heat capacity of independent two-level units. The entropy saturates when both levels become equally populated, while the heat capacity peaks only in the crossover region where the occupation changes rapidly with temperature.

3.8 Worked example: independent two-level systems

Take N independent units with energies 0 and ϵ . The one-unit partition function is

$$z_1 = 1 + e^{-\beta\epsilon}. \quad (3.50)$$

Independence makes the N -unit partition function factorize: each unit contributes either 1 or $e^{-\beta\epsilon}$, so

$$Z = z_1^N, \quad F = -Nk_B T \ln(1 + e^{-\beta\epsilon}). \quad (3.51)$$

The excitation probability and mean energy are

$$p_1 = \frac{1}{1 + e^{\beta\epsilon}}, \quad U = \frac{N\epsilon}{1 + e^{\beta\epsilon}}. \quad (3.52)$$

Here $p_1 = e^{-\beta\epsilon}/z_1$, and the total excitation number is a sum of N identical Bernoulli variables. Differentiating U gives

$$\frac{\partial U}{\partial \beta} = -N\epsilon^2 \frac{e^{\beta\epsilon}}{(1 + e^{\beta\epsilon})^2}, \quad \frac{\partial \beta}{\partial T} = -\frac{1}{k_B T^2}.$$

The heat capacity is

$$C_V = Nk_B(\beta\epsilon)^2 \frac{e^{\beta\epsilon}}{(1 + e^{\beta\epsilon})^2}. \quad (3.53)$$

It vanishes exponentially at low temperature because excitations are frozen out. At high temperature the populations approach 1/2, and changing the temperature hardly changes them, so C_V again vanishes. The peak between these limits is the Schottky anomaly.

Using $S = k_B(\ln Z + \beta U)$, the entropy is

$$S = Nk_B \left[\ln(1 + e^{-\beta\epsilon}) + \frac{\beta\epsilon}{1 + e^{\beta\epsilon}} \right]. \quad (3.54)$$

For a nondegenerate ground state, $S \rightarrow 0$ as $T \rightarrow 0$. At high temperature, $S \rightarrow Nk_B \ln 2$.

The same system makes the relation between ensembles explicit. In the microcanonical description, a shell with exactly m excitations has multiplicity $\binom{N}{m}$ and entropy

$$S_B(m) = k_B \ln \binom{N}{m}. \quad (3.55)$$

Writing $x = m/N$ and using Stirling's formula gives $S_B/(Nk_B) = h_2(x) + \mathcal{O}(\ln N/N)$. Since $E = m\epsilon = N\epsilon x$, $\partial x/\partial E = 1/(N\epsilon)$. Therefore the large- N microcanonical temperature follows from

$$\frac{\partial S_B}{\partial E} = Nk_B h'_2(x) \frac{\partial x}{\partial E}, \quad h'_2(x) = \ln \frac{1-x}{x},$$

which gives

$$\frac{1}{T} = \frac{\partial S_B}{\partial E} \simeq \frac{k_B}{\epsilon} \ln \frac{1-x}{x}. \quad (3.56)$$

Writing the left side as $k_B\beta$ and solving gives $x = (1 + e^{\beta\epsilon})^{-1}$, the same large- N excitation fraction as the canonical mean. For $x > 1/2$, the entropy decreases as energy increases, so $\partial S/\partial E < 0$ and the temperature is negative.

Negative temperature does not mean “colder than zero.” Let two systems exchange a small energy δE from system A to system B . Then

$$\Delta S_{\text{tot}} = -\frac{\delta E}{T_A} + \frac{\delta E}{T_B} = k_B(\beta_B - \beta_A)\delta E.$$

Energy can flow spontaneously from A to B when $\beta_A < \beta_B$. Every negative inverse temperature is below every positive one, so a negative-temperature system is hotter in this operational ordering. Such equilibrium states require an upper-bounded energy spectrum; otherwise a Boltzmann factor with $\beta < 0$ cannot be normalized.

3.9 Exercises

Exercise 3.1. Compute the microcanonical temperature of N two-level units at fixed excitation number m . Show that negative temperatures occur when $m > N/2$. Explain why a bounded spectrum is needed.

Exercise 3.2. Derive the finite-reservoir correction to the canonical distribution through the quadratic term in the reservoir entropy. Express it in terms of C_R .

Exercise 3.3. Show that $\partial_\beta \langle E \rangle = -\text{Var}(E)$ and derive (3.23). Repeat the argument for particle-number fluctuations in the grand-canonical ensemble.

Exercise 3.4. For a classical harmonic oscillator, calculate Z , U , S , and C_V . Keep the phase-space cell h explicit. Compare with the quantum answer at high temperature.

Exercise 3.5. Use the grand partition function to show that $\langle N \rangle = z\partial_z \ln \Xi$. Derive the corresponding expression for $\text{Var}(N)$.

Exercise 3.6 (*). Let $s(e)$ have a convex segment. Draw its concave envelope and show geometrically how the canonical ensemble performs a Maxwell construction. Identify the latent heat.

Exercise 3.7. For the two-level system, find the temperature at which C_V is maximal numerically or graphically. Explain why the high-temperature heat capacity vanishes even though the entropy is maximal.

Exercise 3.8. Prove the classical free-energy identity (3.44) for a continuous phase-space density relative to Liouville measure.

3.10 Further reading

Kardar gives a direct derivation of the ensembles and fluctuation formulas [23]. Pathria and Reichl cover variants and applications [42, 47]. Touchette explains ensemble equivalence through large deviations and Legendre–Fenchel transforms [58].

Part II

Classical systems and irreversibility

Chapter 4

Classical gases and interacting particles

Lecture 11–13. The ideal gas, molecular degrees of freedom, the classical limit, and the virial expansion.

What you should be able to do

- Derive the Sackur–Tetrode entropy and explain both factors $N!$ and h^{3N} .
- Obtain velocity and energy distributions from the canonical ensemble.
- Identify where equipartition works and where quantum level spacing defeats it.
- Calculate the second virial coefficient from a pair potential.

4.1 Monatomic ideal gas

Consider N identical nonrelativistic particles in a three-dimensional box of volume V ,

$$H_N = \sum_{a=1}^N \frac{\mathbf{p}_a^2}{2m}. \quad (4.1)$$

The classical canonical partition function is

$$Z_N = \frac{1}{N! h^{3N}} \int_V \prod_{a=1}^N d^3 q_a d^3 p_a \exp \left(-\beta \sum_a \frac{\mathbf{p}_a^2}{2m} \right) \quad (4.2)$$

$$= \frac{1}{N!} \left(\frac{V}{\lambda_T^3} \right)^N, \quad (4.3)$$

because every position integral gives V and every one-particle momentum integral gives

$$\int_{\mathbb{R}^3} d^3 p e^{-\beta \mathbf{p}^2 / (2m)} = \left(\int_{-\infty}^{\infty} dp_x e^{-\beta p_x^2 / (2m)} \right)^3 \quad (4.4)$$

$$= \left(\frac{2\pi m}{\beta} \right)^{3/2}. \quad (4.5)$$

The one-dimensional Gaussian integral is derived in Eq. (A.4). Defining

$$\lambda_T = \frac{h}{\sqrt{2\pi m k_B T}} = \sqrt{\frac{2\pi \hbar^2}{m k_B T}}. \quad (4.6)$$

This definition turns $(2\pi m/\beta)^{3/2}/h^3$ into $1/\lambda_T^3$. The length λ_T is the thermal de Broglie wavelength: it estimates the spatial scale over which the wave nature of a particle with thermal momentum cannot be ignored.

The dimensionless phase-space density is $n\lambda_T^3$, where $n = N/V$. Roughly, n^{-1} is the volume available per particle while λ_T^3 is the volume of a thermal wave packet. The classical dilute regime requires

$$n\lambda_T^3 \ll 1, \quad n = \frac{N}{V}. \quad (4.7)$$

Under this condition the wave packets overlap weakly, and Bose or Fermi exchange corrections are small. Low density and high temperature both favor this condition.

Take the logarithm of the partition function and use $\ln N! = N \ln N - N + \mathcal{O}(\ln N)$:

$$\ln Z_N = -\ln N! + N \ln V - 3N \ln \lambda_T \quad (4.8)$$

$$= N \left[\ln \left(\frac{V}{N\lambda_T^3} \right) + 1 \right] + \mathcal{O}(\ln N). \quad (4.9)$$

The thermodynamic identities from Chapter 3 now give

$$F = -k_B T \ln Z_N = -Nk_B T \left[\ln \left(\frac{V}{N\lambda_T^3} \right) + 1 \right], \quad (4.10)$$

$$P = - \left(\frac{\partial F}{\partial V} \right)_{T,N} = \frac{Nk_B T}{V}, \quad (4.11)$$

$$U = - \frac{\partial \ln Z_N}{\partial \beta} = \frac{3}{2} Nk_B T, \quad (4.12)$$

$$S = \frac{U - F}{T} = Nk_B \left[\ln \left(\frac{V}{N\lambda_T^3} \right) + \frac{5}{2} \right]. \quad (4.13)$$

For the energy derivative we used $\lambda_T^3 \propto \beta^{3/2}$, hence $\partial_\beta \ln Z_N = -3N/(2\beta)$. The last line is the Sackur–Tetrode entropy in its thermodynamic-limit form. Exercise 4.1 asks for an independent reconstruction starting from the phase-space integral.

The four results are often summarized as

$$PV = Nk_B T, \quad C_V = \frac{3}{2} Nk_B, \quad C_P = \frac{5}{2} Nk_B. \quad (4.14)$$

where the last relation uses the ideal-gas enthalpy $U + PV$.

The constant $5/2$ has two sources. Stirling's $-N$ term contributes $+N$ to $\ln Z_N$, while the three translational quadratic degrees of freedom give $\beta U = 3N/2$. Their sum is $5N/2$. No experimental fitting enters this constant.

Status: controlled approximation

The phase-space calculation is semiclassical. It assumes $n\lambda_T^3 \ll 1$ and neglects exchange corrections. Stirling's approximation drops terms of order $\ln N$ compared with the order- N entropy.

4.1.1 Gibbs paradox

Without the factor $1/N!$, the entropy would contain $N \ln V$ rather than $N \ln(V/N)$. Mixing equal samples of the same gas would then produce a spurious entropy increase, and the entropy would fail to be extensive.

The calculation makes the problem visible. Take two portions of the same gas, with (N_1, V_1) and (N_2, V_2) at equal temperature and density $N_1/V_1 = N_2/V_2 = n$. Removing a partition gives $N = N_1 + N_2$ and $V = V_1 + V_2$, so $V/N = 1/n$ both before and after mixing. The Sackur–Tetrode entropy therefore gives

$$S(N, V) - S(N_1, V_1) - S(N_2, V_2) = 0.$$

Without $1/N!$, each portion would instead gain access to the larger volume, and one would obtain the false increase

$$\Delta S_{\text{labelled}} = k_B \left[N_1 \ln \frac{V}{V_1} + N_2 \ln \frac{V}{V_2} \right].$$

Quantum mechanics identifies permutations of identical particles as the same physical configuration. Dividing by $N!$ retains that identification in the classical limit. For two different species, each species has its own factorial $N_A!N_B!$; after removing the partition, both species gain access to the full volume. For equal initial volumes and equal particle numbers this gives the physical entropy of mixing $\Delta S = 2Nk_B \ln 2$.

4.2 Maxwell distribution

Momentum factors separate in the ideal-gas Boltzmann weight. For one Cartesian component, normalization by the same Gaussian integral used above gives

$$p(v_x) = \sqrt{\frac{m}{2\pi k_B T}} \exp\left(-\frac{mv_x^2}{2k_B T}\right). \quad (4.15)$$

The change from momentum to velocity uses $p_x = mv_x$ and $dp_x = m dv_x$. Multiplying the three independent component densities gives the normalized velocity density

$$p(\mathbf{v}) = \left(\frac{m}{2\pi k_B T}\right)^{3/2} \exp\left(-\frac{m\mathbf{v}^2}{2k_B T}\right). \quad (4.16)$$

The probability that the speed lies in $[v, v + dv]$ is the integral over a spherical shell in velocity space. Since $d^3v = v^2 dv d\Omega$ and $\int d\Omega = 4\pi$, the speed density is

$$f(v) = 4\pi v^2 \left(\frac{m}{2\pi k_B T}\right)^{3/2} \exp\left(-\frac{mv^2}{2k_B T}\right), \quad v \geq 0. \quad (4.17)$$

The factor v^2 is geometric. It makes $f(0) = 0$ even though the vector density $p(\mathbf{v})$ is largest at $\mathbf{v} = 0$.

Three different summaries of this distribution are common. The maximum solves

$$\frac{d}{dv} \ln f(v) = \frac{2}{v} - \frac{mv}{k_B T} = 0,$$

while the moments follow from

$$\langle v^r \rangle = \int_0^\infty v^r f(v) dv = \frac{2}{\sqrt{\pi}} \left(\frac{2k_B T}{m}\right)^{r/2} \Gamma\left(\frac{r+3}{2}\right). \quad (4.18)$$

Here $\Gamma(s) = \int_0^\infty t^{s-1} e^{-t} dt$ is the Gamma function. For $r = 1$, $\Gamma(2) = 1$; for $r = 2$, $\Gamma(5/2) = 3\sqrt{\pi}/4$. Thus

$$v_{\text{mp}} = \sqrt{\frac{2k_B T}{m}}, \quad \langle v \rangle = \sqrt{\frac{8k_B T}{\pi m}}, \quad v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}}. \quad (4.19)$$

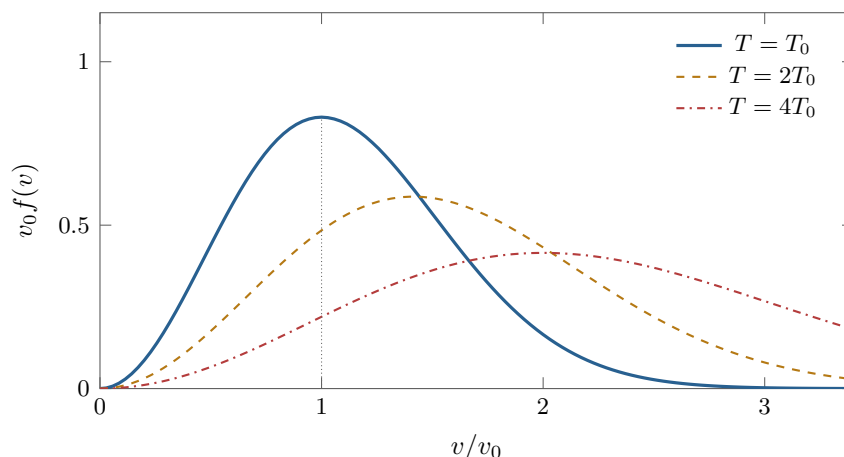


Figure 4.1: Maxwell speed distributions at fixed particle mass, with $v_0 = \sqrt{2k_B T_0/m}$. Raising the temperature shifts probability toward larger speeds and broadens the distribution; each curve remains normalized. The dotted line marks the most probable speed at T_0 .

Molecular speeds at room temperature

For carbon monoxide at $T = 300$ K, using $m \simeq 28.01$ atomic mass units,

$$v_{\text{mp}} \simeq 422 \text{ m s}^{-1}, \quad \langle v \rangle \simeq 476 \text{ m s}^{-1}, \quad v_{\text{rms}} \simeq 517 \text{ m s}^{-1}. \quad (4.20)$$

These are three summaries of the same broad distribution, not three competing predictions for a molecule's speed. Substitution in the formulas above gives $v_{\text{mp}} < \langle v \rangle < v_{\text{rms}}$.

Exercise 4.3 derives the three characteristic speeds directly from the Maxwell density.

4.2.1 Distribution of kinetic energy

The kinetic energy of one particle is $\varepsilon = mv^2/2$. Probability is unchanged by a change of variable, so $g(\varepsilon) d\varepsilon = f(v) dv$. Since $v = \sqrt{2\varepsilon/m}$ and $dv/d\varepsilon = (2m\varepsilon)^{-1/2}$,

$$g(\varepsilon) = \frac{2}{\sqrt{\pi}} \frac{\sqrt{\varepsilon}}{(k_B T)^{3/2}} e^{-\varepsilon/(k_B T)}, \quad \varepsilon \geq 0. \quad (4.21)$$

This is a gamma distribution with shape parameter $3/2$. Direct integration, or Eq. (4.18) with $r = 2$, gives $\langle \varepsilon \rangle = 3k_B T/2$. The most probable kinetic energy is $k_B T/2$; it differs from the mean because the distribution has a long tail.

4.3 Equipartition theorem

Suppose the Hamiltonian contains a coordinate x only through $ax^2/2$, with $x \in \mathbb{R}$. Integration by parts or direct Gaussian integration gives

$$\left\langle \frac{1}{2} ax^2 \right\rangle = \frac{1}{2} k_B T. \quad (4.22)$$

For example, if $I(a) = \int_{-\infty}^{\infty} e^{-\beta ax^2/2} dx = \sqrt{2\pi/(\beta a)}$, then

$$\left\langle \frac{1}{2} ax^2 \right\rangle = -\frac{\partial \ln I}{\partial \beta} = \frac{1}{2\beta} = \frac{1}{2} k_B T.$$

Each independent quadratic term therefore contributes $k_B T/2$ to the mean energy and, when its coefficient is temperature independent, $k_B/2$ to the heat capacity.

More generally, let x_i denote any canonical coordinate or momentum. Assume that the integration domain and the Boltzmann weight make the boundary term $x_i e^{-\beta H}$ vanish at both ends of the x_j integration. Then

$$\left\langle x_i \frac{\partial H}{\partial x_j} \right\rangle = k_B T \delta_{ij}. \quad (4.23)$$

To verify this identity, write

$$\int d\Gamma x_i \frac{\partial H}{\partial x_j} e^{-\beta H} = -\frac{1}{\beta} \int d\Gamma x_i \frac{\partial}{\partial x_j} e^{-\beta H} = \frac{\delta_{ij}}{\beta} \int d\Gamma e^{-\beta H},$$

where the last equality is integration by parts and uses $\partial x_i / \partial x_j = \delta_{ij}$. Division by the full partition integral proves the result. Taking $i = j$ and a quadratic contribution $ax_i^2/2$ reproduces Eq. (4.22).

Status: exact

Equipartition is exact for the stated classical canonical integrals. It is not a theorem about arbitrary quantum systems, bounded variables, or degrees of freedom whose level spacing exceeds $k_B T$.

4.3.1 Classical harmonic solid

A three-dimensional harmonic solid has three momenta and three quadratic displacements per atom. Hence

$$U = 3Nk_B T, \quad C_V = 3Nk_B. \quad (4.24)$$

This is the Dulong–Petit law. Its failure at low temperature was an early sign that classical equipartition could not describe quantized vibrations.

Lecture 11 ends here. Lecture 12 starts with molecular rotations and vibrations.

4.4 Diatomic molecules

A dilute diatomic gas has translational, rotational, vibrational, and electronic degrees of freedom. If the molecular Hamiltonian separates as a sum of terms acting on those degrees of freedom, the Boltzmann factor factorizes. The one-molecule partition function is then

$$z_1 = z_{\text{trans}} z_{\text{rot}} z_{\text{vib}} z_{\text{el}}. \quad (4.25)$$

The N -molecule ideal-gas partition function is $Z_N = z_1^N / N!$. The translational factor is $z_{\text{trans}} = V/\lambda_T^3$ with the molecular mass in λ_T ; the remaining factors determine the internal energy and internal heat capacity.

For a linear rigid rotor with moment of inertia I , quantum mechanics gives

$$E_J = \frac{\hbar^2}{2I} J(J+1), \quad g_J = 2J+1. \quad (4.26)$$

For a heteronuclear molecule, the rotational partition function is

$$z_{\text{rot}} = \sum_{J=0}^{\infty} (2J+1) e^{-\theta_{\text{rot}} J(J+1)/T}, \quad \theta_{\text{rot}} = \frac{\hbar^2}{2Ik_B}. \quad (4.27)$$

Here $J = 0, 1, 2, \dots$, and the degeneracy $2J + 1$ counts the allowed magnetic quantum numbers. At $T \gg \theta_{\text{rot}}$, adjacent levels are closely spaced on the thermal scale. Replacing the sum by an integral and setting $y = J(J + 1)$, so that $dy = (2J + 1) dJ$, gives

$$\int_0^\infty (2J + 1) e^{-(\theta_{\text{rot}}/T)J(J+1)} dJ = \int_0^\infty e^{-(\theta_{\text{rot}}/T)y} dy = \frac{T}{\theta_{\text{rot}}}. \quad (4.28)$$

The molecular symmetry number divides the semiclassical count: $\sigma = 2$ for a homonuclear molecule and $\sigma = 1$ for a heteronuclear molecule. A rotation by π maps a homonuclear molecule onto the same physical orientation, so the classical orientation integral counts each configuration twice. This is the rotational counterpart of the identical-configuration quotient $1/N!$ in Eq. (3.3). Thus, for $T \gg \theta_{\text{rot}}$,

$$z_{\text{rot}} \simeq \frac{T}{\sigma \theta_{\text{rot}}}, \quad U_{\text{rot}} \simeq -N \frac{\partial \ln z_{\text{rot}}}{\partial \beta} = N k_B T. \quad (4.29)$$

At low temperature, a homonuclear molecule must instead be treated with its allowed rotational levels and nuclear-spin statistics; simply dividing the exact quantum sum by two would be incorrect.

A harmonic vibrational mode has levels $E_n = \hbar\omega(n + 1/2)$ for $n = 0, 1, 2, \dots$. Summing the geometric series gives

$$z_{\text{vib}} = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}}, \quad (4.30)$$

and

$$U_{\text{vib}} = -N \frac{\partial}{\partial \beta} \ln z_{\text{vib}} = \frac{N\hbar\omega}{2} + \frac{N\hbar\omega}{e^{\beta\hbar\omega} - 1}. \quad (4.31)$$

The thermal energy above the zero point is therefore

$$U_{\text{vib}}^{\text{th}} = \frac{N\hbar\omega}{e^{\beta\hbar\omega} - 1}. \quad (4.32)$$

Differentiating the thermal part with respect to T , with $x = \theta_{\text{vib}}/T = \beta\hbar\omega$, gives

$$\frac{\partial}{\partial T} \frac{\hbar\omega}{e^x - 1} = k_B x^2 \frac{e^x}{(e^x - 1)^2}.$$

The vibrational heat capacity is therefore

$$C_{V,\text{vib}} = N k_B \left(\frac{\theta_{\text{vib}}}{T} \right)^2 \frac{e^{\theta_{\text{vib}}/T}}{(e^{\theta_{\text{vib}}/T} - 1)^2}, \quad \theta_{\text{vib}} = \frac{\hbar\omega}{k_B}. \quad (4.33)$$

Rotations activate near $T \sim \theta_{\text{rot}}$; vibrations activate at the usually larger scale $T \sim \theta_{\text{vib}}$. “Frozen degrees of freedom” means that thermal occupation of excited levels is exponentially suppressed.

Separated molecular energy scales

Spectroscopic constants for carbon monoxide give approximately $\theta_{\text{rot}} = 2.78 \text{ K}$ and $\theta_{\text{vib}} = 3.12 \times 10^3 \text{ K}$ [38]. At room temperature its rotations are well inside their classical regime, while the vibrational mode is mostly in its ground state. This separation explains why a rigid-rotor heat capacity can work well at temperatures for which vibrational equipartition fails badly.

4.5 No classical orbital magnetism

For particles of charge q in a static vector potential $\mathbf{A}$,

$$H = \sum_a \frac{[\mathbf{p}_a - q\mathbf{A}(\mathbf{q}_a)]^2}{2m} + U(\mathbf{q}_1, \dots, \mathbf{q}_N). \quad (4.34)$$

In the classical partition function, shift each momentum integration variable by $q\mathbf{A}$. The Jacobian is one and the integration range remains all of momentum space. Therefore Z is independent of the magnetic field, and the equilibrium orbital magnetization vanishes.

More explicitly, at each fixed position define $\mathbf{p}'_a = \mathbf{p}_a - q\mathbf{A}(\mathbf{q}_a)$. Then

$$\int_{\mathbb{R}^3} d^3p_a \exp \left[-\frac{\beta}{2m} (\mathbf{p}_a - q\mathbf{A})^2 \right] = \int_{\mathbb{R}^3} d^3p'_a e^{-\beta \mathbf{p}'_a{}^2 / (2m)},$$

which contains no $\mathbf{A}$. Consequently $\mathbf{M} = -(\partial F / \partial \mathbf{B})_{T,V,N} = 0$. This argument assumes a classical phase-space integral over all canonical momenta and an equilibrium system without velocity-dependent constraints.

This is the Bohr–van Leeuwen theorem. Quantum mechanics defeats the momentum-shift argument through the operator algebra and boundary conditions.

Lecture 13 begins here with interacting gases and the density expansion.

4.6 Interacting gases

Let

$$H_N = \sum_a \frac{\mathbf{p}_a^2}{2m} + \sum_{a<b} u(r_{ab}), \quad r_{ab} = |\mathbf{q}_a - \mathbf{q}_b|. \quad (4.35)$$

The momenta still integrate out:

$$Z_N = \frac{1}{N! \lambda_T^{3N}} Q_N, \quad Q_N = \int_V \prod_a d^3q_a \exp \left[-\beta \sum_{a<b} u(r_{ab}) \right]. \quad (4.36)$$

Interactions are now contained in the configurational integral Q_N .

4.6.1 Mayer function and the second virial coefficient

Define

$$f(r) = e^{-\beta u(r)} - 1. \quad (4.37)$$

The Mayer function vanishes wherever the pair interaction vanishes. It is negative in a repulsive region and positive in an attractive region. Its integral

$$b(T) = \int_{\mathbb{R}^3} d^3r f(r). \quad (4.38)$$

This quantity has dimensions of volume. Then

$$e^{-\beta \sum_{a<b} u(r_{ab})} = \prod_{a<b} [1 + f(r_{ab})]. \quad (4.39)$$

The first terms of the product are

$$1 + \sum_{a<b} f_{ab} + \sum_{(a<b)<(c<d)} f_{ab} f_{cd} + \dots, \quad f_{ab} = f(r_{ab}).$$

For one selected pair, translational invariance and a box much larger than the interaction range give

$$\int_V d^3q_a d^3q_b f(|\mathbf{q}_a - \mathbf{q}_b|) \simeq V \int_{\mathbb{R}^3} d^3r f(r) = Vb(T). \quad (4.40)$$

The other $N - 2$ positions each give V . Since there are $\binom{N}{2}$ choices of pair,

$$Q_N = V^N \left[1 + \binom{N}{2} \frac{b(T)}{V} + \dots \right].$$

This expression cannot simply be truncated at fixed density: its displayed correction grows with N . Products of disconnected pair graphs generate the powers needed to exponentiate it. Taking $\ln Q_N$ removes those disconnected repetitions, leaving connected clusters. A connected cluster of three particles first contributes at order Nn^2 . The linked-cluster expansion therefore gives

$$\ln Q_N = N \ln V + \frac{N(N-1)}{2V} b(T) + \mathcal{O}(Nn^2).$$

Exercise 4.8 constructs this expansion systematically. Since the momentum factor has no volume dependence,

$$\frac{P}{k_B T} = \frac{\partial \ln Q_N}{\partial V} = \frac{N}{V} - \frac{N(N-1)}{2V^2} b(T) + \dots \quad (4.41)$$

$$= n - \frac{1}{2} b(T) n^2 + \mathcal{O}(n^3), \quad (4.42)$$

in the thermodynamic limit. Thus $B_2 = -b/2$, and

$$\frac{P}{k_B T} = n + B_2(T) n^2 + \mathcal{O}(n^3), \quad (4.43)$$

where

$$B_2(T) = -\frac{1}{2} \int_{\mathbb{R}^3} d^3r [e^{-\beta u(r)} - 1]. \quad (4.44)$$

For hard spheres of diameter σ ,

$$u(r) = \begin{cases} +\infty, & r < \sigma, \\ 0, & r > \sigma, \end{cases} \quad (4.45)$$

so $f = -1$ inside the excluded sphere and zero outside. Hence

$$b(T) = 4\pi \int_0^\sigma (-1) r^2 dr = -\frac{4\pi}{3} \sigma^3,$$

and

$$B_2 = \frac{2\pi}{3} \sigma^3. \quad (4.46)$$

Repulsion produces a positive correction to the pressure. Attractive regions can make B_2 negative at low enough temperature.

Worked example: hard core plus square-well attraction

Consider

$$u(r) = \begin{cases} +\infty, & 0 < r < \sigma, \\ -\epsilon, & \sigma < r < R, \\ 0, & r > R. \end{cases} \quad (4.47)$$

The Mayer function equals -1 in the core, $e^{\beta\epsilon} - 1$ in the well, and zero outside. Direct radial integration gives

$$b(T) = 4\pi \left[-\int_0^\sigma r^2 dr + (e^{\beta\epsilon} - 1) \int_\sigma^R r^2 dr \right] \quad (4.48)$$

$$= \frac{4\pi}{3} \left[-\sigma^3 + (R^3 - \sigma^3)(e^{\beta\epsilon} - 1) \right]. \quad (4.49)$$

Using $B_2 = -b/2$ gives

$$B_2(T) = \frac{2\pi}{3} \left[\sigma^3 - (R^3 - \sigma^3)(e^{\beta\epsilon} - 1) \right]. \quad (4.50)$$

At high temperature the excluded-volume term wins and $B_2 > 0$. At low temperature the attractive shell wins and $B_2 < 0$. The Boyle temperature, at which the first interaction correction vanishes, obeys

$$k_B T_B = \frac{\epsilon}{\ln[1 + \sigma^3/(R^3 - \sigma^3)]}. \quad (4.51)$$

The gas is not ideal at T_B ; only the coefficient of n^2 happens to vanish. Higher virial terms remain.

Status: controlled approximation

The virial series is controlled at low density when higher clusters are suppressed. A truncated series should not be trusted near a liquid–gas critical point without further analysis.

4.6.2 van der Waals model

The equation

$$\left(P + a \frac{N^2}{V^2} \right) (V - Nb) = Nk_B T. \quad (4.52)$$

models short-range repulsion by an excluded volume b per particle and attraction by a mean cohesive energy $-aN^2/V$. In terms of the volume per particle $v = V/N$,

$$P(T, v) = \frac{k_B T}{v - b} - \frac{a}{v^2}. \quad (4.53)$$

The constant b here is the van der Waals excluded-volume parameter, not the temperature-dependent Mayer integral $b(T)$ used above. Expanding at small density, $v = 1/n$, gives

$$\frac{P}{k_B T} = n + \left(b - \frac{a}{k_B T} \right) n^2 + \mathcal{O}(n^3).$$

Thus the model predicts the second virial coefficient $B_2^{\text{vdW}}(T) = b - a/(k_B T)$. The critical isotherm has a horizontal inflection point. Solving $(\partial P/\partial v)_T = 0$ and $(\partial^2 P/\partial v^2)_T = 0$ gives

$$-\frac{k_B T_c}{(v_c - b)^2} + \frac{2a}{v_c^3} = 0, \quad (4.54)$$

$$\frac{2k_B T_c}{(v_c - b)^3} - \frac{6a}{v_c^4} = 0. \quad (4.55)$$

Divide the second equation by the first after moving one term in each to the other side:

$$\frac{2}{v_c - b} = \frac{3}{v_c}, \quad \text{hence} \quad v_c = 3b.$$

Substitution into Eq. (4.54), followed by evaluation of $P(T_c, v_c)$, yields

$$v_c = 3b, \quad k_B T_c = \frac{8a}{27b}, \quad P_c = \frac{a}{27b^2}. \quad (4.56)$$

Below T_c , part of the algebraic isotherm has negative compressibility and cannot describe a stable homogeneous equilibrium. The Maxwell construction replaces that part by a coexistence pressure.

The model is physically instructive but not a systematic consequence of the second-virial truncation over the whole phase diagram. We will return to its free-energy geometry in [chapter 9](#).

4.7 Collisions and the dilute-gas length scale

For hard spheres of diameter σ , a collision occurs when the relative trajectory passes within distance σ of the other center. The geometric cross section is therefore $\pi\sigma^2$. In a dilute equilibrium gas, neglect correlations between incoming particles. Each Cartesian component of $\mathbf{v}_{\text{rel}} = \mathbf{v}_1 - \mathbf{v}_2$ is the difference of two independent Gaussians, each with variance $k_B T/m$. Its variance is $2k_B T/m$, the Maxwell variance corresponding to the reduced mass $\mu = m/2$. Replacing m by $m/2$ in the mean-speed formula gives

$$\langle v_{\text{rel}} \rangle = \sqrt{\frac{8k_B T}{\pi(m/2)}} = \sqrt{2} \langle v \rangle.$$

During a short time dt , a tagged particle sweeps out an effective collision volume $\pi\sigma^2 v_{\text{rel}} dt$. Multiplication by the number density and averaging over relative speeds gives the collision rate

$$\nu = n\pi\sigma^2 \langle v_{\text{rel}} \rangle = \sqrt{2} n\pi\sigma^2 \langle v \rangle. \quad (4.57)$$

The mean time between collisions is $\tau = 1/\nu$. The mean free path, defined as $\ell = \langle v \rangle \tau$, is therefore

$$\ell = \frac{\langle v \rangle}{\nu} = \frac{1}{\sqrt{2} n\pi\sigma^2}. \quad (4.58)$$

Two small parameters now describe different approximations. The condition $n\lambda_T^3 \ll 1$ suppresses quantum exchange, while $n\sigma^3 \ll 1$ makes simultaneous many-particle collisions rare and supports the binary-collision picture. A gas can satisfy one condition and not the other.

4.8 Exercises

Exercise 4.1. Derive (4.3) and every thermodynamic quantity in (4.13). Keep the $N!$ until the last step.

Exercise 4.2. Calculate the entropy change on mixing (a) equal samples of the same ideal gas and (b) equal samples of two different ideal gases. Explain the discontinuity when two species become physically indistinguishable.

Exercise 4.3. Derive the Maxwell speed distribution and calculate v_{mp} , $\langle v \rangle$, and v_{rms} .

Exercise 4.4. Prove equipartition by integration by parts for a general canonical Hamiltonian. List two ways the proof can fail.

Exercise 4.5. Obtain the high- and low-temperature limits of $C_{V,\text{vib}}$. Sketch the heat capacity of a diatomic gas as successive degrees of freedom become active.

Exercise 4.6. Carry out the momentum shift that proves the Bohr–van Leeuwen theorem. Why is the same shift not an innocuous change of variables for quantum operators on a ring?

Exercise 4.7. Calculate $B_2(T)$ for a square-well potential with a hard core of diameter σ , well depth ϵ , and range $R > \sigma$. Find the Boyle temperature where $B_2 = 0$.

Exercise 4.8 (*). Starting from the Mayer expansion, derive the order- n^2 pressure correction and identify why only connected diagrams contribute to $\ln Z$.

4.9 Further reading

The ideal-gas and molecular calculations follow standard treatments in Kardar, Pathria, and Reichl [23, 42, 47]. Kardar gives a direct introduction to Mayer clusters and virial coefficients [23].

Chapter 5

Microscopic dynamics and irreversibility

Lecture 14–16. Liouville evolution, coarse graining, Markov dynamics, entropy production, the Boltzmann equation, and nonequilibrium work.

What you should be able to do

- Explain why the fine-grained entropy of an isolated system is constant, and name the additional assumption behind each entropy that increases.
- Derive the master equation from a Markov process and prove that the relative entropy to its stationary state decreases.
- Construct the entropy production of a jump process from local detailed balance.
- Build the Boltzmann collision integral by counting collisions, and locate molecular chaos in the H -theorem.
- Derive the Jarzynski equality from Liouville's theorem and obtain the second-law inequality from it.

5.1 The apparent contradiction

Newton's equations do not distinguish the past from the future. Consider particles that interact through forces depending only on their positions, so that the Hamiltonian is even in the momenta, $H(\mathbf{q}, -\mathbf{p}) = H(\mathbf{q}, \mathbf{p})$. If $\mathbf{q}(t), \mathbf{p}(t)$ solves Hamilton's equations, then so does the reversed motion

$$\tilde{\mathbf{q}}(t) = \mathbf{q}(-t), \quad \tilde{\mathbf{p}}(t) = -\mathbf{p}(-t), \quad (5.1)$$

in which every particle retraces its path with reversed velocity. To check this, note that $\partial H / \partial \mathbf{p}$ is odd in the momenta when H is even, so that $\dot{\tilde{\mathbf{q}}}(t) = -\dot{\mathbf{q}}(-t) = -\partial_{\mathbf{p}} H(\mathbf{q}(-t), \mathbf{p}(-t)) = \partial_{\mathbf{p}} H(\tilde{\mathbf{q}}(t), \tilde{\mathbf{p}}(t))$, and similarly for $\tilde{\mathbf{p}}$. A film of any mechanical motion, played backward, shows another allowed motion.

Thermodynamic processes have a direction. A gas released into the empty half of a box spreads over the whole box and never gathers itself back into the original half, although the reversed trajectory, in which every molecule retraces its path, obeys the same equations. The second law summarizes this asymmetry: the entropy of an isolated system does not decrease.

Boltzmann believed in 1872 that he had derived this law from mechanics [5]. He wrote an equation for the distribution of molecular velocities in a dilute gas and proved that a functional

of that distribution, his H , can only decrease (Section 5.6). Loschmidt objected in 1876 that such a result cannot follow from mechanics alone [34]: reverse all velocities at some instant and, by Eq. (5.1), the gas retraces its history while H increases. Twenty years later Zermelo raised a second objection [64], based on a theorem of Poincaré: a mechanical system confined to a bounded region of phase space returns, for almost every initial state, arbitrarily close to that state, so no function of the microscopic state can decrease forever. Both objections are valid statements about mechanics. The way out, which occupies this chapter, is that every entropy that grows in time is the entropy of a description that has discarded some information about the microscopic state, and that the direction of time enters through an assumption about which information is discarded.

To make this precise we distinguish four levels of description, each with its own entropy. The most detailed is the microscopic state itself, a point in phase space that moves according to Hamilton's equations. Next comes a probability density on phase space, which describes what we know about the microscopic state; its entropy is the fine-grained Gibbs entropy, and Section 5.2 shows that it never changes. A coarse-grained description keeps only part of that information, for instance the probability of each of a finite number of cells of phase space, or the distribution of a single molecule in a gas. Finally, a thermodynamic description keeps a handful of macroscopic variables. Section 5.3 shows that the entropy of a coarse-grained description grows when the system is prepared at the resolution of that description. Sections 5.4 and 5.5 study the dynamics that results when the discarded information is assumed never to return, and Section 5.6 applies the same idea to a dilute gas. Section 5.7 returns to exact Hamiltonian dynamics, where the second law survives as a statement about averages over repeated experiments.

5.2 Liouville's theorem

We start at the microscopic level. Write a phase-space point as $\Gamma = (q_1, \dots, q_n, p_1, \dots, p_n)$, where n counts all canonical coordinate pairs; for N particles in three dimensions, $n = 3N$. For a classical Hamiltonian $H(\mathbf{q}, \mathbf{p})$, trajectories obey

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}. \quad (5.2)$$

It is useful to read the right-hand sides as a velocity field $\mathbf{v}(\Gamma) = (\dot{q}_1, \dots, \dot{q}_n, \dot{p}_1, \dots, \dot{p}_n)$ on the $2n$ -dimensional phase space. Every phase-space point moves with the velocity that Hamilton's equations assign to it, much as a fluid element moves with the local velocity of a fluid. We denote by φ_t the flow map, which sends an initial point Γ_0 to the point $\Gamma_t = \varphi_t(\Gamma_0)$ reached at time t . It is invertible, since running the equations backward from Γ_t recovers Γ_0 .

5.2.1 Volume preservation

The Hamiltonian velocity field has zero divergence,

$$\nabla_{\Gamma} \cdot \mathbf{v} = \sum_i \left(\frac{\partial \dot{q}_i}{\partial q_i} + \frac{\partial \dot{p}_i}{\partial p_i} \right) = \sum_i \left(\frac{\partial^2 H}{\partial q_i \partial p_i} - \frac{\partial^2 H}{\partial p_i \partial q_i} \right) = 0, \quad (5.3)$$

because the mixed second derivatives of a twice differentiable Hamiltonian are equal. A flow with zero divergence preserves volume. To see this, follow a small region as it moves. During a short interval δt each point moves from Γ to $\Gamma + \mathbf{v}(\Gamma) \delta t$. The Jacobian matrix of this map is $\delta_{ab} + (\partial v_a / \partial \Gamma_b) \delta t$, and its determinant is

$$\det \left(\delta_{ab} + \frac{\partial v_a}{\partial \Gamma_b} \delta t \right) = 1 + (\nabla_{\Gamma} \cdot \mathbf{v}) \delta t + \mathcal{O}(\delta t^2),$$

since to first order in δt only the product of the diagonal elements contributes, and the first-order part of that product is the trace of $\partial v_a / \partial \Gamma_b$, which is the divergence. The Jacobian determinant of the full flow, $J_t = \det(\partial \Gamma_t / \partial \Gamma_0)$, is the product of the determinants of successive short steps, so $J_{t+\delta t} = [1 + (\nabla_\Gamma \cdot \mathbf{v}) \delta t] J_t$, that is,

$$\frac{dJ_t}{dt} = (\nabla_\Gamma \cdot \mathbf{v}) J_t. \quad (5.4)$$

Since $J_0 = 1$ and the divergence vanishes, $J_t = 1$ for all t . This is Liouville's theorem: Hamiltonian flow preserves phase-space volume, $d\Gamma_t = d\Gamma_0$. A region can be stretched, sheared, and folded by the flow, but its volume never changes.

5.2.2 The Liouville equation

Volume preservation constrains how probability can move. Let $\rho(\Gamma, t)$ be a normalized density, $\int \rho(\Gamma, t) d\Gamma = 1$, so that $\rho d\Gamma$ is the probability that the system lies in the volume element $d\Gamma$ at time t . Probability is carried along by the flow, so the probability inside a fixed region V of phase space can change only by flowing across its boundary ∂V ,

$$\frac{d}{dt} \int_V \rho d\Gamma = - \oint_{\partial V} \rho \mathbf{v} \cdot \mathbf{n} dS = - \int_V \nabla_\Gamma \cdot (\rho \mathbf{v}) d\Gamma, \quad (5.5)$$

where $\mathbf{n}$ is the outward normal and the second equality is Gauss's theorem in $2n$ dimensions. Since V is arbitrary, the integrands on the two sides must agree, which gives the continuity equation

$$\frac{\partial \rho}{\partial t} + \nabla_\Gamma \cdot (\rho \mathbf{v}) = 0, \quad (5.6)$$

the same equation that expresses conservation of mass in a fluid. Now expand the divergence, $\nabla_\Gamma \cdot (\rho \mathbf{v}) = \mathbf{v} \cdot \nabla_\Gamma \rho + \rho \nabla_\Gamma \cdot \mathbf{v}$. The second term vanishes by Eq. (5.3), and Hamilton's equations turn the first into

$$\mathbf{v} \cdot \nabla_\Gamma \rho = \sum_i \left(\dot{q}_i \frac{\partial \rho}{\partial q_i} + \dot{p}_i \frac{\partial \rho}{\partial p_i} \right) = \sum_i \left(\frac{\partial \rho}{\partial q_i} \frac{\partial H}{\partial p_i} - \frac{\partial \rho}{\partial p_i} \frac{\partial H}{\partial q_i} \right) = \{\rho, H\},$$

where the Poisson bracket is

$$\{A, B\} = \sum_i \left(\frac{\partial A}{\partial q_i} \frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i} \frac{\partial B}{\partial q_i} \right).$$

The continuity equation therefore becomes the Liouville equation

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0. \quad (5.7)$$

Its meaning is clearest along a trajectory. The rate of change of ρ seen by an observer who moves with a phase-space point is $\partial_t \rho + \mathbf{v} \cdot \nabla_\Gamma \rho = \partial_t \rho + \{\rho, H\}$, which vanishes. The value of the density is therefore carried unchanged along each trajectory,

$$\rho(\varphi_t(\Gamma_0), t) = \rho(\Gamma_0, 0). \quad (5.8)$$

The probability cloud can be deformed, stretched, and folded, but since volumes are preserved it is never diluted or concentrated.

5.2.3 Conservation of the fine-grained entropy

Relative to a fixed Liouville reference measure, the fine-grained Gibbs information entropy is

$$h_{\text{fg}}[\rho] = - \int d\Gamma \rho \ln \rho. \quad (5.9)$$

As discussed for continuous entropy in Chapter 2, the logarithm requires a reference density to be dimensionless. Here the reference is a constant phase-space cell measured with the invariant Liouville volume. Rescaling that cell shifts h_{fg} by a constant and does not affect its time derivative. An arbitrary noninvariant reference density would not have this property. Under Hamiltonian evolution this entropy is constant,

$$\frac{dh_{\text{fg}}}{dt} = 0. \quad (5.10)$$

The reason is already contained in Eq. (5.8): the flow moves the values of ρ to new places without changing them, and it preserves the volume that carries each value. Explicitly, change the integration variable from Γ to $\Gamma_0 = \varphi_t^{-1}(\Gamma)$, which has unit Jacobian:

$$h_{\text{fg}}[\rho(t)] = - \int d\Gamma \rho(\Gamma, t) \ln \rho(\Gamma, t) = - \int d\Gamma_0 \rho(\Gamma_0, 0) \ln \rho(\Gamma_0, 0) = h_{\text{fg}}[\rho(0)].$$

The same conclusion follows by differentiating and using the Liouville equation, and this route shows which boundary behavior is needed:

$$\frac{dh_{\text{fg}}}{dt} = - \int d\Gamma \partial_t \rho (\ln \rho + 1) = \int d\Gamma \{\rho, H\} (\ln \rho + 1) = \int d\Gamma \{\rho \ln \rho, H\}.$$

The last step is the chain rule, $\partial(\rho \ln \rho) = (\ln \rho + 1) \partial \rho$, applied to each derivative in the bracket. For any function G we have $\{G, H\} = \mathbf{v} \cdot \nabla_{\Gamma} G = \nabla_{\Gamma} \cdot (G \mathbf{v})$, again because $\nabla_{\Gamma} \cdot \mathbf{v} = 0$. The last integral is therefore the flux of $\rho \ln \rho$ through the boundary of phase space, and it vanishes for periodic boundaries or for densities that decay fast enough at large distances.

Status: exact

Equation (5.10) is exact for an isolated Hamiltonian system with a regular normalized density and suitable boundary behavior. Chaos does not change it.

This is the sharp form of the contradiction described in Section 5.1. The entropy of the most detailed statistical description of an isolated system is a constant of the motion, for every Hamiltonian. Whatever entropy increases when a gas fills its container must belong to a less detailed description.

5.3 Coarse graining

Liouville's theorem leaves us with a sharp question. The fine-grained entropy of an isolated system never changes, Eq. (5.10), yet a gas released into an empty chamber ends in a state to which every thermodynamic measurement assigns a larger entropy. Something other than h_{fg} must be increasing. Figure 5.1 shows what it is. A patch of probability prepared in a small region is carried by the Hamiltonian flow into long thin filaments. Their total area and the value of the density on them never change, so h_{fg} stays constant. But the filaments soon become thinner than anything an observer with finite resolution can distinguish, and to such an observer the probability looks spread uniformly over the whole region. The entropy that increases is the entropy of this blurred description. In this section we make that statement precise and identify the assumption that supplies the direction of time.

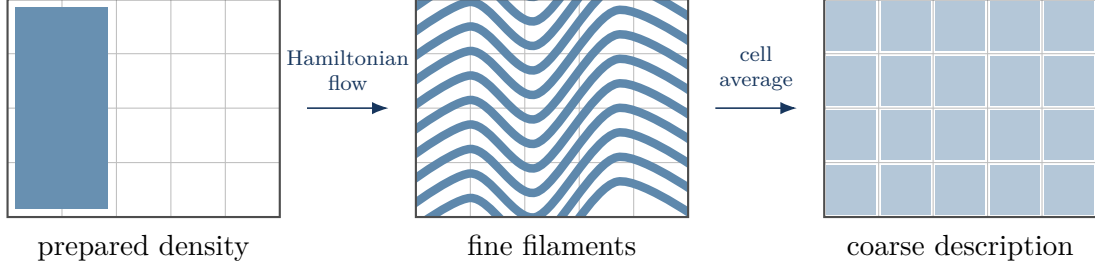


Figure 5.1: Hamiltonian flow preserves the area and fine-grained density of the initial patch but can stretch it into structures finer than the observation cells. Cell averaging deletes the filament positions and raises the coarse entropy.

5.3.1 Cells and the smoothed density

For an isolated system the accessible region of phase space is the energy shell of Chapter 3, the set of points with energy between E and $E + \Delta E$. We denote it by Γ_E and assume that its volume $|\Gamma_E|$ is finite. Partition Γ_E into nonoverlapping cells C_α whose union is the whole accessible region, and let $|C_\alpha|$ be the Liouville volume of C_α . The cells represent the resolution of the observer: two points in the same cell cannot be told apart. An observation that records only the cell label sees the coarse probabilities

$$p_\alpha(t) = \int_{C_\alpha} \rho(\Gamma, t) \, d\Gamma, \quad \sum_\alpha p_\alpha(t) = 1. \quad (5.11)$$

The least committal density compatible with these numbers, the one of maximum entropy in the sense of Section 2.7, spreads each p_α uniformly over its cell,

$$\bar{\rho}(\Gamma, t) = \frac{p_\alpha(t)}{|C_\alpha|}, \quad \Gamma \in C_\alpha, \quad (5.12)$$

which is normalized and reproduces every cell probability, since $\int_{C_\alpha} \bar{\rho} \, d\Gamma = p_\alpha$. The map $\rho \mapsto \bar{\rho}$ is a projection: it keeps every p_α , erases the structure of ρ inside each cell, and applying it twice gives the same result as applying it once. It acts on our description of the system, and the Hamiltonian plays no role in it.

The coarse-grained entropy in Gibbs's sense is the entropy of the smoothed density, $h_{\text{fg}}[\bar{\rho}]$. Because $\ln \bar{\rho}$ is constant on each cell, the phase-space integral collapses to a sum over cells,

$$h_{\text{fg}}[\bar{\rho}] = - \sum_\alpha p_\alpha \ln \frac{p_\alpha}{|C_\alpha|} = h_{\text{cg}} + \sum_\alpha p_\alpha \ln |C_\alpha|, \quad h_{\text{cg}} = - \sum_\alpha p_\alpha \ln p_\alpha, \quad (5.13)$$

where h_{cg} is the Shannon entropy of the discrete cell label. The first quantity is the continuous entropy of a density and depends on the reference volume, as in Chapter 2; the second is a dimensionless discrete entropy. For cells of equal volume v_c they differ by the constant $\ln v_c$, so every change in one is the same change in the other. We will use whichever form is more convenient and call both the coarse-grained entropy.

5.3.2 Averaging never lowers the entropy

Compare the exact density with its smoothed version at a single instant. Start from the relative entropy and split the logarithm,

$$D(\rho \parallel \bar{\rho}) = \int \rho \ln \frac{\rho}{\bar{\rho}} \, d\Gamma = -h_{\text{fg}}[\rho] - \int \rho \ln \bar{\rho} \, d\Gamma. \quad (5.14)$$

On each cell $\ln \bar{\rho}$ equals the constant $\ln(p_\alpha/|C_\alpha|)$, so the last integral only sees how much probability each cell holds,

$$\int \rho \ln \bar{\rho} \, d\Gamma = \sum_{\alpha} \ln \frac{p_{\alpha}}{|C_{\alpha}|} \int_{C_{\alpha}} \rho \, d\Gamma = \sum_{\alpha} p_{\alpha} \ln \frac{p_{\alpha}}{|C_{\alpha}|} = -h_{\text{fg}}[\bar{\rho}].$$

Note that ρ and $\bar{\rho}$ give the same value for this integral, which is the whole content of the calculation. Substitution gives

$$h_{\text{fg}}[\bar{\rho}] - h_{\text{fg}}[\rho] = D(\rho||\bar{\rho}) \geq 0. \quad (5.15)$$

Thus smoothing a density over the cells can only raise its entropy, and the increase equals the relative entropy between the exact and the smoothed densities. Equality holds exactly when ρ is already uniform inside every cell. The gap $D(\rho||\bar{\rho})$ measures the information about positions inside the cells that the smoothing discards.

Equation (5.15) compares two densities at the same instant. By itself it says nothing about time evolution.

5.3.3 Where the arrow of time enters

The time dependence comes from combining Eq. (5.15) with Liouville's theorem. Suppose the system is prepared at the resolution of the observer, that is, the initial density is uniform within each cell. A gas held by a partition in the left half of a box is the standard case: the preparation fixes which cells are occupied and nothing finer. Then $\bar{\rho}(0) = \rho(0)$ and, for every later time t ,

$$h_{\text{fg}}[\bar{\rho}(t)] \geq h_{\text{fg}}[\rho(t)] = h_{\text{fg}}[\rho(0)] = h_{\text{fg}}[\bar{\rho}(0)]. \quad (5.16)$$

The first step is Eq. (5.15) applied at time t , the second is Liouville conservation, and the third is the preparation assumption. The coarse-grained entropy at any later time is therefore at least its initial value. The excess, $D(\rho(t)||\bar{\rho}(t))$, is the information that the flow has carried into structures finer than the cells.

Two features of this argument deserve emphasis. First, the arrow of time is supplied entirely by the preparation. The Hamiltonian flow is reversible, and Eq. (5.16) holds equally well for $t < 0$: running the dynamics backward from the same specially prepared state also raises the coarse entropy. The coarse entropy is smallest at the moment of preparation, and the future is distinguished by our practice of preparing low-entropy states and then watching them evolve.

Second, Eq. (5.16) compares time t with time 0 only. It does not give $h_{\text{fg}}[\bar{\rho}(t_2)] \geq h_{\text{fg}}[\bar{\rho}(t_1)]$ for $0 < t_1 < t_2$, because $\rho(t_1)$ is in general not uniform inside the cells and the argument cannot be restarted at t_1 . The coarse entropy can decrease between two later times; Poincaré recurrences and specially prepared correlations, discussed in Section 5.6.4, are examples. The result also depends on the partition, since the preparation assumption refers to a particular set of cells. The thermodynamic statement concerns reproducible macroscopic preparations and overwhelmingly likely behavior for large systems.

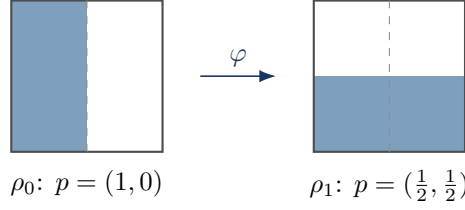
Worked example: the baker's map

Take the unit square as phase space and the baker's map

$$\varphi(x, y) = \begin{cases} (2x, y/2), & 0 \leq x < \frac{1}{2}, \\ (2x - 1, (y + 1)/2), & \frac{1}{2} \leq x < 1, \end{cases}$$

which stretches the square horizontally by a factor 2, compresses it vertically by the same factor, cuts it in half, and stacks the right half on top of the left. Its Jacobian is one and

it is invertible, so it is a discrete-time model of an area-preserving, reversible flow. Nearby points separate as 2^n along x after n steps, which makes it a standard model of chaotic mixing.



Prepare $\rho_0 = 2$ on the left half and zero on the right, and observe with the two cells $C_L = \{x < \frac{1}{2}\}$ and $C_R = \{x \geq \frac{1}{2}\}$, shown dashed. Then $p = (1, 0)$, the preparation is uniform within the cells so $\bar{\rho}_0 = \rho_0$, and $h_{\text{fg}}[\rho_0] = -\ln 2$ because the density equals 2 on an area $\frac{1}{2}$. After one step the left half has become the bottom half, $\rho_1 = 2$ for $y < \frac{1}{2}$ across the full width, and still $h_{\text{fg}}[\rho_1] = -\ln 2$. But each cell now holds half the probability, so $p = (\frac{1}{2}, \frac{1}{2})$, $\bar{\rho}_1 = 1$ everywhere, and

$$h_{\text{fg}}[\bar{\rho}_1] = 0, \quad h_{\text{fg}}[\bar{\rho}_1] - h_{\text{fg}}[\rho_1] = \ln 2 = D(\rho_1 \| \bar{\rho}_1).$$

The coarse entropy has risen by $\ln 2$ while the fine-grained entropy is unchanged. With $|C_\alpha| = \frac{1}{2}$, Eq. (5.13) reads $0 = \ln 2 + \ln \frac{1}{2}$. Further steps cut the bottom strip into horizontal stripes of width 2^{-n} , each spanning the full width, so $p = (\frac{1}{2}, \frac{1}{2})$ from then on and the coarse entropy stays at its two-cell maximum $\ln 2$.

Now observe the same evolution with the horizontal cells $C_B = \{y < \frac{1}{2}\}$ and $C_T = \{y \geq \frac{1}{2}\}$. The initial density fills the left half of each, so $p = (\frac{1}{2}, \frac{1}{2})$ at $t = 0$ and $p = (1, 0)$ after one step: h_{cg} falls from $\ln 2$ to 0. Nothing here contradicts Eq. (5.16). Its preparation hypothesis fails, because relative to these cells ρ_0 carries structure, the left–right split, that the observer cannot resolve. Finally, applying φ^{-1} to ρ_1 returns ρ_0 exactly, which is Loschmidt’s reversibility objection in miniature.

5.3.4 A closed dynamics for the cell probabilities

The observer of Section 5.3.1 records only the cell probabilities $p_\alpha(t)$. A macroscopic description works with variables of this kind alone: the equations of hydrodynamics, for instance, evolve the density and the temperature of a fluid without ever referring to the positions of its molecules. It is natural to ask whether the p_α also obey an equation of their own, one that gives $p(t + \tau)$ from $p(t)$ without returning to the full density ρ .

The exact evolution does not supply such an equation. Over an interval τ , Eq. (5.8) gives $\rho(\Gamma, t + \tau) = \rho(\varphi_\tau^{-1}(\Gamma), t)$, so the probability in cell C_α at time $t + \tau$ is

$$p_\alpha(t + \tau) = \int_{C_\alpha} \rho(\varphi_\tau^{-1}(\Gamma), t) \, d\Gamma = \int_{\varphi_\tau^{-1}(C_\alpha)} \rho(\Gamma, t) \, d\Gamma. \quad (5.17)$$

In the second step we changed the integration variable to $\varphi_\tau^{-1}(\Gamma)$, which has unit Jacobian, and $\varphi_\tau^{-1}(C_\alpha)$ denotes the set of points that the flow carries into C_α during the interval. This set generally cuts across many cells and covers only parts of them. The integral therefore depends on how $\rho(t)$ is distributed inside the cells, which is precisely the information that the cell probabilities do not contain. The baker’s map has already shown the consequence. With horizontal cells, the prepared density ρ_0 has $p = (\frac{1}{2}, \frac{1}{2})$ and moves to $p = (1, 0)$ in one step, while the uniform density $\rho = 1$, which has the same cell probabilities, is invariant under the

map and keeps $p = (\frac{1}{2}, \frac{1}{2})$ forever. Two densities with equal cell probabilities at one time have different cell probabilities one step later. The p_α at a given time do not determine their own future.

A closed description requires an additional assumption. The simplest one supposes that, at the beginning of each interval, the density is uniform inside every cell, so that $\rho(t)$ may be replaced by its smoothed version $\bar{\rho}(t)$ of Eq. (5.12). The physical picture is that the flow stirs the probability inside each cell into structures so fine that, after a time τ , they no longer influence where probability goes next. The interval τ is then a memory time: after it, the arrangement of probability inside the cells has been forgotten and only the cell probabilities matter. We will call this the assumption of repeated coarse graining. It concerns both the dynamics and the choice of cells, and nothing guarantees it in general.

With this assumption one step can be computed. Insert $\bar{\rho} = p_\beta/|C_\beta|$ on each cell C_β into Eq. (5.17):

$$p_\alpha(t + \tau) = \sum_\beta \frac{p_\beta(t)}{|C_\beta|} |\varphi_\tau^{-1}(C_\alpha) \cap C_\beta| = \sum_\beta T_{\alpha\beta} p_\beta(t), \quad T_{\alpha\beta} = \frac{|\varphi_\tau^{-1}(C_\alpha) \cap C_\beta|}{|C_\beta|}. \quad (5.18)$$

The matrix element $T_{\alpha\beta}$ has a direct meaning. The set $\varphi_\tau^{-1}(C_\alpha) \cap C_\beta$ contains the points of C_β that the flow carries into C_α , so $T_{\alpha\beta}$ is the fraction of the volume of C_β that ends up in C_α . Since probability is spread uniformly over C_β , this fraction is also the probability of finding the system in C_α at time $t + \tau$, given that it was in C_β at time t . In the notation of Eq. (2.70), $T_{\alpha\beta} = T_{\alpha|\beta}$, and T is a stochastic map that takes the cell distribution at one time to the cell distribution at the next.

The matrix T has two properties. First, every point of C_β is carried into exactly one cell, so the sets $\varphi_\tau^{-1}(C_\alpha) \cap C_\beta$ for different α divide C_β into nonoverlapping pieces, and

$$\sum_\alpha T_{\alpha\beta} = \frac{1}{|C_\beta|} \sum_\alpha |\varphi_\tau^{-1}(C_\alpha) \cap C_\beta| = 1. \quad (5.19)$$

This is conservation of probability. Second, a consequence of Liouville's theorem, the cells that feed C_α must supply exactly its volume. Let

$$\pi_\alpha = \frac{|C_\alpha|}{|\Gamma_E|} \quad (5.20)$$

be the fraction of the accessible region occupied by C_α , that is, the probability of C_α in the microcanonical distribution of Chapter 3. Then

$$\sum_\beta T_{\alpha\beta} \pi_\beta = \frac{1}{|\Gamma_E|} \sum_\beta |\varphi_\tau^{-1}(C_\alpha) \cap C_\beta| = \frac{|\varphi_\tau^{-1}(C_\alpha)|}{|\Gamma_E|} = \frac{|C_\alpha|}{|\Gamma_E|} = \pi_\alpha, \quad (5.21)$$

where the third equality is Liouville's theorem applied to the set $\varphi_\tau^{-1}(C_\alpha)$. Thus $T\pi = \pi$: the microcanonical cell distribution is stationary, and Eq. (5.21) is Liouville's theorem written at the resolution of the cells. When all M cells have the same volume, $\pi_\alpha = 1/M$ and the condition reduces to $\sum_\beta T_{\alpha\beta} = 1$. A matrix whose columns and rows all add to one is called doubly stochastic.

Under repeated coarse graining the cell probabilities obey $p(n+1) = Tp(n)$, where n counts intervals of length τ . The coarse entropy then increases at every step, and the data-processing inequality of Chapter 2 shows why. Note first that the coarse-grained Gibbs entropy is a relative entropy in disguise. Writing $p_\alpha/|C_\alpha| = (p_\alpha/\pi_\alpha)/|\Gamma_E|$ in Eq. (5.13),

$$h_{\text{fg}}[\bar{\rho}] = - \sum_\alpha p_\alpha \ln \frac{p_\alpha}{|C_\alpha|} = \ln |\Gamma_E| - D(p||\pi). \quad (5.22)$$

For M cells of equal volume this reads $h_{\text{cg}} = \ln M - D(p||u)$, with $u_\alpha = 1/M$. Now apply Eq. (2.74) to the pair $(p(n), \pi)$ and use $T\pi = \pi$:

$$D(p(n+1)||\pi) = D(Tp(n)||T\pi) \leq D(p(n)||\pi). \quad (5.23)$$

By Eq. (5.22), $h_{\text{fg}}[\bar{\rho}(n+1)] \geq h_{\text{fg}}[\bar{\rho}(n)]$ for every n . The coarse entropy grows step after step and stops growing only at a distribution that T leaves unchanged.

Compare this with the evolution of Section 5.3.3, in which the exact density evolves and the observer only looks at it through the cells. There the coarse entropy is guaranteed only to exceed its value at the preparation, and it can rise and fall at later times. Under repeated coarse graining it rises at every step. The difference comes entirely from the assumption. Replacing ρ by $\bar{\rho}$ at each step erases the arrangement of probability inside the cells, and in the exact evolution that arrangement can later return to the level of the cells and lower the coarse entropy (Exercise 5.2 asks for an example). The monotonic growth has been built into the description. Its physical justification is the separation between the memory time τ and the time over which the p_α change appreciably.

The sequence of cells occupied at the times $0, \tau, 2\tau, \dots$ is then a Markov chain in the sense of Eq. (2.75): the probability of the next cell depends only on the present one, through the fixed matrix T . Section 5.4 formulates dynamics of this kind in continuous time.

Worked example: a three-point phase space

The smallest model that shows both the failure of closure and the effect of repeated coarse graining has three microstates, 1, 2, and 3, and the deterministic dynamics $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$, one step per unit time. A permutation of a finite set is the discrete counterpart of a Hamiltonian flow: it is invertible, and it preserves the number of points in every set, which plays the role of volume. Take the two cells $A = \{1\}$ and $B = \{2, 3\}$, with volumes $|A| = 1$ and $|B| = 2$, so that $\pi = (\frac{1}{3}, \frac{2}{3})$.

Start with certainty in microstate 1. The exact evolution visits $1, 2, 3, 1, \dots$, and the cell probabilities are

$$p(0) = (1, 0), \quad p(1) = (0, 1), \quad p(2) = (0, 1), \quad p(3) = (1, 0), \quad \dots$$

The cell distribution $(0, 1)$ is followed by $(0, 1)$ at the first step and by $(1, 0)$ at the second. Which one occurs depends on whether the system sits in microstate 2 or 3, a distinction that the cells do not record. With $h_{\text{fg}}[\bar{\rho}] = -\sum_\alpha p_\alpha \ln(p_\alpha/|C_\alpha|)$, the coarse entropy runs through $0, \ln 2, \ln 2, 0, \dots$ and returns to zero every three steps.

Under repeated coarse graining, Eq. (5.18) gives the matrix directly. Cell A contains only microstate 1, which moves to B , so $T_{AA} = 0$ and $T_{BA} = 1$. Of the two points of B , microstate 2 stays in B and microstate 3 moves to A , so $T_{AB} = T_{BB} = \frac{1}{2}$. Thus

$$T = \begin{pmatrix} 0 & \frac{1}{2} \\ 1 & \frac{1}{2} \end{pmatrix}, \quad T\pi = \pi,$$

and the columns add to one, as Eqs. (5.19) and (5.21) require. Iterating $p(n+1) = Tp(n)$ from $p(0) = (1, 0)$ gives $p_A(n+1) = \frac{1}{2}[1 - p_A(n)]$, whose solution is $p_A(n) = \frac{1}{3} + \frac{2}{3}(-\frac{1}{2})^n$:

n	0	1	2	3	4
$p(n)$	(1, 0)	(0, 1)	($\frac{1}{2}, \frac{1}{2}$)	($\frac{1}{4}, \frac{3}{4}$)	($\frac{3}{8}, \frac{5}{8}$)
$h_{\text{fg}}[\bar{\rho}(n)]$	0	0.693	1.040	1.082	1.095

The coarse entropy now increases at every step toward its maximum $\ln |\Gamma_E| = \ln 3 \approx 1.099$, reached at $p = \pi$. The exact evolution and the Markov chain agree for the first step, while the state is still the one prepared at the resolution of the cells, and they separate afterward.

5.3.5 Ergodicity, mixing, and chaos

The assumption of repeated coarse graining is a statement about the dynamics: the flow must scramble the inside of each cell quickly. The exact evolution of Section 5.3.3 raises a related question. Equation (5.16) guarantees that the coarse entropy at later times is at least its initial value, but it does not say that the cell probabilities approach their equilibrium values π_α . For all that equation tells us, probability could keep sloshing between the cells forever. Three properties of dynamical systems are commonly invoked at this point, and we need to know what each of them delivers.

Throughout this subsection the system is isolated and confined to the energy shell Γ_E . Energy is conserved, so the flow maps the shell into itself, and by Liouville's theorem it preserves the volume of every subset of the shell. The microcanonical average of a function A is its average over the shell with uniform weight,

$$\langle A \rangle_{\text{mc}} = \frac{1}{|\Gamma_E|} \int_{\Gamma_E} A(\Gamma) \, d\Gamma. \quad (5.24)$$

Ergodicity. Boltzmann justified the microcanonical ensemble with the idea that a single trajectory, followed for a long time, visits every part of the energy shell and spends in each region a fraction of the time proportional to its volume. The mathematical form of this idea is ergodicity: for every integrable function A and almost every initial point Γ on the shell,

$$\lim_{t \rightarrow \infty} \frac{1}{t} \int_0^t A(\varphi_s(\Gamma)) \, ds = \langle A \rangle_{\text{mc}}. \quad (5.25)$$

The left side is the time average of A along one trajectory, and the words “almost every” allow exceptional initial points, such as periodic orbits, provided they occupy zero volume. Taking for A the indicator function $\mathbf{1}_C$ of a set C , equal to one inside C and zero outside, shows what the definition says: the fraction of time that a typical trajectory spends in C equals $|C|/|\Gamma_E|$.

Ergodicity concerns time averages along single trajectories, and by itself it does not make a probability density spread out. The simplest example is the translation $\varphi_t(x, y) = (x + t, y + \gamma t)$ on the unit torus, the unit square with opposite sides identified, with an irrational slope γ . It preserves area, and every trajectory is a straight line that winds around the torus without ever closing and covers it uniformly, so the flow is ergodic. But it moves every point by the same amount. A patch of probability prepared in one cell is translated rigidly, and the cell probabilities keep oscillating forever.

Mixing. What coarse-grained relaxation needs is that every prepared patch becomes evenly spread over the shell at the resolution of the cells. Gibbs described this with the image of a drop of ink stirred into water [16]: the volume of ink is conserved, but after enough stirring every macroscopic portion of the water contains ink in proportion to its size. A flow is called mixing if, for every pair of regions R and C of the shell,

$$\frac{|\varphi_t(R) \cap C|}{|R|} \longrightarrow \frac{|C|}{|\Gamma_E|} \quad (t \rightarrow \infty). \quad (5.26)$$

Read R as the region where probability is initially placed and C as an observation cell: the fraction of the evolved patch that lies in C approaches the fraction of the shell that C occupies. Writing $|\varphi_t(R) \cap C| = \int \mathbf{1}_C(\varphi_t(\Gamma)) \mathbf{1}_R(\Gamma) \, d\Gamma$ and allowing general functions in place of the two indicator functions gives the equivalent form

$$\langle A(\varphi_t(\Gamma)) B(\Gamma) \rangle_{\text{mc}} \longrightarrow \langle A \rangle_{\text{mc}} \langle B \rangle_{\text{mc}} \quad (t \rightarrow \infty), \quad (5.27)$$

which says that the value of an observable at a late time becomes uncorrelated with the value of another observable at time zero.

Mixing is the property that brings the cell probabilities to equilibrium in the exact evolution. Let $\rho(\Gamma, 0)$ be a regular normalized density on the shell. By Eq. (5.8) and a change of variables with unit Jacobian,

$$p_\alpha(t) = \int_{C_\alpha} \rho(\Gamma, t) \, d\Gamma = \int \mathbf{1}_{C_\alpha}(\varphi_t(\Gamma)) \rho(\Gamma, 0) \, d\Gamma = |\Gamma_E| \langle \mathbf{1}_{C_\alpha}(\varphi_t(\Gamma)) \rho(\Gamma, 0) \rangle_{\text{mc}}.$$

Equation (5.27) with $A = \mathbf{1}_{C_\alpha}$ and $B = \rho(\cdot, 0)$, together with $\langle \mathbf{1}_{C_\alpha} \rangle_{\text{mc}} = |C_\alpha|/|\Gamma_E|$ and $\langle \rho(\cdot, 0) \rangle_{\text{mc}} = 1/|\Gamma_E|$, gives

$$p_\alpha(t) \longrightarrow \frac{|C_\alpha|}{|\Gamma_E|} = \pi_\alpha \quad (t \rightarrow \infty). \quad (5.28)$$

For every fixed partition and every regular initial density, the coarse probabilities approach their microcanonical values and the coarse entropy approaches its maximum, while the fine-grained entropy stays constant throughout. The baker's map is mixing, and in the example above the limit is reached after a single step. Mixing supplies the limit, but neither a time scale nor monotonic approach: on the way, the p_α can overshoot and oscillate.

Chaos. The third property concerns how fast fine structure forms. A flow is chaotic when nearby trajectories separate exponentially,

$$|\delta\Gamma(t)| \approx e^{\lambda t} |\delta\Gamma(0)|, \quad (5.29)$$

over a long range of times, where $\lambda > 0$ is the largest Lyapunov exponent. Since volume is conserved, stretching in one direction must be compensated by contraction in another, and a small patch is drawn into a filament whose length grows roughly as $e^{\lambda t}$ while its width shrinks. This gives an estimate of the memory time of repeated coarse graining. A patch of linear size δ_0 must be stretched to the size ℓ of a cell before its arrangement inside the cells ceases to matter, which takes a time of order

$$t_\ell \sim \frac{1}{\lambda} \ln \frac{\ell}{\delta_0}. \quad (5.30)$$

The logarithm keeps this time short even when δ_0 is many orders of magnitude smaller than ℓ . In the baker's map nearby points separate as 2^n along x , so $\lambda = \ln 2$ per step, and a vertical strip of width 2^{-k} is stretched across the full width of the square in $k = \ln(2^k)/\ln 2$ steps, as Eq. (5.30) predicts.

The three notions are related but distinct. Ergodicity concerns time averages along single trajectories. Mixing makes every regular density spread over the shell at any fixed resolution, and it implies ergodicity. Chaos, through λ , estimates how fast fine structure forms. None of the three changes the fine-grained entropy, and none of them alone produces the step-by-step growth of Section 5.3.4, which needed the assumption of repeated coarse graining on top of the dynamics. For realistic many-particle Hamiltonians these properties are rarely proven. Mixing has been established for model systems such as the baker's map and certain billiards, and for gases and liquids it remains a well-supported working assumption.

Status: heuristic

The picture of a probability cloud that is stretched, folded, and becomes macroscopically uniform is useful, and Eq. (5.30) is an order-of-magnitude estimate. The entropy increase appears only after finite resolution or restriction to chosen observables.

Lecture 14 ends here. Lecture 15 starts from a Markov description of mesoscopic transitions.

5.4 Master equations

Section 5.3.4 produced a description in which a coarse state evolves through a fixed matrix of conditional probabilities. The same structure appears wherever a system jumps between distinguishable states while the details of each jump remain hidden: a molecule switching between two conformations, an electron hopping between impurity sites in a semiconductor, a spin flipped by lattice vibrations, a chemical reaction changing the number of molecules of each species. In these situations the dynamics is formulated in continuous time, through transition rates. In this section we derive the corresponding evolution equation, the master equation, and study where it leads.

5.4.1 From transition probabilities to rates

Let the system have a finite number of mesostates, labeled $i = 1, \dots, M$. “Mesostate” is a convenient name for a state of an intermediate description: a cell of phase space, a conformation of a molecule, the number of particles in a region. We assume that the evolution is Markovian. That is, the probability of finding the system in state i at time $t + \Delta t$ depends on its past only through the state it occupies at time t . The evolution over an interval Δt is then described by the conditional probabilities

$$T_{ij}(\Delta t) = \mathbb{P}(\text{state } i \text{ at time } t + \Delta t \mid \text{state } j \text{ at time } t), \quad (5.31)$$

which do not depend on t when the environment of the system does not change in time, and

$$p_i(t + \Delta t) = \sum_j T_{ij}(\Delta t) p_j(t). \quad (5.32)$$

As for the matrix of Section 5.3.4, every column adds to one, $\sum_i T_{ij}(\Delta t) = 1$, because a system that starts in j must end up in some state.

In continuous time, a short interval produces a small change. We assume that the probability of a jump to a different state is proportional to the length of a short interval,

$$T_{ij}(\Delta t) = W_{ij} \Delta t + o(\Delta t), \quad i \neq j, \quad (5.33)$$

where $o(\Delta t)$ denotes terms that vanish faster than Δt . This defines the transition rate $W_{ij} \geq 0$ from j to i , the probability per unit time of a jump $j \rightarrow i$. Radioactive decay is the familiar example: a nucleus decays during a short interval dt with probability dt/τ_d , whatever its age, so its decay rate is $1/\tau_d$. The probability of staying in j follows from the normalization of column j ,

$$T_{jj}(\Delta t) = 1 - \sum_{i \neq j} W_{ij} \Delta t + o(\Delta t). \quad (5.34)$$

The sum $\sum_{i \neq j} W_{ij}$ is the total rate of leaving j , and its inverse is the mean time that the system spends in j before jumping.

Now insert these short-time probabilities into Eq. (5.32), separating the term $j = i$:

$$\begin{aligned} p_i(t + \Delta t) &= T_{ii}(\Delta t) p_i(t) + \sum_{j \neq i} T_{ij}(\Delta t) p_j(t) \\ &= p_i(t) - \sum_{j \neq i} W_{ji} \Delta t p_i(t) + \sum_{j \neq i} W_{ij} \Delta t p_j(t) + o(\Delta t). \end{aligned} \quad (5.35)$$

Subtracting $p_i(t)$, dividing by Δt , and letting $\Delta t \rightarrow 0$ gives the master equation

$$\dot{p}_i = \sum_{j \neq i} (W_{ij} p_j - W_{ji} p_i). \quad (5.36)$$

This is a balance equation for probability, and its two terms have a direct reading. The first, $W_{ij}p_j$, is the rate at which probability flows into i from j : the probability of being in j times the probability per unit time of jumping from j to i . The second, $W_{ji}p_i$, is the rate at which probability flows out of i into j . The probability of state i changes by the total gain minus the total loss, which is why Eq. (5.36) is also called a gain–loss equation. In terms of the net probability current from j to i ,

$$J_{ij} = W_{ij}p_j - W_{ji}p_i = -J_{ji}, \quad (5.37)$$

the master equation reads

$$\dot{p}_i = \sum_{j \neq i} J_{ij}. \quad (5.38)$$

This is the discrete counterpart of the continuity equation (5.6). The probability of a state changes only through the currents that flow into it from other states, just as the probability of a region of phase space changes only through the flux across its boundary.

Two properties follow at once. Summing Eq. (5.38) over i , every current appears twice with opposite signs, so $\sum_i \dot{p}_i = 0$ and normalization is preserved. And probabilities stay nonnegative: if $p_i = 0$ at some instant, then $\dot{p}_i = \sum_{j \neq i} W_{ij}p_j \geq 0$, and p_i cannot become negative.

Pauli wrote the first equation of this kind for a quantum system in 1928 [43], with transition rates computed from time-dependent perturbation theory, the rates now known as Fermi's golden rule. His derivation assumed that the phases of the quantum amplitudes are randomized at every step, the quantum counterpart of repeated coarse graining. In every application the Markov property is an assumption about time scales. Equation (5.33) requires Δt to be long compared with the memory time of the hidden degrees of freedom, so that successive jumps are independent, and short compared with the time over which the p_i change appreciably. The master equation is accurate when these two scales are well separated.

5.4.2 Stationary states and detailed balance

Where does the evolution lead? A distribution π is stationary if the master equation leaves it unchanged,

$$\sum_{j \neq i} (W_{ij} \pi_j - W_{ji} \pi_i) = 0 \quad \text{for every } i, \quad (5.39)$$

so that $p(0) = \pi$ implies $p(t) = \pi$ at all later times. Stationarity says that at every state the total inflow equals the total outflow, $\sum_j J_{ij} = 0$, in the same way that Kirchhoff's current law says that no charge accumulates at a node of an electric circuit. Currents may still flow, provided they circulate.

Figure 5.2 shows a simple example, a ring of three states with jumps $1 \rightarrow 2 \rightarrow 3 \rightarrow 1$ at rate k_+ and reverse jumps at rate k_- , that is, $W_{21} = W_{32} = W_{13} = k_+$ and $W_{12} = W_{23} = W_{31} = k_-$. The uniform distribution $\pi_i = \frac{1}{3}$ is stationary, since each state receives $(k_+ + k_-)/3$ per unit time from its two neighbors and loses the same amount. But the current from each state to the next,

$$J = J_{21} = J_{32} = J_{13} = \frac{k_+ - k_-}{3}, \quad (5.40)$$

is nonzero whenever $k_+ \neq k_-$. Probability circulates around the ring at a steady rate while the distribution stays fixed. Such a nonequilibrium steady state describes, for instance, a molecular motor that consumes one fuel molecule at each turn of its cycle. Something must keep the current going, and Section 5.5 identifies the cost.

Thermal equilibrium is more restrictive. In equilibrium no net current flows between any pair of states,

$$W_{ij} \pi_j = W_{ji} \pi_i \quad \text{for every pair } i, j, \quad (5.41)$$

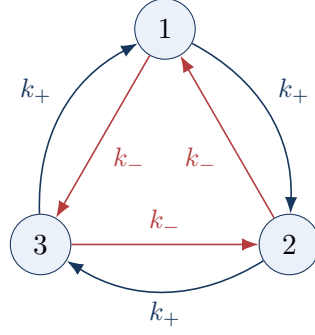


Figure 5.2: A ring of three states with clockwise rates k_+ and counterclockwise rates k_- . The uniform distribution is stationary for any rates, but for $k_+ \neq k_-$ it carries the circulating current $J = (k_+ - k_-)/3$.

a condition called detailed balance. Summing it over j gives Eq. (5.39), so detailed balance implies stationarity, while the ring shows that stationarity does not imply detailed balance: Eq. (5.41) holds there only for $k_+ = k_-$. The physical content of detailed balance is that in equilibrium every elementary process occurs as often as its reverse. A film of the jumps shown backward is then statistically indistinguishable from the film shown forward.

Detailed balance is inherited from the time-reversal symmetry of the microscopic dynamics, and the construction of Section 5.3.4 lets us see how. Let $\mathcal{R}(\mathbf{q}, \mathbf{p}) = (\mathbf{q}, -\mathbf{p})$ reverse all momenta. For a Hamiltonian even in the momenta, Eq. (5.1) states that reversing the momenta, evolving for a time τ , and reversing them again undoes the evolution, $\mathcal{R}\varphi_\tau\mathcal{R} = \varphi_\tau^{-1}$. Multiplying by $\mathcal{R}$ on the left and using $\mathcal{R}^2 = 1$ gives $\mathcal{R}\varphi_\tau^{-1} = \varphi_\tau\mathcal{R}$. Suppose also that every cell is invariant under $\mathcal{R}$, as happens when the cells record positions or energies and ignore the directions of the momenta. Then, since $\mathcal{R}$ and φ_τ preserve volume,

$$\begin{aligned} |\varphi_\tau^{-1}(C_\alpha) \cap C_\beta| &= |\mathcal{R}\varphi_\tau^{-1}(C_\alpha) \cap \mathcal{R}C_\beta| = |\varphi_\tau(C_\alpha) \cap C_\beta| \\ &= |C_\alpha \cap \varphi_\tau^{-1}(C_\beta)|. \end{aligned} \quad (5.42)$$

In the first step we applied $\mathcal{R}$ to the whole set; in the second we used $\mathcal{R}\varphi_\tau^{-1}(C_\alpha) = \varphi_\tau(\mathcal{R}C_\alpha) = \varphi_\tau(C_\alpha)$ and $\mathcal{R}C_\beta = C_\beta$; in the third we applied φ_τ^{-1} to the whole set. The volume carried from C_β into C_α equals the volume carried from C_α into C_β . Dividing by $|\Gamma_E|$ and comparing with Eq. (5.18), this reads $T_{\alpha\beta}\pi_\beta = T_{\beta\alpha}\pi_\alpha$ with $\pi_\alpha = |C_\alpha|/|\Gamma_E|$, which is detailed balance with respect to the microcanonical distribution. For a system that exchanges energy with a heat bath, the same reasoning applied to system and bath together leads to detailed balance with respect to the canonical distribution; Section 5.5 reaches this conclusion by a more direct route.

5.4.3 Relaxation and relative entropy

We now show that the master equation drives every distribution toward the stationary one, and we identify a quantity that decreases along the way. For repeated coarse graining that quantity was the relative entropy $D(p||\pi)$, by Eq. (5.23), and the same holds for every master equation, with or without detailed balance.

Let π be stationary with $\pi_i > 0$ for every state, and write $r_i = p_i/\pi_i$ for the ratio of the actual to the stationary probability. Differentiating $D(p||\pi) = \sum_i p_i \ln r_i$ and using $\sum_i \dot{p}_i = 0$,

$$\frac{d}{dt}D(p||\pi) = \sum_i \dot{p}_i (\ln r_i + 1) = \sum_i \dot{p}_i \ln r_i.$$

Insert the master equation and, in the loss term, rename the summation indices $i \leftrightarrow j$:

$$\begin{aligned} \frac{d}{dt} D(p||\pi) &= \sum_{i \neq j} W_{ij} p_j \ln r_i - \sum_{i \neq j} W_{ji} p_i \ln r_i \\ &= \sum_{i \neq j} W_{ij} p_j (\ln r_i - \ln r_j) = \sum_{i \neq j} W_{ij} \pi_j r_j \ln \frac{r_i}{r_j}. \end{aligned} \quad (5.43)$$

Each logarithm is bounded by the elementary inequality $\ln u \leq u - 1$ of Eq. (2.63), with $u = r_i/r_j$:

$$\frac{d}{dt} D(p||\pi) \leq \sum_{i \neq j} W_{ij} \pi_j r_j \left(\frac{r_i}{r_j} - 1 \right) = \sum_{i \neq j} W_{ij} \pi_j (r_i - r_j).$$

The right side vanishes by stationarity. In the term with r_i sum first over j ; in the term with r_j rename $i \leftrightarrow j$ and then sum over j . The result is $\sum_i r_i \sum_{j \neq i} (W_{ij} \pi_j - W_{ji} \pi_i)$, which is zero by Eq. (5.39). Therefore

$$\frac{d}{dt} D(p(t)||\pi) \leq 0. \quad (5.44)$$

Equality requires equality in $\ln u \leq u - 1$, that is, $r_i = r_j$ for every pair of states connected by a nonzero rate. Suppose the process is irreducible, meaning that every state can be reached from every other through a chain of allowed jumps. Then equality forces all r_i to be equal, and normalization gives $p = \pi$. The relative entropy is thus a Lyapunov function: it decreases strictly until the stationary distribution is reached. In particular, an irreducible process has a single stationary distribution. If π' were another one, the constant solution $p(t) = \pi'$ would give $dD(\pi'||\pi)/dt = 0$, which forces $\pi' = \pi$. A further argument, which we omit, shows that every solution converges to π .

The same result is the data-processing inequality in another guise. The solution of the master equation can be written $p(t) = K_t p(0)$, where K_t is a product of many short-time matrices $T(\Delta t)$ and is therefore itself a matrix of conditional probabilities, with $K_t \pi = \pi$. Then

$$D(p(t)||\pi) = D(K_t p(0)||K_t \pi) \leq D(p(0)||\pi),$$

exactly as in Eq. (5.23). The master equation is the continuous-time version of repeated coarse graining, and the relative entropy to the stationary state decreases for the same reason: each step of the dynamics loses information about where the system came from.

When the stationary state is the canonical distribution $\pi_i = e^{-\beta E_i}/Z$ with $T > 0$, the free-energy identity (3.44) gives $k_B T D(p||\pi) = \mathcal{F}_T[p] - F_{\text{eq}}$, and Eq. (5.44) becomes

$$\frac{d}{dt} \mathcal{F}_T[p(t)] \leq 0. \quad (5.45)$$

A system relaxing in contact with a heat bath lowers its nonequilibrium free energy until it reaches equilibrium. This is the statistical counterpart of the minimum principle for F that Chapter 1 derived from the second law. For a normalizable Gibbs distribution at $T < 0$ the relative entropy still decreases, but the factor $k_B T < 0$ reverses the free-energy inequality.

Status: exact

Equation (5.44) holds for every master equation with a positive stationary distribution. Deriving a master equation from reversible microscopic dynamics is a separate step, which requires a separation of time scales and the loss of memory described above.

Worked example: two-state relaxation

Let $0 \rightleftharpoons 1$ have rates $a = W_{10}$ and $b = W_{01}$. Writing $p(t) = p_1(t)$, the master equation is

$$\dot{p} = a(1 - p) - bp = -(a + b)(p - \pi_1), \quad \pi_1 = \frac{a}{a + b}. \quad (5.46)$$

Therefore

$$p(t) = \pi_1 + [p(0) - \pi_1]e^{-(a+b)t}. \quad (5.47)$$

The probability approaches stationarity with relaxation time $(a + b)^{-1}$. With $\pi_0 = b/(a + b)$ we have $a\pi_0 = b\pi_1$, so this process satisfies detailed balance, as every two-state process does. The relative entropy is

$$D(p||\pi) = p \ln \frac{p}{\pi_1} + (1 - p) \ln \frac{1 - p}{1 - \pi_1}. \quad (5.48)$$

Differentiating gives

$$\dot{D} = \dot{p} \ln \frac{p(1 - \pi_1)}{\pi_1(1 - p)} \quad (5.49)$$

$$= -(a + b)(p - \pi_1) \ln \frac{p(1 - \pi_1)}{\pi_1(1 - p)} \leq 0. \quad (5.50)$$

The logarithm vanishes at $p = \pi_1$ and increases with p , so it has the same sign as $p - \pi_1$. This model separates the relaxation rate $a + b$, the stationary state π , and the Lyapunov function $D(p||\pi)$.

Worked example: the Ehrenfest urn model

In 1907 Paul and Tatiana Ehrenfest proposed a model to show how relaxation and recurrence coexist [13]. Here is a continuous-time version. N numbered balls are distributed between two urns, A and B , and each ball, independently of the others, jumps to the other urn at rate ν . A microstate specifies the urn of every ball, so there are 2^N microstates. The coarse description keeps only the number n of balls in urn A .

The rates for n follow by counting. A jump $n \rightarrow n - 1$ occurs when any of the n balls in A moves, and a jump $n \rightarrow n + 1$ when any of the $N - n$ balls in B moves:

$$W_{n-1,n} = n\nu, \quad W_{n+1,n} = (N - n)\nu.$$

Two balls jump within the same short interval Δt with probability of order $(\nu\Delta t)^2$, so no other transitions survive in Eq. (5.33). The master equation is

$$\dot{p}_n = \nu(N - n + 1)p_{n-1} + \nu(n + 1)p_{n+1} - \nu N p_n.$$

Its stationary distribution is the fraction of microstates with n balls in A ,

$$\pi_n = \binom{N}{n} 2^{-N},$$

and it satisfies detailed balance: the identity $(n + 1)\binom{N}{n+1} = (N - n)\binom{N}{n}$ gives $W_{n+1,n}\pi_n = W_{n,n+1}\pi_{n+1}$. The binomial distribution has mean $N/2$ and standard deviation $\sqrt{N}/2$, so for large N the urns are almost always nearly balanced.

The mean number in A changes because each of the $N - n$ balls in B enters A at rate ν , while each of the n balls in A leaves it at rate ν :

$$\frac{d\langle n \rangle}{dt} = \nu(N - \langle n \rangle) - \nu\langle n \rangle, \quad \langle n \rangle(t) - \frac{N}{2} = \left[\langle n \rangle(0) - \frac{N}{2} \right] e^{-2\nu t}.$$

The same equation follows by multiplying the master equation by n and summing. Starting from all balls in A , the imbalance decays on the time scale $1/(2\nu)$, independently of N , and $D(p(t)||\pi)$ decreases from $\ln(1/\pi_N) = N \ln 2$ toward zero.

Yet the initial state recurs. In the stationary regime the system spends a fraction $\pi_N = 2^{-N}$ of the time in the state $n = N$, and each visit lasts on average $1/(N\nu)$, the inverse of the total rate of leaving. Successive visits to $n = N$ are therefore separated, on average, by

$$t_{\text{rec}} = \frac{1/(N\nu)}{\pi_N} = \frac{2^N}{N\nu}.$$

For $N = 100$ and $\nu = 1 \text{ s}^{-1}$ this is $1.3 \times 10^{28} \text{ s}$, about 4×10^{20} years, or 3×10^{10} times the age of the universe, while the relaxation takes half a second. Kac analyzed the model in these terms [21].

5.5 Entropy production in jump processes

The decrease of $D(p||\pi)$ is a second law for systems that relax toward a stationary state, and it has two limitations. It refers to a stationary distribution that must be known in advance. And in a nonequilibrium steady state, such as the driven ring of Fig. 5.2, $D(p||\pi)$ is already zero while the system keeps converting fuel into heat. We want a quantity with the meaning that the entropy production $d_i S$ has in Chapter 1: the total entropy change of the system and its environment, nonnegative in every process and zero only in equilibrium. We construct it from the master equation in three steps, the entropy of the system, the entropy of the environment, and their sum.

5.5.1 Entropy of the system

The dimensionless Shannon entropy of the system, $h(p) = -\sum_i p_i \ln p_i$, changes at the rate

$$\begin{aligned} \dot{h} &= -\sum_i \dot{p}_i (\ln p_i + 1) = -\sum_i \dot{p}_i \ln p_i = -\sum_{i,j} J_{ij} \ln p_i \\ &= \frac{1}{2} \sum_{i,j} J_{ij} \ln \frac{p_j}{p_i}. \end{aligned} \quad (5.51)$$

The first line uses $\sum_i \dot{p}_i = 0$ and the master equation in the form (5.38), with $J_{ii} = 0$. For the second line, rename $i \leftrightarrow j$ in the last sum of the first line, which turns it into $-\sum_{i,j} J_{ji} \ln p_j = \sum_{i,j} J_{ij} \ln p_j$ because $J_{ji} = -J_{ij}$, and average the two forms. Each pair of states contributes the net current from j to i times the logarithm of p_j/p_i . When probability flows from a more probable to a less probable state, the distribution spreads and h increases. But h can also decrease, for instance when a cold bath draws the system into its ground state. To find the quantity that never decreases we must follow the entropy of the environment as well.

5.5.2 Entropy of the environment: local detailed balance

Every jump of the system is caused by its environment and leaves a trace there. Consider first a system with energy levels E_i in contact with a single heat bath at temperature T , with no other agent acting on it. Total energy is conserved, so in a jump $j \rightarrow i$ the bath energy changes by $-(E_i - E_j)$. To relate the rate of a jump to that of its reverse, we use the counting argument of Section 3.3. Let the bath have $\Omega_B(E_B)$ microstates at energy E_B . The jump $j \rightarrow i$ takes the bath from one microstate at energy E_B to any of the microstates at energy

$E_B - (E_i - E_j)$, and the reverse jump returns it from energy $E_B - (E_i - E_j)$ to energy E_B . Suppose that, as in Fermi's golden rule for a weak coupling, the rate of a jump is proportional to the number of final states available to the bath, with a coupling strength that is the same for a transition and its reverse. Then the ratio of the two rates is a ratio of bath multiplicities, which is the exponential of a bath entropy change:

$$\frac{W_{ij}}{W_{ji}} = \frac{\Omega_B(E_B - (E_i - E_j))}{\Omega_B(E_B)} = \exp \left[\frac{S_B(E_B - (E_i - E_j)) - S_B(E_B)}{k_B} \right] = e^{-\beta(E_i - E_j)}. \quad (5.52)$$

The last step uses $\partial S_B / \partial E_B = 1/T$ and neglects the curvature of S_B , as in the derivation of the canonical distribution in Chapter 3.

Equation (5.52) is called local detailed balance. It is local because it constrains each pair of rates separately, through the energy exchanged in that particular jump. It describes the bath, which is assumed to remain in equilibrium, and it makes no assumption about the state of the system. With a single bath it implies detailed balance (5.41) with respect to the canonical distribution $\pi_i \propto e^{-\beta E_i}$. Its general form is

$$\ln \frac{W_{ij}}{W_{ji}} = \frac{\Delta S_{\text{env}}(j \rightarrow i)}{k_B}, \quad (5.53)$$

where $\Delta S_{\text{env}}(j \rightarrow i)$ is the entropy change of the environment in the jump $j \rightarrow i$ [52]. In this form it continues to hold when the system is driven or coupled to several reservoirs, situations in which no global equilibrium exists. Consider the ring of Fig. 5.2 as a motor driven by a fuel that releases the free energy $\Delta\mu$ per turn into a bath at temperature T , with equal energies for the three states. Each turn returns the system to its initial state and delivers the entropy $\Delta\mu/T$ to the environment, so Eq. (5.53) applied around the cycle gives $(k_+/k_-)^3 = e^{\beta\Delta\mu}$.

Local detailed balance tells us how much entropy each jump delivers to the environment. Jumps $j \rightarrow i$ occur at the rate $W_{ij}p_j$, each changing the environment entropy by $k_B \ln(W_{ij}/W_{ji})$, and the reverse jumps occur at the rate $W_{ji}p_i$, each producing the opposite change. The net rate of entropy flow into the environment, in units of k_B , is therefore

$$\dot{h}_{\text{env}} = \sum_{i < j} (W_{ij}p_j - W_{ji}p_i) \ln \frac{W_{ij}}{W_{ji}} = \frac{1}{2} \sum_{i,j} J_{ij} \ln \frac{W_{ij}}{W_{ji}}, \quad (5.54)$$

where we wrote the sum over pairs as half the sum over ordered pairs, since the terms (i, j) and (j, i) are equal. With a single bath, Eq. (5.52) gives $\dot{h}_{\text{env}} = -\beta\dot{Q}$, where $\dot{Q} = \sum_{i < j} J_{ij}(E_i - E_j)$ is the rate at which the system absorbs heat: the bath loses at temperature T the energy that the system gains.

5.5.3 Total entropy production

Adding the entropy changes of system and environment gives the total entropy production rate

$$\dot{\Sigma} = \dot{h} + \dot{h}_{\text{env}} = \frac{1}{2} \sum_{i,j} J_{ij} \ln \frac{W_{ij} p_j}{W_{ji} p_i} \geq 0. \quad (5.55)$$

Every term is nonnegative. With $x = W_{ij}p_j$ and $y = W_{ji}p_i$, the term for the ordered pair (i, j) is $(x - y) \ln(x/y)$, and since the logarithm is increasing, $x - y$ and $\ln(x/y)$ always have the same sign. Equality holds only when $x = y$ for every pair, that is, when all currents vanish and p satisfies detailed balance. Terms with $x = y = 0$ contribute zero. A jump allowed in one direction only would give an infinite contribution, and accordingly local detailed balance requires every allowed jump to have an allowed reverse. Multiplying $\dot{\Sigma}$ by k_B restores physical entropy units.

Equation (5.55) is the second law at the level of the master equation. Its structure is that of nonequilibrium thermodynamics: a sum of currents J_{ij} , each multiplied by a conjugate force $\ln(W_{ij}p_j/W_{ji}p_i)$ with which it always shares its sign. Two cases illustrate it.

For a system relaxing in contact with one bath, the entropy production is the rate at which the relative entropy of Section 5.4.3 decreases. Local detailed balance (5.52) gives $W_{ij}/W_{ji} = \pi_i/\pi_j$ for the canonical π , so, with $r_i = p_i/\pi_i$,

$$\dot{\Sigma} = \frac{1}{2} \sum_{i,j} J_{ij} \ln \frac{r_j}{r_i} = - \sum_{i,j} J_{ij} \ln r_i = - \sum_i \dot{p}_i \ln r_i = - \frac{d}{dt} D(p||\pi) = - \frac{1}{k_B T} \frac{d\mathcal{F}_T}{dt},$$

where the second equality uses the same renaming as in Eq. (5.51). The free energy lost by the system, divided by T , is the entropy produced.

In the nonequilibrium steady state of the driven ring, nothing changes in the system, $\dot{h} = 0$, and all the entropy production goes to the environment. Each of the three links contributes $J \ln(k_+/k_-)$, so

$$\dot{\Sigma} = \dot{h}_{\text{env}} = 3J \ln \frac{k_+}{k_-} = (k_+ - k_-) \ln \frac{k_+}{k_-} = J \beta \Delta\mu,$$

which is positive for $k_+ \neq k_-$; the last form uses $(k_+/k_-)^3 = e^{\beta \Delta\mu}$. The current J is the net number of turns per unit time, so the motor converts fuel free energy into heat at the rate $J \Delta\mu$. A steady current requires a steady production of entropy.

Stochastic thermodynamics extends heat, work, and entropy production from these averages to individual trajectories of the jump process [52]. Its statements depend on the chosen mesostates and reservoirs, which must be specified as part of the model.

5.6 The Boltzmann equation

Lecture 16 begins here with kinetic theory and continues through the fluctuation relations.

The master equation describes jumps between a finite set of mesostates with given rates. The first equation of this kind was written by Boltzmann in 1872 for a dilute gas [5], where the jumps are collisions that change molecular velocities. Studying it serves three purposes. It was the first derivation of irreversible behavior from molecular mechanics, it shows exactly where an assumption about correlations enters, and it remains the basis of the kinetic theory of gases, from which transport coefficients such as viscosity and thermal conductivity are computed.

5.6.1 The one-particle distribution and free streaming

The complete statistical description of a gas of N molecules is a density on its $6N$ -dimensional phase space. For a dilute gas a much smaller object suffices for most purposes, the one-particle distribution $f(\mathbf{r}, \mathbf{v}, t)$. It is defined so that $\int f d^3r d^3v$ is the mean number of molecules in the volume element d^3r around $\mathbf{r}$ with velocities in d^3v around $\mathbf{v}$, and it is normalized by

$$\int d^3r d^3v f(\mathbf{r}, \mathbf{v}, t) = N. \quad (5.56)$$

The local number density, mean velocity, and kinetic-energy density are integrals of f over velocities, for instance $n(\mathbf{r}, t) = \int f d^3v$, and so are the fluxes of momentum and energy that determine pressure, viscosity, and heat conduction. In equilibrium f is the number density times the Maxwell distribution of Chapter 4,

$$f_{\text{eq}}(\mathbf{v}) = n \left(\frac{m}{2\pi k_B T} \right)^{3/2} \exp \left(- \frac{m|\mathbf{v}|^2}{2k_B T} \right). \quad (5.57)$$

Our task is to find how f evolves.

Consider first molecules that neither collide nor feel external forces. Each moves with constant velocity, so the molecules found at $(\mathbf{r}, \mathbf{v})$ at time $t + dt$ are those that were at $(\mathbf{r} - \mathbf{v} dt, \mathbf{v})$ at time t : $f(\mathbf{r}, \mathbf{v}, t + dt) = f(\mathbf{r} - \mathbf{v} dt, \mathbf{v}, t)$. Expanding to first order in dt gives $\partial_t f + \mathbf{v} \cdot \nabla_{\mathbf{r}} f = 0$, which is the Liouville equation (5.7) for a single free particle. Collisions change this picture because they change velocities abruptly. In a dilute gas a collision lasts much less than the time between collisions, and its effect on f is to remove a molecule from one velocity and place it at another, at practically the same position. The evolution of f therefore takes the form

$$\frac{\partial f}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} f = C[f], \quad (5.58)$$

where the collision term $C[f]$ is the net rate, per unit volume of the one-particle phase space, at which collisions bring molecules into the velocity element around $\mathbf{v}$ at the point $\mathbf{r}$. Like the master equation, this is a balance equation. The collision term is a gain, from collisions that produce a molecule with velocity $\mathbf{v}$, minus a loss, from collisions that remove one. The velocity elements play the role of the mesostates and the collisions play the role of the jumps. What remains is to count the collisions.

5.6.2 Counting collisions

We consider molecules of mass m that interact through short-range forces, so that collisions are elastic events involving two molecules at a time. We first describe a single collision and then count how many occur in the gas.

Kinematics. Two molecules with incoming velocities $\mathbf{v}_1$ and $\mathbf{v}_2$ leave a collision with outgoing velocities $\mathbf{v}'_1$ and $\mathbf{v}'_2$. Conservation of momentum and kinetic energy takes its simplest form in terms of the center-of-mass velocity and the relative velocity,

$$\mathbf{V} = \frac{\mathbf{v}_1 + \mathbf{v}_2}{2}, \quad \mathbf{g} = \mathbf{v}_1 - \mathbf{v}_2, \quad \text{for which} \quad |\mathbf{v}_1|^2 + |\mathbf{v}_2|^2 = 2|\mathbf{V}|^2 + \frac{1}{2}|\mathbf{g}|^2. \quad (5.59)$$

Momentum conservation keeps $\mathbf{V}$ fixed, and energy conservation then keeps $g = |\mathbf{g}|$ fixed. The collision can only rotate the relative velocity, so

$$\mathbf{v}'_1 = \mathbf{V} + \frac{g}{2} \hat{\mathbf{n}}', \quad \mathbf{v}'_2 = \mathbf{V} - \frac{g}{2} \hat{\mathbf{n}}', \quad (5.60)$$

where $\hat{\mathbf{n}}'$ is the direction of the outgoing relative velocity. Figure 5.3 shows this geometry. Given the incoming velocities, the outgoing ones are fixed by the direction $\hat{\mathbf{n}}'$, which we describe by a solid angle Ω measured from the incoming direction $\hat{\mathbf{g}}$; its polar angle χ is the scattering angle. Which direction occurs in a given collision depends on its impact parameter and on the interaction.

Cross section. The interaction enters through the differential cross section, which is defined by a scattering experiment. A uniform beam of molecules, all with relative velocity $\mathbf{g}$, falls on one target molecule. Let Φ be the flux density of the beam, the number of molecules that cross a unit area perpendicular to $\mathbf{g}$ per unit time. The number $d\dot{N}_{\text{sc}}$ of molecules scattered per unit time with outgoing relative velocity in the solid angle $d\Omega$ is proportional to both Φ and $d\Omega$, and the proportionality factor defines the differential cross section,

$$d\dot{N}_{\text{sc}} = \Phi \frac{d\sigma}{d\Omega} d\Omega. \quad (5.61)$$

It has dimensions of area: $(d\sigma/d\Omega) d\Omega$ is the part of the cross-sectional area of the beam whose molecules are deflected into $d\Omega$. For hard spheres of diameter a , the scattering is isotropic in the center-of-mass frame, $d\sigma/d\Omega = a^2/4$, and the total cross section is the geometric area πa^2 of Section 4.7.

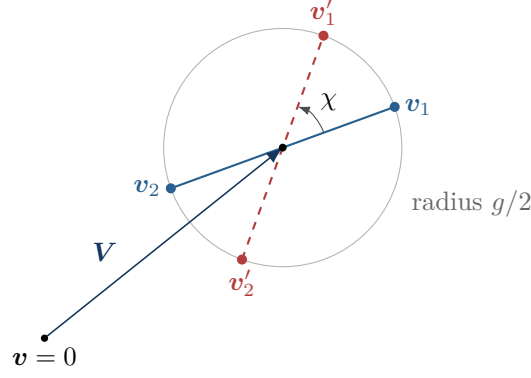


Figure 5.3: An elastic collision of two molecules of equal mass, drawn in velocity space. Momentum conservation fixes the center-of-mass velocity $\mathbf{V}$, the midpoint of the segment that joins the two velocities, and energy conservation fixes the length g of that segment. Incoming velocities (solid) and outgoing velocities (dashed) therefore lie at opposite ends of two diameters of the same sphere of radius $g/2$, and the collision rotates the relative velocity by the scattering angle χ .

The loss term. Now count collisions in the gas. Fix a molecule with velocity $\mathbf{v}_1$ at the point $\mathbf{r}$. The molecules near $\mathbf{r}$ with velocities in d^3v_2 form a beam that approaches it with relative speed $g = |\mathbf{v}_1 - \mathbf{v}_2|$. Their density is $f(\mathbf{r}, \mathbf{v}_2, t) d^3v_2$, so their flux density relative to the first molecule is $\Phi = g f(\mathbf{r}, \mathbf{v}_2, t) d^3v_2$. By Eq. (5.61), the first molecule undergoes collisions with partners from d^3v_2 , scattering into $d\Omega$, at the rate

$$g \frac{d\sigma}{d\Omega} f(\mathbf{r}, \mathbf{v}_2, t) d^3v_2 d\Omega.$$

This is the counting of Section 4.7, where the collision rate was the density times the total cross section times the mean relative speed, now resolved by the velocity of the partner and by the direction of scattering. Each such collision removes the first molecule from its velocity element. Multiplying by the number $f(\mathbf{r}, \mathbf{v}_1, t) d^3r d^3v_1$ of molecules in $d^3r d^3v_1$, and integrating over the velocity of the partner and the direction of scattering, gives the number of molecules that leave $d^3r d^3v_1$ per unit time,

$$d^3r d^3v_1 \int d^3v_2 \int d\Omega g \frac{d\sigma}{d\Omega} f(\mathbf{r}, \mathbf{v}_1, t) f(\mathbf{r}, \mathbf{v}_2, t). \quad (5.62)$$

Molecular chaos. This step contains an assumption. The number of pairs of molecules at the same point with velocities near $\mathbf{v}_1$ and $\mathbf{v}_2$ is given exactly by the two-particle distribution $f^{(2)}(\mathbf{r}, \mathbf{v}_1; \mathbf{r}, \mathbf{v}_2; t)$, and Eq. (5.62) replaced it by the product $f(\mathbf{r}, \mathbf{v}_1, t) f(\mathbf{r}, \mathbf{v}_2, t)$. The replacement is correct only if the velocities of two molecules about to collide are statistically independent. Boltzmann called this the *Stosszahlansatz*, the assumption about the number of collisions, and it is now called the assumption of molecular chaos. For a dilute gas it is plausible. Two molecules about to collide have typically never met, or met so long ago, and after so many other collisions, that their correlation has been spread over many other molecules. Molecules that have just collided, on the other hand, are certainly correlated, since Eq. (5.60) ties their outgoing velocities together. The assumption is made for incoming pairs only, and this asymmetry between before and after is the source of irreversibility in the Boltzmann equation. Without it, the equation for f would involve $f^{(2)}$, whose own equation would involve the three-particle distribution, and so on, in a chain that never closes.

The gain term. Molecules enter the element d^3v_1 through collisions whose outgoing velocities are $\mathbf{v}_1$ and some $\mathbf{v}_2$. Call the incoming velocities of such a collision $\mathbf{v}'_1$ and $\mathbf{v}'_2$. This is the inverse collision $(\mathbf{v}'_1, \mathbf{v}'_2) \rightarrow (\mathbf{v}_1, \mathbf{v}_2)$, whose incoming velocities are related to the outgoing ones by Eq. (5.60) with primed and unprimed velocities exchanged. We count inverse collisions as we counted the loss, again assuming molecular chaos. Incoming pairs with velocities in $d^3v'_1 d^3v'_2$ scatter into $d\Omega$ at the rate $g' (d\sigma'/d\Omega) f(\mathbf{r}, \mathbf{v}'_1, t) f(\mathbf{r}, \mathbf{v}'_2, t) d^3v'_1 d^3v'_2 d\Omega$ per unit volume, where g' and $d\sigma'/d\Omega$ refer to the inverse collision, and this rate must be summed over all incoming velocities whose collisions end in d^3v_1 . Three properties of elastic collisions reduce the result to the form of the loss term.

1. The relative speed is the same before and after the collision, $g' = |\mathbf{v}'_1 - \mathbf{v}'_2| = |\mathbf{v}_1 - \mathbf{v}_2| = g$, by Eq. (5.60).
2. The collision preserves volume in velocity space, $d^3v'_1 d^3v'_2 = d^3v_1 d^3v_2$. The change of variables from $(\mathbf{v}_1, \mathbf{v}_2)$ to $(\mathbf{V}, \mathbf{g})$ has unit Jacobian, and the collision keeps $\mathbf{V}$ while rotating $\mathbf{g}$, which preserves volume. This is Liouville's theorem for a two-body collision.
3. The inverse collision has the same cross section as the direct one. This property, called microscopic reversibility, follows from the invariance of the interaction under time reversal and under spatial inversion, and it holds for central forces.

With these properties the number of molecules that enter $d^3r d^3v_1$ per unit time becomes

$$d^3r d^3v_1 \int d^3v_2 \int d\Omega g \frac{d\sigma}{d\Omega} f(\mathbf{r}, \mathbf{v}'_1, t) f(\mathbf{r}, \mathbf{v}'_2, t), \quad (5.63)$$

where $\mathbf{v}'_1$ and $\mathbf{v}'_2$ are now the incoming velocities of the collision that produces $(\mathbf{v}_1, \mathbf{v}_2)$ with scattering angle Ω .

The collision integral. Dividing the gain minus the loss by $d^3r d^3v_1$ gives the collision term,

$$C[f]_1 = \int d^3v_2 \int d\Omega g \frac{d\sigma}{d\Omega} (f'_1 f'_2 - f_1 f_2), \quad (5.64)$$

with the abbreviations $f_1 = f(\mathbf{r}, \mathbf{v}_1, t)$, $f'_2 = f(\mathbf{r}, \mathbf{v}'_2, t)$, and so on. Every factor now has a meaning. The weight $g (d\sigma/d\Omega) d\Omega d^3v_2$ is the rate at which a given molecule meets partners from d^3v_2 and scatters into $d\Omega$, the product $f_1 f_2$ counts the pairs whose collisions remove a molecule from $\mathbf{v}_1$, and $f'_1 f'_2$ counts the pairs whose collisions bring one in. Together, Eqs. (5.58) and (5.64) form the Boltzmann equation. It is a closed equation for f , and a nonlinear one, because the rate at which a molecule leaves a velocity element depends on how many partners are available, that is, on f itself.

5.6.3 The H -theorem

Boltzmann looked for a functional of f that changes monotonically under his equation and found

$$H[f] = \int d^3r d^3v f \ln f. \quad (5.65)$$

As with any continuous entropy, the logarithm is defined relative to a fixed one-particle phase-space reference; changing that reference adds a constant proportional to the conserved particle number and does not change dH/dt . The meaning of H follows from the case of uncorrelated molecules. If the N -particle density is a product of N identical factors f/N , its fine-grained entropy is

$$-N \int d^3r d^3v \frac{f}{N} \ln \frac{f}{N} = -H[f] + N \ln N.$$

Thus $-k_B H$ is, up to a constant, the entropy of the gas computed as if the molecules were uncorrelated, which is the entropy of the description that molecular chaos retains. In equilibrium it reproduces the thermodynamic entropy. Inserting the Maxwell distribution (5.57) gives

$$-k_B H[f_{\text{eq}}] = N k_B \left[\ln \frac{1}{n} + \frac{3}{2} \ln \frac{2\pi k_B T}{m} + \frac{3}{2} \right],$$

which differs from the Sackur–Tetrode entropy (4.13) by $N k_B [3 \ln(h/m) - 1]$, a term proportional to N that depends on neither T nor V . It records the reference measure of the logarithm, $d^3 r d^3 v$ in place of $d^3 r d^3 p/h^3$, and the factor $N!$ of Chapter 3.

Boltzmann's theorem states that, for a spatially homogeneous gas, H never increases, and that it stays constant only for a Maxwell distribution. Two symmetries of the collision integral do all the work. Take any function $\phi(\mathbf{v})$ and consider the rate at which collisions change its total amount per unit volume,

$$I[\phi] = \int d^3 v_1 \phi_1 C[f]_1 = \int d^3 v_1 d^3 v_2 d\Omega g \frac{d\sigma}{d\Omega} (f'_1 f'_2 - f_1 f_2) \phi_1, \quad (5.66)$$

with $\phi_1 = \phi(\mathbf{v}_1)$, $\phi'_1 = \phi(\mathbf{v}'_1)$, and so on.

The first symmetry is the exchange of the two molecules. Renaming $\mathbf{v}_1 \leftrightarrow \mathbf{v}_2$, which also exchanges $\mathbf{v}'_1 \leftrightarrow \mathbf{v}'_2$, leaves the integration measure, the weight $g d\sigma/d\Omega$, and the factor $f'_1 f'_2 - f_1 f_2$ unchanged, and replaces ϕ_1 by ϕ_2 . Averaging the two forms,

$$I[\phi] = \frac{1}{2} \int d^3 v_1 d^3 v_2 d\Omega g \frac{d\sigma}{d\Omega} (f'_1 f'_2 - f_1 f_2) (\phi_1 + \phi_2).$$

The second symmetry is the exchange of a collision with its inverse. Rename the integration variables $(\mathbf{v}_1, \mathbf{v}_2)$ as $(\mathbf{v}'_1, \mathbf{v}'_2)$, that is, label each collision by its outgoing velocities. By properties 1 to 3 of the gain term, the measure and the weight are unchanged, while $f'_1 f'_2 - f_1 f_2$ changes sign and $\phi_1 + \phi_2$ becomes $\phi'_1 + \phi'_2$. Averaging again,

$$I[\phi] = \frac{1}{4} \int d^3 v_1 d^3 v_2 d\Omega g \frac{d\sigma}{d\Omega} (f'_1 f'_2 - f_1 f_2) (\phi_1 + \phi_2 - \phi'_1 - \phi'_2). \quad (5.67)$$

Two consequences follow. If ϕ is conserved in every collision, $\phi_1 + \phi_2 = \phi'_1 + \phi'_2$, then $I[\phi] = 0$: collisions do not change the total amount of ϕ . The functions 1, $m\mathbf{v}$, and $\frac{1}{2}m|\mathbf{v}|^2$ are such collision invariants, and $I[\phi] = 0$ for them expresses the conservation of particle number, momentum, and kinetic energy.

The second consequence is the H -theorem. For a spatially homogeneous gas f does not depend on $\mathbf{r}$, the streaming term vanishes, and

$$\frac{dH}{dt} = \int d^3 r d^3 v_1 \partial_t f_1 (\ln f_1 + 1) = \int d^3 r (I[\ln f] + I[1]) = \int d^3 r I[\ln f].$$

With $\phi = \ln f$, the last factor in Eq. (5.67) is $\ln(f_1 f_2) - \ln(f'_1 f'_2)$, so

$$\frac{dH}{dt} = -\frac{1}{4} \int d^3 r d^3 v_1 d^3 v_2 d\Omega g \frac{d\sigma}{d\Omega} (f'_1 f'_2 - f_1 f_2) \ln \frac{f'_1 f'_2}{f_1 f_2} \leq 0. \quad (5.68)$$

The inequality is the one met in Section 5.5. With $x = f'_1 f'_2$ and $y = f_1 f_2$, the integrand contains $(x - y) \ln(x/y) \geq 0$, multiplied by the nonnegative weight $g d\sigma/d\Omega$. Thus

$$\frac{dH}{dt} \leq 0, \quad (5.69)$$

and $-k_B H$, the entropy of the uncorrelated description, can only increase. For an inhomogeneous gas in a box with specularly reflecting walls, or with periodic boundaries, the streaming

term adds $-\int d^3r d^3v \nabla_r \cdot (\mathbf{v} f \ln f)$, a surface integral that vanishes, and the theorem holds as well.

Equality in Eq. (5.68) requires $f'_1 f'_2 = f_1 f_2$ for every collision, that is,

$$\ln f_1 + \ln f_2 = \ln f'_1 + \ln f'_2,$$

so $\ln f$ must itself be a collision invariant. A standard result of kinetic theory, which we quote, states that under mild regularity assumptions every collision invariant is a linear combination of 1, $\mathbf{v}$, and $|\mathbf{v}|^2$ [47]. Therefore

$$\ln f = a_0 + \mathbf{b} \cdot \mathbf{v} + c|\mathbf{v}|^2, \quad c < 0,$$

where $c < 0$ is required for normalizability. Completing the square writes this as $f \propto \exp[-m|\mathbf{v} - \mathbf{u}|^2/(2k_B T)]$, a Maxwell distribution moving with mean velocity $\mathbf{u}$, with T , $\mathbf{u}$, and the prefactor fixed by the density, momentum, and energy of the gas. Maxwell had used the condition $f_1 f_2 = f'_1 f'_2$ in 1867 to show that his distribution is unchanged by collisions [37]. Boltzmann's theorem adds that every other distribution evolves toward it.

Status: controlled approximation

The H -theorem is exact for solutions of the Boltzmann equation. The Boltzmann equation itself rests on the dilute-gas limit and on molecular chaos for incoming pairs, so it is an approximate description of finite- N Hamiltonian mechanics.

5.6.4 Loschmidt, Zermelo, and the arrow of time

We can now answer Loschmidt's objection. Reverse every velocity in a gas that has been relaxing for some time. The microscopic motion retraces its path, and H increases. The mechanics is correct, and what fails is molecular chaos in the reversed state. During the forward evolution, every collision created correlations between the outgoing velocities of the two molecules involved. After the reversal these outgoing pairs become incoming pairs, correlated in exactly the way needed for each collision to undo an earlier one. Replacing their two-particle distribution by $f_1 f_2$ discards precisely the correlations that make the reversed evolution possible. The H -theorem holds for gases whose incoming pairs are uncorrelated, a condition that is natural for states prepared by ordinary macroscopic means and that fails for the velocity-reversed state.

The arrow of time therefore enters as it did in Section 5.3. The Boltzmann equation keeps only the one-particle distribution, a coarse-grained description of the N -particle density. Collisions continually create correlations, and the equation assumes that these correlations never return to influence the collisions that follow. This is the kinetic counterpart of repeated coarse graining. The equation discards the correlations that collisions have already produced and assumes that molecules about to collide carry none, and this choice selects the direction of time.

Nevertheless the Boltzmann equation is much more than a plausible model. Lanford proved in 1975 that for hard spheres, in the low-density limit with $N a^2$ held fixed, the motion governed by Newton's equations from an uncorrelated initial state is described by the Boltzmann equation [29]. His proof covered only a short time, a fraction of the mean time between collisions. Deng, Hani, and Ma have extended the derivation to arbitrarily long times, as long as the solution of the Boltzmann equation exists [12]. In this limit molecular chaos is a property that the dynamics propagates from an uncorrelated initial state, and the assumption reduces to a statement about how the gas was prepared.

Zermelo's objection rests on the recurrence theorem that Poincaré proved in 1890 [44]: for a dynamics that preserves volume on a region of finite volume, almost every initial point returns

arbitrarily close to itself, and does so infinitely often. An isolated gas in a box has an energy shell of finite volume, so its microscopic state recurs, and so does every function of that state. No function of the microscopic state can decrease forever. The answer lies in the time scales, which the Ehrenfest model makes explicit. There the macroscopic imbalance relaxes in a time of order $1/(2\nu)$, independent of N , while the initial state recurs after a time $2^N/(N\nu)$, already 3×10^{10} times the age of the universe for one hundred balls jumping once per second. For a macroscopic number of molecules recurrence times are beyond any meaningful comparison, and the thermodynamic description, which concerns far shorter times, remains accurate on them. The Ehrenfests introduced their model precisely to show that relaxation and recurrence are compatible [13]. Recurrence does show that strict monotonicity for all times cannot follow from finite Hamiltonian dynamics alone.

5.7 Nonequilibrium work

The previous sections obtained increasing entropies by discarding information: averaging over cells, assuming Markov dynamics, assuming molecular chaos. In this last section we return to exact Hamiltonian dynamics and ask what form the second law takes there. The answer, found by Jarzynski in 1997 [18], is an exact equality for the fluctuating work done on a system driven out of equilibrium. The familiar inequality follows from it as a statement about averages.

5.7.1 Work on a fluctuating system

Start from thermodynamics. A system in contact with a heat bath at temperature $T > 0$ is initially in equilibrium, and an external agent changes a parameter λ of its Hamiltonian from λ_0 to λ_τ : the position of a piston, say, or the stiffness of an optical trap that holds a colloidal bead. Let W be the work done on the system, positive when the agent supplies energy. This is the opposite of the convention of Chapter 1, where W was the work done by the system. Let Q be the heat absorbed from the bath, and $\Delta F = F(\lambda_\tau) - F(\lambda_0)$ the difference between the equilibrium free energies at the bath temperature. After the protocol, let the system relax to equilibrium at fixed λ_τ , which requires no work. The Clausius inequality (1.18) with $T_b = T$ gives $Q \leq T\Delta S$, and the first law, $\Delta E = Q + W$ in the present convention, then gives

$$W \geq \Delta E - T\Delta S = \Delta F, \quad (5.70)$$

with equality for a reversible process. The work needed to change λ is at least the free-energy difference, and the excess $W - \Delta F$ is dissipated.

For a macroscopic system W hardly varies from one repetition of the process to the next. For a single molecule stretched by optical tweezers it does: the bath kicks the molecule differently in each run, and W fluctuates by amounts of order $k_B T$. Some runs then have $W < \Delta F$. At this scale the second law must be a statement about the distribution of W , and the Jarzynski equality is that statement.

To define the work in a single run, let the Hamiltonian depend on the controlled parameter, $H(x; \lambda)$, where x denotes the microscopic state, and let the agent follow a protocol λ_t , $0 \leq t \leq \tau$. Along a trajectory x_t the energy $H(x_t; \lambda_t)$ changes for two reasons, because the state moves and because the function H itself changes with λ :

$$\frac{d}{dt}H(x_t; \lambda_t) = \nabla_x H \cdot \dot{x}_t + \dot{\lambda}_t \frac{\partial H(x_t; \lambda_t)}{\partial \lambda}.$$

The second term is the energy supplied by the agent, present only while λ changes. The work along the trajectory $\gamma = \{x_t : 0 \leq t \leq \tau\}$ is therefore

$$W[\gamma] = \int_0^\tau \dot{\lambda}_t \frac{\partial H(x_t; \lambda_t)}{\partial \lambda} dt, \quad (5.71)$$

while the first term, the energy exchanged as the state moves at fixed λ , is the heat absorbed from the environment. If the system is isolated while λ changes, Hamilton's equations make the first term vanish, $\nabla_x H \cdot \dot{x} = \sum_i (\partial_{q_i} H \dot{q}_i + \partial_{p_i} H \dot{p}_i) = 0$, and the work is the change of energy,

$$W[\gamma] = H(x_\tau; \lambda_\tau) - H(x_0; \lambda_0). \quad (5.72)$$

From here on, W without indices denotes work, which should be kept apart from the transition rates W_{ij} of Section 5.4.

5.7.2 The Jarzynski equality

Consider the simplest setting, which Jarzynski treated first. The system is prepared in canonical equilibrium at λ_0 by contact with the bath, it is then isolated, and λ is changed at any speed. Different runs start from different microstates x_0 , drawn from $\rho_{\text{eq}}(x_0; \lambda_0) = e^{-\beta H(x_0; \lambda_0)} / Z(\lambda_0)$, and the work (5.72) is a function of x_0 that fluctuates from run to run. Average $e^{-\beta W}$ over the initial distribution:

$$\langle e^{-\beta W} \rangle = \int dx_0 \frac{e^{-\beta H(x_0; \lambda_0)}}{Z(\lambda_0)} e^{-\beta [H(x_\tau; \lambda_\tau) - H(x_0; \lambda_0)]} = \frac{1}{Z(\lambda_0)} \int dx_0 e^{-\beta H(x_\tau; \lambda_\tau)}.$$

The initial Boltzmann factor cancels against part of the work factor, leaving the final Boltzmann factor evaluated at the final point x_τ , which depends on x_0 through the flow. Now change the integration variable from x_0 to x_τ . The flow of a time-dependent Hamiltonian also preserves volume, because the divergence (5.3) vanishes at every instant, whatever the value of λ . Thus $dx_0 = dx_\tau$ and

$$\langle e^{-\beta W} \rangle = \frac{1}{Z(\lambda_0)} \int dx_\tau e^{-\beta H(x_\tau; \lambda_\tau)} = \frac{Z(\lambda_\tau)}{Z(\lambda_0)}.$$

With $F = -k_B T \ln Z$, this is the Jarzynski equality

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}. \quad (5.73)$$

The derivation used canonical initial conditions, Hamiltonian dynamics during the protocol, and Liouville's theorem, and no assumption about the speed of the protocol. At time τ the system is in general far from equilibrium, yet the equality relates its work statistics to the free energies of two equilibrium states, the initial one and the one it would reach at λ_τ . The same result holds when the system stays in contact with the bath during the protocol, if the bath is weakly coupled and included in the Hamiltonian dynamics, or if it is modeled by a Markov process obeying local detailed balance [18, 10]. The equality has been used to obtain equilibrium free-energy differences from nonequilibrium experiments, for instance by unfolding single RNA molecules with optical tweezers at different speeds [33].

The second law follows. The exponential is a convex function, so the average of $e^{-\beta W}$ is at least the exponential of the average (Jensen's inequality, Appendix A):

$$e^{-\beta \langle W \rangle} \leq \langle e^{-\beta W} \rangle = e^{-\beta \Delta F}.$$

Taking the logarithm and dividing by $-\beta < 0$, which reverses the inequality,

$$\langle W \rangle \geq \Delta F, \quad (5.74)$$

which is Eq. (5.70) for the mean work. Individual runs can have $W < \Delta F$, but the equality limits how often. Markov's inequality (2.24) applied to $Y = e^{-\beta W}$ with $a = e^{-\beta(\Delta F - \epsilon)}$ gives

$$\mathbb{P}(W \leq \Delta F - \epsilon) \leq e^{-\beta \epsilon}, \quad (5.75)$$

so a run that beats the bound ΔF by many times $k_B T$ is exponentially rare.

Worked example: a sudden change of oscillator stiffness

A classical harmonic oscillator begins in equilibrium with

$$H_0 = \frac{p^2}{2m} + \frac{k_0 x^2}{2}. \quad (5.76)$$

Change k_0 instantaneously to $k_1 > 0$. Position and momentum do not move during the quench, so

$$W = \frac{k_1 - k_0}{2} x^2. \quad (5.77)$$

The initial position is Gaussian, $p_0(x) = \sqrt{\beta k_0 / (2\pi)} \exp(-\beta k_0 x^2 / 2)$. Therefore

$$\langle e^{-\beta W} \rangle = \sqrt{\frac{\beta k_0}{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\beta k_1 x^2}{2}\right) dx = \sqrt{\frac{k_0}{k_1}}. \quad (5.78)$$

The classical oscillator partition function factorizes into momentum and position integrals:

$$Z(k) = \frac{1}{h} \sqrt{\frac{2\pi m}{\beta}} \sqrt{\frac{2\pi}{\beta k}} = \frac{2\pi}{\beta h} \sqrt{\frac{m}{k}}.$$

Only the $k^{-1/2}$ factor changes during the quench, so

$$\Delta F = \frac{k_B T}{2} \ln \frac{k_1}{k_0}, \quad e^{-\beta \Delta F} = \sqrt{\frac{k_0}{k_1}}. \quad (5.79)$$

Jarzynski's equality is satisfied although the protocol is instantaneous. By contrast, equipartition in the initial state gives $\langle x^2 \rangle = k_B T / k_0$, and hence

$$\langle W \rangle = \frac{k_B T}{2} \left(\frac{k_1}{k_0} - 1 \right) \geq \Delta F, \quad (5.80)$$

where the last inequality is $x - 1 \geq \ln x$ with $x = k_1 / k_0$. Equality holds only for $k_1 = k_0$ in this sudden protocol. Jarzynski's identity uses an exponential average over the whole work distribution; the second-law bound constrains its first moment.

5.7.3 Crooks' relation and the arrow of time

The Jarzynski equality averages over all runs. Crooks found in 1999 a finer statement, which compares each trajectory with its time reverse [10]. For the isolated protocol above it follows from the same ingredients. Besides the forward experiment, consider a reverse experiment, in which the system starts in canonical equilibrium at λ_τ and the agent runs the protocol backward, $\lambda_t^R = \lambda_{\tau-t}$. Let $\mathcal{R}$ reverse all momenta and assume, as in Section 5.1, that $H(\mathcal{R}x; \lambda) = H(x; \lambda)$ for every λ . By time-reversal symmetry, if the forward experiment takes x_0 to x_τ , the reverse experiment started at $\mathcal{R}x_\tau$ retraces the same path and ends at $\mathcal{R}x_0$. We call this path $\tilde{\gamma}$, the time reverse of γ . Its work is

$$W[\tilde{\gamma}] = H(\mathcal{R}x_0; \lambda_0) - H(\mathcal{R}x_\tau; \lambda_\tau) = -W[\gamma].$$

Now compare probabilities. The forward run starts in the volume element dx_0 around x_0 with probability $\rho_{\text{eq}}(x_0; \lambda_0) dx_0$, and since the dynamics is deterministic, this is also the probability $P_F[\gamma]$ of the whole trajectory. The reverse run starts in the image of that element, whose volume is $dx_\tau = dx_0$ by Liouville's theorem, with probability $P_R[\tilde{\gamma}] =$

$\rho_{\text{eq}}(\mathcal{R}x_\tau; \lambda_\tau) dx_\tau = \rho_{\text{eq}}(x_\tau; \lambda_\tau) dx_0$. Their ratio is

$$\frac{P_F[\gamma]}{P_R[\tilde{\gamma}]} = \frac{e^{-\beta H(x_0; \lambda_0)} / Z(\lambda_0)}{e^{-\beta H(x_\tau; \lambda_\tau)} / Z(\lambda_\tau)} = \exp\{\beta[W[\gamma] - \Delta F]\}. \quad (5.81)$$

This is Crooks' fluctuation relation. A trajectory whose work exceeds ΔF is more probable than its time reverse by the factor $e^{\beta W_{\text{diss}}}$, where $W_{\text{diss}} = W - \Delta F$ is the dissipated work. Crooks also proved the relation for stochastic dynamics obeying local detailed balance [10, 52], and Exercise 5.10 derives it for a discrete-time Markov chain by comparing forward and reverse path probabilities step by step.

The Jarzynski equality is a consequence. Rearrange Eq. (5.81) as $P_F[\gamma] e^{-\beta(W[\gamma] - \Delta F)} = P_R[\tilde{\gamma}]$ and sum over all forward trajectories, writing $\int \mathcal{D}\gamma$ for this sum, which in the deterministic case is an integral over x_0 :

$$\langle e^{-\beta(W - \Delta F)} \rangle_F = \int \mathcal{D}\gamma P_F[\gamma] e^{-\beta(W[\gamma] - \Delta F)} = \int \mathcal{D}\gamma P_R[\tilde{\gamma}] = 1,$$

because time reversal maps the set of all trajectories one-to-one onto itself and the reverse probabilities are normalized.

Crooks' relation also gives the dissipated work a statistical meaning. Its logarithm defines the entropy production of a single trajectory,

$$\Sigma[\gamma] = \beta W_{\text{diss}}[\gamma] = \ln \frac{P_F[\gamma]}{P_R[\tilde{\gamma}]}, \quad (5.82)$$

and its average over forward trajectories is a relative entropy,

$$\langle \Sigma \rangle_F = \int \mathcal{D}\gamma P_F[\gamma] \ln \frac{P_F[\gamma]}{P_R^{\text{rev}}[\gamma]} = D(P_F \| P_R^{\text{rev}}) \geq 0, \quad (5.83)$$

where $P_R^{\text{rev}}[\gamma] = P_R[\tilde{\gamma}]$ is the probability that the reverse experiment produces the time reverse of γ . The mean dissipated work is $k_B T$ times the relative entropy between the forward process and the time-reversed backward process, the quantity that, according to Section 2.5, measures how well two distributions can be told apart from data [24]. Imagine being shown a film of one run and asked whether it is played forward or backward. When the mean dissipation is many times $k_B T$, the answer is almost always easy; when the process is nearly reversible, the two possibilities can hardly be distinguished. The inequality $\langle W \rangle \geq \Delta F$ is the positivity of this relative entropy.

Status: exact

For isolated Hamiltonian dynamics with a canonical initial state and a time-reversal-invariant Hamiltonian, Eqs. (5.73) and (5.81) were derived above without approximation, for protocols of any speed. For stochastic dynamics they require local detailed balance, and we quote the general proofs. Estimating the exponential average from finite data may be difficult, because rare runs with low work dominate it [18, 10].

5.8 Summary: which entropy increases?

We can now return to the question of Section 5.1. Table 5.1 collects the entropies met in this chapter. In every row where something increases, the increase comes from an assumption that discards information or selects a class of initial states.

The same question returns in Chapter 7. The von Neumann entropy of an isolated quantum system is constant under unitary evolution, and the entropy that grows is that of a subsystem, which discards its correlations with the environment.

Description	Quantity	Behavior	Source of the behavior
Phase-space density of an isolated system	$h_{\text{fg}}[\rho]$	constant	Liouville's theorem, Eq. (5.10)
Cell probabilities, exact evolution	$h_{\text{fg}}[\bar{\rho}(t)]$	never below its initial value	preparation uniform within the cells, Eq. (5.16)
Cell probabilities, repeated coarse graining	$h_{\text{fg}}[\bar{\rho}]$	increases at every step	arrangement inside the cells discarded at every step, Eq. (5.23)
Master equation	$D(p\ \pi)$	decreases	Markov property, Eq. (5.44)
System and reservoirs	$\dot{\Sigma} = \dot{h} + \dot{h}_{\text{env}}$	nonnegative	local detailed balance, Eq. (5.55)
Dilute gas	$-k_{\text{B}}H[f]$	increases	molecular chaos for incoming pairs, Eq. (5.68)
Driven system, exact dynamics	$\langle W \rangle - \Delta F$	nonnegative	canonical initial state, Eq. (5.73)

Table 5.1: The entropies of this chapter and the ingredient behind the behavior of each.

5.9 Exercises

Exercise 5.1. Derive Liouville's equation and prove conservation of $h_{\text{fg}}[\rho]$. State the boundary assumption used in the integration by parts.

Exercise 5.2. Construct a reversible permutation of four microstates for which a two-cell coarse-grained entropy first increases and later decreases.

Exercise 5.3. For the baker's map of Section 5.3, prepare $\rho_0 = 4$ on the strip $0 \leq x < \frac{1}{4}$ and observe with four vertical cells of width $\frac{1}{4}$. Compute h_{cg} after $n = 0, 1, 2, 3$ steps and verify Eqs. (5.15) and (5.16). Repeat with four horizontal cells of height $\frac{1}{4}$ and explain why the coarse entropy now decreases before it increases.

Exercise 5.4. For a two-state Markov process with rates $W_{10} = a$ and $W_{01} = b$, solve $p_1(t)$. Find the stationary state and verify monotonic decrease of $D(p(t)\|\pi)$.

Exercise 5.5. Prove nonnegativity of (5.55). Show that detailed balance makes every stationary current vanish.

Exercise 5.6. Starting from the collision integral (5.64), list every symmetry and inequality needed for the H -theorem. Identify the step that uses molecular chaos.

Exercise 5.7. Apply Jensen's inequality to (5.73). When does equality hold in (5.74)?

Exercise 5.8. A sudden quench changes a harmonic oscillator stiffness from k_0 to k_1 while the initial position is sampled thermally. Calculate $\langle e^{-\beta W} \rangle$ and verify the Jarzynski equality.

Exercise 5.9 (*). Use Crooks' relation to show that the forward and reverse work distributions cross at $W = \Delta F$.

Exercise 5.10 (*). For a discrete-time Markov chain driven through control values $\lambda_0, \dots, \lambda_n$, assume local detailed balance at each step and equilibrium preparation for both forward and reverse protocols. Compare the probability of a forward trajectory with that of its time reverse and derive Crooks' relation.

5.10 Further reading

Reichl gives a broad treatment of kinetic theory and irreversibility [47], and Kardar derives the Boltzmann equation from the hierarchy of reduced distributions [23]. Lebowitz's short article is a clear account of how Boltzmann's ideas explain the arrow of time [30]. Seifert develops stochastic entropy production and fluctuation relations [52]; the original nonequilibrium work identity is due to Jarzynski [18].

Chapter 6

Information is physical

Lecture 18. Maxwell’s demon and Szilard’s engine, the work bound for states out of equilibrium, Landauer’s erasure bound, measurement and feedback, and the entropy ledger of a complete cycle.

What you should be able to do

- Explain why Szilard’s one-molecule engine seems to violate the second law, and identify the part of the device that this conclusion leaves out.
- Distinguish a logical state from the microstates that realize it, and logical from thermodynamic reversibility.
- Derive the bound $W \geq \Delta\mathcal{F}$ for a system that starts out of equilibrium, and read the dissipated work as discarded information.
- Derive Landauer’s bound and its versions for biased bits and asymmetric memories.
- Bound the work extracted by feedback with mutual information, and close the entropy ledger of a complete Szilard cycle.

Core versus further reading

The lecture follows the main line of the chapter: Szilard’s engine in Section 6.3, the work bound of Section 6.4, Landauer’s bound for a random bit, the feedback bound, and the ledger of Table 6.1. The worked example on measurement errors, the processing of records in Section 6.6.3, and the discussion of asymmetric memories complete the picture and can be read after the lecture.

6.1 Maxwell’s demon

In a letter to Tait in December 1867 [25], and in print four years later in his *Theory of Heat* [36], Maxwell imagined a being able to follow the motion of individual molecules. A gas at uniform temperature fills a vessel divided by a wall into two chambers, A and B , and the being controls a small hole in the wall with a shutter that it can open and close without doing work. It opens the shutter only to let the faster molecules pass from A to B and the slower ones from B to A . As time goes on the gas in B becomes hotter and the gas in A colder, although no work has been spent. Heat has passed from a colder to a hotter body with no other change, which is exactly what the Clausius statement of the second law forbids (Chapter 1). Maxwell’s conclusion was that the second law holds for bodies that we can only handle in bulk, without

perceiving or manipulating their separate molecules. Thomson called the being a demon in 1874 [57], and the name stuck.

The demon turns a question about thermodynamics into a question about information: what does it cost to know where the molecules are, and to act on that knowledge? Part of the answer came from Smoluchowski in 1912 [55]. A demon without intelligence, for instance a light trapdoor held by a spring that opens only when a molecule strikes it from one side, does not work, because the trapdoor is itself in thermal contact with the gas. It fluctuates, opens at random, and lets molecules through in both directions. A device that sorts molecules must find out where they are and act according to what it finds.

Szilard reduced this to its simplest form in 1929 [56]. His engine, which we study in Section 6.3, has one molecule and one binary measurement, and it extracts the work $k_B T \ln 2$ from a single heat bath in every cycle. Szilard concluded that the measurement must produce at least the entropy $k_B \ln 2$, and Brillouin later proposed a mechanism [6]: in a vessel at uniform temperature the demon cannot see the molecules against the thermal radiation, and the light it needs in order to see them raises the entropy by at least k_B for every observation. For about fifty years the measurement was held responsible for saving the second law.

Landauer located the cost elsewhere in 1961, while studying the heat generated by computers [28]. Operations that destroy information, such as resetting a register to zero whatever it contained, must deliver at least $k_B T \ln 2$ of heat per bit to the environment. Bennett showed in 1973 that any computation can be organized in logically reversible steps, which keep the information and have no minimum cost [1], and in 1982 he drew the consequence for the demon [2]. A measurement can be performed reversibly, but the demon's memory fills up with records, and resetting it for the next cycle costs at least what the demon gained. Sagawa and Ueda extended this accounting to measurements with errors, with the mutual information between the system and the record as the measure of what the demon knows [48, 49], and experiments with colloidal particles and single electrons now realize information engines and erasure at the scale of $k_B T$ [59, 4, 26].

This chapter follows that path with the tools of the previous chapters. Section 6.2 describes a memory as a physical system and uses Liouville's theorem to show where the information erased from a memory must go. Section 6.3 sets up Szilard's engine and locates the apparent violation of the second law. Section 6.4 turns the nonequilibrium free energy of Chapter 3 into an exact bound on work, from which Landauer's bound (Section 6.5) and the bound on the work extracted by feedback (Section 6.6) follow in a few lines. Section 6.7 puts the pieces together in the entropy ledger of a complete cycle.

6.2 A memory is a physical system

6.2.1 Logical states and microstates

A bit of information has to be stored in something: the charge of a capacitor, the magnetization of a small domain, the position of a particle. All such memories have one feature in common. A physical degree of freedom has two regions of its state space that can be told apart reliably and that do not exchange their occupants on the time scale of operation. The standard model is a particle in a double-well potential, shown on the left of Fig. 6.1. When the particle is in the left well the memory holds the value 0, and when it is in the right well, the value 1. A barrier of height ΔU separates the wells, and thermally activated jumps over it occur at a rate proportional to $e^{-\Delta U/k_B T}$, so a barrier of a few tens of $k_B T$ keeps the bit for times much longer than any operation on it. A second model, which we will use for calculations because the ideal-gas law makes its thermodynamics explicit, is a single molecule in a box divided by a partition, with the bit given by the side the molecule is on.

The value of the bit is a *logical state*, which we denote by m . Each logical state is realized

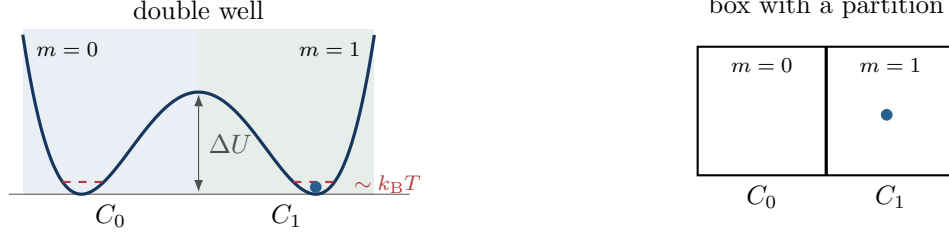


Figure 6.1: Two physical memories holding the value 1. Left: a particle in a double-well potential. The logical states are the two cells C_0 and C_1 , the barrier $\Delta U \gg k_B T$ prevents thermally activated jumps between them, and the particle explores only the bottom of its well, within a few $k_B T$ (dashed). Right: a single molecule in a box divided by a partition, the memory used in the calculations of this chapter.

by a whole region C_m of the memory's phase space, for instance all the positions and momenta of the particle inside one well. The logical value is therefore a coarse-grained variable in the sense of Section 5.3.1, and the regions C_m are its cells. Reliability requires the cells to be distinct and well separated, and each of them may contain an enormous number of microstates.

A memory that has been in contact with a bath for longer than the time needed to equilibrate inside a well, but for much less than the time needed to cross the barrier, is in *local equilibrium*. Inside each cell its phase-space density has the canonical form, while the probabilities p_m of the cells are fixed by the history of the memory, by what was written into it, and the bath has no time to change them. Written out,

$$\rho(x) = p_m \frac{e^{-\beta H(x)}}{Z_m} \quad \text{for } x \in C_m, \quad Z_m = \int_{C_m} dx e^{-\beta H(x)}, \quad (6.1)$$

where x is the microstate of the memory, H its Hamiltonian, and the integral uses the counting measure (3.3). We denote by $\rho_m = e^{-\beta H}/Z_m$ the canonical density restricted to C_m , and by E_m and $F_m = -k_B T \ln Z_m$ its mean energy and its free energy, which we call the *local* free energy of the logical state m .

The nonequilibrium free energy of Chapter 3, $\mathcal{F} = E - k_B T h$, takes a simple form for such a memory. The mean energy is the average of the cell energies, $E = \sum_m p_m E_m$. For the entropy we use the chain rule (2.56): specifying the microstate amounts to specifying first the cell, with probabilities p_m , and then the point inside the cell, with density ρ_m , so that

$$h[\rho] = h(p) + \sum_m p_m h[\rho_m], \quad h(p) = - \sum_m p_m \ln p_m. \quad (6.2)$$

This is the decomposition (5.13) of Chapter 5, for a density that is canonical, rather than uniform, inside each cell. Since each ρ_m is a canonical density, its own free energy is $E_m - k_B T h[\rho_m] = F_m$. Combining the three pieces,

$$\mathcal{F} = E - k_B T h[\rho] = \sum_m p_m F_m - k_B T h(p). \quad (6.3)$$

The free energy of a memory in local equilibrium is the average of the local free energies of its logical states, lowered by $k_B T$ times the Shannon entropy of the logical value. In a *symmetric* memory, whose two cells are physical copies of each other, with equal mean energies $E_0 = E_1$ and equal internal entropies, and therefore equal local free energies $F_0 = F_1$, only the second term depends on what is stored. A bit known to be 0 then carries the free energy $k_B T \ln 2$ above a random bit, which is the worked example on one bit of nonequilibrium free energy in Section 3.5, now seen from inside the memory. A memory left in contact with the bath

for much longer than the time needed to cross the barrier reaches the Gibbs distribution of its Hamiltonian, whose cell probabilities are $p_m^{\text{eq}} = Z_m/Z$ with $Z = \sum_m Z_m$. Inserting $F_m = -k_B T \ln(p_m^{\text{eq}} Z)$ into Eq. (6.3) gives $\mathcal{F} - F_{\text{eq}} = k_B T D(p||p^{\text{eq}})$, which is Eq. (3.44) written for the logical value alone.

6.2.2 Logical and thermodynamic reversibility

A logical operation is a map from input logical states to output logical states. It is *logically reversible* when the output determines the input. Negation, $0 \mapsto 1$ and $1 \mapsto 0$, is reversible. So is copying, which will be our model of a measurement. With x the bit to be copied and m a memory bit, the map

$$(x, m) \mapsto (x, m \oplus x), \quad (6.4)$$

where $\oplus$ denotes addition modulo two, sends $(0, 0) \mapsto (0, 0)$, $(1, 0) \mapsto (1, 1)$, $(0, 1) \mapsto (0, 1)$, and $(1, 1) \mapsto (1, 0)$. It permutes the four joint states and is its own inverse. Starting from $m = 0$ it produces $m = x$, a perfect copy, and leaves the value of x untouched. The reset, or erasure, of a bit to the standard value 0,

$$0 \mapsto 0, \quad 1 \mapsto 0, \quad (6.5)$$

behaves differently. It sends two inputs to the same output, so the output does not reveal the input, and the operation is logically irreversible. Most gates of an ordinary computer are of this kind: the AND gate sends three of its four inputs to 0. Exercise 6.5 embeds it in a reversible map by keeping its inputs alongside the output.

Thermodynamic reversibility is a property of the physical process that implements an operation. A process is thermodynamically reversible when the total entropy of the system and of everything it interacts with does not change, so that it can be run backward without leaving a trace, and it is approached by protocols slow enough to keep every part close to equilibrium. The two notions are independent. A logically reversible gate operated at finite speed dissipates energy like any other driven system. And Section 6.5 shows that a logically irreversible erasure of a random bit can be carried out with zero total entropy production. What logical irreversibility does force is a transfer of entropy out of the logical degree of freedom, and the next subsection uses Liouville's theorem to find where it goes.

6.2.3 Where the erased information goes

Microscopically, the memory and the bath form one Hamiltonian system, on which the agent acts through a time-dependent Hamiltonian, and its flow preserves phase-space volume by Liouville's theorem (Section 5.2), also while the Hamiltonian changes. Logically, erasure merges the cells C_0 and C_1 into C_0 . The two statements are compatible only if the volume that the logical state of the memory loses is gained somewhere else, and Fig. 6.2 shows where.

Consider a symmetric memory whose two cells have the same volume v in the relevant range of energies, and a bath that starts in an energy shell of volume $\Omega_B(E_B)$. Before the erasure the composite can be anywhere in a set of volume $2v \Omega_B(E_B)$, since the memory can be in either cell. A reliable erasure must bring every one of these microstates to a microstate in which the memory is in C_0 . Suppose that the memory has the same energy distribution at the end as at the beginning, so that the work W supplied by the agent ends up in the bath and raises its energy to $E_B + W$. The final microstates then lie in a set of volume $v \Omega_B(E_B + W)$. By Liouville's theorem the image of the initial set has the volume of that set, and it must fit in the final set, so

$$v \Omega_B(E_B + W) \geq 2v \Omega_B(E_B).$$

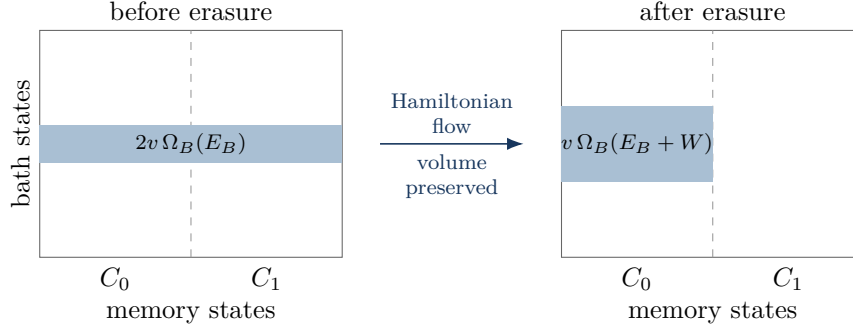


Figure 6.2: Erasure in phase space. Before erasure the memory can be in either cell, each of volume v , and the bath is in an energy shell of volume $\Omega_B(E_B)$. After erasure the memory is in C_0 only. The Hamiltonian flow of memory and bath preserves volume, so the bath must end in a set at least twice as large, which costs the heat $k_B T \ln 2$.

With $S_B = k_B \ln \Omega_B$ and $\partial S_B / \partial E_B = 1/T$, and neglecting the curvature of S_B as in the derivation of local detailed balance, Eq. (5.52), the ratio of the two multiplicities is $e^{W/k_B T}$, and the inequality becomes $e^{W/k_B T} \geq 2$, that is,

$$W \geq k_B T \ln 2. \quad (6.6)$$

The factor two of phase-space volume that the memory loses must be gained by the bath, and at temperature T this requires the heat $k_B T \ln 2$, which raises the bath entropy by $k_B \ln 2$. The same count shows why copying can be free: the map (6.4) sends the four cells of the pair (x, m) one to one onto four cells of equal volume, so no volume has to be handed to the bath.

Nevertheless, the argument assumed cells of equal volume, the same work in every run, and a large bath. When the cells differ, part of the volume can be absorbed inside the memory itself, if the standard cell is the larger one, and Section 6.7.1 shows what this changes. Section 6.5 derives the bound for symmetric memories without the other simplifications, from an exact inequality for the work done on a system that starts out of equilibrium.

6.3 Szilard's engine

Szilard's engine is a Maxwell demon reduced to one molecule and one bit [56]. A single molecule moves in a box of volume V in contact with a bath at temperature $T > 0$, and a controller with a one-bit memory, the demon, runs the cycle of Fig. 6.3 in three steps.

1. *Insertion.* A thin partition is inserted in the middle of the box. The molecule ends up in the left half, $x = 0$, or in the right half, $x = 1$, with probability $1/2$ each.
2. *Measurement.* The demon finds out which half holds the molecule and records the result in its memory, which starts in the standard state $m = 0$. A perfect measurement leaves $m = x$.
3. *Feedback.* Guided by the record, the demon couples the partition to a weight in the direction in which the molecule can push it, and the molecule pushes it quasistatically to the opposite wall, lifting the weight. The partition, now at the wall, is then taken out, and the molecule again moves in the whole box.

The work comes from the third step. For one molecule the ideal-gas equation of state is $P(V') = k_B T / V'$, and the isothermal expansion from $V/2$ to V delivers to the weight

$$W_{\text{ext}} = \int_{V/2}^V \frac{k_B T}{V'} dV' = k_B T \ln 2, \quad (6.7)$$

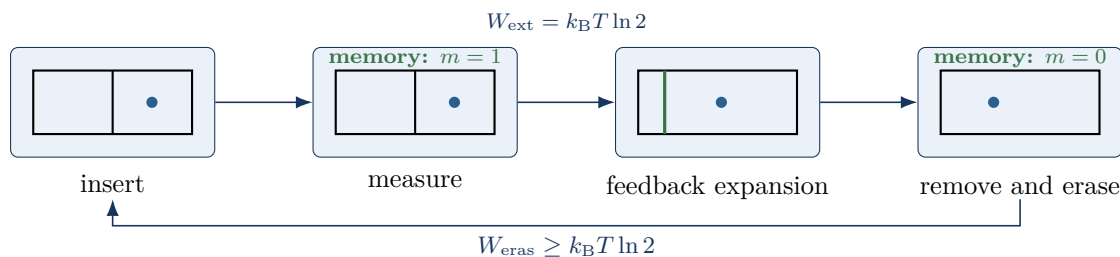


Figure 6.3: One branch of the Szilard cycle, with the molecule found in the right half. The record $m = 1$ selects the side on which the weight is attached. The expansion delivers $k_B T \ln 2$ and leaves the memory holding a bit that is no longer correlated with the molecule; erasing the memory closes the cycle.

whichever side the molecule was on. The energy of an ideal gas at fixed temperature does not change, so the molecule absorbs from the bath, as heat, the energy $k_B T \ln 2$ that it delivers as work. From here on we use the convention of Section 5.7: W is the work done on the system by the external agent, Q is the heat absorbed by the system from the bath, and $\Delta E = W + Q$. The extracted work is $W_{\text{ext}} = -W$. In this notation the expansion has $W = -k_B T \ln 2$ and $Q = k_B T \ln 2$.

The record is what makes the expansion possible. A controller that couples the weight to the same side in every cycle, without looking, finds the molecule on the wrong side half of the time. Moving the partition then compresses the molecule instead of letting it expand, and a quasistatic compression into a vanishing volume costs an unbounded amount of work. The measurement selects which of two protocols to run. An action chosen according to the result of a measurement is what we call *feedback*.

Insertion and removal cost nothing in the ideal model. A point particle has zero probability of being exactly where an infinitely thin partition is placed, and a partition resting against a wall has no gas on its far side. These statements belong to the idealization, and a partition of finite thickness inserted in finite time dissipates work.

Consider now the balance after one cycle. The molecule is again spread over the whole box at temperature T , the partition is back outside, the bath has lost the energy $k_B T \ln 2$ as heat, and the weight has gained it as potential energy. If nothing else had changed, the engine would convert heat from a single bath completely into work in a cycle, which is the process forbidden by the Kelvin–Planck statement of the second law (Chapter 1). In terms of entropy, the bath has lost $k_B \ln 2$, and nothing seems to have gained it.

But something else has changed. The memory began the cycle in its standard state 0, and it ends holding the result of the measurement, 0 or 1 with probability 1/2 each. After the expansion the molecule fills the box, so the record no longer says anything about it, and the memory holds one random bit, with Shannon entropy $\ln 2$. If this Shannon entropy, multiplied by k_B , counts as thermodynamic entropy, the cycle has moved the entropy $k_B \ln 2$ from the bath to the memory, and the total has not decreased. The device has not returned to its initial state either, since its memory has changed. To run another cycle we must either supply a fresh memory in its standard state every time, or reset the old one.

The first option shows what the engine really does. By the worked example of Section 3.5, a bit known to be 0 carries the nonequilibrium free energy $k_B T \ln 2$ above a random bit, and the engine converts that free energy, stored in a tape of blank memory cells, into the work done on the weight. The second option leads to the cost of a reset. Two questions therefore decide whether the second law survives. Does the Shannon entropy of a memory enter the second law in the same way as thermodynamic entropy? And what is the minimum work needed to reset a memory? The next two sections answer both.

Four distinct steps

Partition insertion creates two possible positions. Measurement correlates the memory with the position. Feedback uses that correlation to extract work. Erasure prepares the memory for the next run. Combining these into the single phrase “the demon measures and extracts work” hides the step at which logical irreversibility enters.

6.4 Work and free energy out of equilibrium

Chapter 5 derived the bound $\langle W \rangle \geq \Delta F$ for a system that starts in canonical equilibrium, Eq. (5.74). The systems of this chapter do not start there. A memory prepared in a definite logical state, a gas whose position has just been measured, a bit about to be erased: each is described by a distribution that differs from the Gibbs distribution of its Hamiltonian. For such states Chapter 3 defined the nonequilibrium free energy at the bath temperature T , which we now write with the control parameter λ of the Hamiltonian displayed,

$$\mathcal{F}[p; \lambda] = \langle H(\cdot; \lambda) \rangle_p - k_B T h(p), \quad (6.8)$$

where $\langle \cdot \rangle_p$ is an average over p and $h(p)$ is the Gibbs–Shannon entropy of p , taken relative to the counting measure (3.3) when p is a phase-space density. Chapter 3 also showed that $\mathcal{F} - F_{\text{eq}} = k_B T D(p \| p^{\text{eq}}) \geq 0$, Eq. (3.44), and left open whether this excess can be converted into work. In this section we derive, directly from Hamiltonian dynamics, the bound that settles the question and replaces $\langle W \rangle \geq \Delta F$, following the analysis of Esposito, Lindenberg, and Van den Broeck [15, 14]. Since we follow probability distributions, W will denote the mean work over the initial distribution, the quantity written $\langle W \rangle$ in Chapter 5.

6.4.1 An exact identity

Let the system have microstates x and Hamiltonian $H_S(x; \lambda)$, where an external agent changes λ from λ_0 to λ_τ during $0 \leq t \leq \tau$. It is coupled to a bath with microstates y and Hamiltonian $H_B(y)$, through an interaction whose energy is negligible at the initial and final times; the interaction may also be switched on after $t = 0$ and off before $t = \tau$. The system starts in an arbitrary distribution $p_0(x)$, and the bath starts in canonical equilibrium, independent of the system,

$$\rho_0(x, y) = p_0(x) \pi_B(y), \quad \pi_B(y) = \frac{e^{-\beta H_B(y)}}{Z_B}. \quad (6.9)$$

System and bath evolve together by Hamilton’s equations into a joint density $\rho_\tau(x, y)$, whose marginals are $p_\tau(x)$ for the system and $q_\tau(y)$ for the bath. We need three facts about these densities.

The first is Liouville’s theorem. The joint flow preserves phase-space volume even while λ changes, as we saw in the derivation of the Jarzynski equality, so the fine-grained entropy of the joint density is constant, Eq. (5.10). For the initial product state the entropy is additive, Eq. (2.57), and therefore

$$h[\rho_\tau] = h[\rho_0] = h(p_0) + h(\pi_B). \quad (6.10)$$

Note that this is the only dynamical input to the entropy balance derived below. The other two facts are identities between entropies, valid for any distributions.

The second is the definition (2.66) of mutual information, applied at time τ ,

$$h(p_\tau) + h(q_\tau) = h[\rho_\tau] + I_\tau(S : B), \quad I_\tau(S : B) \geq 0. \quad (6.11)$$

The motion has, in general, correlated the system with the bath, and the sum of the two marginal entropies exceeds the joint entropy by the amount of that correlation.

The third concerns the bath alone. For any bath distribution q , the relative entropy to the canonical distribution is

$$D(q||\pi_B) = \sum_y q(y) \ln \frac{q(y)}{e^{-\beta H_B(y)}/Z_B} = -h(q) + \beta \langle H_B \rangle_q + \ln Z_B,$$

where for a classical bath the sum stands for an integral over its phase space. The relative entropy vanishes for $q = \pi_B$, and subtracting the cases $q = q_\tau$ and $q = \pi_B$ gives

$$h(q_\tau) - h(\pi_B) = \beta \Delta E_B - D(q_\tau||\pi_B), \quad (6.12)$$

where $\Delta E_B = \langle H_B \rangle_{q_\tau} - \langle H_B \rangle_{\pi_B}$ is the energy received by the bath. A bath that remains close to equilibrium changes its entropy by $\beta \Delta E_B$, the heat it receives divided by its temperature, as in thermodynamics. The relative entropy measures how far the process has pushed it away from equilibrium.

Now combine the three facts. Equation (6.11) expresses $h(p_\tau)$ through the joint entropy, which Eq. (6.10) fixes, and Eq. (6.12) supplies the bath entropy:

$$h(p_\tau) - h(p_0) = h(\pi_B) - h(q_\tau) + I_\tau(S : B) = -\beta \Delta E_B + D(q_\tau||\pi_B) + I_\tau(S : B).$$

Rearranged,

$$\Delta h_S + \beta \Delta E_B = I_\tau(S : B) + D(q_\tau||\pi_B) \geq 0, \quad (6.13)$$

where $\Delta h_S = h(p_\tau) - h(p_0)$. The left-hand side is the entropy production of the process in units of k_B : the change of the system's entropy plus the entropy delivered to the bath as heat. The right-hand side says what it is made of, namely the correlation built up between system and bath and the departure of the bath from equilibrium. The two terms combine into a single relative entropy. Expanding the logarithm as in Eq. (2.68),

$$D(\rho_\tau||p_\tau \pi_B) = -h[\rho_\tau] + h(p_\tau) - \langle \ln \pi_B \rangle_{q_\tau} = I_\tau(S : B) + D(q_\tau||\pi_B), \quad (6.14)$$

where the second step uses $h(p_\tau) - h[\rho_\tau] = I_\tau(S : B) - h(q_\tau)$ and $D(q||\pi_B) = -h(q) - \langle \ln \pi_B \rangle_q$. The entropy production is therefore the relative entropy between the actual joint state and the description that keeps the distribution of the system and replaces the bath by an equilibrium reservoir uncorrelated with it. It measures the information that this description discards. This is the lesson of Chapter 5 once more: entropy is produced when information is discarded, and here we can see exactly which information.

To bring in the work, note that the agent supplies energy only through λ , and that system and bath are otherwise isolated. The work done by the agent is then the change of the total energy, Eq. (5.72) applied to system and bath together, and with the interaction energy neglected at the endpoints,

$$W = \Delta E_S + \Delta E_B, \quad \Delta E_S = \langle H_S(\cdot; \lambda_\tau) \rangle_{p_\tau} - \langle H_S(\cdot; \lambda_0) \rangle_{p_0}. \quad (6.15)$$

The heat absorbed by the system is $Q = -\Delta E_B$, and this is the first law $\Delta E_S = W + Q$. Multiply Eq. (6.13) by $k_B T$ and replace ΔE_B by $W - \Delta E_S$:

$$W - (\Delta E_S - k_B T \Delta h_S) = k_B T [I_\tau(S : B) + D(q_\tau||\pi_B)].$$

The quantity in parentheses is the change $\Delta \mathcal{F} = \mathcal{F}[p_\tau; \lambda_\tau] - \mathcal{F}[p_0; \lambda_0]$ of the nonequilibrium free energy (6.8). Therefore

$$W - \Delta \mathcal{F} = k_B T [I_\tau(S : B) + D(q_\tau||\pi_B)] \geq 0, \quad (6.16)$$

and in particular

$$W \geq \Delta \mathcal{F}. \quad (6.17)$$

The work needed to take a system from one state to another in contact with a bath at temperature T is at least the change of its nonequilibrium free energy, whatever the initial and final states. The excess, the dissipated work, is $k_B T$ times the information discarded when the bath is treated as an equilibrium reservoir uncorrelated with the system.

Status: exact

Equations (6.13)–(6.17) are exact for a system and a bath at temperature $T > 0$ that evolve together by Hamilton’s equations, with the bath initially canonical and uncorrelated with the system, and with negligible interaction energy at the initial and final times. No assumption is made about the speed of the protocol, the size of the bath, or the initial and final states of the system. For a quantum system and bath the same derivation holds with von Neumann entropies and unitary evolution [15, 46], in the language of Chapter 7. If the system starts correlated with the bath, or with another system that the protocol can exploit, the bound changes, and Section 6.6 treats the case of a measurement record.

6.4.2 Three readings of the bound

The first reading is the Clausius inequality. With $Q = -\Delta E_B$, Eq. (6.13) reads

$$k_B \Delta h_S \geq \frac{Q}{T}. \quad (6.18)$$

This is the Clausius inequality (1.18) for a single bath, with the Shannon entropy of the system’s distribution, multiplied by k_B , in place of the thermodynamic entropy, and it holds for initial and final states arbitrarily far from equilibrium. It answers the first question of Section 6.3: the Shannon entropy of a memory enters the second law exactly as thermodynamic entropy does. In the Szilard engine the inequality must be applied to the molecule and the memory together, because the feedback protocol depends on the record. Over the measurement and the feedback their joint entropy grows by $k_B \ln 2$, the entropy of the random bit left in the memory, which is exactly the entropy lost by the bath (Section 6.7.2).

The second reading recovers Chapter 5. If the system starts in equilibrium at λ_0 , then $\mathcal{F}[p_0; \lambda_0] = F(\lambda_0)$, while the variational principle of Chapter 3 gives $\mathcal{F}[p_\tau; \lambda_\tau] \geq F(\lambda_\tau)$ for any final state. Equation (6.17) then gives $W \geq \Delta F$, which is Eq. (5.74).

The third reading is the one announced in Chapter 3. Suppose the agent returns λ at the end to its initial value, or to any value with the same equilibrium free energy. The extracted work obeys

$$W_{\text{ext}} = -W \leq \mathcal{F}[p_0] - \mathcal{F}[p_\tau] \leq \mathcal{F}[p_0] - F_{\text{eq}} = k_B T D(p_0 \| p^{\text{eq}}),$$

where the second inequality is again the variational principle. The free energy that a nonequilibrium state carries above equilibrium, $k_B T D(p_0 \| p^{\text{eq}})$, is the largest work that can be extracted from it with the help of a bath at temperature T . A bit known to be 0 in a symmetric memory, for instance, can deliver at most $k_B T \ln 2$, and the Szilard expansion, which starts from a molecule known to be in one half of the box, delivers exactly this amount.

6.4.3 Reaching the bound

Equality in Eq. (6.17) requires $I_\tau(S : B) = 0$ and $q_\tau = \pi_B$. The second condition fixes $\Delta E_B = 0$, so a process that exchanges no heat, such as a sudden change of the Hamiltonian, can reach the bound exactly, while a process that exchanges heat only approaches it, in the limit of slow driving and of a bath large enough for the heat it receives to leave it practically in

equilibrium [46]. A protocol that starts out of equilibrium must first deal with the mismatch between p_0 and the Hamiltonian, and this can be done in two stages. At $t = 0$ the Hamiltonian is changed suddenly to

$$H'(x) = -k_B T \ln p_0(x), \quad (6.19)$$

for which p_0 is the Gibbs distribution, with partition function $Z' = 1$ and free energy $F' = 0$. The change is too fast for the distribution to move or for heat to flow, so its work is the change of mean energy at fixed p_0 . The entropy terms of $\mathcal{F}[p_0; H']$ and $\mathcal{F}[p_0; \lambda_0]$ are equal, and p_0 is the equilibrium state of H' , so

$$W_1 = \langle H' - H_S(\cdot; \lambda_0) \rangle_{p_0} = \mathcal{F}[p_0; H'] - \mathcal{F}[p_0; \lambda_0] = F' - \mathcal{F}[p_0; \lambda_0].$$

The Hamiltonian is then changed quasistatically from H' to $H_S(\cdot; \lambda_\tau)$, keeping the system in equilibrium at every instant, which costs $W_2 = F(\lambda_\tau) - F'$. The total work,

$$W_1 + W_2 = F(\lambda_\tau) - \mathcal{F}[p_0; \lambda_0],$$

is the change of nonequilibrium free energy for a process that ends in equilibrium. The sudden stage reaches the bound exactly, and the quasistatic stage approaches it. Where p_0 vanishes, H' is infinite, which is a wall. The partitions of the Szilard engine, and those of the erasure protocol of Section 6.5, do precisely this job: they are walls placed where the molecule is known not to be, so that its state becomes an equilibrium state that can then be changed quasistatically.

6.5 Landauer's bound

6.5.1 Erasing a random bit

We can now compute the cost of a reset. Take a symmetric memory holding a random bit, with $p_m = 1/2$ for both values of m , uncorrelated with anything else and in local equilibrium with a bath at temperature T . The erasure leaves it in local equilibrium in the cell C_0 , and its Hamiltonian and controls are returned to their initial settings, so the memory ends as it began except for the stored value. By Eq. (6.3),

$$\mathcal{F}_{\text{initial}} = F_0 - k_B T \ln 2, \quad \mathcal{F}_{\text{final}} = F_0,$$

and Eq. (6.17) gives

$$W \geq \Delta \mathcal{F} = k_B T \ln 2. \quad (6.20)$$

The two cells of a symmetric memory have the same mean energy, so the mean energy of the memory is the same before and after, $\Delta E_S = 0$, and the first law gives $W = -Q$. The heat delivered to the bath obeys

$$-Q \geq k_B T \ln 2. \quad (6.21)$$

This is Landauer's bound [28]. The bath entropy grows by at least $k_B \ln 2$, the Shannon entropy removed from the memory, as the counting argument of Section 6.2.3 anticipated. Here it follows for any symmetric memory, without assuming the same work in every run or a large bath.

6.5.2 A reversible erasure

The bound can be reached. Take as memory a molecule in a box with a partition in the middle, the right panel of Fig. 6.1, and run the protocol of Fig. 6.4.

1. Remove the partition. Since the bit is random, the molecule was already spread uniformly over the box, so removing the partition changes neither its distribution nor its free energy, and costs no work.
2. Push a piston quasistatically from the right wall to the middle. The molecule is compressed from V to $V/2$ at temperature T , which costs $W = k_B T \ln 2$ and delivers the same energy to the bath as heat.
3. Insert the partition in the middle, next to the piston, and withdraw the piston, which now has empty space behind it. The memory holds $m = 0$.

The total work is $k_B T \ln 2$, and the bound (6.20) is reached in the quasistatic limit. Every step is either quasistatic or leaves the distribution unchanged, so the total entropy production vanishes: this logically irreversible erasure is thermodynamically reversible. The entropy $k_B \ln 2$ of the random bit has been moved into the bath, which is what logical irreversibility forces in a symmetric memory. The protocol is the feedback expansion of the Szilard engine, in the branch with the molecule on the left, run backward. The expansion turns the knowledge of one bit into the work $k_B T \ln 2$, and the compression turns the work $k_B T \ln 2$ into one bit of knowledge.

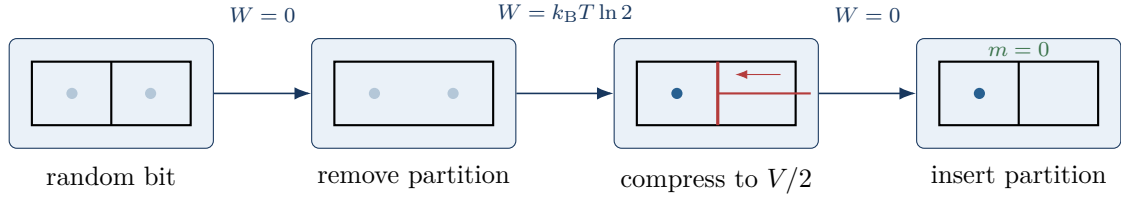


Figure 6.4: Reversible erasure of a random bit stored in a one-molecule box. The faint dots mark the two equally likely positions. Removing the partition does not change the distribution of the molecule, which was already spread over the whole box. The quasistatic compression costs $k_B T \ln 2$, delivered to the bath as heat, and reinserting the partition leaves the memory in $m = 0$. The protocol is the mirror image of the feedback expansion of Fig. 6.3, run backward.

The same protocol applied to a bit known to be 1 is wasteful. Its first step is then the free expansion of a molecule from the right half into the whole box, an irreversible process that produces the entropy $k_B \ln 2$ and destroys the free energy $k_B T \ln 2$ that the known bit carried, and the compression must then pay $k_B T \ln 2$ again. A known bit is better reset by a logically reversible operation chosen with that knowledge: if the bit is 1, translate the occupied half of the box to the left at constant volume, which costs no work. More generally, suppose the agent holds a record y correlated with the bit, for instance a copy made earlier. It can choose the protocol according to y . In a symmetric memory, Eq. (6.3) applied to the conditional distribution $p(m|y)$, together with Eq. (6.17), bounds the work from below by $k_B T h(p_{M|Y=y})$ for each y . Averaging over y with Eq. (2.54),

$$W \geq k_B T h(M|Y). \quad (6.22)$$

The cost of erasure is set by what the agent performing it does not know about the bit, and a perfect copy reduces it to zero.

Bérut and collaborators realized Landauer's erasure with a colloidal bead in a double-well potential created by optical tweezers [4]. The protocol lowered the barrier, pushed the bead toward one well, and raised the barrier again. The mean heat released approached $k_B T \ln 2$ as the protocol was made slower, with an excess that decreased with its duration, and later experiments in feedback traps reached higher precision [20].

The Landauer energy at room temperature

At $T = 300\text{ K}$,

$$k_{\text{B}}T \ln 2 \simeq 2.87 \times 10^{-21} \text{ J} \simeq 1.79 \times 10^{-2} \text{ eV}. \quad (6.23)$$

This is a lower bound for an ideal quasistatic reset, not a forecast for the power used by a present-day computer. Real devices operate in finite time, use error margins, and dissipate energy in control circuitry.

6.5.3 Biased bits and asymmetric memories

If the bit equals 1 with probability p , the Shannon entropy of the logical value is the binary entropy (2.35),

$$h_2(p) = -p \ln p - (1-p) \ln(1-p). \quad (6.24)$$

In a symmetric memory, Eq. (6.3) gives $\mathcal{F}_{\text{initial}} = F_0 - k_{\text{B}}T h_2(p)$ and $\mathcal{F}_{\text{final}} = F_0$, so that

$$W \geq k_{\text{B}}T h_2(p). \quad (6.25)$$

A nearly known bit costs little to reset, because little information is removed. Exercise 6.7 constructs a box protocol that reaches this bound: before the partition is removed, it is moved quasistatically until the volumes of the two sides are proportional to their probabilities. The symmetric protocol of Fig. 6.4 costs $k_{\text{B}}T \ln 2$ for every p , and for a biased bit it wastes $k_{\text{B}}T[\ln 2 - h_2(p)]$ in the free expansion of its first step.

In an asymmetric memory the logical states have different local free energies, $F_m = E_m - TS_m^{\text{int}}$, where E_m is the mean energy and $S_m^{\text{int}} = k_{\text{B}}h[\rho_m]$ the internal entropy of the cell C_m . A memory whose logical state m has probability p_m and is reset into the standard state 0 has $\mathcal{F}_{\text{final}} = F_0$, and Eq. (6.3) gives

$$W \geq F_0 - \sum_m p_m F_m + k_{\text{B}}T h(p). \quad (6.26)$$

The first two terms are the cost of moving probability between logical states with different local free energies, and the last is the information term. When the local free energies differ, $k_{\text{B}}T \ln 2$ is no longer the cost of erasing a random bit. The bound can be smaller or larger, and it can even be negative, when the excess energy of an occupied excited state pays for the reset (Exercise 6.3). The heat follows from Eq. (6.13), with the entropy of the memory split as in Section 6.2 into the logical part and the internal entropies of the cells. The heat released obeys

$$-Q \geq k_{\text{B}}T h(p) - T \left[S_0^{\text{int}} - \sum_m p_m S_m^{\text{int}} \right]. \quad (6.27)$$

In a symmetric memory the internal entropies are equal, and this is Eq. (6.21) extended to a biased bit. If the standard cell has more internal entropy than the others, part of the logical entropy is absorbed inside the memory, and the heat released can be smaller than $k_{\text{B}}T h(p)$; Exercise 6.8 gives a memory in which it vanishes. Equal local free energies are therefore not enough for the heat form (6.21), and Exercise 6.2 builds a memory with $F_0 = F_1$ whose random bit can be erased without releasing any heat. What remains universal in an asymmetric memory is the combined cost of writing a record and erasing it, which Section 6.7.1 derives.

Status: exact

Landauer's bound, in each of the forms (6.20)–(6.27), is exact under the assumptions of Section 6.4.1, with a bath at $T > 0$ that starts in equilibrium and uncorrelated with the memory, and with the memory Hamiltonian and controls restored after the reset. For a

symmetric memory it also bounds the heat released when information is erased without being used, Eq. (6.21). It does not say that every logical operation dissipates $k_B T \ln 2$, since logically reversible operations have no minimum cost, nor that measurement itself must cost this amount.

6.6 Measurement and feedback

Szilard's demon uses its record to choose a protocol. To bound the work that such feedback can extract we need two things: a measure of how much the record says about the system, and a version of Eq. (6.17) for protocols chosen after a measurement.

6.6.1 Measurement creates correlation

Let X be the measured variable of the system, with distribution $p(x)$, and M the record. The measuring device produces the record m with probability $p(m|x)$ when the system is in x , which is a stochastic map, Eq. (2.70), acting on x , and the joint distribution is $p(x, m) = p(x)p(m|x)$. The information that the record carries about the system is the mutual information (2.66), which we write in the form (2.69),

$$I(X : M) = h(M) - h(M|X). \quad (6.28)$$

The first term is the uncertainty of the record before x is known. The second is the uncertainty that remains once x is specified, which is the noise of the measuring device.

For a perfect measurement of a fair bit, $m = x$, so $h(M) = \ln 2$ and $h(M|X) = 0$, and $I(X : M) = \ln 2$. Now suppose that a binary symmetric measurement reports the wrong value with probability ϵ . A fair input still produces a fair record, so $h(M) = \ln 2$. Given either value of x , the record is right with probability $1 - \epsilon$ and wrong with probability ϵ , so $h(M|X) = h_2(\epsilon)$, and Eq. (6.28) gives

$$I(X : M) = \ln 2 - h_2(\epsilon). \quad (6.29)$$

The information decreases from $\ln 2$ at $\epsilon = 0$ to zero at $\epsilon = 1/2$, where the record is independent of the system. A record that is always wrong, $\epsilon = 1$, is as useful as a perfect one, once its labels are exchanged.

What does the controller know after reading m ? Bayes' theorem, Eq. (2.6), replaces the prior $p(x)$ by the posterior $p(x|m) = p(m|x)p(x)/p(m)$. The average relative entropy between posterior and prior is the mutual information,

$$\sum_m p(m) D(p_{X|M=m} \| p_X) = \sum_{x,m} p(x, m) \ln \frac{p(x|m)}{p(x)} = I(X : M), \quad (6.30)$$

since $p(x|m)/p(x) = p(x, m)/[p(x)p(m)]$ turns the middle expression into the relative entropy (2.66). Here x may be the full microstate. When the device responds only to a coarse variable, such as the side of the box that holds the molecule, the record depends on the microstate only through that variable, and the mutual information with the microstate equals the mutual information with the coarse variable. If the system was in equilibrium before the measurement, its distribution $p(x)$ over microstates is the Gibbs distribution, and by Eq. (3.44) each posterior has the nonequilibrium free energy $F + k_B T D(p_{X|M=m} \| p_X)$. For an agent who knows the outcome, the measurement has raised the free energy of the system, on average, by $k_B T I(X : M)$.

6.6.2 The feedback bound

Let the system start in equilibrium at λ_0 , uncorrelated with the bath as in Eq. (6.9), so that $p(x) = p^{\text{eq}}(x; \lambda_0)$. An instantaneous measurement, which does not disturb the system, produces the record m with probability $p(m|x)$, and the agent then runs a protocol $\lambda_t^{(m)}$ that depends on m and ends at a common final value λ_τ . In the runs with outcome m the system starts the protocol in the posterior $p(x|m)$, still uncorrelated with the bath, so Eq. (6.17) applies to them,

$$W(m) \geq \mathcal{F}[p_\tau(\cdot|m); \lambda_\tau] - \mathcal{F}[p(\cdot|m); \lambda_0],$$

where $W(m)$ is the mean work in these runs and $p_\tau(\cdot|m)$ their final distribution. The variational principle bounds the first term from below by $F(\lambda_\tau)$, and Eq. (3.44) gives the second as $F(\lambda_0) + k_B T D(p_{X|M=m} \| p_X)$. Thus $W(m) \geq \Delta F - k_B T D(p_{X|M=m} \| p_X)$. Averaging over the outcomes with $p(m)$ and using Eq. (6.30),

$$\langle W \rangle \geq \Delta F - k_B T I(X : M), \quad \text{or} \quad \langle W_{\text{ext}} \rangle \leq -\Delta F + k_B T I(X : M), \quad (6.31)$$

where the average includes the measurement outcomes. This is the bound of Sagawa and Ueda [48], who proved it also for quantum measurements. A measurement lowers the minimum work of an isothermal process by at most $k_B T$ per nat of information that the record carries about the system. Without a record, $I = 0$, and the inequality reduces to $\langle W \rangle \geq \Delta F$. Note that the bound counts only the work done on the system during the feedback. The cost of producing the record, and of erasing it afterward, falls on the memory, and Section 6.7 adds it to the account.

Equality requires two things. Each conditional process must be reversible in the sense of Eq. (6.16), and it must end in the equilibrium state at λ_τ , $p_\tau(\cdot|m) = p^{\text{eq}}(\cdot; \lambda_\tau)$ for every m . The final state must keep no trace of the outcome, since a system that ends correlated with the record has left part of the information unused. The Szilard engine satisfies both conditions. Its perfect measurement has $I = \ln 2$, the box is the same at the beginning and at the end so that $\Delta F = 0$, and the bound $\langle W_{\text{ext}} \rangle \leq k_B T \ln 2$ is reached by Eq. (6.7), with the molecule finally spread over the box independently of the record. Szilard's engine is an optimal feedback engine.

Status: exact

The feedback bound (6.31) is exact for a system initially in equilibrium, a measurement that does not disturb it, and conditional protocols that satisfy the assumptions of Section 6.4.1 [48]. Sagawa and Ueda later found an equality for the fluctuating work under feedback, from which Eq. (6.31) follows in the same way as Eq. (5.74) follows from the Jarzynski equality [50], and Toyabe and collaborators tested it with a colloidal particle driven by feedback [59].

Worked example: a Szilard engine with measurement errors

Let the measurement of the side of the molecule be wrong with probability $\epsilon < 1/2$. After reading the record, the controller moves the partition quasistatically until the side named by the record occupies the fraction ξ of the box, with $1/2 < \xi < 1$, and then removes the partition. When the record is right, which happens with probability $1 - \epsilon$, the molecule's chamber expands from $V/2$ to ξV and delivers $k_B T \ln(2\xi)$. When it is wrong, the molecule is on the other side, its chamber shrinks from $V/2$ to $(1 - \xi)V$, and the delivered work is $k_B T \ln[2(1 - \xi)] < 0$. The mean extracted work is

$$\frac{\langle W_{\text{ext}} \rangle}{k_B T} = (1 - \epsilon) \ln(2\xi) + \epsilon \ln[2(1 - \xi)] = \ln 2 + (1 - \epsilon) \ln \xi + \epsilon \ln(1 - \xi), \quad (6.32)$$

plotted in Fig. 6.5. Setting its derivative with respect to ξ to zero, $(1 - \epsilon)/\xi = \epsilon/(1 - \xi)$, gives $\xi = 1 - \epsilon$, and the maximum is

$$\begin{aligned}\langle W_{\text{ext}} \rangle_{\text{max}} &= k_{\text{B}}T [\ln 2 + (1 - \epsilon) \ln(1 - \epsilon) + \epsilon \ln \epsilon] \\ &= k_{\text{B}}T [\ln 2 - h_2(\epsilon)] = k_{\text{B}}TI(X : M),\end{aligned}\quad (6.33)$$

by Eq. (6.29). The feedback bound (6.31), with $\Delta F = 0$, is reached. Note that the optimal share of the box, $\xi = 1 - \epsilon$, is the posterior probability that the molecule is on the side named by the record: the controller gives each side a volume proportional to the probability that the molecule is there. The reason why this is optimal is visible in the final state. At $\xi = 1 - \epsilon$ the conditional density of the molecule is $(1 - \epsilon)/[(1 - \epsilon)V] = 1/V$ on the side named by the record and $\epsilon/(\epsilon V) = 1/V$ on the other side. It is uniform, so removing the partition changes nothing, and every conditional process ends in equilibrium, as the equality condition demands. A controller that trusts the record completely, $\xi \rightarrow 1$, compresses the molecule into a vanishing volume in the runs with a wrong record, and its mean work diverges to $-\infty$ for every $\epsilon > 0$. Sagawa and Ueda worked out this engine in their analysis of feedback control [51], and Horowitz and Parrondo related the optimality of feedback to its thermodynamic reversibility [17].

For $\epsilon = 0.10$,

$$h_2(0.10) = -0.10 \ln 0.10 - 0.90 \ln 0.90 \simeq 0.325,$$

so $I \simeq 0.368$ nats, or $I/\ln 2 \simeq 0.531$ bits, and the partition stops at $\xi = 0.9$. At 300 K the largest mean work per cycle is

$$k_{\text{B}}TI(X : M) \simeq 1.52 \times 10^{-21} \text{ J}, \quad (6.34)$$

a little more than half of the Szilard value. The missing part is the uncertainty about the molecule that remains after the record has been read. Exercise 6.4 extends the example to a partition inserted off center, where the work extracted after each record turns out to be $k_{\text{B}}T$ times the relative entropy between posterior and prior, the quantity averaged in Eq. (6.30).

6.6.3 Processing the record

Suppose the raw record M is copied, compressed, or transmitted through a noisy channel into a new record M' before the controller uses it. The new record depends on the system only through M , so $X \rightarrow M \rightarrow M'$ is a Markov chain, and the data-processing inequality (2.76) gives

$$I(X : M') \leq I(X : M). \quad (6.35)$$

The feedback bound computed with M' is therefore never larger than the one computed with M . Processing a record cannot create extractable work, and uncontrolled noise acting on the record can only waste it.

Data processing for relative entropy, Eq. (2.74), also protects the free energy stored in a nonequilibrium state. Let a stochastic map $\mathcal{T}$ act on the system and leave its equilibrium distribution invariant, $\mathcal{T}p^{\text{eq}} = p^{\text{eq}}$, as contact with the bath at fixed Hamiltonian does. Then the inequality, with $q = p^{\text{eq}}$, gives

$$\mathcal{F}[\mathcal{T}p] - F_{\text{eq}} = k_{\text{B}}TD(\mathcal{T}p||p^{\text{eq}}) \leq k_{\text{B}}TD(p||p^{\text{eq}}) = \mathcal{F}[p] - F_{\text{eq}}. \quad (6.36)$$

Such processes can only degrade the free energy that a state carries above equilibrium. A memory left in contact with the bath behind a finite barrier slowly loses its bit, together with

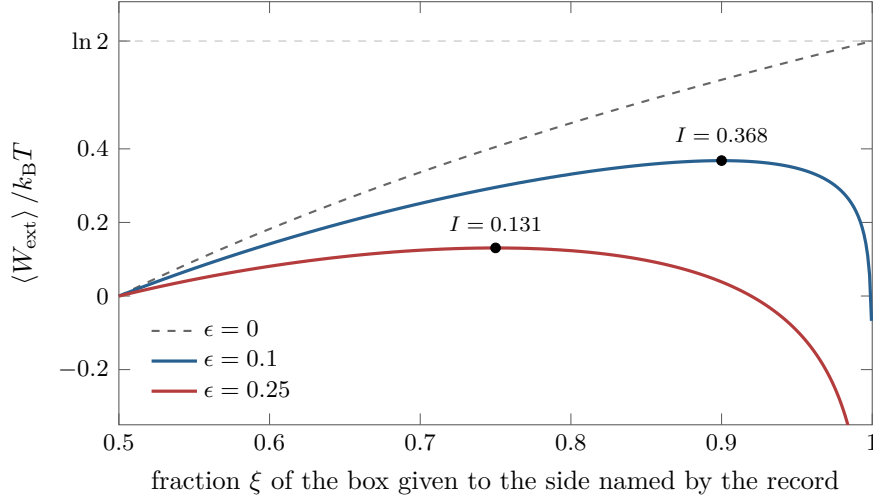


Figure 6.5: Mean work extracted by the Szilard engine with measurement errors, when the partition is moved until the side named by the record occupies the fraction ξ of the box. For an error probability ϵ the maximum lies at $\xi = 1 - \epsilon$ and equals the information bound $k_B T I(X : M) = k_B T [\ln 2 - h_2(\epsilon)]$, marked by the dots. Trusting the record completely, $\xi \rightarrow 1$, makes the mean work diverge to $-\infty$ for every $\epsilon > 0$.

the free energy $k_B T \ln 2$ that the bit carried, and the master equations of Section 5.4 describe this loss, Eq. (5.45).

6.7 Closing the cycle

The feedback bound counts the work gained from a record, but it says nothing about the cost of producing the record and of disposing of it afterward. A demon that runs in a cycle must do both. In this section we bound these costs with the same inequality, add them to the feedback bound, and close the entropy ledger of the Szilard engine.

6.7.1 The cost of measurement and erasure

Let the memory have logical states m with local free energies F_m , as in Section 6.2, and the standard state 0. The measurement acts on the system X and the memory M together. It starts from $p(x)$ with a blank memory and ends in $p(x, m) = p(x)p(m|x)$, without changing the distribution of the system. We assume that system and memory do not interact at the beginning and at the end of the measurement, and that the memory is in local equilibrium inside each cell whatever the state of the system. For the joint free energy we then apply the chain rule as in Eq. (6.3), specifying first the microstate of the system, then the logical state of the memory given that microstate, and finally the microstate of the memory inside its cell. This gives

$$\mathcal{F}_{XM} = \mathcal{F}_X + \sum_m p(m) F_m - k_B T h(M|X),$$

where $\mathcal{F}_X$ is the free energy of the system alone. Before the measurement $h(M|X) = 0$ and the memory contributes F_0 , so Eq. (6.17) applied to system and memory together gives

$$W_{\text{meas}} \geq \sum_m p(m) F_m - F_0 - k_B T h(M|X). \quad (6.37)$$

Erasing the record afterward acts on the memory and the bath only, whatever correlation with the system the feedback has left, and it is the operation of Section 6.5, Eq. (6.26), with the

probabilities $p(m)$ of the records,

$$W_{\text{eras}} \geq F_0 - \sum_m p(m) F_m + k_B T h(M). \quad (6.38)$$

The local free energies enter the two bounds with opposite signs, and they cancel in the sum,

$$W_{\text{meas}} + W_{\text{eras}} \geq k_B T [h(M) - h(M|X)] = k_B T I(X : M). \quad (6.39)$$

Producing a record and erasing it costs together at least $k_B T$ times the information that the record carries. This is the result of Sagawa and Ueda [49].

How the cost is divided between the two steps depends on the memory. In a symmetric memory with an error-free measurement, $h(M|X) = 0$ and the local free energies are equal, so Eq. (6.37) gives $W_{\text{meas}} \geq 0$. The reversible copy (6.4) reaches this bound, and the whole cost $k_B T \ln 2$ falls on the erasure, which is the picture of Landauer and Bennett. In an asymmetric memory part of the cost can be moved to the measurement. Exercise 6.8 works out the box memory with which Sagawa and Ueda made this point, in which all of the cost falls on the measurement and the erasure is free, which is the picture of Szilard and Brillouin. The question of which step pays has no universal answer; Eq. (6.39) shows that the second law fixes only the total.

Now add the three steps of a cycle. The feedback gains at most what Eq. (6.31) allows, and the memory costs at least what Eq. (6.39) demands, so the total work done on the engine and its demon obeys

$$W_{\text{cycle}} = W_{\text{meas}} + \langle W \rangle_{\text{fb}} + W_{\text{eras}} \geq k_B T I(X : M) + [\Delta F - k_B T I(X : M)] = \Delta F. \quad (6.40)$$

For a cyclic engine $\Delta F = 0$, and no cycle of measurement, feedback, and erasure extracts net work from a single bath. The Kelvin–Planck statement holds for the engine together with its demon. Mutual information is the quantity that passes from one step to the next: the measurement creates it, the feedback spends it, and creating and erasing the record costs at least as much as the feedback can gain.

6.7.2 The entropy ledger of the Szilard cycle

To follow the entropy through the cycle step by step, apply Eq. (6.13) to the system and the memory together. Multiplied by k_B , it states that the total entropy change, made of $k_B \Delta h_{XM}$ and the bath entropy change $\Delta S_{\text{bath}} = \Delta E_B/T$, is never negative. The joint entropy of a correlated pair is the sum of the marginal entropies minus the mutual information, Eq. (2.66),

$$S_{XM} = k_B h(X, M) = S_X + S_M - k_B I(X : M), \quad (6.41)$$

with $S_X = k_B h(X)$ and $S_M = k_B h(M)$, so the entropy balance of each step reads

$$\Delta S_{\text{tot}} = \Delta S_X + \Delta S_M - k_B \Delta I(X : M) + \Delta S_{\text{bath}} + \Delta S_{\text{other}} \geq 0, \quad (6.42)$$

where ΔS_{other} collects any other reservoir or device that exchanges entropy with the engine. Each additional reservoir, initially in equilibrium at its own temperature, enters the derivation of Eq. (6.13) in the same way as the bath and contributes its own term $\Delta E/T$. If X and M are uncorrelated at both endpoints of a step, $\Delta I = 0$ and the familiar sum of subsystem entropy changes is recovered. During measurement and feedback the correlation term must be kept.

Table 6.1 lists the entries for the ideal Szilard cycle, and it is worth reading column by column. The molecule's own entropy never changes: its marginal distribution is uniform over the box before the insertion, after the measurement, and after the feedback. The measurement

Step	W	ΔS_X	ΔS_M	$-k_B \Delta I$	ΔS_{bath}	ΔS_{tot}
Insertion	0	0	0	0	0	0
Measurement	0	0	$+k_B \ln 2$	$-k_B \ln 2$	0	0
Feedback	$-k_B T \ln 2$	0	0	$+k_B \ln 2$	$-k_B \ln 2$	0
Erasure	$+k_B T \ln 2$	0	$-k_B \ln 2$	0	$+k_B \ln 2$	0
Cycle	0	0	0	0	0	0

Table 6.1: Entropy ledger of the ideal Szilard cycle, with a perfect measurement, a symmetric memory, and quasistatic feedback and erasure. X is the molecule, M the memory, W the work done on the engine, and each entry is the change during the step. The columns are the terms of Eq. (6.42).

raises the memory's entropy by $k_B \ln 2$, since the record is a random bit, and creates the correlation $I = \ln 2$ between record and molecule, so that the joint entropy is unchanged and the reversible copy produces no entropy. The feedback consumes the correlation and draws $k_B \ln 2$ of entropy from the bath, together with the heat $k_B T \ln 2$ that becomes work. The erasure returns this entropy to the bath and the memory to its standard state. Every row adds up to zero, and the ideal cycle is reversible as a whole. Its net work vanishes, since the work delivered by the feedback pays exactly for the erasure. In a real engine some rows have positive entropy production, and the cycle costs net work, as Exercise 6.6 shows for a noisy measurement. A gas that is followed together with the record gains the entropy $k_B \ln 2$ during the expansion. That is the entropy of the molecule given the record, $S_X - k_B I$, which grows by $k_B \ln 2$ as the correlation is consumed; the table splits the same change into $\Delta S_X = 0$ and $-k_B \Delta I = +k_B \ln 2$, and Exercise 6.1 compares the two descriptions for a box divided into N chambers. This is Bennett's resolution of Szilard's paradox [2]: the feedback converts heat from the bath into work by consuming the correlation between record and molecule, which leaves the memory holding a random bit, and resetting that bit returns the entropy to the bath.

6.7.3 Maxwell's demon revisited

The sorting demon of Section 6.1 fits the same accounting. Each time a molecule approaches the hole, the demon measures whether it is fast or slow, stores the result, and opens the shutter or keeps it closed accordingly, which is feedback with a protocol that costs no work. The gas is isolated while it is sorted, and for each record the motion of the shutter is a Hamiltonian process, which preserves the entropy of the gas conditioned on the record. After a decision the entropy of the gas is therefore at least this conditional entropy, which is the entropy before the decision minus I , by Eq. (2.69), where I is the information that the record carries about the gas. Each decision lowers the entropy of the gas by at most $k_B I$ and leaves its energy unchanged, so it raises the nonequilibrium free energy of the gas at its initial temperature T by at most $k_B T I$. Producing and erasing the record costs at least $k_B T I$, by Eq. (6.39), if the memory is reset in contact with a bath at the same temperature. Whatever work a heat engine later extracts from the temperature difference has therefore been paid for, at least in full, by the time the demon's memory is reset. Experiments now realize such demons at the scale of $k_B T$. Toyabe and collaborators made a Brownian particle climb a potential staircase by switching the potential according to measurements of its position, converting information into free energy [59]. Koski and collaborators built a Szilard engine in which the role of the molecule is played by a single electron that can occupy either of two small metallic islands, and extracted about three quarters of $k_B T \ln 2$ per cycle [26].

Checking an information engine

When analyzing an information engine, first choose the thermodynamic boundary and state the level of description. A microscopic closed Hamiltonian or unitary model preserves fine-grained entropy. The nondecreasing quantity in the second law is the thermodynamic or appropriately coarse-grained entropy of the system, memory, reservoirs, and any omitted work or control devices. In this chapter it is the left-hand side of Eq. (6.13), the Shannon entropy of the system plus the entropy delivered to the reservoirs as heat. Then ask:

- Did the device return to its original physical state?
- Where are the measurement records stored?
- Which correlations were created or consumed?
- Was a reservoir or work source omitted from the description?

Many apparent violations of the second law come from drawing the boundary around only the subsystem that benefits.

6.8 Summary

Table 6.2 collects the bounds of this chapter. All of them are Eq. (6.17) applied to different systems and initial states: the memory alone for erasure, the system in the posterior state for feedback, system and memory together for measurement. What changes from row to row is the information available to the agent and the system to which the bound is applied. A probability distribution, a physical memory, and the bath are separate objects. One must state which entropy belongs to which object and follow the full cycle.

Process	Bound	Conditions
Any isothermal process	$W \geq \Delta \mathcal{F}$, Eq. (6.17)	bath initially canonical and uncorrelated with the system
Erasure of a random bit	$W \geq k_B T \ln 2$, Eq. (6.20)	symmetric memory, controls restored
Erasure, general memory	Eq. (6.26)	local equilibrium inside each logical state
Erasure with a record Y	$W \geq k_B T h(M Y)$, Eq. (6.22)	symmetric memory, protocol chosen according to Y
Feedback	$\langle W \rangle \geq \Delta F - k_B T I(X : M)$, Eq. (6.31)	system initially in equilibrium, measurement that does not disturb it
Measurement and erasure	$W_{\text{meas}} + W_{\text{eras}} \geq k_B T I(X : M)$, Eq. (6.39)	memory in local equilibrium, erasure acting on memory and bath only
Complete cycle	$W_{\text{cycle}} \geq \Delta F$, Eq. (6.40)	all of the above

Table 6.2: The bounds of this chapter. Each is Eq. (6.17) applied to a particular system and initial state, and each is approached by quasistatic protocols that leave no correlation with the bath and no unused correlation with the record.

The same bookkeeping holds for quantum devices, with von Neumann entropies and quantum mutual information in place of their classical counterparts, once the tools of Chapter 7 are available.

6.9 Exercises

Exercise 6.1. Divide the box of the Szilard engine into N equal chambers with $N - 1$ thin partitions, and let the demon record perfectly which chamber holds the molecule. In the feedback, the partitions that do not bound the occupied chamber are removed, and the molecule then pushes the walls of its chamber quasistatically to the ends of the box. Find the extracted work and compare it with $I(X : M)$. Compute the entropy change of the gas during the expansion in two ways, given the record and without it, and show how both fit in a ledger like Table 6.1. What is the minimum work needed to reset the N -state memory, assumed symmetric?

Exercise 6.2. A one-molecule box memory has the cell C_0 of volume $4V/5$ and the cell C_1 of volume $V/5$, and a uniform potential energy $-k_B T \ln 4$ acts on the molecule inside C_1 only. Show that the two cells have equal local free energies but different mean energies and internal entropies. Derive the heat bound (6.27) from Eq. (6.13) and the entropy decomposition of Section 6.2. For a random bit reset to 0, find the minimum work and the minimum heat released, and explain why the heat form (6.21) of Landauer's bound fails for this memory while the work form (6.20) holds.

Exercise 6.3. A memory bit has two nondegenerate states with energies $E_0 = 0$ and $E_1 = \epsilon$, where in this exercise $\epsilon > 0$ is an energy. It initially occupies these states with probabilities $(1 - p, p)$ and is reset to 0. Find the minimum work in terms of its initial nonequilibrium free energy.

Exercise 6.4. The partition of a Szilard engine is inserted so that the left chamber has volume $(1 - q)V$ and the right chamber volume qV , and a binary symmetric measurement reports the side of the molecule with error probability ϵ . Compute $I(X : M)$. For each record m , find the posterior probabilities and the quasistatic partition move that maximizes the extracted work, and show that this work is $k_B T D(p_{X|M=m} \| p_X)$. Average over the records and compare with the bound (6.31). Evaluate everything for $q = 1/4$ and $\epsilon = 0.1$.

Exercise 6.5. Design a logically reversible version of the AND operation by retaining enough input information in auxiliary bits. Explain why logical reversibility does not guarantee zero dissipation at finite speed.

Exercise 6.6 (*). Repeat the entropy ledger of Table 6.1 for a Szilard engine whose measurement errs with probability ϵ , with the optimal feedback of the worked example in Section 6.6.2, a symmetric memory, and a quasistatic erasure. Assume that the measuring device makes its errors spontaneously, without exchanging work or heat. Find the entropy production of each step and the net work per cycle. Which step is irreversible, and what does Eq. (6.37) say about a measurement carried out reversibly?

Exercise 6.7. A one-molecule box memory with its partition in the middle holds the value 1, molecule on the right, with probability p . Construct a protocol of quasistatic partition moves, partition removal, and isothermal compression that resets it to 0 with mean work $k_B T h_2(p)$, reaching the bound (6.25). Show that the protocol of Fig. 6.4 costs $k_B T \ln 2$ for every p , and find the step in which it produces the entropy $k_B [\ln 2 - h_2(p)]$.

Exercise 6.8 (*). A one-molecule box memory has its partition placed so that the cell C_0 has volume αV and the cell C_1 has volume $(1 - \alpha)V$. It stores the result of a perfect measurement of the side of a Szilard molecule, and it is reset to 0 after the feedback. Use Eqs. (6.37) and (6.38) to find the minimum work of the measurement and of the erasure as functions of α , and check that their sum is $k_B T \ln 2$ for every α . Show that the erasure is free for $\alpha = 4/5$. In which step was the work needed for the reset paid?

6.10 Further reading

Leff and Rex collect the original papers on Maxwell's demon, from Maxwell and Szilard to Landauer and Bennett, with an introduction that follows the debate [31]. Bennett's review remains a clear account of reversible computation and of the resolution of the paradox [2], and his later notes answer the objections that the resolution has met [3]. Maruyama, Nori, and Vedral review the subject including its quantum side [35], and Parrondo, Horowitz, and Sagawa give a modern account in the language of stochastic thermodynamics [41]. Cover and Thomas supply the information identities used in this chapter [9].

Part III

Quantum statistics

Chapter 7

Quantum states, entropy, and equilibrium

Lecture 19–20. Density operators, composite systems, von Neumann entropy, quantum relative entropy, Gibbs states, purification, and entanglement.

What you should be able to do

- Distinguish a coherent superposition from a classical mixture.
- Translate probabilities, expectations, and marginalization into operator language.
- Calculate reduced states and entanglement entropy using the Schmidt decomposition.
- Use quantum relative entropy to prove the Gibbs variational principle.
- State precisely which quantum entropies remain constant under unitary evolution.

Core versus further reading

For the two scheduled lectures, the required core ends with the Gibbs variational principle, mutual information, pure-state entanglement entropy, and unitary invariance. Rényi entropies, channel dilations, canonical typicality, and ETH form the guided reading that connects this chapter to current quantum-statistical mechanics.

7.1 From probabilities to operators

The quantum formalism changes what counts as an alternative, but it does not discard the probability theory developed earlier. A state assigns probabilities to measurement outcomes. If a density operator is diagonal in a basis $\{|i\rangle\}$,

$$\rho = \sum_i p_i |i\rangle\langle i|, \quad (7.1)$$

then a measurement with projectors $\Pi_i = |i\rangle\langle i|$ has probabilities

$$\mathbb{P}(i) = \text{Tr}(\rho \Pi_i) = p_i.$$

The entropy $s(\rho)$ defined below then equals the Shannon entropy $h(p)$. The new features appear when no single basis diagonalizes all states and observables of interest, and when the state of a composite system cannot be reconstructed from the states of its parts.

This chapter is organized in three layers. First come states, partial traces, and spectral entropy. The second layer introduces relative entropy and correlations. The third applies these tools to Gibbs states, open evolution, and thermalization. Readers whose quantum-mechanics course did not emphasize density operators should work through the two qubit examples before using the general inequalities.

Three questions to keep separate

1. *Preparation uncertainty*: which density operator describes the system?
2. *Measurement uncertainty*: what is the outcome distribution for a specified measurement?
3. *Entanglement*: is mixedness of a subsystem caused by correlations with another quantum system?

A pure state can have uncertain measurement outcomes, so zero von Neumann entropy does not mean that every observable has a certain value. Conversely, a mixed state can arise from classical preparation uncertainty or from tracing out an entangled partner.

7.2 Density operators

A pure state is represented by a unit vector $|\psi\rangle$, modulo an overall phase. Its density operator is the rank-one projector

$$\rho_\psi = |\psi\rangle\langle\psi|. \quad (7.2)$$

If a preparation device selects $|\psi_a\rangle$ with probability p_a , the expectation of an observable A is

$$\langle A \rangle = \sum_a p_a \langle \psi_a | A | \psi_a \rangle = \text{Tr} \left[\left(\sum_a p_a |\psi_a\rangle\langle\psi_a| \right) A \right].$$

This motivates the density operator of the mixture,

$$\rho = \sum_a p_a |\psi_a\rangle\langle\psi_a|. \quad (7.3)$$

Every density operator is positive and normalized:

$$\rho \geq 0, \quad \text{Tr } \rho = 1. \quad (7.4)$$

Positivity means $\langle \phi | \rho | \phi \rangle \geq 0$ for every vector $|\phi\rangle$; it guarantees nonnegative measurement probabilities. Trace one guarantees that the probabilities of a complete projective measurement sum to one. Conversely, every positive trace-one operator is a valid finite-dimensional state.

Expectation values are

$$\langle A \rangle = \text{Tr}(\rho A). \quad (7.5)$$

A state is pure exactly when

$$\rho^2 = \rho, \quad \text{or equivalently} \quad \text{Tr } \rho^2 = 1. \quad (7.6)$$

To verify the equivalence, diagonalize $\rho = \sum_i \lambda_i |i\rangle\langle i|$. Its eigenvalues obey $0 \leq \lambda_i \leq 1$ and $\sum_i \lambda_i = 1$. Hence

$$\text{Tr } \rho^2 = \sum_i \lambda_i^2 \leq \sum_i \lambda_i = 1.$$

Equality requires every nonzero eigenvalue to satisfy $\lambda_i^2 = \lambda_i$. Normalization then leaves one eigenvalue equal to one and all others equal to zero, which also gives $\rho^2 = \rho$. Mixed states satisfy $\text{Tr } \rho^2 < 1$ in finite dimension.

The decomposition (7.3) is not unique. Different ensembles can produce the same density operator and cannot be distinguished by measurements on the system alone. The state is ρ , not the particular story used to prepare it.

Worked example: superposition versus mixture

In the basis $\{|0\rangle, |1\rangle\}$, compare

$$\rho_+ = |+\rangle\langle+| = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad |+\rangle = \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \quad (7.7)$$

$$\rho_{\text{mix}} = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1| = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad (7.8)$$

Both give probabilities $(1/2, 1/2)$ for a measurement in the computational basis. They differ in the x basis:

$$\text{Tr}(\rho_+ \sigma_x) = 1, \quad \text{Tr}(\rho_{\text{mix}} \sigma_x) = 0. \quad (7.9)$$

The off-diagonal entries encode coherence relative to the chosen basis. The spectra are $(1, 0)$ and $(1/2, 1/2)$, so $s(\rho_+) = 0$ while $s(\rho_{\text{mix}}) = \ln 2$.

7.2.1 Qubit review

The Hermitian matrices $\{\mathbb{I}, \sigma_x, \sigma_y, \sigma_z\}$ form a basis for 2×2 Hermitian operators. Trace one fixes the coefficient of $\mathbb{I}$, so every qubit state can be written

$$\rho = \frac{1}{2}(\mathbb{I} + \mathbf{r} \cdot \boldsymbol{\sigma}), \quad |\mathbf{r}| \leq 1. \quad (7.10)$$

The Pauli identity $(\mathbf{r} \cdot \boldsymbol{\sigma})^2 = |\mathbf{r}|^2 \mathbb{I}$ shows that $\mathbf{r} \cdot \boldsymbol{\sigma}$ has eigenvalues $\pm |\mathbf{r}|$. Thus the eigenvalues of ρ are

$$\lambda_{\pm} = \frac{1 \pm |\mathbf{r}|}{2}. \quad (7.11)$$

Positivity of both eigenvalues is exactly the condition $|\mathbf{r}| \leq 1$. Pure states have spectrum $(1, 0)$ and therefore lie on the Bloch sphere, $|\mathbf{r}| = 1$; the maximally mixed state has spectrum $(1/2, 1/2)$ and sits at the center, $\mathbf{r} = 0$.

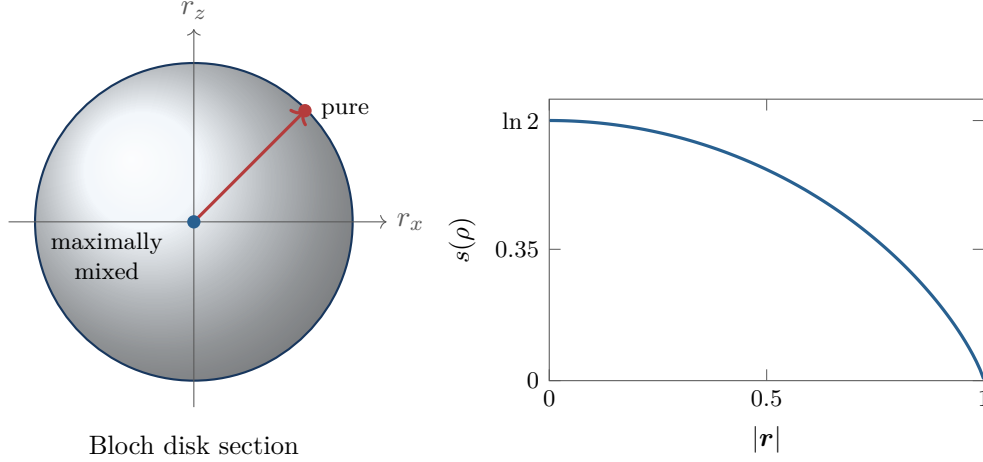


Figure 7.1: A qubit is specified by a Bloch vector inside the unit ball. Purity is radial: the boundary contains pure states and the center is maximally mixed. The entropy depends only on the radius, not on the direction.

7.3 Composite systems and partial trace

For systems A and B ,

$$\mathcal{H}_{AB} = \mathcal{H}_A \otimes \mathcal{H}_B. \quad (7.12)$$

The reduced state of A is

$$\rho_A = \text{Tr}_B \rho_{AB}, \quad (7.13)$$

defined by the requirement

$$\text{Tr}_A(\rho_A A) = \text{Tr}_{AB}[\rho_{AB}(A \otimes \mathbb{I}_B)] \quad (7.14)$$

for every observable A on A . In an orthonormal basis $\{|b\rangle_B\}$, this requirement is implemented by

$$\text{Tr}_B X_{AB} = \sum_b (\mathbb{I}_A \otimes \langle b|) X_{AB} (\mathbb{I}_A \otimes |b\rangle). \quad (7.15)$$

The result does not depend on the chosen basis. The basis-free expectation identity above determines every matrix element of ρ_A by choosing $A = |j\rangle\langle i|$, so it determines ρ_A uniquely; Eq. (7.15) is one basis representation of that unique operator. It is the operator analogue of summing a joint probability $p(a, b)$ over the unobserved variable b .

Consider the Bell state

$$|\Phi^+\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}. \quad (7.16)$$

The composite state is pure, yet

$$\rho_A = \rho_B = \frac{1}{2}\mathbb{I}. \quad (7.17)$$

Local mixedness can therefore arise from entanglement with another system, without ignorance about the global state.

Worked example: performing a partial trace

Expand the Bell projector:

$$\rho_{\Phi^+} = \frac{1}{2}(|00\rangle\langle 00| + |00\rangle\langle 11| + |11\rangle\langle 00| + |11\rangle\langle 11|). \quad (7.18)$$

On elementary operators,

$$\mathrm{Tr}_B(|i\rangle\langle j|_A \otimes |k\rangle\langle \ell|_B) = \delta_{k\ell} |i\rangle\langle j|_A. \quad (7.19)$$

The two cross terms have $k \neq \ell$ and vanish. The diagonal terms give

$$\rho_A = \frac{1}{2} (|0\rangle\langle 0| + |1\rangle\langle 1|) = \frac{\mathbb{I}}{2}. \quad (7.20)$$

Nothing random was added to the pair. The local state is mixed because local expectation values cannot access the phase relation stored in the A – B correlations.

7.3.1 Purification

Let

$$\rho_A = \sum_i \lambda_i |i\rangle\langle i|. \quad (7.21)$$

Introduce a reference system R with orthonormal states $|i\rangle_R$ and define

$$|\Psi\rangle_{AR} = \sum_i \sqrt{\lambda_i} |i\rangle_A |i\rangle_R. \quad (7.22)$$

Expanding the projector and tracing R gives

$$\mathrm{Tr}_R |\Psi\rangle\langle\Psi| = \sum_{i,j} \sqrt{\lambda_i \lambda_j} |i\rangle\langle j|_A \mathrm{Tr}(|i\rangle\langle j|_R) \quad (7.23)$$

$$= \sum_i \lambda_i |i\rangle\langle i|_A = \rho_A, \quad (7.24)$$

because $\mathrm{Tr}(|i\rangle\langle j|) = \delta_{ij}$. Every finite-dimensional mixed state therefore admits a purification. With reference spaces large enough, any two purifications are related by an isometry on the purifying system; this quoted uniqueness statement is the unitary freedom in the Schmidt bases derived in Section 7.8 [39].

Purification does not assert that every laboratory mixture has a unique hidden partner. It says that any mixed-state calculation can be embedded in a larger pure-state calculation. The freedom to act isometrically on R is exactly the nonuniqueness of that embedding. This viewpoint is useful in statistical mechanics because a thermal state of a subsystem may be represented as part of a pure entangled state, without claiming that the whole universe is mixed.

7.4 von Neumann entropy

The dimensionless von Neumann entropy is

$$s(\rho) = -\mathrm{Tr}(\rho \ln \rho). \quad (7.25)$$

The operator logarithm is defined spectrally: if $\rho = \sum_i \lambda_i |i\rangle\langle i|$, then $\ln \rho = \sum_{\lambda_i > 0} (\ln \lambda_i) |i\rangle\langle i|$ on the support of ρ . With the continuous convention $0 \ln 0 = 0$,

$$s(\rho) = -\sum_i \lambda_i \ln \lambda_i. \quad (7.26)$$

Thus von Neumann entropy is the Shannon entropy of the spectrum. Thermodynamic units are restored by $S_{\mathrm{vN}} = k_{\mathrm{B}} s$.

For a d -dimensional system,

$$0 \leq s(\rho) \leq \ln d. \quad (7.27)$$

This is the Shannon bound from Chapter 2 applied to the eigenvalues of ρ . The lower bound holds for pure states; the upper bound holds for $\rho = \mathbb{I}/d$.

Two properties follow directly from eigenvalues. A unitary conjugation changes eigenvectors but not eigenvalues, so

$$s(U\rho U^\dagger) = s(\rho).$$

If ρ_A and ρ_B have eigenvalues $\{a_i\}$ and $\{b_j\}$, the product state has eigenvalues $a_i b_j$. Therefore

$$\begin{aligned} s(\rho_A \otimes \rho_B) &= - \sum_{i,j} a_i b_j \ln(a_i b_j) \\ &= - \sum_i a_i \ln a_i - \sum_j b_j \ln b_j = s(\rho_A) + s(\rho_B), \end{aligned}$$

where $\sum_i a_i = \sum_j b_j = 1$ was used. Two further properties are

$$s(\rho_{AB}) \leq s(\rho_A) + s(\rho_B), \quad (7.28)$$

$$|s(\rho_A) - s(\rho_B)| \leq s(\rho_{AB}). \quad (7.29)$$

The last two are subadditivity and the Araki–Lieb inequality.

Introduce the quantum relative entropy

$$D(\rho \parallel \sigma) = \text{Tr}[\rho(\ln \rho - \ln \sigma)].$$

Klein’s inequality, quoted with its support condition in the next section, states that this quantity is nonnegative. To apply it, use

$$\ln(\rho_A \otimes \rho_B) = \ln \rho_A \otimes \mathbb{I}_B + \mathbb{I}_A \otimes \ln \rho_B$$

on the product support. The required support inclusion is automatic:

$$\text{supp } \rho_{AB} \subseteq \text{supp } \rho_A \otimes \text{supp } \rho_B.$$

For example, if $|a\rangle$ lies in the kernel of ρ_A , then $\text{Tr}[\rho_{AB}(|a\rangle\langle a| \otimes \mathbb{I}_B)] = 0$; positivity of ρ_{AB} means that it has no support on $|a\rangle \otimes \mathcal{H}_B$. The same argument applies to the kernel of ρ_B . The defining property of the partial trace then gives

$$D(\rho_{AB} \parallel \rho_A \otimes \rho_B) = \text{Tr}(\rho_{AB} \ln \rho_{AB}) - \text{Tr}(\rho_A \ln \rho_A) - \text{Tr}(\rho_B \ln \rho_B) \quad (7.30)$$

$$= s(\rho_A) + s(\rho_B) - s(\rho_{AB}) \geq 0. \quad (7.31)$$

This is subadditivity. The Araki–Lieb inequality follows by purifying ρ_{AB} to a pure state on ABC . The Schmidt decomposition of a pure bipartite state, derived in Section 7.8, implies that complementary reductions have the same nonzero spectrum. Applied to the cuts $AB|C$ and $AC|B$, this gives $s(AB) = s(C)$ and $s(AC) = s(B)$. Subadditivity on AC then reads

$$s(B) = s(AC) \leq s(A) + s(C) = s(A) + s(AB),$$

or $s(B) - s(A) \leq s(AB)$. Exchanging A and B gives the opposite difference and completes the absolute-value bound.

Strong subadditivity is a quoted theorem. For any tripartite state it states

$$s(\rho_{ABC}) + s(\rho_B) \leq s(\rho_{AB}) + s(\rho_{BC}). \quad (7.32)$$

It is one of the main structural inequalities of quantum information. We will use it but not prove it.

Here $s(AB)$ abbreviates $s(\rho_{AB})$, and similarly for other subsystems. Define the conditional mutual information

$$I(A : C|B) = s(AB) + s(BC) - s(B) - s(ABC). \quad (7.33)$$

Strong subadditivity says $I(A : C|B) \geq 0$. The theorem was proved by Lieb and Ruskai [32]. Unlike subadditivity, its general proof does not reduce to the scalar log-sum inequality; noncommuting operators are the central difficulty.

Status of the entropy inequalities

Unitary invariance and product additivity follow directly from the spectrum. Subadditivity follows from positivity of quantum relative entropy. Araki–Lieb follows from subadditivity plus purification. Strong subadditivity is a theorem cited above; its proof is outside these notes. Later uses of $I(A : C|B) \geq 0$ depend on that theorem.

7.5 Quantum relative entropy

For density operators ρ and σ ,

$$D(\rho||\sigma) = \text{Tr}[\rho(\ln \rho - \ln \sigma)]. \quad (7.34)$$

The support of a positive operator is the span of its eigenvectors with nonzero eigenvalues. Equation (7.34) is finite when $\text{supp } \rho \subseteq \text{supp } \sigma$; otherwise we define it to be $+\infty$. This convention records the fact that ρ assigns nonzero probability to a direction that σ declares impossible.

The quoted Klein inequality states

$$D(\rho||\sigma) \geq 0, \quad (7.35)$$

with equality exactly when $\rho = \sigma$. We use Klein’s inequality as a quoted operator theorem; proofs may be found in Refs. [39, 60]. Its scalar reduction is the Gibbs inequality proved in Chapter 2.

When ρ and σ commute, choose a common eigenbasis and

$$D(\rho||\sigma) = \sum_i p_i \ln \frac{p_i}{q_i}, \quad (7.36)$$

so the quantum expression reduces to the classical one. For noncommuting states there is no joint list of eigenvalue pairs. The order inside the trace and the support condition are part of the definition.

A quantum channel $\mathcal{N}$ is a linear, completely positive, trace-preserving map. Trace preservation keeps total probability equal to one. Complete positivity means that $\mathcal{N} \otimes \mathbb{I}_R$ maps positive operators to positive operators for every untouched reference system R ; ordinary positivity on the system alone would not suffice for entangled inputs.

We quote the quantum data-processing theorem:

$$D(\mathcal{N}(\rho)||\mathcal{N}(\sigma)) \leq D(\rho||\sigma). \quad (7.37)$$

The partial trace is a channel, so discarding a subsystem cannot make two preparations more distinguishable. Unlike the classical proof in Chapter 2, a proof of Eq. (7.37) must handle noncommuting operator logarithms; we refer to Refs. [39, 60].

A projective measurement with projectors $\{\Pi_x\}$ defines a channel from operators to a classical register,

$$\mathcal{M}(\rho) = \sum_x \text{Tr}(\Pi_x \rho) |x\rangle\langle x|. \quad (7.38)$$

Data processing then gives

$$D(p_\rho^{\mathcal{M}} \| p_\sigma^{\mathcal{M}}) \leq D(\rho \| \sigma). \quad (7.39)$$

Here $p_\rho^{\mathcal{M}}(x) = \text{Tr}(\Pi_x \rho)$ and likewise for σ . No fixed measurement can reveal more statistical distinguishability than is present in the quantum states. This statement does not identify a measurement that attains the quantum relative entropy in one shot; it is an upper bound valid for every measurement.

Status: exact

Klein's inequality and quantum data processing are exact operator theorems quoted here; Refs. [39, 60] give proofs. We do not prove their noncommutative forms. Equality in data processing has a stronger quoted characterization: under suitable support conditions, a recovery map reconstructs both states [60].

7.6 Quantum Gibbs state and free energy

Maximize $s(\rho)$ subject to $\text{Tr} \rho = 1$ and $\text{Tr}(\rho H) = U$. Introduce multipliers α and β' and vary

$$\mathcal{L}(\rho) = s(\rho) - \alpha(\text{Tr} \rho - 1) - \beta'[\text{Tr}(\rho H) - U].$$

For a full-rank stationary state, the trace derivative derived in Appendix A, $\delta s = -\text{Tr}[\delta \rho(\ln \rho + \mathbb{I})]$, gives

$$\delta \mathcal{L} = -\text{Tr}\{\delta \rho[\ln \rho + \mathbb{I} + \alpha \mathbb{I} + \beta' H]\}.$$

This vanishes for every Hermitian $\delta \rho$ only if $\ln \rho = -(1 + \alpha)\mathbb{I} - \beta' H$. Exponentiating gives $\rho = C e^{-\beta' H}$, and normalization fixes $C^{-1} = \text{Tr} e^{-\beta' H}$. The energy constraint fixes β' ; for equilibrium with a bath at temperature T , it equals the thermodynamic inverse temperature $\beta = 1/(k_B T)$. Thus

$$\rho_\beta = \frac{e^{-\beta H}}{Z}, \quad Z = \text{Tr} e^{-\beta H}. \quad (7.40)$$

Define the nonequilibrium free energy

$$F_T(\rho) = \text{Tr}(\rho H) - k_B T s(\rho). \quad (7.41)$$

For the free-energy minimum statement below we assume a bath at $T > 0$. The identity itself remains algebraically valid for a normalizable negative-temperature Gibbs state, but its inequality then reverses. Since $\ln \rho_\beta = -\beta H - \ln Z$,

$$k_B T D(\rho \| \rho_\beta) = k_B T \text{Tr}(\rho \ln \rho) - k_B T \text{Tr}(\rho \ln \rho_\beta) \quad (7.42)$$

$$= -k_B T s(\rho) + \text{Tr}(\rho H) + k_B T \ln Z \quad (7.43)$$

$$= F_T(\rho) - F_T(\rho_\beta). \quad (7.44)$$

The last equality uses $F_T(\rho_\beta) = -k_B T \ln Z$, which follows by substituting $\ln \rho_\beta = -\beta H - \ln Z$ into the entropy. For $T > 0$, Klein's inequality now gives $F_T(\rho) \geq F_T(\rho_\beta)$. At positive finite T , ρ_β has full support, and equality in Klein's inequality holds only for $\rho = \rho_\beta$. The Gibbs state therefore uniquely minimizes F_T . Among states satisfying the same energy constraint $\text{Tr}(\rho H) = U$, the energy term in F_T is fixed. Minimizing $F_T = U - k_B T s(\rho)$ is then exactly the same as maximizing $s(\rho)$, so the stationary solution above is the global entropy maximum rather than merely a local extremum.

Maximum entropy and minimum free energy

The constrained calculation fixes the mean energy and maximizes entropy. At $T > 0$, the relative-entropy identity compares arbitrary states at a fixed bath temperature and minimizes free energy. They select the same Gibbs family, but they answer different variational questions.

Worked example: thermal qubit and the free energy of coherence

Let $H = \epsilon|1\rangle\langle 1|$. The Gibbs state is

$$\rho_\beta = \begin{pmatrix} 1 - p_\beta & 0 \\ 0 & p_\beta \end{pmatrix}, \quad p_\beta = \frac{1}{1 + e^{\beta\epsilon}}. \quad (7.45)$$

Indeed, $Z = 1 + e^{-\beta\epsilon}$ and the excited-state probability is $e^{-\beta\epsilon}/Z = 1/(1 + e^{\beta\epsilon})$. Now compare the pure coherent state $\rho_+ = |+\rangle\langle +|$ with its energy-basis dephasing

$$\Delta(\rho_+) = \frac{1}{2}|0\rangle\langle 0| + \frac{1}{2}|1\rangle\langle 1|. \quad (7.46)$$

Both have mean energy $\epsilon/2$. Their entropies are 0 and $\ln 2$, so

$$F_T(\rho_+) - F_T(\Delta\rho_+) = k_B T \ln 2. \quad (7.47)$$

This free-energy difference is associated with coherence in the energy basis. The identity alone does not guarantee that all of it can be extracted under a restricted set of operations; the allowed controls and any phase reference must also be specified.

Entropy units

$s(\rho)$, $D(\rho\|\sigma)$, and quantum mutual information are dimensionless. The factor $k_B T$ turns relative entropy into a free-energy difference. Keeping this convention prevents the mutual-information identity from acquiring a hidden unit mismatch.

7.7 Quantum mutual information

Define

$$I(A : B)_\rho = s(\rho_A) + s(\rho_B) - s(\rho_{AB}). \quad (7.48)$$

It has the relative-entropy form

$$I(A : B)_\rho = D(\rho_{AB} \| \rho_A \otimes \rho_B). \quad (7.49)$$

The equality is precisely the trace calculation in Eq. (7.31). Klein's inequality therefore makes $I(A : B)_\rho$ nonnegative, with equality exactly when $\rho_{AB} = \rho_A \otimes \rho_B$. It measures total correlation, classical and quantum.

Quantum conditional entropy is

$$s(A|B) = s(AB) - s(B). \quad (7.50)$$

Unlike its classical counterpart, it can be negative. For a maximally entangled pair of qubits,

$$s(A|B) = -\ln 2. \quad (7.51)$$

Indeed, the Bell state is pure, so $s(AB) = 0$, while its reduced state is $\rho_B = \mathbb{I}/2$, so $s(B) = \ln 2$. Thus $s(A|B) = 0 - \ln 2$. Negative conditional entropy has an operational meaning in quantum communication protocols [60]; it is not a negative number of classical alternatives.

Worked example: the same local states, different correlations

Compare the classically correlated state

$$\rho_{cc} = \frac{1}{2}|00\rangle\langle 00| + \frac{1}{2}|11\rangle\langle 11| \quad (7.52)$$

with the Bell state ρ_{Φ^+} . Both have $\rho_A = \rho_B = \mathbb{I}/2$. For ρ_{cc} , the joint eigenvalues are $(1/2, 1/2, 0, 0)$, and hence

$$s(AB) = \ln 2, \quad I(A : B) = \ln 2, \quad s(A|B) = 0. \quad (7.53)$$

For the Bell state,

$$s(AB) = 0, \quad I(A : B) = 2 \ln 2, \quad s(A|B) = -\ln 2. \quad (7.54)$$

The local entropy alone cannot tell whether the correlation is classical or quantum. Mutual information sees the amount of total correlation but is not, for mixed states, an entanglement measure.

7.8 Entanglement entropy

Every pure bipartite state admits a Schmidt decomposition

$$|\psi\rangle_{AB} = \sum_{i=1}^r \sqrt{\lambda_i} |i\rangle_A |i\rangle_B. \quad (7.55)$$

To obtain it, choose product bases and expand

$$|\psi\rangle = \sum_{i,j} C_{ij} |i\rangle_A |j\rangle_B.$$

Treat the coefficients C_{ij} as a matrix and take a singular-value decomposition $C = U\sqrt{\Lambda}V^\dagger$. Absorbing U and the complex conjugate of V into new orthonormal bases leaves only the nonnegative singular values $\sqrt{\lambda_i}$, giving Eq. (7.55). Normalization yields $\sum_i \lambda_i = 1$. Expanding $|\psi\rangle\langle\psi|$ and using $\text{Tr}_B(|i\rangle\langle j|_B) = \delta_{ij}$ gives

$$\rho_A = \sum_i \lambda_i |i\rangle_A \langle i|, \quad \rho_B = \sum_i \lambda_i |i\rangle_B \langle i|. \quad (7.56)$$

The reduced states share the nonzero spectrum $\{\lambda_i\}$, so

$$s(\rho_A) = s(\rho_B). \quad (7.57)$$

For a pure bipartite state, this common value is the entanglement entropy:

$$s_{\text{ent}}(A) = s(\rho_A). \quad (7.58)$$

It vanishes exactly for product states and equals $\ln d$ for a maximally entangled state with Schmidt rank d .

For a mixed global state, $s(\rho_A)$ is not an entanglement measure by itself. It includes local thermal or classical uncertainty. Mixed-state entanglement requires other constructions.

What entanglement entropy does not measure

For pure bipartite states, $s(\rho_A)$ is both the local mixedness and the entanglement entropy. For mixed ρ_{AB} these roles separate. A product thermal state can have large $s(\rho_A)$ and no entanglement; a classically correlated state can have nonzero mutual information and still be separable. Statements about entanglement entropy must therefore specify that the global state is pure or introduce a mixed-state measure.

7.9 Rényi entropies

For $\alpha > 0$, $\alpha \neq 1$, define

$$s_\alpha(\rho) = \frac{1}{1-\alpha} \ln \text{Tr } \rho^\alpha. \quad (7.59)$$

As $\alpha \rightarrow 1$, both numerator and denominator vanish. L'Hôpital's rule gives

$$\lim_{\alpha \rightarrow 1} s_\alpha(\rho) = - \left. \frac{\text{Tr}(\rho^\alpha \ln \rho)}{\text{Tr } \rho^\alpha} \right|_{\alpha=1} = -\text{Tr}(\rho \ln \rho) = s(\rho). \quad (7.60)$$

The value $\alpha = 2$ is related to purity:

$$s_2(\rho) = -\ln \text{Tr } \rho^2. \quad (7.61)$$

Different Rényi indices probe different parts of the spectrum. In lattice models and field theory, integer Rényi entropies often admit replica constructions.

7.10 Dynamics: closed and open

For a closed quantum system,

$$\rho(t) = U(t)\rho(0)U^\dagger(t), \quad (7.62)$$

so its eigenvalues and von Neumann entropy remain constant:

$$s(\rho(t)) = s(\rho(0)). \quad (7.63)$$

For a subsystem A of an evolving pure state AB , $s(\rho_A)$ can grow because entanglement moves information into correlations with B . The global entropy stays zero.

We quote the finite-dimensional Kraus representation theorem: every quantum channel can be written [39]

$$\rho \mapsto \mathcal{N}(\rho) = \sum_a K_a \rho K_a^\dagger, \quad \sum_a K_a^\dagger K_a = \mathbb{I}. \quad (7.64)$$

The operators K_a are Kraus operators. Their completeness relation ensures trace preservation:

$$\text{Tr } \mathcal{N}(\rho) = \sum_a \text{Tr}(K_a^\dagger K_a \rho) = \text{Tr } \rho.$$

The Kraus form also makes complete positivity manifest. Every such finite-dimensional channel can be represented by adjoining an environment in a fixed state, applying a joint unitary, and discarding the environment:

$$\mathcal{N}(\rho) = \text{Tr}_E \left[U(\rho \otimes |0\rangle\langle 0|_E) U^\dagger \right]. \quad (7.65)$$

To construct the dilation, define

$$V|\psi\rangle = \sum_a K_a |\psi\rangle |a\rangle_E.$$

For arbitrary $|\phi\rangle$ and $|\psi\rangle$,

$$\langle\phi|V^\dagger V|\psi\rangle = \sum_a \langle\phi|K_a^\dagger K_a|\psi\rangle = \langle\phi|\psi\rangle.$$

Thus $V^\dagger V = \mathbb{I}$: V preserves inner products and is an isometry. It can be extended to a unitary U on a sufficiently large system–environment space, a standard finite-dimensional linear-algebra fact discussed in Ref. [39]. Choosing the extension so that $U(|\psi\rangle|0\rangle_E) = V|\psi\rangle$ gives $K_a = {}_E\langle a|U|0\rangle_E$. Expanding in the basis $\{|a\rangle_E\}$ and tracing E produces Eq. (7.65). Any apparent nonunitarity of the system comes from interaction with and subsequent neglect of E .

Von Neumann entropy does not increase under every channel. An amplitude-damping channel can purify its system as energy and entropy pass to the environment. A channel is *unital* when $\mathcal{N}(\mathbb{I}) = \mathbb{I}$, or equivalently $\mathcal{N}(\mathbb{I}/d) = \mathbb{I}/d$ in dimension d . Since

$$D(\rho||\mathbb{I}/d) = \text{Tr}(\rho \ln \rho) + \ln d = \ln d - s(\rho),$$

data processing gives

$$\begin{aligned} \ln d - s(\mathcal{N}(\rho)) &= D(\mathcal{N}(\rho)||\mathcal{N}(\mathbb{I}/d)) \\ &\leq D(\rho||\mathbb{I}/d) = \ln d - s(\rho). \end{aligned}$$

After cancelling $\ln d$ and reversing signs,

$$s(\mathcal{N}(\rho)) \geq s(\rho). \quad (7.66)$$

Thus unital channels do not decrease entropy in finite dimension.

Worked example: dephasing and amplitude damping

A qubit dephasing channel can be written

$$\mathcal{D}_p(\rho) = (1-p)\rho + p\sigma_z\rho\sigma_z. \quad (7.67)$$

It is unital. Acting on $|+\rangle$, it changes the Bloch vector from $(1, 0, 0)$ to $(1-2p, 0, 0)$; the entropy rises from 0 to $\ln 2$ as p goes from 0 to $1/2$. The output eigenvalues are p and $1-p$, so this entropy is $h_2(p)$ on that interval.

Amplitude damping has Kraus operators

$$K_0 = |0\rangle\langle 0| + \sqrt{1-\gamma}|1\rangle\langle 1|, \quad K_1 = \sqrt{\gamma}|0\rangle\langle 1|. \quad (7.68)$$

Starting from $|1\rangle$, the output is

$$(1-\gamma)|1\rangle\langle 1| + \gamma|0\rangle\langle 0|. \quad (7.69)$$

Its eigenvalues are γ and $1-\gamma$, so its entropy is $h_2(\gamma)$. It rises from zero to $\ln 2$ at $\gamma = 1/2$ and returns to zero at $\gamma = 1$, when the system has relaxed to the pure ground state. The environment carries the emitted excitation and the information needed to keep the joint evolution unitary. System entropy is therefore not a general Lyapunov function for open dynamics.

7.11 Typicality and thermalization

7.11.1 Canonical typicality: a statement about most vectors

Let $\mathcal{H}_R \subset \mathcal{H}_S \otimes \mathcal{H}_B$ be a large constrained subspace, such as an energy shell. Its microcanonical state and the corresponding subsystem state are

$$\Omega_R = \frac{\Pi_R}{d_R}, \quad \Omega_S = \text{Tr}_B \Omega_R. \quad (7.70)$$

Here Π_R is the projector onto $\mathcal{H}_R$ and $d_R = \text{Tr} \Pi_R$ is its dimension. Choose a unit vector $|\psi\rangle$ with the unitarily invariant, or Haar, probability measure on this subspace and define $\rho_S(\psi) = \text{Tr}_B |\psi\rangle\langle\psi|$.

We quote a canonical-typicality theorem [45]. Let $d_S = \dim \mathcal{H}_S$, define $\Omega_B = \text{Tr}_S \Omega_R$, and set $d_B^{\text{eff}} = 1/\text{Tr} \Omega_B^2$. Then

$$\mathbb{E}_\psi \|\rho_S(\psi) - \Omega_S\|_1 \leq \sqrt{\frac{d_S}{d_B^{\text{eff}}}}, \quad d_B^{\text{eff}} = \frac{1}{\text{Tr} \Omega_B^2}. \quad (7.71)$$

Here $\|A\|_1 = \text{Tr} \sqrt{A^\dagger A}$ is the trace norm and $\mathbb{E}_\psi$ denotes the Haar average. If $d_B^{\text{eff}} \gg d_S$, the right-hand side is small. Applying Markov's inequality, Eq. (2.24), to the nonnegative random variable $\|\rho_S(\psi) - \Omega_S\|_1$ gives, for every $\delta > 0$,

$$\mathbb{P}_\psi(\|\rho_S(\psi) - \Omega_S\|_1 \geq \delta) \leq \frac{1}{\delta} \sqrt{\frac{d_S}{d_B^{\text{eff}}}}. \quad (7.72)$$

Sharper results give exponential concentration around this scale [45].

This result reconciles local thermality with global purity. It does not require the global vector to have large von Neumann entropy; its entropy is zero. Instead, entanglement makes the reduced state nearly indistinguishable from the reduced microcanonical state for measurements confined to S .

Status: kinematic theorem

Canonical typicality is a theorem about the measure of vectors in a constrained Hilbert subspace. Its conclusion depends on the relative effective dimensions and on the chosen constraint subspace. It does not assert that a time-evolving state becomes a typical vector.

7.11.2 Time evolution and the diagonal ensemble

Write an initial state in the energy basis,

$$|\psi(0)\rangle = \sum_n c_n |E_n\rangle. \quad (7.73)$$

Under $U(t) = e^{-iHt/\hbar}$,

$$|\psi(t)\rangle = \sum_n c_n e^{-iE_n t/\hbar} |E_n\rangle.$$

For an observable O , define $O_{mn} = \langle E_m | O | E_n \rangle$. Substitution into $\langle \psi(t) | O | \psi(t) \rangle$ gives

$$\langle O(t) \rangle = \sum_{m,n} c_m^* c_n e^{i(E_m - E_n)t/\hbar} O_{mn}. \quad (7.74)$$

If the energies present in the state are nondegenerate, then $E_m - E_n \neq 0$ for $m \neq n$, and

$$\lim_{\tau \rightarrow \infty} \frac{1}{\tau} \int_0^\tau e^{i(E_m - E_n)t/\hbar} dt = \lim_{\tau \rightarrow \infty} \frac{e^{i(E_m - E_n)\tau/\hbar} - 1}{i(E_m - E_n)\tau/\hbar} = 0.$$

Define $\overline{\langle O \rangle} = \lim_{\tau \rightarrow \infty} \tau^{-1} \int_0^\tau \langle O(t) \rangle dt$. The preceding integral removes every off-diagonal term and leaves

$$\overline{\langle O \rangle} = \sum_n |c_n|^2 O_{nn} = \text{Tr}(\rho_{\text{diag}} O), \quad \rho_{\text{diag}} = \sum_n |c_n|^2 |E_n\rangle\langle E_n|. \quad (7.75)$$

This is dephasing under a time average, not convergence of the global density operator in trace norm. If energy levels are degenerate, coherences within each degenerate block remain. Recurrences remain possible, and the diagonal ensemble retains every energy-basis population $|c_n|^2$ in the nondegenerate case shown.

7.11.3 The eigenstate thermalization hypothesis

The eigenstate thermalization hypothesis (ETH) proposes, for suitable few-body observables in a nonintegrable many-body system,

$$O_{mn} = \overline{O}(E) \delta_{mn} + e^{-S(E)/(2k_B)} f_O(E, \omega) R_{mn}, \quad (7.76)$$

where $E = (E_m + E_n)/2$, $\omega = E_m - E_n$, $S(E)$ is the thermodynamic entropy near energy E , $\overline{O}(E)$ varies smoothly on thermodynamic energy scales, and R_{mn} has irregular order-one entries. The diagonal part makes nearby energy eigenstates agree on macroscopic observables. The ansatz assigns an exponentially small scale to off-diagonal matrix elements, which suppresses late-time fluctuations in the systems for which ETH applies.

The connection to equilibrium averages has two separate steps. The exact diagonal-ensemble result is $\overline{\langle O \rangle} = \sum_n |c_n|^2 O_{nn}$. If $|c_n|^2$ is concentrated in a narrow energy window around E_0 and the ETH diagonal elements satisfy $O_{nn} \simeq \overline{O}(E_n)$, then smoothness across that window gives

$$\overline{\langle O \rangle} \simeq \sum_n |c_n|^2 \overline{O}(E_n) \simeq \overline{O}(E_0). \quad (7.77)$$

The final quantity agrees with the microcanonical prediction when $\overline{O}(E)$ has that thermodynamic interpretation. Equivalence of ensembles can then replace it by a canonical expectation for the appropriate class of local observables.

ETH is well supported for broad classes of nonintegrable models, but it fails in integrable systems, many-body localized regimes, and models with additional constraints or special nonthermal eigenstates. A review and precise variants are given in [11].

Status: structured hypothesis, not a universal theorem

Equation (7.76) is an ansatz whose domain must be checked for the Hamiltonian and observable in question. The exact statement in (7.74) is only the energy-basis expansion. The time-average formula shown above assumes nondegenerate energies; bounds on temporal fluctuations often require stronger conditions on energy gaps. Replacement of the diagonal ensemble by a thermal ensemble adds ETH or another thermalization mechanism.

7.12 Further reading

Nielsen and Chuang introduce density operators, purification, and entanglement [39]. Wilde develops quantum entropies and data processing in detail [60]. Witten offers a physics-focused discussion of classical and quantum information measures [62].

7.13 Exercises

Exercise 7.1. For a qubit with Bloch vector $\mathbf{r}$, calculate $s(\rho)$ and $s_2(\rho)$. Examine the pure and maximally mixed limits.

Exercise 7.2. Compute the reduced state and entanglement entropy of $\sqrt{p}|00\rangle + \sqrt{1-p}|11\rangle$.

Exercise 7.3. Show that two different ensembles of pure qubit states can represent $\rho = \mathbb{I}/2$. Explain why no measurement on the qubit can identify which ensemble decomposition was used.

Exercise 7.4. Derive the Gibbs state by constrained variation of $s(\rho)$. Treat the trace constraint carefully.

Exercise 7.5. Prove the free-energy identity (7.44). Use it to show uniqueness of the Gibbs state when $T > 0$.

Exercise 7.6. For the Bell state, calculate $s(A)$, $s(B)$, $s(AB)$, $I(A : B)$, and $s(A|B)$.

Exercise 7.7. Give a channel that decreases system entropy and a unital channel that increases it. In each example, identify the environment's role.

Exercise 7.8 (*). Use monotonicity of quantum relative entropy under partial trace to prove monotonicity of mutual information under a local channel on B .

Exercise 7.9. For $\rho = \text{diag}(p, 1-p)$ and $\sigma = \text{diag}(q, 1-q)$, calculate $D(\rho||\sigma)$ and show that it is the classical relative entropy of two Bernoulli distributions. State when the answer is infinite.

Exercise 7.10. Compare the Bell state $|\Phi^+\rangle$ with the classically correlated state $\rho_{cc} = (|00\rangle\langle 00| + |11\rangle\langle 11|)/2$. Compute the marginal entropies and mutual information of both states, and explain why equal marginals do not imply equal correlations.

Exercise 7.11. Construct a purification of a three-level state with eigenvalues (p_1, p_2, p_3) . Trace out each subsystem explicitly and identify the minimum dimension required for the purifying space when exactly r eigenvalues are nonzero.

Exercise 7.12 (*). Repeat the time-average calculation of Section 7.11.2 when the Hamiltonian has degenerate energy eigenspaces. Show that coherences between different energies vanish in the infinite-time average while coherences within a degenerate block can remain.

Chapter 8

Identical particles and ideal quantum gases

Lecture 21–24. Fock space, Bose and Fermi distributions, degenerate Fermi gases, Bose–Einstein condensation, photons, and the classical limit.

What you should be able to do

- Build grand partition functions from occupation numbers.
- Derive the Fermi–Dirac and Bose–Einstein distributions, including their mode fluctuations.
- Calculate the low-temperature thermodynamics of a Fermi gas.
- State the dimensional condition for Bose–Einstein condensation.
- Derive Planck’s law and photon-gas thermodynamics.
- Recover the classical ideal gas and its leading exchange correction.

8.1 Identical particles

For identical particles, labels attached to individual particles have no observable meaning. We take the symmetrization postulate from quantum mechanics. If $U(P)$ implements a permutation P , then

$$U(P)|\Psi\rangle = \begin{cases} |\Psi\rangle, & \text{bosons,} \\ \text{sgn}(P)|\Psi\rangle, & \text{fermions.} \end{cases} \quad (8.1)$$

Here $\text{sgn}(P) = +1$ for an even permutation and -1 for an odd one. We do not derive the postulate; Ref. [42] discusses the fully symmetric and fully antisymmetric sectors.

The Pauli principle follows immediately. If two fermions occupied the same one-particle state $|\phi\rangle$, exchanging them would leave the product $|\phi\rangle|\phi\rangle$ unchanged. Antisymmetry would require the same vector to acquire a minus sign, so it must vanish. A nonzero fermionic state therefore has at most one particle in each one-particle state.

Occupation numbers give a natural basis. The integer n_r records how many particles occupy the one-particle state r . If its energy is ϵ_r , a many-body basis vector and its additive observables are

$$|n_1, n_2, \dots\rangle, \quad E = \sum_r n_r \epsilon_r, \quad N = \sum_r n_r. \quad (8.2)$$

For bosons, $n_r = 0, 1, 2, \dots$; for fermions, $n_r = 0, 1$.

The symmetric or antisymmetric Fock-space structure should not be identified automatically with operational entanglement between particle labels. Identical particles do not define distinguishable tensor factors in the same way as two spatially separated qubits. Mode entanglement and particle entanglement require a specified observable algebra or subsystem split.

8.2 Grand partition function

For a noninteracting quantum gas,

$$H - \mu N = \sum_r (\epsilon_r - \mu) n_r. \quad (8.3)$$

Write $\beta = 1/(k_B T)$ and define the fugacity

$$z = e^{\beta\mu}. \quad (8.4)$$

The grand partition function is a trace over all allowed occupation-number configurations:

$$\Xi = \text{Tr} e^{-\beta(H - \mu N)} \quad (8.5)$$

$$= \sum_{\{n_r\}} \prod_r e^{-\beta(\epsilon_r - \mu)n_r} \quad (8.6)$$

$$= \prod_r \sum_{n_r} \left(z e^{-\beta\epsilon_r} \right)^{n_r} \equiv \prod_r \Xi_r. \quad (8.7)$$

The second line factorizes because the energy and particle number are sums of independent mode contributions. The only difference between fermions and bosons is the allowed range of n_r .

For fermions $n_r = 0, 1$, so

$$\Xi_r^{(F)} = \sum_{n=0}^1 (z e^{-\beta\epsilon_r})^n = 1 + z e^{-\beta\epsilon_r}, \quad \ln \Xi_F = \sum_r \ln(1 + z e^{-\beta\epsilon_r}). \quad (8.8)$$

For bosons $n_r = 0, 1, 2, \dots$, and the geometric series gives

$$\Xi_r^{(B)} = \sum_{n=0}^{\infty} (z e^{-\beta\epsilon_r})^n = \frac{1}{1 - z e^{-\beta\epsilon_r}}, \quad \ln \Xi_B = - \sum_r \ln(1 - z e^{-\beta\epsilon_r}). \quad (8.9)$$

The geometric series converges only if $z e^{-\beta\epsilon_r} < 1$ for every mode. The strongest restriction comes from the lowest energy ϵ_0 :

$$z e^{-\beta\epsilon_0} < 1 \quad \Longleftrightarrow \quad \mu < \epsilon_0.$$

In the thermodynamic limit a Bose gas can approach $\mu = \epsilon_0$ from below; this limiting behavior underlies condensation.

For either statistic, the mean occupation follows from a logarithmic derivative of the single-mode partition function:

$$\langle n_r \rangle = \frac{\sum_{n_r} n_r x_r^{n_r}}{\sum_{n_r} x_r^{n_r}} = x_r \frac{\partial \ln \Xi_r}{\partial x_r}, \quad x_r = z e^{-\beta\epsilon_r}.$$

For a fermionic mode, the two weights 1 and x_r give $\langle n_r \rangle = x_r/(1 + x_r)$. For a bosonic mode, differentiating the geometric sum gives $\langle n_r \rangle = x_r/(1 - x_r)$. Since $x_r^{-1} = e^{\beta(\epsilon_r - \mu)}$, both cases can be written

$$\langle n_r \rangle = \frac{1}{e^{\beta(\epsilon_r - \mu)} \pm 1}, \quad \begin{array}{l} + : \text{ fermions,} \\ - : \text{ bosons,} \end{array} \quad (8.10)$$

The formula can be read as a modified Boltzmann weight. For $e^{\beta(\epsilon-\mu)} \gg 1$, the ± 1 is negligible and both distributions become $\langle n \rangle \simeq e^{-\beta(\epsilon-\mu)}$. Near unit occupation the signs matter: exclusion caps a fermionic mode at one particle, while stimulated occupation allows a bosonic mode to become macroscopically populated.

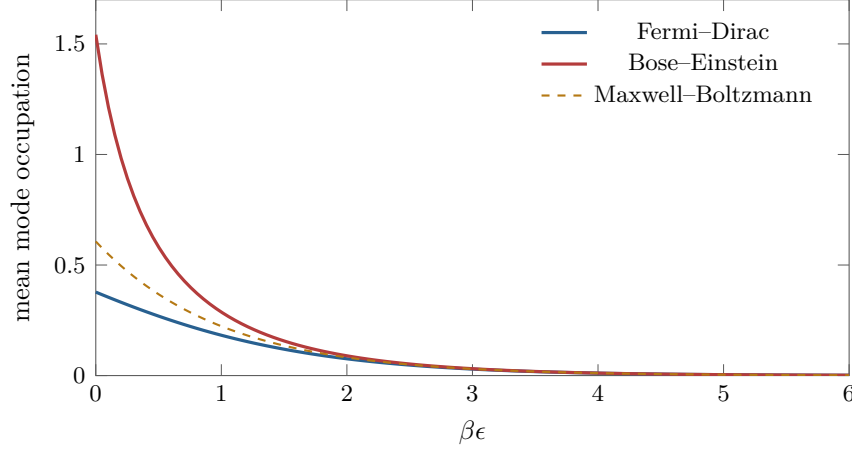


Figure 8.1: Mode occupations at the same fugacity, $\beta\mu = -0.5$. Quantum statistics is most visible in the highly occupied low-energy modes. In the dilute high-energy tail all three distributions agree.

Number fluctuations in one mode are

$$\text{Var}(n_r) = \langle n_r \rangle (1 \mp \langle n_r \rangle), \quad (8.11)$$

where the upper sign corresponds to fermions. Pauli exclusion suppresses fluctuations; Bose enhancement increases them.

The fluctuation identity can be derived from the mode weights. With $x_r = ze^{-\beta\epsilon_r}$,

$$\begin{aligned} x_r \frac{\partial \langle n_r \rangle}{\partial x_r} &= \frac{\sum_n n^2 x_r^n}{\sum_n x_r^n} - \left(\frac{\sum_n n x_r^n}{\sum_n x_r^n} \right)^2 \\ &= \langle n_r^2 \rangle - \langle n_r \rangle^2 = \text{Var}(n_r). \end{aligned}$$

At fixed ϵ_r and T , x_r is proportional to z , so

$$\langle n_r \rangle = z \frac{\partial \ln \Xi_r}{\partial z}, \quad \text{Var}(n_r) = z \frac{\partial \langle n_r \rangle}{\partial z}. \quad (8.12)$$

Differentiating $\langle n_r \rangle = x_r/(1 \pm x_r)$ gives $\langle n_r \rangle (1 - \langle n_r \rangle)$ for fermions and $\langle n_r \rangle (1 + \langle n_r \rangle)$ for bosons. The sign of the statistical correlation is already visible in one-mode number noise.

8.3 Density of states

Take a cubic box of side L and volume $V = L^3$, with periodic boundary conditions. Each allowed momentum occupies a cell of volume $(2\pi\hbar/L)^3$ in momentum space. Including an internal degeneracy g , the large-volume replacement is

$$\sum_{\mathbf{p}} \rightarrow \frac{gV}{(2\pi\hbar)^3} \int d^3p, \quad (8.13)$$

where $g = 2$ for spin-1/2 particles when both spin states are available. Define $D(\epsilon)$ so that $D(\epsilon) d\epsilon$ counts one-particle states with energies in $[\epsilon, \epsilon + d\epsilon]$. For the nonrelativistic dispersion $\epsilon = p^2/(2m)$, the number of states below momentum p is

$$\mathcal{N}(p) = \frac{gV}{(2\pi\hbar)^3} \frac{4\pi p^3}{3}. \quad (8.14)$$

Since $p = (2m\epsilon)^{1/2}$,

$$\begin{aligned} D(\epsilon) &= \frac{d\mathcal{N}}{d\epsilon} = \frac{gV}{2\pi^2\hbar^3} p^2 \frac{dp}{d\epsilon} \\ &= \frac{gV}{2\pi^2\hbar^3} p^2 \frac{m}{p} = \frac{gV}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \epsilon^{1/2}. \end{aligned}$$

Thus

$$D(\epsilon) = \frac{gV}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \epsilon^{1/2}. \quad (8.15)$$

The total mean particle number and energy sum the occupation of every level. Replacing those sums by energy integrals gives

$$N = \int_0^\infty d\epsilon \frac{D(\epsilon)}{e^{\beta(\epsilon-\mu)} \pm 1}, \quad U = \int_0^\infty d\epsilon \frac{D(\epsilon)\epsilon}{e^{\beta(\epsilon-\mu)} \pm 1}. \quad (8.16)$$

Lecture 22 begins here with the degenerate Fermi gas and its low-temperature expansion.

8.4 Degenerate Fermi gas

For fermions, the limit $T \rightarrow 0$ turns Eq. (8.10) into a step. If $\epsilon < \mu$, then $\beta(\epsilon - \mu) \rightarrow -\infty$ and $\langle n \rangle \rightarrow 1$; if $\epsilon > \mu$, then $\langle n \rangle \rightarrow 0$. The zero-temperature chemical potential is called the Fermi energy, $\mu(0) = \epsilon_F$, so

$$\langle n(\epsilon) \rangle = \Theta(\epsilon_F - \epsilon). \quad (8.17)$$

Define the Fermi momentum p_F and Fermi wave number k_F by $p_F = \hbar k_F$ and $\epsilon_F = p_F^2/(2m)$. Counting all states inside the Fermi sphere with Eq. (8.14) gives

$$N = \frac{gV}{6\pi^2\hbar^3} p_F^3 = \frac{gV}{6\pi^2} k_F^3, \quad k_F = \left(\frac{6\pi^2 n}{g} \right)^{1/3}, \quad (8.18)$$

where $n = N/V$. The Fermi energy is therefore

$$\epsilon_F = \frac{\hbar^2 k_F^2}{2m}. \quad (8.19)$$

For spin-1/2 particles, $g = 2$ and $k_F = (3\pi^2 n)^{1/3}$.

The ground-state energy is

$$U_0 = \int_0^{\epsilon_F} \epsilon D(\epsilon) d\epsilon = \frac{3}{5} N \epsilon_F. \quad (8.20)$$

To see the coefficient without carrying the density-of-states constant, write $D(\epsilon) = C\epsilon^{1/2}$. Then

$$N = C \int_0^{\epsilon_F} \epsilon^{1/2} d\epsilon = \frac{2C}{3} \epsilon_F^{3/2}, \quad U_0 = C \int_0^{\epsilon_F} \epsilon^{3/2} d\epsilon = \frac{2C}{5} \epsilon_F^{5/2},$$

and their ratio is $U_0/N = 3\epsilon_F/5$.

The relation $PV = 2U/3$ can be derived directly from the grand partition function. Let $\eta = +1$ for fermions and $\eta = -1$ for bosons. In the continuum,

$$\ln \Xi = \eta \int_0^\infty d\epsilon D(\epsilon) \ln(1 + \eta z e^{-\beta\epsilon}).$$

For $f(\epsilon) = \ln(1 + \eta z e^{-\beta\epsilon})$,

$$\frac{df}{d\epsilon} = -\beta\eta \langle n(\epsilon) \rangle.$$

Since $D(\epsilon) = C\epsilon^{1/2}$ and $(2/3)\epsilon D(\epsilon)$ is an antiderivative of $D(\epsilon)$, integration by parts gives

$$\begin{aligned} \ln \Xi &= \eta \left[\frac{2}{3} \epsilon D(\epsilon) f(\epsilon) \right]_0^\infty - \frac{2\eta}{3} \int_0^\infty \epsilon D(\epsilon) f'(\epsilon) d\epsilon \\ &= \frac{2\beta}{3} \int_0^\infty d\epsilon D(\epsilon) \epsilon \langle n(\epsilon) \rangle = \frac{2\beta U}{3}. \end{aligned}$$

The boundary term vanishes: the factor $\epsilon D(\epsilon)$ vanishes at the lower limit and the logarithm decays exponentially at the upper limit. For a homogeneous system, the grand potential satisfies both $\Phi = -k_B T \ln \Xi$ and $\Phi = -PV$, so $PV = k_B T \ln \Xi = \ln \Xi / \beta$. It follows that $PV = 2U/3$. This continuum thermodynamic-limit relation ignores a separately isolated zero-energy Bose ground mode; that mode contributes no energy and no extensive pressure. For the Fermi gas at zero temperature, $U_0 = (3/5)N\epsilon_F$, so

$$P_0 = \frac{2}{5} n \epsilon_F. \quad (8.21)$$

This degeneracy pressure remains nonzero at zero temperature.

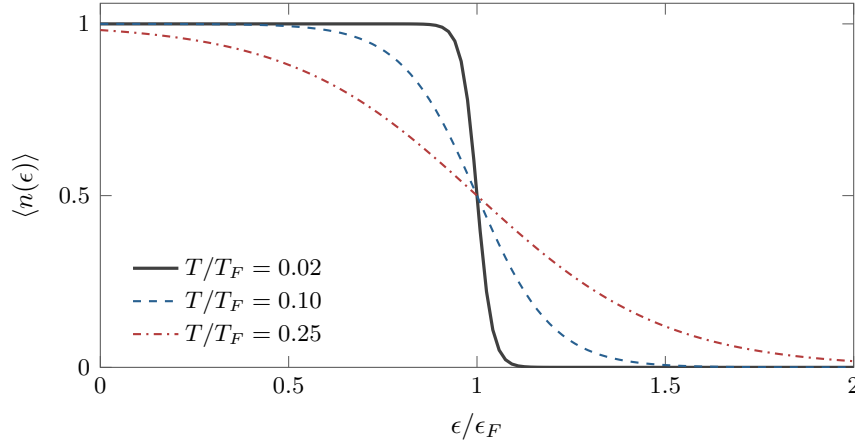


Figure 8.2: Fermi occupation near the Fermi surface, with the small temperature dependence of μ omitted in the plot. Thermal changes are confined to an energy window of order $k_B T$ around ϵ_F .

A metallic-density electron gas

For spin-1/2 electrons at number density $n = 10^{29} \text{ m}^{-3}$,

$$k_F \simeq 1.44 \times 10^{10} \text{ m}^{-1}, \quad \epsilon_F \simeq 7.86 \text{ eV}, \quad T_F = \frac{\epsilon_F}{k_B} \simeq 9.12 \times 10^4 \text{ K}. \quad (8.22)$$

The Fermi velocity is $v_F = \hbar k_F / m_e \simeq 1.66 \times 10^6 \text{ m s}^{-1}$. At room temperature $T/T_F \simeq 3.3 \times 10^{-3}$, so the gas is deeply degenerate even though the material is not cold on an everyday scale.

8.4.1 Sommerfeld expansion

For a smooth function $\phi(\epsilon)$,

$$\int_0^\infty \frac{\phi(\epsilon) d\epsilon}{e^{\beta(\epsilon-\mu)} + 1} = \int_0^\mu \phi(\epsilon) d\epsilon + \frac{\pi^2}{6} (k_B T)^2 \phi'(\mu) + \frac{7\pi^4}{360} (k_B T)^4 \phi^{(3)}(\mu) + \dots \quad (8.23)$$

To derive the coefficients, subtract a sharp step at the same value of μ and set $x = |\epsilon - \mu|/(k_B T)$. When $\mu \gg k_B T$, extending the lower-side integral to $x = \infty$ changes the result only by terms exponentially small in $\mu/(k_B T)$. Thus

$$I(T, \mu) - \int_0^\mu \phi(\epsilon) d\epsilon = k_B T \int_0^\infty \frac{dx}{e^x + 1} [\phi(\mu + k_B T x) - \phi(\mu - k_B T x)].$$

The symmetric difference contains only odd derivatives:

$$\phi(\mu + k_B T x) - \phi(\mu - k_B T x) = 2k_B T x \phi'(\mu) + \frac{(k_B T x)^3}{3} \phi^{(3)}(\mu) + \dots$$

The needed integrals follow from

$$\frac{1}{e^x + 1} = \sum_{\ell=1}^{\infty} (-1)^{\ell+1} e^{-\ell x}, \quad \int_0^\infty x^q e^{-\ell x} dx = \frac{q!}{\ell^{q+1}}.$$

For $q = 1$ and $q = 3$ this gives

$$\begin{aligned} \int_0^\infty \frac{x dx}{e^x + 1} &= \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}}{\ell^2} = (1 - 2^{-1})\zeta(2) = \frac{\pi^2}{12}, \\ \int_0^\infty \frac{x^3 dx}{e^x + 1} &= 6 \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell+1}}{\ell^4} = 6(1 - 2^{-3})\zeta(4) = \frac{7\pi^4}{120}. \end{aligned}$$

Substituting these two values into the odd-derivative expansion produces the T^2 and T^4 terms in Eq. (8.23). The expansion assumes $\mu \gg k_B T$ and smoothness of ϕ across the thermal window.

Keeping N fixed in three dimensions gives

$$\mu(T) = \epsilon_F \left[1 - \frac{\pi^2}{12} \left(\frac{k_B T}{\epsilon_F} \right)^2 + \mathcal{O} \left(\left[\frac{k_B T}{\epsilon_F} \right]^4 \right) \right]. \quad (8.24)$$

To impose fixed N , write $D(\epsilon) = C\epsilon^{1/2}$. Applying Eq. (8.23) to the number integral gives

$$N = C \left[\frac{2}{3} \mu^{3/2} + \frac{\pi^2}{12} (k_B T)^2 \mu^{-1/2} + \mathcal{O}((k_B T)^4 \mu^{-5/2}) \right]. \quad (8.25)$$

At zero temperature $N = (2C/3)\epsilon_F^{3/2}$. Put

$$\mu = \epsilon_F (1 + at^2 + \dots), \quad t = \frac{k_B T}{\epsilon_F}.$$

Using $(1 + at^2)^{3/2} = 1 + (3a/2)t^2 + \mathcal{O}(t^4)$ and $(1 + at^2)^{-1/2} = 1 + \mathcal{O}(t^2)$ in the fixed- N equation gives

$$\frac{N}{C\epsilon_F^{3/2}} = \frac{2}{3} + \left(a + \frac{\pi^2}{12} \right) t^2 + \mathcal{O}(t^4).$$

The left side is exactly $2/3$, so $a = -\pi^2/12$, which proves the displayed formula for $\mu(T)$.

For the energy integral, use $\phi(\epsilon) = C\epsilon^{3/2}$ in Eq. (8.23):

$$\begin{aligned} U(T) &= C \left[\frac{2}{5} \mu^{5/2} + \frac{\pi^2}{4} (k_B T)^2 \mu^{1/2} + \mathcal{O}((k_B T)^4 \mu^{-3/2}) \right] \\ &= \frac{2C}{5} \epsilon_F^{5/2} + \frac{\pi^2 C}{6} \epsilon_F^{5/2} t^2 + \mathcal{O}(C \epsilon_F^{5/2} t^4). \end{aligned}$$

In the second line we inserted $\mu/\epsilon_F = 1 - \pi^2 t^2/12 + \mathcal{O}(t^4)$. Since $U_0 = (2C/5)\epsilon_F^{5/2}$ and $N = (2C/3)\epsilon_F^{3/2}$,

$$U(T) = U_0 + \frac{\pi^2}{4} N \frac{(k_B T)^2}{\epsilon_F} + \mathcal{O}\left(N \frac{(k_B T)^4}{\epsilon_F^3}\right), \quad (8.26)$$

$$C_V = \frac{\pi^2}{2} N k_B \frac{k_B T}{\epsilon_F} + \mathcal{O}\left[N k_B \left(\frac{k_B T}{\epsilon_F}\right)^3\right]. \quad (8.27)$$

The heat capacity is the fixed-volume, fixed-particle-number derivative $C_V = (\partial U/\partial T)_{V,N}$; differentiating the T^2 term produces the factor of two in the second line. Only fermions within an energy window of order $k_B T$ around the Fermi surface can be thermally excited. This explains the linear rather than classical constant heat capacity.

Status: controlled approximation

The Sommerfeld series is controlled by $k_B T/\epsilon_F \ll 1$ and smoothness of the density of states near ϵ_F .

Lecture 23 begins here with the saturation of excited Bose states.

8.5 Ideal Bose gas and condensation

Set the ground-state energy to zero and let g denote a finite internal degeneracy common to the one-particle levels. Separate the discrete ground-state contribution from the continuum of excited states. Since one zero-energy mode has occupation $z/(1-z)$, a g -fold ground-state subspace has

$$N = N_0 + N_{\text{ex}}, \quad N_0 = \frac{gz}{1-z}, \quad (8.28)$$

while

$$N_{\text{ex}} = \int_0^\infty d\epsilon \frac{D(\epsilon)}{z^{-1}e^{\beta\epsilon} - 1}. \quad (8.29)$$

Define the thermal wavelength, as in Eq. (4.6), by

$$\lambda_T = \sqrt{\frac{2\pi\hbar^2}{mk_B T}}. \quad (8.30)$$

In three dimensions the excited population can be written

$$N_{\text{ex}} = \frac{gV}{\lambda_T^3} g_{3/2}(z), \quad g_s(z) = \sum_{\ell=1}^{\infty} \frac{z^\ell}{\ell^s}. \quad (8.31)$$

To evaluate the integral, expand the Bose factor:

$$\frac{1}{z^{-1}e^{\beta\epsilon} - 1} = \sum_{\ell=1}^{\infty} z^\ell e^{-\ell\beta\epsilon}. \quad (8.32)$$

The remaining integral is

$$\int_0^\infty \epsilon^{1/2} e^{-\ell\beta\epsilon} d\epsilon = \frac{\sqrt{\pi}}{2} (\ell\beta)^{-3/2}. \quad (8.33)$$

Write the density of states as $D(\epsilon) = C\epsilon^{1/2}$, with C read from Eq. (8.15). Substitution gives

$$\begin{aligned} N_{\text{ex}} &= C \sum_{\ell=1}^{\infty} z^\ell \frac{\sqrt{\pi}}{2} (\ell\beta)^{-3/2} \\ &= \frac{gV}{\lambda_T^3} \sum_{\ell=1}^{\infty} \frac{z^\ell}{\ell^{3/2}} = \frac{gV}{\lambda_T^3} g_{3/2}(z). \end{aligned}$$

The second equality follows by inserting C and $\beta^{-1} = k_B T$; its prefactor is

$$C \frac{\sqrt{\pi}}{2} \beta^{-3/2} = gV \left(\frac{mk_B T}{2\pi\hbar^2} \right)^{3/2} = \frac{gV}{\lambda_T^3}.$$

The condensation threshold is an infrared saturation effect: the continuum of excited modes cannot hold an arbitrarily large density at fixed T . The excited states have a finite maximum capacity as $z \rightarrow 1$:

$$N_{\text{ex}}^{\text{max}} = \frac{gV}{\lambda_T^3} \zeta(3/2). \quad (8.34)$$

When N exceeds this value, the excess occupies the ground-state subspace macroscopically. At the threshold $N = N_{\text{ex}}^{\text{max}}$, so

$$n = g\zeta(3/2) \left(\frac{mk_B T_c}{2\pi\hbar^2} \right)^{3/2}.$$

Raising both sides to the power $2/3$ and solving for T_c gives

$$T_c = \frac{2\pi\hbar^2}{mk_B} \left[\frac{n}{g\zeta(3/2)} \right]^{2/3}. \quad (8.35)$$

For $T < T_c$ in the thermodynamic limit,

$$\frac{N_0}{N} = 1 - \left(\frac{T}{T_c} \right)^{3/2}. \quad (8.36)$$

For $T < T_c$, the thermodynamic-limit chemical potential remains at zero, so $z = 1$. Since $\lambda_T^{-3} \propto T^{3/2}$,

$$\frac{N_{\text{ex}}(T)}{N} = \frac{gV\zeta(3/2)/\lambda_T^3}{gV\zeta(3/2)/\lambda_{T_c}^3} = \left(\frac{T}{T_c} \right)^{3/2}.$$

Using $N_0 = N - N_{\text{ex}}$ yields Eq. (8.36).

A dilute rubidium gas

For a uniform gas of ^{87}Rb atoms in one internal state with density $n = 10^{20} \text{ m}^{-3}$, the ideal-gas formula gives

$$T_c \simeq 3.98 \times 10^{-7} \text{ K} = 398 \text{ nK}. \quad (8.37)$$

This scale is low because the atomic mass is large and the density is dilute. In experiments, trapping, finite particle number, and interactions shift the relation between the observed crossover and the homogeneous ideal value.

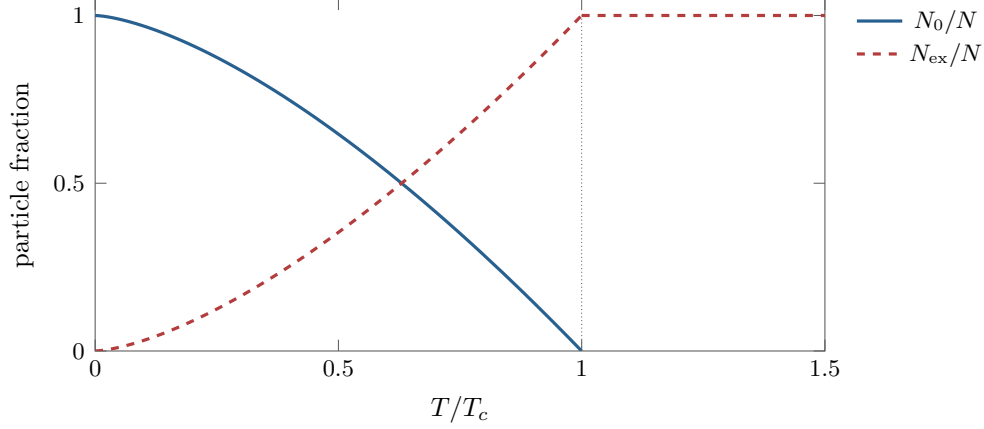


Figure 8.3: Condensate and excited-state fractions for the homogeneous three-dimensional ideal Bose gas. Above T_c the condensate fraction vanishes; below T_c the excited-state population is saturated.

For a dispersion $\epsilon = Ap^s$ in spatial dimension d , the number of momentum states below p scales as p^d . Since $p = (\epsilon/A)^{1/s}$,

$$\mathcal{N}(\epsilon) \propto p^d \propto \epsilon^{d/s}.$$

Differentiating with respect to energy gives the low-energy density of states,

$$D(\epsilon) \propto \epsilon^{d/s-1}. \quad (8.38)$$

The excited-state integral at $\mu = 0$ converges near $\epsilon = 0$ only if

$$\frac{d}{s} > 1. \quad (8.39)$$

Thus a homogeneous nonrelativistic ideal gas ($s = 2$) has no finite-temperature condensate in $d \leq 2$.

The criterion follows from the small- ϵ behavior at $\mu = 0$:

$$\frac{D(\epsilon)}{e^{\beta\epsilon} - 1} \sim \frac{\epsilon^{d/s-1}}{\beta\epsilon} = \frac{1}{\beta} \epsilon^{d/s-2}. \quad (8.40)$$

The integral $\int_0^\infty \epsilon^\alpha d\epsilon$ converges only for $\alpha > -1$, which gives $d/s > 1$. This is a statement about the homogeneous thermodynamic limit; an external trap changes the low-energy density of states and therefore changes the criterion.

Status: thermodynamic-limit statement

Bose–Einstein condensation means macroscopic occupation of one mode, or of a finite degenerate ground-state subspace, in the thermodynamic limit. Finite trapped systems show a crossover, and interactions modify the ideal-gas formulas. The ideal calculation remains the reference point.

Below T_c ,

$$U = \frac{3}{2} k_B T \frac{gV}{\lambda_T^3} \zeta(5/2). \quad (8.41)$$

To derive it, insert an extra factor of ϵ into the excited-state integral and use

$$\int_0^\infty \epsilon^{3/2} e^{-\ell\beta\epsilon} d\epsilon = \Gamma(5/2) (\ell\beta)^{-5/2} = \frac{3\sqrt{\pi}}{4} (\ell\beta)^{-5/2}.$$

Comparing with the number prefactor leaves an extra factor $(3/2)k_B T$ and changes $g_{3/2}(1)$ to $g_{5/2}(1) = \zeta(5/2)$. Since $\lambda_T^{-3} \propto T^{3/2}$, one has $U \propto T^{5/2}$ and therefore $C_V = (5/2)U/T \propto T^{3/2}$ below T_c .

Lecture 24 begins here with blackbody radiation and ends by comparing the three ideal-gas statistics.

8.6 Photon gas and blackbody radiation

Photon number is not a conserved charge in thermal equilibrium with matter: absorption and emission can change it while conserving total energy. In the grand-canonical weight, a chemical potential constrains the mean value of a conserved particle number. No such constraint is imposed for equilibrium photons, so

$$\mu_\gamma = 0. \quad (8.42)$$

Photons have two transverse polarizations and dispersion $\epsilon = \hbar\omega = \hbar ck$. The number of modes with wave number below k is

$$\mathcal{N}(k) = 2 \frac{V}{(2\pi)^3} \frac{4\pi k^3}{3} = \frac{V k^3}{3\pi^2}.$$

Using $k = \omega/c$ and differentiating gives the density of modes per angular-frequency interval,

$$D(\omega) = \frac{V\omega^2}{\pi^2 c^3}. \quad (8.43)$$

Equation (8.10) with the bosonic sign and $\mu_\gamma = 0$ gives the mean occupation $\langle n(\omega) \rangle = 1/(e^{\beta\hbar\omega} - 1)$. The temperature-dependent mean energy per mode is therefore

$$\langle \epsilon(\omega) \rangle = \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1}. \quad (8.44)$$

The zero-point term $\hbar\omega/2$ has been omitted because it is independent of temperature and is not part of the thermal occupation energy considered here. Multiplying the mode density per volume by the mean mode energy gives Planck's law:

$$u(\omega) = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\beta\hbar\omega} - 1}. \quad (8.45)$$

Thus $u(\omega) d\omega$ is the energy per volume carried by modes with angular frequencies in $[\omega, \omega + d\omega]$. With $x = \beta\hbar\omega$,

$$\frac{U}{V} = \frac{(k_B T)^4}{\pi^2 \hbar^3 c^3} \int_0^\infty \frac{x^3}{e^x - 1} dx. \quad (8.46)$$

Expanding $(e^x - 1)^{-1} = \sum_{\ell \geq 1} e^{-\ell x}$ and integrating term by term gives

$$\int_0^\infty \frac{x^3}{e^x - 1} dx = 6 \sum_{\ell=1}^\infty \frac{1}{\ell^4} = 6\zeta(4) = \frac{\pi^4}{15}. \quad (8.47)$$

Using the evaluated integral gives

$$\frac{U}{V} = aT^4, \quad a = \frac{\pi^2 k_B^4}{15 \hbar^3 c^3}. \quad (8.48)$$

The pressure and entropy do not require a separate kinetic argument. From the bosonic grand partition function at $z = 1$,

$$\ln \Xi = - \int_0^\infty \frac{V\omega^2 d\omega}{\pi^2 c^3} \ln(1 - e^{-\beta\hbar\omega}).$$

The substitution $x = \beta\hbar\omega$ shows that $\ln \Xi = V\beta^{-3}C_\gamma$, where C_γ is independent of β and V . Therefore, at fixed volume,

$$U = -\frac{\partial \ln \Xi}{\partial \beta} = 3V\beta^{-4}C_\gamma = \frac{3}{\beta} \ln \Xi,$$

$$PV = k_B T \ln \Xi = \frac{1}{\beta} \ln \Xi.$$

Eliminating $\ln \Xi$ gives $U = 3PV$. With $U = aVT^4$ and $\mu_\gamma = 0$, the Euler relation, Eq. (1.33), reads $TS = U + PV$. Hence

$$P = \frac{1}{3} \frac{U}{V}, \quad S = \frac{4}{3} aVT^3. \quad (8.49)$$

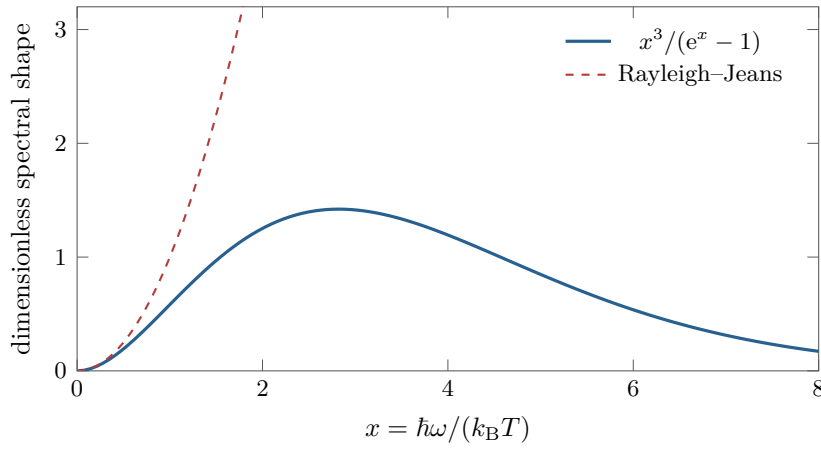


Figure 8.4: Planck's spectral shape and its low-frequency Rayleigh–Jeans approximation. The dotted continuation shows why the classical approximation cannot be integrated over all frequencies.

Thermal radiation at 300 K

The radiation constant is $a \simeq 7.57 \times 10^{-16} \text{ J m}^{-3} \text{ K}^{-4}$. At room temperature,

$$\frac{U}{V} \simeq 6.13 \times 10^{-6} \text{ J m}^{-3}, \quad P \simeq 2.04 \times 10^{-6} \text{ Pa}. \quad (8.50)$$

The equilibrium photon pressure is tiny compared with atmospheric pressure, but the T^4 law makes radiation dominant at sufficiently high temperature.

For $\beta\hbar\omega \ll 1$, $e^{\beta\hbar\omega} - 1 \simeq \beta\hbar\omega$, so the mean mode energy approaches the classical equipartition value,

$$\langle \epsilon(\omega) \rangle \simeq k_B T, \quad u_{\text{RJ}}(\omega) \simeq \frac{k_B T \omega^2}{\pi^2 c^3}.$$

This is the Rayleigh–Jeans law. Extending it to arbitrarily high frequency makes the total energy proportional to $\int_0^\infty \omega^2 d\omega$, which diverges. The full Bose occupation supplies the exponential high-frequency suppression and removes the ultraviolet catastrophe.

The final part of Lecture 24 compares Bose, Fermi, and Maxwell–Boltzmann statistics in the dilute regime.

8.7 Classical limit and exchange corrections

The classical dilute regime has small mean occupation in every mode. A sufficient condition is $z \ll 1$, for which

$$\langle n_r \rangle \simeq z e^{-\beta \epsilon_r}, \quad (8.51)$$

and both quantum distributions reduce to Maxwell–Boltzmann statistics. For a spinless gas,

$$\frac{P}{n k_B T} = 1 \mp \frac{n \lambda_T^3}{2^{5/2}} + \dots, \quad (8.52)$$

with the upper minus for bosons and lower plus for fermions. Exchange acts like an effective attraction for bosons and repulsion for fermions at this order, even without a mechanical pair potential.

The size of the leading quantum correction is set by the phase-space density $n \lambda_T^3$. To derive it, use the series

$$-\ln(1-x) = x + \frac{x^2}{2} + \dots, \quad \ln(1+x) = x - \frac{x^2}{2} + \dots$$

in the bosonic and fermionic grand partition functions. After integrating each Boltzmann factor over momentum,

$$\int \frac{d^3 p}{(2\pi\hbar)^3} e^{-\ell \beta p^2/(2m)} = \frac{1}{\lambda_T^3 \ell^{3/2}}. \quad (8.53)$$

The term of order z^ℓ carries an additional factor $1/\ell$ from the logarithm, so its total power is $\ell^{-5/2}$. Keeping $\ell = 1, 2$ gives

$$\frac{\ln \Xi}{V} = \frac{1}{\lambda_T^3} \left(z \pm \frac{z^2}{2^{5/2}} + \mathcal{O}(z^3) \right), \quad (8.54)$$

where the upper sign is for bosons. Equivalently, the bosonic series is the expansion of $g_{5/2}(z)$ used above. For fermions, the alternating logarithm produces the same powers with alternating signs. The density follows from $N = z \partial_z \ln \Xi$:

$$n = z \frac{\partial}{\partial z} \frac{\ln \Xi}{V} = \frac{1}{\lambda_T^3} \left(z \pm \frac{z^2}{2^{3/2}} + \mathcal{O}(z^3) \right). \quad (8.55)$$

Set $y = n \lambda_T^3$. Multiplying the last equation by λ_T^3 gives

$$y = z \pm \frac{z^2}{2^{3/2}} + \mathcal{O}(z^3).$$

Seek the inverse in the form $z = y + c y^2 + \mathcal{O}(y^3)$. Substitution gives $c = \mp 2^{-3/2}$, so

$$z = y \mp \frac{y^2}{2^{3/2}} + \mathcal{O}(y^3).$$

Now use $P = k_B T \ln \Xi / V$:

$$\begin{aligned} \frac{P \lambda_T^3}{k_B T} &= z \pm \frac{z^2}{2^{5/2}} + \mathcal{O}(z^3) \\ &= y \mp \frac{y^2}{2^{3/2}} \pm \frac{y^2}{2^{5/2}} + \mathcal{O}(y^3) \\ &= y \mp \frac{y^2}{2^{5/2}} + \mathcal{O}(y^3). \end{aligned}$$

Dividing by $y = n \lambda_T^3$ proves Eq. (8.52). Exercise 8.8 asks you to rederive the fugacity inversion and interpret the Bose and Fermi signs. The calculation is controlled by $n \lambda_T^3 \ll 1$, the same parameter that marked the classical limit in [chapter 4](#).

8.8 Exercises

Exercise 8.1. Derive the factorized grand partition functions and the distribution (8.10). Calculate single-mode fluctuations.

Exercise 8.2. Derive the density of states (8.15), including the degeneracy factor g . Generalize to dimension d and dispersion $\epsilon = Ap^s$.

Exercise 8.3. Calculate k_F , ϵ_F , U_0 , and P_0 for a three-dimensional spin-1/2 Fermi gas.

Exercise 8.4 (*). Use the Sommerfeld expansion to derive the chemical potential and heat capacity in (8.27).

Exercise 8.5. Derive the condensation criterion (8.39). Discuss nonrelativistic gases in dimensions one, two, and three.

Exercise 8.6. Derive T_c and the condensate fraction (8.36). Show that the ground-state contribution to pressure vanishes in the thermodynamic limit.

Exercise 8.7. Derive Planck's law and the Stefan–Boltzmann energy density. Use thermodynamics to obtain $P = U/(3V)$ and the entropy.

Exercise 8.8 (*). Expand the Bose and Fermi grand potentials at small fugacity and derive the exchange virial correction (8.52).

8.9 Further reading

Pathria gives a detailed account of ideal quantum gases and condensation [42]. Kardar emphasizes the occupation-number derivation and classical limit [23]; Reichl supplies additional applications [47].

Part IV

Collective behavior and scale

Chapter 9

Phase transitions, order parameters, and scaling

Lecture 25. Thermodynamic singularities, coexistence, Landau theory, correlations, and critical exponents.

What you should be able to do

- Explain why genuine phase-transition singularities require a thermodynamic limit.
- Distinguish first-order coexistence from a continuous critical point.
- Derive mean-field exponents from a Landau free energy and check which stationary points are stable.
- Connect susceptibility to integrated connected correlations and the correlation length.
- Extract critical exponents from a homogeneous singular free energy by explicit differentiation.

9.1 Nonanalyticity requires infinitely many degrees of freedom

What does a thermodynamic singularity mean? For example, a discontinuity in entropy is a kink in the free energy, and a divergent heat capacity is a divergent second derivative. Such behavior cannot occur in an ordinary finite system. For a finite classical system, suppose the Boltzmann integrand is dominated by an integrable function throughout an open complex neighborhood of the inverse temperature under study. The parameter integral then defines an analytic $Z(\beta)$. For a finite-dimensional quantum system, Z is a finite sum of exponentials of energy levels and is analytic. At real positive temperature $Z > 0$, so $\ln Z$ and its finite derivatives are also analytic. Sharp phase transitions can therefore arise only after a limit such as

$$N, V \rightarrow \infty, \quad \frac{N}{V} = n. \quad (9.1)$$

has been taken. Here $n = N/V$ is held fixed. Other thermodynamic limits are possible, but some measure of system size must diverge.

The order of operations matters. One first calculates the finite- N free energy and then takes $N \rightarrow \infty$. Differentiating the limiting function need not give the limit of the finite- N derivatives at a point where convergence is nonuniform. That failure of uniform convergence is how a sequence of analytic functions can approach a nonanalytic thermodynamic potential.

At finite N , one observes rounded crossovers. The rounding narrows as the system grows, and finite-size scaling can recover critical properties from the way the finite system approaches the thermodynamic limit.

Status: thermodynamic-limit statement

A phase transition is a nonanalyticity of a thermodynamic potential in the thermodynamic limit. A peak or rapid change in a finite system may signal the nearby limit, but it is not itself a mathematical singularity.

Worked example: how a kink emerges from analytic partition functions

Suppose two families of macrostates have free-energy densities $f_1(T)$ and $f_2(T)$. A finite- N approximation to the partition function is

$$Z_N = e^{-\beta N f_1} + e^{-\beta N f_2}. \quad (9.2)$$

The finite-size free-energy density is

$$f_N = -\frac{k_B T}{N} \ln Z_N = \min(f_1, f_2) - \frac{k_B T}{N} \ln [1 + e^{-\beta N |f_1 - f_2|}]. \quad (9.3)$$

It is smooth for every finite N . As $N \rightarrow \infty$, the second term vanishes and $f = \min(f_1, f_2)$. If the two branches cross with different slopes, the minimum has a kink and the entropy density $-\partial f / \partial T$ jumps. The rounding occurs in a temperature window where $N|f_1 - f_2| \lesssim k_B T$, hence typically shrinks as $1/N$.

9.2 First-order transitions and coexistence

At a first-order transition, a first derivative of a thermodynamic potential jumps. For fixed particle number, the Gibbs free energy obeys

$$dG = -S dT + V dP + \mu dN. \quad (9.4)$$

A jump in entropy between the two equilibrium phases produces latent heat. In the molar convention used below,

$$L = T \Delta s. \quad (9.5)$$

A jump in V produces a density discontinuity.

Along a coexistence curve, the molar Gibbs free energies of phases a and b are equal. Write $\Delta s = s_b - s_a$ and $\Delta v = v_b - v_a$. Differentiating $g_a(T, P) = g_b(T, P)$ while moving along the coexistence curve gives

$$-s_a dT + v_a dP = -s_b dT + v_b dP. \quad (9.6)$$

Moving all terms belonging to the same phase to one side gives

$$\frac{dP}{dT} = \frac{\Delta s}{\Delta v} = \frac{L}{T \Delta v}. \quad (9.7)$$

This is the Clapeyron equation. Its sign has a direct interpretation: if phase b has both larger entropy and larger volume, the coexistence pressure rises with temperature.

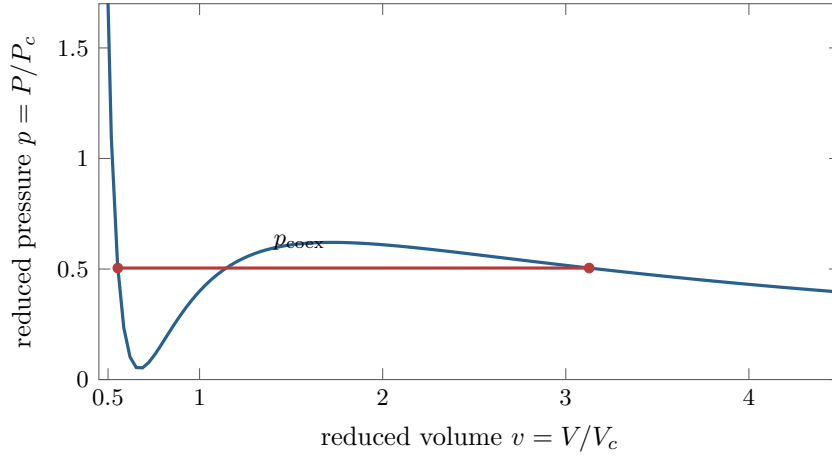


Figure 9.1: Reduced van der Waals isotherm at $T/T_c = 0.85$. The horizontal line has equal signed areas above and below the loop and joins the coexisting liquid and gas volumes.

9.2.1 van der Waals loop

Below its critical temperature, the van der Waals isotherm contains a region with $\partial P/\partial V > 0$. The isothermal compressibility is $\kappa_T = -(1/V)(\partial V/\partial P)_T$, so this region has $\kappa_T < 0$ and is unstable. The equilibrium isotherm replaces the loop by a constant-pressure coexistence segment. Equality of Gibbs free energies produces the Maxwell equal-area construction

$$\int_{V_l}^{V_g} [P(V, T) - P_{\text{coex}}] dV = 0. \quad (9.8)$$

Geometrically, this construction replaces the nonconvex part of the candidate free energy by its convex envelope.

To derive the area rule, use $P = -(\partial F/\partial V)_T$. Equality of the two Gibbs free energies at the same pressure means

$$0 = G(V_g) - G(V_l) \quad (9.9)$$

$$= F(V_g) - F(V_l) + P_{\text{coex}}(V_g - V_l) \quad (9.10)$$

$$= - \int_{V_l}^{V_g} P(V, T) dV + P_{\text{coex}}(V_g - V_l). \quad (9.11)$$

This is exactly the stated equal-area condition. The unstable part of the loop is not traversed by an equilibrium homogeneous phase; it is replaced by a mixture whose phase fraction changes at fixed pressure.

9.3 Continuous transitions and order parameters

A continuous transition has no latent heat, but response functions and correlation length can diverge. An order parameter vanishes in one phase and takes a nonzero value in another. Examples include magnetization for a ferromagnet, density difference for a fluid, and a complex condensate amplitude for a superfluid.

The order parameter should reflect a symmetry. For an Ising ferromagnet at zero external field, spin reversal gives

$$m \mapsto -m. \quad (9.12)$$

The disordered phase respects the symmetry with $m = 0$; the ordered phase chooses one of two symmetry-related values.

In a finite system at zero field with symmetry-preserving boundary conditions, pairing every configuration with its spin reverse forces $\langle m \rangle = 0$. Spontaneous symmetry breaking requires the order of limits

$$m_\star = \lim_{h \rightarrow 0^+} \lim_{N \rightarrow \infty} \langle m \rangle. \quad (9.13)$$

Reversing the limits gives zero.

9.4 Landau theory

For a scalar order parameter with $m \mapsto -m$ symmetry, odd powers of m are absent at $h = 0$. Assuming the free energy is analytic in a small, spatially uniform m , keep the lowest powers needed for stability:

$$f(m; T, h) = f_0(T) + \frac{a}{2}tm^2 + \frac{b}{4}m^4 - hm + \dots, \quad (9.14)$$

where $t = (T - T_c)/T_c$, $a > 0$, and $b > 0$. The function $f_0(T)$ is a smooth background independent of m . The linear term $-hm$ describes the field conjugate to m . Equilibrium minimizes f , so a stationary point satisfies

$$atm + bm^3 - h = 0. \quad (9.15)$$

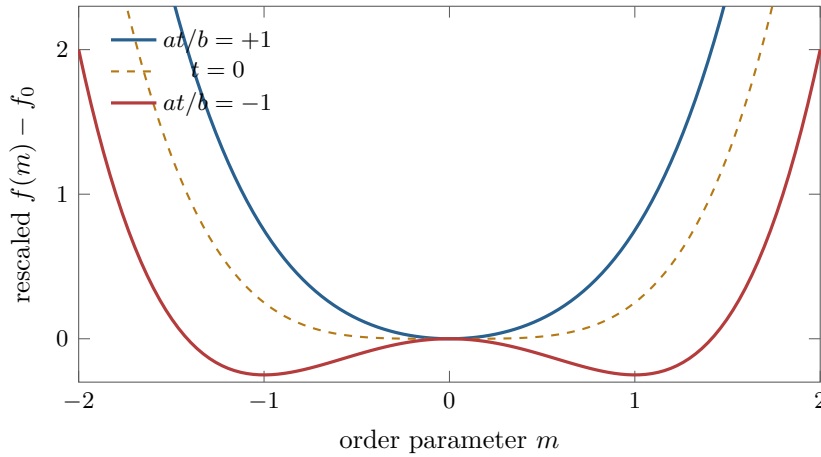


Figure 9.2: Landau free energy at zero field after rescaling. The single minimum at $m = 0$ splits continuously into two symmetry-related minima as t changes sign.

At $h = 0$, factorizing the stationarity equation gives

$$m = 0 \quad \text{or} \quad m^2 = -\frac{a}{b}t. \quad (9.16)$$

Stationarity alone does not identify a minimum. The curvature is

$$\frac{\partial^2 f}{\partial m^2} = at + 3bm^2. \quad (9.17)$$

For $t > 0$, $m = 0$ has positive curvature and is the only minimum. For $t < 0$, $m = 0$ has negative curvature and is a local maximum, while the two nonzero stationary points have curvature $-2at > 0$. Thus

$$m_\star = \pm \sqrt{\frac{-at}{b}}. \quad (9.18)$$

Thus the order-parameter exponent is

$$m_\star \sim (-t)^\beta, \quad \beta_{\text{MF}} = \frac{1}{2}. \quad (9.19)$$

For $t > 0$ and small h , the order parameter is small. Dropping the cubic term in (9.15) gives $m \simeq h/(at)$, so

$$\chi = \left(\frac{\partial m}{\partial h} \right)_T \sim t^{-\gamma}, \quad \gamma_{\text{MF}} = 1. \quad (9.20)$$

At $t = 0$, the linear and cubic terms balance exactly within the quartic Landau theory:

$$h = bm^3, \quad \delta_{\text{MF}} = 3. \quad (9.21)$$

Substitution of $m_\star^2 = -at/b$ gives

$$f(m_\star) - f_0 = \frac{at}{2} \left(-\frac{at}{b} \right) + \frac{b}{4} \left(\frac{a^2 t^2}{b^2} \right) \quad (9.22)$$

$$= -\frac{a^2 t^2}{4b} \quad (t < 0), \quad (9.23)$$

while the minimizing value of this Landau contribution is zero for $t > 0$. Since $t = (T - T_c)/T_c$, two derivatives with respect to T give a finite constant below T_c and zero above it. The heat capacity therefore has a finite jump rather than a power-law divergence, giving $\alpha_{\text{MF}} = 0$.

Here C_h is the heat capacity at fixed conjugate field, ξ is the correlation length, $G(r; 0, 0)$ is the connected two-point correlation at $t = h = 0$, d is the spatial dimension, and η measures its anomalous power. The exponent notation is defined by the limiting paths

$$C_h \sim |t|^{-\alpha}, \quad m(t, 0) \sim (-t)^\beta \quad (t < 0), \quad (9.24)$$

$$\chi(t, 0) \sim |t|^{-\gamma}, \quad m(0, h) \sim |h|^{1/\delta} \text{sgn } h, \quad (9.25)$$

$$\xi(t, 0) \sim |t|^{-\nu}, \quad G(r; 0, 0) \sim r^{-(d-2+\eta)}. \quad (9.26)$$

The symbol β here denotes a critical exponent, not inverse temperature. Context normally distinguishes them, but in calculations near criticality it is safer to write inverse temperature explicitly as $1/(k_B T)$.

9.4.1 First-order behavior from the same expansion

Let the quartic coefficient be negative and stabilize the theory with a sixth-order term:

$$f(m) = f_0 + \frac{r}{2}m^2 - \frac{u}{4}m^4 + \frac{v}{6}m^6, \quad u, v > 0. \quad (9.27)$$

At a nonzero stationary point, $r - um^2 + vm^4 = 0$. Coexistence with the $m = 0$ minimum also requires $f(m) = f(0)$. Put $x = m^2 > 0$. Dividing the coexistence equation by x gives a second equation,

$$r - ux + vx^2 = 0, \quad (9.28)$$

$$\frac{r}{2} - \frac{u}{4}x + \frac{v}{6}x^2 = 0. \quad (9.29)$$

Twice the second equation is $r - ux/2 + vx^2/3 = 0$. Subtracting it from the first gives $-ux/2 + 2vx^2/3 = 0$, so $x = 3u/(4v)$. Substitution into the stationary equation gives $r = ux - vx^2 = 3u^2/(16v)$. Hence

$$m_{\text{coex}}^2 = \frac{3u}{4v}, \quad r_{\text{coex}} = \frac{3u^2}{16v}. \quad (9.30)$$

The curvature check confirms that the degenerate stationary points are minima. At $m = 0$, $f''(0) = r_{\text{coex}} > 0$. At the nonzero solution $x = m^2 = 3u/(4v)$, use $r = ux - vx^2$ to obtain

$$f''(m) = r - 3ux + 5vx^2 = -2ux + 4vx^2 = ux > 0. \quad (9.31)$$

The order parameter therefore jumps while the two minima have equal free energy. Between coexistence and a spinodal, one minimum is metastable: it is locally stable but not the global minimum. Nucleation, which requires a spatially varying droplet and an interfacial free-energy cost, determines how the metastable phase decays. The uniform polynomial locates the competing extrema but does not calculate the nucleation rate.

Status: mean field

Landau theory assumes that a smooth local free-energy expansion and a spatially uniform order parameter capture the transition. It gives exact exponents only at or above the upper critical dimension, apart from logarithmic corrections at the boundary. Below it, long-wavelength fluctuations change the exponents.

9.5 Spatial fluctuations and correlation length

Promote m to a field and include the lowest gradient penalty:

$$\mathcal{F}[m] = \int d^d x \left[\frac{c}{2} (\nabla m)^2 + \frac{a}{2} t m^2 + \frac{b}{4} m^4 - h m \right]. \quad (9.32)$$

In the disordered phase $t > 0$, fluctuations are centered on $m = 0$. Use the Fourier convention $m(\mathbf{x}) = \int d^d k e^{i\mathbf{k} \cdot \mathbf{x}} m_{\mathbf{k}} / (2\pi)^d$. Neglecting the quartic interaction gives the Gaussian functional

$$\mathcal{F}_G[m] = \frac{1}{2} \int \frac{d^d k}{(2\pi)^d} (ck^2 + at) |m_{\mathbf{k}}|^2, \quad (9.33)$$

where $m_{\mathbf{k}}$ is the Fourier amplitude of the order-parameter field. Each quadratic mode has a thermal variance inversely proportional to its stiffness, hence

$$\langle |m_{\mathbf{k}}|^2 \rangle \propto \frac{1}{ck^2 + at}. \quad (9.34)$$

Factoring the denominator as $c(k^2 + \xi^{-2})$ defines the mean-field correlation length:

$$\xi = \sqrt{\frac{c}{at}} \sim t^{-1/2}, \quad \nu_{\text{MF}} = \frac{1}{2}. \quad (9.35)$$

In $d = 3$, the Fourier transform can be done explicitly:

$$G(r) = k_B T \int \frac{d^3 k}{(2\pi)^3} \frac{e^{i\mathbf{k} \cdot \mathbf{r}}}{ck^2 + at} = \frac{k_B T}{4\pi c r} e^{-r/\xi}. \quad (9.36)$$

The Fourier integral is evaluated in Appendix A, Eq. (A.42). The exponential gives the decay scale, while the factor $1/r$ is the three-dimensional Gaussian power law. At criticality $\xi \rightarrow \infty$ and the exponential cutoff disappears.

9.5.1 When is mean field self-consistent?

Mean field replaces the order parameter by its minimizing value. This can work only if fluctuations averaged over a correlation volume are small compared with $m_\star^2$. Modes with wavelength at least ξ have $k \lesssim \xi^{-1}$, so their contribution scales as

$$\langle (\delta m)^2 \rangle_\xi \sim k_B T \int^{\xi^{-1}} \frac{d^d k}{(2\pi)^d} \frac{1}{ck^2 + c\xi^{-2}} \quad (9.37)$$

$$\sim \frac{k_B T}{c} \frac{\xi^{-d}}{\xi^{-2}} = \frac{k_B T}{c} \xi^{2-d}. \quad (9.38)$$

Below T_c , the squared mean-field order parameter scales as

$$m_\star^2 = \frac{a|t|}{b} \sim \frac{c}{b\xi^2}, \quad (9.39)$$

where the second relation uses the scaling $\xi^2 \sim c/(a|t|)$. Numerical factors depend on whether ξ is evaluated on the symmetric or broken side of the transition and do not affect the Ginzburg power count. Their ratio is therefore

$$\frac{\langle (\delta m)^2 \rangle_\xi}{m_\star^2} \sim \frac{k_B T b}{c^2} \xi^{4-d}. \quad (9.40)$$

As $\xi \rightarrow \infty$, fluctuations overwhelm the mean for $d < 4$, remain marginal at $d = 4$, and become relatively small for $d > 4$. This Ginzburg estimate identifies $d_c = 4$ as the upper critical dimension of the scalar ϕ^4 theory. It is a scaling argument, not an exact value for the width of a material's non-mean-field region.

Away from mean field, define the connected correlation function by subtracting the disconnected product:

$$G(\mathbf{r}) = \langle m(\mathbf{r})m(0) \rangle - \langle m(\mathbf{r}) \rangle \langle m(0) \rangle. \quad (9.41)$$

Near criticality its scaling form is

$$G(r) = \langle m(\mathbf{r})m(0) \rangle_c \sim \frac{1}{r^{d-2+\eta}} \mathcal{G}(r/\xi). \quad (9.42)$$

Let $M = \int d^d x m(\mathbf{x})$ be coupled linearly to a uniform field through $-hM$. Differentiating $\langle M \rangle$ with respect to h gives the static fluctuation–response identity $\partial_h \langle M \rangle = \beta_T \text{Var}(M)$, where $\beta_T = 1/(k_B T)$ in this sentence. Translation invariance then gives

$$\frac{1}{V} \text{Var}(M) = \frac{1}{V} \int d^d x \int d^d y G(\mathbf{x} - \mathbf{y}) \quad (9.43)$$

$$= \int d^d r G(r). \quad (9.44)$$

Thus, apart from the nonsingular factor β_T , susceptibility is the integrated connected correlation:

$$\chi \propto \int d^d r G(r). \quad (9.45)$$

If the scaling function cuts off the integral at $r \sim \xi$, using (9.42) gives

$$\chi \propto \int_0^\infty dr r^{d-1} r^{-(d-2+\eta)} \mathcal{G}(r/\xi) \quad (9.46)$$

$$= \xi^{2-\eta} \int_0^\infty du u^{1-\eta} \mathcal{G}(u), \quad u = r/\xi. \quad (9.47)$$

The remaining integral is independent of ξ , so

$$\chi \sim \xi^{2-\eta}, \quad \gamma = \nu(2 - \eta). \quad (9.48)$$

9.6 Scaling hypothesis

Separate the free-energy density into a regular part, which remains analytic at the transition, and a singular part f_s . Let t and h be the two relevant scaling fields near a simple critical point. In this section $b > 1$ denotes a change of length scale; it is unrelated to the quartic coefficient b in (9.14). The scaling hypothesis assumes

$$f_s(t, h) = b^{-d} f_s(b^{y_t} t, b^{y_h} h). \quad (9.49)$$

Choosing $b = |t|^{-1/y_t}$ expresses all singular behavior as powers of t . The renormalization group will explain the origin of this homogeneity and identify y_t and y_h as eigenvalues of the flow near a fixed point.

Differentiate (9.49) with respect to h while holding t fixed:

$$m(t, h) = -\frac{\partial f_s(t, h)}{\partial h} \quad (9.50)$$

$$= -b^{-d} b^{y_h} \frac{\partial f_s(b^{y_t} t, b^{y_h} h)}{\partial (b^{y_h} h)} \quad (9.51)$$

$$= b^{y_h-d} m(b^{y_t} t, b^{y_h} h). \quad (9.52)$$

At $h = 0$, choose $b = |t|^{-1/y_t}$. The rescaled temperature is then $b^{y_t} t = \text{sgn } t$, a number of order one. Provided $m(-1, 0)$ is finite and nonzero on the ordered side, all dependence on $|t|$ sits in the prefactor:

$$m(t, 0) \sim |t|^{(d-y_h)/y_t}, \quad \beta = \frac{d - y_h}{y_t}. \quad (9.53)$$

Thus the homogeneity law fixes $\beta = (d - y_h)/y_t$. The other exponents come from the same law. Since ξ rescales as a length,

$$\nu = \frac{1}{y_t}. \quad (9.54)$$

For the heat capacity, two t derivatives contribute b^{2y_t} , so $\partial_t^2 f_s \sim |t|^{(d-2y_t)/y_t}$. For the susceptibility, two h derivatives contribute b^{2y_h} . Thus

$$\alpha = 2 - \frac{d}{y_t} = 2 - d\nu, \quad (9.55)$$

$$\gamma = \frac{2y_h - d}{y_t}. \quad (9.56)$$

To obtain the critical-isotherm exponent, set $t = 0$ and choose $b = |h|^{-1/y_h}$ in (9.52). Then

$$m(0, h) \sim |h|^{(d-y_h)/y_h} \text{sgn } h, \quad \delta = \frac{y_h}{d - y_h}. \quad (9.57)$$

Exercise 9.6 asks for these differentiations in a form suitable for checking the signs and powers independently. Substitution makes the exponent relations transparent:

$$\alpha + 2\beta + \gamma = 2 - \frac{d}{y_t} + 2\frac{d - y_h}{y_t} + \frac{2y_h - d}{y_t} = 2, \quad (9.58)$$

$$\frac{\gamma}{\beta} = \frac{2y_h - d}{d - y_h} = \delta - 1. \quad (9.59)$$

Therefore

$$\alpha + 2\beta + \gamma = 2, \quad \gamma = \beta(\delta - 1). \quad (9.60)$$

Hyperscaling, such as $2 - \alpha = d\nu$, can fail above the upper critical dimension because an irrelevant coupling can still appear in a denominator or amplitude of a singular observable. Such a coupling flows toward zero but cannot simply be set to zero; it is called *dangerously irrelevant*. Scaling relations should therefore be quoted together with their assumptions.

9.7 Exercises

Exercise 9.1. Derive the Clapeyron equation (9.7). Apply it to a transition whose denser phase has lower entropy and determine the sign of the coexistence-curve slope.

Exercise 9.2. Find the critical point of the van der Waals equation by imposing $(\partial P/\partial V)_T = (\partial^2 P/\partial V^2)_T = 0$.

Exercise 9.3. Minimize the Landau free energy and derive the mean-field exponents β , γ , δ , and α .

Exercise 9.4. Add a positive m^6 term and let the quartic coefficient become negative. Draw the resulting free-energy shapes and locate the first-order transition.

Exercise 9.5. Derive the Gaussian correlation function in momentum space from (9.32) and identify the correlation length.

Exercise 9.6 (*). Starting from (9.49), derive two scaling relations among α , β , γ , δ , ν , and η .

9.8 Further reading

Kardar develops Landau theory and scaling with a field-theory viewpoint [22]. Sethna gives a physical account of order parameters and universality [53]. Reichl discusses coexistence and thermodynamic classification [47].

Chapter 10

The Ising model

Lecture 26–27. Transfer matrices, domain walls, mean field, the two-dimensional critical point, Monte Carlo, and finite-size scaling.

What you should be able to do

- Solve the one-dimensional Ising model exactly with a transfer matrix.
- Reproduce the same one-dimensional physics by counting domain walls.
- Derive the mean-field self-consistency equation and identify its failure in low dimension.
- Verify detailed balance for Metropolis sampling and account for correlated measurements.
- Distinguish quoted two-dimensional exact results from results derived here.

10.1 Definition and symmetry

On a lattice with N sites, place a two-valued spin $s_i = \pm 1$ at each site. The Ising Hamiltonian is

$$H = -J \sum_{\langle ij \rangle} s_i s_j - h \sum_i s_i, \quad (10.1)$$

where $\langle ij \rangle$ denotes a nearest-neighbor bond, $J > 0$ favors parallel spins, and h is an external field. At $h = 0$, the model has a global $\mathbb{Z}_2$ symmetry $s_i \mapsto -s_i$. Thermal averages use the Gibbs probability $p(s) = e^{-\beta H(s)}/Z$. The magnetization per site is

$$m = \frac{1}{N} \sum_i \langle s_i \rangle. \quad (10.2)$$

The model strips a ferromagnet down to the competition that matters: energy favors alignment; entropy favors the many disordered configurations. It also describes the universality class of other systems with a single scalar order parameter and short-range interactions.

10.2 One-dimensional transfer matrix

Take a periodic chain of N spins, so $s_{N+1} = s_1$. Define dimensionless couplings

$$K = \beta J, \quad \eta = \beta h. \quad (10.3)$$

The Boltzmann exponent is

$$-\beta H = \sum_{i=1}^N \left[K s_i s_{i+1} + \frac{\eta}{2} (s_i + s_{i+1}) \right]. \quad (10.4)$$

Each field contribution appears in two neighboring bond terms, each time with a factor one half. Define the Boltzmann weight of one bond as the matrix element

$$T_{ss'} = \exp \left[K s s' + \frac{\eta}{2} (s + s') \right]. \quad (10.5)$$

In the basis $s = +1, -1$,

$$T = \begin{pmatrix} e^{K+\eta} & e^{-K} \\ e^{-K} & e^{K-\eta} \end{pmatrix}. \quad (10.6)$$

Each sum over an intermediate spin is a matrix multiplication. Periodicity closes the last index, so

$$Z_N = \sum_{s_1, \dots, s_N} T_{s_1 s_2} T_{s_2 s_3} \cdots T_{s_N s_1} \quad (10.7)$$

$$= \sum_{s_1} (T^N)_{s_1 s_1} = \text{Tr } T^N. \quad (10.8)$$

The eigenvalues follow from the characteristic polynomial. The trace and determinant are

$$\text{Tr } T = 2e^K \cosh \eta, \quad (10.9)$$

$$\det T = e^{2K} - e^{-2K}. \quad (10.10)$$

Therefore

$$\lambda_{\pm} = \frac{\text{Tr } T}{2} \pm \sqrt{\left(\frac{\text{Tr } T}{2} \right)^2 - \det T} \quad (10.11)$$

$$= e^K \cosh \eta \pm \sqrt{e^{2K} \sinh^2 \eta + e^{-2K}}. \quad (10.12)$$

Since the trace of a power is the sum of the powered eigenvalues,

$$Z_N = \text{Tr } T^N = \lambda_+^N + \lambda_-^N, \quad (10.13)$$

and Exercise 10.1 checks this algebra directly. For $T > 0$, every entry of T is positive. The explicit eigenvalues then satisfy $\lambda_+ > |\lambda_-|$, so $N^{-1} \ln Z_N \rightarrow \ln \lambda_+$ and the thermodynamic-limit free energy per spin is

$$f = -k_B T \ln \lambda_+. \quad (10.14)$$

At zero field,

$$\lambda_+ = 2 \cosh K, \quad \lambda_- = 2 \sinh K, \quad (10.15)$$

and

$$f = -k_B T \ln(2 \cosh K). \quad (10.16)$$

This is analytic for every $T > 0$. Hence the one-dimensional short-range Ising model has no finite-temperature phase transition.

At nonzero field, write

$$\lambda_+ = e^K (\cosh \eta + R), \quad R = \sqrt{\sinh^2 \eta + e^{-4K}}. \quad (10.17)$$

Since $\partial_\eta R = \sinh \eta \cosh \eta / R$,

$$\frac{\partial \ln \lambda_+}{\partial \eta} = \frac{\sinh \eta + (\sinh \eta \cosh \eta) / R}{\cosh \eta + R} = \frac{\sinh \eta}{R}. \quad (10.18)$$

Why is this derivative the magnetization? From $\eta = \beta h$,

$$\frac{1}{N} \frac{\partial \ln Z_N}{\partial \eta} = \frac{1}{N} \left\langle \sum_i s_i \right\rangle = m. \quad (10.19)$$

In the thermodynamic limit only λ_+ contributes to $N^{-1} \ln Z_N$, so the derivative above is the exact magnetization:

$$m = \frac{\sinh \eta}{\sqrt{\sinh^2 \eta + e^{-4K}}}. \quad (10.20)$$

For any $T > 0$, taking $\eta \rightarrow 0$ gives $m = 0$. The zero-field susceptibility per spin is nevertheless large at low temperature:

$$m = \frac{\eta + \mathcal{O}(\eta^3)}{\sqrt{\eta^2 + e^{-4K} + \mathcal{O}(\eta^4)}} = \eta e^{2K} + \mathcal{O}(\eta^3). \quad (10.21)$$

Since $\eta = \beta h$, this gives

$$\chi = \left(\frac{\partial m}{\partial h} \right)_{h=0} = \beta e^{2K}. \quad (10.22)$$

The chain forms long aligned domains and responds strongly to a tiny field, yet it has no spontaneous magnetization. A large response is not the same as an ordered phase.

Status: model-specific exact result

The transfer-matrix solution is exact for the one-dimensional nearest-neighbor model. The absence of a transition is not a statement that no one-dimensional model can ever order; long-range interactions or additional constraints can change the conclusion.

10.2.1 Correlations

At $h = 0$, the normalized eigenvectors are

$$|+\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \quad |-\rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad (10.23)$$

with eigenvalues $\lambda_+ = 2 \cosh K$ and $\lambda_- = 2 \sinh K$. The spin matrix $S = \text{diag}(1, -1)$ interchanges them: $S|+\rangle = |-\rangle$ and $S|-\rangle = |+\rangle$. Inserting two spin operators into the transfer-matrix trace gives

$$\langle s_0 s_r \rangle = \frac{\text{Tr}(S T^r S T^{N-r})}{\text{Tr } T^N}. \quad (10.24)$$

In the eigenbasis, the numerator is

$$\lambda_-^r \lambda_+^{N-r} + \lambda_+^r \lambda_-^{N-r}. \quad (10.25)$$

For fixed r and $N \rightarrow \infty$, divide numerator and denominator by λ_+^N . Since $|\lambda_-/\lambda_+| < 1$, the second numerator term and the λ_-^N denominator term vanish. Hence

$$\langle s_0 s_r \rangle = \left(\frac{\lambda_-}{\lambda_+} \right)^r = (\tanh K)^r. \quad (10.26)$$

Exercise 10.2 repeats this insertion and keeps the finite- N terms. Writing this as $e^{-r/\xi}$ gives

$$\xi^{-1} = -\ln(\tanh K). \quad (10.27)$$

The correlation length is finite for every $T > 0$ and diverges only as $T \rightarrow 0$.

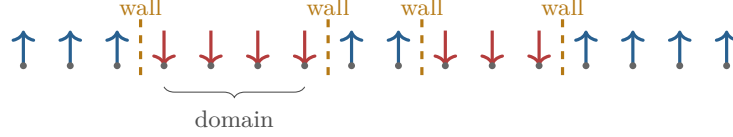


Figure 10.1: A one-dimensional spin configuration as domains separated by walls. Each wall costs $2J$, but it can be placed at order N bonds. At every nonzero temperature the equilibrium wall density is nonzero.

10.2.2 Domain-wall picture

A domain wall is a bond with $s_i \neq s_{i+1}$. A parallel bond contributes $-J$ to the energy, while an antiparallel bond contributes $+J$; one wall therefore costs $2J$. Its Boltzmann weight relative to a parallel bond is the fugacity

$$y = e^{-2\beta J}. \quad (10.28)$$

For a chain of length N , there are N possible wall positions. The energy cost stays fixed as N grows, but the number of placements grows. At any nonzero temperature the equilibrium wall density is therefore nonzero, and the chain breaks into finite domains. This is the physical reason behind the analytic transfer-matrix result.

The wall counting also reproduces the partition function. A configuration with n_w walls has $N - n_w$ parallel bonds and n_w antiparallel bonds, so

$$e^{-\beta H} = e^{K(N-n_w)} e^{-Kn_w} = e^{KN} e^{-2Kn_w}. \quad (10.29)$$

Choose the n_w wall bonds in $\binom{N}{n_w}$ ways. Once those bonds and the first spin are fixed, every other spin is fixed. The first spin gives a factor two. Periodicity requires an even number of flips, hence an even n_w :

$$Z_N = 2e^{KN} \sum_{\substack{n_w=0 \\ n_w \text{ even}}}^N \binom{N}{n_w} e^{-2Kn_w} \quad (10.30)$$

$$= (2 \cosh K)^N + (2 \sinh K)^N. \quad (10.31)$$

The resummation follows by adding the binomial expansions of $(1+x)^N$ and $(1-x)^N$; odd powers cancel and even powers double:

$$\sum_{k \text{ even}} \binom{N}{k} x^k = \frac{1}{2} [(1+x)^N + (1-x)^N]. \quad (10.32)$$

In the infinite chain the probability that a given bond is a wall follows from its indicator $w_i = (1 - s_i s_{i+1})/2$. Using the nearest-neighbor case of (10.26),

$$p_w = \langle w_i \rangle = \frac{1 - \langle s_i s_{i+1} \rangle}{2} = \frac{1 - \tanh K}{2} = \frac{e^{-2K}}{1 + e^{-2K}}. \quad (10.33)$$

At low temperature the mean domain length is $p_w^{-1} \simeq e^{2K}$, while $\xi \simeq (2p_w)^{-1} \simeq e^{2K}/2$. The factor two appears because crossing each random wall flips the sign of a correlation.

Correlation length in the one-dimensional chain

Measured in lattice spacings, (10.27) gives

$K = J/(k_B T)$	ξ
0.5	1.30
1	3.67
2	27.3

The rapid low-temperature growth can mimic ordering on a finite chain shorter than ξ , even though the infinite chain remains disordered for every $T > 0$.

Lecture 26 ends here. Lecture 27 starts with mean-field theory, then turns to the square lattice and finite-size scaling.

10.3 Mean-field theory

On a lattice with coordination number z , write $s_i = m + \delta s_i$, where $m = \langle s_i \rangle$ is assumed uniform. Then

$$s_i s_j = (s_i - m)(s_j - m) + m(s_i + s_j) - m^2. \quad (10.34)$$

Neglect the quadratic fluctuation term $(s_i - m)(s_j - m)$. Summing the remaining terms over bonds gives

$$H_{\text{MF}} = -(h + Jzm) \sum_i s_i + \frac{1}{2} N J z m^2. \quad (10.35)$$

The factor z counts the neighbors of one spin, while the factor one half in the constant avoids counting every bond twice. The spins are now independent in the effective field $h_{\text{eff}} = h + Jzm$. One spin has partition function $2 \cosh(\beta h_{\text{eff}})$ and mean

$$\langle s \rangle = \frac{e^{\beta h_{\text{eff}}} - e^{-\beta h_{\text{eff}}}}{e^{\beta h_{\text{eff}}} + e^{-\beta h_{\text{eff}}}} = \tanh(\beta h_{\text{eff}}). \quad (10.36)$$

Self-consistency requires this mean to equal the m used to define the effective field:

$$m = \tanh[\beta(Jzm + h)]. \quad (10.37)$$

At $h = 0$, the transition occurs when a nonzero solution first branches from $m = 0$. Near that point m is small, so $\tanh x = x - x^3/3 + \mathcal{O}(x^5)$ gives

$$m = \beta J z m - \frac{1}{3} (\beta J z m)^3 + \dots. \quad (10.38)$$

The linearized equation has a nonzero solution when its slope reaches one. Thus

$$k_B T_c^{\text{MF}} = Jz. \quad (10.39)$$

Keeping the cubic term shows both the exponent and its leading amplitude. Dividing the nonzero solution by m gives

$$1 = \beta J z - \frac{1}{3} (\beta J z)^3 m^2 + \dots, \quad (10.40)$$

so

$$m^2 \simeq \frac{3(\beta J z - 1)}{(\beta J z)^3} \simeq 3 \frac{T_c - T}{T_c}. \quad (10.41)$$

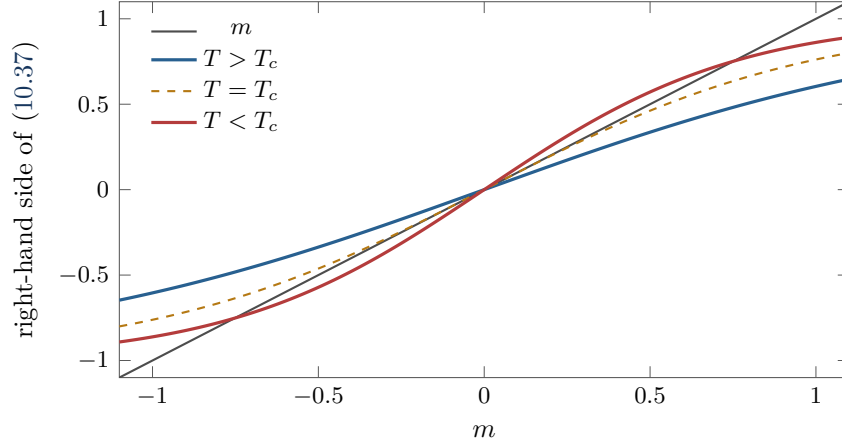


Figure 10.2: Graphical solution of the zero-field mean-field equation. Above T_c only $m = 0$ intersects the diagonal. Below T_c two stable nonzero intersections appear while the central solution becomes unstable.

Here the last expression keeps only the leading power of $(T_c - T)/T_c$. Hence $m \sim (T_c - T)^{1/2}$. The slope criterion provides the same transition: iteration of the self-consistency map near $m = 0$ is stable when $\beta Jz < 1$ and unstable when $\beta Jz > 1$.

One convenient mean-field potential is

$$\tilde{f}_{\text{MF}}(m) = \frac{1}{2}Jzm^2 - k_B T \ln [2 \cosh \beta(Jzm + h)], \quad (10.42)$$

and its derivative is

$$\frac{\partial \tilde{f}_{\text{MF}}}{\partial m} = Jz \{m - \tanh[\beta(Jzm + h)]\}. \quad (10.43)$$

Its stationary points therefore obey (10.37). A direct variational functional at fixed trial magnetization is energy minus mixing entropy:

$$f_{\text{MF}}(m) = -\frac{1}{2}Jzm^2 - hm - k_B T h_2\left(\frac{1+m}{2}\right). \quad (10.44)$$

Its stationarity also reproduces (10.37). The two functionals have the same value on self-consistent solutions, although they are not identical away from stationarity.

The mixing-entropy term follows from independent spins with probabilities $p_{\pm} = (1 \pm m)/2$. Their entropy per site is the binary entropy

$$h_2(p_+) = -p_+ \ln p_+ - p_- \ln p_-. \quad (10.45)$$

The trial energy per site is $-Jzm^2/2 - hm$, where the factor one half avoids counting each bond twice. Thus (10.44) is the free energy of a product-state trial distribution. The approximation discards connected correlations between different sites; this makes its variational meaning more precise than the phrase “replace neighbors by their average.”

The stationarity check is short but useful. Since

$$\frac{d}{dm} h_2\left(\frac{1+m}{2}\right) = -\frac{1}{2} \ln \frac{1+m}{1-m}, \quad (10.46)$$

$\partial_m f_{\text{MF}} = 0$ becomes

$$\frac{1}{2} \ln \frac{1+m}{1-m} = \beta(Jzm + h). \quad (10.47)$$

Applying $\tanh$ to both sides recovers (10.37).

Mean field incorrectly predicts a finite T_c in one dimension because it neglects long-wavelength domain fluctuations. The exact one-dimensional result exposes this failure and provides a stringent test of the approximation.

10.4 Two-dimensional square lattice

For the square-lattice Ising model at zero field, Kramers–Wannier duality maps K to K^* with $\sinh(2K) \sinh(2K^*) = 1$. Assuming a single transition, it identifies the self-dual candidate

$$\sinh(2K_c) = 1, \quad K_c = \frac{1}{2} \ln(1 + \sqrt{2}). \quad (10.48)$$

Onsager’s exact solution confirms a continuous transition at this value [27, 40]. The spontaneous magnetization was obtained later by Yang [63].

The logic of the duality can be seen without deriving the full finite-lattice identity. At high temperature, write each bond factor as

$$e^{K s_i s_j} = \cosh K [1 + s_i s_j \tanh K]. \quad (10.49)$$

After expanding the product over bonds, every selected bond contributes one factor of the spin at each endpoint. The sum $\sum_{s_i=\pm 1} s_i^q$ vanishes for odd q and equals two for even q . A term therefore survives the spin sums only when an even number of selected bonds meets at every site. Surviving graphs are closed loops, each occupied bond carrying weight $\tanh K$.

At low temperature, excitations above an aligned state are domain walls on the dual lattice. These boundaries are closed loops of the same combinatorial class as the surviving high-temperature graphs. A wall of length ℓ breaks ℓ bonds and carries weight $e^{-2K^*\ell}$. Matching the two loop expansions bond by bond gives

$$e^{-2K^*} = \tanh K, \quad (10.50)$$

which is equivalent to $\sinh(2K) \sinh(2K^*) = 1$. Boundary conditions and overall factors matter for the exact finite-size partition functions, but not for locating the self-dual coupling. Indeed, $e^{-2K^*} = \tanh K$ gives $e^{2K^*} = \coth K$, and hence

$$\sinh(2K^*) = \frac{1}{2}(\coth K - \tanh K) = \frac{1}{\sinh(2K)}. \quad (10.51)$$

Status: exact mapping plus a phase-structure assumption

Kramers–Wannier duality exactly relates high- and low-temperature partition functions, with finite-size qualifications. The equation $K = K^*$ identifies a self-dual point. Concluding that this point is the critical point also uses the fact that there is a single transition separating the two phases; duality alone does not prove uniqueness or continuity.

The spontaneous magnetization is

$$m(T) = \begin{cases} [1 - \sinh^{-4}(2K)]^{1/8}, & T < T_c, \\ 0, & T \geq T_c. \end{cases} \quad (10.52)$$

The critical exponents are

$$\alpha = 0 \text{ (logarithmic)}, \quad \beta = \frac{1}{8}, \quad \gamma = \frac{7}{4}, \quad \nu = 1, \quad \eta = \frac{1}{4}, \quad \delta = 15. \quad (10.53)$$

Quoted result

We derive the one-dimensional transfer matrix and mean-field equation. We do not derive Onsager's two-dimensional free energy or the magnetization formula in this course. They are quoted exact results used to test scaling and universality.

10.5 Metropolis sampling

For a finite two-dimensional lattice, a Markov chain generates a sequence of configurations in which the distribution of the next configuration depends only on the present one. This use of “Markov” names a memoryless stochastic process and is unrelated to Markov's probability inequality. The Metropolis rule aims to make the Gibbs distribution stationary:

1. choose a site i uniformly;
2. calculate the energy change ΔE for $s_i \mapsto -s_i$;
3. accept the flip with probability $\min(1, e^{-\beta\Delta E})$;
4. repeat and measure after equilibration.

For the proposed flip at site i , only terms containing s_i change. Therefore

$$\Delta E = 2s_i \left(J \sum_{j \sim i} s_j + h \right). \quad (10.54)$$

The proposal is symmetric: the probability of proposing $C \rightarrow C'$ equals that of proposing the reverse flip. The acceptance rule obeys

$$\frac{a(C \rightarrow C')}{a(C' \rightarrow C)} = e^{-\beta[E(C') - E(C)]}.$$

If $\Delta E > 0$, the forward acceptance is $e^{-\beta\Delta E}$ and the reverse acceptance is one. If $\Delta E < 0$, the same statement holds with forward and reverse interchanged. In both cases the displayed ratio follows. Multiplying by $p_{\text{eq}}(C) \propto e^{-\beta E(C)}$ gives

$$p_{\text{eq}}(C)P(C \rightarrow C') = p_{\text{eq}}(C')P(C' \rightarrow C), \quad (10.55)$$

which is detailed balance and implies stationarity after summing over C . Explicitly,

$$\sum_C p_{\text{eq}}(C)P(C \rightarrow C') = p_{\text{eq}}(C') \sum_C P(C' \rightarrow C) = p_{\text{eq}}(C'). \quad (10.56)$$

Near criticality, local updates suffer critical slowing down. Cluster algorithms improve the dynamics but are not needed for the first numerical experiment.

Detailed balance establishes stationarity, but a simulation also needs ergodicity: every configuration of nonzero Gibbs weight must be reachable. Single-spin flips satisfy this on a finite lattice at nonzero temperature. One *sweep* means N attempted single-spin updates. After an initial equilibration period, measurements taken after successive sweeps are still correlated. If A_t is the value recorded after sweep t , define the normalized autocorrelation function and integrated autocorrelation time

$$\rho_A(k) = \frac{\langle A_t A_{t+k} \rangle - \langle A \rangle^2}{\text{Var}(A)}, \quad \tau_{\text{int},A} = \frac{1}{2} + \sum_{k=1}^{\infty} \rho_A(k). \quad (10.57)$$

For M recorded sweeps, define the sample mean by

$$\bar{A} = \frac{1}{M} \sum_{t=1}^M A_t. \quad (10.58)$$

For a stationary chain with M much larger than the correlation time, summing the covariances of all pairs in this sample mean gives approximately

$$\text{Var}(\bar{A}) \simeq \frac{2\tau_{\text{int},A}}{M} \text{Var}(A). \quad (10.59)$$

Reporting error bars as if all sweeps were independent can therefore underestimate uncertainty badly near criticality.

For a d -dimensional hypercubic box of linear size L , the number of sites is $N = L^d$. Useful finite-size observables are

$$m_L = \frac{1}{L^d} \sum_i s_i, \quad (10.60)$$

$$\chi_L = \beta L^d (\langle m_L^2 \rangle - \langle m_L \rangle^2), \quad (10.61)$$

$$\chi_L^{(|m|)} = \beta L^d (\langle m_L^2 \rangle - \langle |m_L| \rangle^2), \quad (10.62)$$

$$U_4 = 1 - \frac{\langle m_L^4 \rangle}{3 \langle m_L^2 \rangle^2}. \quad (10.63)$$

The first susceptibility is the fluctuation–response quantity. The second is a frequently used finite-size estimator that removes fluctuations caused only by switching between the two magnetized peaks; it should be labeled separately. The Binder cumulant U_4 for several L values has a common crossing near T_c , up to corrections to scaling.

10.6 Finite-size scaling

Define the reduced temperature by $t = (T - T_c)/T_c$. In the bulk, [section 9.6](#) chooses the rescaling factor from $\xi \sim |t|^{-\nu}$. A finite box introduces the dimensionless ratio $L/\xi \sim L|t|^\nu$, or equivalently the scaling variable $tL^{1/\nu}$. At $t = 0$ the only large length is L , so the bulk powers of ξ become powers of L . This gives

$$\langle |m_L| \rangle = L^{-\beta/\nu} \mathcal{M}(tL^{1/\nu}), \quad (10.64)$$

$$\chi_L(t) = L^{\gamma/\nu} \mathcal{X}(tL^{1/\nu}). \quad (10.65)$$

Data collapse tests both T_c and the exponent ratios. It also shows why a finite simulation has rounded rather than singular curves.

The susceptibility maximum occurs when its scaling argument is order one, so the pseudo-critical temperature shifts as

$$|T_c(L) - T_c| \sim L^{-1/\nu}. \quad (10.66)$$

At the peak, $\chi_L^{\text{max}} \sim L^{\gamma/\nu}$, while at the critical point $\langle |m_L| \rangle \sim L^{-\beta/\nu}$. These three relations provide separate checks. A visually pleasing collapse alone is not a reliable exponent measurement if the temperature window, corrections to scaling, or autocorrelations have not been tested.

10.7 Exercises

Exercise 10.1. Construct the transfer matrix and derive the eigenvalues (10.12).

Exercise 10.2. Use transfer-matrix eigenvalues to derive (10.26) and (10.27).

Exercise 10.3. Calculate the one-dimensional energy and heat capacity at zero field. Show that the heat capacity is analytic for $T > 0$.

Exercise 10.4. Estimate the typical domain size at low temperature from the domain-wall fugacity and compare with the asymptotic form of ξ .

Exercise 10.5. Derive the mean-field equation and free energy. Calculate the susceptibility above and below T_c^{MF} .

Exercise 10.6. Verify that the exact two-dimensional exponents satisfy $\gamma = \nu(2 - \eta)$ and the Rushbrooke relation $\alpha + 2\beta + \gamma = 2$.

Exercise 10.7 (*). Implement Metropolis sampling for the square-lattice model. Plot energy, $|m|$, heat capacity, susceptibility, and Binder cumulant for at least three lattice sizes. Do not infer a sharp T_c from a single finite lattice.

10.8 Further reading

Kardar treats transfer matrices, mean field, and scaling in a unified way [23, 22]. The exact two-dimensional solution is due to Onsager [40], with the critical point anticipated by duality [27].

Chapter 11

Renormalization and universality

Lecture 28–29. Coarse graining, an exact one-dimensional decimation, fixed points, relevance, scaling, and entropy under loss of resolution.

What you should be able to do

- Perform one explicit real-space decimation and read the resulting coupling flow.
- Distinguish an exact coarse-graining identity from a truncation of the effective Hamiltonian.
- Linearize a renormalization map near a fixed point and classify its scaling fields.
- Derive critical exponents from RG eigenvalues and perform Gaussian power counting.
- Explain what “information loss under coarse graining” does and does not imply for entropy.

11.1 Why scale enters

Near a continuous phase transition, the correlation length becomes much larger than the microscopic lattice spacing. A long-distance observer cannot resolve every lattice variable separately. The useful question is how the couplings change when the observation scale changes. Systems with different microscopic Hamiltonians can share the same critical exponents when their coarse-grained couplings approach the same scale-invariant theory.

The renormalization group turns this observation into a procedure:

1. remove or average short-distance variables;
2. rescale lengths to restore the original cutoff;
3. rescale fields and compare the new Hamiltonian with the old one.

The Hamiltonian changes because the remaining variables must reproduce the statistics of the variables that were retained. Repeating the procedure generates a flow in the space of all symmetry-allowed couplings. The word “group” is historical: coarse graining usually cannot be inverted, so the transformations form a semigroup unless extra information is retained.

11.2 Exact definition of a coarse-grained Hamiltonian

Let microscopic configurations be s and block variables be S . A coarse-graining kernel $P(S|s)$ is the conditional probability of assigning block configuration S to microscopic configuration s ; it satisfies

$$P(S|s) \geq 0, \quad \sum_S P(S|s) = 1. \quad (11.1)$$

In this chapter H is dimensionless: it denotes β times an energy, or more generally minus the logarithm of a Boltzmann weight. Define the effective Hamiltonian by

$$e^{-H'(S)} = \sum_s P(S|s) e^{-H(s)}. \quad (11.2)$$

Summing both sides over S shows what is exact:

$$Z' = \sum_S e^{-H'(S)} \quad (11.3)$$

$$= \sum_s \left[\sum_S P(S|s) \right] e^{-H(s)} \quad (11.4)$$

$$= \sum_s e^{-H(s)} = Z. \quad (11.5)$$

An additive constant in H' carries part of the free energy and cannot be discarded when thermodynamic potentials are wanted. The resulting H' usually contains every interaction allowed by symmetry. Equation (11.2) itself is an identity; replacing H' by a short list of couplings is the approximation.

11.3 One-dimensional Ising decimation

Take an even number N of sites so that alternate-spin decimation closes on a periodic chain. At zero field, the dimensionless Hamiltonian is

$$-H = K \sum_i s_i s_{i+1}. \quad (11.6)$$

Keep the even spins $S_i = s_{2i}$ and sum over the odd spins. An odd spin s_{2i+1} couples only to its two retained neighbors, so its local sum is

$$\sum_{s_{2i+1}=\pm 1} \exp[K S_i s_{2i+1} + K s_{2i+1} S_{i+1}] \quad (11.7)$$

$$= 2 \cosh[K(S_i + S_{i+1})]. \quad (11.8)$$

Different odd-spin sums share no eliminated variable, so the full partition sum factors into one such expression per block. Write the local factor as a new nearest-neighbor interaction:

$$A(K) e^{K' S_i S_{i+1}}. \quad (11.9)$$

For parallel block spins, $S_i S_{i+1} = +1$ and the factor is $2 \cosh 2K$. For antiparallel spins, $S_i S_{i+1} = -1$ and the factor is 2. Thus

$$A e^{K'} = 2 \cosh 2K, \quad A e^{-K'} = 2. \quad (11.10)$$

Dividing these equations determines K' ; multiplying them determines A :

$$K' = \frac{1}{2} \ln \cosh(2K), \quad A = 2 \sqrt{\cosh(2K)}. \quad (11.11)$$

At fixed K' , the factor A cancels from normalized probabilities of the retained spins, so it does not change their correlation functions. It does contribute to the free energy and to thermodynamic derivatives with respect to the original coupling K .

For small K , use $\ln \cosh x = x^2/2 - x^4/12 + \mathcal{O}(x^6)$:

$$K' = K^2 - \frac{2}{3}K^4 + \mathcal{O}(K^6), \quad (11.12)$$

so finite high-temperature couplings flow toward $K = 0$. At low temperature, introduce the domain-wall fugacity $y = e^{-2K}$. Equation (11.11) gives

$$\cosh(2K) = \frac{1}{2}e^{2K} (1 + e^{-4K}), \quad (11.13)$$

$$K' = K - \frac{1}{2} \ln 2 + \mathcal{O}(e^{-4K}). \quad (11.14)$$

The fugacity recursion can be written exactly:

$$y' = e^{-2K'} = \frac{1}{\cosh 2K} = \frac{2y}{1 + y^2}. \quad (11.15)$$

Expanding it at small y gives

$$y' = 2y - 2y^3 + \mathcal{O}(y^5). \quad (11.16)$$

The linear coefficient is larger than one, so domain walls are relevant under blocking: any nonzero temperature flows away from the ordered zero-temperature point. This recovers the absence of a finite-temperature transition. Exercise 11.3 checks the correction term.

Status: model-specific exact result

The decimation formula is exact for the zero-field nearest-neighbor chain because summing alternate spins generates only a new nearest-neighbor interaction and a constant. In higher dimensions, decimation generates multispin and longer-range couplings.

11.3.1 Fixed points and the free-energy contribution

A fixed point is a coupling left unchanged by the recursion. This map has two fixed points when the endpoints are included:

$$K^* = 0, \quad K^* = \infty. \quad (11.17)$$

For $K > 0$, $\cosh(2K) < e^{2K}$, so $0 < K' < K$; every finite ferromagnetic coupling therefore flows to zero. The point $K = \infty$ represents the zero-temperature ordered chain and is unstable to a nonzero wall fugacity.

One decimation step doubles the lattice spacing, so $b = 2$. After n steps the linear term gives $y_n \simeq 2^n y$. The flow leaves the low-temperature neighborhood when y_n becomes order one, or $2^n \sim y^{-1}$. The corresponding physical length is $b^n = 2^n$, so

$$\xi \sim y^{-1} \sim e^{2K}. \quad (11.18)$$

Equivalently, the linearized map has eigenvalue $b^{y_t} = 2$, hence $y_t = 1$ and $\nu_y = 1/y_t = 1$. This agrees with the exact low-temperature correlation length in Eq. (10.27), whose leading form is $\xi \simeq e^{2K}/2$. The exponent refers to $y = e^{-2K}$ as the scaling field near the $T = 0$ fixed point; writing a power law in T itself would be misleading.

The factor $A(K)$ is needed for thermodynamics. For even N ,

$$Z_N(K) = A(K)^{N/2} Z_{N/2}(K'). \quad (11.19)$$

If $\varphi(K) = -(1/N) \ln Z_N$ is the dimensionless free energy per original site, divide the logarithm of the preceding identity by $-N$:

$$\varphi(K) = -\frac{1}{N} \left[\frac{N}{2} \ln A(K) + \ln Z_{N/2}(K') \right] \quad (11.20)$$

$$= -\frac{1}{2} \ln A(K) + \frac{1}{2} \varphi(K'). \quad (11.21)$$

Iterating until K is small reconstructs the free energy as a sum of constants generated at successive scales. Dropping those constants would preserve correlation recursions but give the wrong thermodynamics.

11.4 Fixed points and scaling fields

For a general system, expand the dimensionless Hamiltonian in all symmetry-allowed operators $\mathcal{O}_i$:

$$H = \sum_i g_i \mathcal{O}_i. \quad (11.22)$$

An RG step with length factor b defines $g'_i = R_i(\mathbf{g})$. A fixed point satisfies

$$\mathbf{g}^* = \mathbf{R}(\mathbf{g}^*). \quad (11.23)$$

Linearize near it:

$$\delta g'_i = \sum_j T_{ij} \delta g_j, \quad T_{ij} = \left. \frac{\partial R_i}{\partial g_j} \right|_{\mathbf{g}^*}. \quad (11.24)$$

Diagonalize the stability matrix. If u_a is the coordinate along one of its eigenvectors, then

$$u'_a = b^{y_a} u_a. \quad (11.25)$$

After n steps, $u_a^{(n)} = b^{ny_a} u_a$. This makes the classification concrete:

- $y_a > 0$: relevant, grows at long distance;
- $y_a < 0$: irrelevant, decays at long distance;
- $y_a = 0$: marginal, decided by nonlinear terms.

A critical surface consists of couplings that flow into a critical fixed point. Each relevant direction must be tuned away to reach that surface. In the simplest Ising transition at $h = 0$, tuning the temperature to T_c sets the thermal relevant field to zero; imposing the $\mathbb{Z}_2$ symmetry sets the magnetic field to zero.

If an operator $\mathcal{O}_a$ has scaling dimension Δ_a , then under $\mathbf{x} = b\mathbf{x}'$ it transforms as $\mathcal{O}_a(\mathbf{x}) = b^{-\Delta_a} \mathcal{O}'_a(\mathbf{x}')$. Its Hamiltonian term becomes

$$\int d^d x g_a \mathcal{O}_a(\mathbf{x}) = \int d^d x' b^{d-\Delta_a} g_a \mathcal{O}'_a(\mathbf{x}'). \quad (11.26)$$

The rescaled coupling is $g'_a = b^{y_a} g_a$ with

$$y_a = d - \Delta_a. \quad (11.27)$$

Relevant operators have $\Delta_a < d$, irrelevant operators have $\Delta_a > d$, and marginal operators have $\Delta_a = d$. This operator language is often easier to transfer from lattice models to field theory.

Lecture 28 ends here. Lecture 29 starts by turning fixed-point eigenvalues into critical exponents and power-counting rules.

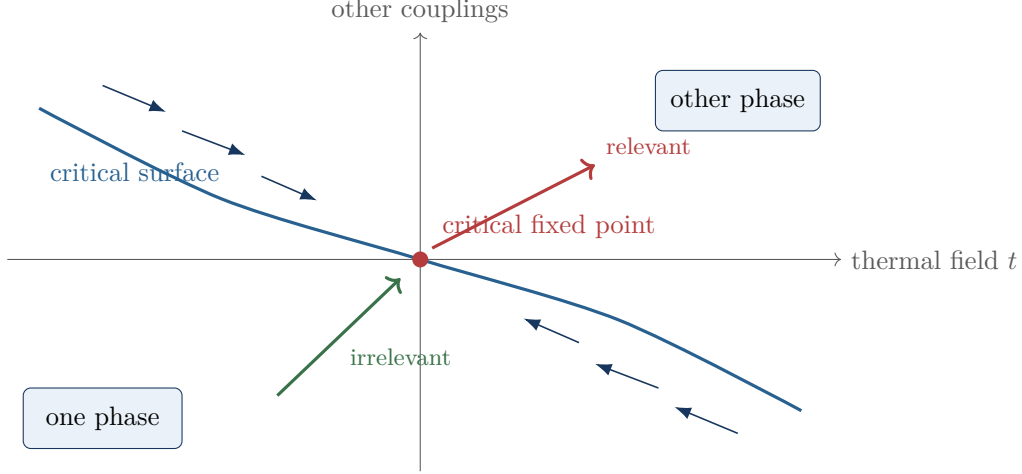


Figure 11.1: Schematic flow near a critical fixed point. Irrelevant directions contract toward the critical surface, while a relevant thermal perturbation carries the system toward one of two phases. Reaching the fixed point requires tuning the relevant field.

11.5 Critical exponents from RG eigenvalues

If the reduced temperature transforms as

$$t' = b^{y_t} t, \quad (11.28)$$

choose $b \sim |t|^{-1/y_t}$ so that t' becomes order one. Since lengths shrink by b under the RG step,

$$\xi(t) = b \xi(t') \sim |t|^{-1/y_t}. \quad (11.29)$$

Therefore

$$\nu = \frac{1}{y_t}. \quad (11.30)$$

If the magnetic field has eigenvalue y_h , the singular free energy satisfies

$$f_s(t, h) = b^{-d} f_s(b^{y_t} t, b^{y_h} h). \quad (11.31)$$

The chain-rule calculation in Eq. (9.52) gives the magnetization exponent. Applying the same operation to the other derivatives gives

$$\beta = \frac{d - y_h}{y_t}, \quad (11.32)$$

$$\gamma = \frac{2y_h - d}{y_t}, \quad (11.33)$$

$$\delta = \frac{y_h}{d - y_h}, \quad (11.34)$$

$$\alpha = 2 - \frac{d}{y_t}, \quad (11.35)$$

provided the homogeneous free-energy form holds without a dangerous irrelevant variable. The Fisher relation $\gamma = \nu(2 - \eta)$ was derived independently from the correlation function in Eq. (9.48).

11.6 Power counting in the Landau–Ginzburg theory

Consider

$$\mathcal{F}[\phi] = \int d^d x \left[\frac{1}{2} (\nabla \phi)^2 + \frac{r}{2} \phi^2 + \frac{u}{4!} \phi^4 - h \phi \right]. \quad (11.36)$$

Rescale $\mathbf{x} = b\mathbf{x}'$. Then $d^d x = b^d d^d x'$ and $\nabla = b^{-1} \nabla'$. Write $\phi(\mathbf{x}) = b^{-\Delta_\phi} \phi'(\mathbf{x}')$. The gradient term transforms as

$$\frac{1}{2} \int d^d x (\nabla \phi)^2 = \frac{1}{2} b^{d-2-2\Delta_\phi} \int d^d x' (\nabla' \phi')^2. \quad (11.37)$$

Keeping its coefficient fixed requires $\Delta_\phi = (d-2)/2$, so

$$\phi(\mathbf{x}) = b^{-(d-2)/2} \phi'(\mathbf{x}'). \quad (11.38)$$

The mass and quartic terms then transform one by one:

$$\frac{r}{2} \int d^d x \phi^2 = \frac{b^2 r}{2} \int d^d x' \phi'^2, \quad (11.39)$$

$$\frac{u}{4!} \int d^d x \phi^4 = \frac{b^{4-d} u}{4!} \int d^d x' \phi'^4. \quad (11.40)$$

Thus at the Gaussian fixed point,

$$r' = b^2 r, \quad u' = b^{4-d} u. \quad (11.41)$$

The linear field term similarly gives

$$h' = b^{d-(d-2)/2} h = b^{(d+2)/2} h. \quad (11.42)$$

The mass parameter has eigenvalue $y_r = 2$, while the Gaussian magnetic eigenvalue is $y_h = (d+2)/2$. The quartic coupling is irrelevant for $d > 4$, relevant for $d < 4$, and marginal at $d = 4$. This is why four is the upper critical dimension of scalar ϕ^4 theory.

For $d < 4$, the interacting fixed point changes the exponents from their mean-field values. The $\epsilon = 4 - d$ expansion makes this statement calculable, but its loop expansion lies beyond the two-lecture RG block [61].

Status: controlled approximation

Power counting at the Gaussian fixed point is exact as dimensional analysis. Using it to calculate non-Gaussian exponents below four dimensions requires a perturbative RG calculation, usually organized in ϵ or another expansion parameter.

11.6.1 What the first epsilon-expansion step says

The one-loop result below is guided reading and is not a required exam derivation. We quote the shell integral and rescaling calculation from Ref. [61]. Let ℓ be the accumulated logarithmic scale, so an infinitesimal RG step changes ℓ by $d\ell$. In a dimensionless normalization of the quartic coupling g , the one-component flow is

$$\frac{dg}{d\ell} = \epsilon g - 3g^2 + \mathcal{O}(g^3), \quad (11.43)$$

$$\frac{dt}{d\ell} = (2 - g)t + \mathcal{O}(g^2 t), \quad \epsilon = 4 - d. \quad (11.44)$$

Here t denotes the thermal scaling field after the critical value of the mass has been subtracted; it is proportional to $T - T_c$ near the fixed point. The right-hand sides of these flow equations

are called beta functions. Their numerical coefficients depend on the normalization of g , while fixed-point exponents do not. Fixed points solve $dg/d\ell = 0$. Factoring the first beta function gives $g(\epsilon - 3g) = 0$, so besides the Gaussian fixed point there is an interacting Wilson–Fisher fixed point

$$g^* = \frac{\epsilon}{3} + \mathcal{O}(\epsilon^2). \quad (11.45)$$

Substituting g^* into the coefficient of t gives the thermal eigenvalue

$$y_t = 2 - \frac{\epsilon}{3} + \mathcal{O}(\epsilon^2). \quad (11.46)$$

Expanding its reciprocal gives

$$\nu = \frac{1}{y_t} = \frac{1}{2 - \epsilon/3} = \frac{1}{2} + \frac{\epsilon}{12} + \mathcal{O}(\epsilon^2). \quad (11.47)$$

The anomalous dimension begins at order ϵ^2 . The other first-order exponents follow from the scaling relations already derived. Two examples are

$$\gamma = \nu(2 - \eta) = 1 + \frac{\epsilon}{6} + \mathcal{O}(\epsilon^2), \quad (11.48)$$

$$\alpha = 2 - (4 - \epsilon)\nu = \frac{\epsilon}{6} + \mathcal{O}(\epsilon^2). \quad (11.49)$$

The Rushbrooke relation then gives

$$\beta = \frac{2 - \alpha - \gamma}{2} = \frac{1}{2} - \frac{\epsilon}{6} + \mathcal{O}(\epsilon^2). \quad (11.50)$$

Collecting the results,

$$\eta = \mathcal{O}(\epsilon^2), \quad \gamma = 1 + \frac{\epsilon}{6} + \mathcal{O}(\epsilon^2), \quad (11.51)$$

$$\beta = \frac{1}{2} - \frac{\epsilon}{6} + \mathcal{O}(\epsilon^2), \quad \alpha = \frac{\epsilon}{6} + \mathcal{O}(\epsilon^2). \quad (11.52)$$

The resulting exponent shifts quantify the effect of fluctuations below four dimensions. The displayed beta functions are quoted one-loop results; the algebra from them to the fixed point and exponents is derived above. The expansion is controlled for $\epsilon \ll 1$. Substituting $\epsilon = 2$ to estimate the two-dimensional Ising model is an extrapolation, not a controlled result.

11.7 Universality

Microscopic Hamiltonians in the same basin of attraction flow toward the same fixed point. For static critical behavior, the usual classification asks for:

- spatial dimension;
- order-parameter components and symmetry;
- range of interactions;

Questions about critical dynamics require extra information, including which quantities the dynamics conserves. Two models can therefore share static exponents but have different dynamic exponents. Irrelevant couplings still affect nonuniversal quantities such as T_c and amplitudes. Universality does not say that microscopic physics disappears from every observable.

11.8 Entropy and coarse graining

RG removes short-distance variables, so it is tempting to say that “entropy must increase under RG.” That sentence is incomplete. Several distinct claims are available:

- A stochastic coarse-graining channel contracts relative entropy by data processing.
- The Shannon entropy of a blocked variable can rise or fall, depending on the map and distribution.
- The thermodynamic entropy of the physical state does not change merely because we rewrite its partition function with effective variables.
- Integrating out modes moves their effects into the effective Hamiltonian and its normalization; dropping generated couplings adds an approximation.

The statement that survives these distinctions is relational: coarse graining reduces distinguishability among microscopic preparations. Relative entropy expresses this exactly through the data-processing inequality (2.76).

This can be tied directly to (11.2). A microscopic Gibbs distribution is

$$p(s) = \frac{e^{-H(s)}}{Z}. \quad (11.53)$$

The induced block distribution is

$$p'(S) = \sum_s P(S|s)p(s) = \frac{e^{-H'(S)}}{Z}. \quad (11.54)$$

Thus the same normalized kernel that defines the effective Hamiltonian defines a stochastic channel on states. For any two microscopic preparations p and q , apply the same kernel to both to obtain

$$D(p' \| q') \leq D(p \| q). \quad (11.55)$$

The statement compares two preparations at the same resolution. It does not compare the Shannon entropy of variables living on spaces of different sizes.

Worked example: why Shannon entropy has no fixed RG direction

Let two independent fair spins (s_1, s_2) be blocked into their product $S = s_1 s_2$. The microscopic Shannon entropy is $2 \ln 2$ and the block entropy is $\ln 2$: it decreases because one bit has been discarded. Conversely, a stochastic blocking rule that ignores its input and outputs a fair bit sends a certain input to entropy $\ln 2$: it increases because the kernel injects noise. Both are valid stochastic maps, and both contract relative entropy between any two input preparations. Data processing, not monotonic marginal entropy, is the invariant statement.

11.9 Entanglement near a one-dimensional quantum critical point

For a pure ground state, the entropy of a region measures entanglement with its complement. In the gapped one-dimensional local systems relevant here, it is typically governed by degrees of freedom close to the two boundary points and approaches a constant when the interval is much longer than the correlation length. At a critical point correlations persist on every scale and the entropy grows logarithmically instead.

For the ground state of a one-dimensional critical quantum system described by a conformal field theory, the dimensionless entanglement entropy of an interval of length ℓ in an infinite system scales as

$$s_A = \frac{c}{3} \ln \frac{\ell}{a} + c_1, \quad (11.56)$$

where the central charge c is a dimensionless number characterizing the conformal field theory, a is a short-distance cutoff, and c_1 is nonuniversal. For $\alpha > 0$ with $\alpha \neq 1$, Rényi entropies obey

$$s_\alpha(A) = \frac{c}{6} \left(1 + \frac{1}{\alpha}\right) \ln \frac{\ell}{a} + c_\alpha. \quad (11.57)$$

These formulas connect scale invariance with the information stored across a spatial boundary [7].

At nonzero temperature, write $\beta_T = 1/(k_B T)$ to avoid confusing inverse temperature with the order-parameter exponent. The conformal result becomes

$$s_A(\ell, \beta_T) = \frac{c}{3} \ln \left[\frac{\beta_T \hbar v}{\pi a} \sinh \left(\frac{\pi \ell}{\beta_T \hbar v} \right) \right] + c_1, \quad (11.58)$$

where v is the characteristic velocity. For $\ell \ll \beta_T \hbar v$, the ground-state logarithm is recovered. For $\ell \gg \beta_T \hbar v$,

$$s_A \simeq \frac{\pi c}{3\beta_T \hbar v} \ell + \text{constant}, \quad (11.59)$$

and the extensive term is the thermal entropy contained in a long interval. The two limits follow directly from $\sinh x \simeq x$ for $x \ll 1$ and $\sinh x \simeq e^x/2$ for $x \gg 1$. In the first limit the factors $\beta_T \hbar v/\pi$ cancel, leaving (11.56); in the second, the logarithm of the exponential produces the term linear in ℓ . The same formula therefore displays the crossover from boundary entanglement to volume-law thermal entropy.

The cutoff a is unavoidable in the entropy of a sharp spatial region. Mutual information between separated regions cancels the leading boundary-local cutoff terms and is often a cleaner continuum quantity. This is another instance in which a relative or correlation measure is less sensitive to the reference resolution than an absolute entropy.

Quoted field-theory result

Equations (11.56), (11.57), and (11.58) are quoted results. Their derivation uses conformal field theory and the replica method, which are outside this course.

11.10 Core results and further reading

The core results of this chapter are the exact one-dimensional decimation (11.11), its fixed points, and the classification of relevant and irrelevant directions. Power counting, the one-loop ϵ expansion, and critical entanglement extend that core and may be treated as further reading.

11.11 Exercises

Exercise 11.1. Derive the decimation recursion (11.11), including the constant $A(K)$.

Exercise 11.2. Plot or iterate $K' = \frac{1}{2} \ln \cosh 2K$. Show that every finite K flows toward the high-temperature fixed point.

Exercise 11.3. Derive the low-temperature domain-wall flow $y' = 2y + \mathcal{O}(y^3)$. Interpret its eigenvalue.

Exercise 11.4. Starting from the free-energy scaling law, derive the exponent formulas in terms of y_t and y_h .

Exercise 11.5. Perform power counting for ϕ^6 and find the dimension at which it becomes marginal at the Gaussian fixed point.

Exercise 11.6. Give an explicit binary coarse-graining map for which Shannon entropy decreases. Verify that relative entropy between any two input distributions cannot increase.

Exercise 11.7 (*). For a finite periodic chain, take the quoted CFT formula

$$s_A = \frac{c}{3} \ln \left[\frac{L}{\pi a} \sin \left(\frac{\pi \ell}{L} \right) \right] + c_1, \quad (11.60)$$

and verify that it is symmetric under $\ell \leftrightarrow L - \ell$ and reduces to (11.56) when $a \ll \ell \ll L$. Explain why dimensional analysis and periodicity alone do not fix the sine dependence.

11.12 Further reading

Wilson and Kogut give the classic account of RG and the ϵ expansion [61]. Kardar develops modern statistical field theory [22]. The critical entanglement formulas are due to Calabrese and Cardy [7].

Chapter 12

One word, several entropies

Lecture 29, second half. A final map of entropy concepts and the assumptions connecting them.

What you should be able to do

- Identify the mathematical object and units behind each use of the word entropy.
- Separate exact identities from equilibrium statements, limiting results, and coarse-grained descriptions.
- Construct an entropy balance for a system, bath, and their correlations.
- Diagnose common claims about missing information, entanglement, and entropy increase.

12.1 The map

The word entropy refers to several functions defined on different objects. Some carry thermodynamic units through a factor k_B ; others are dimensionless. Some belong to one state, while relative entropy compares two states. Confusion begins when the object, units, or chosen description is left implicit.

Name	Definition	Object and role	Main qualification
Thermodynamic	$dS = \delta Q_{\text{rev}}/T$	Equilibrium state function	For irreversible paths use $\Delta S \geq \int \delta Q/T_b$; a local differential form needs internal equilibrium
Boltzmann	$S_B = k_B \ln \Omega(M)$	Multiplicity of microstates in macrostate M	Depends on macrovariables, shell, and coarse graining

Name	Definition	Object and role	Main qualification
Gibbs	$S_G = -k_B \sum_i p_i \ln p_i$	Probability distribution over discrete microstates	Fine-grained value is conserved by Hamiltonian flow; continuous variables require a reference measure
Shannon	$h(X) = -\sum_x p(x) \ln p(x)$	Classical random variable; uncertainty and coding	Dimensionless and description-dependent
Relative entropy	$D(p q) = \sum_i p_i \ln(p_i/q_i)$	Distinguishability of two distributions	Asymmetric divergence, not a state entropy; infinite if p has support outside q
von Neumann	$s(\rho) = -\text{Tr } \rho \ln \rho$	Spectrum of a quantum state	Dimensionless here; conserved by unitary evolution
Quantum relative	$D(\rho \sigma) = \text{Tr } \rho(\ln \rho - \ln \sigma)$	Quantum distinguishability and free-energy excess	Requires a support condition
Mutual information	$I(A : B) = s_A + s_B - s_{AB}$	Total correlations between subsystems	Contains classical and quantum correlation
Entanglement	$s(\rho_A)$ for pure AB	Entanglement across a bipartition	Reduced entropy is not an entanglement measure for a general mixed AB state
Rényi	$s_\alpha = (1 - \alpha)^{-1} \ln \text{Tr } \rho^\alpha$	Family probing the spectrum	Different indices answer different questions
Entropy production	$\Sigma[\gamma] = \ln(P_F[\gamma]/P_R[\tilde{\gamma}])$, with mean $\langle \Sigma \rangle$	Irreversibility of a specified stochastic or open process	The forward/reverse experiment, time reversal, and reservoirs must be named

12.2 Exact coincidences

12.2.1 Uniform microcanonical shell

If the probability is uniform on exactly Ω allowed microstates,

$$p_i = \frac{1}{\Omega}, \quad (12.1)$$

then

$$S_G = -k_B \sum_{i=1}^{\Omega} \frac{1}{\Omega} \ln \frac{1}{\Omega} = k_B \ln \Omega = S_B. \quad (12.2)$$

The equality is exact after the shell and its measure have been fixed. It does not say that every equilibrium distribution is uniform on the full state space; uniformity holds on the specified microcanonical shell.

12.2.2 Equilibrium Gibbs distribution

For $p_i^{\text{eq}} = e^{-\beta E_i} / Z$,

$$S_G[p^{\text{eq}}] = -k_B \sum_i p_i^{\text{eq}} \ln \left(\frac{e^{-\beta E_i}}{Z} \right) \quad (12.3)$$

$$= k_B \sum_i p_i^{\text{eq}} (\beta E_i + \ln Z) \quad (12.4)$$

$$= k_B (\beta U + \ln Z). \quad (12.5)$$

For a temperature-independent Hamiltonian, $F = -k_B T \ln Z$. Differentiating at fixed V, N gives

$$-\left(\frac{\partial F}{\partial T} \right)_{V,N} = k_B \ln Z + k_B T \frac{\partial \ln Z}{\partial T} \quad (12.6)$$

$$= k_B \ln Z + \frac{U}{T} = S_G[p^{\text{eq}}], \quad (12.7)$$

where $\partial_T \ln Z = U/(k_B T^2)$ follows by differentiating the canonical partition function. The statistical and thermodynamic entropies agree for the equilibrium model. Agreement with a physical material still depends on the Hamiltonian and the equilibrium assumptions being accurate.

12.2.3 Free-energy excess

For a classical distribution p , define $\mathcal{F}_T[p] = \sum_i p_i E_i - T S_G[p]$. For a density operator ρ , define $F_T(\rho) = \text{Tr}(\rho H) - k_B T s(\rho)$. Let p_β and ρ_β denote the corresponding Gibbs states. Then

$$\mathcal{F}_T[p] - \mathcal{F}_T[p_\beta] = k_B T D(p \| p_\beta), \quad (12.8)$$

$$F_T(\rho) - F_T(\rho_\beta) = k_B T D(\rho \| \rho_\beta). \quad (12.9)$$

These are exact identities for a fixed Hamiltonian and nonzero temperature; their derivations appear in Eqs. (3.44) and (7.44). For $T > 0$, positivity of relative entropy proves that the Gibbs state minimizes free energy. For a normalizable negative-temperature Gibbs state, the factor $k_B T < 0$ reverses the inequality and the Gibbs state maximizes these functionals.

12.2.4 Correlation as distinguishability from a product

Classically and quantum mechanically,

$$I(X : Y) = D(p_{XY} \| p_X p_Y), \quad (12.10)$$

$$I(A : B) = D(\rho_{AB} \| \rho_A \otimes \rho_B). \quad (12.11)$$

Mutual information measures how distinguishable a joint state is from the product state with the same marginals. It vanishes exactly when the joint state is a product.

12.3 Statements that require a limit or approximation

12.3.1 Boltzmann entropy and thermodynamic entropy

For a large additive equilibrium system, a macrostate M carries a multiplicity $\Omega(M)$. When $\ln \Omega$ is extensive, ratios of multiplicities behave as exponentials of system size. The maximizing macrostate then dominates sums over M , and $k_B \ln \Omega(M_*)$ reproduces thermodynamic entropy to leading order. At finite size, the choice of shell width, surface versus volume count, and subextensive corrections remain visible.

12.3.2 Equivalence of ensembles

Microcanonical and canonical predictions agree for many local observables when the entropy density is concave, correlations are sufficiently short-ranged, and the thermodynamic limit is regular. The canonical energy distribution then concentrates near the energy selected by $\partial S / \partial E = 1/T$. Long-range interactions or a nonconcave entropy density can prevent this concentration and break equivalence. At first-order coexistence, even additive short-range systems require care because the canonical state can mix phases while a microcanonical constraint fixes the total energy.

12.3.3 Entropy increase

Fine-grained classical and quantum entropies do not increase under reversible closed dynamics. Increasing entropy appears after one of the following has been stated:

- coarse graining or limited observables;
- a Markov approximation and discarded memory;
- molecular chaos in kinetic theory;
- an open-system channel with an environment;
- typical macroscopic preparation and a thermodynamic-limit argument.

No single item should be silently substituted for another.

These mechanisms answer different questions. Coarse graining changes the description; an open-system channel changes the reduced state; molecular chaos closes a kinetic equation by discarding correlations. A derivation must name which operation it uses.

12.3.4 An exact system–bath entropy balance

There is a useful identity that gathers several of these ideas. Let a system start in an arbitrary state ρ_S and an initially uncorrelated bath start in the Gibbs state $\rho_B^\beta = e^{-\beta H_B} / Z_B$. Apply a joint unitary and call the final state ρ'_{SB} . Define heat absorbed by the system as the bath energy loss,

$$Q = \text{Tr}(H_B \rho_B^\beta) - \text{Tr}(H_B \rho'_B). \quad (12.12)$$

Write $\Delta s_S = s(\rho'_S) - s(\rho_S)$. Unitary invariance and the initial product state give

$$D(\rho'_{SB} \| \rho'_S \otimes \rho_B^\beta) = \Delta s_S - \beta Q \geq 0. \quad (12.13)$$

To verify the identity, use $\ln \rho_B^\beta = -\beta H_B - \ln Z_B$ to expand the relative entropy:

$$D = -s(\rho'_{SB}) + s(\rho'_S) + \beta \text{Tr}(H_B \rho'_B) + \ln Z_B.$$

Unitarity and the initial product state give $s(\rho'_{SB}) = s(\rho_S) + s(\rho_B^\beta)$, while the Gibbs identity gives $s(\rho_B^\beta) = \beta \text{Tr}(H_B \rho_B^\beta) + \ln Z_B$. Substitution leaves $\Delta s_S - \beta Q$. The right-hand side is

the dimensionless entropy production for this specified system–bath boundary. The relative entropy on the left can be split further:

$$D(\rho'_{SB} \| \rho'_S \otimes \rho_B^\beta) = I(S : B)_{\rho'} + D(\rho'_B \| \rho_B^\beta). \quad (12.14)$$

To verify the split, add and subtract $\ln \rho'_B$ inside the trace. The terms containing $\ln \rho'_{SB} - \ln \rho'_S - \ln \rho'_B$ give $I(S : B)_{\rho'}$; the remaining terms give $D(\rho'_B \| \rho_B^\beta)$. Entropy production therefore records both correlations created with the bath and the bath’s displacement from its reference Gibbs state.

Status: exact identity under a specified boundary

Equation (12.13) assumes an initially factorized state, a Gibbs reference bath, joint unitary evolution, and heat defined by the bath energy change. With appreciable endpoint interaction energy, the algebraic identity for that definition still holds, but interpreting Q as system heat and separating heat from work needs extra bookkeeping. Initial correlations change the identity itself.

12.4 Three false shortcuts and their counterexamples

12.4.1 “Entropy is missing information”

This phrase is usable only after the missing variables and reference description are named. For example, the pure qubit $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$ has von Neumann entropy zero, although a measurement in the $\{|0\rangle, |1\rangle\}$ basis has Shannon entropy $\ln 2$. The first quantity belongs to the state spectrum; the second belongs to the outcome distribution of a chosen measurement. For a macroscopic gas, the missing microscopic information likewise depends on the macrovariables used to define the macrostate.

12.4.2 “Entanglement is thermal entropy”

A subsystem of a pure entangled state can be exactly Gibbs for a chosen Hamiltonian, but this requires a special relation between its Schmidt spectrum and that Hamiltonian. On the support of a general reduced state, define the modular Hamiltonian $K_A = -\ln \rho_A$, so $\rho_A = e^{-K_A}$ there. Calling this thermal at a physical temperature is meaningful only when K_A is related to the physical energy and a thermodynamic protocol.

12.4.3 “Entropy always increases after discarding information”

Relative entropy between two preparations cannot increase when the same channel acts on both. The entropy of one output state has no such universal direction: a unital channel cannot lower finite-dimensional von Neumann entropy, while amplitude damping can. A deterministic block map can reduce Shannon entropy by shrinking the alphabet, while added noise can raise it. The operation and the entropy functional must both be stated.

12.5 Four diagnostic examples

12.5.1 Free expansion of an ideal gas

An equilibrium gas initially occupies half a container. Removing a partition doubles the accessible volume. The worked example in [section 1.6](#) obtains the same state-function change from a reversible path. Thermodynamics gives

$$\Delta S = Nk_B \ln 2. \quad (12.15)$$

Liouville evolution preserves the fine-grained Gibbs entropy of the phase-space distribution. The thermodynamic entropy compares the initial equilibrium macrostate with the final equilibrium macrostate that occupies the larger volume. Treating the evolved, filamented distribution as locally equilibrated discards the fine correlations needed to reverse the expansion. These statements refer to two different levels of description and therefore do not conflict.

12.5.2 A thermal qubit

For $H = \epsilon|1\rangle\langle 1|$,

$$\rho_\beta = \frac{|0\rangle\langle 0| + e^{-\beta\epsilon}|1\rangle\langle 1|}{1 + e^{-\beta\epsilon}}. \quad (12.16)$$

Its von Neumann entropy is also the Shannon entropy of the energy measurement outcomes because ρ_β is diagonal in the energy basis. Coherence breaks this equality. For instance, the pure state $(|0\rangle + |1\rangle)/\sqrt{2}$ has equal energy populations and energy-measurement entropy $\ln 2$, but its von Neumann entropy is zero.

12.5.3 A Bell pair

For $|\Phi^+\rangle = (|00\rangle + |11\rangle)/\sqrt{2}$, the joint state is pure, so $s(AB) = 0$. Tracing out either qubit gives $\rho_A = \rho_B = \mathbb{I}/2$, hence $s(A) = s(B) = \ln 2$. This local entropy is entanglement entropy, not thermodynamic entropy unless an additional physical setting establishes that relation.

12.5.4 A critical chain

Near a one-dimensional quantum critical point, interval entanglement grows logarithmically with size according to Eq. (11.56). The coefficient contains the central charge, while the additive constant depends on the cutoff. This is an information measure of ground-state correlations, not the thermal entropy of the zero-temperature pure state.

12.6 A decision procedure

Before writing an entropy symbol, ask:

1. What is the object: an equilibrium macrostate, a distribution, a density operator, a subsystem, or a trajectory ensemble?
2. Which measure, partition, or subsystem split defines it?
3. Is the quantity dimensionless or multiplied by k_B ?
4. Does the claimed equality hold exactly, at equilibrium, in a thermodynamic limit, or after an approximation?
5. If entropy increases, where did information go and what dynamics describes that loss?

These questions prevent most entropy paradoxes before any calculation begins.

12.7 Synthesis exercises

Exercise 12.1. For each entropy in the table, name its mathematical object, units, and behavior under reversible closed dynamics.

Exercise 12.2. A qubit has Hamiltonian $H = \epsilon\sigma_z/2$ and state ρ . Express its free-energy excess above equilibrium as a quantum relative entropy. Evaluate the result for a pure energy eigenstate.

Exercise 12.3. Compare free expansion using Boltzmann, canonical Gibbs, and fine-grained phase-space entropy. State why the numerical answers appear different and which physical description each uses.

Exercise 12.4. An isolated pure state of AB evolves unitarily while $s(A)$ increases. Construct a consistent entropy ledger for A , B , and AB .

Exercise 12.5 (*). Choose one result from each status class used in these notes: exact, thermodynamic limit, controlled approximation, mean field, model-specific exact, and heuristic. State the assumptions and describe one failure mode.

Part V

Appendices

Appendix A

Mathematical tools

This appendix supplies calculations used in several chapters. Results derived below are shown with their intermediate steps. When a theorem is only quoted, the text says so explicitly.

A.1 Stirling's formula

For large N , the Euler–Maclaurin expansion gives

$$\ln N! = N \ln N - N + \frac{1}{2} \ln(2\pi N) + \frac{1}{12N} + \mathcal{O}(N^{-3}). \quad (\text{A.1})$$

We quote the Euler–Maclaurin result rather than derive the summation formula here; Ref. [42] records the same expansion in the statistical-mechanics setting. The first two terms are extensive when N scales with system size. The logarithmic term supplies the leading finite-size correction.

For $m = Nx$, apply the formula separately to $N!$, $(Nx)!$, and $[N(1-x)]!$. The terms proportional to $N \ln N$ and N cancel:

$$\begin{aligned} \ln \binom{N}{Nx} &= N \ln N - Nx \ln(Nx) - N(1-x) \ln[N(1-x)] \\ &\quad + \frac{1}{2} \ln(2\pi N) - \frac{1}{2} \ln(2\pi Nx) - \frac{1}{2} \ln[2\pi N(1-x)] + \mathcal{O}(N^{-1}) \end{aligned} \quad (\text{A.2})$$

$$= Nh_2(x) - \frac{1}{2} \ln[2\pi Nx(1-x)] + \mathcal{O}(N^{-1}). \quad (\text{A.3})$$

Here $h_2(x) = -x \ln x - (1-x) \ln(1-x)$. This calculation is valid when both Nx and $N(1-x)$ grow; it is not uniform at the endpoints $x = 0, 1$.

A.2 Gaussian integrals

For $a > 0$,

$$I(a) = \int_{-\infty}^{\infty} e^{-ax^2} dx = \sqrt{\frac{\pi}{a}}. \quad (\text{A.4})$$

The standard proof squares the integral and changes to polar coordinates:

$$I(a)^2 = \int_{\mathbb{R}^2} dx dy e^{-a(x^2+y^2)} = 2\pi \int_0^{\infty} r e^{-ar^2} dr = \frac{\pi}{a}. \quad (\text{A.5})$$

The integral is positive, so taking the positive square root proves Eq. (A.4).

Differentiation with respect to a generates even moments:

$$(-\partial_a)^n I(a) = \int_{-\infty}^{\infty} x^{2n} e^{-ax^2} dx. \quad (\text{A.6})$$

In n dimensions, a real symmetric positive-definite matrix A can be diagonalized by an orthogonal change of variables. The integral then factorizes into one-dimensional Gaussian integrals:

$$\int_{\mathbb{R}^n} d^n x \exp\left(-\frac{1}{2} \mathbf{x}^\top A \mathbf{x}\right) = \frac{(2\pi)^{n/2}}{\sqrt{\det A}}. \quad (\text{A.7})$$

For a normalized one-dimensional Gaussian with mean zero and variance σ^2 ,

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-x^2/(2\sigma^2)}, \quad (\text{A.8})$$

the generating function is

$$\langle e^{tx} \rangle = e^{\sigma^2 t^2/2}. \quad (\text{A.9})$$

The calculation is a completion of the square:

$$-\frac{x^2}{2\sigma^2} + tx = -\frac{(x - \sigma^2 t)^2}{2\sigma^2} + \frac{\sigma^2 t^2}{2}. \quad (\text{A.10})$$

The shifted Gaussian still integrates to one. Expanding the generating function in powers of t and comparing coefficients shows that all odd moments vanish and that

$$\langle x^{2n} \rangle = (2n-1)!! \sigma^{2n}. \quad (\text{A.11})$$

The double factorial means $(2n-1)!! = (2n-1)(2n-3)\cdots 3 \cdot 1$. Odd moments vanish because the integrand is odd under $x \mapsto -x$. These formulas underlie the Maxwell distribution, canonical fluctuations, and Gaussian field theory.

A.3 Gamma, zeta, and Bose–Fermi integrals

The Gamma function is

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx, \quad s > 0. \quad (\text{A.12})$$

Integration by parts gives the recurrence

$$\Gamma(s+1) = \int_0^\infty x^s e^{-x} dx = s \int_0^\infty x^{s-1} e^{-x} dx = s\Gamma(s), \quad (\text{A.13})$$

where the boundary term vanishes for $s > 0$. Therefore

$$\Gamma(n+1) = n!, \quad \Gamma(1/2) = \sqrt{\pi}. \quad (\text{A.14})$$

The factorial identity follows by repeating the recurrence from $\Gamma(1) = 1$. For the half-integer value, set $x = t^2$ and use Eq. (A.4) with $a = 1$:

$$\Gamma(1/2) = 2 \int_0^\infty e^{-t^2} dt = \sqrt{\pi}. \quad (\text{A.15})$$

The substitution $y = ax$ gives

$$\int_0^\infty x^{s-1} e^{-ax} dx = \frac{\Gamma(s)}{a^s}, \quad a > 0. \quad (\text{A.16})$$

For $\Re(s) > 1$, the Riemann zeta function has both a series and a Bose integral:

$$\zeta(s) = \sum_{\ell=1}^{\infty} \ell^{-s}, \quad (\text{A.17})$$

$$\int_0^{\infty} \frac{x^{s-1}}{e^x - 1} dx = \Gamma(s) \zeta(s). \quad (\text{A.18})$$

The proof expands $(e^x - 1)^{-1} = \sum_{\ell \geq 1} e^{-\ell x}$ and uses the rescaled Gamma integral term by term. For $\Re(s) > 1$ the terms are nonnegative when s is real, and Tonelli's theorem justifies the interchange of sum and integral; the complex case follows by absolute convergence. The Fermi analogue is

$$\int_0^{\infty} \frac{x^{s-1}}{e^x + 1} dx = (1 - 2^{1-s}) \Gamma(s) \zeta(s), \quad (\text{A.19})$$

obtained from

$$\frac{1}{e^x + 1} = \sum_{\ell=1}^{\infty} (-1)^{\ell-1} e^{-\ell x}, \quad \sum_{\ell=1}^{\infty} \frac{(-1)^{\ell-1}}{\ell^s} = (1 - 2^{1-s}) \zeta(s). \quad (\text{A.20})$$

At nonzero fugacity, define

$$g_s(z) = \sum_{\ell=1}^{\infty} \frac{z^{\ell}}{\ell^s}. \quad (\text{A.21})$$

For $|z| < 1$ this is the polylogarithm $\text{Li}_s(z)$. Useful identities are

$$z \frac{\partial g_s(z)}{\partial z} = g_{s-1}(z), \quad g_s(1) = \zeta(s) \quad (s > 1). \quad (\text{A.22})$$

The divergence of $g_s(1)$ for $s \leq 1$ is the analytic form of the infrared obstruction to homogeneous ideal-gas condensation in low dimension.

A.4 Laplace's method

Let $f(x)$ have a unique interior maximum at x_* , with $f''(x_*) < 0$. Let g be continuous and nonnegative, with $g(x_*) > 0$. Assume the contribution outside a fixed neighborhood of x_* is exponentially smaller than the contribution from that neighborhood. Then

$$f(x) = f(x_*) - \frac{|f''(x_*)|}{2} (x - x_*)^2 + \mathcal{O}((x - x_*)^3). \quad (\text{A.23})$$

The exponential confines the relevant region to $x - x_* = \mathcal{O}(N^{-1/2})$. Across that region, $g(x) = g(x_*) + o(1)$. Substituting Eq. (A.23) and using the Gaussian integral yields

$$\int dx g(x) e^{Nf(x)} \sim g(x_*) e^{Nf(x_*)} \sqrt{\frac{2\pi}{N |f''(x_*)|}}. \quad (\text{A.24})$$

The symbol $\sim$ means that the ratio of the two sides approaches one as $N \rightarrow \infty$ under the assumptions just stated. Consequently

$$\frac{1}{N} \ln \int dx g(x) e^{Nf(x)} \longrightarrow f(x_*). \quad (\text{A.25})$$

This logarithmic form explains why equilibrium thermodynamics keeps the dominant macrostate while finite-size physics retains the prefactor.

For several variables $\mathbf{x} \in \mathbb{R}^n$, let the Hessian of $-f$ at the maximum be the positive matrix A . Then

$$\int d^n x g(\mathbf{x}) e^{Nf(\mathbf{x})} \sim g(\mathbf{x}_\star) e^{Nf(\mathbf{x}_\star)} \left(\frac{2\pi}{N} \right)^{n/2} \frac{1}{\sqrt{\det A}}. \quad (\text{A.26})$$

If the nonnegative integrand is normalized to a probability density, the inverse Hessian sets the covariance of fluctuations:

$$\text{Cov}(x_i, x_j) \simeq \frac{1}{N} (A^{-1})_{ij}. \quad (\text{A.27})$$

Thus the same curvature that determines thermodynamic stability determines the leading Gaussian fluctuations.

If two isolated maxima have equal height, both saddle contributions must be retained. Their relative prefactors affect finite-size probabilities even though $(1/N) \ln Z$ keeps only their common height. This is the saddle-point version of phase coexistence. If the maximum lies at an integration boundary or has $f''(\mathbf{x}_\star) = 0$, the Gaussian formula fails and a different scaling analysis is required; critical points are precisely where such naive curvature expansions become unreliable.

Worked example: binomial fluctuations from the saddle

For a fair binary variable, the probability of $m = Nx$ contains $\binom{N}{Nx} 2^{-N}$. Around $x = 1/2 + \delta$,

$$h_2(1/2 + \delta) = \ln 2 - 2\delta^2 - \frac{4}{3}\delta^4 + \dots. \quad (\text{A.28})$$

Keeping the quadratic term gives

$$P_N(Nx) \propto e^{-2N\delta^2}. \quad (\text{A.29})$$

This is a Gaussian with $\text{Var}(x) = 1/(4N)$, or $\text{Var}(m) = N/4$, as obtained directly from independent Bernoulli variables. The quartic term controls the first non-Gaussian correction.

A.5 Legendre and Legendre–Fenchel transforms

For a differentiable convex function $\phi(k)$, its convex conjugate is

$$\phi^\star(x) = \sup_k \{kx - \phi(k)\}. \quad (\text{A.30})$$

At a smooth interior optimum k_x , differentiate the quantity being maximized:

$$0 = \frac{\partial}{\partial k} [kx - \phi(k)]_{k=k_x} = x - \phi'(k_x). \quad (\text{A.31})$$

Thus $x = \phi'(k_x)$ and $\phi^\star(x) = k_x x - \phi(k_x)$. The supremum definition remains valid when ϕ is not strictly convex or differentiable, even though this stationary-point calculation may then fail.

For a concave entropy density $s(e)$, thermodynamics uses

$$-\beta f(\beta) = \sup_e \left[\frac{s(e)}{k_B} - \beta e \right]. \quad (\text{A.32})$$

Under the usual closedness conditions, the Legendre–Fenchel biconjugation theorem returns the concave envelope of s . With the supremum convention used here, this is the least concave majorant: the smallest concave function that lies nowhere below s . A nonconcave segment is therefore invisible to the canonical ensemble. We quote the convex-analysis theorem; Ref. [58] explains its statistical-mechanical use.

A.6 Delta functions and density of states

If the zeros x_i of f are isolated and simple, meaning $f'(x_i) \neq 0$, then

$$\delta[f(x)] = \sum_{x_i: f(x_i)=0} \frac{\delta(x - x_i)}{|f'(x_i)|} \quad (\text{A.33})$$

converts momentum integrals into energy integrals. In d dimensions,

$$\int d^d p = S_{d-1} \int_0^\infty p^{d-1} dp, \quad S_{d-1} = \frac{2\pi^{d/2}}{\Gamma(d/2)}. \quad (\text{A.34})$$

For $\epsilon = Ap^s$ with $A > 0$, $s > 0$, and $\epsilon > 0$, let $p_\epsilon = (\epsilon/A)^{1/s}$. Apart from volume and degeneracy factors, the density of states is

$$D(\epsilon) \propto \int_0^\infty p^{d-1} \delta(\epsilon - Ap^s) dp \quad (\text{A.35})$$

$$= \frac{p_\epsilon^{d-1}}{As p_\epsilon^{s-1}} \propto \epsilon^{d/s-1}. \quad (\text{A.36})$$

The denominator is the Jacobian $|d(Ap^s)/dp|_{p=p_\epsilon}$.

A.7 Fourier transforms and Gaussian correlations

We use the continuum convention

$$\phi(\mathbf{x}) = \int \frac{d^d k}{(2\pi)^d} e^{i\mathbf{k} \cdot \mathbf{x}} \phi_{\mathbf{k}}, \quad (\text{A.37})$$

$$\phi_{\mathbf{k}} = \int d^d x e^{-i\mathbf{k} \cdot \mathbf{x}} \phi(\mathbf{x}). \quad (\text{A.38})$$

Then

$$\int d^d x (\nabla \phi)^2 = \int \frac{d^d k}{(2\pi)^d} k^2 |\phi_{\mathbf{k}}|^2. \quad (\text{A.39})$$

For the Gaussian functional

$$\mathcal{F}[\phi] = \frac{1}{2} \int \frac{d^d k}{(2\pi)^d} (ck^2 + r) |\phi_{\mathbf{k}}|^2, \quad (\text{A.40})$$

each independent real quadratic mode contributes according to equipartition, giving

$$\langle \phi_{\mathbf{k}} \phi_{\mathbf{q}} \rangle = (2\pi)^d \delta^{(d)}(\mathbf{k} + \mathbf{q}) \frac{k_B T}{ck^2 + r}. \quad (\text{A.41})$$

The denominator can be written $c(k^2 + \xi^{-2})$ with $\xi = \sqrt{c/r}$. In three dimensions,

$$\int \frac{d^3 k}{(2\pi)^3} \frac{e^{i\mathbf{k} \cdot \mathbf{r}}}{k^2 + \xi^{-2}} = \frac{e^{-r/\xi}}{4\pi r}. \quad (\text{A.42})$$

To see the factors, align the polar axis with $\mathbf{r}$ and perform the angular integral:

$$\int \frac{d^3 k}{(2\pi)^3} \frac{e^{i\mathbf{k} \cdot \mathbf{r}}}{k^2 + \xi^{-2}} = \frac{1}{2\pi^2 r} \int_0^\infty dk \frac{k \sin(kr)}{k^2 + \xi^{-2}} \quad (\text{A.43})$$

$$= \frac{1}{2\pi^2 r} \frac{\pi}{2} e^{-r/\xi}. \quad (\text{A.44})$$

The last one-dimensional integral follows by closing the contour in the upper half-plane for $r > 0$; its pole is at $k = i/\xi$. In general dimension the large- r behavior is an exponential cutoff times a dimension-dependent power. At $r = 0$ the continuum integral can diverge at large k ; a lattice spacing or momentum cutoff remembers the microscopic theory.

A.8 Cumulants, sources, and response

For a partition function with source J coupled to an observable A ,

$$Z(J) = \text{Tr} \exp[-\beta H + JA], \quad (\text{A.45})$$

derivatives of $W(J) = \ln Z(J)$ generate connected moments when the operators commute or in the corresponding classical problem:

$$W'(0) = \langle A \rangle, \quad W''(0) = \langle A^2 \rangle - \langle A \rangle^2. \quad (\text{A.46})$$

For several sources, mixed derivatives give connected cross-correlations. In a single-variable setting, suppose $a = W'(J)$ is locally invertible and $W''(J) \neq 0$, and define $\Gamma(a) = Ja - W(J)$. Differentiating while remembering that J depends on a gives

$$\Gamma'(a) = J, \quad \Gamma''(a) = \frac{dJ}{da} = \frac{1}{W''(J)}. \quad (\text{A.47})$$

Thus the curvature of the Legendre transform is the inverse susceptibility. A flat thermodynamic potential accompanies a large response near criticality.

For noncommuting quantum observables, derivatives generate imaginary-time ordered correlations rather than an ordinary product. The first derivative of $\ln Z$ remains simple by cyclicity of the trace, but the second derivative must be handled with the Duhamel formula below.

A.9 Operator derivatives and traces

For noncommuting $H(\lambda)$ and $\partial_\lambda H$, one cannot generally write

$$\partial_\lambda e^{-\beta H} \neq -\beta(\partial_\lambda H)e^{-\beta H} \quad \text{in general.}$$

The correct identity is the Duhamel formula

$$\partial_\lambda e^{-\beta H} = - \int_0^\beta ds e^{-(\beta-s)H} (\partial_\lambda H) e^{-sH}. \quad (\text{A.48})$$

It follows directly from the power series. For a general operator $A(\lambda)$, expand the proposed right-hand side:

$$\int_0^1 du e^{(1-u)A} (\partial_\lambda A) e^{uA} = \sum_{m,n \geq 0} \frac{A^m (\partial_\lambda A) A^n}{m!n!} \int_0^1 (1-u)^m u^n du \quad (\text{A.49})$$

$$= \sum_{m,n \geq 0} \frac{A^m (\partial_\lambda A) A^n}{(m+n+1)!}. \quad (\text{A.50})$$

This is exactly the term-by-term derivative of $\sum_{r \geq 0} A^r / r!$: for each power A^r , the derivative can occupy any one of the r positions. Setting $A = -\beta H$ and $s = \beta u$ proves Eq. (A.48).

Cyclicity of the trace removes the s dependence after tracing:

$$\partial_\lambda Z = - \int_0^\beta ds \text{Tr}[e^{-(\beta-s)H} (\partial_\lambda H) e^{-sH}] \quad (\text{A.51})$$

$$= -\beta \text{Tr}[(\partial_\lambda H) e^{-\beta H}]. \quad (\text{A.52})$$

Division by Z gives the partition-function derivative

$$\partial_\lambda \ln Z = -\beta \langle \partial_\lambda H \rangle_\beta. \quad (\text{A.53})$$

For a perturbation $H_f = H - fA$, differentiate the normalized expectation $\langle B \rangle_f = Z_f^{-1} \text{Tr}(B e^{-\beta H_f})$. The derivative of the numerator comes from Eq. (A.48); the derivative of Z_f^{-1} subtracts the product of the two means. At $f = 0$ this gives the static response

$$\chi_{BA} = \left. \frac{\partial \langle B \rangle_f}{\partial f} \right|_{f=0} = \int_0^\beta ds \left\langle e^{sH} \delta A e^{-sH} \delta B \right\rangle_\beta, \quad \delta A = A - \langle A \rangle_\beta, \quad \delta B = B - \langle B \rangle_\beta. \quad (\text{A.54})$$

This is the Kubo–Mori, or Duhamel, two-point function. If A , B , and H commute, the integrand is independent of s and Eq. (A.54) reduces to $\chi_{BA} = \beta \text{Cov}(B, A)$.

A.10 Schmidt decomposition

Write a bipartite pure state as

$$|\psi\rangle = \sum_{ij} C_{ij} |i\rangle_A |j\rangle_B. \quad (\text{A.55})$$

Apply the singular-value decomposition $C = U \sqrt{\Lambda} V^\dagger$, where $\Lambda = \text{diag}(\lambda_a)$ and the singular values imply $\lambda_a \geq 0$. Absorbing U and V^* into new local bases gives

$$|\psi\rangle = \sum_a \sqrt{\lambda_a} |a\rangle_A |a\rangle_B. \quad (\text{A.56})$$

Normalization gives $\sum_a \lambda_a = 1$. Taking either partial trace explicitly gives

$$\rho_A = \sum_a \lambda_a |a\rangle_A \langle a|_A, \quad \rho_B = \sum_a \lambda_a |a\rangle_B \langle a|_B. \quad (\text{A.57})$$

The reduced states therefore have the same nonzero eigenvalues λ_a .

A.11 Useful inequalities

For $x > 0$,

$$\ln x \leq x - 1, \quad (x - y) \ln(x/y) \geq 0 \quad (x, y > 0). \quad (\text{A.58})$$

The first inequality follows because $x - 1 - \ln x$ has its unique minimum at $x = 1$. The second follows from the monotonicity of the logarithm: $x - y$ and $\ln x - \ln y$ always have the same sign.

Jensen’s inequality for a differentiable convex function φ is

$$\varphi(\langle X \rangle) \leq \langle \varphi(X) \rangle. \quad (\text{A.59})$$

To prove it, put $m = \langle X \rangle$. The tangent-line property of a convex function gives $\varphi(x) \geq \varphi(m) + \varphi'(m)(x - m)$. Taking the expectation removes the last term and leaves Jensen’s inequality. Taking $\varphi(x) = e^x$ converts the Jarzynski equality into the average-work form of the second law.

A.11.1 Log-sum inequality and classical data processing

For nonnegative numbers a_i, b_i , the log-sum inequality is

$$\sum_i a_i \ln \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \ln \frac{\sum_i a_i}{\sum_i b_i}. \quad (\text{A.60})$$

To prove it, first suppose $A = \sum_i a_i > 0$ and $B = \sum_i b_i > 0$, and set $p_i = a_i/A$ and $q_i = b_i/B$. The difference between the two sides is $AD(p||q) \geq 0$. We use the continuous conventions

$0 \ln(0/b) = 0$ for $b \geq 0$, including $0 \ln(0/0) = 0$. If $A = 0$, both sides vanish; if $B = 0 < A$, both sides are infinite under the support convention for relative entropy. The remaining boundary cases follow by continuity.

Let $T_{y|x}$ be a stochastic matrix and define $p'_y = \sum_x T_{y|x} p_x$, with the analogous definition for q' . Apply (A.60) separately for each y , using $a_x = T_{y|x} p_x$ and $b_x = T_{y|x} q_x$. Terms with $T_{y|x} = 0$ contribute zero and are omitted from the logarithmic form below. Summing over y gives

$$D(p||q) = \sum_{\substack{x,y \\ T_{y|x} > 0}} T_{y|x} p_x \ln \frac{T_{y|x} p_x}{T_{y|x} q_x} \quad (\text{A.61})$$

$$\geq \sum_y p'_y \ln \frac{p'_y}{q'_y} = D(Tp||Tq). \quad (\text{A.62})$$

This proof shows both the inequality and the role of normalization. It also covers deterministic coarse graining as a special stochastic matrix.

A.11.2 Norms and distinguishability

For a classical signed vector v , $\|v\|_1 = \sum_i |v_i|$. The total-variation distance is

$$\text{TV}(p, q) = \frac{1}{2} \|p - q\|_1. \quad (\text{A.63})$$

For density operators, the trace distance is

$$T(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1, \quad \|X\|_1 = \text{Tr} \sqrt{X^\dagger X}. \quad (\text{A.64})$$

It equals the largest total-variation distance produced by a two-outcome measurement and is therefore an operational measure of one-shot distinguishability. Channels contract trace distance. These are quoted finite-dimensional theorems; proofs are given in Refs. [39, 60]. Relative entropy is more sensitive to support and governs asymptotic discrimination rates, so the two quantities should not be interchanged.

Pinsker's inequality, with natural logarithms, reads

$$D(p||q) \geq 2 \text{TV}(p, q)^2. \quad (\text{A.65})$$

The quantum version replaces total variation by trace distance. We quote both forms here; Refs. [9, 60] give proofs. Thus small relative entropy guarantees small distinguishability, though the converse requires additional assumptions.

A.12 Partial traces and operator logarithms

For product-basis operators,

$$\text{Tr}_B(|i\rangle\langle j|_A \otimes |\mu\rangle\langle \nu|_B) = \delta_{\mu\nu} |i\rangle\langle j|_A. \quad (\text{A.66})$$

Linearity determines the partial trace on every finite-dimensional operator. It preserves trace and positivity but not, by itself, the spectrum.

For a positive operator

$$\rho = \sum_i \lambda_i |i\rangle\langle i|, \quad (\text{A.67})$$

spectral calculus defines

$$\ln \rho = \sum_{\lambda_i > 0} (\ln \lambda_i) |i\rangle \langle i|. \quad (\text{A.68})$$

The convention $0 \ln 0 = 0$ makes $-\text{Tr} \rho \ln \rho$ finite in finite dimension. In $D(\rho \|\sigma)$, however, a component of ρ outside the support of σ encounters $\ln 0$ and makes the relative entropy infinite.

If $\rho(\lambda)$ is differentiable and full rank,

$$\frac{d}{d\lambda} \text{Tr}[\rho \ln \rho] = \text{Tr}[(\partial_\lambda \rho)(\ln \rho + \mathbb{I})]. \quad (\text{A.69})$$

For trace-preserving variations, $\text{Tr}(\partial_\lambda \rho) = 0$, leaving only the $\ln \rho$ term. This is the step used in the constrained variation that produces a Gibbs state.

A.13 Constants and scale checks

Quantity	Symbol	Value
Boltzmann constant	k_B	$1.380\,649 \times 10^{-23} \text{ J K}^{-1}$
Planck constant	h	$6.626\,070\,15 \times 10^{-34} \text{ J s}$
Reduced Planck constant	$\hbar$	$1.054\,571\,817 \times 10^{-34} \text{ J s}$
Speed of light	c	$2.997\,924\,58 \times 10^8 \text{ m s}^{-1}$
Electron mass	m_e	$9.109\,383\,7 \times 10^{-31} \text{ kg}$
Atomic mass unit	m_u	$1.660\,539\,1 \times 10^{-27} \text{ kg}$

Exact defining values are shown to a useful number of digits. Before trusting a numerical result, check at least one independent scale: $k_B T$ has energy units, λ_T has length units, $n\lambda_T^3$ is dimensionless, and a density of states $D(\epsilon)$ has units of inverse energy. Dimensional analysis catches many missing factors before any comparison with data.

Appendix B

Hints for selected exercises

These hints identify a first step. They are deliberately shorter than solutions.

Chapter 1

For the stability problem, write the second variation of the total entropy of two subsystems exchanging E or V . Express the curvature through C_V and κ_T only after establishing the sign required at the maximum.

Chapter 2

For data processing, apply the log-sum inequality separately to each output value of the stochastic matrix. For the biased-walk rate function, first calculate $\lambda(k) = \ln(pe^k + qe^{-k})$.

Chapter 3

For a nonconcave entropy, the common tangent touches the entropy curve at two energies. Its slope is the coexistence inverse temperature; the energy gap is the latent heat per particle.

Chapter 4

In the Mayer problem, expand the configurational integral through one factor f_{ij} . When taking $\ln Z$, disconnected products cancel against the exponential of lower-order terms.

Chapter 5

For the sudden oscillator quench, $W = (k_1 - k_0)x^2/2$ and the initial x distribution is Gaussian. The work-distribution crossing follows by integrating Crooks' trajectory relation over paths with fixed work. For the Markov-chain proof of Crooks' relation, write each path probability as its initial equilibrium weight times a product of transition probabilities. Divide the forward product by the reversed product. Local detailed balance supplies the heat contribution from the transition factors; the endpoint Gibbs weights supply the free-energy difference. The first law then combines heat and the system-energy change into work.

Chapter 6

For the N -chamber engine, remove only partitions that separate empty regions; the occupied chamber then expands from V/N to V . Given the record the molecule's entropy grows, while its marginal distribution is uniform over the box before and after the expansion.

For the memory with a potential offset, compute each local free energy from the configurational integral of $e^{-\beta U}$ over the cell. The offset changes the energy of C_1 , while the volume of each cell sets its internal entropy.

For the off-center partition, give each side a final volume proportional to its posterior probability. The work for the record m is then a relative entropy, and its average is Eq. (6.30).

For the entropy ledger with a noisy measurement, keep the controller memory explicit after measurement and use the joint entropy $S_{XM} = S_X + S_M - k_B I(X : M)$. Spontaneous errors raise the joint entropy of molecule and memory by $k_B h_2(\epsilon)$ with no compensating change in the bath; the extra entropy is correlation with the source of the errors, which the ledger discards. Do not reset the memory until the expansion work has been extracted.

For the erasure of a biased bit, move the partition before removing it, until the probability of each side is proportional to its volume. The density of the molecule is then uniform, and removing the partition changes nothing.

For the asymmetric memory, the local free energy of a cell of volume V_m holding one molecule is $-k_B T \ln V_m$ plus a constant common to both cells.

Chapter 7

For local data processing of mutual information, use

$$I(A : B) = D(\rho_{AB} \| \rho_A \otimes \rho_B) \quad (\text{B.1})$$

and apply the channel $\text{id}_A \otimes \mathcal{N}_B$ to both arguments.

Chapter 8

For the exchange virial coefficient, use

$$-\ln(1-x) = \sum_{\ell \geq 1} \frac{x^\ell}{\ell} \quad \text{for bosons,} \quad \ln(1+x) = \sum_{\ell \geq 1} \frac{(-1)^{\ell+1} x^\ell}{\ell} \quad \text{for fermions.}$$

Perform the Gaussian momentum integral term by term and eliminate z in favor of density.

Chapter 9

For exponent relations, choose $b = |t|^{-1/y_t}$ in the scaling form and differentiate with respect to h . The correlation relation follows by integrating $G(r)$ up to $r \sim \xi$.

Chapter 10

For the numerical exercise, discard an equilibration interval and estimate autocorrelation times between successive recorded sweeps. Error bars based on treating every recorded sweep as independent will be too small near T_c .

Chapter 11

For the finite-circle entanglement form, check $\sin[\pi(L - \ell)/L] = \sin(\pi\ell/L)$ and $\sin(\pi\ell/L) \simeq \pi\ell/L$ when $\ell \ll L$. These checks establish the required symmetry and short-interval limit. The sine dependence itself is quoted conformal-field-theory input; periodicity and dimensional analysis do not derive it.

Chapter 12

For the status-class exercise, a suitable set is: free-energy identity, ensemble equivalence, Sommerfeld expansion, Landau exponents, one-dimensional Ising transfer matrix, and the mixing picture of coarse-grained relaxation.

Short checks for selected calculations

These results are supplied so that a student can check an intermediate calculation. They are not substitutes for the derivations.

Thermodynamics

For two bodies with constant heat capacities C_A and C_B in an isolated enclosure,

$$T_f = \frac{C_A T_A + C_B T_B}{C_A + C_B}, \quad (\text{B.2})$$

and

$$\Delta S = C_A \ln \frac{T_f}{T_A} + C_B \ln \frac{T_f}{T_B} \geq 0. \quad (\text{B.3})$$

Set $w_A = C_A/(C_A + C_B)$ and $w_B = 1 - w_A$, so $T_f = w_A T_A + w_B T_B$. Then

$$\frac{\Delta S}{C_A + C_B} = \ln T_f - w_A \ln T_A - w_B \ln T_B \geq 0$$

by concavity of the logarithm.

Probability and information

For a fair input bit sent through a binary symmetric channel with error probability ϵ ,

$$I(X : Y) = \ln 2 - h_2(\epsilon). \quad (\text{B.4})$$

It equals $\ln 2$ at $\epsilon = 0$, vanishes at $\epsilon = 1/2$, and rises again if $\epsilon > 1/2$ because a reliably flipped bit can be corrected.

For the binomial rate function,

$$D(x||p) = \frac{(x-p)^2}{2p(1-p)} + \mathcal{O}((x-p)^3). \quad (\text{B.5})$$

The exponent ND therefore gives variance $p(1-p)/N$ for the fraction x .

Ensembles

For a one-dimensional classical harmonic oscillator,

$$Z = \frac{1}{h} \int dx dp e^{-\beta[p^2/(2m) + m\omega^2 x^2/2]} = \frac{1}{\beta\hbar\omega}. \quad (\text{B.6})$$

Hence $U = k_B T$ and $C_V = k_B$. The quantum partition function is

$$Z_q = \frac{e^{-\beta\hbar\omega/2}}{1 - e^{-\beta\hbar\omega}}, \quad (\text{B.7})$$

whose high-temperature limit reproduces the classical result.

Classical and quantum gases

For mixing N_A and N_B particles of two ideal species at equal T and P ,

$$\Delta S_{\text{mix}} = -k_B (N_A \ln x_A + N_B \ln x_B), \quad x_i = \frac{N_i}{N_A + N_B}. \quad (\text{B.8})$$

For the same species, removal of a fictitious label makes the physical entropy of mixing zero.

For a dispersion $\epsilon = Ap^s$ in dimension d ,

$$D(\epsilon) \propto \epsilon^{d/s-1}. \quad (\text{B.9})$$

At $\mu = 0$, the Bose number integral behaves as $\int_0 \epsilon^{d/s-2} d\epsilon$ and converges only for $d > s$.

Information thermodynamics

For a memory with $E_0 = 0$, $E_1 = \epsilon$ and initial distribution $(1 - p, p)$, resetting to the pure state 0 obeys

$$W_{\text{on}} \geq -p\epsilon + k_B T h_2(p) \quad (\text{B.10})$$

in the ideal nonequilibrium free-energy accounting. A negative lower bound means that the energetic bias can, in principle, supply work during reset; it does not require work extraction in every protocol.

Phase transitions and Ising models

For the van der Waals equation,

$$V_c = 3Nb, \quad k_B T_c = \frac{8a}{27b}, \quad P_c = \frac{a}{27b^2}. \quad (\text{B.11})$$

For the one-dimensional zero-field Ising chain,

$$\frac{U}{N} = -J \tanh K, \quad \frac{C_V}{Nk_B} = K^2 \text{sech}^2 K. \quad (\text{B.12})$$

The Ising energy and heat capacity are analytic at every $T > 0$.

Renormalization

At the Gaussian fixed point, $\Delta_\phi = (d - 2)/2$. A coupling $g_{2n}\phi^{2n}$ has eigenvalue

$$y_{2n} = d - n(d - 2). \quad (\text{B.13})$$

For ϕ^6 , $y_6 = 6 - 2d$, so the interaction is marginal at $d = 3$. This is power counting at the Gaussian fixed point; an interacting tricritical fixed point requires the corresponding RG analysis.

Appendix C

Formula and assumption summary

This appendix is a retrieval aid. A formula is safe to use only with the assumptions printed beside it; the chapter references contain the derivations and the definitions of all symbols.

C.1 Thermodynamics

For a simple one-component equilibrium system,

$$dE = T dS - P dV + \mu dN, \quad (\text{C.1})$$

$$dF = -S dT - P dV + \mu dN, \quad (\text{C.2})$$

$$dG = -S dT + V dP + \mu dN, \quad (\text{C.3})$$

$$d\Phi = -S dT - P dV - N d\mu. \quad (\text{C.4})$$

The natural variables are respectively (S, V, N) , (T, V, N) , (T, P, N) , and (T, V, μ) . For a homogeneous extensive phase,

$$E = TS - PV + \mu N, \quad \Phi = -PV, \quad S dT - V dP + N d\mu = 0. \quad (\text{C.5})$$

Extensivity can fail for long-range interactions, surface-dominated systems, and constrained nonadditive models.

With heat signed positive into a cyclic device, the cyclic Clausius inequality is

$$\sum_{\alpha} \frac{Q_{\alpha}}{T_{\alpha}} \leq 0, \quad (\text{C.6})$$

with equality for a reversible cycle. For any process between equilibrium endpoints,

$$S(B) - S(A) \geq \int_A^B \frac{\delta Q}{T_b}. \quad (\text{C.7})$$

Here T_b is the temperature of the reservoir or boundary through which the heat crosses. Only when an entropy state can be assigned at every instant may one use the local shorthand

$$dS \geq \frac{\delta Q}{T_b}, \quad dS = \frac{\delta Q}{T_b} + d_i S, \quad d_i S \geq 0. \quad (\text{C.8})$$

With several simultaneous heat contacts, replace the exchange term by $\sum_{\alpha} \delta Q_{\alpha}/T_{\alpha}$.

For an ordinary stable additive phase in canonical or isothermal mechanical contact, useful response definitions and stability signs are

$$C_V = \left(\frac{\partial E}{\partial T} \right)_{V,N} > 0, \quad \kappa_T = -\frac{1}{V} \left(\frac{\partial V}{\partial P} \right)_{T,N} > 0. \quad (\text{C.9})$$

At a first-order coexistence line,

$$\frac{dP}{dT} = \frac{\Delta s}{\Delta v} = \frac{L}{T \Delta v}. \quad (\text{C.10})$$

C.2 Equilibrium ensembles

The microcanonical, canonical, and grand-canonical weights are

$$\rho_{\text{mc}} \propto \mathbf{1}_{[E, E+\Delta E]}(H), \quad (\text{C.11})$$

$$p_i^{\text{can}} = \frac{e^{-\beta E_i}}{Z}, \quad Z = \sum_i e^{-\beta E_i}, \quad (\text{C.12})$$

$$p_i^{\text{gc}} = \frac{e^{-\beta(E_i - \mu N_i)}}{\Xi}, \quad \Xi = \sum_i e^{-\beta(E_i - \mu N_i)}. \quad (\text{C.13})$$

Their thermodynamic potentials are

$$S_B = k_B \ln \Omega, \quad F = -k_B T \ln Z, \quad \Phi = -k_B T \ln \Xi. \quad (\text{C.14})$$

For a temperature-independent Hamiltonian, canonical derivatives give

$$U = -\partial_\beta \ln Z, \quad \text{Var}(E) = \partial_\beta^2 \ln Z = k_B T^2 C_V. \quad (\text{C.15})$$

Grand-canonical number fluctuations obey

$$\text{Var}(N) = \frac{1}{\beta} \left(\frac{\partial \langle N \rangle}{\partial \mu} \right)_{T, V}. \quad (\text{C.16})$$

The fluctuation identities are exact inside the stated ensemble. Relative fluctuations scale as $N^{-1/2}$ only with ordinary extensivity and away from critical singularities.

For any normalized classical distribution p over energy states, define $\mathcal{F}_T[p] = \sum_i p_i E_i - k_B T h(p)$. Then, for $T \neq 0$,

$$\mathcal{F}_T[p] - F_{\text{eq}} = k_B T D(p \| p^{\text{eq}}). \quad (\text{C.17})$$

For $T > 0$ the right-hand side is nonnegative, so the Gibbs distribution minimizes $\mathcal{F}_T$. For a normalizable Gibbs distribution at $T < 0$, the inequality reverses.

C.3 Classical ideal and dilute gases

For N identical monatomic particles in three dimensions,

$$Z_N = \frac{1}{N!} \left(\frac{V}{\lambda_T^3} \right)^N, \quad \lambda_T = \frac{h}{\sqrt{2\pi m k_B T}}, \quad (\text{C.18})$$

$$P = n k_B T, \quad U = \frac{3}{2} N k_B T, \quad (\text{C.19})$$

$$S = N k_B \left[\ln \left(\frac{V}{N \lambda_T^3} \right) + \frac{5}{2} \right]. \quad (\text{C.20})$$

The semiclassical dilute condition is $n \lambda_T^3 \ll 1$. The displayed entropy is the Sackur–Tetrode expression after Stirling's approximation. It is accurate to leading extensive order; subextensive corrections of order $k_B \ln N$ have been omitted.

The Maxwell speed density and its three standard scales are

$$f(v) = 4\pi v^2 \left(\frac{m}{2\pi k_B T} \right)^{3/2} e^{-mv^2/(2k_B T)}, \quad (\text{C.21})$$

$$v_{\text{mp}} = \sqrt{\frac{2k_B T}{m}}, \quad \langle v \rangle = \sqrt{\frac{8k_B T}{\pi m}}, \quad v_{\text{rms}} = \sqrt{\frac{3k_B T}{m}}. \quad (\text{C.22})$$

For a pair potential $u(r)$, the second virial coefficient is

$$B_2(T) = -\frac{1}{2} \int d^3r [e^{-\beta u(r)} - 1], \quad (\text{C.23})$$

and

$$\frac{P}{k_B T} = n + B_2(T)n^2 + \mathcal{O}(n^3). \quad (\text{C.24})$$

The truncation is a low-density expansion, not an equation of state near a critical point.

C.4 Information and nonequilibrium relations

For normalized classical distributions p and q ,

$$h(p) = -\sum_i p_i \ln p_i, \quad (\text{C.25})$$

$$D(p\|q) = \sum_i p_i \ln \frac{p_i}{q_i} \geq 0, \quad (\text{C.26})$$

$$I(X : Y) = D(p_{XY} \| p_X p_Y). \quad (\text{C.27})$$

Terms with $p_i = 0$ are defined by continuity. If $p_i > 0$ for some state with $q_i = 0$, then $D(p\|q) = +\infty$. For every stochastic map $\mathcal{T}$,

$$D(\mathcal{T}p\|\mathcal{T}q) \leq D(p\|q). \quad (\text{C.28})$$

This is a statement about distinguishability of two preparations. It does not say that $h(\mathcal{T}p)$ must be larger than $h(p)$.

For a Markov jump process with current

$$J_{ij} = W_{ij}p_j - W_{ji}p_i, \quad (\text{C.29})$$

the dimensionless total entropy-production rate, in units of k_B , is

$$\dot{\Sigma} = \frac{1}{2} \sum_{ij} J_{ij} \ln \frac{W_{ij}p_j}{W_{ji}p_i} \geq 0. \quad (\text{C.30})$$

Terms with both directed fluxes zero are interpreted by continuity. A strictly positive flux in one direction and zero reverse flux gives an infinite contribution. Here $\dot{\Sigma}$ is a rate; $\Sigma[\gamma]$ in Chapter 5 is the entropy production of one trajectory. The rates, mesostates, and reservoirs are part of the physical model.

For work performed on a system initially in canonical equilibrium,

$$\langle e^{-\beta W} \rangle = e^{-\beta \Delta F}, \quad \langle W \rangle \geq \Delta F. \quad (\text{C.31})$$

The Jarzynski equality needs canonical initialization for the forward protocol and microscopic reversibility, or the corresponding local-detailed-balance condition. No reverse experiment has to be sampled. Crooks' stronger forward–reverse relation additionally defines a reverse protocol initialized in equilibrium at the final control value. The displayed mean-work inequality follows from Jensen's inequality for $\beta > 0$; at $\beta < 0$ its direction reverses.

For a system that starts in an arbitrary distribution, coupled to a bath at $T > 0$ that starts in canonical equilibrium and uncorrelated with it, with negligible interaction energy at the initial and final times, the work done on the system obeys

$$W - \Delta F = k_B T [I(S : B) + D(q_B \| \pi_B)] \geq 0, \quad \mathcal{F} = \langle H \rangle - k_B T h, \quad (\text{C.32})$$

where $I(S : B)$ is the final system–bath mutual information, q_B is the final distribution of the bath, and $D(q_B \| \pi_B)$ measures its departure from the initial canonical distribution π_B . Ideal erasure of a symmetric, degenerate, equiprobable bit into a bath at $T > 0$, with the memory Hamiltonian and controls restored, obeys

$$W_{\text{erase}} \geq k_B T \ln 2. \quad (\text{C.33})$$

For biased or energetically asymmetric memories, use the full nonequilibrium free-energy change instead. Feedback after a measurement with record M , on a system initially in equilibrium, and the combined cost of producing and erasing the record obey

$$\langle W \rangle \geq \Delta F - k_B T I(X : M), \quad W_{\text{meas}} + W_{\text{eras}} \geq k_B T I(X : M). \quad (\text{C.34})$$

When a system and memory are correlated, their joint entropy is

$$S_{XM} = S_X + S_M - k_B I(X : M). \quad (\text{C.35})$$

An entropy ledger may add marginal entropies only when the correlation terms vanish at the accounting endpoints.

C.5 Quantum entropy and correlations

For a density operator ρ ,

$$s(\rho) = -\text{Tr}(\rho \ln \rho), \quad (\text{C.36})$$

$$D(\rho \| \sigma) = \text{Tr}[\rho(\ln \rho - \ln \sigma)], \quad (\text{C.37})$$

$$I(A : B) = s(A) + s(B) - s(AB) \quad (\text{C.38})$$

$$= D(\rho_{AB} \| \rho_A \otimes \rho_B). \quad (\text{C.39})$$

Quantum relative entropy is finite only when $\text{supp } \rho \subseteq \text{supp } \sigma$. For a channel $\mathcal{N}$,

$$D(\mathcal{N}\rho \| \mathcal{N}\sigma) \leq D(\rho \| \sigma). \quad (\text{C.40})$$

Strong subadditivity is

$$I(A : C|B) = s(AB) + s(BC) - s(B) - s(ABC) \geq 0. \quad (\text{C.41})$$

It is a theorem, not an assumption about the state. A pure bipartite state has the Schmidt decomposition

$$|\psi\rangle = \sum_i \sqrt{\lambda_i} |i_A\rangle |i_B\rangle, \quad \lambda_i \geq 0, \quad \sum_i \lambda_i = 1. \quad (\text{C.42})$$

The λ_i are Schmidt probabilities, equivalently the nonzero eigenvalues of either reduced state, and

$$s_{\text{ent}}(A) = s(\rho_A) = -\sum_i \lambda_i \ln \lambda_i. \quad (\text{C.43})$$

For a mixed global state, $s(\rho_A)$ alone is not an entanglement measure.

Define $F_T(\rho) = \text{Tr}(\rho H) - k_B T s(\rho)$. The Gibbs state and quantum free-energy identity are

$$\rho_\beta = \frac{e^{-\beta H}}{Z}, \quad F_T(\rho) - F_T(\rho_\beta) = k_B T D(\rho \| \rho_\beta). \quad (\text{C.44})$$

For $T > 0$, the Gibbs state minimizes F_T ; for a normalizable Gibbs state at $T < 0$, it maximizes F_T .

For an initially factorized state $\rho_S \otimes \rho_B^\beta$, a Gibbs bath $\rho_B^\beta = e^{-\beta H_B}/Z_B$, and joint unitary evolution to ρ'_{SB} , define heat absorbed by the system as the bath energy loss, $Q = \text{Tr}[H_B(\rho_B^\beta - \rho'_B)]$. Then

$$D(\rho'_{SB} \| \rho'_S \otimes \rho_B^\beta) = \Delta s_S - \beta Q = I(S : B)_{\rho'} + D(\rho'_B \| \rho_B^\beta) \geq 0. \quad (\text{C.45})$$

This exact identity assumes the stated initial product, Gibbs reference bath, joint unitary, and heat convention. Initial correlations require a modified balance. Closed unitary evolution preserves $s(\rho)$; a subsystem entropy can change because entanglement with its complement changes.

C.6 Ideal quantum gases

For one-particle levels ϵ_r ,

$$\langle n_r \rangle = \frac{1}{e^{\beta(\epsilon_r - \mu)} \pm 1}, \quad (\text{C.46})$$

with plus for fermions and minus for bosons. For bosons, $\mu \leq \epsilon_0$, where ϵ_0 is the lowest one-particle energy. The one-mode fluctuation is

$$\text{Var}(n_r) = \begin{cases} \langle n_r \rangle [1 - \langle n_r \rangle], & \text{fermions,} \\ \langle n_r \rangle [1 + \langle n_r \rangle], & \text{bosons.} \end{cases} \quad (\text{C.47})$$

For a three-dimensional quadratic dispersion,

$$D(\epsilon) = \frac{gV}{4\pi^2} \left(\frac{2m}{\hbar^2} \right)^{3/2} \epsilon^{1/2}. \quad (\text{C.48})$$

Here g is the internal degeneracy, such as the number of spin states.

For a spin-1/2 Fermi gas at $T = 0$,

$$k_F = (3\pi^2 n)^{1/3}, \quad \epsilon_F = \frac{\hbar^2 k_F^2}{2m}, \quad (\text{C.49})$$

$$U_0 = \frac{3}{5} N \epsilon_F, \quad P_0 = \frac{2}{5} n \epsilon_F. \quad (\text{C.50})$$

At $k_B T \ll \epsilon_F$,

$$C_V = \frac{\pi^2}{2} N k_B \frac{k_B T}{\epsilon_F} + \mathcal{O} \left[N k_B \left(\frac{k_B T}{\epsilon_F} \right)^3 \right]. \quad (\text{C.51})$$

For the homogeneous three-dimensional ideal Bose gas,

$$T_c = \frac{2\pi\hbar^2}{mk_B} \left[\frac{n}{g\zeta(3/2)} \right]^{2/3}, \quad (\text{C.52})$$

$$\frac{N_0}{N} = 1 - \left(\frac{T}{T_c} \right)^{3/2} \quad (T < T_c). \quad (\text{C.53})$$

For dispersion $\epsilon \propto p^s$, finite-temperature condensation in a homogeneous ideal gas requires $d > s$.

For photons,

$$u(\omega) = \frac{\hbar\omega^3}{\pi^2 c^3} \frac{1}{e^{\beta\hbar\omega} - 1}, \quad (\text{C.54})$$

where $u(\omega)$ is the energy density per angular-frequency interval. With $a = \pi^2 k_B^4 / (15\hbar^3 c^3)$,

$$U/V = aT^4, \quad P = U/(3V), \quad S = \frac{4}{3} aVT^3. \quad (\text{C.55})$$

C.7 Critical phenomena and renormalization

In this section, $t = (T - T_c)/T_c$ is the reduced temperature, h is the field conjugate to the order parameter m , and $G(r)$ is the connected two-point correlation function. The symbol β in $m \sim (-t)^\beta$ is a critical exponent. It is unrelated to the inverse temperature $1/(k_B T)$ used in the ensemble formulas; the two meanings never occur in the same equation below.

The standard critical limits are

$$m(t, 0) \sim (-t)^\beta \quad (t < 0), \quad C_h(t, 0) \sim |t|^{-\alpha}, \quad \chi \sim |t|^{-\gamma}, \quad \xi \sim |t|^{-\nu}, \quad (\text{C.56})$$

$$m(0, h) \sim |h|^{1/\delta} \operatorname{sgn} h, \quad G(r)|_{t=h=0} \sim r^{-(d-2+\eta)}. \quad (\text{C.57})$$

The scaling hypothesis, differentiated explicitly for the magnetization in [section 9.6](#),

$$f_s(t, h) = b^{-d} f_s(b^{y_t} t, b^{y_h} h) \quad (\text{C.58})$$

uses a coarse-graining factor $b > 1$, unrelated to the excluded-volume parameter in the van der Waals equation. It gives, when hyperscaling applies,

$$\nu = \frac{1}{y_t}, \quad \beta = \frac{d - y_h}{y_t}, \quad \gamma = \frac{2y_h - d}{y_t}, \quad \alpha = 2 - \frac{d}{y_t}, \quad \delta = \frac{y_h}{d - y_h}, \quad \gamma = \nu(2 - \eta). \quad (\text{C.59})$$

For the one-dimensional Ising chain, let $K = J/(k_B T)$. Its exact zero-field decimation is

$$K' = \frac{1}{2} \ln \cosh(2K), \quad A = 2\sqrt{\cosh(2K)}. \quad (\text{C.60})$$

Here A is the multiplicative normalization contributed by each retained block. All finite K flow toward the high-temperature fixed point. In terms of the wall fugacity $y = e^{-2K}$,

$$y' = \frac{2y}{1 + y^2} = 2y - 2y^3 + \mathcal{O}(y^5), \quad (\text{C.61})$$

showing that nonzero temperature is relevant near $K = \infty$.

At the Gaussian fixed point of scalar ϕ^4 theory,

$$r' = b^2 r, \quad u' = b^{4-d} u. \quad (\text{C.62})$$

The quartic interaction is relevant below four dimensions, and mean-field exponents cease to be exact. Computing the interacting fixed point requires an RG expansion; power counting alone gives no corrected exponent.

References

- [1] C. H. Bennett. “Logical Reversibility of Computation”. In: *IBM Journal of Research and Development* 17 (1973), pp. 525–532. DOI: [10.1147/rd.176.0525](#).
- [2] Charles H. Bennett. “The Thermodynamics of Computation: A Review”. In: *International Journal of Theoretical Physics* 21 (1982), pp. 905–940. DOI: [10.1007/BF02084158](#).
- [3] Charles H. Bennett. “Notes on Landauer’s principle, reversible computation, and Maxwell’s Demon”. In: *Studies in History and Philosophy of Modern Physics* 34 (2003), pp. 501–510. DOI: [10.1016/S1355-2198\(03\)00039-X](#). arXiv: [physics/0210005](#).
- [4] Antoine Bérut, Artak Arakelyan, Artyom Petrosyan, Sergio Ciliberto, Raoul Dillenschneider, and Eric Lutz. “Experimental verification of Landauer’s principle linking information and thermodynamics”. In: *Nature* 483 (2012), pp. 187–189. DOI: [10.1038/nature10872](#).
- [5] Ludwig Boltzmann. “Weitere Studien über das Wärmegleichgewicht unter Gasmolekülen”. In: *Sitzungsberichte der Akademie der Wissenschaften in Wien* 66 (1872), pp. 275–370.
- [6] L. Brillouin. “Maxwell’s Demon Cannot Operate: Information and Entropy. I”. In: *Journal of Applied Physics* 22 (1951), pp. 334–337. DOI: [10.1063/1.1699951](#).
- [7] Pasquale Calabrese and John Cardy. “Entanglement Entropy and Quantum Field Theory”. In: *Journal of Statistical Mechanics: Theory and Experiment* 2004 (2004), P06002. DOI: [10.1088/1742-5468/2004/06/P06002](#). arXiv: [hep-th/0405152](#).
- [8] Herbert B. Callen. *Thermodynamics and an Introduction to Thermostatistics*. 2nd ed. Wiley, 1985.
- [9] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. 2nd ed. Wiley, 2006. DOI: [10.1002/047174882X](#).
- [10] Gavin E. Crooks. “Entropy Production Fluctuation Theorem and the Nonequilibrium Work Relation for Free Energy Differences”. In: *Physical Review E* 60 (1999), pp. 2721–2726. DOI: [10.1103/PhysRevE.60.2721](#).
- [11] Luca D’Alessio, Yariv Kafri, Anatoli Polkovnikov, and Marcos Rigol. “From Quantum Chaos and Eigenstate Thermalization to Statistical Mechanics and Thermodynamics”. In: *Advances in Physics* 65 (2016), pp. 239–362. DOI: [10.1080/00018732.2016.1198134](#). arXiv: [1509.06411](#).
- [12] Yu Deng, Zaher Hani, and Xiao Ma. “Long Time Derivation of the Boltzmann Equation from Hard Sphere Dynamics”. In: *Annals of Mathematics* (2026). arXiv: [2408.07818](#). Forthcoming.
- [13] Paul Ehrenfest and Tatiana Ehrenfest. “Über zwei bekannte Einwände gegen das Boltzmannsche H-Theorem”. In: *Physikalische Zeitschrift* 8 (1907), pp. 311–314.
- [14] M. Esposito and C. Van den Broeck. “Second law and Landauer principle far from equilibrium”. In: *EPL (Europhysics Letters)* 95 (2011), p. 40004. DOI: [10.1209/0295-5075/95/40004](#). arXiv: [1104.5165](#).

-
- [15] Massimiliano Esposito, Katja Lindenberg, and Christian Van den Broeck. “Entropy production as correlation between system and reservoir”. In: *New Journal of Physics* 12 (2010), p. 013013. DOI: [10.1088/1367-2630/12/1/013013](https://doi.org/10.1088/1367-2630/12/1/013013). arXiv: [0908.1125](https://arxiv.org/abs/0908.1125).
 - [16] J. Willard Gibbs. *Elementary Principles in Statistical Mechanics*. New York: Charles Scribner’s Sons, 1902.
 - [17] Jordan M. Horowitz and Juan M. R. Parrondo. “Thermodynamic reversibility in feedback processes”. In: *EPL (Europhysics Letters)* 95 (2011), p. 10005. DOI: [10.1209/0295-5075/95/10005](https://doi.org/10.1209/0295-5075/95/10005). arXiv: [1104.0332](https://arxiv.org/abs/1104.0332).
 - [18] C. Jarzynski. “Nonequilibrium Equality for Free Energy Differences”. In: *Physical Review Letters* 78 (1997), pp. 2690–2693. DOI: [10.1103/PhysRevLett.78.2690](https://doi.org/10.1103/PhysRevLett.78.2690). arXiv: [cond-mat/9610209](https://arxiv.org/abs/cond-mat/9610209).
 - [19] E. T. Jaynes. “Information Theory and Statistical Mechanics”. In: *Physical Review* 106 (1957), pp. 620–630. DOI: [10.1103/PhysRev.106.620](https://doi.org/10.1103/PhysRev.106.620).
 - [20] Yonggun Jun, Momčilo Gavrilov, and John Bechhoefer. “High-Precision Test of Landauer’s Principle in a Feedback Trap”. In: *Physical Review Letters* 113 (2014), p. 190601. DOI: [10.1103/PhysRevLett.113.190601](https://doi.org/10.1103/PhysRevLett.113.190601). arXiv: [1408.5089](https://arxiv.org/abs/1408.5089).
 - [21] Mark Kac. “Random Walk and the Theory of Brownian Motion”. In: *American Mathematical Monthly* 54 (1947), pp. 369–391.
 - [22] Mehran Kardar. *Statistical Physics of Fields*. Cambridge University Press, 2007. DOI: [10.1017/CB09780511815881](https://doi.org/10.1017/CB09780511815881).
 - [23] Mehran Kardar. *Statistical Physics of Particles*. Cambridge University Press, 2007. DOI: [10.1017/CB09780511815898](https://doi.org/10.1017/CB09780511815898).
 - [24] R. Kawai, J. M. R. Parrondo, and C. Van den Broeck. “Dissipation: The Phase-Space Perspective”. In: *Physical Review Letters* 98 (2007), p. 080602. DOI: [10.1103/PhysRevLett.98.080602](https://doi.org/10.1103/PhysRevLett.98.080602). arXiv: [cond-mat/0701397](https://arxiv.org/abs/cond-mat/0701397).
 - [25] Cargill Gilston Knott. *Life and Scientific Work of Peter Guthrie Tait: Supplementing the Two Volumes of Scientific Papers Published in 1898 and 1900*. Cambridge: Cambridge University Press, 1911.
 - [26] Jonne V. Koski, Ville F. Maisi, Jukka P. Pekola, and Dmitri V. Averin. “Experimental realization of a Szilard engine with a single electron”. In: *Proceedings of the National Academy of Sciences of the United States of America* 111 (2014), pp. 13786–13789. DOI: [10.1073/pnas.1406966111](https://doi.org/10.1073/pnas.1406966111). arXiv: [1402.5907](https://arxiv.org/abs/1402.5907).
 - [27] H. A. Kramers and G. H. Wannier. “Statistics of the Two-Dimensional Ferromagnet. Part I”. In: *Physical Review* 60 (1941), pp. 252–262. DOI: [10.1103/PhysRev.60.252](https://doi.org/10.1103/PhysRev.60.252).
 - [28] Rolf Landauer. “Irreversibility and Heat Generation in the Computing Process”. In: *IBM Journal of Research and Development* 5.3 (1961), pp. 183–191. DOI: [10.1147/rd.53.0183](https://doi.org/10.1147/rd.53.0183).
 - [29] Oscar E. Lanford. “Time Evolution of Large Classical Systems”. In: *Dynamical Systems, Theory and Applications*. Ed. by J. Moser. Vol. 38. Lecture Notes in Physics. Berlin: Springer, 1975, pp. 1–111.
 - [30] Joel L. Lebowitz. “Boltzmann’s Entropy and Time’s Arrow”. In: *Physics Today* 46.9 (1993), pp. 32–38.
 - [31] Harvey S. Leff and Andrew F. Rex, eds. *Maxwell’s Demon 2: Entropy, Classical and Quantum Information, Computing*. Bristol and Philadelphia: Institute of Physics Publishing, 2003.

- [32] Elliott H. Lieb and Mary Beth Ruskai. “Proof of the Strong Subadditivity of Quantum-Mechanical Entropy”. In: *Journal of Mathematical Physics* 14 (1973), pp. 1938–1941. DOI: [10.1063/1.1666274](https://doi.org/10.1063/1.1666274).
- [33] Jan Liphardt, Sophie Dumont, Steven B. Smith, Ignacio Tinoco, and Carlos Bustamante. “Equilibrium Information from Nonequilibrium Measurements in an Experimental Test of Jarzynski’s Equality”. In: *Science* 296 (2002), pp. 1832–1835. DOI: [10.1126/science.1071152](https://doi.org/10.1126/science.1071152).
- [34] Josef Loschmidt. “Über den Zustand des Wärmegleichgewichtes eines Systems von Körpern mit Rücksicht auf die Schwerkraft”. In: *Sitzungsberichte der Akademie der Wissenschaften in Wien* 73 (1876), pp. 128–142.
- [35] Koji Maruyama, Franco Nori, and Vlatko Vedral. “Colloquium: The physics of Maxwell’s demon and information”. In: *Reviews of Modern Physics* 81 (2009), pp. 1–23. DOI: [10.1103/RevModPhys.81.1](https://doi.org/10.1103/RevModPhys.81.1). arXiv: [0707.3400](https://arxiv.org/abs/0707.3400).
- [36] J. Clerk Maxwell. *Theory of Heat*. London: Longmans, Green, and Co., 1871.
- [37] James Clerk Maxwell. “On the Dynamical Theory of Gases”. In: *Philosophical Transactions of the Royal Society of London* 157 (1867), pp. 49–88.
- [38] National Institute of Standards and Technology. *Carbon Monoxide: Constants of Diatomic Molecules*. NIST Chemistry WebBook, SRD 69. URL: <https://webbook.nist.gov/cgi/cbook.cgi?ID=C630080&Mask=1000> (visited on 07/14/2026).
- [39] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- [40] Lars Onsager. “Crystal Statistics. I. A Two-Dimensional Model with an Order-Disorder Transition”. In: *Physical Review* 65 (1944), pp. 117–149. DOI: [10.1103/PhysRev.65.117](https://doi.org/10.1103/PhysRev.65.117).
- [41] Juan M. R. Parrondo, Jordan M. Horowitz, and Takahiro Sagawa. “Thermodynamics of information”. In: *Nature Physics* 11 (2015), pp. 131–139. DOI: [10.1038/nphys3230](https://doi.org/10.1038/nphys3230).
- [42] R. K. Pathria and Paul D. Beale. *Statistical Mechanics*. 3rd ed. Elsevier, 2011.
- [43] Wolfgang Pauli. “Über das H-Theorem vom Anwachsen der Entropie vom Standpunkt der neuen Quantenmechanik”. In: *Probleme der modernen Physik: Arnold Sommerfeld zum 60. Geburtstag gewidmet von seinen Schülern*. Leipzig: S. Hirzel, 1928, pp. 30–45.
- [44] Henri Poincaré. “Sur le problème des trois corps et les équations de la dynamique”. In: *Acta Mathematica* 13 (1890), pp. 1–270.
- [45] Sandu Popescu, Anthony J. Short, and Andreas Winter. “Entanglement and the Foundations of Statistical Mechanics”. In: *Nature Physics* 2 (2006), pp. 754–758. DOI: [10.1038/nphys444](https://doi.org/10.1038/nphys444). arXiv: [quant-ph/0511225](https://arxiv.org/abs/quant-ph/0511225).
- [46] David Reeb and Michael M. Wolf. “An improved Landauer principle with finite-size corrections”. In: *New Journal of Physics* 16 (2014), p. 103011. DOI: [10.1088/1367-2630/16/10/103011](https://doi.org/10.1088/1367-2630/16/10/103011). arXiv: [1306.4352](https://arxiv.org/abs/1306.4352).
- [47] Linda E. Reichl. *A Modern Course in Statistical Physics*. 4th ed. Wiley-VCH, 2016.
- [48] Takahiro Sagawa and Masahito Ueda. “Second Law of Thermodynamics with Discrete Quantum Feedback Control”. In: *Physical Review Letters* 100 (2008), p. 080403. DOI: [10.1103/PhysRevLett.100.080403](https://doi.org/10.1103/PhysRevLett.100.080403). arXiv: [0710.0956](https://arxiv.org/abs/0710.0956).
- [49] Takahiro Sagawa and Masahito Ueda. “Minimal Energy Cost for Thermodynamic Information Processing: Measurement and Information Erasure”. In: *Physical Review Letters* 102 (2009), p. 250602. DOI: [10.1103/PhysRevLett.102.250602](https://doi.org/10.1103/PhysRevLett.102.250602). arXiv: [0809.4098](https://arxiv.org/abs/0809.4098).

-
- [50] Takahiro Sagawa and Masahito Ueda. “Generalized Jarzynski Equality under Nonequilibrium Feedback Control”. In: *Physical Review Letters* 104 (2010), p. 090602. DOI: [10.1103/PhysRevLett.104.090602](https://doi.org/10.1103/PhysRevLett.104.090602). arXiv: [0907.4914](https://arxiv.org/abs/0907.4914).
 - [51] Takahiro Sagawa and Masahito Ueda. “Nonequilibrium thermodynamics of feedback control”. In: *Physical Review E* 85 (2012), p. 021104. DOI: [10.1103/PhysRevE.85.021104](https://doi.org/10.1103/PhysRevE.85.021104). arXiv: [1105.3262](https://arxiv.org/abs/1105.3262).
 - [52] Udo Seifert. “Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines”. In: *Reports on Progress in Physics* 75 (2012), p. 126001. DOI: [10.1088/0034-4885/75/12/126001](https://doi.org/10.1088/0034-4885/75/12/126001). arXiv: [1205.4176](https://arxiv.org/abs/1205.4176).
 - [53] James P. Sethna. *Statistical Mechanics: Entropy, Order Parameters, and Complexity*. 2nd ed. Oxford University Press, 2021.
 - [54] Claude E. Shannon. “A Mathematical Theory of Communication”. In: *Bell System Technical Journal* 27 (1948), pp. 379–423, 623–656. DOI: [10.1002/j.1538-7305.1948.tb01338.x](https://doi.org/10.1002/j.1538-7305.1948.tb01338.x).
 - [55] M. v. Smoluchowski. “Experimentell nachweisbare, der üblichen Thermodynamik widersprechende Molekularphänomene”. In: *Physikalische Zeitschrift* 13 (1912), pp. 1069–1080.
 - [56] L. Szilard. “Über die Entropieverminderung in einem thermodynamischen System bei Eingriffen intelligenter Wesen”. In: *Zeitschrift für Physik* 53 (1929), pp. 840–856. DOI: [10.1007/BF01341281](https://doi.org/10.1007/BF01341281).
 - [57] William Thomson. “Kinetic Theory of the Dissipation of Energy”. In: *Nature* 9 (1874), pp. 441–444. DOI: [10.1038/009441c0](https://doi.org/10.1038/009441c0).
 - [58] Hugo Touchette. “The Large Deviation Approach to Statistical Mechanics”. In: *Physics Reports* 478 (2009), pp. 1–69. DOI: [10.1016/j.physrep.2009.05.002](https://doi.org/10.1016/j.physrep.2009.05.002). arXiv: [0804.0327](https://arxiv.org/abs/0804.0327).
 - [59] Shoichi Toyabe, Takahiro Sagawa, Masahito Ueda, Eiro Muneyuki, and Masaki Sano. “Experimental demonstration of information-to-energy conversion and validation of the generalized Jarzynski equality”. In: *Nature Physics* 6 (2010), pp. 988–992. DOI: [10.1038/nphys1821](https://doi.org/10.1038/nphys1821). arXiv: [1009.5287](https://arxiv.org/abs/1009.5287).
 - [60] Mark M. Wilde. *Quantum Information Theory*. 2nd ed. Cambridge University Press, 2017. DOI: [10.1017/9781316809976](https://doi.org/10.1017/9781316809976).
 - [61] Kenneth G. Wilson and John Kogut. “The Renormalization Group and the ϵ Expansion”. In: *Physics Reports* 12 (1974), pp. 75–199. DOI: [10.1016/0370-1573\(74\)90023-4](https://doi.org/10.1016/0370-1573(74)90023-4).
 - [62] Edward Witten. “A Mini-Introduction to Information Theory”. In: *Rivista del Nuovo Cimento* 43 (2020), pp. 187–227. DOI: [10.1007/s40766-020-00004-5](https://doi.org/10.1007/s40766-020-00004-5). arXiv: [1805.11965](https://arxiv.org/abs/1805.11965).
 - [63] C. N. Yang. “The Spontaneous Magnetization of a Two-Dimensional Ising Model”. In: *Physical Review* 85 (1952), pp. 808–816. DOI: [10.1103/PhysRev.85.808](https://doi.org/10.1103/PhysRev.85.808).
 - [64] Ernst Zermelo. “Ueber einen Satz der Dynamik und die mechanische Wärmetheorie”. In: *Annalen der Physik* 57 (1896), pp. 485–494.