Lectures on Stochastic Thermodynamics

Chem 220B @ UC Berkeley
Spring 2023

Dr. Gavin E. Crooks

https://github.com/gecrooks/LectureNotes220b

Contents

1 Lectures on Stochastic Thermodynamics

Instructor: Dr. Gavin E. Crooks
Email: gavincrooks@gmail.com
Office: 319A Gilman
Office Hours: 11-12 Tu, 3:30-4:30 Thr (Or by appointment)

GSI: Aditya Singh
Email: ansingh@berkeley.edu

Synopsis: In this advanced graduate-level course we will discuss contemporary thermodynamics. In contrast to traditional equilibrium thermodynamics and statical mechanics, we are increasingly interested in the dynamical properties of microscopic systems away from thermodynamic equilibrium. We will review driven non-equilibrium thermodynamics, both in linear response and far-from equilibrium and the various dynamics that are used to model reality, and physics of information.

Prerequisites: A solid understanding of equilibrium thermodynamics and statistical dynamics (i.e. Chem 220a). Basic quantum mechanics, linear algebra, and programming.

Grading: Grades will be based on 7-8 problem sets (65%) and a final report (35%), due during reading week. We will discuss the details of this report as the semester progresses.

Discussion Sections: In addition to the regularly schedule lectures, discussion sections will generally be held on a biweekly basis. These sections will be taught by a GSI and will serve as a complement to the lecture with a special emphasis on topics relevant to the problem sets.

Textbook: We will not follow a specific textbook. Lecture notes and synopses will be posted through the semester. However, the following books may be useful:

- Luca Peliti & Simone Pigolotti, Stochastic Thermodynamics: An Introduction [1]
- Robert Zwanzig, Nonequilibrium Statistical Mechanics [2]

Syllabus Outline (tentative):

1. Information Theory (Including Maxwell's demon and Landauer's principle)
2. Thermodynamics (We will review the fundamentals and those aspects that need reconsideration in light of recent developments: work and heat, entropy and free energy, and the nature of thermodynamic equilibrium)
3. Fluctuation Theorems (microscopic reversibility and detailed balance; microscopic, detailed, and integrated fluctuation theorems; various consequences thereof; and experimental realizations)

4. Linear response (response to perturbations in the near-equilibrium regime, fluctuation-dissipation theorems, thermodynamic geometry)

5. Dynamics (Interspersed through the semester, we will discuss various models of reality, including DTMC, CTMC, classical dynamics, Langevin dynamics, molecular dynamics, and quantum processes)

6. Advanced Topics (TBD, depending on time)

7. Guest Lectures (Contemporary thermodynamics as practiced by those at the frontiers of research)

2 Lecture Summaries

2.1 Probability and ensembles

- Probabilities are contextual (e.g. probability of a fair coin flip)
- Probabilities depend on compute power (e.g. probability that a number is prime)
- Human's are often bad at probabilities (Monte Hall problem, Three card problem)
- Probability notation
- Ensembles
- random variables
- joint, marginal, and conditional probabilities
- Bayes rule

2.2 Bits and Bytes

- Entropy, $S(A) = -\sum_a P_A(a) \log P_A(a)$
- review history (Clausis, Boltzmann, Gibbs, Shannon, Jaynes, Bennet and Landauer)
- review how logs work
- units (bits, nats, ban, deciban)
- units, kT
- orders of magnitude

Further Reading

Review Chapter 2 of Cover and Thomas (2006) [3].
Optional: Shannon's original paper is well worth reading, Shannon (1948) [4].

2.3 Entropy and Information

- joint, marginal, conditional entropies
- mutual information
- nature of information. Coin example, correlations
- information diagrams,
- 3 variable mutual information
- information diagram, 3 variable
- relative entropy
  - naming
  - notation

  - intuition
  - not a distance
  - related to mutual information
  - cross entropy
- Jensens inequality
  - Statement
  - Convex and concave functions
  - $< \exp x >\geqslant exp(< x >)$
  - Relative entropy is non-negative
  - Mutual information is non-negative
  - Max $S(A) = \ln N$
- Relative entropy and free energy
- differential entropy
  - normal
  - change of variable problems
  - problems with classical entropy of continuous system
  - differential relative entropy invariant

Further Reading

Chapter 2 of Cover and Thomas (2006) [3]. See also Crooks (2015-2021) [5], which summaries the information measures we covered, as well as many other less common measures of information.

2.4 Maxwell's demon and Szilard's Engine

2.5 Thermodynamics and Stat. Mech.

2.6 Free Energy, work and heat

2.7 Discrete Time Markov Chains (DTMC)

Further Reading

The book by Norris provides a good introduction to discrete and continuous time Markov Chains Norris (1997) [6].

2.8 Jarzynski Identity

2.9 Fluctuation Theorems

2.10 Continuous Time Markov Chains (CTMC)

Further Reading

The book by Norris provides a good introduction to discrete and continuous time Markov Chains Norris (1997) [6].

3 Probability and ensembles

3.1 The Coin Toss

Suppose I present to you a coin, one side of which we will call heads, and one tails. I propose to perform a fair coin toss. Let's assume that it's not a trick coin, and I'm not a magician. What are the chances that the coin will come down heads up? Under these circumstances, with this information, the only reasonable answer is that the coin with come up heads or tails equally often. Of course, the coin may be biased, and come down on one side or the other more often. But we can't know which side is more likely unless we preform experiments on this particular coin. So I have to assign equal chances to both possibilities, since the labels are essentially arbitrary and changing the labels shouldn't change anything.

Now suppose I toss the coin, snatch it out of the air, and slap the coin onto the back of my left hand, concealed the top face with my right hand, as tradition dictates. Neither you nor I have fast enough perception to see which way the coin fell. So what now are the chances that the coin is heads? The only reasonable answer is that the chances have not changed.

I now peak at the coin under my hand, but continue to conceal the coin from yourself. What are the chances that the coin is heads up? It depends. For myself, the coin is now either definitely heads up, or definitely tails up. But for yourself, the chances remain even.

We represent probabilities by numerical measures between 0 and 1 (with 0 representing certainly false, and 1 certainly true). These probabilities are not properties of the system itself (here the system of interest is the coin). Rather probabilities *contextual*[1]; they depend on what we know about the system, and since you and I can have different knowledge our probabilities for the same event can be different.

3.2 Prime Numbers

Suppose I present to you a random 30 digit number:

$$192339819110572236368487297967$$

What are the chances that this number is prime? In a Platonic sense this number is either prime, or not prime, and has always been prime (or not prime) since the beginning of time. Asking if this number is prime is purely a question about our mathematical ignorance. But if we know a bit of mathematics, then we'd know the *prime number theorem*, which states that the probability of a random number N being prime is about

$$P(\mathrm{N\ is\ prime}) \approx \frac{1}{\ln \mathrm{N}} \tag{1}$$

which is about 1 in 70 for a 30 digit number.

Proving that a large number is prime is hard. However, number theorists have developed fast prime number tests that can tell you with near certainty if a number is prime (or not), provided we are able and willing to expend a modest amount of computational power. Our number turns out to be almost certainly prime [2].

[1] Jaynes described probabilities as *subjective* rather than *contextual*. But the word *subjective* carries a lot of unwarranted connotations. We're not irrational, just ill-informed and short of time.

[2] Probably. We also have to factor in the odds that I was able to faithfully transcribed the original prime to the page or blackboard.

Probabilities depend not just on our knowledge of the system, but also on our understanding of reality (in this case, how much number theory we know), and on how much compute we throw at the problem.

3.3 Probability theory

Probability theory is a physical theory of plausible reasoning in a classical universe. Probability theory is often treated as branch of pure mathematics, but the fundamental assumptions include a physical model of reality corresponding to classical, pre-quantum physics. Probabilities are contextual, in that one's assignment of probabilities depends on what you know. But objective in that different rational observers with the same information, model, and computational resources should come to the same conclusions, and assign the same probabilities.

Propositions A *proposition* or *statement* is a logical assertion that may be true or false. Examples include "All men are mortal", "Socrates is a man", and "When I flip this coin, it will land with the tail side up". I'll tend to use lower case letters from the beginning of the alphabet for propositions: a, b, c. (Upper case letters are common in the literature, but we'll reserve those for ensembles, defined below.) A *conditional proposition* $a \mid h$ ("a given h") is an assertion 'a' premised on some other data or hypothesis 'h'. In principle all propositions are conditional, although conditions constant across an expression are typically not stated explicitly.

Boolean algebra The elementary Boolean operations are negation (not), conjugation (and) and disjunction (or). Note that logical 'or' means 'either or both', which differs from colloquial English usage. Below are the truth tables of the three basic logic elements, along with some of the different notations that you may encounter.

a	b	not a	a and b	a or b
		$\neg a$	$a \vee b$	$a \wedge b$
		$\sim a$	$a \,\&\, b$	$a \mid b$
false	false	true	false	false
false	true		false	true
true	false	false	false	true
true	true		true	true

Probability A *probability* is a numerical measure of the plausibility of a logical proposition a, given the hypothesis (data or evidence) h, written $P(a \mid h)$ and verbalized "the probability of a given h". Probability satisfies the following three rules.

(1) Convexity rule:

$$P(a \mid h) \text{ is a real number between zero and one.}$$
$$\text{with one representing certain truth.}$$

(2) Product rule:

$$P(a \text{ and } b \mid h) = P(b \mid a \text{ and } h)\, P(a \mid h)$$

(3) Sum rule:

$$P(a \text{ or } b \mid h) = P(a \mid h) + P(b \mid h) - P(a \text{ and } b \mid h)$$

One immediate and useful consequence is *Bayes' rule*

$$P(a \mid b) = \frac{P(b \mid a)P(a)}{P(b)} \tag{2}$$

A comma between propositions is taken to be equivalent to a conjunction of the propositions.

$$P(a, b \mid h) \equiv P(a \text{ and } b \mid h) \tag{3}$$

4 Entropy

The concept of entropy was first discovered by Rudolf Clausius in the 1860s [0, 0]. He realized that an equilibrium thermodynamic system was characterized both a conserved quantity, the energy of the system, but also by another quantity, which he named entropy, that is not conserved, but instead tends towards a maximum.

Clausius envisioned entropy as a physical substance, much like energy, and this perspective persisted in macroscopic thermodynamics for many years. But the early pioneers of probabilistic thinking in physics (most notable Maxwell, Boltzmann, and Gibbs), soon realized that entropy is essentially statistical in nature. We see the first formulations of entropy in terms of probabilities in Boltzmann's works [0, 0] (and the formula $S = \ln \Omega$ is inscribed on his tomb), followed by the general definition of thermodynamic entropy in Gibbs's magnum opus laying down the foundations of statistical mechanics [0].

As thermodynamics and statistical mechanics developed during the first half of the 20th century, physicists continued to think of entropy as primarily a property of physical systems at thermodynamic equilibrium. This began to change with Claude Shannon's development of information theory in the 1940s. Shannon was concerned with developing a mathematical theory of communication. He needed a quantitative measure of the information of a message to be transmitted, and in doing so he rediscovered the entropy formulas of Boltzmann and Gibbs, but shorn of all irrelevant physics [3]. Entropy is not a property of a system as such, but rather of the ensemble, the probabilities that the system of interest can be found in different states.

It took many years for Shannon's insights to be fully incorporated into thermodynamics. Jaynes[0, 0] was an early and vocal proponent of placing Shannon's conception of entropy at the heart of statistical mechanics. The equivalence of Shannon and thermodynamic entropies was most convincingly demonstrated by the work of Bennet and Landauer [0, 0]. The equivalence of Shannons information entropy and Clausius thermodynamic entropy is now known as Landauer's principle. They have to be equivalent or thermodynamics itself stops being consistence.

4.1 Bits and bytes

A *bit* is the fundamental irreducible unit of information storage. Essentially any physical system with two distinct, stable states can constitute a bit. We could label these states head and tails (if the bit is a coin), or up and down (for an electron spin), or on and off, or yes and no, or true and false. Most commonly we call the two states zero and one.

Suppose we want to record the state of some other physical system. How many bits of memory will we need? For instance, suppose we have a collection of N coins,

[3]

each of which is either heads up or tails up. Since these coins are themselves two-state systems it seems evident that we'll require N bits. But what if, instead of a coins, we have N 10-sided dice? The dice system has 10^N distinct states, so we'll need enough bits S so that the total number state of the bits 2^S is greater or equal to the states of the dice 10^N. Taking binary logarithms, the number of bits is at least

$$S = \log_2 10^N = N \log_2 10 \tag{4}$$

If S is fractional, we'll have to round up to the next integer. We need 3 bits for 1 die, 10 bits for 3 dice, 20 bits for 6 dice, as so on. Asymptotically, for large N we need $\log_2 10 \approx 3.32$ bits per digit. More generally, we say that the *entropy* of a system with N distinct, equiprobable states is

$$S = \log_b N \tag{5}$$

where b is the base of the logarithm. The units of entropy depend on which logarithmic base we choose. If we use base 10, then entropy is measured in digits (or bans). In computer science we most commonly use base 2, and measure entropy in bits (binary digits), whereas in physics it turns out to be most convenient to use natural logarithms (base e), and measure entropy in nats (natural units).

[properties of logarithms]

To convert between bits, nats, and bans, it is useful to remember the approximation that

$$2^{10} \approx e^7 \approx 10^3 \tag{6}$$

The exact numbers are $1024 \approx 1096.63\ldots \approx 1000$. This means that 10 bits is about 7 nats is about 3 digits. Also a kilobyte is about 2^{10} bytes, a megabyte is about 2^{20} bits, a gigabyte is about 2^{30} bytes, and so on. A byte is the normal unit of computer memory, which in modern usage is 8 bits. This size was chosen because $8 = 2^3$ is a nice round number, and $2^8 = 256$ is more than enough code points to encode all the digits, and upper and lower case Roman letters, plus a bunch of punctuation marks and control codes.

Note that the word 'bit' is overloaded. Bit refers both to the physical storage medium (i.e. a two state system) and the unit of entropy. It perfectly reasonable to have half a bucket of water, but half a bucket is nonsensical (Either it's not a bucket, if you cut vertically, or a smaller bucket, if you cut horizontally). Similarly, we can't have half a bit, but we can have half a bit of entropy[4].

[Physical units] [Kibibytes]

4.2 Logarithms

It's worth reviewing the properties of logarithms. The logarithm is the inverse function to exponentiation,

$$\log_b(b^x) = x, \qquad b^{\log_b x} = x \tag{7}$$

where $b > 1$ is the *base* of the logarithm[5]. We'll write base 2 "binary" logarithms as $\lg(x) \equiv \log_2(x)$, and base $e(\approx 2.71\ldots)$ "natural" logarithms as $\ln(x) \equiv \log_e(x)$.

[4] Some helpful pedantics have suggested that we replace *bit* with *shannon* as the unit of entropy. Thankfully, this proposal has not caught on.

[5] You could define the base to be any positive number, but we exclude fractional bases because we never need them, and some logarithmic properties become more complicated for sub-unity bases.

Table 1: Units of entropy

deciban	$\frac{1}{10}\log_2(10) \approx 0.33$ bits	tenth of a ban
bit (shannon)	1 bit	
nat (nit, nepit)	$\log_2(e) \approx 10/7$ bits	natural digit
trit	$\log_2(3) \approx 1.6$ bits	ternary digit
quad	2 bits	
ban (digit, hartly)	$\log_2(10) \approx 10/3$ bits	decimal digit
nibble (nybble)	4 bits	half a byte
byte	8 bits	
kilobyte	10^3 bytes	$\approx 2^{10}$ bytes
megabyte	10^6 bytes	$\approx 2^{20}$ bytes
gigabyte	10^9 bytes	$\approx 2^{30}$ bytes
terabyte	10^{12} bytes	$\approx 2^{40}$ bytes

When the base doesn't matter (except being consistent across an expression), we'll simple write $\log(x)$.

[tk figure]

Logarithms turn products into sums and quotients into differences,

$$\log_b(xy) = \log_b(x) + \log_b(y)$$
$$\log_b(\frac{x}{y}) = \log_b(x) - \log_b(y)$$

and exponents of the argument scale the logarithm.

$$\log_b(x^k) = k\log_b(x) \tag{8}$$

Conversely exponentiation turns addition into multiplication

$$b^{xy} = b^x\, b^y = (b^x)^y \tag{9}$$

Scaling a logarithm by a constant changes the base of the logarithm.

$$\log_b(x) = \log_b(c)\log_c(x) \tag{10}$$

To see why, exponentiate both sides to base b,

$$\begin{aligned} b^{\log_b(x)} &= b^{\log_b(c)\log_c(x)} \\ x &= (b^{\log_b(c)})^{\log_c(x)} \\ x &= c^{\log_c(x)} \\ x &= x \end{aligned}$$

The conversion factor between base 2 and base e is about $7/10$, and between base 2 and base 10 about $3/10$.

$$\begin{aligned} \ln(x) &= \ln(2)\lg(x) &&\approx 0.693\lg(x) \\ \log_{10}(x) &= \log_{10}(2)\lg(x) &&\approx 0.301\lg(x) \end{aligned}$$

The derivative of a logarithm is

$$\frac{d}{dx}\log_b(x) = \frac{1}{\ln(b)}\frac{1}{x}$$
$$= \frac{1}{x} \text{ for natural logarithms}$$

and the integral

$$\frac{d}{dx}\log_b(x) = \frac{1}{\ln(b)}\frac{1}{x}$$
$$= \frac{1}{x} \text{ for natural logarithms}$$

4.3 Entropy

[transition. Averages]

The *entropy* is a measure of the inherent information content of an ensemble. It is the average number of bits (or other unit of information) that is needed to record the state of the system.

$$S(A) = -\sum_a P(A_a)\log_b P(A_a) \tag{11}$$

In information theory the entropy is typically denoted by the symbol H, a notation that dates back to Boltzmann and his H-theorem [7], and adopted by Shannon [4]. The notation S is due to Clausius and the original discovery of entropy in thermodynamics [8][6], and adopted by Gibbs [9] for use in statistical mechanics. I tend to use S since I care about the physics of information, and we often need the symbol H to denote the Hamiltonian.

Entropies of discrete distributions are non-negative and bounded. The minimum occurs when one outcome is certain (and all other states have zero probability. The maximum arrises when all states of the ensemble have the same probability [7].

$$0 \leqslant S(A) \leqslant \ln|\Omega_A|$$

Note that the entropy of a continuous distribution can be negative, as we'll discuss in sec??.

Given a joint probability distribution $P_{AB}(x,y)$ then the *joint entropy* is

$$S(A,B) = -\sum_{x,y} P_{AB}(x,y)\ln P_{AB}(x,y) \tag{12}$$

This joint entropy can be readily generalized to any number of variables.

$$S(A_1, A_2, \ldots, A_n)$$
$$= -\sum_{x_1,x_2,\ldots,x_n} P_{A_1A_2\ldots A_n}(x_1,x_2,\ldots,x_n)\ln P_{A_1A_2\ldots A_n}(x_1,x_2,\ldots,x_n)$$

The *marginal entropy* is the entropy of a marginal ensemble. Thus $S(A)$, $S(B)$, $S(C)$, $S(A,B)$, $S(B,C)$ and $S(A,C)$ are all marginal entropies of the joint entropy

[6]Nobody knows why Clausius chose S for entropy. It's been suggested that his choice may have been in tribute to Sadi Carnot, but that's pure speculation.

[7]We'll postpone proving the such inequalities until we discuss relative entropy and Jensen's inequality **??**

$S(A, B, C)$.

The conditional entropy measures how uncertain we are (on average) about A when we know the state of B.

$$S(A \mid B) = -\sum_b P_B(b) \sum_a P_{AB}(a \mid b) \ln P_{AB}(a \mid b) \tag{13}$$
$$= -\sum_{ab} P_{AB}(a, b) \ln P_{AB}(a \mid b)$$

The conditional entropy is non-negative, since it is the expectation of non-negative entropies.

The *chain rule for entropies* [4, 3] follows from the probability chain rule (**??**),

$$S(A, B) = S(A \mid B) + S(B)\ . \tag{14}$$

It follows that conditioning always reduces entropy, $S(A \mid B) \leqslant S(A)$ and therefore that entropy is *subadditive*: The joint entropy is less than the sum of the individual entropies (with equality if and only if A and B are independent).

$$S(A, B) \leqslant S(A) + S(B) \tag{15}$$

5 Information

5.1 Mutual Information

$$I(A:B) = \sum_{x,y} P_{AB}(x,y) \ln \frac{P_{AB}(x,y)}{P_A(x)P_B(y)} \tag{16}$$

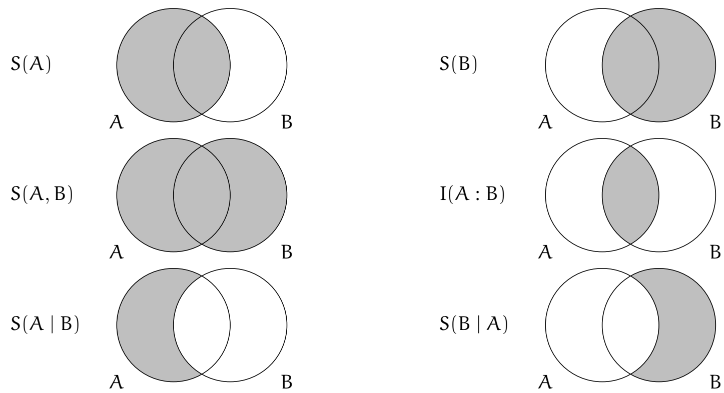

Figure 1: Two-variable Information diagrams [10, 11].

5.2 Relative Entropy

$$D(A\|B) = \sum_{x} P_A(x) \ln \frac{P_A(x)}{P_B(x)} \tag{17}$$

5.3 Differential entropy

5.4 Order and disorder

5.5 Further Reading

6 Jarzynski

6.1 Markov's Inequality

Markov's inequality[8] states that for a nonnegative random variable X the probability of observing a value above some threshold a is at most the mean of X divided by the threshold

$$P(X \geqslant a) \leqslant \frac{\langle X \rangle}{a} \tag{18}$$

The proof is straightforward.

$$\begin{aligned}
\langle X \rangle &= \int_{-\infty}^{\infty} P(x)\, x \, dx = \int_{0}^{\infty} P(x)\, x \, dx && (19a)\\
&= \int_{0}^{a} P(x)\, x\, dx + \int_{a}^{\infty} P(x)\, x\, dx && (19b)\\
&\geqslant \int_{a}^{\infty} P(x)\, x\, dx && (19c)\\
&\geqslant \int_{a}^{\infty} P(x)\, a\, dx && (19d)\\
&= a \int_{a}^{\infty} P(x)\, dx = a\, P(X \geqslant a) && (19e)
\end{aligned}$$

(a) We write the mean for a non-negative variable; (b) Split the average into two parts, before and after the threshold; (c) and throw away the first part, which must be positive; (d) replace the variable with the threshold inside the average; and (e) pull the threshold from out of the integral, and recognize the remaining expression as the desired tail probability.

6.2 Chernoff bound

Chernoff bound is an extension of Markov's inequality to all distributions, not just non-negative distributions.

$$P\left(X \geqslant a\right) \leqslant M_X(t)\, e^{-ta} \qquad \text{for } t > 0 \tag{20}$$

where $M_X(t)$ is the moment generating function $\langle e^{tx} \rangle$. This inequality follows by applying Markov's inequality to e^{tX}, with t positive.

$$\begin{aligned}
P\left(X \geqslant a\right) &= P\left(e^{tX} \geqslant e^{ta}\right) \qquad \text{for } t > 0\\
&\leqslant \frac{\langle e^{tX} \rangle}{e^{+ta}} = M_X(t)\, e^{-ta}
\end{aligned}$$

Similarly for the left tail we get

$$P\left(X \leqslant a\right) \leqslant M_X(t)\, e^{-ta} \qquad \text{for } t < 0 \tag{21}$$

[8] Named after the same Russian mathematician as for Markov chains.

6.3 Bounds on entropy production

Given the Jarzynski identity $\langle e^{-\Sigma} \rangle = 1$, we can apply the Chernoff bound to bound the probability that entropy decreases.

$$\begin{aligned}
P\,(\Sigma \leqslant \epsilon) &\leqslant \langle e^{+t\Sigma} \rangle \, e^{-t\epsilon} \quad \text{for } t < 0 && \text{(22a)}\\
&\leqslant \langle e^{-\Sigma} \rangle \, e^{+\epsilon} \quad \text{with } t = -1 && \text{(22b)}\\
&\leqslant e^{+\epsilon} && \text{(22c)}
\end{aligned}$$

7 Further Reading

I got another quarter hundred weight of books on the subject last night. I have not read them all through.

William Thomson (Lord Kelvin) Lecture IX, p87

Stochastic Thermodynamics: For a recent introduction to stochastic thermodynamics see Peliti and Pigolotti (2021) [1] Other general monographs and reviews include Evans and Searles (2002) [12], Harris and Schütz (2007) [13], Seifert (2012) [14], Spinney and Ford (2013) [15], Van den Broeck and Esposito (2015) [16].

Foundations – Thermodynamics and statistical mechanics Efficiency of heat engines and the foundation of thermodynamics: Carnot (1824) [17]; First law of thermodynamics: von Helmholtz (1847) [18]; Second law of thermodynamics Thomson (Lord Kelvin) [19] Clausius (1865) [8]; Entropy: Clausius (1865) [8]; Statistical definition of entropy: Boltzmann (1872) [7], Boltzmann (1898) [20], Planck (1901) [21], Gibbs (1902) [9], Shannon (1948) [4], Jaynes (1957) [22], Jaynes (1957) [23]; Foundations of statistical mechanics: Maxwell (1871) [24], Boltzmann (1896) [25], Boltzmann (1898) [20], Gibbs (1902) [9].

Bibliography

[0] [citation needed].

[1] Luca Peliti and Simone Pigolotti. *Stochastic Thermodynamics: An Introduction*. Princeton University Press (2021).

[2] Robert W. Zwanzig. *Nonequilibrium statistical mechanics*. Oxford University Press, New York (2001).

[3] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. Wiley, New York, 2nd edition (2006).
◦ Standard go-to reference for information theory. [Cover2006a]

[4] Claude E. Shannon. A mathematical theory of communication. *Bell Syst. Tech. J.*, 27:379–423, 623–656 (1948). doi:10.1002/j.1538-7305.1948.tb01338.x.
★ ◦ Foundation of information theory. Latter reprinted in book form with the subtle change in title "The Mathematical Theory of Communication". Entropy; mutual information; channel coding theorem; first use of word 'bit' in print. [Shannon1948a]

[5] Gavin E. Crooks. On measures of entropy and information. Technical report, (2015-2021). Tech. Note 009 v0.8 http://threeplusone.com/info.

[6] James R. Norris. *Markov Chains*. Cambridge University Press, Cambridge, England (1997).

[7] Ludwig Boltzmann. Weitere studien über das Wärmegleichgewicht unter Gasmolekülen (Further studies on the thermal equilibrium of gas molecules). *Sitzungsberichte Akad. Wiss., Vienna, part II*, 66:275–370 (1872).
◦ The Boltzmann equation and Boltzmann's H-theorem. Statistical definition of entropy, Eq 7. Boltzmann's derivation of the Mazwell-Boltzmann velocity distribution. First paper on non-equilibrium thermodynamics. [Boltzmann1872a]

[8] Rudolf Clausius. Ueber verschiedene für die anwendung bequeme formen der hauptgleichungen der mechanischen wärmetheorie. *Annalen der Physik und Chemie*, 201(7):353–400 (1865). doi:10.1002/andp.18652010702.
◦ Introduction of the term entropy, and statement of Clausius's form of the second law of thermodynamics. Quote: "We now seek an appropriate name for S . . . We would call S the transformation content of the body. However I have felt it more suitable to take names of important scientific quantities from the ancient languages in order that they may appear unchanged in all contemporary languages. Hence I propose that we call S the entropy of the body after Greek word . . . meaning "transformation". I have intentionally formed the word entropy to be as similar as possible to the word energy, since the two quantities that are given these names are so closely related in their physical significance that a certain likeness in their names has seemed appropriate." The paper concludes with concise statements of the First and Second Laws of thermodynamics: "1. The energy of the universe is a constant. 2) The entropy of the universe tends to a maximum." [Clausius1865a]

[9] J. Willard Gibbs. *Elementary principles in statistical mechanics*. Charles Scribner's Sons, New York (1902).
◦ Genesis and foundation of modern statistical mechanics. Nonstandard notations: energy: ϵ; temperature: θ; free energy: ψ; index of probability: $\nu \equiv \ln p_x$; entropy: $-\bar{\nu}$. Canonical ensemble (91). Statistical definition of entropy (111) p44. Second law expressed as an average work inequality (481). [Gibbs1902a]

[10] Raymond W. Yeung. A new outlook on Shannon's information measures. *IEEE Trans. Inf. Theory*, 37(3):466–474 (1991). doi:10.1109/18.79902.

[11] Ryan G. James, Christopher J. Ellison, and James P. Crutchfield. Anatomy of a bit: Information in a time series observation. *Chaos*, 21(3):037109 (2011). doi:10.1063/1.3637494.

[12] Denis J. Evans and Debra J. Searles. The fluctuation theorem. *Adv. Phys.*, 51(7):1529–1585 (2002). doi:10.1080/00018730210155133.

[13] R. J. Harris and G. M. Schütz. Fluctuation theorems for stochastic dynamics. *J. Stat. Mech.: Theor. Exp.*, page P07020 (2007). doi:10.1088/1742-5468/2007/07/P07020. Cond-mat/0702553v1.

∘ A good review that ties together a larage number of results in a single framework.w [Harris2007a]

[14] Udo Seifert. Stochastic thermodynamics, fluctuation theorems, and molecular machines. *Rep. Prog. Phys.*, 75(12):126001 (2012). doi:10.1088/0034-4885/75/12/126001. arXiv:1205.4176.

★ ∘ [Seifert2012a]

[15] Richard Spinney and Ian Ford. Fluctuation relations: A pedagogical overview. In Rainer Klages, Wolfram Just, and Christopher Jarzynski, editors, *Nonequilibrium Statistical Physics of Small Systems: Fluctuation Relations and Beyond*, chapter 1, pages 3–56. Wiley-VCH, Weinheim (2013). doi:10.1002/9783527658701.ch1.

[16] Christian Van den Broeck and Massimiliano Esposito. Ensemble and trajectory thermodynamics: A brief introduction. *Physica A*, 418:6–16 (2015). doi:10.1016/j.physa.2014.04.035.

[17] Sadi Carnot. *Réflexions sur la puissance motrice du feu et sur les machines propres a développer cette puissance*. Chez Bachelier, Libraire, Paris (1824). Reflections on the motive power of fire and machines fitted to develop that power.

∘ Foundation of thermodynamcs. Heat engines and Carnot efficiency. "The driving power of heat is independent of the agents used to realize it; its value is uniquely fixed by the temperatures of the bodies between which the transfer of caloric is made" [Carnot1824a]

[18] Hermann von Helmholtz. *Über die Erhaltung der Kraft* (*On the Conservation of Force*). Druck und Verlag, Berlin (1847). On the Conservation of Force.

∘ Definitive statement of the conservation of energy (And thereby the first law of thermodynamcis). (Terminology was not yet settled: Helmholtz's uses the word *force* for what we now call *energy*.) [Helmholtz1847a]

[19] William Thomson (Lord Kelvin). On the dynamical theory of heat, with numerical results deduced from Mr. Joule's equivalent of a thermal unit, and M. Regnault's observations on steam. *Trans. Roy. Soc. Edin.*, XX (part II):261–268; 289–298 (1851).

∘ First statement of (essentially) the second law of thermodynamics. "It is impossible, by means of inanimate material agency, to derive mechanical effect from any portion of matter by cooling it below the temperature of the coldest of the surrounding objects." [Thomson1851a]

[20] Ludwig Boltzmann. *Vorlesungen über Gastheorie* (*Lectures on gas theory*), volume II. J. A. Barth (1898). Translated in [26].

[21] M. Planck. Ueber das gesetz der energieverteilung im normalspektrum (on the law of distribution of energy in the normal spectrum). *Annalen der Physik*, 309(3):553–563 (1901). doi:10.1002/andp.19013090310.

[22] Edwin T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 106:620–630 (1957). doi:10.1103/PhysRev.106.620.

[23] Edwin T. Jaynes. Information theory and statistical mechanics II. *Phys. Rev.*, 108:171–190 (1957). doi:10.1103/PhysRev.108.171.

[24] James C. Maxwell. *Theory of heat*. Appleton, London (1871).

∘ Invocation of Maxwell's demon. [Maxwell1871a]

[25] Ludwig Boltzmann. *Vorlesungen über Gastheorie* (*Lectures on gas theory*), volume I. J. A. Barth (1896). Translated in [26].

[26] Ludwig Boltzmann. *Lectures on gas theory*. Univ. of Calif. Press (1964). Translation of [25] and [20].