Notes on Entropy

Jonathan Bain
Dept of Technology, Culture and Society; NYU-Tandon

These notes review the notion of entanglement entropy and other concepts of entropy including thermodynamic entropy, statistical mechanical entropy, Shannon entropy, von Neumann entropy, and black hole entropy.

1. Thermodynamic Entropy
In thermodynamics, one analyzes a single system (like a gas) in terms of its "macroscopic" thermodynamical properties (volume, pressure, temperature, etc.). A *thermodynamic equilibrium state* is a state of a single system in which its thermodynamical properties are constant.

Def. 1 (*Thermodynamic entropy*). The **thermodynamic entropy** $S_{\text{TD}}(\sigma_2)$ of a state σ_2 of a physical system is the ratio of the change in heat δQ_R to temperature T of a reversible process that connects an initial state σ_1 to σ_2:
$$S_{TD}(\sigma_2) \equiv \int_{\sigma_1}^{\sigma_2} \delta Q_R / T + const.$$

Note 1: $S_{\text{TD}}(\sigma_2)$ takes a maximum value when σ_2 is a thermodynamic equilibrium state.

Note 2: S_{TD} does *not* refer to information, particles, probability, uncertainty, disorder, *etc.*

Note 3: The 2nd Law of Thermodynamics applies to S_{TD} and it states that $\Delta S_{\text{TD}} = S_{\text{TD}}(\sigma_2) - S_{\text{TD}}(\sigma_1) \geq 0$. This is an empirical law: no thermodynamical system has ever been observed to violate it.

2. Entropy in Classical Statistical Mechanics
The goal of statistical mechanics is to provide a microphysical basis for thermodynamical systems. There are two ways of doing this: an approach due to Boltzmann, and an approach due to Gibbs.

2.1. Boltzmannian Statistical Mechanics
References: Frigg (2008); Frigg & Werndl (2011)

In Boltzmann's approach to statistical mechanics, one analyzes a single isolated system at constant energy (like a gas) in terms of its microphysical constituents (i.e., it's particles). Its thermodynamical properties are interpreted as macroproperties that reduce to the microproperties (i.e., positions, momenta) of its particles. A *microstate* of a system is represented by a point x in phase space Γ (labeled by positions and momenta). Let Γ_E be the subregion of Γ consisting of all microstates of the system with the same energy. A *macrostate* of the system is represented by a subset Γ_M of Γ_E. A *macroproperty* is then represented by a function on phase space, $f : \Gamma_E \to \mathbb{R}$. A *Boltzmann equilibrium macrostate* is the macrostate of the system with the greatest phase space volume. (There's no guarantee that the system will remain in it; i.e., no guarantee that macroproperties of this state remain constant.)

Def. 2 (*Boltzmann entropy*). The **Boltzmann entropy** $S_{\text{Boltz}}(\Gamma_M)$ of a macrostate Γ_M with size $|\Gamma_M|$ is given by:
$$S_{\text{Boltz}}(\Gamma_M) \equiv \ln|\Gamma_M|$$

Note 1: Insofar the number of microstates in Γ_M depends on its size, $S_{\text{Boltz}}(\Gamma_M)$ can be considered a measure of the number of microstates in Γ_M.

Note 2: The equilibrium macrostate is the macrostate that maximizes S_{Boltz}. So it's the macrostate that has the greatest size (or, equivalently, the greatest number of microstates).

Note 3 *(probability interpretation)*: Suppose our system consists of N particles. So Γ is $6N$-dim (3 position variables and 3 momentum variables for each particle). Its points represent the microstates of our system (each microstate is a description of all the N particles in terms of their positions and momenta at an instant in time). We can also consider the 6-dim *single-particle* phase space, call it Γ_μ, with $\Gamma = \Gamma_\mu \times \Gamma_\mu \times \cdots$ (N times). A point x_μ of Γ_μ represents a microstate of a single particle of our system. Suppose we "coarse-grain" Γ_μ by dividing it into ℓ cells. Now let $\{n_1, ..., n_\ell\}$ be a "Boltzmann distribution" of single-particle microstates, where n_i is the number of single-particle microstates in the ith cell, and let $\{p_1, ..., p_\ell\}$, $p_i = n_i/N$, be a probability distribution over the single-particle microstates (p_i is the probability of finding a single-particle microstate in the ith cell). Then (after a bit of algebra) the Boltzmann entropy $S_{\text{Boltz}}(\Gamma_M)$ of Γ_M can be re-expressed as
$$S_{\text{Boltz}}(\Gamma_M) = -N\sum_i p_i \ln p_i$$
or, in continuous notation,
$$S_{\text{Boltz}}(\Gamma_M) = -N\int_{\Gamma_\mu} \rho_\mu(x_\mu)\ln\rho_\mu(x_\mu)dx_\mu$$
where $\rho_\mu(x_\mu)$ is a probability distribution over single-particle microstates in Γ_μ. Note that this is an integral over the single-particle phase space Γ_μ, as opposed to the N-particle phase space Γ of our system. This should be compared with the expression for the Gibbs entropy below.

Note 4 (uncertainty interpretation): Suppose $-\ln p_i$ is the information gained if a particle is found to be in a microstate in the ith cell of Γ_μ (motivation: the greater p_i, the more certain that the microstate of a particle is in the ith cell of Γ_μ and the less information associated with this result). Then $S_{\text{Boltz}}(\Gamma_M)/N$ can be interpreted as the expected information gain upon finding a particle of an N-particle system to be in a microstate in the ith cell of Γ_μ. So the larger $S_{\text{Boltz}}(\Gamma_M)$, the greater the information gained upon finding a particle's microstate in the ith cell of Γ_μ, hence the greater the uncertainty of this microstate.

2.2. Gibbsian Statistical Mechanics
References: Frigg (2008); Frigg & Werndl (2011); Greiner, _et al._ (1995, Chaps 6 & 7); Reif (1965, Chap 6)

In Gibbs' approach to statistical mechanics, one analyzes an _ensemble_ of infinitely many copies of a multi-particle system. The motivation for this is that observed macroscopic properties of a system are time averages of the properties of its microstates. But time averages are hard to calculate (try to keep track of the positions and momenta of all the particles in a gas)! So Gibbs' idea was to replace the _time average_ of a single system over a period of time by an _ensemble average_ of infinitely many copies of it at an instant of time. In this approach, a point in phase space Γ represents a microstate of a member of an ensemble (at an instant in time), and the state of the entire ensemble is represented by a "Gibbs distribution"; i.e., a probability distribution $\rho(x)$ defined on Γ (note that this different from the single-particle probability distribution $\rho_\mu(x_\mu)$ defined on Γ_μ in Note 3 of Section 2.1 above). The probability at time t of finding the system's microstate in a region S of Γ is represented by the integral $\int_S \rho(x)dx$ of $\rho(x)$ over S. You might think of Γ as a big (continuous) set of possible states, and $\rho(x)$ as assigning probabilities to all of these states. For any property represented by a function f on Γ, we can now define an _ensemble average_ $\langle f \rangle$ as the weighted sum of f over all points in Γ, weighted by the probability distribution ρ:

$$\langle f \rangle \equiv \int_\Gamma f(x)\rho(x)dx$$

$\langle f \rangle$ is constant just when ρ is stationary (i.e., doesn't change in time). So, if thermodynamic properties are represented by ensemble averages, then they do not change in time for stationary distributions. The Gibbs approach now adopts the following principle:

Averaging Principle: The measured value of a thermodynamic property f of a system in thermodynamic equilibrium is the ensemble average $\langle f \rangle$ of an ensemble in statistical equilibrium (i.e., an ensemble characterized by a stationary distribution).

This is usually justified by the following argument: A measurement takes a finite amount of time, which is infinite when compared to molecular processes. So what actually gets measured in a lab is an _infinite time average_ $f^*(x_0) \equiv \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t_0}^{t_0+\tau} f(\varphi_t(x_0))dt$, where x_0 is the initial state of the system at time t_0, and φ_t is a time-evolution operator. And for sufficiently chaotic (or "ergodic") systems, one can show that $\langle f \rangle = f^*(x_0)$.

In the Gibbs approach, entropy is defined by the following:

> **Def. 3** (*Gibbs entropy*). The **Gibbs entropy** $S_{\text{Gibbs}}(\rho)$ of an ensemble distribution ρ is the ensemble average of the quantity $-\ln\rho$:
> $$S_{\text{Gibbs}}(\rho) \equiv -\int_{\Gamma} \rho(x)\ln\rho(x)dx$$

Note 1 (*fine-grained vs. coarse-grained Gibbs entropy*):

Since a Gibbs distribution ρ is required to be stationary (constant in time), the Gibbs entropy $S_{\text{Gibbs}}(\rho)$ is constant in time, too. So how can it increase over time? One solution is to coarse-grain it. We can divide the full phase space Γ into cells ω of size $\delta\omega$ (just like Boltzmann did with the single-particle phase space Γ_μ), and for a given Gibbs distribution ρ, we then define a coarse-grained version of it ρ_{coarse} by

$$\rho_{\text{coarse}}(x) \equiv \frac{1}{\delta\omega}\int_{\omega(x)} \rho(x')dx'$$

where $\omega(x)$ is the cell in which the microstate x lies, and the integral is over all microstates x' in that cell. This definition entails that ρ_{coarse} assigns to every microstate x in a given cell ω the average of the values that ρ assigns to all microstates in ω. (So whereas ρ may assign different values to all the microstates in a given cell, ρ_{coarse} just assigns one value to them; namely, their "ρ-average".) One can then show that, for any stationary Gibbs distribution ρ, $S_{\text{Gibbs}}(\rho) \leq S_{\text{Gibbs}}(\rho_{\text{coarse}})$. So we might say that, if the Gibbs entropy at an initial time is given by the fine-grained Gibbs entropy, and if it's given at a later time by the coarse-grained entropy, then it either remains unchanged or increases. (Which might not be all *that* helpful. What might be more helpful is the fact that ρ_{coarse} is not necessarily stationary: Technically, stationarity follows from Liouville's theorem, which is a general characteristic of a phase space, and you can show that Liouville's theorem doesn't apply to ρ_{coarse}. You can then adopt various assumptions (like ergodicity, or "molecular randomness", etc.) that entail that $S_{\text{Gibbs}}(\rho_{\text{coarse}})$ remains unchanged or increases over time.)

Note 2: In Note 3 of Section 2.1, S_{Boltz} was expressed in a form that is similar to S_{Gibbs}, but that form was an integral over the single-particle phase space Γ_μ of the "constant-energy" subregion Γ_E of phase space Γ. In contrast, S_{Gibbs} is an integral over the full phase space Γ (and not even the restriction of Γ to Γ_E). We can, however, show that S_{Gibbs} reduces to S_{Boltz} for the case of a system of N identical, non-interacting particles at constant energy. Such a system is characterized by $\rho(x) = \prod_{i=1}^{N} \rho_\mu^i(x_\mu)$, where $\rho_\mu^i(x_\mu)$ is the ith single-particle probability distribution, defined on Γ_μ, and $\rho_\mu^i = \rho_\mu^j$ for all i, j. Then

$$\begin{aligned}
S_{\text{Gibbs}}(\rho) &= -\int_{\Gamma} [\rho_\mu^1(x_\mu^1)\cdots\rho_\mu^N(x_\mu^N)]\ln[\rho_\mu^1(x_\mu^1)\cdots\rho_\mu^N(x_\mu^N)]dx \\
&= -\int_{\Gamma_\mu} \rho_\mu^1(x_\mu^1)dx_\mu^1 \cdots \int_{\Gamma_\mu} \rho_\mu^N(x_\mu^N)dx_\mu^N \ln\rho_\mu^1(x_\mu^1) - \cdots \\
&\quad - \int_{\Gamma_\mu} \rho_\mu^1(x_\mu^1)dx_\mu^1 \cdots \int_{\Gamma_\mu} \rho_\mu^N(x_\mu^N)dx_\mu^N \ln_\mu^N(x_\mu^N) \\
&= -\int_{\Gamma_\mu} \rho_\mu^1(x_\mu^1)dx_\mu^1 \ln\rho_\mu^1(x_\mu^1) - \cdots - \int_{\Gamma_\mu} \rho_\mu^N(x_\mu^N)dx_\mu^1 \ln\rho_\mu^N(x_\mu^N) \\
&= -N\int_{\Gamma_\mu} \rho_\mu(x_\mu)\ln\rho_\mu(x_\mu)dx_\mu \\
&= S_{\text{Boltz}}(\Gamma_M)
\end{aligned}$$

where in the second line, $dx = dx_\mu^1\cdots dx_\mu^N$, recalling that $\Gamma = \Gamma_\mu \times \Gamma_\mu \times \cdots$ (N times), and in the third line $\int_{\Gamma_\mu} \rho_\mu^i(x_\mu^i)dx_\mu^i = 1$.

Note 3 _(uncertainty interpretation)_: Suppose $-\ln\rho(x)$ is the information gained if our system is found to be in microstate x. Then $S_{\text{Gibbs}}(\rho)$ can be interpreted as the expected information gain upon finding an N-particle system to be in microstate x. So the larger $S_{\text{Gibbs}}(\rho)$, the greater the information gained upon finding an N-particle system to be in microstate x, hence the greater the uncertainty of this microstate.

Note 4: In principle for any given physical system, there can be many different ways to define a Gibbs ensemble distribution ρ. Minimally, we want a ρ that is stationary and that maximizes S_{Gibbs}. Two standard choices are the _microcanonical distribution_ ρ_{mc} and the _canonical distribution_ ρ_{c}.

Microcanonical Distribution
The _microcanonical distribution_ ρ_{mc} describes an isolated system with constant energy $H(x) = E$ (or more precisely, with $H(x)$ in the range $[E, E+\Delta E]$). It takes the form:

$$\rho_{\text{mc}}(x) = \begin{cases} 1/\Omega(E), & \text{for } H(x) \in [E, E + \Delta E] \\ 0, & \text{otherwise} \end{cases}$$

where $\Omega(E) = \int_{[E, E+\Delta E]} dx$ is the number of microstates with $H(x) \in [E, E+\Delta E]$. The microcanonical Gibbs entropy is given by:

$$S_{\text{Gibbs}}(\rho_{\text{mc}}) = \ln\Omega(E)$$

Note 1: The motivation for $\rho_{\text{mc}}(x)$ comes from the assumption that all microstates with the same energy have equal probability. So if there are $\Omega(E)$ of them, then each one should be assigned the probability $1/\Omega(E)$.

Note 2: The Gibbs entropy of $\rho_{\text{mc}}(x)$ is explicitly a measure of the number of possible microstates of the system:

$$\begin{aligned}
S_{\text{Gibbs}}(\rho_{\text{mc}}) &= -\int_{\Gamma} \rho_{\text{mc}}(x)\ln\rho_{\text{mc}}(x)dx \\
&= -\int_{[E, E+\Delta E]} \frac{1}{\Omega(E)}\ln\frac{1}{\Omega(E)}dx \\
&= -\frac{1}{\Omega(E)}\ln\frac{1}{\Omega(E)}\int_{[E, E+\Delta E]} dx \\
&= \ln\Omega(E)
\end{aligned}$$

Note 3: The Boltzmann entropy $S_{\text{Boltz}}(\Gamma_M)$ of a macrostate of an isolated physical system at constant energy might be thought of as the Gibbs entropy $S_{\text{Gibbs}}(\rho_{\text{mc}})$ of the microcanonical distribution for a system of N identical, non-interacting particles.

Canonical Distribution
The _canonical distribution_ ρ_{c} describes a composite closed system, consisting of a heat bath R and a subsystem P, at fixed temperature T and constant total energy $H(x) = E = E_R + E_P$, $E_S \ll E$. It takes the form:

$$\rho_{\text{c}}(x) = Z^{-1}e^{-\beta H(x)},$$

where $Z = \int e^{-\beta H(x)} dx$ and $\beta = 1/T$. The canonical Gibbs entropy is given by:

$$S_{\text{Gibbs}}(\rho_c) = \beta \langle H \rangle + \ln Z$$

<u>*Note* 1</u>: Here's how to derive the form of ρ_c. First suppose that P is in a microstate i with energy E_i. This means that R must be in a microstate with energy $E_R = E - E_i$. It also means that the number of microstates $\Omega_{(P+R)}(E)$ that the combined system $P+R$ can be in is just the number $\Omega_R(E_R)$ that R can be in (since P is already in a particular microstate). So $\Omega_{(P+R)}(E) = \Omega_R(E_R) = \Omega_R(E - E_i)$. Now assume that the probability p_i that P is in microstate i is proportional to the number of microstates that the combined system $P+R$ can be in, when P is in the microstate i. In other words, $p_i \propto \Omega_R(E - E_i)$. Since $E_i \ll E$, we can Taylor expand $\ln \Omega_R(E - E_i)$ about E:

$$\ln \Omega_R(E - E_i) \approx \ln \Omega_R(E) - (\partial/\partial E) k \ln \Omega_R(E) E_i + \cdots$$
$$= \ln \Omega_R(E) - (\partial S_R/\partial E) E_i + \cdots \qquad \text{where } S_R = \ln \Omega_R(E)$$
$$= \ln \Omega_R(E) - (1/T) E_i + \cdots \qquad \text{where } \partial S_R/\partial E = 1/T$$

So $\Omega_R(E - E_i) \approx \Omega_R(E) e^{-\beta E_i}$, where $\beta = 1/T$, and $\Omega_R(E) = $ const. And this means $p_i \propto e^{-\beta E_i}$. If we impose the normalization condition $\sum_i p_i = 1$, then we get

$$p_i = e^{-\beta E_i}/\sum_i e^{-\beta E_i} = Z^{-1} e^{-\beta E_i}$$

where $Z = 1/\sum_i e^{-\beta E_i}$. $\rho_c(x)$ is then the continuous version of p_i.

<u>*Note* 2</u>: $S_{\text{Gibbs}}(\rho_c)$ is obtained by taking the ensemble average $\langle -\ln \rho_c \rangle$:

$$S_{\text{Gibbs}}(\rho_c) = \langle -\ln \rho_c \rangle = \int_\Gamma \rho_c [-\ln \rho_c] dx$$
$$= \int_\Gamma \rho_c [-\ln (e^{-\beta H(x)}/Z)] dx$$
$$= \int_\Gamma \rho_c [-(\ln e^{-\beta H(x)} - \ln Z)] dx$$
$$= \int_\Gamma \rho_c [\beta H(x) + \ln Z] dx$$
$$= \beta \langle H \rangle + \ln Z \int_\Gamma \rho_c dx$$
$$= \beta \langle H \rangle + \ln Z \qquad \text{where } \int_\Gamma \rho_c dx = 1$$

<u>*Note* 3</u>: Note that $-\partial Z/\partial \beta = \int_\Gamma H(x) e^{-\beta H(x)} dx = Z \langle H \rangle$. So $\langle H \rangle = Z^{-1}(-\partial Z/\partial \beta) = -(\partial/\partial \beta) \ln Z$. So another way of expressing $S_{\text{Gibbs}}(\rho_c)$ is

$$S_{\text{Gibbs}}(\rho_c) = -(\beta \partial/\partial \beta - 1) \ln Z$$

3. Entropy in Classical Information Theory: Shannon Entropy

Reference: Frigg & Werndl (2011)

In Gibbs' approach to statistical mechanics, we considered an ensemble of classical states, where each state is a point in phase space Γ, and a Gibbs distribution ρ is a continuous probability distribution on Γ that assigns a probability to each state. The Gibbs entropy is the ensemble average of the quantity $-\ln \rho$. The Shannon entropy can initially be thought of as a generalization of Gibbs' approach: We view $-\ln \rho$ as a measure of "information", and

we generalize the notion of a classical phase space Γ of microstates x to a random variable X with possible values x.

Def. 4 (*Shannon entropy*). Let X be a random variable with possible values $\{x_1, ..., x_\ell\}$ and probability distribution $\{p_1, ..., p_\ell\}$. The **Shannon entropy** $S_{\text{Shan}}(X)$ of X is given by

$$S_{\text{Shan}}(X) \equiv -\sum_i p_i \log_2 p_i$$

<u>Note</u> 1: The expected value of X is $\sum_i p_i x_i$. The information gained if X is measured to have value x_i is $-\log_2 p_i$ (motivation: the greater p_i, the more certain that the value of X is x_i, and the less information associated with this result). Thus $S_{\text{Shan}}(X)$ is the expected information gain upon measurement of X. So the larger $S_{\text{Shan}}(X)$, the greater the information gained upon measuring X, and the greater the uncertainty of its measured value.

<u>Note</u> 2 *(comparison)*:
- S_{Shan}, S_{Gibbs}, and S_{Boltz} can all be interpreted as measuring uncertainty:
  - The greater $S_{\text{Shan}}(X)$, the greater the uncertainty of the value of the random variable X.
  - The greater $S_{\text{Gibbs}}(\rho)$, the greater the uncertainty of finding the system in a microstate in Γ.
  - The greater $S_{\text{Boltz}}(\Gamma_M)$, the greater the uncertainty of finding a particle of an N-particle system in a single-particle microstate in Γ_μ.
- S_{TD} has nothing to do with probability or uncertainty.
- The 2nd Law only applies to S_{TD}. It does not apply to S_{Shan}, S_{Boltz}, or S_{Gibbs}.
- Note that under certain circumstances (large number of weakly interacting distinguishable particles undergoing reversible process) it is consistent to identify S_{Boltz} and S_{Gibbs} with S_{TD}; but conceptually they measure different quantities.

4. Entropy in Quantum Mechanics: von Neumann Entropy, Entanglement Entropy

Recall from the "Notes on Quantum Entanglement" that in quantum mechanics a density operator ρ represents an ensemble $\{|\psi_i\rangle, p_i\}$ of vector states $|\psi_i\rangle$, each with a given probability p_i. The same symbol "$\rho(x)$" also appears in Gibbs' approach to classical statistical mechanics, where you can think of it as representing a "continuous" ensemble of classical states: each classical state is represented by a point x in phase space Γ, and $\rho(x)$ assigns a probability to each x. So density operators in quantum mechanics are the correlates of Gibbs distributions in Gibbsian classical statistical mechanics.

From the "Notes on Quantum Entanglement", recall the following definitions:

Def. 5 (*von Neumann entropy*). The **von Neumann entropy** $S_{\text{vN}}(\rho)$ of a density operator state ρ is defined by

$$S_{\text{vN}}(\rho) \equiv -\text{Tr}(\rho \ln \rho)$$

Note: Recall that this can also be written as $S_{vN}(\rho) = -\sum_i p_i \ln p_i$, where p_i is the probability that ρ assigns to $|\psi_i\rangle$.

Def. 6 (*Entanglement entropy*). For a bipartite system AB with density operator ρ_{AB}, the **entanglement entropy** S_A of subsystem A is defined to be the von Neumann entropy of ρ_A:
$$S_A \equiv S(\rho_A) = -\text{Tr}(\rho_A \ln \rho_A)$$

Note 1: Evidence for taking entanglement entropy as thermodynamic entropy comes from arguments (e.g., due to von Neumann) that claim (a) a measurement in quantum mechanics is associated with an increase in thermodynamic entropy, and (b) a post measurement state is a mixed state (Prunkl 2020).

Note 2: Recall that $S_{vN}(\rho)$ varies from zero (for a pure state) to $\ln n$ for a maximally mixed state $\rho_{max} = (1/n) I_n$, where n is the dimension of the Hilbert space $\mathcal{H}$. Note that the dimension of a Hilbert space is the number of distinct possible states contained in it. Now recall that the Gibbs entropy of a microcanonical distribution is given by $S_{Gibbs}(\rho_{mc}) = \ln \Omega(E)$, where $\Omega(E)$ is the number of distinct possible states with energy E. So the von Neumann entropy of the state of an isolated physical system at constant energy E is bounded above by the microcanonical Gibbs entropy of that state:

$$S_{vN}(\rho) \leq S_{Gibbs}(\rho_{mc})$$

Of course this assumes the physical system is a quantum system, so that the von Neumann entropy of one of its states makes sense. So we have to assume that the classical Gibbs entropy can be applied to quantum density operator states. This relation between the von Neumann entropy and the Gibbs entropy (or the "microcanonical" entropy) of an isolated system at constant energy is important in Page's (1993) account of how the entanglement entropy of the Hawking radiation of an evaporating black hole behaves.

5. Black Hole Entropy

Def. 7 (*Black hole entropy*). The **Bekenstein-Hawking entropy** of a black hole is given by
$$S_{BH} \equiv \text{Area(horizon)}/4G$$
where G is the Newtonian gravitational constant.

Note 1: The form for S_{BH} can be motivated either by thermodynamical or statistical mechanical considerations:
- *Thermodynamical motivations*:
  - Small-scale interactions involving black holes can be characterized as "reversible" and "irreversible" with $A = \text{Area(horizon)}$ playing the role of S_{TD}.
  - Stationary black holes satisfy $dM = (1/8\pi)\kappa dA + \cdots$, where κ is surface gravity. This is formally identical to $dU = T dS_{TD} + \cdots$ if $S_{TD} = A/4$ and $T = (1/2\pi)\kappa$.
  - Hawking's (1971) area theorem: $\Delta A \geq 0$ in finite processes involving black holes.
- *Statistical mechanical motivations*:

- If black holes are thermodynamical objects, they should have a statistical mechanical description, too. In particular, an isolated black hole with constant energy E should have a Gibbs microcanonical entropy $S_{\text{Gibbs}}(\rho_{\text{mc}}) = \ln \Omega(E)$. If this is equated with the black hole's entropy S_{BH}, we obtain a formula for the number of possible states of the black hole: $\Omega_{\text{BH}}(E) = e^{A/4G}$. Evidence for this formula comes from calculations in effective field theory, string theory, and AdS/CFT.

Note 2: Evidence for taking S_{BH} as thermodynamic entropy comes from arguments that appeal to the 2nd Law of thermodynamics (Bekenstein 1973), and arguments that indicate that a black hole can function as an element in a Carnot cycle (Prunkl & Timpson 2019).

6. Entanglement Entropy in Quantum Field Theory
The entanglement entropy S_A of the degrees of freedom of a massless scalar field inside a spherical region A of spacetime with respect to its degrees of freedom outside A is proportional to the area of the boundary ∂A of A (Srednicki 1993; Callan & Wilczek 1994).
Note 1: The boundary ∂A is called the *entangling surface*: it is a surface that divides the scalar field into two subsystems, and we can thing of the entanglement between these subsystems as entanglement across ∂A.

7. Entanglement Entropy in AdS/CFT: The RT Formula
References: von Raamsdonk (2017, Sec 3.3-3.5); Harlow (2016, Sec VI.J); Headrick (2019, Sec 5.2)

Entanglement entropy and black hole entropy are related in the AdS/CFT correspondence by means of the Ryu-Takayanagi (RT) formula.

Def. 8 (*RT Formula with quantum correction*). Let $\rho_{A\bar{A}}$ be the density operator for a bipartite CFT system localized on a boundary spatial region $A \cup \bar{A}$. Then the entanglement entropy S_A of the subsystem localized on A is given by
$$S_A \equiv \text{Area}(\gamma_A)/4G + S_{\Sigma_A}$$
where γ_A is the minimal area bulk spatial surface with the same boundary as A; Σ_A is the bulk surface bounded by $\gamma_A \cup A$; and S_{Σ_A} is the entanglement entropy for a bulk subsystem, localized in Σ_A, of a bulk bipartite system localized in $\Sigma_A \cup \bar{\Sigma}_A$.

Note 1: Recall from Def. 6 that to calculate the entanglement entropy of a subsystem A of a bipartite system, we need to know the reduced density operator state ρ_A of the subsystem (and then we calculate S_A according to the formula $S_A = -\text{Tr}(\rho_A \ln \rho_A)$). The RT Formula gives us a way of calculating S_A without needing to know what ρ_A is. Rather, as long as we can view the subsystem A as located on a boundary spatial subregion of a bulk spacetime, we can calculate S_A in terms of the area of a particular bulk surface γ_A, and the entanglement entropy S_{Σ_A} of bulk subsystems in regions separated by γ_A.

Note 2: If the bulk spacetime contains a black hole, then in the limit in which A encompasses the entire boundary spatial region, γ_A becomes the event horizon, and the RT

formula (without the "quantum" correction) reduces to the Bekenstein-Hawking black hole entropy.

Note 3: As seen in Section 6, the RT formula has a precedent in quantum field theory. Note that the entangling surface in the RT formula is γ_A: this surface separates the boundary system into two parts, and it also separates the bulk system into two parts.

Note 4: Note 2 suggests that the Bekenstein-Hawking black hole entropy might be interpreted as the entanglement entropy between interior degrees of freedom and exterior degrees of freedom, separated by the event horizon.

References

Bekenstein, J. (1973) 'Black Holes and Entropy', *Physical Review D7*, pp. 2333-2346.

Callan, C. and F. Wilczek (1994) 'On Geometric Entropy', *Physics Letters B*, pp. 55-61.

Frigg, R. (2008) 'A Field Guide to Recent Work on the Foundations of Statistical Mechanics', in D. Rickles (ed.) *The Ashgate Companion to Contemporary Philosophy of Physics*, Ashgate.

Frigg, R. and C. Werndl (2011) 'Entropy: A Guide for the Perplexed', in C. Beisbart & S. Hartmann (eds.) *Probabilities in Physics*, Oxford Univ. Press, pp. 115-42.

Greiner, W., L. Neise, H. Stocker (1995) *Thermodynamics and Statistical Mechanics*, Springer.

Headrick, M. (2019) 'Lectures on Entanglement Entropy in Field Theory and Holography', arXiv:1907.08126v1.

Harlow, D. (2016) 'Jerusalem Lectures on Black Holes and Quantum Information', *Reviews of Modern Physics 88*, 015002.

Page, D. (1993) 'Information in Black Hole Radiation', *Physical Review Letters 71*, 3743-3746.

Prunkl, C. (2020) 'On the Equivalence of von Neumann and Thermodynamic Entropy', *Philosophy of Science 87*, 262-280.

Prunkl, C. and C. Timpson (2019) 'Black Hole Entropy is Thermodynamic Entropy', arXiv:1903.06276v1.

Reif, F. (1965) *Fundamentals of Statistical and Thermal Physics*, McGraw-Hill.

Srednicki, M. (1993) 'Entropy and Area', *Physical Review Letters 71*, pp. 666-669.

von Raamsdonk, M. (2017) 'Lectures on Gravity and Entanglement', in J. Polchinski, P. Vieria, O. DeWolfe (eds.) *TASI 2015: New Frontiers in Fields and Strings*, World Scientific, pp. 297-351.