



General Physics I

Lecture 27: Entropy and Information



Outline

Introducing entropy

The meaning of entropy

- **Reversibility**
- **Disorder**
- **Information**

Selected topics on information



1st & 2nd Laws of Thermodynamics

The 1st law specifies that we **cannot get more** energy out of a cyclic process by work than the amount of energy we put in.

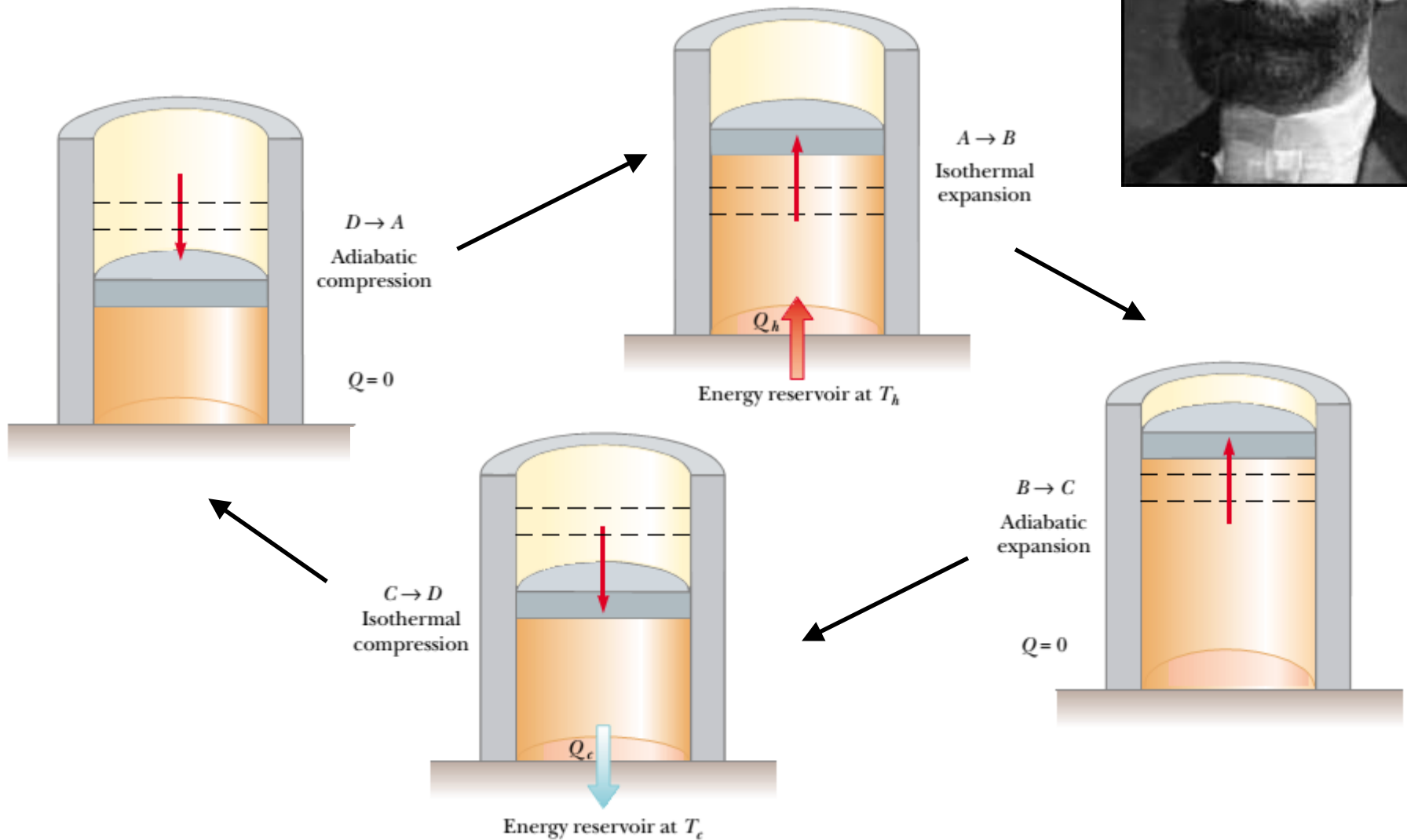
$$\Delta U = Q - W$$

The 2nd law states that we **cannot break even** because we must put more energy in, at the higher temperature, than the net amount of energy we get out by work.

$$\varepsilon = \frac{W}{Q_h} = 1 - \frac{|Q_c|}{Q_h} \leq \varepsilon_{\text{carnot}} = 1 - \frac{T_c}{T_h}$$

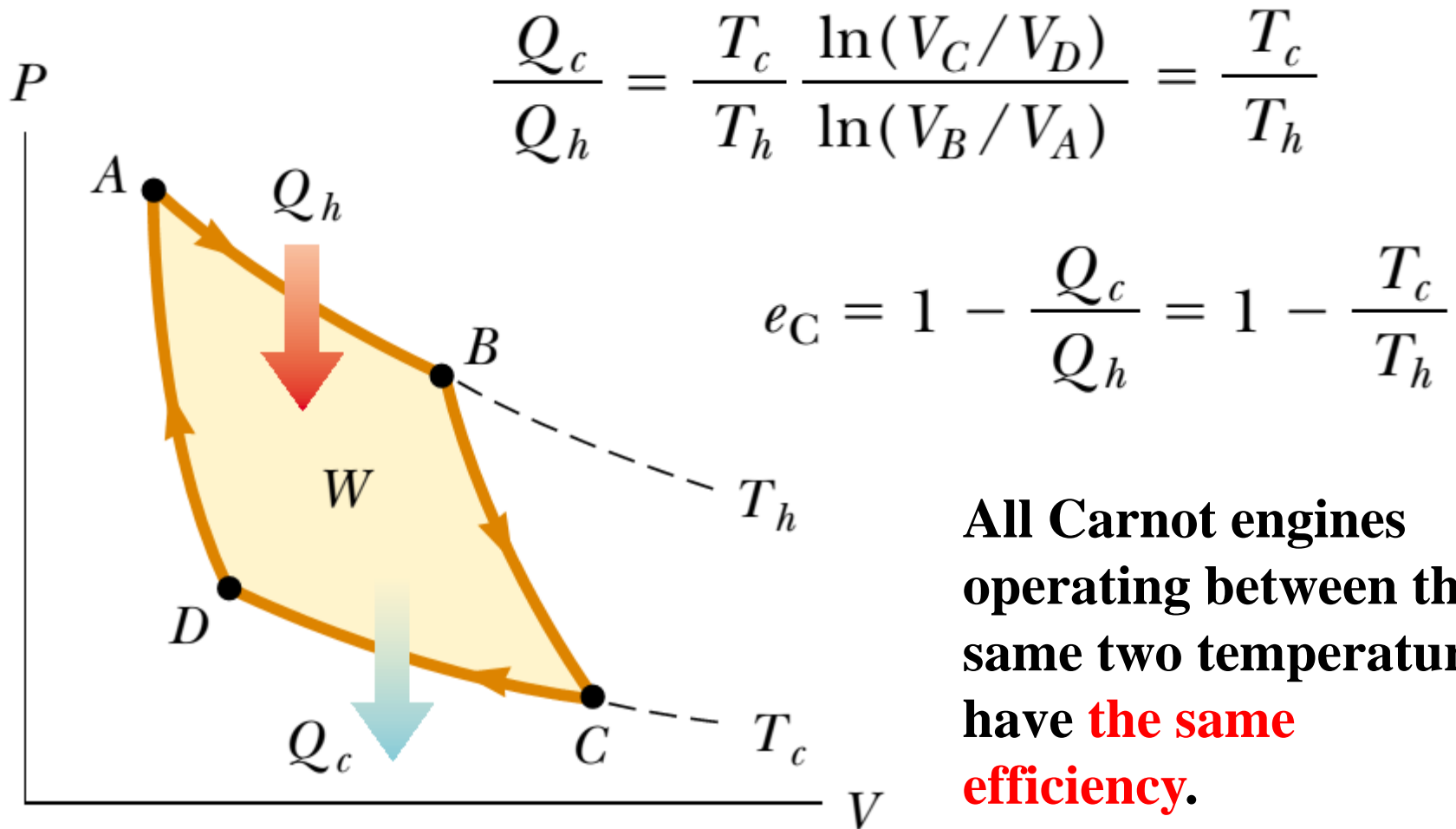


Carnot's Engine



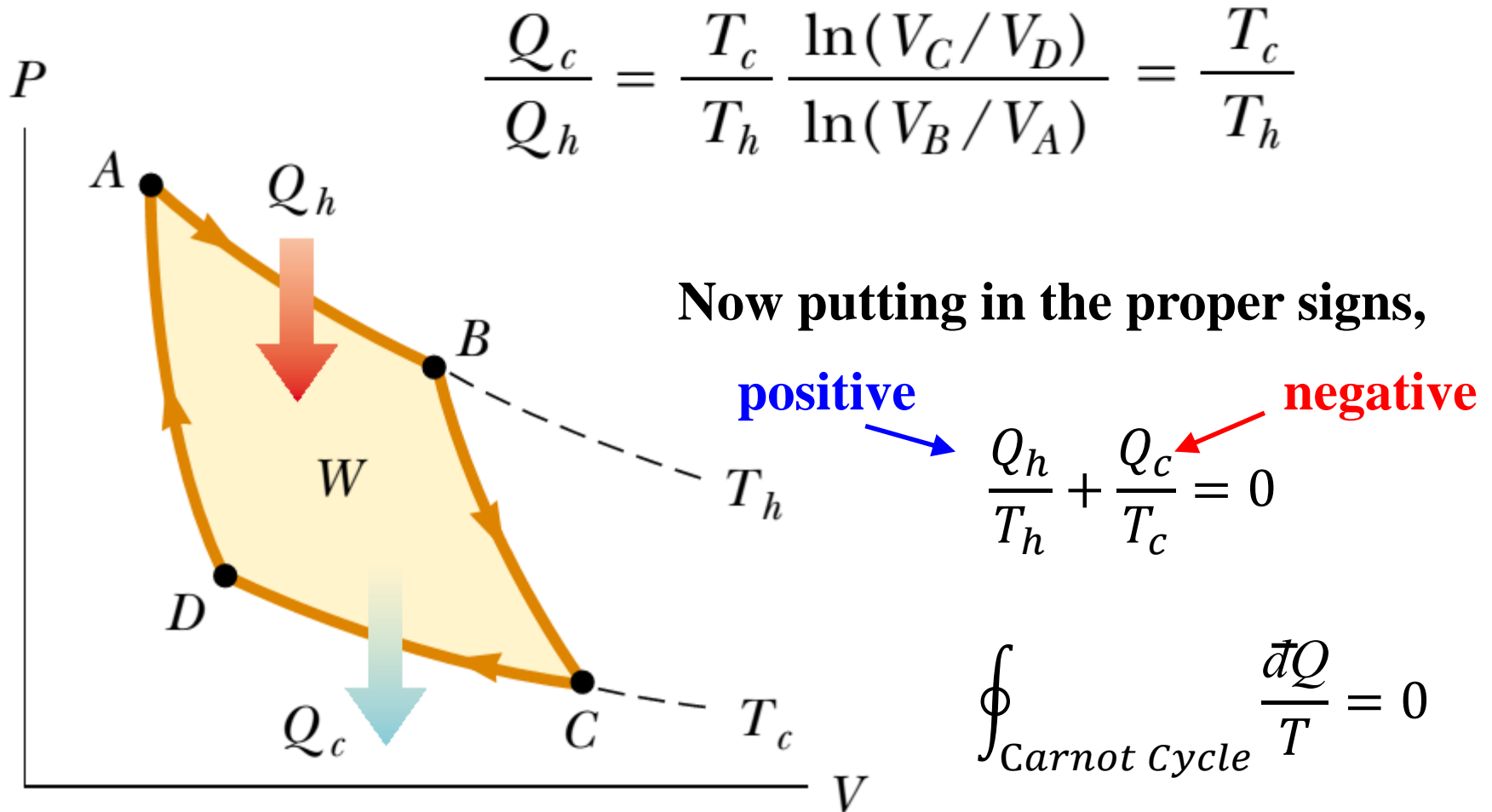


Efficiency of a Carnot Engine



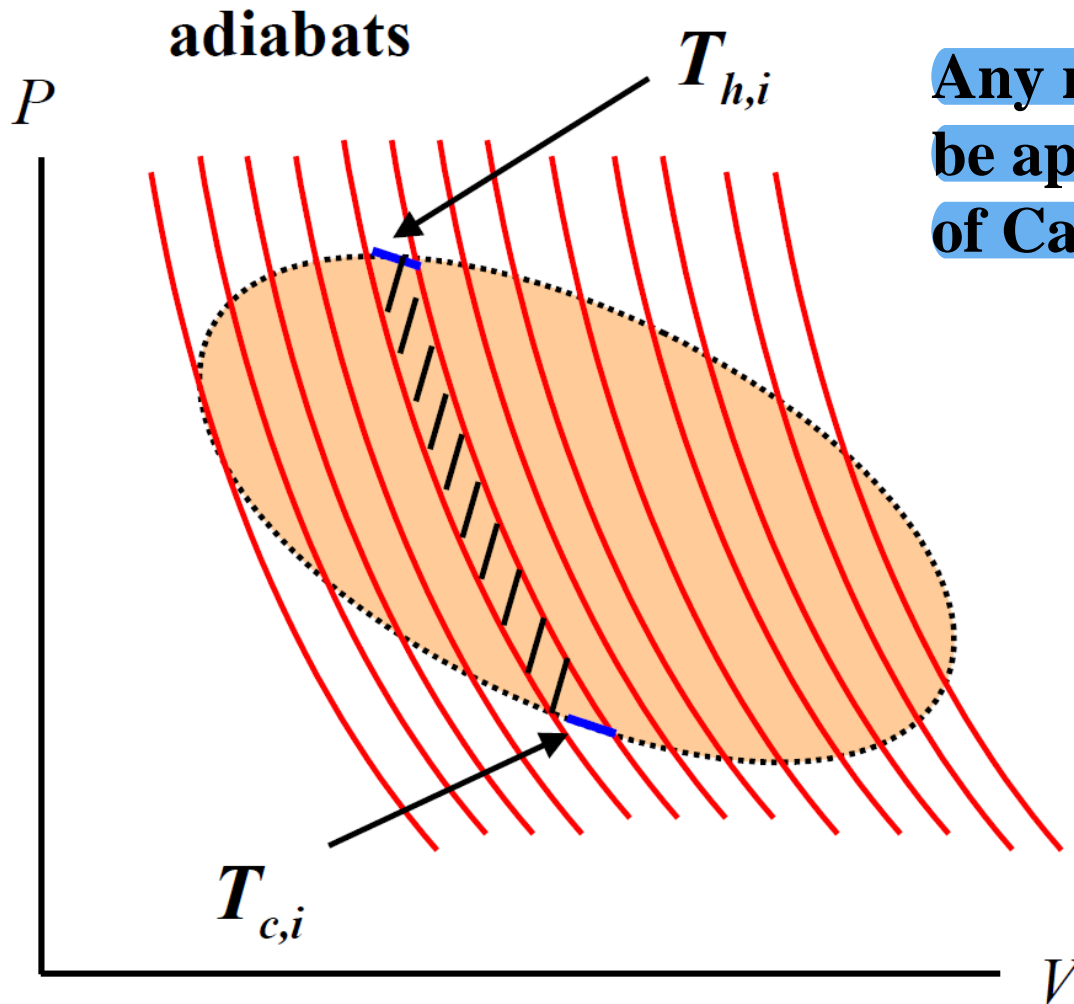


An Equality





A Sum of Carnot Cycles



Any reversible process can be approximated by a sum of Carnot cycles, hence

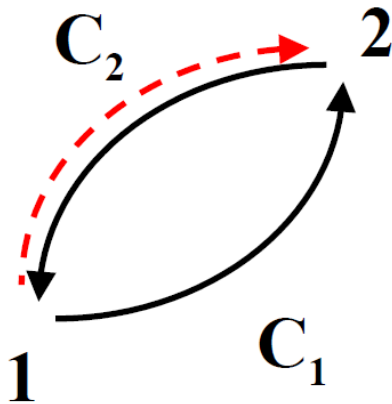
$$\sum_i \frac{Q_{h,i}}{T_{h,i}} + \frac{Q_{c,i}}{T_{c,i}} = 0$$

$$\oint_C \frac{\delta Q}{T} = 0$$



Clausius Definition of Entropy

Entropy is a **state function**, the change in entropy during a process depends only on the end points and is independent of the actual path followed.



$$dS = \frac{\delta Q_{\text{reversible}}}{T}$$

$$\int_{C_1, 1 \rightarrow 2} dS + \int_{C_2, 2 \rightarrow 1} dS = \oint_C dS = 0$$

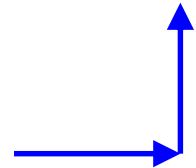
$$S_2 - S_1 = \int_{C_1, 1 \rightarrow 2} dS = - \int_{C_2, 2 \rightarrow 1} dS = \int_{C_2, 1 \rightarrow 2} dS$$



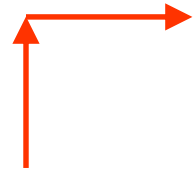
Return to Inexact Differential

Assume $\bar{d}g = dx + \frac{x}{y} dy$

$$\left(\int_{(1,1)}^{(2,1)} + \int_{(2,1)}^{(2,2)} \right) \left(dx + \frac{x}{y} dy \right) = 1 + 2\ln 2$$



$$\left(\int_{(1,1)}^{(1,2)} + \int_{(1,2)}^{(2,2)} \right) \left(dx + \frac{x}{y} dy \right) = \ln 2 + 1$$



Note: $df \equiv \frac{\bar{d}g}{x} = \frac{dx}{x} + \frac{dy}{y}$ is an exact differential.



Integrating factor

$$f(x, y) = \ln x + \ln y + f_0$$



Back to the First Law

Heat is path dependent.

$$\vec{d}Q = dU + PdV$$

Therefore, $1/T$ is really the **integrating factor** for the differential form of heat. Now we can recast the **1st law of thermodynamics** as

$$dU = TdS - PdV$$

Entropy is also a **state function**, as is the internal energy or volume.



Entropy of an Ideal Gas (1 mole)

$$p(T, V) = \frac{RT}{V}$$

$$U(T) = C_V^{mol} T = \frac{fR}{2} T$$

$$dS = \frac{1}{T} (dU + p dV) = \frac{C_V^{mol} dT}{T} + \frac{R dV}{V}$$

Integrating from (T_0, V_0) to (T, V)

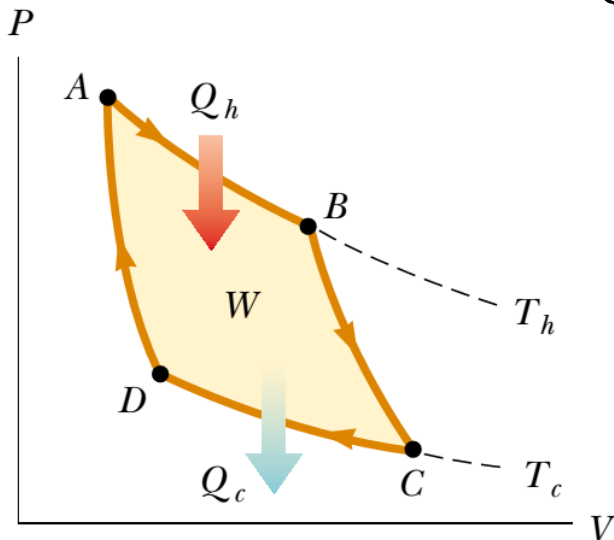
$$S(T, V) = S_0 + C_V^{mol} \ln \frac{T}{T_0} + R \ln \frac{V}{V_0}$$



Carnot's Theorem

No real heat engine operating between two energy reservoirs can be more efficient than Carnot's engine operating between the same two reservoirs.

$$e' = 1 + \frac{Q'_c}{Q'_h} \overset{\text{positive}}{<} 1 - \frac{T_c}{T_h} \Rightarrow \overset{\text{negative}}{\frac{Q'_h}{T_h} + \frac{Q'_c}{T_c}} < 0$$

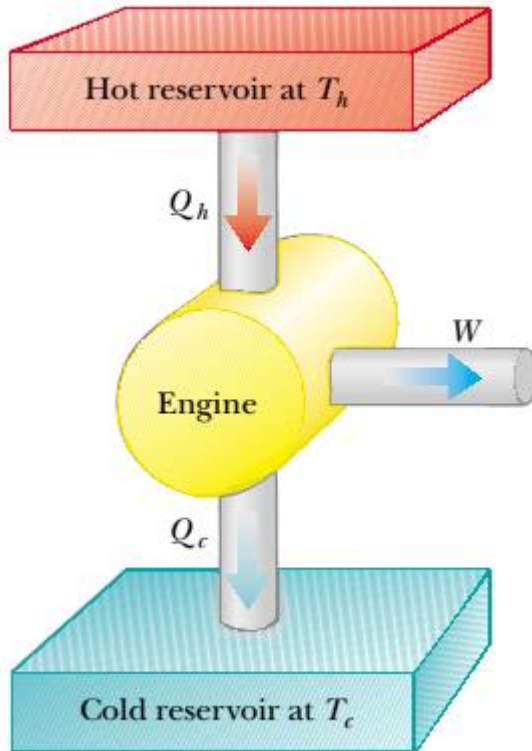


What does this mean? Still, for any engine in a cycle (**S is a state function!**)

$$\oint dS = 0$$



Counting the Heat Baths in



hot reservoir 熵的该变量

$$\Delta S_h = \frac{-Q'_h}{T_h} \quad Q'_h > 0$$

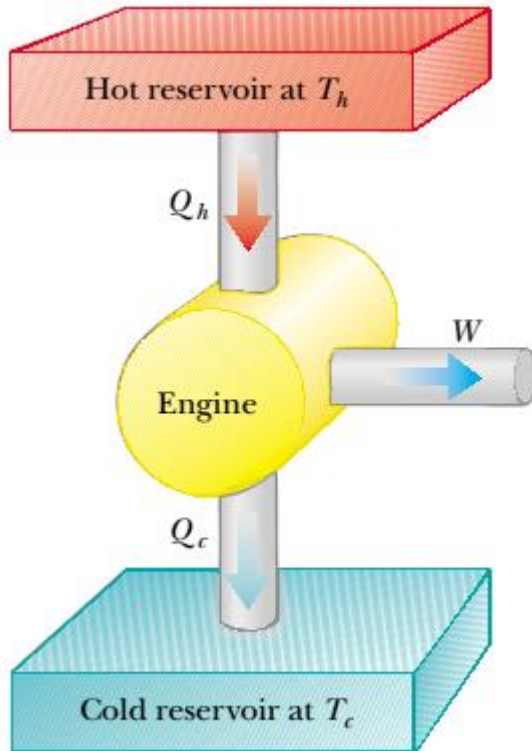
$$\Delta S_{\text{gas}} = \oint dS = 0 \quad \text{after a cycle}$$

$$\Delta S_c = \frac{-Q'_c}{T_c} \quad Q'_c < 0$$

$$\Delta S = \Delta S_h + \Delta S_{\text{gas}} + \Delta S_c = \frac{-Q'_h}{T_h} + 0 + \frac{-Q'_c}{T_c} > 0$$



Counting the Heat Baths in



$$\Delta S_h = \frac{-Q'_h}{T_h} \quad Q'_h > 0$$

$$\Delta S_{\text{gas}} = \oint dS = 0 \quad \text{after a cycle}$$

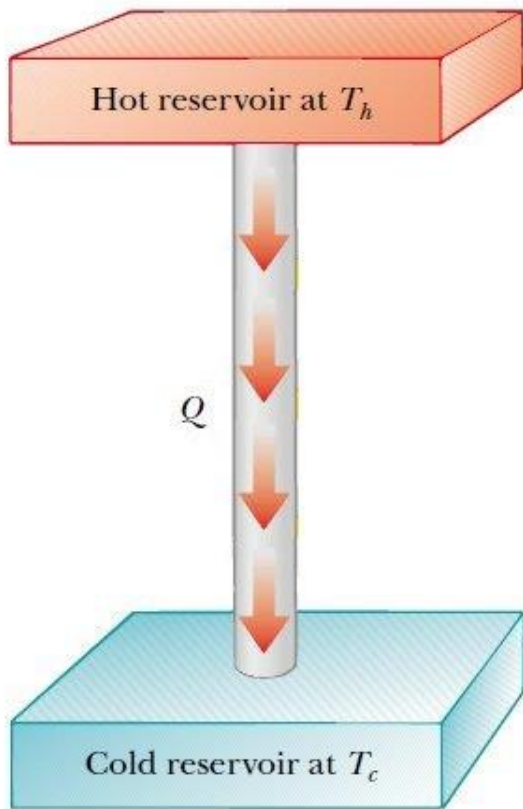
$$\Delta S_c = \frac{-Q'_c}{T_c} \quad Q'_c < 0$$

isolated system! ! !

The total entropy of an isolated system that undergoes a change can never decrease.



Example 1: Clausius Statement



$$\Delta S_h = \frac{-Q}{T_h}$$

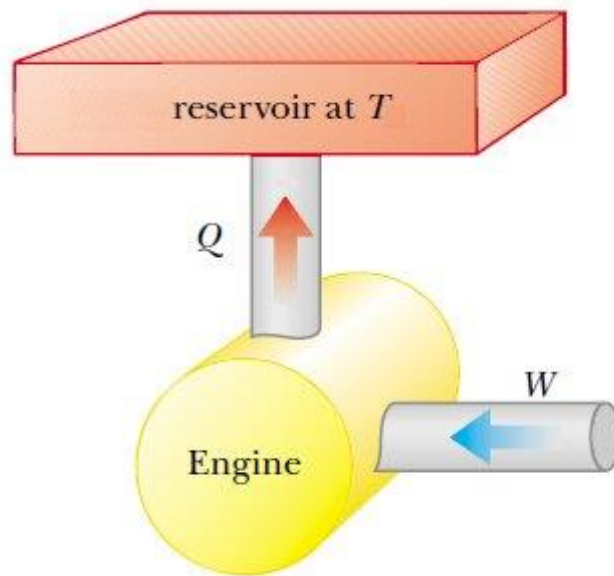
$$\Delta S_c = \frac{Q}{T_c}$$

$$\Delta S = \Delta S_h + \Delta S_c = -\frac{Q}{T_h} + \frac{Q}{T_c} > 0$$

Irreversible!



Example 2: Kelvin Statement



$$\Delta S = \frac{Q}{T} > 0$$

Irreversible!



Specific Heat

$$c \equiv \frac{Q}{m\Delta T}$$

$$Q = mc\Delta T$$

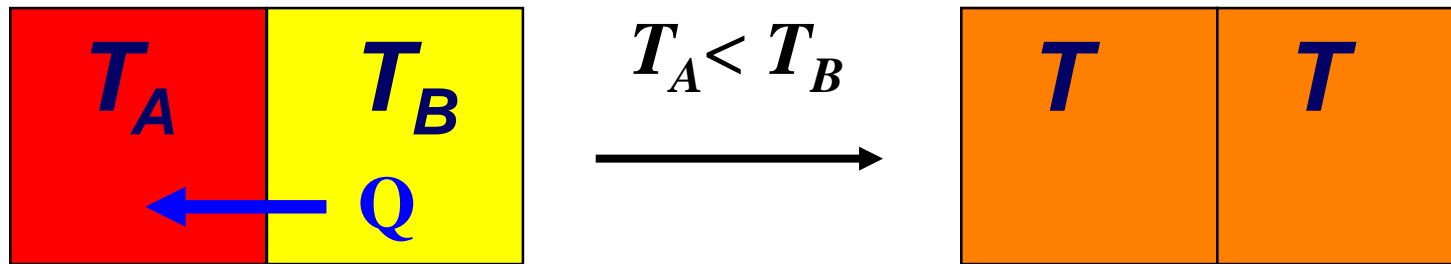
Note: Last time we defined molar specific heat. In physics, we also use specific heat per particle.

Specific Heats of Some Substances at 25°C and Atmospheric Pressure

Substance	Specific Heat c	
	J/kg · °C	cal/g · °C
Elemental Solids		
Aluminum	900	0.215
Beryllium	1 830	0.436
Cadmium	230	0.055
Copper	387	0.092 4
Germanium	322	0.077
Gold	129	0.030 8
Iron	448	0.107
Lead	128	0.030 5
Silicon	703	0.168
Silver	234	0.056
Other Solids		
Brass	380	0.092
Glass	837	0.200
Ice (− 5°C)	2 090	0.50
Marble	860	0.21
Wood	1 700	0.41
Liquids		
Alcohol (ethyl)	2 400	0.58
Mercury	140	0.033
Water (15°C)	4 186	1.00
Gas		
Steam (100°C)	2 010	0.48



Example 3: Mixing Water



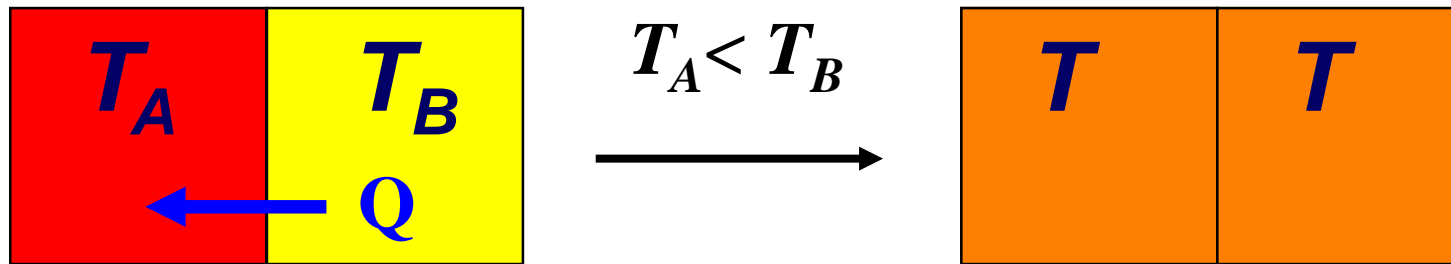
$$T_A \rightarrow T: |Q| = cm_A (T - T_A)$$

$$T_B \rightarrow T: |Q| = cm_B (T_B - T)$$

$$\frac{m_A}{m_B} = \frac{T_B - T}{T - T_A} \Rightarrow T = \frac{m_A T_A + m_B T_B}{m_A + m_B}$$



Example 3: Mixing Water



$$T_A \rightarrow T: \Delta S_A = \int \frac{cm_A dT}{T} = cm_A \ln(T/T_A) > 0$$

为什么可以计算熵改变?

$$T_B \rightarrow T: \Delta S_B = \int \frac{cm_B dT}{T} = cm_B \ln(T/T_B) < 0$$

For simplicity, assume $m_A = m_B = m, T = (T_A + T_B)/2$

$$\Delta S = \Delta S_A + \Delta S_B = cm \ln(T^2/T_A T_B) > 0$$

Irreversible!



The Second Law in terms of Entropy

The total entropy of an isolated system that undergoes a change can never decrease.

- If the process is **irreversible**, then the total entropy of an isolated system **always increases**.
- In a **reversible** process, the total entropy of an isolated system **remains constant**.

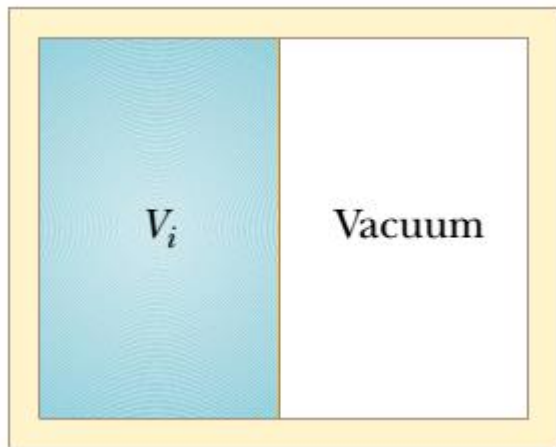
The change in entropy of the Universe must be greater than zero for an irreversible process and equal to zero for a reversible process.

$$\Delta S_{Universe} \geq 0$$

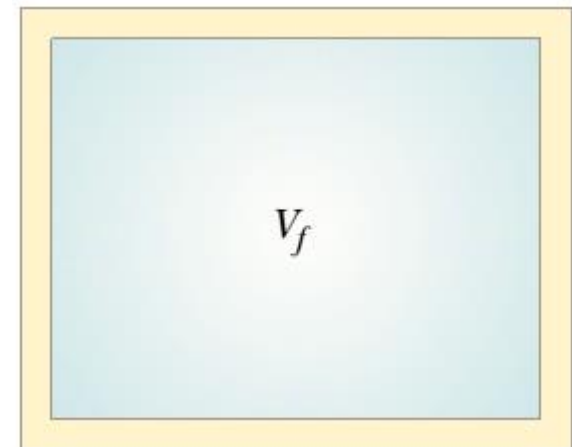
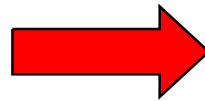


Order versus Disorder

Isolated systems tend toward disorder and that entropy is a measure of this disorder.



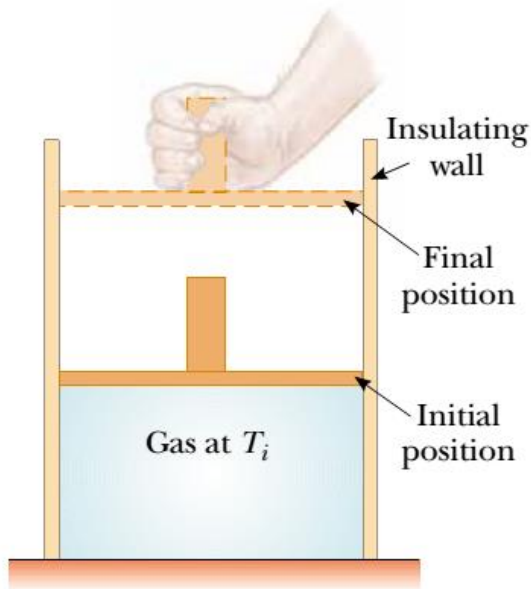
**Ordered: all molecules
on the left side**



**Disordered: molecules
on the left and right**



Example 4: Free Expansion



Energy reservoir at T_i

(a)

$$\Delta U = Q = W = 0 \Rightarrow \Delta S = 0$$

We can only calculate ΔS with a reversible process! In this case, we replace the free expansion by the **isothermal process** with the same initial and final states.

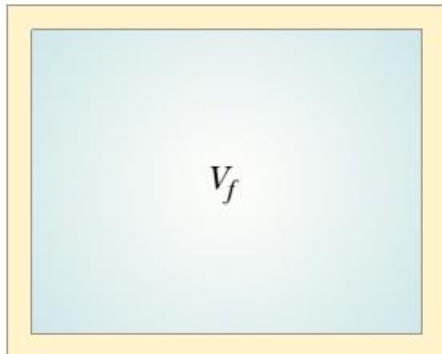
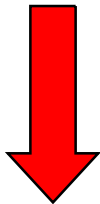
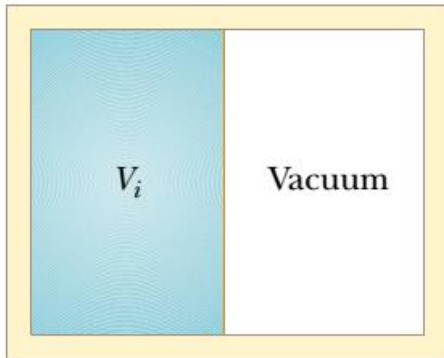
这样计算，因为不是孤立系统了，所以可逆过程【准静态？】，但是熵增加即用可逆过程才能计算熵的改变，找到与初末状态相同的可逆过程

$$\Delta S = \int_{v_i}^{v_f} \frac{dQ}{T} = \int_{v_i}^{v_f} \frac{PdV}{T} = \int_{v_i}^{v_f} \frac{nRdV}{V} = nR \ln(V_f/V_i) > 0$$

Irreversible!



Entropy: A Measure of Disorder



We assume that each molecule occupies some microscopic volume V_m .

$$W_i = (V_i/V_m)^N$$

$$W_f = (V_f/V_m)^N$$

W:可能出现的位置个数

$$\Rightarrow W_f/W_i = (V_f/V_i)^N$$

$$\Delta S = Nk_B \ln(V_f/V_i) = Nk_B \ln 2$$

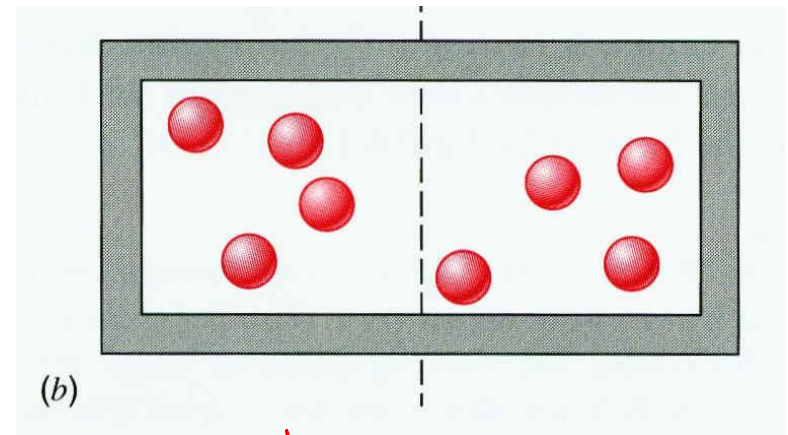
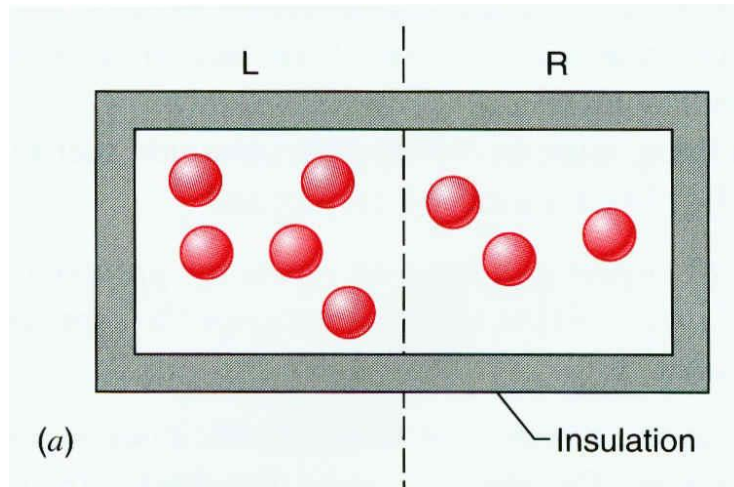
suggesting $S = k_B \ln W$

(Boltzmann)





A statistical view of entropy



$$N = 200$$

The multiplicity of configuration is

$$w = \frac{N!}{N_1! N_2!}$$

$$\ln N! \approx N \ln N - N$$

$$1) N_1 = 100; N_2 = 100$$

$$w_1 = \frac{200!}{100! 100!} = 1.6 \times 10^{60}$$

$$2) N_1 = 150; N_2 = 50$$

$$w_2 = 6.97 \times 10^{48}$$

$$\therefore w_1 / w_2 = 2.3 \times 10^{11}$$



Information, An Example

Consider the following three statements.

- 1. Newton's birthday falls on a particular day of the year. ($P = 1$)**
- 2. Newton's birthday falls in the second half of the year. ($P = 1/2$)**
- 3. Newton's birthday falls on the 25th of a month. ($P = 12/365$)**

As P decreases, the information content increases.

Combining 2 and 3, $P = 1/2 * 12/365 = 6/365$.

It is natural to assume information content is additive.

The information content Q of a statement (with probability P) can then be defined by

$$Q = -\log P$$



Information and Entropy

(1927) Bell Labs, Ralph Hartley

- Measure for information in a message
- Logarithm: 8 bit = $2^8 = 256$ different numbers

(1948) Bell Labs, Claude Shannon

- “A mathematical theory of communication”
- Probability of a particular message
- Average information ~ entropy?

$$S_{\text{information}} = - \sum_i P_i \log P_i$$

Shannon entropy



Shannon Entropy

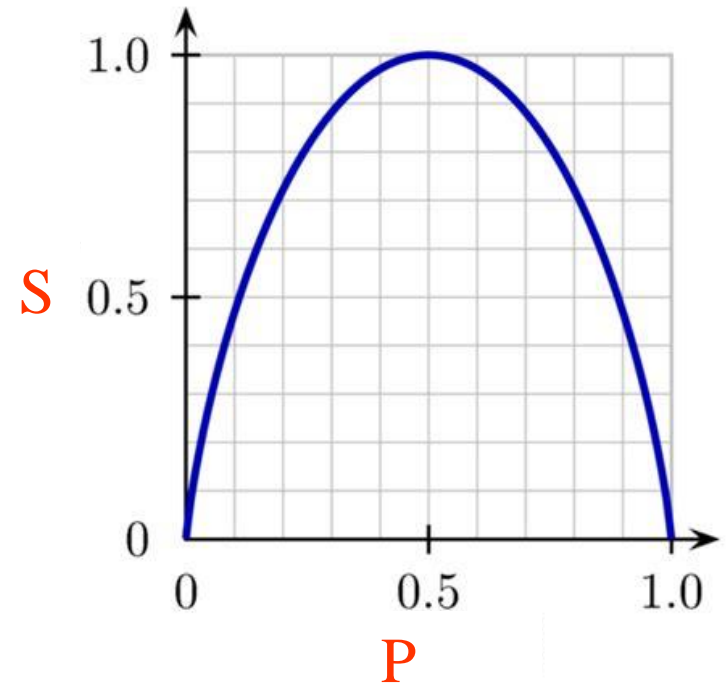
“It is already in use under that name. ... and besides, it will give you great edge in debates because nobody really knows what entropy is anyway.”

– John von Neumann



Entropy of A Bernoulli Trial

Consider the simplest case:
Bernoulli trial –
a two-outcome random
variable with probability
 p and $1-p$ of the two
outcomes.



$$S = - \sum_i P_i \log P_i = -p \log p - (1 - p) \log(1 - p)$$



The Entropy of Mixing

Consider two **different** ideal gases (call them 1 and 2) which are in separate vessels with volume xV and $(1-x)V$ respectively at the same pressure p and temperature T . If the pipe that connects the two vessels is opened, the gas will spontaneously mix, resulting an increase in entropy.

如果是两种同样的gas，则无熵变化

$$\begin{aligned}\Delta S &= xNk_B \int \frac{dV_1}{V_1} + (1-x)Nk_B \int \frac{dV_2}{V_2} \\ &= -Nk_B [x \ln x + (1-x) \ln(1-x)]\end{aligned}$$

Distinguishability is an important piece of information!



Entropy, from Information Viewpoint

Entropy is a measure of our **uncertainty** of a system, based on our **limited knowledge** of its properties and ignorance about which of its microstates it is in.

In making inferences on the basis of partial information, we can assign probabilities on the basis that we **maximize entropy subject to the constraints** provided by what is known about the system.

Exercise: **Maximize** entropy $S = -k_B \sum_i P_i \ln P_i$ subject to $1 = \sum_i P_i$ and $U = \sum_i E_i P_i$. **What do you expect?**



Solution

Using the method of Lagrange multipliers, in which we maximize

$$S/k_B - \alpha \left(\sum_i P_i - 1 \right) - \beta \left(\sum_i P_i E_i - U \right)$$

where α and β are Lagrange multipliers. We vary this expression with respect to one of the probability P_j and get

$$\frac{\partial}{\partial P_j} \left(\sum_i -P_i \ln P_i - \alpha P_i - \beta P_i E_i \right) = 0$$

so that

$$-\ln P_j - 1 - \alpha - \beta E_j = 0$$

This can be rearranged to give

$$P_j = \frac{e^{-\beta E_j}}{e^{1+\alpha}}$$

Boltzmann probability



Selected Topics

Maxwell's demon

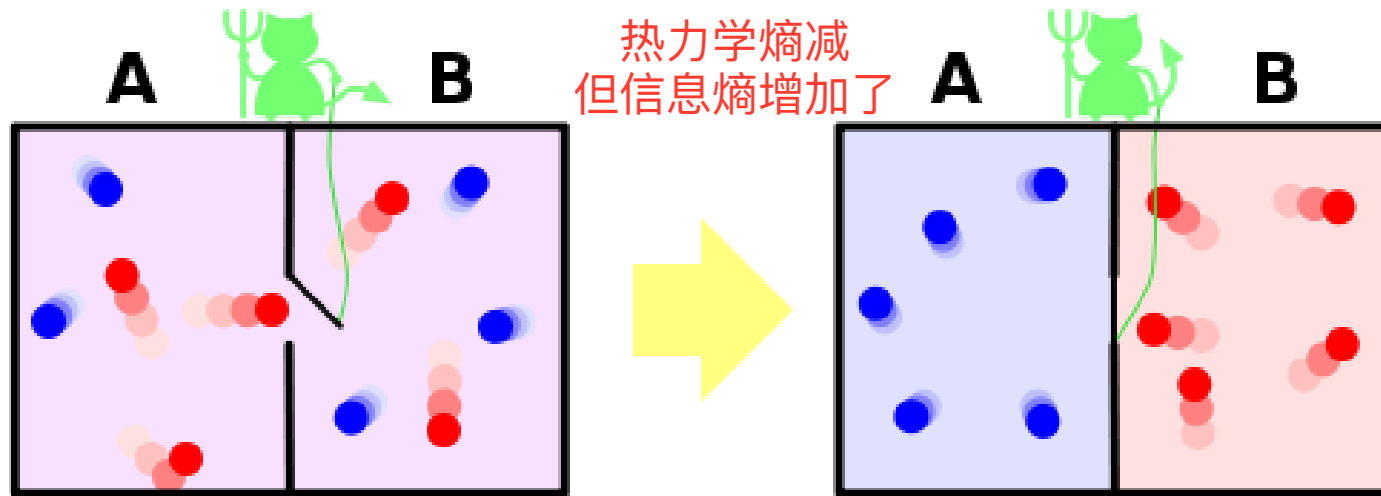
Data compression

Relative entropy and cross entropy

Quantum information



I. Maxwell's Demon



To determine whether to let a molecule through, the demon must acquire information about the state of the molecule. However well prepared, the demon will eventually run out of information storage space and must begin to erase the information it has previously gathered. Erasing information is a thermodynamically irreversible process that increases the entropy of a system.



Information is Physical

Consider a device which has stored 1 bit of information and is connected to a thermal reservoir of temperature T . The bit can be either 1 or 0. We can erase it by setting the bit to zero.

- Before erasure, $p_1 = 1/2$, $p_2 = 1/2$. $S = \log 2$.
- After erasure, $p_1 = 1$, $p_2 = 0$. $S = 0$.
- Erasure is irreversible. Entropy goes down by $k_B \ln 2$.

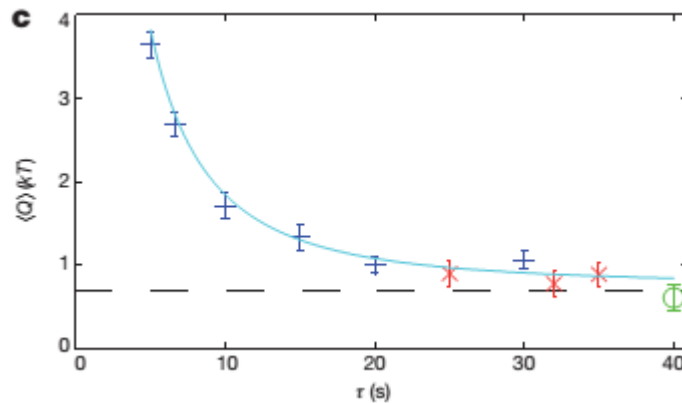
For the total entropy of the universe not to decrease, the entropy of the surroundings must go up by at least $k_B \ln 2$ and we must dissipate heat in the surroundings.



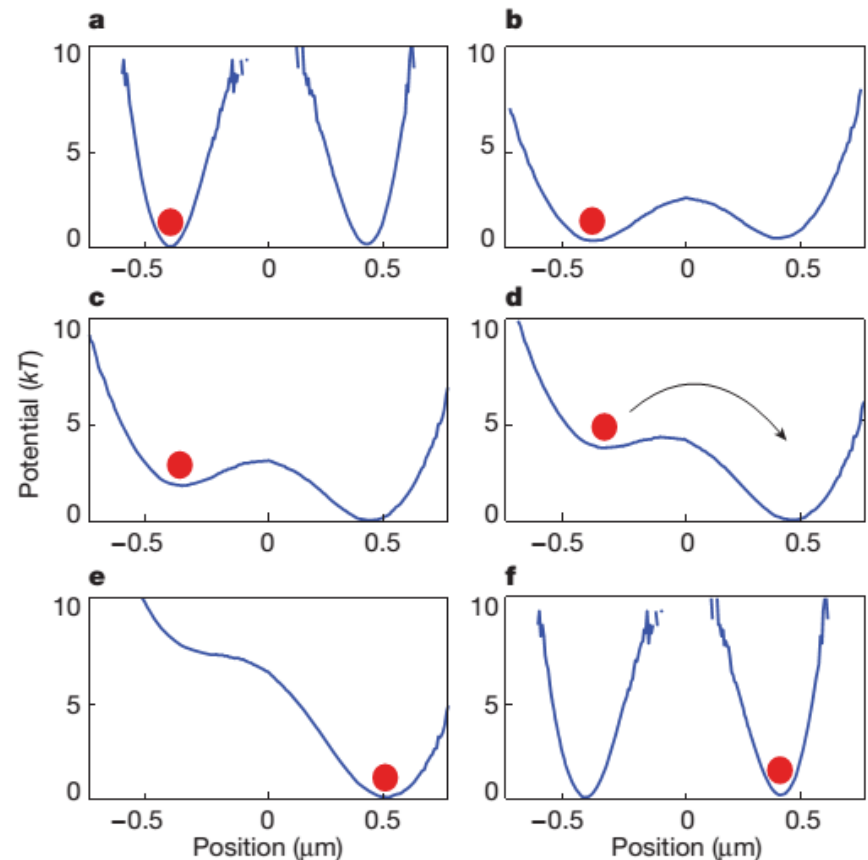
Landauer's Principle & Verification

Computation needs to involve heat dissipation only when you do something irreversible with the information.

Lutz group (2012)



$$\frac{\langle Q \rangle}{k_B T} = \ln 2 \approx 0.693$$





For Those Who Are Interested

Reading (downloadable from website):

- Charles Bennett and Rolf Landauer, *The fundamental physical limits of computation*.
- Antoine Bérut *et al.*, *Experimental verification of Landauer's principle linking information and thermodynamics*, Nature (2012).
- Seth Lloyd, *Ultimate physical limits to computation*, Nature (2000).

<http://zimp.zju.edu.cn/~xinwan/courses/physl18/references.php>

Dare to adventure where you have not been!



II. Data Compression

Information must be stored/transmitted. Compressible?

Example: Classical information is stored in the form of binary digits 0 (with probability p) and 1 (with probability $1-p$). Suppose $p = 0.9$, the data contain less information than in the case of $p = 0.5$.

Two-bit encoding:

00 \rightarrow 0

10 \rightarrow 10

01 \rightarrow 110

11 \rightarrow 1110

$$p^2 = 0.81$$

$$p(1-p) = 0.09$$

$$p(1-p) = 0.09$$

$$(1-p)^2 = 0.01$$

← Code efficiently the typical sequence!

$$\text{Average length: } 0.81 + 2*0.09 + 3*0.09 + 4*0.01 = 1.3$$



Generalization to n-bit Encoding

The probability of finding a sequence x_1, \dots, x_n is

$$P(x_1, x_2, \dots, x_n) = P(x_1)P(x_2)\dots P(x_n) \approx p^{np}(1-p)^{n(1-p)}$$

The information content is

$$-\log P(x_1, x_2, \dots, x_n) = -np \log p - n(1-p) \log(1-p) = nS$$

where S is the entropy for a Bernoulli trial with probability p . Hence

$$P(x_1, x_2, \dots, x_n) = \frac{1}{2^{nS}}$$

This shows that there are at most 2^{nS} typical sequences and hence it only requires nS bits to code them.



Noiseless Channel Coding Theorem

A compression algorithm will take a typical sequence of n terms x_1, \dots, x_n and turn them into a string of length nR . Hence the smaller R is, the greater the compression.

If we have a source of information with entropy S , and if $R > S$, then there exists a reliable compression scheme of compression factor R . Conversely, if $R < S$, then any compression scheme will not be reliable.

The entropy S sets the ultimate compression limit on a set of data.



III. Relative Entropy

The **Kullback–Leibler divergence** (also called **relative entropy**) is a measure of how one probability distribution diverges from a second, expected probability distribution.

$$D_{KL}(P|Q) = - \sum P_i \log \frac{Q_i}{P_i} = \underbrace{- \sum P_i \log Q_i}_{\text{cross entropy}} - \underbrace{\left(- \sum P_i \log P_i \right)}_{\text{(Shannon) entropy}}$$

- The KL divergence is always non-negative (Gibbs' inequality), with $D_{KL} = 0$ if and only if $P = Q$ almost everywhere.
- KL divergence and cross entropy can be used to define the loss function in machine learning and optimization. The true probability P is the true label, and the given distribution Q is the predicted value of the current model.



IV. Quantum Information

Quantum information is composed of quantum bits (known as qubits), which are two-level quantum systems that can be represented by linear combinations of the states $|0\rangle$ and $|1\rangle$.

Quantum entanglement: no classical counterpart

- Bell state $(|01\rangle + |10\rangle)/\sqrt{2}$
- **Entanglement entropy**
- Information transmission faster than the speed of light? No!

Quantum no-cloning theorem: It is impossible to make a copy a non-orthogonal quantum mechanical states.