The Carnot Cycle: from Classical Thermo to Stat Thermo

Addison Ault, Department of Chemistry, Cornell College, Mt. Vernon IA

The Carnot cycle is usually described in terms of classical thermodynamics, and the requirement that heat be dumped to the "cold" reservoir is presented as the requirement that the cyclic process not involve a decrease in *entropy*. In this supplement I review the classical thermodynamic explanation and then present a parallel explanation from statistical mechanics in which the requirement that heat be dumped to the "cold" reservoir is presented as the requirement that the cyclic process not involve a decrease in *multiplicity*.

The efficiency of a heat engine

The merit of any engine is the degree to which the energy put into the engine can be converted to work. In the case of a heat engine, this merit, or *efficiency*, is expressed as the degree to which the thermal energy, or *heat*, absorbed by the engine can be converted to mechanical energy, or *work*. It is a fact that heat is unique as the one form of energy that cannot, even in principle, be *completely* converted to work. This fact is often called the Second Law of Thermodynamics:

It is impossible to convert heat (thermal energy) <u>completely</u> into work in a <u>cyclic process</u>. *William Thompson, Lord Kelvin*

A more general form of the Second Law is attributed to Clausius:

The entropy of the world tends toward a maximum.

Rudolph Clausius

The heat engine illustrates these equivalent versions of the second law.

What is it that determines the efficiency of a heat engine?

The French engineer Sadi Carnot (1796-1832) has been given the credit for being the first to answer this question. Mendoza describes the situation in this way. (Reference 1).

"The problem occupying Carnot was how to design good steam engines. Steam power already had many uses - draining water from mines, excavating ports, forging iron, grinding grain, and spinning and weaving cloth - but it was inefficient. The import into France of advanced engines after the war with Britain showed Carnot how far French design had fallen behind. It irked him particularly that the British had progressed so far through the genius of a few engineers who lacked formal scientific education. British engineers had also accumulated and published reliable data about the efficiency of many types of engines under actual running conditions; and they vigorously argued the merits of low- and high-pressure engines and of single-cylinder and multi-cylinder engines."

Carnot imagined a heat engine of a particular design, and he then calculated its efficiency. While real engines run fast and under sub-optimal conditions, and real engines lose mechanical energy to friction, Carnot's analysis of an ideal heat engine indicates that there is an upper limit to the efficiency of any real engine.

Carnot's insight, as more fully interpreted by Clausius, showed that the efficiency of a heat engine is completely determined by a *pair of temperatures*. One temperature is that of the hot source of the heat, and the other temperature is that of the cold sink into which it is required that *some* heat be dumped. The *requirement* for dumping makes it *impossible* for a heat engine, which must operate in a cycle, to convert to work all of the heat taken in at the start of the cycle.

The Big Question

The big question, of course, is Why must *any* thermal energy be dumped to the sink. One answer, as we will see, is To repay the entropic cost of the heat taken in at the start of the cycle. The other answer, as we will also see, is To cancel the increase in multiplicity that accompanies the transfer of thermal energy at the start of the cycle.

Our example

So that we can do the math in our heads our example will consist of a hot source at 600 K, and a cold sink at 300 K.

The upper limit to the efficiency of any heat engine that operates between these temperatures equals the *difference* between the high temperature, T_{high}, and the low temperature, T_{low}, *divided by* T_{high}.

$$\text{maximum possible efficiency} = \frac{T_{high} - T_{low}}{T_{high}}$$

For these temperatures the maximum possible efficiency of a heat engine will be 50%.

$$\text{maximum possible efficiency} = \frac{600 - 300}{600} = 1/2 = 50\%$$

The efficiency of any *real* heat engine that operates between these two temperatures will be less than 50%.

Since our example assumes a heat input of 6,000 energy units, the maximum amount of work that an ideal heat engine can produce from this input will be 3,000 energy units. The reason 3,000 units of energy is the upper limit for work is that 3,000 units of heat *must* be dumped to the low temperature sink. We will see that the reason half the heat must be dumped is that the temperature of the sink is half that of the source.

If the temperature of the sink were 400 K, then 4,000 of the 6,000 units of heat would have to be dumped to the sink.

If the sink is at the *same* temperature as the source, *no* work can be produced. On the other hand, if the sink were at absolute zero *all* of the 6,000 units of heat could be converted to work.

Carnot's heat engine

Carnot's engine consists of a piston in a cylinder that confines the steam. As the steam pushes the piston out, the piston does work on something. As the piston is pushed in, work is done on the steam. A complete "cycle" consists of an outward "work" stroke, and followed by an inward "recovery" stroke. Each stroke has two parts, and so there are two pairs of parts, or four parts in all, to a "Carnot cycle."

The value of this particular model of a heat engine is that, first, one can calculate its efficiency and, second, no other engine can be more efficient than the Carnot engine. The Carnot cycle sets an upper limit to expectations for the efficiency of any conceivable heat engine, regardless of its mechanism or the identity of the "working fluid", superheated water vapor in the case of the "steam" engine. The upper limit of efficiency depends entirely on the temperature of the heat source: coal-fired boiler, gas turbine, or nuclear reactor, and the temperature of the heat sink: the air, or the water of a river or the water of a lake.

Two answers for the Big Question

The big question is Why must *any* thermal energy be dumped to the sink.

The answer according to classical thermodynamics has to do with *entropy*, and the observation, according to the second law, that for anything to actually happen the corresponding change in *entropy* must be positive.

This was the original answer, and for many years the only answer. This answer does not make use of any assumptions about the nature of material substances; it makes no mention of atoms or molecules. This answer is just as true now as it has always been, but no more insightful.

There is now, however, another type of answer, the answer according to statistical thermodynamics. This new answer has to do with *multiplicity* and the realization that, according statistical mechanics, for anything to actually happen the new state has be more probable, and the corresponding change in *multiplicity* must be positive.

This new answer is based on the assumption that both matter and energy are granular in nature; that atoms and quanta exist. Multiplicity is the number of possible microstates, or combinations of atoms and quanta, and a macrostate with more microstates is (overwhelmingly) more likely.

We will look at the Carnot cycle from both these points of view.

The Carnot cycle according to classical thermodynamics

The work stroke of the Carnot engine has two parts.

In the *first* part of the work stroke the steam expands and delivers its push; the steam does *work*. In order that the temperature of the expanding steam not fall, *heat* is delivered to the steam in such a way that the temperature of the expanding steam is maintained at its initial, "high", temperature of 600 K.

This first part of the work stroke is an *isothermal* process. Although work, w, is exported, the simultaneous import of an energetically equivalent amount of heat, q, keeps the temperature, T, and the internal energy, E, constant.

Since heat is imported there must be an increase in entropy. If we say that 6,000 units of energy are imported at a constant temperature of 600 K the increase in entropy, ΔS, will be $q/T = 6{,}000/600 = 10$ entropy units.

In the *second* part of the work stroke the steam continues to expand and to deliver more of its push, but the temperature, and thus the pressure, of the steam is allowed to fall. When the temperature of the steam has fallen to the "low" temperature, the work stroke is done.

This second part of the work stroke is an *adiabatic* process. Work is done, but there is no compensating uptake of heat, q, and therefore no change in entropy. The internal energy will thus decrease, and the temperature will fall. When the temperature of the steam has fallen to the "low" temperature of 300 K, the temperature of the sink, the second part of the work stroke is done, and the recovery stroke is initiated.

The recovery stroke

The recovery stroke is not merely the reverse of the work stroke. If the recovery stroke were the reverse of the work stroke the recovery stroke would undo the effect of the work stroke. Nothing would have been accomplished.

The first part of the recovery stroke is, instead, compression of the steam at the temperature of the heat sink, 300 K. This compression is done as an *isothermal* process, and the work energy transmitted to the steam is passed on as thermal energy to the heat sink. There will therefore be no change in internal energy, E. However, since heat is exported, there must be a decrease in entropy. If 3,000 energy units of heat, q, are exported at 300 K there will be a decrease in entropy ΔS will be $q/T = 3{,}000/300 = 10$ entropy units. This is how the "entropy debt" incurred in the first half of the work stroke is repaid.

The second part of the recovery stroke is further compression of the steam to its initial pressure. This recompression is *adiabatic:* no heat is removed during this compression and so the temperature of the steam will rise to the original, "high" temperature. The steam is now ready for a second cycle.

Net work

The work produced in the second half of the work stroke, the *adiabatic expansion*, is exactly equal to the work consumed in the second step of the recovery stroke, the *adiabatic compression*. There is no net production of work from this pair of adiabatic steps.

In contrast, the recompression at 300 K, consumes only half as much work as is produced in the original expansion at 600 K because the pressure at 300 K is only half the pressure at 600 K. This pair of *isothermal* steps provides, in this example, a *net production* of work equivalent to half of the thermal energy imported in the first step of the work stroke.

The secret of a successful heat engine

The ideal cyclic process must be "entropy-neutral" because to return to the initial state, which is what "cyclic" means, is to return the system, or engine, to its initial entropy as well as to its initial energy, temperature, pressure, and volume.

The "secret of success" for the ideal heat engine of this example is to pay back the entropy "cost" of 6,000 units of heat (6,000/600 = 10 entropy units) with only 3,000 units of heat. This is done by making the 3,000 heat unit payment at a temperature of only 300 K, a temperature at which it takes only 3,000 heat units to be worth 10 entropy units. (3,000/300 = 10 entropy units). We could call this the "buy high, sell low" strategy.

More secrets of a successful heat engine

The heat engine of the example is an ideal heat engine. It will have no losses to friction, and its parts will move very slowly. It is only under these ideal conditions that the efficiency of a real heat engine will approach the ideal, or maximum possible efficiency.

Losses to friction

Friction is the conversion of mechanical energy to heat energy, and we know that there is no way to completely convert heat back to work.

Irreversible conditions

For maximum efficiency change must take place very slowly. Using the language of classical thermodynamics a maximally efficient change must take place under quasi-static or "reversible" conditions. The lack of balance between forces that must exist for change to occur must be as small as possible. Since real engines run fast, this condition is not met, and the real efficiency will be correspondingly less.

To anticipate the statistical mechanical view of the Carnot cycle, the particles of the working fluids of real engines are not always distributed according to the Boltzmann distribution.

The Carnot cycle according to statistical mechanics.

The work stroke of the Carnot engine has two parts.

In the *first* part of the work stroke the steam expands and delivers its push; the steam does work. As the steam expands, as the volume of the steam increases and the translational energy levels of the water molecules become more closely spaced. Translational energy levels that were not originally accessible are now accessible. In order that the temperature of the expanding steam not fall, heat must be delivered to the steam so that the newly accessible energy levels can be populated according to the Boltzmann distribution that corresponds to a temperature of 600 K. It is in this way that the temperature of the system is maintained at its initial, "high", temperature of 600 K.

This first part of the work stroke is an *isothermal* process. Although work, w, is exported, and additional translational energy levels become accessible, the simultaneous import of an amount of heat, q, equivalent to the work done, w, keeps the internal energy, E, and the temperature, T, constant.

The fact that the newly accessible energy levels are populated as they become accessible means that the partition function, Q, will *increase.*

The *second* part of the work stroke is an *adiabatic* process. In this second part of the work stroke the steam continues to expand and to deliver more of its push, but this time the temperature, and thus the pressure, is allowed to fall. In this second part of the work stroke the translational energy levels become still more closely spaced and therefore still more energy levels become accessible, but this time there is no simultaneous import of heat to populate these new levels and thereby maintain the temperature and the internal energy. During this second part of the work stroke the populations of all levels remain the same even though the spacings of the levels are diminished. The distribution over levels will remain a Boltzmann distribution, but the distribution will correspond to a lower temperature. When the temperature of the steam has fallen to the “low” temperature of 300 K, the temperature of the sink, the second part of the work stroke is done, and the recovery stroke is initiated.

Again, in this second part of the work stroke the fact that more energy levels become accessible as the volume increases means that the partition function, Q, will *increase.*

The recovery stroke

The recovery stroke is not merely the reverse of the work stroke. If the recovery stroke were the reverse of the work stroke the recovery stroke would undo the effect of the work stroke. Nothing would have been accomplished.

The first part of the recovery stroke is, instead, an *isothermal* recompression of the steam at the temperature of the heat sink, 300 K. As the recompression takes place the volume of the steam will decrease, and the translational energy levels of the system will move up and apart.

At the same time, however, heat will be transferred from the system to the sink at 300 K, and in this way both the energy and the temperature of the system will remain constant. However, during this first part of the *recovery* stroke, the fact that translational energy levels become *less* accessible as the volume *decreases* means that during this isothermal *compression* the partition function, Q, will *decrease.*

As we shall see this *decrease* in Q for the first part of the recovery stroke is, for this example, exactly the same size, but of opposite sign, as the *increase* in Q for the first part of the work stroke *even though* q for the first part of the work stroke, 6,000 energy units, is twice the size of q for this first part of the recovery stroke, 3,000 energy units.

The second part of the recovery stroke is further compression to the initial volume and pressure of the steam. This time, however, no heat is removed during recompression; this recompression is *adiabatic*. The energy of the system, E, will therefore increase and the temperature of the system, T, will rise to the temperature of the heat source, 600 K.

As the volume of the steam gets smaller during recompression the translational energy levels will increase in both energy and in spacing, and will therefore become less accessible. Thus during this adiabatic compression the partition function for the system, Q, will *decrease*. This *decrease* in Q will be exactly equal in magnitude, but opposite in sign, to the *increase* in Q that accompanied the adiabatic expansion of the second half of the power stroke.

Net work

It's the same system and so the maximum efficiency and net work will be the same.

The secret of a successful heat engine

Again, it's the same system and so the secret must be contained in the relationship between the two isothermal steps.

According to classical thermodynamics $\Delta S = q/T$. You can get the same entropy change, ΔS, for a different q and a different T so long as the ratio of q to T is the same. The Carnot cycle takes advantage of this fact by accepting heat at a high temperature and dumping heat at a low temperature.

According to statistical mechanics a unit of thermal energy, represented here as E/RT, a dimensionless ratio, corresponds to the temperature of the system, T, times the rate of change of the log of the partition function with temperature at that temperature.

$$\frac{E}{RT} = T \frac{\partial \ln Q}{\partial T}$$

Thus you can get the same change in Q with different amounts of thermal energy, depending upon the temperature of the system.

At a *higher* temperature, larger T and, by implication, larger Q, the *fractional* rate of change of Q with T, $\partial \ln Q / \partial T$, will be *smaller* for the transfer of a given amount of heat.

At a *lower* temperature and, by implication, smaller Q the fractional rate of change of Q with T will be *larger* for the transfer of a given amount of heat.

In our example 3,000 units of energy dumped at 300 K results in an absolute change in Q of exactly the same size, but opposite sign, as 6,000 units of energy accepted at 600 K.

The Boltzmann distribution

The typical calculations of statistical thermodynamics are for systems at equilibrium. At equilibrium the distribution of particles over energy levels will be a Boltzmann distribution, and the corresponding multiplicities will be the largest possible multiplicities. Thus the estimates of maximum possible efficiency include the assumption that the Boltzmann distributions will be maintained at all times.

This is equivalent to the assumption of "reversible" or "quasi-static" conditions: Boltzmann or close-enough-to-Boltzmann distributions. The faster an engine is run the farther the distributions will be from Boltzmann, or equilibrium, distributions, and the greater the degree by which the efficiency will fall short of the maximum possible efficiency.

Summary

In these parallel stories the "secret of success" is found in the isothermal steps. Transfer of a given amount of heat at a higher temperature corresponds to a smaller change in both entropy and multiplicity.

References

1. E. Mendoza, "The problem occupying Carnot", www; many places.