On the Foundations of Classical Thermodynamics, and the Tolman-Ehrenfest Effect

William Rupush

Bachelor's Thesis in Physics

Department of Physics
Stockholm University
Stockholm, Sweden 2013

Department of Physics

Stockholm University

Contents

A theory is the more impressive the greater the simplicity of its premises, the more different kinds of things it relates, and the more extended its area of applicability. Therefore the deep impression that classical thermodynamics made upon me. It is the only physical theory of universal content which I am convinced will never be overthrown, within the framework of applicability of its basic concepts.

— Albert Einstein

Abstract

Classical thermodynamics is commonly formulated as a theory of cyclic processes and heat engines, particularly in how the law of the increase of entropy is derived from the Kelvin-Planck statement which asserts the impossibility of perfectly efficient conversion of heat into mechanical work. The reason for this is that much of the development of classical thermodynamics was closely tied with the research and development of heat engines during the 18:nth century, and the Second Law itself arose out of efforts to maximize their efficiency. Much of the available empirical data on thermal phenomena concerned the operations and performance of such engines, and consequently the theoretical framework that was built to accomodate that data became formulated in those terms as well. In this thesis I will provide a summary of a possible foundations for classical thermodynamics by reviewing three empirical principles, namely statements of the laws of thermodynamics (excluding the Third Law). There are a number of different possible formulations of the laws approaching the subject from various perspectives. I will take an approach that emphasizes the role of heat engines. At the end I will also discuss the Tolman-Ehrenfest effect, the phenomenon of non-uniform temperature distributions at equilibrium for systems in curved space-times. It is no wonder that the laws of thermodynamics are among the most venerated of all scientific laws. It is awe-inspiring how simple they are, and how yet in spite of their simplicity they together serve to erect the vast structure of projectible generalizations of thermal phenomena that are so successfully and diversely employed.

1 Useful Terminology

In any domain of physics one is careful to separate the object of interest from everything else, and thermodynamics is no exception. This will be done by using two kinds of *walls*. If a wall is completely impenetrable by flows of matter and energy we call it an *adiabatic wall*, and if not, we call it a *diathermal wall*. Any region of space (empty of not) enclosed by walls (real or not) is referred to as a *thermal system*, and the rest of the universe is referred to as the *surroundings*. If a thermal system is completely enclosed by adiabatic walls it is said to be *isolated*, if not, it is said to be in *thermal contact* with its surroundings (or other systems). The state S of the system is an ordered n-tuple of the coordinates $X_1, X_2, , X_n$ which will be called *state parameters*, and the set of all states is an n-dimensional manifold called *state space*.

2 Introduction

Thermodynamics, in a broad sense, is the study of the properties of matter insofar as they are sensitive to changes in temperature, and of the relationships between thermal and mechanical energy transformations. There are two main approaches one might take to this study, one is *statistical*, and the other *phenomenological*. In the statistical approach one starts from a fundamental postulate that all accessible microstates of the system at equilibrium are obtained with equal probabilty, and postulates a relation between the multiplicity Ω of microstates to the thermodynamic entropy S through Boltzmann's equation $S = k \log \Omega$, where k is Boltzmann's constant. All microscopic details in the state description are coarse-grained over by statistical averaging, in a way exemplified by the central limit theorem, and the quantities of thermodynamics are then explained as emerging from the statistical action of matter in aggregation [5] (p.455). The phenomenological approach however, which is the one I'll be following, is (generally considered) not explanatory in the sense that it does not provide much insight into the underlying mechanisms which gives rise to the observed phenomena, it only deals with the phenomena itself. A phenomenological theory can therefore be said to be "merely" a codification of experimental results [13]. Classical thermodynamics is a prime example of a phenomenological theory, and provides an illumination of the road starting from meager stimulations on our sensory surfaces, and culminating in a scientific theory.

It is taken as a methodological starting point in classical thermodynamics to not make use of any hypotheses regarding the microscopic constitution of matter [8] (p.13). Instead it is based on empirical principles, statements generalizing known facts of experience. A great benefit of this approach is generality. By defining the central theoretical terms (such as temperature, energy and entropy) in a phenomenological way, one avoids commitment of the theory to any specific model of microscopic matter, and is able to accomodate a wider range of phenomena into the theory than otherwise. In this thesis, three such principles will be presented, and each will be shown to imply the existence of a function of state. Those functions of state will then serve to completely characterize the thermodynamic system.

The main bulk of this presentation, namely the review of classical thermodynamics, is based to a large extent on the works of A. B. Pippard [1] and C. J. Adkins [2]. I refer the reader interested in a more mathematically rigorous presentation of classical thermodynamics based on heat engines to the work of D.R. Owen [14]. An alternative and more axiomatic approach which began in 1909 with the work "Investigations on the Foundations of Thermodynamics" by C. Carathéodory [15] which emphasizes the role of the entropy function as a codification of possible state changes. This approach is based on the notion of *adiabatic accessibility* , a relation which concerns the fact that for a system under thermal isolation there are some states which can never be reached, namely the states of lower entropy. This relation can be used to establish the notion of entropy without recourse to either heat engines or statistical mechanics. This approach was pioneered by H. A. Buchdahl [2] and M. Born [16] (p.1)

among others, and in 1999 E. H. Lieb and J. Yngvason [11] provided perhaps the most promising axiomatization of classical thermodynamics in these terms. The foundations of classical thermodynamics is today a subject with no shortage of impressive textbooks, and this thesis but scratches the surface of the available litterature.

A question that might arise in some people is why one should bother with axiomatic foundations at all. Intellectual curiosity is one answer, another would be that today statistical mechanics is among the few fields of physics where many practitioners are not using the same set of principles in their work, but rather the subject is divided into several "schools", each with its own research programmes and technical tools [8] (p.4). The foundations of statistical mechanics, and its relationship to classical thermodynamics, is today a hotly debated topic among physicists and philosophers of science alike. A reductionistic methodological standpoint concerning these issues has been the norm, and the reduction of classical thermodynamics to statistical mechanics has been seen as wholly uncontroversial. Recently however, the significance of that reduction has been challenged somewhat, and there has emerged more interest among philosophers of science in phenomenological theories in general, and in their descriptive and explanatory power [12] (p.2). A deeper understanding of phenomenological thermodynamics as an autonomous theory in its own right is also beneficial for practical reasons, as in many practical situations the phenomenological description provides the starting point from which to do further analysis on a thermal system, prior to (or in absence of) any statistical-mechanistic understanding of it.

3 Thermal Equilibrium, the Zeroth Law and Empirical Temperature

Our starting point is the observation that isolated thermal systems will tend towards a terminal state at which no further changes are macroscopically perceptible. For example, if a body is heated and then placed in thermal contact with its surroundings and left to itself, then heat will diverge from the hot body into the colder surroundings until the body has cooled down and the heat gradient has vanished. After uniformity of hotness has been established, no more changes will occur, and at this point the body is said to have reached *thermal equilibrium with its surroundings*. The meaning of the term "hotness" will eventually need clarification, but for now we can simply identify it with our sensory experience of heat. Likewise, if two (or more) systems with (generally) different state parameters are brought together into thermal contact, they will tend towards a state of *equilibrium with one another*. We can characterize equilibria as being states of time translational invariance.

An operative word in the previous description is "macroscopically", if we had sense organs capable of resolving shorter intervals of space and time we would percieve an equilibrium state as constantly changing (Brownian motion).

Our actual (limited) sense organs however blur out all the atomic degrees of freedom in the system leaving only a few parameters perceptible to us. These coarse grained averages of the atomic coordinates are what we will use as state parameters. An example of such a macroscopic parameter is pressure, which corresponds to an average of the momentum transfer per unit area into the walls of the system by molecular collisions. But notions concerning microscopic underpinnings, while being of explanatory utility, are not of essential importance to us here and instead the system is regarded as a "black box", in the sense that we do not know what is contained within its boundaries, and can only measure its response at the boundaries to interactions with other systems. So one should not regard the laws of thermodynamics as fundamental, but rather as auxiliary hypotheses which constrain the set of allowed dynamical processes of the more fundamental theory. While this macroscopic description might lack fundamental significance it does hold an epistemological high ground by virtue of being directly suggested to us by our senses.

While it takes on the order of 10^{24} bits of Shannon information to specify the microscopic state of a litre of air (at 1 atm, and assuming the effective granularity of phase space imposed by the uncertainty principle), everything of thermodynamic interest to us can be specified by just two parameters : pressure P, and volume V. This is also true for any simple fluid, which means that we can state the equilibrium conditions of two simple fluid systems with states $S_1 = (P_1, V_1)$ and $S_2 = (P_2, V_2)$ in a formal way as

$$F(P_1, V_1, P_2, V_2) = 0, \tag{1}$$

for some function F. This equation can then be solved for any of the state variables, for example P_1, yielding

$$P_1 = f(V_1, P_2, V_2), \tag{2}$$

for some other function f. It is clear that mutual equilibrium is some relation $\sim$ between equilibrium systems, that is, it is some subset of the set of ordered pairs of equilibriated systems $\Sigma \times \Sigma$, where $\Sigma = P \times V$ is the state space. If the system is not a fluid, then the state space will be a Cartesian product of some other set of state parameters. It is clear that this relation has to be reflexive and symmetric, that is, $S_1 \sim S_1$ and $S_2 \sim S_1 \leftrightarrow S_1 \sim S_2$ for all $S_1, S_2 \in \Sigma$. For this to be an *equivalence relation*, however, we also need the property that $S_1 \sim S_3$ and $S_2 \sim S_3$ implies $S_1 \sim S_2$ (transitivity). These properties are intuitively obvious, and thus the Zeroth Law did not gain the status of a thermodynamic law until the 1930s when it was accepted that it allowed for a proper empirical definition of temperature, which is independent of the concept of entropy. I will state the law now, and provide an empirical justification for it at the end of the section.

The Zeroth Law *: "Being in equilibrium with one another" is an equivalence relation on states of static systems.*

From the transitive property of equivalence relations we know that if two systems S_1 and S_2 are separately in equilibrium with a third system S_3, then S_1 and S_2 are also in equilibrium with one another. Solving $F(P_i, V_i, P_3, V_3) = 0$ for P_3, for the two cases $i = 1, 2$ yields two functions f' and $\hat{f}$. When we set $f' = \hat{f}$ (since S_1 and S_2 are in equilibrium with one another) we obtain

$$f'(P_1, V_1, V_3) = \hat{f}(P_2, V_2, V_3). \tag{3}$$

This immediately implies that

$$F_1(P_1, V_1, P_2, V_2) = f'(P_1, V_1, V_3) - \hat{f}(P_2, V_2, V_3) = 0, \tag{4}$$

which is true if and only if the V_3 dependence cancels out. So f' can be written as

$$f'(P_1, V_1, V_3) = \theta(P_1, V_1)\xi(V_3) - \eta(V_3), \tag{5}$$

and similarly for $\hat{f}$, from which it follows that

$$\theta(P_1, V_1) = \theta(P_2, V_2), \tag{6}$$

is a condition for equilibrium. We have just established the existence of a state function $\theta(P_i, V_i)$ which we call the *empirical temperature* which is homogeneous across systems at equilibrium. In this derivation we assumed that our system was a fluid so that we could characterize it by just pressure and volume, but the same derivation can be performed for any thermal system to yield the equilibrium condition

$$\theta(X_{1,i}, X_{2,i}, ..., X_{n,i}) = \theta_i. \tag{7}$$

where $X_1, ..., X_n$ can be any number of state parameters. The surface defined by this equation is called an *isothermal surface*. We can define the equilibrium relation by stating that $S_1 \sim S_2$ if and only if $\theta_1 = \theta_2$, in which case the systems belong to the same equivalence class

$$[S_k] = \{S \in \Sigma : S \sim S_k\}. \tag{8}$$

The empirical temperature is simply a numerical value, so we have a *total ordering* $<$ between systems with different values of θ_i such that $\theta_1 < \theta_2$ can be interpreted as "S_2 is hotter than S_1". The association of empirical temperature with our sensation of hotness might seem phenomenologically unjustified at this point, but the correspondance will be established later on. Using θ_i as an ordering parameter we can perform a partitioning of state space into a union of isothermal surfaces

$$\Sigma = [S_1] \cup [S_2] \cup [S_3]... \tag{9}$$

with $\theta_1 < \theta_2 < \theta_3$. Equivalence classes are disjoint, this means that isothermal surfaces do not intersect. If we choose a system S_1 as a reference system, and

let another system S be in equilibrium with S_1, then $[S_1]$ is interpreted as the set of all states that $S \in \Sigma$ can be deformed into (have its state parameters continuously varied) such that it maintains equilibrium with S_1. A process occuring at constant temperature is called an isothermal process. If we assume a fluid system $S = (P, V)$, the set $[S_1]$ will be the integral curve of $\theta(P, V) = \theta_1$, that is

$$[S_1] = \{(P, V) \in \Sigma : \theta(P, V) = \theta_1\}. \tag{10}$$

Constructing an empirical temperature scale amounts to choosing a standard system (which could be done arbitrarily) which we call a *thermometer*, and adopting some set of rules for labeling the isothermal surfaces with numerical values [17] (p.10-11). If the state of our thermometer is a function of X and Y, we can construct it such that Y is held constant, and X is allowed to vary. Then the temperature scale will be defined by the function $\theta(X)$ and we call X the *thermometric property*. In general, we may choose the temperature to be any real-valued, continuous and monotonic function of the thermometric property. A convenient choice would be

$$\theta(X) = aX, \tag{11}$$

where a is a real number. Then the temperatures of two systems S_1 and S_2 will be related by

$$\frac{\theta(X_1)}{\theta(X_2)} = \frac{X_1}{X_2}. \tag{12}$$

For temperature to function as a common measure of hotness we need agreement on a calibration point which is reproducable to a high degree of precision. The most common calibration point to use is 273.16 Kelvin (0.01 C) . This is the triple-point of water, the single combination of pressure and temperature at which phases of liquid water, ice and water vapour can be in stable equilibrium with one another. After the thermometer has been brought into thermal contact with a body of water at the triple-point, and the combined system has settled into equilibrium, we measure the thermometric property X_{triple}. The thermometer is now calibrated, and can be removed. This gives us the empirical temperature scale

$$\theta(X) = 273.16 \frac{X}{X_{triple}}. \tag{13}$$

There are many different kinds of thermometers that one may use to define this temperature scale, and the choice of which one to use is done based on convenience (e.g. a gas with the thermometric property of pressure, or a black-body radiator with the property of radiant emittance). This is why the temperature scale is "empirical". Different thermometers generally give different numerical values of $\theta(X)$, even slightly among thermometers with similar construction. What they all must agree on however is the *ordering of hotness*. The discrepancy in numerical values, however, can be minimized by using gas thermometers

at low pressures. Even though $\theta(P)$ will depend on the nature of the specific gas at ordinary pressures, all gases indicate the same temperature as the pressure goes towards zero. For this reason, the gas at vanishing pressure is used to define the empirical temperature scale by the equation [17] (p.17)

$$\theta = 273.16 \lim_{P \to 0} \frac{P}{P_{triple}}. \tag{14}$$

Although this temperature scale is independent on the nature of the specific gas involved, it does depend on the properties of gases in general, which is unsatisfying for a general definition of temperature. But before this can be remedied, additional structure need to be introduced into the theory.

To verify the Zeroth Law, prepare three systems, S_1 , S_2 and S_3. Bring S_1 and S_2 into thermal contact, and let equilibrium settle in while S_2 is kept at constant temperature, as measured by any thermometer (e.g. a helium gas thermometer). Then separate them. Now repeat the process with S_3 instead of S_1. After this procedure, if S_1 and S_3 are brought into thermal contact, one can observe that no heat will flow between the systems. This proves the transitive property of equilibria, as for the symmetry and reflexivity properties, they are obvious to the point of lacking any empirical content. This is why many formulations of the Zeroth Law only mention transitivity. But the essential content of the Zeroth Law is that equilibrium is an equivalence relation, even though we are willing to accept two of the properties as manifestly obvious.

4 The First Law, and Internal Energy

It was previously thought that heat was a fluid endowed with very unusual properties, such as, being weightless and self-repellant. This fluid is referred to as *caloric*. It was thought that if a body contained more caloric, it was hotter. Since caloric was percieved as a material substance, it was thought to be indestructible, so heat conservation was an essential part of the theory. The heat produced by rubbing two bodies together was thought to be due to the caloric being squeezed out of the bodies [18].

The caloric theory had considerable predictive success, most notably through the development of the theory of heat engines and the science of calorimetry. How it is that accurate predictions could successfully be extracted from a theory whose central theoretical term, by and large, fails to refer to anything real, is an interesting topic of philosophy of science. It is a common view that the caloric theory, together with other abandoned theories of the past which once yielded novel and successful predictions such as the ether theory, are simply and plainly false. I consider that view to be grossly mistaken. While there was an ontological discontinuity in the transition between the caloric and the kinetic theories of heat, there was also some retention of structure, over and above the retention of empirical content. Structural continuity across theory change was first emphasized by Henri Poincaré within the context of the transition from Fresnel's theory of light to Maxwell's [19]. A part of the mathematical structure

retained from the caloric theory is the definition of differential heat flow, which is as valid today as it was back then.

The science of calorimetry studies thermal changes in systems undergoes exchanges of heat. This field was developed in particular by Joseph Black in the eighteenth century, who was also the first person to recognize that there was a difference between heat and temperature [8] (p.14). The tools of trade included differential calculus, which allowed a definition of differential heat flow into (or out of) a system as

$$\delta Q = c_V d\theta + \Lambda_\theta dV. \tag{15}$$

Here C_V is the *heat capacity*, which specifies the amount of heat required to raise the temperature of the system (which is held at constant volume), and Λ_θ the *latent heat*, which specifies the amount of heat the system exchanges during an isothermal process. The reason why the heat flow is written δQ and not dQ, is because it's an *inexact differential*. That is, we do not assume Q to be a well defined function of state, for it makes no sense to ask how much heat is contained in a system (however, if the caloric theory was true and a system contained a fixed amount of caloric, that assumption would be plausible). Heat is (like work) a quantity associated with processes rather than states. We can obtain the net heat flow during a *quasistatic process* $\mathcal{P}$ by the integral

$$Q(\mathcal{P}) = \int_{\mathcal{P}} \delta Q = \int_{\mathcal{P}} (c_V d\theta + \Lambda_\theta dV). \tag{16}$$

I take "quasistatic process" to mean a parametrized curve through equilibrium state space Σ, written in a more formal manner as

$$\mathcal{P} = \{S_t \in \Sigma : t_i \leqslant t \leqslant t_f\}, \tag{17}$$

where S_t represents the instantaneous state of the system at time t, t_i the inital state and t_f the final state. For a process to be quasistatic, every state the system passes through during the process must be an equilibrium state. This is essential since the state function θ is ill-defined for non-equilibrium states, and the heat integral would be impossible to carry out otherwise. This is an idealized description, since any heat flow between systems at a finite temperature difference would be non-quasistatic, so the process would have to proceed at an infinitely slow pace. A process during which $Q(\mathcal{P}) = 0$ is called an *adiabatic process*.

For the compression of a gas to be quasistatic it must happen slowly enough for equilibrium to settle in after every infinitesimal increment of compression, to avoid the generation of sound waves propagating through the gas irreversibly dissipating energy. An ideal quasistatic process can of course never exist, but as long as the compression happens fairly slowly the process will be approximately quasistatic. If this criteria is met, the work done by external forces can be described in terms of the state parameters. For example, if our system is a gas which is being compressed quasi-statically by a frictionless piston then the

work performed on the gas is equal to. $W = -PdV$.

That the caloric theory is empirically inadequate can be seen by observing that there is no limit to how much heat can be produced by frictional work (rubbing two bodies together). If caloric is a conserved substance, then the system should eventually run out. But the production of heat depends only on the continued performance of work, and not on any properties of the system itself.

If an amount of work W is performed on a system, then the associated temperature rise will be proportional to W and independent of the nature of the process that accomplished the change of state (e.g. frictional heating or electrical work). Moreover, if in the presence of a temperature gradient a heat flow Q is utilized to perform work (e.g. through the expansion of a gas), then the amount of work performed will be proportional to the supply of heat, again in a path-independent manner. This implies a direct equivalence between heat and work as forms of energy transfer, and these observations can be generalized in the following statement [8] (p.15)

The First Law *: For any cyclic process $\mathcal{C}$, the amount of heat $Q(\mathcal{C})$ absorbed by the system is proportional to the work $W(\mathcal{C})$ performed by the system.*

Cyclic is taken to mean that the process forms a closed loop in Σ. If we take the work $W(\mathcal{C})$ performed as being positive, we can state the first law as

$$JQ(\mathcal{C}) + W(\mathcal{C}) = 0, \tag{18}$$

where $J \approx 4.2$ Nm/cal is *Joule's constant*. James Joule found the value of the constant named after him by taking measurements on the amount of frictional work required to raise the temperature of a gram water by one degree Celsius (though he used units of pounds and Farenheit), and published the result in 1845 in his article "*On the Mechanical Equivalent of Heat*" [20]. This value is today known to be simply the specific heat of water, and J is set to unity by modern conventions, in which the joule is used as a unit of energy instead of the calorie.

We can partition the cycle into infinitesimal elements and integrate over the associated differential quantities through the complete cycle to obtain

$$\oint_{\mathcal{C}} (\delta Q + \delta W) = 0. \tag{19}$$

Since this must be true for any cyclic process $\mathcal{C}$ through Σ, the gradient theorem for line integrals guarantees us the existence of a continuous state function U on Σ, which we will call the *internal energy*, such that

$$dU = \delta Q + \delta W. \tag{20}$$

This is the main content of the first law. It states that heat, like work, is a process of energy transfer between systems, and that the energy of closed

thermal systems are seen to be conserved as soon as we recognize this. The first law relates the change in the energy of a system to the energy transfers occuring at its boundaries, but provides no natural zero-point of energy. There is no natural distinction between heat and work (since they're equivalent), the distinction is rather a pragmatic one. The forms of energy transfer which we can keep track of (e.g compression of a gas by a piston) we designate as "work", while energy which is being transported through microscopic degrees of freedom are labeled "heat" [5] (p.8).

The internal energy is a state function, but the decomposition of dU into Q and W is path dependent. However, as was noted before, if the process is quasistatic then W can be written in terms of the state parameters of the system. Then the differential form of the first law can be written as

$$dU = \sum_i X_i dx_i + \delta Q. \tag{21}$$

The X_i and dx_i are pairs of generalized forces and their conjugate displacements, and these terms represent all the ways in which work can be performed on (or by) the system. They can be pressure P and change in volume dV, magnetic field strength B and change in magnetization dM, surface tension and change in area dA, and so forth.

5 The Second Law, and Entropy

While the first law provides powerful restrictions on the behavior of thermal systems, it still leaves open a manifold of possibilities for their time-evolution. Conservation of energy does not require that systems tend towards equilibrium, or that heat does not spontaneosly flow from a colder body to a hotter one. Merely restricting a closed system to a constant-energy surface in state space does not reduce the dynamical possibilities enough for thermodynamics to have sufficient predictive power. An additional principle is needed to generalize the observed regularities in heat flow, and provide a constraint on the allowed forms of energy transfer. From the First Law we know that mechanical work can be fully transformed into heat, but it gives us little insight regarding the extent to which the converse is possible. A large body of observational evidence regarding the efficiencies of heat engines, and an observed inability to construct a perpetuum mobile of the second kind, compels us to make the following statement [1] (p.30) :

The Second Law (Kelvin-Planck formulation) *: No cyclic process is possible whose sole result is the absorbtion of heat from a reservoir and its complete conversion into mechanical work.*

Taking the term "possible" to mean "allowed by the theory" is favoured by modern philosophers of science [9] (p.9). More specifically I take it to mean that a process is possible if and only if one can specify a model (a set-theoretic

structure satisfying the axioms) of the theory in which it occurs [21]. By "reservoir" I mean a system large enough so that any change of state it might undergo during interactions between it and our system of interest is negligible. The connection between the Second Law and heat flowing against temperature gradients is more vivid in another formulation by Clausius, which asserts the impossibility of cyclic processes which produce no other effect than a transfer of heat from a colder to a hotter body. Many textbook authors claim that these two formulations are logically equivalent, and they show this by proving that a violation of the KP-statement requires a violation of the C-statement, and vice versa. In the proof it is assumed that anti-KP and anti-C engines can be coupled to Carnot engines (a term to be explained) without restrictions, which is not obviously true, especially in a world with negative temperatures. When introducing negative temperatures the Clausius statement is unchanged if hotter is taken to be defined by the direction of heat flow (in this sense, negative temperatures are hotter than positive ones), but the Kelvin-Planck statement is violated. Apart from this, they're equivalent [22].

Central to thermodynamics is the notion of *reversilibity*, which unfortunately (like several other terms) has acquired several meanings in the thermodynamic litterature. Here I follow Planck [9] (p.13) in defining a process $\mathcal{P}$ which induces a transition $(S_i, \mathcal{Z}_i) \xrightarrow{\mathcal{P}} (S_f, \mathcal{Z}_f)$ as reversible if and only if a process $\mathcal{P}'$ is possible which induces the transition $(S_f, \mathcal{Z}_f) \xrightarrow{\mathcal{P}'} (S_i, \mathcal{Z}_i)$, where the $\mathcal{Z}_i$s are thermodynamic states of the surroundings. So the criterion for a process to be reversible is that the original state of the system under consideration can be recovered by another process. If a process is quasistatic, and occurs without frictional losses or hysteresis, then it is also reversible.

A *heat engine* is a machine for converting a temperature gradient into mechanical work. It consists of a working substance, auxiliary parts for manipulating the working substance, and it operates between two reservoirs. Heat exchanges with the reservoirs together with various operations of the auxiliary parts (e.g. moving or compressing the working substance) drive the thermodynamic state of the working substance around a closed path in state space. At the end of each cycle, an amount of work will have been generated. If each part of the cycle is a reversible process, then we call the engine a *reversible engine*, and the total work performed by the systen can be written [3] (p.52)

$$W(\mathcal{C}) = \sum_i \oint_{\mathcal{C}} X_i dx_i. \tag{22}$$

With a simple fluid as a working substance, we can construct a reversible engine by connecting a sequence of four reversible processes such that at the end of stage four the system has returned to its original state. This can be done only in one way.

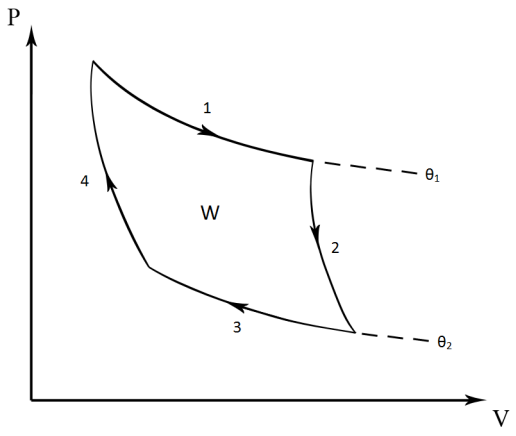
1 : The system expands isothermally and reversibly at a source of temperature θ_1 , while being supplied with an amount of heat Q_1.

2 : The system expands adiabatically and reversibly until it reaches the temperature of the sink θ_2.

3 : The system undergoes reversible isothermal compression while depositing an amount of heat Q_2 at the sink.

4 : The system undergoes reversible adiabatic compression back to temperature θ_1.

This cycle consists of two *isotherms* connected by two *adiabatics*, where an adiabatic is the sets of points $S \in \Sigma = P \times V$ that the state traces through as the system undergoes an adiabatic process, and an isotherm is a path completely contained within an isothermal surface. An engine undergoing such a cycle is called a *Carnot Engine*, and since each part of the cycle is reversible the entire cycle must also be reversible. If we want the cycle to be fully reversible, all heat flows between the engine and the reservoirs must occur while they're at the same temperature, so the system must proceed along the isotherms in an infinitely slow pace. The perfect Carnot engine is therefore impossible to build in practice, but this idealized engine is still very useful in theoretical investigations, and will serve as our prototype reversible engine. Using the Carnot engine one can construct an operational definition of temperature which is independent of the properties of any particular substance, and deduce the existence of another state function which determines the direction of thermodynamic processes. The work performed by this engine is given by the area enclosed by the two isotherms and the two adiabatics. Below is a diagram of the Carnot cycle (references at the end of the thesis).



Uniqueness of isotherms have already been proved, but we also need to assure ourselves that for any adiabatic starting at a specified point along an isotherm

it will always intersect the other isotherm at a unique point. For if this were not the case, the total work done during the cycle would be ill-defined. Or more generally that the set of points that are reachable by a reversible adiabatic process defines a unique surface (or line for $n = 2$). The reason for caution is that what we require is the existence of a state function which is constant for any process with $\delta Q = 0$, which is not obviously the case, since δQ is an inexact differential. The uniqueness of adiabatics is easily proved in the case of two-parameter systems. Consider a fluid undergoing an adiabatic change of state. We may write a differential change in the internal energy as

$$dU = (\frac{\partial U}{\partial P})_V dP + (\frac{\partial U}{\partial V})_P dV. \tag{23}$$

If we insert this expression into the the first law we obtain the adiabatic equation

$$\delta Q = dU + PdV = (\frac{\partial U}{\partial P})_V dP + [P + (\frac{\partial U}{\partial V})_P]dV = 0. \tag{24}$$

The coefficients of dP and dV are both functions of state, which guarantees the existence of unique adiabatics for simple fluids. This can be seen more vividly by representing any change of state by a vector (dP, dV) in P-V space. Then the above condition implies that (dP, dV) is orthogonal to a vector $(f_1(P, V), f_2(P, V))$, which defines a unique line. This argument, however, only holds for systems with a two-dimensional state space, which is therefore what we'll restrict ourselves to in the following discussion. That unique adiabatic surfaces exist for systems with an arbitrary number of state parameters will be proven later on. But one must not assume it beforehand, since the adiabatic equation for an n-parameter system will be given by a linear differential form (or Pfaffian equation)

$$\sum_{i=1}^{n} Y_i dy_i = 0, \tag{25}$$

and such an equation cannot in general be obtained by taking the total differential of a function of state. It is by virtue of the Second Law that δQ has an integrating factor for systems with $n > 2$. The most imporant conclusions of the Second Law can be derived from the existence of unique reversible adiabatic surfaces, as we will be shown later.

The extent to which heat can be converted into mechanical work can be quantified by defining the *thermodynamic efficiency* of the engine as the ratio of the heat Q_H supplied from the hot reservoir, to the amount of energy W successfully converted into mechanical work. By the First Law we know that the heat absorbed from the source must be equal to the sum of the work performed by the engine and the heat deposited at the sink, so we can write

$$\eta = \frac{W}{Q_H} = \frac{Q_H - Q_C}{Q_H} = 1 - \frac{Q_C}{Q_H} \in [0, 1]. \tag{26}$$

It turns out that this quantity is a universal function of the empirical temperatures of the reservoirs *only*. It is independent of the properties of any particular

substance and depends only on the notion of the reversible engine, and can therefore serve to operationally define the absolute temperature scale.

To see that this is the case recall the Clausius statement, that no cyclic process is possible whose sole result is the transfer of heat from a colder body to a hotter one. Since Carnot engines are reversible, one can run them in reverse to act as a refrigerator. Consider two engines S and S' operating between a source at temperature θ_H and a sink at temperature θ_C, with $\theta_C < \theta_H$. Let S be an arbitrary heat engine that supplies work to run a Carnot engine S' in reverse. The arbitrary engine S withdraws heat Q_H from the source, performs work on S', and deposits heat Q_C at the sink. The Carnot engine S', supplied with external work, withdraws heat Q'_C from the sink, and deposits heat Q'_H at the source. For the Second Law to hold, the net flow of heat at the source cannot be negative, which implies that $Q_H \geqslant Q'_H$. Assuming the magnitude of work is the same for both engine, this implies the inequalities

$$\frac{W}{Q_H} \leqslant \frac{W}{Q'_H} \leftrightarrow \eta(S) \leqslant \eta(S'). \tag{27}$$

An immediate corollary is that all Carnot engines are equally efficient, since each can be used to run the other in reverse. If the magnitude of work varies between different Carnot engines the argument still applies, for there is no requirement that the engines operate at the same time, or at the same rate. If their ratio is some rational number $Q_1/Q'_1 = n'/n$, then we consider the composite engines consisting of S executing n cycles of operation, and S' executing n' cycles of operation. Since we assume that the engines are equally efficient each cycle, the results obtained by considering the composite engines apply for the ordinary case as well. So we can state this as a general theorem valid for any reversible engine.

Carnot's Theorem *: No heat engine can be more efficient than a reversible engine operating between the same two temperatures, and all such reversible engines are equally efficient.*

If all reversible heat engines are equally efficient, then this implies that their efficiency must be a universal function of empirical temperature only. This provides for us a way of constructing a general definition of temperature which is independent of the properties of any particular substance. This implies that

$$\frac{Q_1}{Q_2} = f(\theta_1, \theta_2), \tag{28}$$

for some function f. An obvious mathematical property of f is that

$$f(\theta_1, \theta_3) = f(\theta_1, \theta_2) f(\theta_2, \theta_3). \tag{29}$$

For this to be true f must factorize as (proof in the appendix)

$$f(\theta_1, \theta_2) = \frac{Q_1}{Q_2} = \frac{T(\theta_1)}{T(\theta_2)}. \tag{30}$$

The heat flows Q_1 and Q_2 are measurable quantities, so this will serve as the operational definition of temperature, which we can state more precisely as:

Definition (Thermodynamic Temperature) *: The ratio of the thermodynamic temperatures of two heat reservoirs is equal to the ratio of the heats exchanged at those two reservoirs by a reversible heat engine operating between them.*

In general any continuous, real-valued and monotonic function $T(\theta)$ will suffice to establish the thermodynamic temperature scale, and the most convenient choice is $T(\theta) = \theta$. Moreover it is possible to show that for an ideal gas it indeed takes the form $T(\theta) = \theta$. It can be shown by explicitly calculating the efficiency of a Carnot engine and showing it to be equal to $1 - \theta_C/\theta_H$, if one assumes the relation $PV^\gamma = constant$ for adiabatic expansion (and compression) as an empirical fact (which one is justified in doing). It can also be shown more elegantly by an appeal to Maxwell's relations and the laws of Boyle and Joule [1] (p.47), but this will be omitted here. Once we've decided the form of $T(\theta)$ we can fix one reference point (e.g. the triple-point of water) to complete the construction of the thermodynamic temperature scale.

We can now express the efficiency of a Carnot engine operating between a sink at T_C and a source at T_H as

$$\eta = 1 - \frac{T_C}{T_H}. \tag{31}$$

Again we see that uniformity of temperature is a condition for equilibrium, for a Carnot engine operating between two reservoirs at equilibrium must have an efficiency of zero, for otherwise the Kelvin-Planck version of the Second Law would be violated. For the efficiency to be equal to zero, the temperatures evidently must be equal. The characterization of equilibrium between two systems by requiring a Carnot engine operating between them to have zero efficiency will prove useful at the end of the thesis, when thermodynamic temperature is defined for systems in curved space-times.

We can also see from Eq. (31) that if an engine were to operate between two reservoirs, one at a positive and one at a negative temperature, then the efficiency would be greater than 1 and we would have a violation of the Kelvin-Planck statement. These considerations raise some doubt as to whether or not the results just derived are valid when negative temperatures are taken into account. However, thermodynamics deals solely with equilibrium states, and its not clear at all that a negative temperature equilibrium state is even physically realizable, even though we know that negative temperature states can exist for short periods of time.

According to Carnot's theorem, the efficiency of an arbitrary heat engine must not exceed that of a reversible engine operating between the same tem-

peratures, a statement we can express formally as

$$1 - \frac{Q_1}{Q_2} \leqslant 1 - \frac{Q_1^R}{Q_2^R} = 1 - \frac{T_1}{T_2}, \tag{32}$$

or equivalently

$$\frac{Q_2}{T_2} \leqslant \frac{Q_1}{T_1}. \tag{33}$$

If we take the heat entering the system as positive, we can generalize this for engines operating between an arbitrary number n of reservoirs of different temperatures T_i as

$$\sum_{i=1}^{n} \frac{Q_i}{T_i} \leqslant 0, \tag{34}$$

with equality for reversible processes. One can generalize this result to arbitrary systems with any number of degrees of freedom, undergoing cycles of any degree of complexity, including ones where the temperature is allowed to vary continuously during the cycle. Since we have not yet proved the existence of unique adiabatics for systems with more than two degrees of freedom we cannot apply the above Carnot's theorem directly. Instead we will split the complex cycle into infinitesimal elements, and imagine that each element involves a Carnot engine operating between the arbitrary system and a reservoir at temperature T_{res}, so that we can apply Carnot's theorem to the Carnot engine [3] (p. 69-70).

Consider an engine S undergoing an arbitrary cycle $\mathcal{C}$, and split $\mathcal{C}$ into infinitesimal elements at which the temperature of the system is T. A Carnot engine S_C is supplied with work δW_C, in order to extract heat δQ_{res} from the reservoir and transfer heat δQ reversibly to the system. During one such element the heat extracted reversibly from the reservoir by S_C is

$$\delta Q_{res} = \frac{T_{res}}{T} \delta Q. \tag{35}$$

We can integrate this expression through the complete cycle to obtain the net heat extracted from the reservoir

$$Q_{res} = \oint_{\mathcal{C}} \delta Q_{res} = T_{res} \oint_{\mathcal{C}} \frac{\delta Q}{T}. \tag{36}$$

At the end of $\mathcal{C}$ the system has returned to its initial state, so the First Law then implies that $Q_{res} = W_C + W = W_{total}$. If this quantity is positive, then we have a violation of the Kelvin-Planck statement. Thus the cycle must perform work, and heat up its surroundings. We can therefore conclude that

$$T_{res} \oint_{\mathcal{C}} \frac{\delta Q}{T} \leqslant 0. \tag{37}$$

If we assume that T_{res} is positive, we can remove it from the inequality. Moreover if the process is reversible we could let the system act as a refrigerator, and after the same lines of reasoning obtain

$$\oint_{\mathcal{C}} \frac{\delta Q}{T} \geqslant 0. \tag{38}$$

Equations (37) and (38) together imply that equality must hold for reversible processes. The results of the preceeding arguments are summarized thus.

Clausius' Theorem : *For an arbitrary cycle $\mathcal{C}$, regardless of its degree of complexity, the following relation holds :*

$$\oint_{\mathcal{C}} \frac{\delta Q}{T} \leqslant 0, \tag{39}$$

with equality if and only if $\mathcal{C}$ is reversible.

Clausius' theorem is the analytical content of the Second Law, one could take this as a starting point instead of the Kelvin-Planck statement. If $\mathcal{C}$ is reversible, which implies that equality holds in Clausius' theorem, then we can invoke the same argument as in the derivation of the internal energy function, namely that if a line integral around a closed path is zero, then it follows that the integral is path-independent, and that the integrand is the exact differential of a function of state

$$dS = \frac{\delta Q}{T}. \tag{40}$$

The differential δQ becomes exact when multiplied by the integrating factor T^{-1}, and therefore the existence of unique reversible adiabatic surfaces are now guaranteed for systems with any number of state parameters : these are the surfaces of constant entropy, or *isentropic* surfaces.

The above result holds for reversible processes only. For an irreversible process $\mathcal{P}$ inducing a change of state $s_i \to s_f$, complete the cycle by connecting $\mathcal{P}$ to any reversible process $\mathcal{R}$ inducing the change of state $s_f \to s_i$. Clausius' theorem for this cycle becomes

$$\int_{\mathcal{P}} \frac{\delta Q}{T} + \int_{\mathcal{R}} \frac{\delta Q}{T} \leqslant 0. \tag{41}$$

Let $\mathcal{R}'$ be the time-reversal of the process $\mathcal{R}$. If they are exchanged the integral changes sign, therefore

$$\int_{\mathcal{P}} \frac{\delta Q}{T} \leqslant \int_{\mathcal{R}'} \frac{\delta Q}{T} = S(s_f) - S(s_i). \tag{42}$$

This can also be written conveniently in differential form

$$dS \geqslant \frac{\delta Q}{T}. \tag{43}$$

This is a general result valid for processes of any degree of complexity, with equality obtaining if and only if the process is reversible. If we consider a case where the system is isolated from its surroundings and hence $\delta Q = 0$ we obtain a more familiar statement of the Second Law.

The Second Law (Entropy Formulation)*: For any process $\mathcal{P}$ occuring inside an adiabatically enclosed system inducing the change of (equilibrium) state $s_i \xrightarrow{\mathcal{P}} s_f$, the entropy of the system either increases as a result of $\mathcal{P}$ or remains constant, formally*

$$\Delta S_{\mathcal{P}} = S(s_f) - S(s_i) \geqslant 0. \tag{44}$$

If equality holds, then $\mathcal{P}$ is reversible. Otherwise $\mathcal{P}$ is irreversible.

To see what this implies, consider two thermal systems S_1 and S_2, with temperatures T_1 and T_2 respectively forming a composite system $S = S_1 \times S_2$ which is isolated, and has an adiabatic wall separating S_1 and S_2. Now replace the adiabatic wall by a diathermal wall, to enable thermal contact between the subsystems. Then an amount of heat $Q > 0$ will flow between the subsystems until equilibrium is established. The composite system is isolated, and hence the net increase of entropy during the process must be positive. If we assume that S_1 is hotter so that the heat flows from S_1 to S_2, then the entropy change will be negative in S_2 and positive in S_2, hence

$$\Delta S = \Delta S_1 + \Delta S_2 = Q(\frac{1}{T_2} - \frac{1}{T_1}) > 0. \tag{45}$$

This implies that $T_1 > T_2$, which establishes the correspondance between hotness and temperature. Heat will flow between systems at different temperature until the composite system reaches a state of *maximum entropy*, which corresponds to the equilibrium state of uniform temperature. For this reason the statement of the tendency of thermal systems to evolve towards equilibrium is most often explained as a consequence of the Second Law. But this line of reasoning seems suspicious since the entropy function is not even defined for non-equilibrium states, which makes it difficult to speak of an increase in entropy during any non-quasi-static processes. A resolution to the problem is to accept that thermodynamics (at least as described here) aims only at a description of equilibrium systems, and that we can only meaningfully speak of an entropy increase between an initial and a final equilibrium state. This means that we should accept the existence of equilibrium states axiomatically. This is more or less the approach taken by for example Callen [5] (p.13,26), who views the "all-encompassing problem of thermodynamics" as the determination of the final equilibrium state that eventually results after the removal of internal constraints in a closed, composite system initially at equilibrium (for example, after removing an adiabatic wall separating two systems of different temperatures). The final state is the one that maximizes the total entropy.

How can one physically interpret the increase in entropy? There are several

answers to that question, Kelvin for example understood it to mean that natural processes tend towards a "degradation" of energy. To see what that could mean, consider again the previous set up of the composite system consisting of two subsystems at different temperatures initially separated by an adiabatic wall. Instead of letting the heat Q flow from S_1 to S_2 once the adiabatic wall is replaced by a diathermal one, situate a Carnot engine between them and let it operate between S_1 and an external reservoir at the temperature temperature $T_{res} < T_2$ utilizing Q to perform an amount of work W. The work performed by the Carnot engine is given by

$$W = Q(1 - \frac{T_{res}}{T_1}). \tag{46}$$

If we instead let the composite system come to an equilibrium at temperature T_f such that $T_2 < T_f < T_1$, and operate the Carnot engine between the composite system at equilibrium and the reservoir, then the work performed when the engine recieves the same amount of heat Q is given by

$$W' = Q(1 - \frac{T_{res}}{T_f}) < W. \tag{47}$$

The difference in the work performed by the Carnot engine before and after thermal dissipation, which is also the maximum work that can be extracted from the given heat flow δQ , is given by

$$\Delta W = W - W' = T_{res}\Delta S. \tag{48}$$

This means that we can interpret the magnitude of the entropy change as a measure of the extent to which useful energy becomes degraded and unavailable for doing work. When the entropy increases, the maximum amount of work one could possibly extract from the system decreases as well. Or as Kelvin expressed it, the law of increase in entropy represents "*the universal tendency in nature to the dissipation of mechanical energy*" [23]. This can be intuitively reconciled with the understanding of entropy in statistical mechanics, where entropy is a measure of the amount of information (understood in the Shannon sense) required to provide a specification of the microscopic state of the system. Or more informally, a measure of "disorder". The increase in entropy can then be intuitively explained by the fact that there are many more ways in which a system can be disordered than there is for it to be orderely arranged, which makes it very likely that a stochastic system of interacting particles will tend towards a disordered state.

Having found an integrating factor $1/T$ for δQ, we now expressions for both Q and W in terms of functions of state, and can write the differential form of the First Law as

$$dU = \sum_i X_i dx_i + TdS. \tag{49}$$

This equation involes only functions of state, and from this we may conclude that it holds regardless of the nature of the process, whether reversible or not. This equation may then be applied to any irreversible process, provided that a reversible change may be found which connects the initial and final states, and is the starting point of many applications of thermodynamics [1] (p.42). This is sometimes called the fundamental equation of thermodynamics [5] (p. 284).

6 Integrability and the Second Law

It was previously mentioned that the most important consequences of the Second Law, namely the law of increase of entropy and the absolute scale of temperature, can be deduced from the existence of a set of non-intersecting surfaces, each representing a locus of points that can be reached by means of reversible adiabatic processes, such that the union of the surfaces cover the whole of state space. The existence of these surfaces is also equivalent to the existence of an integrating factor $1/\lambda$ for δQ such that $\delta Q/\lambda$ is en exact differential. I will merely derive the operational definition of temperature, but obtaining the law of increase of entropy from there is a fairly straightforward matter, and a discussion can be found in e.g. Adkins. But before that, a short review of the integrability conditions for linear differential forms is in order. A more in-depth exposition can be found in Buchdahl [2] (p.55-65). A generic linear differential form of n independent variables $x_1, x_2, , x_n$ can be written

$$\delta L = \sum_{i=1}^{n} X_i dx_i. \tag{50}$$

where $X_i = X_i(x_1, x_2, , x_n)$ for all $i = 1, 2, , n$, with the δ to emphasize the fact that its not necessarily exact. Since the X_i:s could take on a variety of different forms, we have no reason a priori to expect that δL, consisting of n arbitrary functions, could be obtained by taking the total differential of a single function. In fact, such cases are exceptions rather then generic. Assume that δL is in fact exact, and equal to dR for some function $R = R(x_1, x_2, , x_n)$, i.e.

$$\sum_{i=1}^{n} X_i dx_i = \sum_{i=1}^{n} \frac{\partial R}{\partial x_i} dx_i \tag{51}$$

which implies that $X_i = \frac{\partial R}{\partial x_i}$. But partial derivatives commute, so this implies that

$$\frac{\partial X_i}{\partial x_j} = \frac{\partial X_j}{\partial x_i} \tag{52}$$

for every $i, j = 1, 2, ..., n$. When this condition obtains, the equation $\delta L = 0$ reduces to $dR = 0$, which has the trivial solution

$$R(x_1, x_2, ..., x_n) = k \tag{53}$$

where k is an integration constant. This represents a family of surfaces (hypersurfaces for $n > 3$) which cover the whole space. A somewhat weaker condition than this will suffice for δL to be exact, namely that δL is proportional to an exact differential.

$$\delta L = qdR. \tag{54}$$

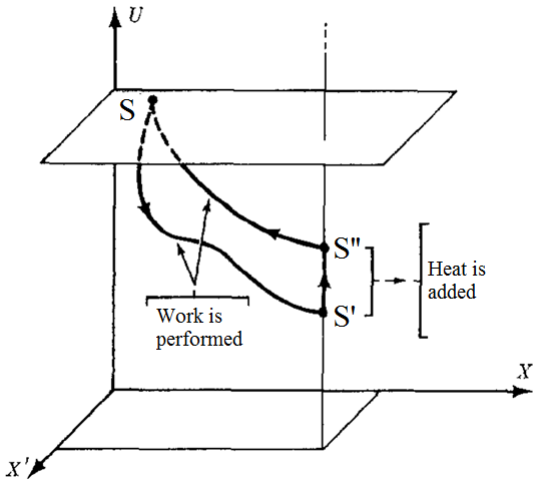
One can verify that in this case a necessary (and in fact sufficient) condition is

$$C_{ijk} = X_i(X_{j,k} - X_{k,j}) + X_j(X_{k,i} - X_{i,k}) + X_k(X_{i,j} - X_{j,i}) = 0. \tag{55}$$

If the adiabatic equation $\delta Q = 0$ is integrable, then we know from Eq. (53) that there exists a state function σ which we will call the *empirical entropy* such that all solution curves to the adiabatic equation for reversible processes will lie on a hypersurface $\sigma = k$ where k is a constant. The converse also holds, namely, that if the locus of points that are reachable by means of reversible adiabatic processes form a continuous surface, and if the union of such (non-intersecting) surfaces cover the whole space, then the adiabatic equation is integrable. The reason why σ is called the "empirical entropy" is because it plays the same role with regards to reversible adiabatic accessibility, meaning the relation of one state being accessible from another by means of a reversible adiabatic process, as that which the empirical temperature plays with regards to the equilibrium relation. Reversible adiabatic accessibility is also an equivalence relation (general adiabatic accessibility however is a pre-order), which guarantees a partitioning of state space into equivalence classes which one can identify as surfaces of constant entropy. One can state the Second Law in terms of the adiabatic accessibility relation, which has been done prominently by Carathéodory. His principle states that in any neighbourhood of an arbitrarily chosen state, there are other states that are inaccessible by means of an adiabatic process (reversible or not). He then makes use of a mathematical theorem (called Carathéodory's Theorem) which states that if for any state S, there are other states in the neighbourhood of S that are inaccessible from S along solution curves to the adiabatic equation $\delta Q_{rev} = \sum_{i=1}^{n} X_i dx_i = 0$, then the adiabatic equation is integrable. It is however possible to prove integrability and the existence of unique reversible adiabatic surfaces directly from the Kelvin-Planck statement, although not in quite as rigorous a manner.

A geometric argument put forward by Zemansky [16] can be used to prove the existence of such surfaces from the Kelvin-Planck statement, and the argument goes as follows. Consider an adiabatically enclosed n-parameter system with the state description $S = (U, x_1, x_2, ..., x_{n-1})$. The state variables $(x_1, x_2, ..., x_{n-1})$ are taken to be the variables which occur as conjugate displacements in each term describing a way in which work can be performed on the system, and U is the internal energy. The state variables $(x'_1, ..., x'_{n-1})$ can be reversibly deformed into any combination of values $(x'_1, ..., x'_{n-1})$ yielding some new value U'. The claim is that for any specified combination of $(x'_1, ..., x'_{n-1})$ only one value U' for the internal energy is possible. To prove this, I will assume the contrary and show that it leads to a violation of the Kelvin-Planck statement.

Then given our initial state S, we can reversibly deform S into two different states $S' = (U', x'_1, ..., x'_{n-1})$ and $S'' = (U'', x'_1, ..., x'_{n-1})$ by means of processes $\mathcal{P}'$ and $\mathcal{P}''$. We can without loss of generality assume that $U'' > U'$. Begin by deforming S into S' by the process $\mathcal{P}'$. Then increase the energy of the system until it reaches U'', while holding the remaining state parameters fixed, for example by performing irreversible work on it. We call this process $\mathcal{W}$, and its consequence is the absorption of heat by the system by an amount given by $Q = U'' - U'$. Since $\mathcal{P}''$ is reversible, there exists another process $\mathcal{P}*$ which induces a transition from S'' to S. Applying $\mathcal{P}*$ after $\mathcal{W}$ brings the system back to its initial state S. Since the system returns to its initial state at the end, and therefore to the same value of the internal energy, the energy supplied during $\mathcal{W}$ must be transferred into the surroundings by means of mechanical work during the processes $\mathcal{P}'$ and $\mathcal{P}*$. Then the combined process $\mathcal{P}'\mathcal{W}\mathcal{P}*$ described a cyclic process whose sole result is the absorption of heat $Q = U'' - U'$ and its complete conversion into mechanical work, in violation of the Kelvin-Planck statement. Therefore we can conclude that only one state corresponding to any set of values for $(x'_1, ..., x'_{n-1})$ can be reached by means of a reversible adiabatic process. Since at least one state can be reached corresponding to $(x'_1, ..., x'_{n-1})$ and since the internal energy function is continuous, the locus of all points reachable from a given state S by means of reversible adiabatic processes must form a continuous surface. Below is a picture illustrating the argument for a system with three state parameters.



That two such surfaces are not allowed to intersect can be shown by a similar argument. If they intersect, then begin at a state situated at the intersection of the two surfaces, proceed along the lower surface (the one which gives the lower value of U), increase U by means of irreversible work to reach the upper surface, and then proceed along that surface back to the initial state. This again represents a cyclic process whose sole consequence is the complete conversion of heat

into mechanical work. Since we now know that the state space Σ is covered by non-intersecting reversible adiabatic surfaces, we can construct a state function σ which labels each surface with a real numerical value.

Consider an n-parameter system. Since σ is a function of state, we may use it as an independent variable in the description of our system, and write $U = U(\sigma, x_1, , x_{n-1})$. Invoking the First Law, the differential of U for a reversible adiabatic change becomes

$$dU = \frac{\partial U}{\partial \sigma} d\sigma + \sum_{i=1}^{n-1} \frac{\partial U}{\partial x_i} dx_i = \delta Q_{rev} + \sum_{i=1}^{n-1} X_i dx_i. \tag{56}$$

Reversibility is assumed so that the work performed can be expressed in terms of the state parameters. For any adiabatic process the heat exchange with the surroundings vanishes, which yields

$$\delta Q_{rev} = \frac{\partial U}{\partial \sigma} d\sigma + \sum_{i=1}^{n-1} (\frac{\partial U}{\partial x_i} - X_i) dx_i = 0. \tag{57}$$

Each state parameter is independent of the others, and δQ_{rev} must be zero whenever $d\sigma$ is zero (since the solution curves lie on surfaces of constant σ), therefore every coefficient in the sum must vanish separately. Consequently

$$\delta Q_{rev} = \frac{\partial U}{\partial \sigma} d\sigma = \lambda d\sigma. \tag{58}$$

We can see that $1/\lambda$ is an integrating factor for δQ_{rev}. Now, it remains to be seen what the physical significance of λ is.

Consider two systems in equilibrium with one another, with states $S' = (x_i', \sigma', \theta)$ and $S'' = (x_i'', \sigma'', \theta)$. Any reversible exchange of heat between the composite system $S = S' \times S''$ and the surroundings can be expressed as

$$\delta Q_{rev} = \delta Q'_{rev} + \delta Q''_{rev} = \lambda' d\sigma' + \lambda'' d\sigma'' = \lambda d\sigma. \tag{59}$$

A priori, without any detailed knowledge of the system, we can expect the λs to be functionally dependent on all state parameters of the given systems, therefore we have the following implicit functions

$$\lambda' = \lambda'(x_i', \sigma', \theta), \tag{60}$$

$$\lambda'' = \lambda''(x_i'', \sigma'', \theta), \tag{61}$$

$$\lambda = \lambda(x_i', x_i'', \sigma', \sigma'', \theta). \tag{62}$$

But if we rewrite Eq. (59) in the following way :

$$d\sigma = \frac{\lambda'}{\lambda} d\sigma' + \frac{\lambda''}{\lambda} d\sigma'', \tag{63}$$

we can see that the entropy of the composite system is a function of the entropies of the subsystems only. Hence it follows that $\partial\sigma/\partial\sigma' = \lambda'/\lambda$ and $\partial\sigma/\partial\sigma'' = \lambda''/\lambda$

are also functions of σ' and σ'' only. For this to be the case, there can be no functional dependence of the λs on the x_is. Hence we obtain the implicit functions

$$\lambda = \lambda(\sigma', \sigma'', \theta), \tag{64}$$

$$\lambda' = \lambda'(\sigma', \theta), \tag{65}$$

$$\lambda'' = \lambda''(\sigma'', \theta). \tag{66}$$

Since λ'/λ and λ''/λ are functions of σ and σ' only (and hence not of θ), it follows that the θ dependence must cancel out, and therefore the λs must factorize into the form

$$\lambda(\theta, \sigma) = T(\theta)F(\sigma). \tag{67}$$

This can be shown more rigorously as follows. Since the λ quotients are independent of θ we obtain

$$\frac{\partial}{\partial\theta}(\frac{\lambda'}{\lambda}) = \frac{\partial}{\partial\theta}(\frac{\lambda''}{\lambda}) = 0. \tag{68}$$

If we carry out the derivative of $\frac{\lambda'}{\lambda}$ we find that

$$\frac{\partial}{\partial\theta}(\frac{\lambda'}{\lambda}) = \frac{\frac{\partial\lambda'}{\partial\theta}\lambda - \lambda'\frac{\partial\lambda}{\partial\theta}}{\lambda^2} = 0. \tag{69}$$

This is true if and only if

$$\frac{\partial\lambda'}{\partial\theta}\frac{1}{\lambda'} = \frac{\partial\lambda}{\partial\theta}\frac{1}{\lambda} \Longleftrightarrow \frac{\partial}{\partial\theta}(\log\lambda) = \frac{\partial}{\partial\theta}(\log\lambda'). \tag{70}$$

Consequently Eq. (68) yields

$$\frac{\partial}{\partial\theta}(\log\lambda) = \frac{\partial}{\partial\theta}(\log\lambda') = \frac{\partial}{\partial\theta}(\log\lambda''). \tag{71}$$

In the second term we have a function of σ' and θ only, and in the third term a function of σ'' and θ only. Since σ' and σ'' are independent variables, it follows that all of the expressions in Eq. (71) must be equal to some universal function $g(\theta)$ of empirical temperature *only* . In other words

$$\frac{\partial}{\partial\theta}(\log\lambda) = g(\theta). \tag{72}$$

If we integrate this expression with respect to θ and solve for λ we obtain

$$\lambda(\sigma, \theta) = F(\sigma)e^{\int g(\theta)d\theta}. \tag{73}$$

where the form of the function $F(\sigma)$ is determined by the choice of empirical entropy scale. Now, boldly make the definitions:

$$T(\theta) = Ce^{\int g(\theta)d\theta}, \tag{74}$$

$$S(\sigma) = \frac{1}{C}\int F(\sigma)d\sigma. \tag{75}$$

where C can be any arbitrary constant. Now the equation for reversible heat transfer has the form

$$\delta Q_{rev} = \lambda d\sigma = e^{\int g(\theta)d\theta} F(\sigma)d\sigma = TdS \tag{76}$$

where $T(\theta)$ is a universal function of the empirical temperature, and $S(\sigma)$ a universal function of the empirical entropy.

Now consider a system undergoing cyclic and reversible operations. The system is initially at a point in the intersection of a reversible isothermal surface at θ_1 and a reversible adiabatic surface at σ_1. It then undergoes an isothermal change shifting its position in state space to a new reversible adiabatic surface at σ_2, while staying at the surface θ_1 and exchanging an amount of heat Q_1 with the surroundings. It then moves along the surface σ_2 to a new reversible isothermal surface θ_2. It then proceeds along θ_2 back to the surface σ_1 while exchanging an amount of heat Q_2 with the surroundings. It then proceeds along σ_1 back to its initial state. Such a process is analogous to a Carnot cycle, and the heat exchanges are given by

$$Q_1 = T(\theta_1) \int_{\sigma_1}^{\sigma_2} F(\sigma)d\sigma = T_1 \int_{S_1}^{S_2} dS \tag{77}$$

$$Q_2 = T(\theta_2) \int_{\sigma_2}^{\sigma_1} F(\sigma)d\sigma = T_2 \int_{S_2}^{S_1} dS \tag{78}$$

Which yields the previously derived operational definition of temperature

$$\frac{Q_1}{Q_2} = \frac{-T_1}{T_2}. \tag{79}$$

7 The Tolman-Ehrenfest Effect

There are three primary ways in which thermodynamics needs modification to become compatible with general relativity: modification to the state-variables to ensure Lorentz invariance, new conditions for thermal equilibria in curved space-time (meaning that a relativistic gravitational field is present), and the incorporation of the mass-energy relation into the First Law. Lorentz transformations of thermal systems is a thorny issue which has garnered a great deal of controversy during the last decade during which several mutually incompatible transformation laws for temperature (and by proxy other quantities) have been derived [4][25]. It is possible that this stems from the fact that thermodynamical quantities are only definied for equilibrium states, and it's not at all clear that a state of equilibrium can be reached for a thermal system in relative motion to a heat bath, through for example the exchange of thermal radiation [24]. For this reason, the issue of the transformation law of temperature might perhaps more properly be dealt with within the context on non-equilibrium thermodynamics, and a discussion will be omitted here.

Incorporation of the mass-energy relation into the First Law by adding a

term Δmc^2 to ΔU will allow one to study, for example, equilibrium between matter and radiation, where the matter could consist of electrons and positrons annihilating to produce radiation, and the radiation could form electron-positron pairs from the quantum vacuum. These matters have been discussed by Tolman [4], among others. Although I will make use of the modification to the First Law, a thorough discussion will also be omitted for the sake of brevity, and instead I will focus on the issue of thermal equilibrium in curved space-time.

Thermal equilibria in the presence of relativistic gravitational fields will have qualitatively different properties from both the non-gravitational case as well as the Newtonian-gravitational case. In developing classical thermodynamics one assumes (although not always explicitly) either that no gravitational fields are present, or in some cases where they are treated they are Newtonian, and thus act on the rest mass of the system only. An effect of introducing a Newtonian gravitational field is that the system will become stratified into layers of constant gravitational potential ϕ [2] (p.143), and the chemical potential will have to be modified to take into account the work required to move an amount of matter to a layer of greater ϕ. The temperature however remains homogeneous across the system at equilibrium even in the presence of a Newtonian gravitational field. This is no longer the case in the presence of relativistic gravitational fields. The reason for this is the *equivalence principle*, which states an equivalence between gravitational and inertial mass. According to the celebrated equation $E = mc^2$ anything which has an energy also has an inertial mass, and this includes heat and radiation. The equivalence principle then implies that the gravitational field will act on all forms of energy in the system, not only on the rest mass of its constituent particles. So heat which flows along the gradient of a gravitational potential will lose energy, and electromagnetic radiation will be redshifted. That this implies a non-uniformity of temperature at equilibrium can be seen by engaging in two thought experiment, courtesy of N.L. Balazs and J.M. Dawson [6].

Consider a cylinder, whose bottom is a black-body radiator, with perfectly reflecting walls and top. Situate the cylinder in a static gravitational field such that the field gradient points along the symmetry axis of the cylinder towards the bottom. Let the interior, consisting of thermal radiation, come to equilibrium with the bottom. To measure the temperature, one uses a radiation thermometer measuring the thermometric property of peak emittance of black-body radiation. At each layer of the cylinder the thermometer will measure a Planckian spectrum whose peak emittance determines the temperature of that layer by the equation $kT = 0.2014hf_{peak}$, where k is Boltzmann's constant, h is Planck's constant and f_{peak} is the peak frequency of the black-body radiation. Due to the equivalence principle f_{peak} will vary as one moves through a potential difference $\Delta\phi$ according to the formula $\Delta f_{peak} = -f_{peak}\Delta\phi/c^2$, where c is the speed of light. This implies that the temperatures of two layers 1 and 2 inside the cylinder separated by $\Delta\phi$ are related by the equation

$$T_1 = T_2(1 - \frac{\Delta\phi}{c^2}). \tag{80}$$

Evidently the temperature will not be uniform at equilibrium, but regions at lower gravitational potential will be slightly hotter. To prove that this result is not merely an artifact of a specific empirical temperature scale but holds for the thermodynamic scale as well, we can apply the Second Law to a Carnot engine operating between reservoirs at different gravitational potentials.

Consider a thermal system in a gravitational field ϕ. Let the engine operate between two layers of the system separated by a potential difference $\Delta\phi$. Let the engine absorb an amount of heat Q from the layer at greater ϕ, call it layer 1. Then lower the engine to the layer 2 and deposit the heat Q there. Complete the cycle by returning it to its initial state. Since the engine returns without the heat Q, a smaller magnitude of work will need to be performed in bringing it back up than was gained by lowering it down the potential, while the engine still contained Q. By the equivalence of energy, inertial mass and gravitational mass, this results in a potential energy gain of $\Delta E = \frac{Q}{c^2}\Delta\phi$, which could be utilized to perform work. In order to avoid violation of the Second Law an additional amount of work equal to the potential energy gain must have been performed when depositing Q, implying that the lower subsystem has a higher temperature. The efficiency of the Carnot engine is given by $\eta = 1 - T_1/T_2$, so setting $W = \Delta E$ yields

$$(1 - \frac{T_1}{T_2})Q = \frac{Q}{c^2}(\phi_1 - \phi_2) \leftrightarrow T_1 = T_2(1 - \frac{\Delta\phi}{c^2}). \tag{81}$$

This result proves that variations in temperature distributions at equilibrium in curved space-times is not merely an artifact of a specific empirical temperature scale, but holds for the thermodynamic scale as well. This is known as the *Tolman-Ehrenfest effect*. A more general result has been proved by Tolman for systems in general static space-times [4][10]. He derived his result by a direct application of Einstein's field equations, using the stress-energy tensor for a radiation gas to represent a radiation thermometer in thermal contact with the systems whose temperature he wants to compare. His result reads

$$T\sqrt{g_{tt}} = k, \tag{82}$$

where g_{tt} is the time-component of the metric and k is a constant. If the metric is Schwarzschild then this expression implies Eq. (81). But using Carnot cycles is preferable to Tolman's approach since it avoids any appeal to Einstein's field equations, in spirit with the methodology of phenomenological thermodynamics, where mention of any underlying dynamics should be avoided in the formulation of the theory, if possible. Alas, there is more to be wanted in that regard, as the derivations using Carnot cycles still make use of the mass-energy relation.

One can obtain a *covariant* expression of the Tolman-Ehrenfest relation, by applying the Kelvin-Planck statement to an engine executing Carnot cycles in a general *stationary* space-time to obtain its maximum efficiency. Analogously to the non-relativistic case, one can obtain a characterization of thermal equilibrium in curved space-time by demanding that the efficiency of a reversible engine operating between two reservoirs at equilibrium with one another is zero.

The following presentation will follow closely a 1971 article by Ebert and Göbel [7].

A space-time is stationary if the metric components can be written in a form in which the time coordinate does not appear, by means of a suitable coordinate transformation. This means that the geometry of space-time appears unchanging with time for some set of privileged observers, which is reasonable to demand of a space-time in which we wish to study thermal equilibrium. If no such observers existed no state of time-translational invariance, including thermal equilibria, would exist either. The world-lines of those observers are a vector field on the space-time manifold called a *time-like Killing field*. Measurements of a system in thermal equilibrium can be done with respect to the reference-frame of any world-line of the Killing field.

Consider a thermal system in equilibrium, situated in a stationary space-time equipped with a time-like Killing field ξ with respect to which the equilibrium is established. This could be for example the terrestrial gravitational field. We imagine that the system is partitioned into stationary subsystems with world-lines given by ξ. We assume that the gravitational potential is uniform across each subsystem, and that each subsystem is large enough to serve as a heat reservoirs for a Carnot engine. The Carnot engine is also assumed to be small enough not to disturb the gravitational field significantly or alter the world-lines of the subsystems. The engine executes Carnot cycles between two subsystems, which we label α and β such that $\phi(\beta) > \phi(\alpha)$. This cycle is identical to the regular Carnot cycle apart from the transportation of heat between the subsystems, and hence across the gravitational potential. The motion of the engine between the subsystems is assumed to be *geodesic motion*, which means free fall. The speed of light is set to unity and the (+ - - -) convention is used for the classification of time-like four-vectors. The cycle consists of four parts.

I : The engine undergoes an isothermal and reversible change of state $s_a \to s_b$ while in contact with the subsystem α, recieving an amount of heat Q_α from α during the process. This increases the mass of the engine to $m_b = m_a + Q_\alpha$.

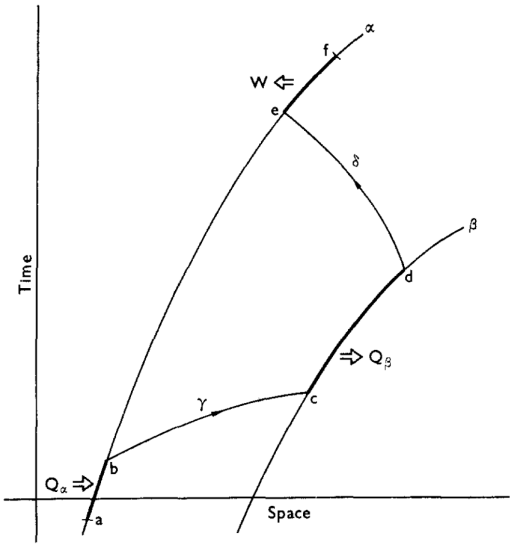
II : The engine is supplied with an amount of kinetic energy E_b^{kin}, which sends it along a geodesic γ to the subsystem β, and undergoes a reversible adiabatic change of state $s_b \to s_c$ reaching the temperature of β.

III : The engine undergoes a reversible isothermal change of state $s_c \to s_d$ while in contact with subsystem β, depositing an amount of heat Q_β at β during the process. The energy of the engine is now $m_d = m_c - Q_\beta$.

IV : The engine falls along a geodesic δ back to the subsystem α gaining the kinetic energy E_d^{kin}, and undergoes an adiabatic and reversible change $s_d \to s_e$. The net work W obtained in the cycle is then stored at α.

s_i represents the thermodynamic state of the system at the space-time point $i = a, b, c, d, e$. One can construct the cycle such that the kinetic energies E_c^{kin}

and E_d^{kin} are equal, in which case the engine only deposits heat at β without performing any work. Below is a space-time diagram of the cycle.



All measurements of the energy of the engine are done with respect to the world-line ξ_α. Let $E_p(\lambda)$ be the energy of the engine (as measured from ξ_α) when located at a space-time point p and moving on a world-line λ. Letting $u_p(\lambda)$ and $\xi/||\xi_\alpha||$ be the unit tangent vectors (or four-velocities) of the engine and observer world-lines respectively, we have the equation

$$E_p(\lambda) = \frac{m_p}{||\xi_\alpha||}\langle u_p(\lambda), \xi\rangle. \tag{83}$$

The net work performed during one cycle is equal to the energy $E_e(\delta)$ of the engine as measured at the end of the cycle, minus the energy $E_a(\alpha)$ as measured at the start and the kinetic energy E_b^{kin} supplied to initiate the process. This gives us

$$W = E_e(\delta) - [E_a(\alpha) + E_b^{kin}] = E_e(\delta) - [E_a(\alpha) + E_b(\gamma) - E_b(\alpha)]. \tag{84}$$

The second equality is seen to hold by noting that E_b^{kin} is the difference of the energies as measured before and after departure from α. Similarly, the kinetic energy of the engine upon arrival at β is the difference between the energies as measured before and after arrival, and likewise for the kinetic energy at departure from β. We previously assumed that $E_c^{kin} = E_d^{kin}$, which can be written

$$E_c(\gamma) - E_c(\beta) = E_d(\delta) - E_d(\beta). \tag{85}$$

The energy of the engine is constant along the geodesics γ and δ, so $E_c(\gamma) = E_b(\gamma)$ and $E_d(\delta) = E_e(\delta)$. Substituting the above relations into Eq. (84) and

using the fact that $E_b(\alpha) - E_a(\alpha) = Q_\alpha$ we obtain

$$W = Q_\alpha + E_d(\beta) - E_c(\beta). \tag{86}$$

The unit four-velocity of the engine at β is given by $\xi/||\xi_\beta||$ (the engine is co-moving with the subsystem β), so the energies at c and d are given by

$$E_i(\beta) = \frac{m_i}{||\xi_\alpha||||\xi_\beta||}\langle \xi, \xi \rangle = m_i \frac{||\xi_\beta||}{||\xi_\alpha||}. \qquad i = c, d \tag{87}$$

The inner product of the Killing fields is evaluated for $\xi = \xi_\beta$, the world-line of the engine at the time of measurement. The engine deposits an amount of heat Q_β at β, so the masses at the points c and d are related by the equation $m_d = m_c - Q_\beta$. Eq. (86) now becomes

$$W = Q_\alpha - Q_\beta \frac{||\xi_\beta||}{||\xi_\alpha||}. \tag{88}$$

If the gravitational field is negligible, or if the reservoirs are at the same gravitational potential, then this expression reduces to the ordinary application of the First Law, namely $W = Q_{source} - Q_{sink}$. We can divide Eq. (88) by Q_α to obtain the thermodynamic efficiency of a Carnot engine operating in a gravitational field :

$$\eta = 1 - \frac{Q_\beta ||\xi_\beta||}{Q_\alpha ||\xi_\alpha||}. \tag{89}$$

Defining the absolute temperature scale for systems in the presence of space-time curvature proceeds in a way analogous to the non-relativistic case, except that the Carnot efficiency is now functionally dependent on $||\xi||$ as well, not only on the empirical temperature θ. Written mathematically

$$\frac{Q_\alpha}{Q_\beta} = f(\theta_\alpha, ||\xi_\alpha||, \theta_\beta, ||\xi_\beta||). \tag{90}$$

The property of the heat ratio that

$$\frac{Q_\alpha}{Q_\beta}\frac{Q_\beta}{Q_\gamma} = \frac{Q_\alpha}{Q_\gamma}, \tag{91}$$

implies the functional equation

$$f(\theta_\alpha, ||\xi||_\alpha, \theta_\beta, ||\xi||_\beta) f(\theta_\beta, ||\xi||_\beta, \theta_\gamma, ||\xi||_\gamma) = f(\theta_\alpha, ||\xi||_\alpha, \theta_\gamma, ||\xi||_\gamma). \tag{92}$$

The theorem used in the non-relativistic definition of temperature (and proved in the appendix) is a special case of a general theorem which applies equally well here. Hence

$$f(\theta_\alpha, ||\xi||_\alpha, \theta_\beta, ||\xi||_\beta) = \frac{g(\theta_\alpha, ||\xi||_\alpha)}{g(\theta_\beta, ||\xi||_\beta)}, \tag{93}$$

where $g(\theta, ||\xi||)$ can be any continuous, real-valued and monotonic function of θ and $||\xi||$. A convenient choice of definition would be one which implies that non-relativistic thermodynamics is valid *locally*. Any observer concerned only with the thermodynamics of his/her local surroundings, and is hence unaware of any gravitational influences, should define the temperature function to be the same as in the non-relativistic case. Therefore we set

$$g(\theta, ||\xi||) = T. \tag{94}$$

This gives us the operational definition of temperature

$$\frac{T_\alpha}{T_\beta} = \frac{Q_\alpha}{Q_\beta}. \tag{95}$$

The Carnot efficiency then becomes

$$\eta = 1 - \frac{T_\beta ||\xi||_\beta}{T_\alpha ||\xi||_\alpha}. \tag{96}$$

In generic situations the ratio of the norms is very close to one, making the ordinary result a good approximation. However in more exotic situations, for example if one considers an engine in the vicinity a rotating black hole, things can become very different. At the Killing horizon of the rotating black hole, the boundary beyond which no observer can remain static, the norm of the Killing field vanishes, enabling the possibility of a perfectly efficient engine.

I take the Kelvin-Planck statement for granted, and define thermal equilibrium between systems in curved space-time by stating that a Carnot engine operating between them must have zero efficiency. This leads us to the covariant form of the Tolman-Ehrenfest relation.

The Tolman-Ehrenfest Relation : *Given a stationary space-time with a Killing field ξ, if two systems α and β are in thermal equilibrium with one another, then their thermodynamic temperatures T_α and T_β are related by*

$$T_\alpha ||\xi_\alpha|| = T_\beta ||\xi_\beta||, \tag{97}$$

where $||\xi_\alpha||$ and $||\xi_\beta||$ are the norms of the Killing fields at the positions of the respective systems.

The Zeroth Law is not violated because there still is a quantity that is uniform at equilium, that quantity however now depends not only on the thermodynamic variables, but also on the gravitational field. One could make the definition $g(\theta, ||\xi||) = T||\xi||$, in which case temperature would be uniform at equilibrium. A drawback of such a definition is that explicitly incorporating gravity, and having to deal with arbitrary constants relating to our freedom in the choosing the zero-point of the gravitational potential, makes things needlessly complicated.

By now it should be clear that general relativity actually predicts that temperature (as ordinarily understood) should not be uniform at equilibrium. Although probably existent, the effect is very small. If we imagine that the experiment to measure the temperature variations is being performed on the earth,

then we may use the Schwarzschild metric for which the time-component of the metric is given by

$$g_{tt} = 1 - \frac{2M_\oplus G}{(R_\oplus + h)c^2}. \tag{98}$$

Here $M_\oplus \approx 5.97 \cdot 10^{24} kg$ and $R_\oplus \approx 6.37 \cdot 10^6 m$ are the mass and radius of the earth, $c \approx 3 \cdot 10^8 m/s$ is the speed of light and $G \approx 6.67 \cdot 10^{-11} Nm^2/kg^2$ is the gravitational constant. And h represents the height above ground level. If we imagine a long tube streching from ground level to $1km$ up in the air where the interior of the tube is in thermal equilibrium, then the ratio of the temperature of the interior at ground level to the temperature at the top is given by

$$\frac{T_{ground}}{T_{h=1000m}} = \frac{\sqrt{1 - \frac{2M_\oplus G}{R_\oplus c^2}}}{\sqrt{1 - \frac{2M_\oplus G}{(R_\oplus + 1000)c^2}}} \approx 1 - 1.09 \times 10^{-13}. \tag{99}$$

This is evidently a very small effect, and to date there is (to my knowledge) no experimental verification of its existence. However, its existence can be derived in several different ways using very plausible assumptions, namely the mass-energy equivalence and the equivalence of inertial and gravitational mass, both of which are hypotheses which have a large amount of empirical data to back them up. Derivations of the Tolman-Ehrenfest effect have also been found using general relativistic statistical mechanics [27], and also using the classical relation $1/T = \partial S/\partial U$ [26].

An interesting recent development is the proposal by C. Rovelli, H.M. Haggard and M. Smerlak [28][29] that the Tolman-Ehrenfest effect can be understood by appealing only to the equivalence principle and the notion of *thermal time*, which is a measure of the number of states a system transits during intervals of proper time. They base this notion on the fact that at equilibrium, the time it takes for a quantum system to evolve to a distinguishable quantum state is proportional only to the temperature. They propose characterizing thermal equilibrium by proportionality of thermal time τ to proper time t according to $\tau = \frac{kT}{h}t$. In natural units T then takes on units of states per seconds, giving temperature the physical interpretation of a measure of the rate at which a system samples its available microstates, or as the rate of thermal time with respect to proper time. Equilibrium between systems is according to them characterized by a vanishing of the net information transfer between the systems, the information transfer being the logarithm of the number of states transited between them. The Tolman-Ehrenfest effect can be intuitively understood in these terms. Consider two thermal systems S_1 and S_2, at equilibrium with one another, with S_1 residing in a stronger gravitational field. S_1 will then be hotter due to the Tolman-Ehrenfest effect, increasing the information flow in states/second from S_1 to S_2. However, due to gravitational time-dilation one second will last longer (to an external observer) in S_1 than in S_2. The two effects balance out and the net information transfer between the systems vanishes.

8 Appendix

If f is a real valued, non-negative and differentiable function then we have that

$$f(x,y) = f(x,z)f(z,y) \leftrightarrow f(x,y) = \frac{T(x)}{T(y)}. \tag{100}$$

PROOF: The implication to the left is trivial. As for the implication to the right, consider the logarithm of f:

$$\log f(x,y) = \log f(x,z) + \log f(z,y). \tag{101}$$

Taking the derivative of both sides with respect to z yields

$$\frac{\partial}{\partial z}\log f(x,y) = \frac{\partial}{\partial z}\log f(x,z) + \frac{\partial}{\partial z}\log f(z,y) = 0. \tag{102}$$

This implies that

$$\frac{\partial}{\partial z}\log f(x,z) = -\frac{\partial}{\partial z}\log f(z,y) = g(z). \tag{103}$$

Integration gives us

$$\log f(x,z) = h(x) + \int_{z_0}^{z} g(z')dz', \tag{104}$$

$$\log f(z,y) = k(y) - \int_{z_0}^{z} g(z')dz'. \tag{105}$$

This gives us two different expressions for $\log f(x,y)$

$$\log f(x,y) = h(x) + \int_{z_0}^{y} g(z')dz' = k(y) - \int_{z_0}^{x} g(z')dz'. \tag{106}$$

Which can be rearranged to yield

$$h(x) + \int_{z_0}^{x} g(z')dz' = k(y) - \int_{z_0}^{y} g(z')dz' = c \rightarrow h(x) = c - \int_{z_0}^{x} g(z')dz'. \tag{107}$$

The second equality holds since the left-hand-side is independent of y, and the right-hand-side is independent of x, so both sides must be equal to a constant c. This gives us

$$\log f(x,y) = h(x) - h(y) + c \rightarrow f(x,y) = e^c \frac{e^{h(x)}}{e^{h(y)}} = e^c \frac{T(x)}{T(y)}. \tag{108}$$

By considering the equation $f(x,x) = f(x,y)f(y,x)$ we see that $e^c = e^{2c}$, therefore $e^c = 1$. In conclusion

$$f(x,z) = f(x,y)f(y,z) \rightarrow f(x,y) = \frac{T(x)}{T(y)}, \tag{109}$$

which proves the theorem.

9 Pictures

The Carnot engine - Picture taken (and modified) from
http://www.codecogs.com/library/engineering/thermodynamics/index.php .
Unique adiabatic surfaces - Picture taken (and modified) from [16].
The Carnot engine in a gravitational field - Picture taken from [7].

References

[1] A. B. Pippard, *The Elements of Classical Thermodynamics*, Cambridge University Press, 1964.

[2] H. A. Buchdahl, *The Concepts of Classical Thermodynamics*, Cambridge University Press, 1966.

[3] C. J. Adkins, *Equilibrium Thermodynamics*, Third edition, Mc Graw-Hill, London, 1983.

[4] R. C. Tolman, *Relativity Thermodynamics and Cosmology*, Dover Puclications Inc, New York, 1934.

[5] H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*, Second edition, John Wiley and Sons Inc. , 1985.

[6] N. L. Balazs and J.M. Dawson, *On Thermodynamic Equilibrium in a Gravitaitonal Field*, Physical Vol 31, 1965.

[7] R. Ebert and R. Göbel, *Carnot Cycles in General Relativity*, GRG Vol.4, No.5, p. 375-386, 1973.

[8] J. Uffink, *Compendium on the Foundations of Classical Statistical Physics*, Handbook for the Philosophy of Physics (J. Butterfield and J. Earman (eds)), Amsterdam Elsevier, pp. 924-1074, 2007.

[9] J. Uffink, *Bluff your way in the Second Law of Thermodynamcs*, Studies in History and Philosophy of Modern Physics 32.3, pp. 305394, 2007.

[10] R. C. Tolman, *Temperature Equilibrium in a Static Gravitational Field*, Physical Review, December 1930.

[11] E. H. Lieb and J. Yngvason, *The Physics and Mathematics of the Second Law of Thermodynamics* 1999, Physics Reports, 1999 - Elsevier.

[12] Robert Batterman, *The Oxford Handbook of Philosophy of Physics*, Oxford University Press, 2013.

[13] J. Thewlis, *Concise dictionary of physics and related subjects* (1st ed.). Oxford: Pergamon Press. p. 248., 1973.

[14] D.R. Owen, *A First Course in the Mathematical Foundations of Thermodynamics*, Springer-Verlag, New York, 1984.

[15] C. Carathéodory, *Investigations into the foundations of thermodynamics*, Dowden, Hutchinson and Ross, 1976.

[16] M. Zemansky, *Kelvin and Caratheodory, A Reconciliation*, The City College of the City University of New York, New York (Received 28 December 1965).

[17] M. Zemansky, *Heat and Thermodynamics*, Fifth edition, McGraw-Hill Book Company Inc. , 1968.

[18] S. Carnot, *Reflections on the Motive Power of Fire and on Machines Fitted to Develop that Power* , Paris, 1824.

[19] J. Ladyman, *Structural Realism*, The Stanford Encyclopedia of Philosophy (Summer 2013 Edition), Edward N. Zalta (ed.), 1. Introduction.

[20] J. Joule, *On the Mechanical Equivalent of Heat*, Philosophical Transactions of the Royal Society of London, Vol. 140, 1850, p.61-82.

[21] Frigg, Roman and Hartmann, Stephan, *Models in Science*, The Stanford Encyclopedia of Philosophy (Fall 2012 Edition), Edward N. Zalta (ed.), 2.3 Set-theoretic structures.

[22] N. F. Ramsey, *Thermodynamics and Statistical Mechanics at Negative Absolute Temperatures*, Physical Review, Vol. 103, July 1 1956, p.23.

[23] Sir W. Thomson (Lord Kelvin), *On a Universal Tendency in Nature to the Dissipation of Mechanical Energy*, Proceedings of the Royal Society of Edinburgh for April 19, 1852, also Philosophical Magazine, Oct. 1852.

[24] P. T. Landsberg, *Laying the ghost of the relativistic temperature transformation*, Physics Letters A Volume 223, Issue 6, 23 December 1996, Pages 401403.

[25] T. K. Nakamura, *Three Views of a Secret in Relativistic Thermodynamics*, Prog. Theor. Phys. 128 (2012), 463-475.

[26] L. D. Landau and E.M. Lifshitz, *Statistical Physics*, Pergamon Press, London and New York, chapter 29, 1959.

[27] G. E. Tauber and J. W. Weinberg, Phys. Rev. 122.

[28] H. M. Haggard and C. Rovelli, *Death and resurrection of the zeroth principle of thermodynamics*, Physical Review D, 2013 - APS.

[29] C. Rovelli and M. Smerlak, *Thermal time and the Tolman-Ehrenfest effect: 'temperature as the speed of time'*, Quantum Gravity Vol. 28 N.7, 2011.