

# A SIMPLE, UNIFIED APPROACH TO THE FIRST AND SECOND LAWS OF THERMODYNAMICS

J. KESTIN

*Department of Engineering, Brown University*

## ABSTRACT

The article gives a new verbal formulation of the second law of thermodynamics. It is claimed that the physical content of this statement as well as the derivation of the mathematical consequences normally referred to as the first and second parts of the second law are simpler and more easily grasped by beginners than the standard formulations. The argument is so designed as to be closely modelled on one which pertains to the derivation of the mathematical formulation of the first law.

## 1. MOTIVATION FOR THIS ARTICLE

There is little advantage, from the point of view of advancing progress in physics, in reopening the question of the optimal formulation of the second law of thermodynamics. However, a case can be made for returning to this fundamental topic in the interests of those who are engaged in transmitting existing knowledge.

Regardless of which primary formulation of the second law is adopted, it is commonly agreed that it must lead, by an easy logical and mathematical derivation, to three statements:

(a) There exists a property called entropy,  $S$ , which is additive for subsystems and which possess the mathematical properties of a potential.

(b) There exists a variable, called thermodynamic temperature,  $T$ , which has the mathematical property of being that integrating denominator, among infinitely many, for an element of heat,  $dQ^0$ , in a *reversible process*<sup>†</sup> which turns the latter into the perfect differential of entropy

$$dS = dQ^0/T \text{ (Carnot's theorem)} \quad (1)$$

The thermodynamic temperature,  $T$ , is a unique function of any empirical temperature,  $t$ .

(c) There exists a quantity called entropy production,  $\theta$ , which is positive in any irreversible process. In an adiabatic irreversible process between an initial state 1 and a final state 2, we define

$$\theta = S_2 - S_1 \quad (2a)$$

and must have

$$\theta > 0 \quad (2b)$$

---

<sup>†</sup> All symbols with the superscript <sup>0</sup> refer to reversible processes.

whereas in any quasistatic irreversible process we must find that

$$dS - dQ/T = d\theta \text{ with } d\theta > 0 \quad (3)$$

To the preceding three requirements one may add the pedagogical *desideratum* that the plan of derivation should be as close as possible in spirit and in the basic appeal to experiment (or intuition) to that of the first law. The present article undertakes to sketch a development of this kind for which the claim is made that it is easily grasped by beginners.

## 2. RECENT WORK WITH SIMILAR MOTIVATION

A similar concern is evidenced in the articles by L. A. Turner<sup>1</sup>, P. T. Landsberg<sup>2-4</sup>, F. W. Sears<sup>5</sup>, and M. W. Zemansky<sup>6</sup>, as well in the latter's recent book<sup>7</sup>. It may even be said to go back to M. Born<sup>8</sup>. In particular, M. W. Zemansky<sup>6,7</sup> ably proceeds to simplify the mathematical apparatus needed in the development, thus considerably reducing the amount of prior preparation required of the student. Questions of mathematical rigour which must be answered in this connection, and which are evaded here owing to present intent, have been investigated, and thoroughly answered by P. T. Landsberg<sup>2-4,9</sup>.

## 3. METHODOLOGY

A review of standard textbooks reveals that there exist two fully equivalent<sup>2-4</sup> and yet pedagogically divergent ways of leading the student to the three conclusions. One stems from R. Clausius and Lord Kelvin, the other from C. Carathéodory and M. Born†. Broadly speaking, the first stream makes the statement that a selected irreversible process *is* irreversible, and develops the theory from a particular case by a discussion of reversible and irreversible cycles. The common objection to this development is a sense of artificiality and the impression of an unmotivated *ad hoc* reasoning given by it to a beginning student. The second method starts with an abstract, common characteristic of all irreversible processes, and derives the same three statements as a result of Carathéodory's mathematical theorem. The objection to this development turns on the fact that the theorem is not normally expounded in courses in mathematics, and that the need to grasp it diverts the student's attention from physics to mathematics.

P. T. Landsberg<sup>3</sup> and, later, M. W. Zemansky<sup>6</sup> achieved a 'reconciliation' of the two streams of thought, and the object here is to suggest a further simplification as well as a closer link with the development of the first law. Thus, in addition to statements (a) to (c) above, we must also show that

(d) There exists a universal function for all systems, called their energy,  $E$ , which has the mathematical properties of a potential.

For the sake of completeness, we must also mention the so-called postulational method which *starts* with the equivalents of statements (a) to (c). The common pedagogical objection to this mode of exposition is that it expects the student to accept statements which are alien to him without first creating an adequate intuitional and physical foundation.

---

† For a parallel exposition of these two streams, the reader may consult Chapters 9 and 10 in ref. 10.

#### 4. EXPERIMENTAL BASIS

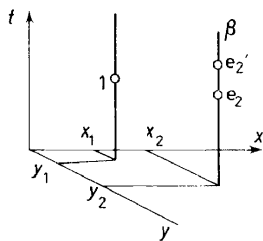
In order to provide an easy intuitive grip on the subject we propose to root the exploration in a single experiment, the famous experiment performed by J. P. Joule. The result of this experiment can be expressed in precise thermodynamic terms as follows:

- A. *Given two arbitrary states of equilibrium 1 and 2 of any closed system it is either possible to reach state 2 from state 1 or state 1 from state 2, but not both, by an adiabatic process involving the performance of work only. The work so performed is independent of the details of the process.*
- B. *If the process is performed at constant volume in a simple system or, generally, with constant initial and final values of the deformation coordinates, work must be performed on the system. (Such a process cannot be carried out without work or in a manner to produce work.)*

#### 5. THE FIRST LAW

We consider a space of the  $n$  independent thermodynamic properties,  $x_1, \dots, x_m$ , of a system. For ease of illustration, we assume in *Figure 1* that  $x, y$  are deformation coordinates and that the third coordinate is the empirical temperature,  $t$ . We now centre attention on an arbitrary state 1 and say, by *definition*, that any state 2 which can be reached from 1 or from which state 1 can be reached adiabatically without the performance of work<sup>†</sup> ( $W_{12} = 0$ ), is called isoenergetic with it. For definiteness, we shall assume that the natural direction of all processes considered henceforth is  $1 \rightarrow 2$ .

We now examine all states for which the deformation variables have given values  $x = x_2$  and  $y = y_2$ ; they lie on the vertical line  $\beta$ . It follows immediately from statement A that there exists only one state 2 on  $\beta$ , denoted by  $e_2$  in *Figure 1*, which is isoenergetic with state 1. If a second such state existed,



*Figure 1.* Uniqueness of isoenergetic point.

say at  $e_2'$ , it is clear from statement B that process  $e_2 \rightarrow e_2'$  or  $e_2' \rightarrow e_2$  would require the performance of work. For the sake of being definite, suppose that negative work is associated with process  $e_2 \rightarrow e_2'$ . It follows that process  $1 \rightarrow e_2 \rightarrow e_2'$  would require the performance of work. Therefore, state  $e_2$  would not be isoenergetic with 1. By continuity<sup>‡</sup>, we now reach the conclusion that the locus of all states which are isoenergetic with an arbitrary

<sup>†</sup> We follow the convention that the work performed on the system is negative and that performed by the system is positive.

<sup>‡</sup> The full implications of the assumption of continuity—understandably evaded in an elementary exposition—are treated rigorously in refs. 2, 3, 4 and 9. Ref. 2 examines this problem in depth.

state 1 form a surface or, more precisely, a hypersurface of  $n - 1$  dimensions in the space of states of  $n$  dimensions.

Varying the temperature of state 1 along the line  $\alpha$  for which  $x = x_1$  and  $y = y_1$ , we can classify all such states according to the quantity of work required to reach them adiabatically from 1 or, for negative work, according to the work required to reach state 1 from them. With each such state,  $1', 1'', \dots$ , there is associated an isoenergetic hypersurface of  $n - 1$  dimensions, *Figure 2*.

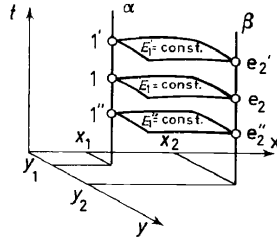


Figure 2. Surfaces of constant energy.

The preceding argument proves the existence of a potential function for any closed thermodynamic system which we can define as

$$E(t, x_1, \dots, x_{n-1}) - E(t^*, x_1^*, \dots, x_{n-1}^*) = -W_{ad} \quad (4)$$

This proves statement (d). Here the parameters  $t^*, x_1^*, \dots, x_{n-1}^*$ ,  $n$  in all, describe an arbitrary reference state, and  $W_{ad}$  is the work needed to perform the adiabatic process to (or from) the current state from (or to) the reference state. This latter is always possible, as asserted by postulate A.

The existence of a single point on a line of constant values of the deformation coordinates which is isoenergetic to a given state proves that surfaces of constant energy cannot intersect.

The generalizations

$$Q_{12} = E_2 - E_1 + W_{12} \quad (5)$$

$$dQ = dE + dW \quad (5a)$$

to non-adiabatic general and non-adiabatic quasi-static processes, respectively, are standard and require no further comment.

## 6. THE SECOND LAW

Having acquired the concept of energy, we can establish an equivalent formulation of postulate B in terms of it:

*B'. It is impossible adiabatically to reduce the energy of a system when its deformation variables retain constant values.*

The equivalence follows at once from equation 4 applied to two states for which the deformation coordinates have equal values.

## 7. CARNOT'S THEOREM

In order to prove statement (a), we apply statement B' to a reversible process (first part of the second law) for which

$$dQ^0 = dE + dW^0 \quad (6)$$

and *define* any state 2, denoted by  $s_2$  in *Figure 3*, which can be reached from a given state 1 reversibly and adiabatically as *isentropic* with it. It is now easy to show, by an argument modelled on the one used earlier in conjunction with *Figure 1*, that there exists only one isentropic point on a given line  $\beta$ . Before we do this, however, it is necessary to point out to a beginner that an isentropic point  $s_2$  is different from an isoenergetic point  $e_2$ . They are both reached by adiabatic processes, a reversible process now and an irreversible process before. However, the work is zero for an isoenergetic point, being different from zero, as seen from equation 6, for an isentropic point.

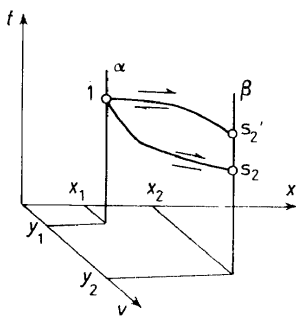


Figure 3. Uniqueness of isentropic point.

Referring to *Figure 3*, we suppose that states  $s_2$  and  $s'_2$  are both isentropic with state 1. Further, for the sake of being definite, we suppose that

$$E(s'_2) > E(s_2) \quad (7)$$

It is now clear that the system would be capable of performing some adiabatic reversible process as well as its reverse process, both symbolized by the full lines in the diagram. In the first case the energy of the system would increase at constant values of the deformation coordinates. However, in the second case, its energy would decrease with  $x_2$  and  $y_2$  reverting to their original values in contradiction to statement B'. Owing to the assumption of reversibility, the contradiction can be removed only by recognizing that states  $s_2$  and  $s'_2$  must be identical. Again, by continuity<sup>2-4,9</sup>, it follows that with any state 1, 1', 1'', ... along  $\alpha$  we may associate a coherent hypersurface, *Figure 4*. The set of such hypersurfaces defines the potential. The resulting family must consist of non-intersecting hypersurfaces, because no point 1', 1'', ... on  $\alpha$  can be reached reversibly from point 1 without exchanging heat, as is easy to prove from equation 6 and the definition of energy. Indeed, for such points we must have  $dW^0 = 0$  but  $dE \neq 0$ .

We can call the resulting potential the empirical entropy,  $\sigma$ , and assert the existence of a family of non-intersecting hypersurfaces

$$\sigma = \sigma(t, x_1, \dots, x_{n-1}) = \text{const.} \tag{8}$$

for any system whatsoever, as shown in *Figure 4*.

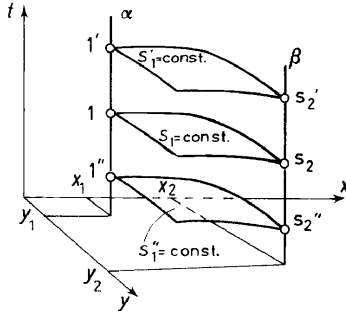


Figure 4. Surfaces of constant entropy.

This family of hypersurfaces intersects the family of hypersurfaces  $E(t, x_1, \dots, x_{n-1})$  from equation 4 along entities of  $n - 2$  dimensions, proving the existence of reversible isothermal-adiabatic processes<sup>6</sup>, as well as of intersecting isentropic lines whenever the number of independent variables,  $n$ , equals or exceeds three.

In the preceding derivation, unlike those in some textbook presentations, we expressly refrained from making an appeal to the statement that it is impossible to design a cycle which consists of two isentropics and one isothermal. Such cycles are possible if  $n \geq 3$ , as shown elsewhere<sup>6</sup>.

To complete the argument, it is now necessary to show that the existence of non-intersecting isentropic surfaces,  $\sigma$ , leads to statement (b) above. The proof can be modelled on that of Carathéodory, and we refrain from giving the details, because a simple version can be found in the literature<sup>3-5</sup>. This reasoning leads naturally to equations 1 and 2.

## 8. THE SECOND PART OF THE SECOND LAW

Statement (c), or the second part of the second law, follows when we extend our inquiry to irreversible adiabatic processes. Thus, we consider an arbitrary adiabatic process which ends at point  $i_2$  in *Figure 5*. We also consider the point  $s_2$  which is isentropic to 1 together with the reversible process. An examination of the combined process (the reverse of which is impossible) in the light of statement B' convinces us that

$$E(i_2) > E(s_2) \tag{9}$$

Reference to equation 1 permits us to integrate for entropy along the reversible path and reference to equation 6 with  $dW^0 = 0$  shows that only positive elements of heat,  $dQ^0 > 0$ , must be summed. This proves that

$$S(i_2) > S(s_2) \tag{10}$$

Generally, we write

$$S_2 - S_1 > 0 \quad (10a)$$

and define the entropy produced as in equation 2a, so that equation 2b follows. This is the principle of entropy increase for adiabatic processes. The generalization to equation 3 is again standard. It suffices to note that  $-dQ/T$  is the change in the entropy of the immediate surroundings, so that  $dS - dQ/T$  is the total change in the entropy of an adiabatic system consisting of the system proper coupled with its immediate surroundings.

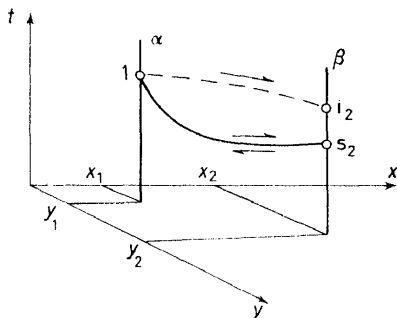


Figure 5. Characteristics of irreversible adiabatic process.

Finally, we note that equation 9 combined with equation 5 and the condition that  $Q_{12} = 0$  proves that of all adiabatic processes which occur between a given initial state and prescribed values of the deformation coordinates at the final state, an isentropic process yields the maximum positive (or minimum negative) work. Indeed, along  $1 \rightarrow s_2$

$$W_{12}^0 = E_1 - E(s_2) \quad (11a)$$

whereas along an arbitrary adiabatic irreversible process for the same  $x_2, y_2$ , we have

$$W_{12} = E_1 - E(i_2) \quad (11b)$$

Reference to equation 9 shows that

$$W_{12}^0 > W_{12} \quad (12)$$

## REFERENCES

- <sup>1</sup> L. A. Turner, *Am. J. Phys.* **28**, 781 (1960); **29**, 40 (1961); **30**, 506 (1962).
- <sup>2</sup> P. T. Landsberg, *Physica Status Solidi*, **1**, 120 (1961).
- <sup>3</sup> P. T. Landsberg, *Nature, London*, **201**, 485 (1964).
- <sup>4</sup> P. T. Landsberg, *Bull. Inst. Phys. and Phys. Soc.* 150 (1964).
- <sup>5</sup> F. W. Sears, *Am. J. Phys.* **31**, 747 (1963); **34**, 665 (1966).
- <sup>6</sup> M. W. Zemansky, *Am. J. Phys.* **34**, 914 (1966).
- <sup>7</sup> M. W. Zemansky, *Heat and Thermodynamics*, Chapter 8, 5th ed. McGraw-Hill: New York (1968).
- <sup>8</sup> M. Born, *Natural Philosophy of Cause and Chance*. Clarendon Press: Oxford (1949).

- <sup>9</sup> P. T. Landsberg, *Thermodynamics with Quantum Statistical Illustrations*. Interscience : New York (1961).
- <sup>10</sup> J. Kestin, *A Course in Thermodynamics*, Vol. I. Blaisdell : New York (1966).
- <sup>11</sup> J. Kestin, *Am. J. Phys.* **29**, 329 (1961).