AFOSR-TN-60-23

THE THERMODYNAMICS OF PHASE EQUILIBRIUM

LASZLO TISZA

TECHNICAL REPORT 360

FEBRUARY 26, 1960

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
RESEARCH LABORATORY OF ELECTRONICS
CAMBRIDGE, MASSACHUSETTS

The Research Laboratory of Electronics is an interdepartmental laboratory of the Department of Electrical Engineering and the Department of Physics.

The research reported in this document was made possible in part by support extended the Massachusetts Institute of Technology, Research Laboratory of Electronics, jointly by the U. S. Army (Signal Corps), the U. S. Navy (Office of Naval Research), and the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), under Signal Corps Contract DA36-039-sc-78108, Department of the Army Task 3-99-20-001 and Project 3-99-00-000; and in part by the U. S. Air Force (Office of Scientific Research, Air Research and Development Command) under Contract AF49(638)-95.

Reprinted from ANNALS OF PHYSICS, Volume 13, No. 1, April 1961
Copyright © by Academic Press Inc. *Printed in U.S.A.*

ANNALS OF PHYSICS: **13,** 1–92 (1961)

The Thermodynamics of Phase Equilibrium*

LASZLO TISZA

Department of Physics, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

Thermodynamics is usually subdivided into a theory dealing with equilibrium and into one concerned with irreversible processes. In the present paper this subdivision is carried further and the Gibbsian thermodynamics of phase equilibrium is distinguished from the thermodynamics of Clausius and Kelvin. The latter was put into an axiomatic form by Carathéodory; the present paper attempts a similar task for the Gibbs theory. The formulation of this theory as an autonomous logical structure reveals characteristic aspects that were not evident until the two logical structures were differentiated. The analysis of the basic assumptions of the Gibbs theory allows the identification and removal of defects that marred the classical formulation. In the new theory thermodynamic systems are defined as conjunctions of spatially disjoint volume elements (subsystems), each of which is characterized by a set of additive conserved quantities (invariants): the internal energy, and the mole numbers of the independent chemical components. For the basic theory, it is convenient to assume the absence of elastic, electric, and magnetic effects. This restriction enables us to define thermodynamic processes as transfers of additive invariants between subsystems. Following Gibbs, we postulate that all thermostatic properties of system are contained in a fundamental equation representing the entropy as a function of the additive invariants. Geometrically, this equation is represented as a surface in a space to which we refer as Gibbs space. In order to make the information contained in the fundamental equation complete, we have to use, in many cases, additional quasi-thermodynamic variables to specify the intrinsic symmetry properties of the system. The walls, or boundaries limiting thermodynamic systems are assumed to be restrictive or nonrestrictive with respect to the transfer of the various invariants. The manipulations of the boundary conditions (imposition and relaxation of constraints) are called thermodynamic operations. In systems with nonrestrictive internal boundaries, the constraints are consistent with infinitely many distributions of the invariants over the subsystems. These virtual states serve as comparison states for the entropy maximum principle. This principle allows us to identify the state of thermodynamic equilibrium, attained asymptotically by real systems. The thermostatic ex-

---

* This research was supported in part by the U. S. Army Signal Corps, the Air Force Office of Scientific Research, and the Office of Naval Research; and in part by the U. S. Air Force (Office of Scientific Research, Air Research and Development Command) under Contract AF49(638)-95. Reproduction in whole or in part is permitted for any purpose of the United States Government.

tremum principle is the basis of a theory of stability. Stability may be normal or critical. In the latter case, the compliance coefficients (specific heat c_p, expansion coefficient, and isothermal compressibility) tend, in general, to infinity. The phases of thermodynamic systems are each represented by a primitive surface in Gibbs space, the points of these surfaces correspond to modifications of the phase. The actual distribution of the invariants of a system over phases is determined by the entropy maximum principle. Among the fundamental theorems are the two phase rules. The first rule specifies the dimension of the set of points in the space of intensities, in which a given number of modifications can coexist. The second rule specifies the dimensionality of the set of critical points. The phase rules are somewhat more general than those of Gibbs because of our use of symmetry considerations. At a critical point two modifications become identical, and we obtain critical points of two kinds: (i) the modifications differ in densities, as for liquid and vapor; (ii) the modifications differ in symmetry, as the two directions of ordering in the Ising model. The second case relates to the well-known λ-points and λ-lines in the $p - T$ diagram, hence these phenomena fit into the framework of the theory without *ad hoc* assumptions. For the case of liquid helium, this interpretation requires that the superfluid ground stage be degenerate. This conclusion is not inconsistent with the third law, but it requires substantiation by quantum-mechanical methods. Another fundamental theorem is the principle of thermostatic determinism: a reservoir of given intensities determines the densities (energy and components per unit volume) of a system which is in equilibrium with it and, conversely, the densities of the system determine the intensities. The mutual determination is unique, except at critical points and at absolute zero. A more satisfactory description of these singular situations calls for the use of statistical methods. The present approach leads to several statistical theories, the simplest of which is developed in the second paper of this series.

INTRODUCTION

This paper is the first in a series in which thermodynamics is developed beyond its usual scope from a new postulational basis.

The postulational basis of classical thermodynamics is firmly established in tradition and a new departure calls for an explanation of the underlying ideas.

It is widely believed that thermodynamics consists essentially of the implications of the first, second, and third law of thermodynamics. Actually, however, few if any significant results can be derived from these postulates without using a number of additional assumptions concerning the properties of material systems, such as the existence of homogeneous phases, validity of equations of state and the like.

The three laws of thermodynamics are presumably of universal validity, a point that is emphasized by their formulation as impossibility axioms. In contrast, most of the additional assumptions involve approximations of a more or less restrictive character. In the traditional presentations these additional assumptions and their limitations receive but scant attention. Therefore the rigor of the theory that the painstaking establishment of the universal principles was meant to ensure is considerably impaired.

The nature of a theory is determined by its postulational basis and it is to be expected that the reformulation of the basic postulates of thermodynamics will have a bearing on the type of thermodynamic theory that will be obtained.

Within judiciously chosen limitations, a set of basic assumptions can be expressed in terms precise enough to generate a mathematical theory with desirable properties. Some of the limitations of the theory can be overcome by the formulation of sets of successively more refined assumptions generating theories that are adapted to deal with the more exacting requirements. Thus thermodynamics becomes a master scheme consisting of a number of closely knit deductive systems devised for different types of situations. This procedure is not only more rigorous than the classical one, but opens up new areas for investigation where the possibilities of the classical theory seemed already exhausted.

An important step leading toward such a pluralistic conception of thermodynamics was the establishment of irreversible thermodynamics as an autonomous discipline (*1*). The classical formalism is recognized to be valid only for equilibrium situations.

We wish to make the further point that even the classical theory of equilibrium is a blend of two essentially different logical structures: we shall distinguish the theory of Clausius and Kelvin, on the one hand, from that of Gibbs, on the other.

In the first of these theories the thermodynamic system is considered as a "black box"; all relevant information is derived from the amount of energy absorbed or provided by idealized auxiliary devices such as heat and work reservoirs coupled to the system. The main achievement of this theory is the establishment of the concepts of internal energy and entropy from observable quantities.

In the Gibbs theory (*2*) attention is turned toward the system. The concepts of internal energy and entropy are taken for granted, and are used to provide a more detailed description of the system in equilibrium, which includes its chemical and phase structure.

The logical-mathematical structure of the Clausius-Kelvin theory received the most satisfactory expression in the axiomatic investigation of Carathéodory (*3*). We shall speak of the Clausius–Kelvin–Carathéodory (CKC) theory.

No comparable axiomatic investigation of the Gibbs theory has been performed thus far. Accordingly, in the existing texts the CKC and the Gibbsian thermodynamics are intricately interwoven, although the Gibbs formalism is emphasized in texts oriented towards physical-chemical applications.

The first objective of this paper is to formulate a postulational basis from which a theory, almost identical to that of Gibbs, is derived as a self-contained logical structure. We shall call this theory thermostatics, or alternately the macroscopic thermodynamics of equilibrium (MTE).[1]

In order to obtain a preliminary impression of the features in which this modernized Gibbs theory differs from that of CKC, it is convenient to compare the geometrical methods utilized by each.

The application of geometrical methods to thermodynamics is based on a thermodynamic *phase space*, i.e., a space spanned by a number of thermodynamic variables. In the CKC theory one considers a thermodynamic phase space spanned by such variables as pressure, volume, and the mole numbers of the chemical components. The guiding idea is that the variables be directly measurable. Beyond this requirement, however, the theory is quite insensitive to the specific choice of phase space. In particular, there is no geometrical significance attached to the distinction between extensive and intensive variables, as indicated by the use of volume and pressure in an interchangeable manner.[2]

In contrast, in the Gibbs theory a particular phase space spanned by the extensive variables: energy, entropy, volume, and mole numbers plays a quite unique role. In order to underline its importance we shall refer to it as *Gibbs space*.[3]

It was a remarkable discovery of Gibbs that a single surface in this space (the primitive surface) provides a compact representation of all thermostatic properties of a phase. Phases are homogeneous extensions of matter that constitute the building blocks for more complex, heterogeneous systems. Moreover, Gibbs based his theory of thermodynamic stability on the analysis of the curvature of the primitive surface.

From the mathematical point of view, the success of this procedure is somewhat surprising because such elementary geometrical concepts as orthogonality and metric, which are necessary both in the elementary and the Riemannian theory of curvature, are lacking in the thermodynamic phase spaces.

Nevertheless Gibbs, acting apparently on physical intuition alone, developed an *ad hoc* theory of curvature, the exact mathematical foundation for which was discovered nearly half a century later by Pick and Blaschke (*7*). This somewhat esoteric theory operates in terms of affine differential geometry in which parallel projection is substituted for the more common orthogonal projection.

---

[1] We use capitalized abbreviations to denote theories with explicitly stated postulational bases, whereas "thermodynamics" is used in a more vague, traditional sense.

[2] The geometrization of thermodynamics along these lines has been further developed by Landsberg (*5*).

[3] It has already been recognized by Ehrenfest (*6*) that the precise distinction between extensive and intensive variables calls for an axiomatic investigation going beyond that of Carathéodory.

The Gibbs space alone, of all thermodynamic phase spaces can be considered as an affine space. Thus the idea of attaching a special importance to Gibbs space is supported by the fact that in this space we can establish a far reaching geometrization of the properties of matter.

In this paper we shall confine ourselves to a few remarks concerning the role of affine geometry in thermostatics, and hope to return to this question in the future.[4]

The formulation of the Gibbs theory of phase equilibrium, as an autonomous logical structure, gives justice to aspects of the theory that are obscured when embedded in the CKC theory. At the same time the analysis of the basic assumptions has also brought to light significant defects in the classical theory.

The most important of these defects concerns the question of identity and distinctness of two phases. It is evidently important to establish the criteria by which we decide whether or not a surface can be considered a phase boundary separating two distinct phases. It is the implicit assumption of the classical theory that two volume elements contain identical phases if and only if they have identical chemical compositions and identical densities of the extensive quantities. Actually, however, volume elements of identical composition and density may still be distinct because of their symmetry properties. Situations

---

[4] The following remarks are for preliminary orientation of the reader who is familiar with differential geometry, we shall not make use of them in the present paper: The elementary theory of curvature of a surface embedded in three-dimensional Euclidian space is based on the two fundamental forms: the first of these (I) determines the line element in terms of the parameters describing the surface; the second form (II) expresses the curvature with respect to the embedding space. Both forms are needed to determine the principal radii of curvature, R_1 and R_2. However, according to a fundamental theorem of Gauss, the product R_1R_2 depends solely on form I, it is invariant with respect to bending of the surface, an operation which preserves the metric I, but changes the embedding specified by II. This theorem is the basis for the development of the Riemannian theory of curvature solely in terms of the intrinsic metric form I of the surface. Neither the elementary, nor the Riemannian theory of curvature can be applied in Gibbs space, in which no physically meaningful metric form is definable. However, in 1917, Pick and Blaschke developed a theory of curvature, within affine differential geometry, based solely on form II. This form is called the affine fundamental form; it is invariant under linear transformations of determinant unity (equiaffine invariance). We shall see, particularly in Section VIII that there is a quadratic form, we call it the fundamental thermostatic form, that plays an important role in the thermostatic theory of stability. This form can be shown to be an affine fundamental form of the primitive surface in Gibbs space. Thus the parallel between the significant thermostatic and geometric concepts is very close. Although the Riemannian and the affine theories of curvature have similar traits, there are also differences. Without entering into geometrical intricacies, Appendix B contains a short discussion of the reduction of a quadratic form in affine space. The treatment and its physical interpretation differ considerably from the eigenvalue method that is customary in spaces in which orthogonal (or unitary) transformations are meaningful.

of this sort arise if there exist two phases that transform into each other by means of inversion or mirroring, as in the case of right and left quartz. Such a symmetry relation ensures the distinctness of the two phases and the exact identity of their densities as well. If the phases are equivalent under a pure rotation, we should not consider them as distinct.

The redefinition of the criterion of distinctness has a significant implication for the phase rule. While the possible types of phase equilibrium predicted by the classical theory remain unchanged, the theory is enriched by a new category: the λ-points and λ-lines arising in order-disorder phenomena. The experimental occurrence of these phenomena was a serious embarrassment to the classical theory. This gap was meant to be filled by the *ad hoc* assumption by Keesom (*8*) and Ehrenfest (*9*) of the concept of higher order transitions.

This scheme has met with wide acceptance in the systematization of experimental facts, although it led to a number of difficulties which could only be removed by making further assumptions of a purely taxonomical nature (*10*).

It is therefore satisfying that the experimental facts in question fit into the present theory without *ad hoc* assumptions.[5] At least this is the case for λ-phenomena in crystals. The λ-anomaly of liquid helium falls into a special class. It fits into the present theory only if the ground-state wave function of the superfluid is assumed to be at least doubly degenerate. The present theory is too phenomenological to provide a substantiation of this conclusion, and the matter requires further study by quantum mechanical methods.

We have mentioned that the key to the continued development of thermodynamics lies in those assumptions that are usually not explicitly mentioned. Among the most significant of these is the assumption of a *constructive principle*.

In the CKC theory systems are considered as undivided units coupled only with auxiliary macroscopic devices, in strong contrast to the constructive principle of statistical mechanics, in which systems are built from constituent invariant particles.

We shall see that thermodynamics has a constructive principle of its own: systems are built up from disjoint volume elements that are described in terms of extensive variables specifying the amount of energy, mass, and chemical components within each volume element.

In MTE these quantities are always assumed to have reached their equilibrium values. However, it is possible to take fluctuations into consideration, and let the extensive variables become random variables. Thus thermodynamics becomes an essentially statistical discipline and the phenomenological formulation appears only as a limiting case.

[5] Some of the ideas concerning the expansion of the classical theory have been reported before (*11–13*). We note also that Landau (Chapter XIV of Ref. *20*) was the first to recognize the connection between symmetry and the λ-lines of crystalline substances. However, we arrive at different conclusions concerning the mathematical character of the singularity.

The simplest instance of statistical thermodynamics will be developed in the second paper of this series (*14*). We shall refer to it as the statistical thermodynamics of equilibrium (STE).

I. DEFINITIONS AND POSTULATES

We now proceed to list the basic definitions and postulates, the *basis* of thermostatics (MTE), amplified by short comments that are required for the clarification of the formal statements. A more extensive discussion of the limits of validity of the postulates, and comparison of the present procedures with traditional ones, follows in the next section.

The postulational basis has a hypothetical character justified by the merits of the theory generated by it. The measure of experimental agreement achieved will be assessed in the final discussion.

A theoretical justification of the postulational basis can be obtained also by reducing the basis to more fundamental theories. Within the context of the present over-all approach this is a program of many stages and is beyond the scope of this paper.

We emphasize that there is a considerable latitude in the formulation of the postulational basis even within thermostatics. The following basis was chosen, partly in view of developing the theory of phase equilibrium, and partly in order to facilitate generalization to the statistical thermodynamics and the quantum mechanics of phase theory.

A. Definitions and Postulates Concerning the Independent Variables of Thermostatics

D a1 *Thermodynamic Simple System:* A finite region in space specified by a set of variables $X_1, X_2, \cdots, X_{r+1}$. The X_i symbolize the physical quantities listed in *P a3*.[6]

D a2 *Composite system:* A conjunction of spatially disjoint simple systems. A composite system is obtained by uniting separate systems or by partitioning a single system. The simple systems that form the composite systems are called its subsystems. The subsystems and their properties will be denoted by primes or superscripts, thus X_i' or $X_i^{(a)}$ denotes the amount of the quantity X_i in the subsystem a. The set of quantities $X_i^{(a)}$ specifies the state of the composite system. Geometrically, it is a point in the, say M-dimensional vector space spanned by the variables $X_i^{(a)}$, $i = 1, 2, \cdots r + 1$; $a = 1, 2, \cdots$.

D a3 *Transfer quantity* X_i^{ab}*:* The amount of X_i transferred from the system a to b during an arbitrary, fixed time interval. Such a transfer is called a *thermodynamic process.*

---

[6] *D a1*, *P a1*, and *C a1* stand for Definition *a1*, Postulate *a1*, and Corollary *a1*, respectively.

P a1 *Conservation postulate:* The quantities represented by the variables X_i are conserved; they obey the continuity equations

$$\Delta X_i^{(a)} + \sum_b X_i^{ab} = 0, \tag{1.1}$$

$i = 1, 2, \cdots, r + 1$, $a = 1, 2, \cdots$, where $\Delta X_i^{(a)}$ is the net increase of the quantity X_i in the system a, and the summation is taken over all systems involved in the exchange of X_i. Both $\Delta X_i^{(a)}$ and X_i^{ab} refer to the same time intervals.

P a2 *Additivity postulate:* In a composite system

$$X_i = \sum_a X_i^{(a)}, \qquad i = 1, 2, \cdots, r + 1, \tag{1.2}$$

where the summation is over the subsystems. The quantity on the left belongs to the composite system.

D a4 *Additive invariant:* a quantity satisfying *P a1* and *P a2*.

D a5 *Wall:* physical system idealized as a surface forming the common boundary of two systems, say a and b, and preventing the transfer of some of the quantities X_k, so that $X_k^{ab} = 0$ regardless of the state or properties of the systems a and b. No such restrictions are valid for the other quantities X_j. The wall is said to be *restrictive of the* X_k and *nonrestrictive of the* X_j.[7] Walls separating two subsystems of a composite system are called *partitions*: those completely enclosing a system are *enclosures*. The set of walls in a system is referred to as the *constraints*. These are said to exert *passive forces* on the distribution of the additive invariants.

D a6 *Thermodynamic Operation:* any imposition or relaxation of a constraint through the uniting or subdividing of systems, or the altering of the type of the enclosures or partitions.

D a7 *Coupled Systems:* systems for which at least one transfer quantity X_i^{ab} can be different from zero in the presence of the existing constraints. The systems are said to be coupled by means of X_i exchange.

C a1 It follows from *P a1* and *P a2* that for a system a in an enclosure restrictive of X_i, the value $X_i^{(a)}$ is constant in time. Hence the conservation laws and the restrictive enclosures impose linear constraints on the variables $X_i^{(a)}$:

$$\sum_a X_i^{(a)} = \xi_i^{(k)} = \text{constant}. \tag{1.3}$$

The summation is taken over the subsystems of an enclosure restrictive of X_i.

---

[7] These expressions are somewhat elliptic. We should say: "restrictive of the transfer quantity associated with X_k, $\cdots$."

The superscript k identifies the enclosure that may coincide with the entire system or with parts of it. The $\xi_i^{(k)}$ are invariants of the composite system.

The conditions (3) may be supplemented by inequalities, e.g., requiring that some of the X_i be positive.

Let us assume that the composite system has b linearly independent invariants, relabeled with single subscripts $\xi_1, \xi_2 \cdots \xi_b$. Equation (3) can be used to express b of the variables $X_i^{(a)}$ in terms of the ξ_j and the remaining $M - b = f$ variables $X_i^{(a)}$ which we relabel $Y_1, Y_2, \cdots Y_f$. It is convenient to substitute the variables ξ_k, Y_i for the original $X_i^{(a)}$, and we define:

D a8 *Fixed variables:* the variables $\xi_1, \xi_2, \cdots, \xi_b$ appearing in Eq. (3); invariants of the composite system.

D a9 *Free variables:* the variables $Y_1, Y_2, \cdots, Y_f$ obtained by the foregoing construction. Within an indicated range the Y_i specify the *virtual states* of the composite system, i.e., those states that are consistent with the constraints, and with the set of fixed variables ξ_j.

D a10 *Virtual process (displacement):* the increments ΔY_i leading from one virtual state to another. Geometrically, the ratio $\Delta Y_1 : \Delta Y_2 : \cdots$ represents a line in the vector space spanned by the $X_i^{(a)}$.

C a2 The linear combination of two virtual displacements is, again, a virtual displacement. The virtual displacements form a linear manifold specified by conditions of the form

$$\sum_a \Delta X_i^{(a)} = 0. \tag{1.4}$$

The range of summation is the same as in (3).

We consider some important special cases. If a composite system consists of isolated subsystems, all variables $X_i^{(a)}$ are fixed, the space of virtual states reduces to a point.

In the case of two subsystems coupled by the transfer of the positive quantity X_i, we have

$$\xi_1 = X_1^{(a)} + X_1^{(b)}, \tag{1.5}$$

and

$$0 \leqq Y_1 = X_1^{(a)} \leqq \xi_1 . \tag{1.5b}$$

The virtual displacement is

$$\Delta Y_1 = \Delta X_1^{(a)} = -\Delta X_1^{(b)} = X_1^{ba}. \tag{1.5c}$$

Note that the displacement uniquely determines a transfer quantity. This is no longer the case for three coupled systems. In fact, the transfer

$$X^{ab} = X^{bc} = X^{ca} \tag{1.6}$$

produces no displacement: $\Delta X^a = \Delta X^b = \Delta X^c = 0$, hence (6) may be added to any set of transfer quantities without changing the displacement.

Up to this point, the variables X_i have been defined only as additive invariants (*P a1*, *P a2*, *Da 4*). It is indeed possible to develop the formalism of MTE on the basis of such an *implicit* definition by postulation. However, the physical interpretation of the theory calls for an *explicit* interpretation of the variables.

P a3 The variables $X_1, X_2, \cdots, X_{r+1}$ are interpreted as the volume V, the internal energy U and the mole numbers of the independent components $N_1, N_2, \cdots, N_c$; hence

$$r = c + 1. \tag{1.7}$$

The joint use of *P a1*, *P a2* with *P a3* gives rise to new sorts of questions: Are the variables enumerated in *P a3* actually additive invariants? Are there actual walls that are restrictive or nonrestrictive of various combinations of these variables? Neither of these questions is answered by an unqualified "yes," and the theory to be developed is subject to limitations that are discussed in Section II. Here we confine ourselves to a short summary of the conclusions reached.

The conservation of internal energy is postulated in the first law of thermodynamics and is sufficiently well substantiated. The additivity of the energy is, in general, not a valid statement, but it represents a sufficiently good approximation if long-range gravitational, electric or magnetic fields and surface effects are only of negligible importance. The thermodynamic theories of electric, magnetic, elastic, gravitational and capillary systems call for special considerations that are not included in this paper.

In mechanics and electrodynamics the non-additive interaction energies are essential to provide a coupling of systems, but in thermodynamics coupling arises even in the case of strict additivity through the exchange of additive invariants.

The material content of thermodynamic systems appears as a collection of particles (molecules, atoms, ions, nuclei, etc.) of definite identity. The numbers of these particles, or the mole numbers of the corresponding chemical species are additive, but not conserved, since they are subject to change due to chemical reactions. This complication can be easily handled, provided that the relevant reactions can be classified into slow and fast ones, in such a way that the advancement of the former is negligible on the time scale of the actual experiment, while the latter can be assumed to have come to a standstill at chemical equilibrium. Under such conditions it is possible to define additive invariants, the so-called *mole numbers* of the (*independent chemical*) *components* (p. 63 of Ref. *2*). The specification of these numbers will be called the *composition*.[8]

---

[8] In this paper we do not consider chemical reactions explicitly and therefore we do not go into a detailed definition of independent components. In the absence of reactions, the number of indestructible particles of each species are additive invariants, and can be used

Although we do not enter into an explicit discussion of chemical reactions, we may note that their existence has an implicit bearing on the scope of the thermostatic formalism. The rates of reactions can be speeded up, or slowed down, to an extreme degree by the variation of the conditions. One can vary the temperature and the pressure, and add, or remove catalysts and anticatalysts. On changing the set of relavant reactions, the number of distinct independent components, used according to *P a3* for the description of the system, will change as well. Therefore, we state:

C a3 The choice of thermodynamic variables and the resulting formalism is relative to the set of chemical reactions, that have been used to define the independent components. The formalism is only adequate under conditions at which these, and only these, reactions proceed at a noticeable rate.

The joint use of *P a3* and *D a5* requires special consideration. First, we note that an enclosure may be restrictive of all the *X-s* listed in *P a3* without precluding a change of the shape of the system. Although the present theory is concerned only with shape independent effects, it is desirable to specify whether or not an enclosure is *rigid*, and precludes a change of *shape* of the enclosed system. In fact, we shall find it convenient to assume, for the time being, that all our walls and enclosures are rigid, and that both volume and shape of our systems are constant. This assumption seems drastic, since it excludes macroscopic compressional work from our discussion. (Other types of macroscopic work are excluded in virtue of *P a3*, which does not provide for elastic, electric, and magnetic variables.) We shall consider macroscopic work in a somewhat different context in Section IV. The temporary exclusion of macroscopic work has a number of advantages.

P a4 The following types of rigid (volume and shape preserving) walls and enclosures are actually available:

Impermeable wall: restrictive of matter.
Semi-permeable wall: restrictive of certain chemical species, nonrestrictive of energy.
Permeable wall: nonrestrictive of matter and energy.
Adiabatic wall: restrictive of matter and energy.
Diathermic wall: restrictive of matter, nonrestrictive of energy.

D a11 *Isolated system:* system enclosed in a completely restrictive enclosure.
Closed system: system in an impermeable enclosure.

---

in *P a3*. Part of the usefulness of the concept of components is that it allows us to "save" the simple formalism for inert species, even in the presence of reactions, by establishing a more sophisticated conceptual interpretation of the variables.

Open system: system in which exchange of matter is not precluded by impermeable enclosure.

D a12 *Heat:* transfer quantity, associated with energy, across a rigid diathermic partition (no work):

$$Q^{ba} = U^{ba}. \tag{1.8}$$

B. Definitions and Postulates on Equilibrium and Entropy

P b1 *Entropy maximum principle.* (Composite systems).

Isolated composite systems tend toward a quiescent asymptotic state called thermodynamic equilibrium, in which the free variables assume constant values specified as solutions of an extremum problem. The extremal function is constructed as follows: To each simple system there is assigned an *entropy function,* or *fundamental equation*

$$S^{(a)} = S^{(a)}(X_1^{(a)}, \cdots X_{r+1}^{(a)}), \tag{1.9}$$

a continuous first order-homogeneous function with continuous first, and piecewise continuous higher derivatives. The fundamental equation of the composite system is

$$S(\xi_1, \xi_2, \cdots, \xi_b) = \max\{\sum_a S^{(a)}(X_1^{(a)}, X_2^{(a)}, \cdots, X_{r+1}^{(a)})\}. \tag{1.10}$$

The comparison states of this maximum problem are the virtual states described by the free variables $Y_1, Y_2, \cdots, Y_f$.

We shall see in Section XII that the solution of the maximum problem (10) is unique, under rather general conditions. Moreover, if the conditions of uniqueness are not satisfied, the various solutions are related to each other by rather obvious symmetry relations. We express this by stating that the solution of Eq. (10) is *almost unique.* The precise meaning of this term will be explained later.

C b1 Let $\bar{Y}_1, \bar{Y}_2, \cdots \bar{Y}_f$ be a set of values of the free variables for which expression (10) is a maximum. In virtue of Eq. (10) the $\bar{Y}_j$ are represented as functions of the ξ_k. By using Eq. (3), we obtain the $\bar{X}_i^{(a)}$ corresponding to the $\bar{Y}_i$ and ξ_k. Then

$$S(\xi_1, \xi_2, \cdots \xi_b) = \sum_a S^{(a)}(\bar{X}_1^{(a)}, \bar{X}_2^{(a)}, \cdots, \bar{X}_{r+1}^{(a)}). \tag{1.11}$$

The terms of this sum are the entropy values assigned to each subsystem *only* in the case of equilibrium. The left-hand side of this equation is the entropy of the composite system; thus the entropy is additive.

C b2 Let the entropy of a composite system in equilibrium be S_i.

Consider a thermodynamic operation in which some of the internal constraints of the system are relaxed, and thus the manifold of virtual states is increased. With the increase of the set of comparison states, the maximum (10) either increases or remains unchanged. Denoting the entropy in the relaxed equilibrium state S_f we have

$$S_f > S_i \quad \text{(a)}$$

or

$$S_f = S_i \quad \text{(b)}. \tag{1.12}$$

In Case (a) the relaxation of constraints triggers a process leading to a redistribution of the additive invariants, i.e., to a new equilibrium. This process involves the increase in entropy:

$$DS = S_f - S_i, \tag{1.13}$$

The symbol D indicates an actual change in contrast to virtual changes denoted by Δ.

An imposition of a constraint in a system in equilibrium leaves the distribution of the X_i and hence the entropy, unchanged. In particular, the reimposition of the initial constraint fails to reestablish the initial situation.

In case (b) the relaxation of constraints leads to no process at all.

D b1 *Irreversible operation:* relaxation of constraints producing an entropy increase $DS = S_f - S_i$. This quantity is called the *measure of irreversibility.*

D b2 *Reversible operation:* imposition of constraints in a system in equilibrium, or a relaxation of constraints that does not produce an entropy increase.

D b3 *Path:* Sequence of operations and processes in which a simple system is transferred from an initial to a final state.

An operation is performed only after the equilibrium is attained in the previous step. As a result, the system is transformed from an initial to a final state while the total measure of irreversibility is

$$\sum_{k=1}^{N} DS_k . \tag{1.14}$$

In general, the initial and final states do not uniquely specify the path. In particular, the increase of the number of steps N may lead to a decrease of the measure of irreversibility (14). The path is called reversible if

$$\lim_{N\to\infty} \sum_{k=1}^{N} DS_k = 0. \tag{1.15}$$

We shall show later that reversible paths do exist in this idealized limiting sense.

D b4 *Gibbs space associated with a simple system:* the $(r + 2)$-dimensional space spanned by the variables $X_1, X_2, \cdots, X_{r+1}, S$. The Gibbs space of a composite system is the direct product of the Gibbs space of its subsystems. It is spanned by the variables X_i^a, $i = 1, 2, \cdots$; $a = 1, 2, \cdots$; S^a, $a = 1, 2, \cdots$.

D b5 *Extensive variables:* variables proportional to the size of the system. Primarily, the additive invariants and the entropy.

C. DEFINITIONS AND POSTULATES ON TEMPERATURE

D c1 The *temperature* T is defined as

$$T = (\partial S/\partial U)^{-1}. \tag{1.16}$$

P c1 The *range* of the *temperature* is

$$0 < T < \infty. \tag{1.17}$$

Zero and infinite temperatures are admitted for limiting considerations.

P c2 The entropy of every thermodynamic system in equilibrium vanishes in the limit of vanishing temperatures:

$$\lim_{T\to 0} S = 0. \tag{1.18}$$

P c3 The internal energy has a lower bound that is reached at $T = 0$.

D. DEFINITIONS AND POSTULATES ON PHASE EQUILIBRIUM

P d1 *Phase postulate.*

The structural elements for building up simple systems in equilibrium are spatially homogeneous extensions of matter called *phases*. A system of given composition exists potentially in a number of phases, each of which is specified by a phase-entropy function or a *primitive fundamental equation*

$$\mathrm{S}^{(\alpha)} = \mathrm{S}^{(\alpha)}(X_1^{(\alpha)}, X_2^{(\alpha)} \cdots X_{r+1}^{(\alpha)}; \eta_1^{(\alpha)}, \eta_2^{(\alpha)}, \cdots), \tag{1.19}$$
$$\alpha = 1, 2, \cdots$$

expressing the entropy of the phase α as a continuous first order homogeneous function of the additive invariants, admitting continuous higher derivatives with respect to these variables. The functions (19) represent $(r + 1)$-dimensional hypersurfaces, called *primitive* or *phase* surfaces in Gibbs space. Each function is defined within a characteristic range of additive invariants. (More precisely: in a characteristic range of the densities defined in *D d2*.) In some cases the functions depend also on additional quasi-thermodynamic parameters $\eta_1, \eta_2, \cdots$. These are introduced for homogeneous states of matter that have identical densities and are nevertheless distinguishable by virtue of their intrinsic symmetry properties (e.g., mirror-image states). The parameters are identified as translational invariants in Appendix B. The set of all primitive surfaces contains all of the thermostatic information concerning a system of given composition.

D d1 *Scale factor:* one additive invariant, say X_{r+1}, singled out to specify the size of the system. Unless otherwise stated, the scale factor is chosen as the volume: $X_{r+1} = V$.

D d2 *Densities:*

$$\begin{aligned} x_i^{(\alpha)} &= X_i^{(\alpha)}/X_{r+1}^{(\alpha)}, \qquad i = 1, 2, \cdots r, \\ \mathfrak{s}^{(\alpha)} &= \mathcal{S}^{(\alpha)}/X_{r+1}^{(\alpha)}, \end{aligned} \tag{1.20}$$

where $X_{r+1}^{(\alpha)} = V^{(\alpha)}$. The intrinsic, size-independent properties of a phase are represented by the r-dimensional primitive surface

$$\mathfrak{s}^{(\alpha)} = \mathfrak{s}^{(\alpha)}(x_1^{(\alpha)}, x_2^{(\alpha)}, \cdots x_r^{(\alpha)}; \eta_1^{(\alpha)}, \eta_2^{(\alpha)}, \cdots). \tag{1.21}$$

We shall refer to the space spanned by the variables (20) also as Gibbs space. If necessary, it may be referred to as the $(r + 1)$-dimensional Gibbs space compared with the $(r + 2)$-dimensional space of *D b4*.

It is possible to choose a *mole number* or the *total mass* as a scale factor. This is particularly convenient in one component systems. The quantities (20) then become *molar* or *specific quantities.*

D d3 *Modifications of a phase:* equilibrium states corresponding to different regions of the surface (1.19); a region may shrink to a single point. Two modifications are *distinct* if they have different densities, or if they differ in the parameters η.

Examples: liquid and vapor are two modifications of the fluid phase. Right and left quartz have identical densities and are distinguished only in terms of the parameters η.

Note that in current usage the terms "phase" and "modification" are used as synonyms.

D d4 *Virtual heterogeneous state:* obtained through the decomposition of an isolated simple system into disjoint volume elements under observation of *P a1* and *P a2*:

$$\sum_{\beta} X^{(\beta)} = X_i, \quad i = 1, 2, \cdots, r + 1, \tag{1.22}$$

where each term corresponds, in ordinary space, to a volume element $X_{r+1}^{(\beta)} = V^{(\beta)}$, and in Gibbs space to a point on one of the phase surfaces, i.e., to a modification of the corresponding phase. The modifications in different volume elements are distinct in the sense of *D d3*.

P d2 *Entropy maximum principle.* (Heterogeneous systems)

An isolated simple system tends toward a heterogeneous equilibrium state that is singled out among the virtual heterogeneous states (*D d4*) by means of the extremum problem

$$S(X_1, X_2, \cdots, X_{r+1}) \\ = \max \{ \sum \mathcal{S}^{(\beta)}(X_1^{(\beta)}, X_2^{(\beta)}, \cdots, X_{r+1}^{(\beta)} ; \eta_1^{(\beta)}, \eta_2^{(\beta)}, \cdots) \}. \tag{1.23}$$

The comparison states for this maximum problem are the virtual heterogeneous states defined in *D d4*.

The remarks following *P b1* apply also in this case, and we state that the solution of the maximum problem (23) is *almost unique*. The precise meaning of this term will be explained later.

C d1 Let the maximum of (23) be attained for the heterogeneous state

$$\bar{X}_1^{(\kappa)}, \bar{X}_2^{(\kappa)}, \cdots, \bar{X}_{r+1}^{(\kappa)} ; \quad \eta_1^{(\kappa)}, \eta_2^{(\kappa)}, \cdots , \quad \kappa = 1, 2, \cdots, m, \tag{1.24}$$

where m is the number of distinct modifications in the actually realized heterogeneous state. If $m = 1$, the state degenerates into a homogeneous equilibrium; only $m \geqq 2$ corresponds to a heterogeneous state in the strict sense of the word.

The translational invariants η are not conserved, and are unconditionally varied in the maximization process.

The expression (23) for the total entropy of the system can be written also as

$$S(X_1, X_2, \cdots, X_{r+1}) = \sum_{\kappa=1}^{m} \mathcal{S}^{(\kappa)}(X_1^{(\kappa)}, X_2^{(\kappa)}, \cdots, X_{r+1}^{(\kappa)} ; \eta_1^{(\kappa)}, \eta_2^{(\kappa)}, \cdots). \tag{1.25}$$

Equations (23) and (25) express the connection between the fundamental equation (9) and the primitive fundamental equation (19).

The heterogeneous equilibrium of a simple system is reminiscent of the composite system in which the distinct properties of the subsystems are maintained by passive forces (*D a5*). In order to emphasize the similarities and dissimilarities of these cases, we introduce the following terminology:

D d5 *Active forces:* phenomenological designation of the sum total of those microscopic factors that bring about and maintain the asymptotically constant values of the free variables in composite and heterogeneous systems.

C d2 It is evident that *the thermostatic formalism does not distinguish between two systems, the variables X_i of which are held constant by active or by passive forces.*

E. CRITERIA OF STABILITY

According to *P b1* and *P d2*, thermodynamic equilibrium is formally associated with the maximization of entropy. The mathematical characterization of the entropy function around its maximum leads to a classification of the states of equilibrium with respect to stability.

Let $S_0(X_{10}, X_{20}, \cdots, X_{r+10})$ be a "point" in Gibbs space (*D b4*) corresponding to the equilibrium of a simple system, and $\delta X_i = X_i - X_{i0}$, $i = 1, 2, \cdots, r$ a virtual displacement (*D a10*) leading to a constrained equilibrium. Because of the insertion of constraints, we now have a composite system. The corresponding virtual entropy change is

$$\Delta S = \delta S + \delta^2 S + \cdots , \tag{1.26}$$

where the terms on the right correspond to the first-, second-, etc. order expansion terms in the virtual displacement.[9] Equilibrium in the widest sense of the word is characterized by the condition

$$\delta S = 0, \tag{1.27}$$

which is valid for all infinitesimal virtual displacements.

The existence of a virtual displacement with $\delta S < 0$, implies that there is a virtual displacement with the opposite sign for which $\delta S > 0$, hence the entropy can actually increase, and there is no equilibrium.

The equilibrium states can be classified in terms of the following relations:

$$\Delta S \leqq 0, \tag{1.28}$$

$$\delta^2 S < 0, \tag{1.29}$$

$$\delta^2 S = 0, \tag{1.30}$$

$$\delta^2 S > 0. \tag{1.31}$$

We define:

D e1 *Equilibrium:* Equation (27) is valid for all (virtual) displacements.

D e2 *Stable equilibrium:* Equations (27) and (28) are valid for all displacement. The stable equilibrium is *normal* if (29) is valid for all displacements, it is *critical,* if (30) is valid for some displacements. The latter are called *critical displacements.*

D e3 *Metastable equilibrium:* Equations (27) and (29) are valid for all displacements; (28) is invalid for some displacements.

D e4 *Essential instability:* Equation (31) is valid for some displacements.

---

[9] We denote infinitesimal and finite virtual displacements by δ and Δ, respectively.

II. DISCUSSION OF THE POSTULATIONAL BASIS

Among the most important concepts of thermodynamics is that of equilibrium, an elusive concept, since there are no purely observational means for deciding whether an apparently quiescent system has actually reached equilibrium, or is merely stranded in a non-equilibrium state while imperceptibly drifting toward equilibrium.

The solution of this difficulty, which is developed in the present paper, is different from the one chosen in the fundamental considerations of the classical theory. Yet, our approach is much closer to the point of view that is implicit in practical procedures.

The classical fundamental approach insists that the basic postulates be of universal validity even if this requirement can only be achieved by advancing comparatively weak statements. Thus, in essence, the second law asserts the impossibility of processes in which systems drift away from equilibrium, but it does not claim that equilibrium is actually reached. Although this procedure has proved fruitful of results, it has some shortcomings. The rigorous implications of the second law are only inequalities, and the derivation of the important thermodynamic equalities always involves the assumption that equilibrium is actually reached, whether or not this is explicitly stated as a postulate.

In this paper, we postulate equilibrium in a much stronger form (*P b1* and *P d2*). Starting from these assumptions, we arrive at a theory of normal equilibrium behavior. In general, the theory does not predict the actual behavior of individual systems, but provides primarily a practical criterion for deciding whether or not equilibrium has been reached in the concrete case. Of course, if past experience justifies us in expecting equilibrium in a given situation, the theory has a very precise predictive value.

The difference between the present and the classical points of view comes to sharp focus in the discussion of phenomena near absolute zero. The absolute predictions of the classical theory become particularly vague because the occurrence of frozen-in nonequilibrium states is very common. The standard discussion of these phenomena is further complicated by the fact that the incomplete equilibrium states can be of a great variety. Thus in a solid that is in equilibrium with respect to the phonon distribution, there may, or may not, be equilibrium with respect to the orientation of molecules, radicals, electronic or nuclear spins, or with respect to the isotopic distribution.

The questions concerning the level at which equilibrium is established can be answered without ambiguity by a systematic application of the present point of view. In fact, the phenomena near absolute zero provide a unique opportunity for putting the concept of equilibrium on a firm basis.

By combining a quantum statistical calculation of the entropy in the ideal gas limit with experimental data, as outlined in Section VII, it is possible to check whether or not the strong version of Nernst's heat theorem (*P c2*) is satisfied

An affirmative answer assures us that equilibrium is indeed established. Moreover, if we remember the fact that entropy is relative to a particular choice of independent components (*C a3*), the procedure can be carried out with various choices of components, with or without taking cognizance of, say, nuclear spin and isotopic differences, and therefore reliable answers are obtained to our questions concerning the level on which equilibrium is established.

This procedure, described here very sketchily, has been extensively used in practice and yielded valuable information not otherwise attainable (*15*).

There is at present a virtually general agreement about the practical applications of Nernst's theorem, called also the third law of thermodynamics (*16*). Nevertheless earlier controversies linger on in formulations that involve weaker statements in order to lay claim to universal validity.

Three of the most common statements of this sort are, in substance, as follows:

(i) It is impossible to attain absolute zero.

(ii) Only entropy differences, rather than the absolute entropies, are required to vanish as $T \to 0$.

(iii) The absolute entropy vanishes in crystalline pure substances.

Statements (i) and (ii) are undoubtedly correct, but too weak to provide the foundation for the practical investigations of low temperature equilibrium mentioned above. Statement (iii) is, in general, not correct, if it is meant as the prediction of the *actual*, rather than of the *normal equilibrium* behavior. In some respects it is not restrictive enough, since there are pure crystalline substances such as ice, which exhibit a finite zero point entropy because of frozen-in orientational disorder (*17*). In other respects it is too restrictive, because ordered alloys and liquid helium are needlessly excluded.

We call attention also to a semantic difficulty, which may have complicated the discussion of the third law. The standard term "absolute entropy" has a double meaning. On the one hand, the entropy of *P c2* is "absolute", in the sense that it involves the *entropy constant* provided by quantum statistics and not only the empirical *entropy differences.* On the other hand, the entropy is *relative* to the choice of independent components (*C a3*). Far from causing difficulties, this circumstance renders the use of *P c2* particularly effective in providing subtle structural information about the level of equilibrium.

We turn now to a methodological problem connected with the fact that the thermostatic formalism deals only with states of equilibrium while nontrivial results can be obtained only if processes are considered at least to a limited extent. The nature of this problem is illustrated by the following paradox: How are we to give a precise meaning to the statement that entropy tends toward a maximum, whereas entropy is defined only for systems in equilibrium? Thus in an isolated simple system (*D a1*, *D a11*) the entropy is constant, if it is defined at all.

This difficulty is resolved in a natural way by the artifice of composite systems

(*D a2*) that enables us to deal with more or less constrained equilibria. Other methods involving phenomena of phase and chemical equilibrium to define constrained equilibria will be considered below.

Gibbs made no explicit use of the concept of composite systems, and his formulation of the increase of entropy is not without a certain obscurity.

The concept of composite systems plays an essential part in Carathéodory's theory (*3*) and is widely used in more recent works (*18*). However, the systematic incorporation of this concept into the foundations of the theory necessitates a considerable revision of the classical conceptual framework. This is evident from the set of definitions grouped under Ia. In particular, we have to consider *thermodynamic operations* as well as *thermodynamic processes*. The latter are always irreversible and quasi-static processes, basic in the traditional theory, and appear now as rather complicated limiting constructs (*D b3*).[10]

The device of composite systems allows us to formulate precise theoretical statements that parallel closely the actual experimental operations. However the quantum-mechanical interpretation of the distinction between thermodynamic processes and operations raises some new problems to be discussed elsewhere.

From the phenomenological point of view, the entropy maximum principle (*P b1*, *P d2*) is so thoroughly corroborated by experiment that we are confident in interpreting any deviation in an actual case as an indication of incomplete thermodynamic equilibrium. The situation is different if we envisage observations on a finer scale and take cognizance of fluctuations. The entropy maximum principle states, in essence, that the continuum of virtual states (*D a9*, *D d4*) is narrowed down by the active forces (*D d5*) to an "almost" uniquely defined actual macroscopic state. (This statement is explained in a remark following *P b1* and in greater detail in Section IX.)

The actual occurrence of fluctuations indicates that this picture is oversimplified[11]: the virtual states have more reality, even in equilibrium, than has been asserted by MTE. In STE (*14*) the phenomenological *P b1* is replaced by statistical postulates. Equilibrium is no longer a state in the thermostatic sense, but rather a distribution over all virtual states. In spite of its statistical character, this theory is still closely related to MTE.

---

[10] The fact that the thermodynamic processes triggered by thermodynamic operations are irreversible, suggests the possibility of a natural generalization of MTE to irreversible thermodynamics. In the latter theory, just as in MTE, we consider the transition from a more constrained to a less constrained equilibrium. However, in MTE, we deal only with the initial and final equilibrium states, ignoring both the intermediate nonequilibrium situations and also the time delay involved in the transition.

[11] We shall see below that critical phenomena manifest themselves as singularities in the mathematical formalism: a danger signal that thermostatics is not entirely correct, appearing already within the macroscopic theory.

According to the postulates of Section Ia, the basic independent variables of the theory are additive invariants (*D a4*), and thermodynamic processes are defined as exchanges of these scalar invariants among coupled systems (*D a3*, *D a7*). The fact that homogeneous phases are the structural elements for building thermodynamic systems (*P d1*) is another aspect of additivity.

The requirement of additive invariance is a characteristic feature of the present theory, in which thermodynamic systems are constructed from spatially disjoint subsystems coupled by the exchange of conserved quantities. Additive invariance is of as much importance to the present formalism as the assumption of nondissipative forces in analytical mechanics. We develop an "analytical thermodynamics" formulated in terms of additive invariants; effects not fitting into this framework are at first excluded, although they can be handled by specialized considerations and corrections.

In *P a3* we have listed the variables that we wish to use in MTE. We have to consider, now, the conditions under which these variables are actually additive invariants. Moreover we have to investigate the status of those elastic, electric, magnetic variables which are not listed in *P a3* although they are generally considered as extensive thermodynamic variables.

The only geometric variable listed in *P a3* is the volume. This means that in the basic theory we neglect surface phenomena and shape-dependent effects. It goes without saying that capillary effects have to be included in a realistic theory of phase equilibrium, and elastic strains are important for the phase equilibrium of crystals. The basic theory of this paper will have to be corrected to include these additional effects.

The volume is obviously an additive variable. For the time being we consider it only as a scale factor (*D d1*), serving merely to define the size of the system. The volume will be introduced as a full-fledged independent variable in Section IV, and we shall consider its "conservation" in Section V.

We turn to the discussion of the internal energy U; this is a variable of unique importance that is relevant for every thermodynamic system. The internal energy is obtained if the total energy is evaluated in the barycentric coordinate system, and if the potential energy arising from external fields is subtracted. For the purposes of MTE, macroscopic kinetic energies will be ignored. We shall assume also the absence of external fields, say, due to gravitation. For small systems, in which the potential energy differences due to internal displacements of matter are negligible compared with the internal energy, the potential energy represents only an irrelevant additive constant. In large systems, however, a gravitational field impairs the spatial homogeneity of the system. It is therefore appropriate in a theory dealing primarily with homogeneous phases to assume the absence of external fields. The generalization to inhomogeneous systems built up from homogeneous volume elements presents no difficulties.

Within the approximations stated there is no distinction between "energy" and "internal energy" and the principle of energy conservation can be invoked to support the validity of the conservation postulate *P a1* for the internal energy. Henceforth, we shall refer to the latter briefly as "energy".

The situation is different for the additivity postulate *P a2*, the validity of which cannot be inferred from general principles. We have to require that the interaction energy between thermodynamic systems be negligible. This assumption is closely related to the homogeneity postulate *P d1*. From the molecular point of view, additivity and homogeneity can be expected to be reasonable approximations for systems containing many particles, provided that the intermolecular forces have a short range character. This conjecture is confirmed by the following phenomenological considerations: As we shall see, thermostatics implies that the stability of phases is an intrinsic property, independent of the size and shape of the system. This result, which follows essentially from additivity (*P a2*) and homogeneity (*P d1*), will be called the "scale invariance" of homogeneous phases. It provides a convenient criterion for gauging the limits of validity of *P a2* and *P d1*.

Within the range of common physical-chemical experimentations scale invariance is satisfied to such an extent that it is usually taken for granted: e.g., equations of state are given for unit amounts of each substance and scaled to any size of interest. Yet, scale invariance breaks down and stability becomes size-dependent in small droplets because of surface effects, in stars because of long-range forces, and in heavy atomic nuclei because of the combination of both factors. In all these cases the basic theory is to be modified.

We turn now to the discussion of the variables describing the distribution of matter. In the absence of chemical reactions, the mole numbers of the chemical species are additive invariants. In the presence of chemical reactions, these numbers do not satisfy the conservation law *P a1*. However, it is possible to define chemical components, the mole numbers of which are invariants of the chemical reactions, and are appropriate thermostatic variables.

The existing presentations of thermodynamics are not as restrictive in the choice of extensive variables as our *P a3*. Instead of requiring additive invariance, the guiding principle is to account for the complete energy balance, and all variables that are needed for describing the performance of work on or by the system are considered. Thus the electric and magnetic polarization vectors, and the components of the eastic strain tensor are joined to the set of independent variables.

We shall refer to these additional extensive variables of the conventional theory as *pseudo-thermodynamic variables*. The variations of these variables describe the performance of work on the system, but these processes cannot be formalized as exchanges of conserved quantities. The pseudo-thermodynamic variables, e.g., magnetic and electric vectors, are subject to boundary conditions that do not conform to the simple type of constraint (1.3). Consequently, the translation of the theorems of MTE to magnetic systems does not always yield correct theorems. The thermodynamics of magnetic systems will not be systematically considered in this paper. In fact, it is advisable to develop the thermodynamics of magnetic systems within a theory that has a certain amount of autonomy. In this theory one has to identify the thermostatic results that can be translated automatically to magnetic systems. Moreover, entirely new problems have to be considered such as the equilibrium of magnetic phases, and, in particular, ferromagnetic domains. These problems have complicated shape dependent aspects, quite foreign to basic MTE (*19*).

The same remarks hold for electric, elastic and capillary systems. None of these will be considered in the present paper.

One of the central ideas of MTE is that spatially extended homogeneous portions of matter are the structural elements for the construction of thermodynamic systems (*P d1*). The spatial homogeneity of phases is taken in TS as an exact mathematical statement. In reality, of course, homogeneity is limited for more than one reason. Because of atomic structure, homogeneity is to be understood in the "coarse grained" sense. (see Appendix B.) A limitation of homogeneity arises also because of the long range interactions discussed above in connection with the additivity of the energy. All of the limitations mentioned leave a wide margin of validity for the homogeneity assumption.

An important application of the phase concept is that it leads to an alternative formulation of the entropy maximum principle (*P d2*). The requirement that a system be in a definite phase is comparable to a constraint imposed in composite systems by restrictive partitions.

In comparing the maximum principles *P b1* with *P d2*, and the entropy function (fundamental equation) with the phase-entropy function (primitive fundamental equation), we are struck by the similarity of the corresponding concepts. It is indeed a characteristic feature of the thermostatic formalism that the same formulas can be used with different sets of variables appropriate to a more or less detailed description of the system. Thus, if the entropy function is expressed as a function of a number of free variables, some of these can be eliminated by maximizing the entropy function with respect to the variables in question, and inserting for these their constant extremal values. Consequently the entropy appears as a function of a reduced number of variables. For example, it would be possible to make our formalism even more detailed and express the primitive fundamental equation in terms of the mole numbers of the chemical species. By maximizing this expression for constant values of the mole numbers of the components, one obtains the equilibrium values of the chemical species. For molecular reactions, these results are standard in chemical thermodynamics, and we do not consider them in this paper.

Granted the similarity of *P b1* and *P d2*, we note the difference that the phase-entropy function (the primitive fundamental equation) is entirely regular with continuous higher derivatives, whereas the second and higher derivatives are only piecewise continuous for the entropy function. These singularities of the entropy function follow from *P d2* and, in fact, the main content of thermostatic phase theory is to derive the topological properties of the sets of singular points in Gibbs space.

The distinction between the primitive fundamental equation and the fundamental equation was recognized by Gibbs. His terminology was "primitive surface" and "surface of dissipated energy". The rationale for the last expression will be explained in Section IV. This distinction between the two surfaces is essential, to account for the phenomenon of metastable phases.

We shall see in Section VII that the phase-entropy function can be established from experimental data,[12] and theoretically, it can be obtained from quantum mechanics, a problem to be discussed elsewhere.

We have seen that MTE is more restrictive in its choice of independent variables than the traditional theory, and we have excluded, at least from the basic MTE, the discussion of pseudo-thermodynamic variables.

In another respect, however, the present theory operates with more variables: namely, with the quasi-thermodynamic parameters of *P d1*, to account for the distinctness of modifications that have identical densities. While the pseudo-thermodynamic variables are bound by more complex constraints than those provided by Eq. (1.3), the quasi-thermodynamic parameters are not bound by any constraints at all. They are discussed in more detail in Appendix A.

The definitions and postulates listed in Ic will be discussed in Section III.

III. THERMAL EQUILIBRIUM

The principal method for drawing inferences from our postulational basis consists of the application of the equilibrium and stability criteria Ie to concrete situations. We consider, at first, the special case of a composite system in which the volume and composition of each subsystem is fixed, and the internal energies $U^{(a)}$ are the only free variables (*D a9*). The problem of establishing the equilibrium distribution of the internal energy is commonly known as that of thermal equilibrium (see *D a12*). The discussion of this case provides an opportunity to clarify the meaning and the limits of validity of the postulates listed in I, C.

Let us consider an isolated composite system divided by diathermic rigid partitions into subsystems of constant volume and composition. Thus the extremum problem (1.10) reduces to

$$S(U) = \max \sum_{a} S^{(a)}(U^{(a)}), \tag{3.1}$$

---

[12] Provided the phase is stable or at least metastable. The absolute value of the phase entropy is obtained from a quantum statistical calculation.

under the constraint

$$\sum_a U^{(a)} = U. \tag{3.2}$$

For the sake of simplicity, we have suppressed the fixed variables $V^{(a)}$ and $N_k^{(a)}$.

Let $\delta U^{(a)}$ be a virtual displacement leading from the equilibrium values $U^{(a)}$ to the constrained equilibrium $U^{(a)} + \delta U^{(a)}$ with

$$\sum \delta U^{(a)} = 0. \tag{3.3}$$

The condition of equilibrium (*D e1*) is

$$\delta S - \lambda \sum \delta U^{(a)} = \sum_a \left(\frac{\partial S^{(a)}}{\partial U^{(a)}} - \lambda \right) \delta U^{(a)} = 0, \tag{3.4}$$

where λ is the Lagrange multiplier associated with the constraint (3). Hence the condition (4) reduces to

$$\frac{\partial S^{(a)}}{\partial U^{(a)}} = \lambda, \qquad a = 1, 2, \cdots . \tag{3.5}$$

The parameter λ is common to all systems that are in thermal equilibrium with each other. It is advantageous from the conceptual point of view to define the temperature as λ^{-1}, i.e., as an intensity that stipulates thermal equilibrium without reference to the properties of any particular system. According to our original definition (*D c1*) the temperature is a zero order homogeneous function of the additive invariants of individual systems. Of course, Eq. (5) ensures the consistency of the two points of view.

Let us now consider a system consisting of two isolated subsystems of slightly different temperatures

$$T' \gtrsim T''. \tag{3.6}$$

After removing the constraint, an actual process associated with an increase of entropy takes place:

$$dS = dS' + dS'' = \frac{\partial S'}{dU'} dU' + \frac{\partial S''}{dU''} dU'' = \left(\frac{1}{T'} - \frac{1}{T''} \right) dU' > 0. \tag{3.7}$$

In view of (6) and *P c1*, we have

$$dU' < 0. \tag{3.8}$$

Hence energy is displaced from the high to the low temperature system. The energy transferred is a heat quantity $đQ$ (*D a12*). We have

$$\frac{1}{T} (dU)_{V,N} = \frac{1}{T} đQ = dS, \tag{3.9}$$

the well-known relation indicating that T^{-1} is the integrating factor of the incomplete differential $đQ$.

The inequality (8) is valid for

$$T' > T'', \tag{3.10}$$

regardless of the magnitude of $T' - T''$, although the expansion (7), broken off after the first term, fails in this case. The flow of energy between the systems produces changes in their temperatures. We claim that the temperature increases, (to be more precise, it does not decrease) for the system that receives the energy:

$$(\partial T/\partial U)_{V,N} \geqq 0. \tag{3.11}$$

This relation can be proved formally from the stability criterion (1.28), as we shall see in Section VIII where the stability problem will be discussed for the general case of many variables. In order to avoid repetition, we invoke at this point only an intuitive argument in favor of (11). This relation assures us that the actual process (8) leads to an equalization of the temperatures T' and T'', and hence to equilibrium. If (11) were violated, the process (8) would produce increasingly different temperatures and violate the condition (5), and *P b1*.

Combining (11) and *D c1*, we have

$$\frac{\partial^2 S}{\partial U^2} = \frac{1}{T^2}\frac{\partial T}{\partial U} \leqq 0, \tag{3.12}$$

and thus the entropy versus energy curve is concave to the U axis.

We define the *heat capacity at constant volume*

$$C_V = (\partial U/\partial T)_V = T(\partial S/\partial T)_V. \tag{3.13}$$

The heat capacity per mole or per unit mass is called *molar heat* and *specific heat*, respectively.

Condition (12) is now written as

$$0 \leqq C_V \leqq \infty. \tag{3.14}$$

In Eq. (13) the temperature is considered to be the independent variable. Thus it is implied that (1.16) which provides the function $T = T(U)$ can be inverted to give $U = U(T)$. This depends on the condition

$$\frac{\partial T}{\partial U} = \frac{\partial^2 S}{\partial U^2} \neq 0 \tag{3.15}$$

or

$$C_V \neq \infty. \tag{3.16}$$

Thus complications may be expected for $C_V = \infty$, that is if (11) is to be taken with the equality sign. We shall consider such singular situations later.

We define the limiting concept of a heat reservoir as a system, the size of which tends to infinity, as compared with the finite systems that are being studied. With the size of the system its extensive variables, in particular U and S, tend to infinity and reservoirs are specified only in terms of their temperatures. If a finite system is coupled to a reservoir, the temperature of the latter stays constant and the same temperature is reached asymptotically by the finite system.

We turn now to the discussion of the postulates listed in Ic. Postulates *P c2* and *P c3*, although phrased in phenomenological terms, are expressions of the quantum mechanical nature of thermodynamic systems.[13] We pointed out already in Section II that *P c2* is of fundamental importance for the concept of equilibrium. We shall return to this question later.

Until recently most texts of thermodynamics have implicitly assumed that the thermodynamic temperature is positive (*P c1*). This statement appears as a theorem in the presentation of Landau-Lifshitz (*20*). It is proved for systems in which the conversion of the internal energy into molecular kinetic energy is a virtual process. If such systems had a negative temperature: $\partial S/\partial U < 0$, an increase of the entropy would be brought about by converting the internal energy into the kinetic energy of internal macroscopic motion. This is a violation of *P b1*.

We should note in this context that Ramsay (*21*) recently demonstrated that negative absolute temperatures are useful for the description of the properties of certain rather unusual systems. The thermodynamic system is a collection of nuclear spins, thermally isolated from the lattice occupying the same region of space. In Ramsay's system the proof mentioned above fails, since there is no conversion of internal energy into molecular kinetic energy (*22*). The concept of two isolated systems (spin and lattice) occupying the same region of space is in conflict with our *D a1*. As in a number of other instances, our basic theory is restricted to a narrower class of systems, in order to yield a larger number of theorems. However, our procedure is consistent with the use of different definitions to suit specialized situations.

After relating the various types of thermostatics definitions of the temperature to each other, the question of measurement and the establishment of an absolute scale is still left open. We shall briefly indicate in Section VII that this

---

[13] The quantum-mechanical character of the third law of thermodynamics is well known. On the other hand, it is often stated that the lowest state of classical systems is the crystalline state. Actually, from the classical point of view, a collection of nuclei and electrons has no stable state, and the energy would tend to negative infinity due to the collapse of opposite charges.

question is soluble in a practical fashion. Meanwhile, we anticipate the result that temperature is measurable on an absolute scale.

Finally, we consider the procedure of heating a thermodynamic system from an initial temperature T_i to a final temperature T_f, and illustrate the existence of different paths characterized by their measure of irreversibility (*D b1* to *3*). Let us first perform the heating in the simplest manner by coupling the system with a heat reservoir of temperature T_f.

The measure of irreversibility defined in (1.13), in this case, is

$$DS = \int_{T_i}^{T_f} \frac{C_V}{T}\, dT - \frac{1}{T_f}\int_{T_i}^{T_f} C_V\, dT, \tag{3.17}$$

where according to (13) the first term represents the entropy increase of the system, the second is the entropy decrease of the reservoir. The net entropy change is obviously positive.[14]

Let us consider now the same initial and final states of the system and perform the transition over a path consisting of n steps by coupling the system consecutively with n reservoirs of temperatures $T_1, T_2, \cdots, T_n$, where

$$T_i < T_1 < T_2 < \cdots < T_n = T_f.$$

The contact with the reservoir T_k is to be brought about after the system has reached its equilibrium at T_{k-1}. The total measure of irreversibility is now

$$DS^{(n)} = \int_{T_i}^{T_f} \frac{C_V}{T}\, dT - \sum_{k=1}^{n} \frac{1}{T_k}\int_{k-1}^{T^k} C_V\, dT, \tag{3.18}$$

hence,

$$\lim_{n\to\infty} DS^{(n)} = 0. \tag{3.19}$$

Thus the path is reversible in the ideal limit in which it is performed in infinitely many infinitesimal steps (*D b3*).

IV. THE ENERGY SCHEME

It is an essential feature of MTE that the fundamental equation and the associated stability considerations can be cast in different forms that depend on the choice of independent variables. Although these forms are all consistent with each other, their suitability for solving various problems differs greatly. Thus before extending the discussion of the last section to the general stability problem, we have to expand the basic formalism of the theory.

As a first step toward developing a systematic transformation theory we solve

---

[14] In the special case in which C_v is constant in the interval $T_i \leqq T \leqq T_f$, the entropy increase is represented geometrically as an area in the diagram in which C_v/T is plotted against T.

the fundamental equation

$$S = S(U, X_2, \cdots, X_{r+1}) \tag{4.1}$$

for the energy:

$$U = U(S_1, X_2, \cdots, X_{r+1}). \tag{4.2}$$

This procedure is applicable also to the primitive fundamental equation and Eq. (1) may represent either (1.9), (1.19), or (1.21). However, we omit the superscripts if the identification of subsystems or phases is not essential.

We shall refer to the formalism based on (1) and (2) as the *entropy* and *energy scheme*, respectively. Although, formally, the transition from (1) to (2) consists merely of the rotation of the Gibbs space resulting in the interchange of the roles of S and U as dependent and independent variables, there are essential differences in the physical interpretation of the two schemes.

First, we note that Eqs. (1) and (2) are mathematically equivalent in the following sense: The power series expansion of Eq. (1) around a "point" U_0, S_0, can be resolved to give the power series expansion of Eq. (2), provided that the condition

$$(\partial S/\partial U)_0 \neq 0 \tag{4.3}$$

is satisfied. Conversely, Eq. (1) can be regained from Eq. (2) if

$$(\partial U/\partial S)_0 \neq 0. \tag{4.4}$$

According to (3) the energy scheme is not unique for $T \rightarrow \infty$, a limitation of little consequence. The neighborhood of absolute zero is of greater practical importance. At first sight, the breakdown of condition (4) would seem to indicate here a less than complete equivalence of the two schemes. Given the energy fundamental equation, the entropy function can be established only with a large uncertainty. However this difficulty is removed by *P c2* that assigns a vanishing entropy to all systems in equilibrium for $T \sim 0$.

We proceed now to compare the thermodynamic processes in the two schemes. Considering the exchange of additive invariants between two subsystems of a composite system, we obtain the virtual processes:

$$\Delta S' + \Delta S'' = 0, \qquad \Delta X_i' + \Delta X_i'' = 0, \qquad i = 2, 3, \cdots, r + 1. \tag{4.5}$$

If we compare these processes with those of Eq. (1.5c), we see that the conservation of energy has been replaced with that of the entropy. The latter condition means that the process is reversible. The nonconservation of the energy is not objectionable for virtual processes. Nevertheless, it is intuitively satisfying to introduce an auxiliary system to provide the excess energy $\Delta U' + \Delta U''$ that may be positive or negative. In order to avoid any irreversible entropy increase, this system is chosen as a "work reservoir," an idealized device that is not the seat of thermal effects. Of course, in *actual* processes the presence of the work reservoir is necessary to maintain the energy balance.

It is interesting to compare the virtual processes (5) with those of the entropy scheme (1.5c). In the second scheme the energy is conserved along with the other additive invariants, while the entropy is allowed to vary. Whereas in the entropy scheme composite systems were essential for the discussion of processes, in the energy scheme one may also consider simple systems coupled with a work reservoir.

The difference between the processes of the two schemes is illustrated also by the following example: in the energy scheme we have adiabatic volume changes $(\Delta V)_S$, e.g., in sound waves; in the entropy scheme, we have free expansions, $(\Delta V)_U$. Evidently, the two schemes deal with different types of processes. We note also the analogy to the Clausius and to the Kelvin–Planck formulations of the traditional theory.

The stability of a thermodynamic system depends on whether or not any of the virtual processes can actually occur. We have seen that this question is decided by the entropy maximum principle, the counterpart of which in the transformed formalism is the *energy minimum principle.* Stating it loosely; in equilibrium the entropy is a maximum at constant energy, while the energy is a minimum at constant entropy.

In Section I we expressed the entropy maximum principle in three different forms (*P b1*, *P d2*, and *Ie*). We shall confine ourselves to the reformulation of the stability criterion *Ie* and we shall prove the equivalence of the two formulations.

Criterion of Stability. Consider a simple system mentally divided into two subsystems by a nonrestrictive mathematical wall. Transfer this state by a virtual process (5) into a constrained equilibrium that is "trapped" by the insertion of a restrictive wall. *The variation of the volume by a movable piston* is now included among the virtual processes. Let the process be associated with the absorption of the virtual work

$$\Delta W = \Delta U' + \Delta U'' \tag{4.6}$$

from the work reservoir.

We consider the relations that are analogous to (1.26) to (1.31):

$$\Delta W = \delta W + \delta^2 W + \cdots, \tag{4.7}$$

$$\delta W = 0, \tag{4.8}$$

$$\Delta W \geqq 0, \tag{4.9}$$

$$\delta^2 W > 0, \tag{4.10}$$

$$\delta^2 W = 0, \tag{4.11}$$

$$\delta^2 W < 0. \tag{4.12}$$

D f1. *Equilibrium:* Equation (8) is valid for all (virtual) displacements.

D f2. *Stable equilibrium:* Equations (8) and (9) are valid for all displacements. The stable equilibrium is *normal* if (10) is valid for all displacements; it is *critical* if (11) holds for some displacements. The latter are called *critical displacements.*

D f3. *Metastable equilibrium:* Equations (8) and (10) are valid for all displacements, (9) is invalid for some displacement.

D f4. *Essential instability:* Equation (12) is valid for some displacement.

Expression (6) represents the minimum work required to produce the displacement (5) leading from the equilibrium to the constrained equilibrium state.

If the roles of initial and finite states are interchanged, $-(\Delta U' + \Delta U'') = | \Delta U' + \Delta U'' |$ is the maximum work (available work) which may be extracted from the system on transferring it from the constrained to the unconstrained equilibrium.

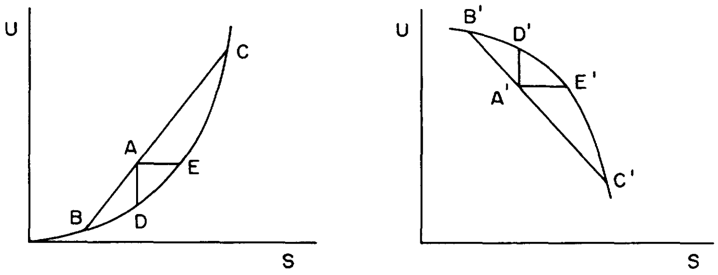
The foregoing criterion is obviously identical in wording to the energy minimum principle of macroscopic physics (mechanical statics and electrostatics) where thermal effects are ignored. Hence the thermostatic energy-minimum principle can be expressed concisely as follows:

The energy-minimum principle of macroscopic physics maintains its validity in thermostatics if the comparison states are provided by reversible virtual processes in which the total entropy is held constant.

We shall prove the equivalence of the criteria. Let us at first assume that a state is instable according to the energy criterion: thus there is a varied state specified by (5) for which $W = \Delta U' + \Delta U'' < 0$ and $\Delta S' + \Delta S'' = 0$. After dissipating the energy $| W |$ in the varied state, the energy is brought back to its original value and, by virtue of *P c1*, the entropy has increased. Hence the state is instable according to the entropy criterion. Conversely, we assume that there is a state with $\Delta S = \Delta S' + \Delta S'' > 0$ and $\Delta U' + \Delta U'' = 0$. Removing the entropy ΔS through contact with a heat reservoir leads to a varied state with $\Delta S = 0$, $\Delta U < 0$, and hence to instability with respect to the energy criterion.

The above proof relies essentially on the assumption that the temperature is positive (*P c1*). Although the discussion of negative temperatures is beyond the scope of this paper, it is instructive to consider such systems for comparison. The positive and negative temperature cases are represented in Figs. 1(a) and 1(b), respectively.

Let us consider a system represented by point A. We interpret this situation as a composite system in a constrained equilibrium, where the states of the separate parts are represented by B and C. The relative amounts of these parts are given by the ratio of the line segments AC/BA (lever rule). If the internal constraint inhibiting the passage of energy is relaxed, the system performs an irreversible process into the state E, and $\Delta S = AE$ is the measure of irreversi-



(a) (b)

FIG. 1. The fundamental equation, $U = U(S)$, with X_i = constant ($i = 2, 3, \cdots$); (a) positive temperatures (b) negative temperatures.

bility. If the system is coupled with a work reservoir and the process is led reversibly, the maximum available work is represented by $-\Delta U = AD$.

The case for negative temperatures is presented in Fig. 1(b). The irreversible attainment of equilibrium is essentially the same as in the positive temperature case, but the reversible transition to equilibrium requires the work input $\Delta U' = A'D'$.

On the other hand, in the negative-temperature case, the system is capable of evolving spontaneously from B' towards E' by simultaneously increasing its entropy and performing work. However, from the kinetic point of view, this process must proceed slowly (long relaxation time) because otherwise the system would not be found at all in a negative-temperature state.

Finally, we note that Fig. 1(a) has already been used by Gibbs (p. 51 of Ref. 2). However, point A is interpreted there as a system the parts of which are in state of macroscopic motion, and the process $A \rightarrow E$ is the dissipation of the kinetic energy. Accordingly Gibbs calls the surface $BDEC$ the "surface of dissipated energy." This approach has the disadvantage that the initial state is not an equilibrium, and strictly speaking, cannot be represented in the diagram.

V. THE INTENSIVE VARIABLES

We write the fundamental equation (4.2) in the differential form

$$dU = \sum_{i=1}^{r+1} P_i \, dX_i , \tag{5.1}$$

where the

$$P_i = \frac{\partial U}{\partial X_i}, \qquad i = 1, 2, \cdots, r + 1 \tag{5.2}$$

are the intensities conjugate to the extensive variables X_i. The entropy is included here among the extensive variables: $X_1 = S$.

Taking into account the physical interpretation of the variables (*P a3*), we obtain corresponding interpretations for the intensities:

$$dU = TdS - pdV + \sum_{j=1}^{c} \mu_j \, dN_j \, . \tag{5.3}$$

The temperature T is the conjugate of the entropy. This relation is equivalent to *D c1*, a definition that can be expressed by saying that in the entropy scheme $1/T$ is conjugate to U. The definition of the intensities within the entropy scheme will be briefly discussed later.

The chemical potentials μ_j are defined as the intensities conjugate to the mole numbers of the chemical components. The intensity conjugate to the volume is the negative pressure $-p$. This follows from the requirement that in the absence of caloric and chemical changes Eq. (3) should reduce to the well-known energy relation of fluid mechanics.

We shall derive a number of formal properties of the intensive variables, and follow it, at the end of this section, by a few more specific results that are related to their physical interpretation.

We consider a composite system in which one of the variables, say X_k, is free to vary on account of nonrestrictive partitions. In other words, we have the virtual processes:

$$\sum_a \delta X_k^{(a)} = 0$$

and (5.4)

$$\delta X_i^{(a)} = 0, \quad \text{for all } i \neq k \text{ and all } a.$$

Applying the equilibrium condition formulated in the last section, we have

$$\delta W = \sum_a \delta U^{(a)} = 0 \tag{5.5}$$

under the constraints (4). By using the method of Lagrange multipliers, we obtain

$$\delta U - \lambda_k \sum_a \delta X_k^{(a)} = \sum_a (P_k^{(a)} - \lambda_k) \delta X_k^{(a)} = 0. \tag{5.6}$$

Hence for all subsystems coupled with each other through the exchange of X_k we arrive at the equilibrium condition

$$P_k^{(a)} = \lambda_k = \text{constant for all } a. \tag{5.7}$$

In other words the free extensive variables will attain those equilibrium values which ensure that their conjugate intensities are constant. This result is immediately generalized to the case of virtual processes involving the simultaneous variation of two or more extensive variables X_i, provided that the different variations $\delta X_i^{(a)}$ and $\delta X_k^{(a)}$ are independent of each other. The conjugate intensities are always constant over the entire composite system.

In order to appreciate the significance of the intensive variables, it is convenient to single out one subsystem for detailed consideration, and join all the others to provide the "surroundings" of this system. We require also that the over-all size of the surroundings be very large, in the limit infinite, compared with the system of interest. The surroundings is specified by the set of its intensities $P_1, P_2, \cdots$. This description is quite schematic; it ignores the densities of the subsystems constituting the surroundings. The specification is adequate, nevertheless, for deciding whether or not a system is in equilibrium when coupled to this surroundings. Moreover, because of the large size of the latter, the intensities are not influenced by any exchange of quantities between system and surroundings. (This plausible statement is formally proved in Section VIII.) We shall refer to "surroundings" also as *(generalized) reservoirs.*

According to Eq. (2) the intensive variables are functions of the extensive variables of the system. Since the intensities are zero order homogeneous functions of the extensive variables, and do not depend on the size of the system, the intensities depend actually only on the densities (1.20). Hence

$$P_i = P_i(x_1, x_2, \cdots, x_r) \qquad i = 1, 2, \cdots, r + 1. \qquad (5.8)$$

Relations of this sort that arise through differentiation of fundamental equations we shall call *generalized equations of state.* We shall see that the relations commonly called equations of state arise from (8) through transformations of the variables. However, the common usage is more restricted and does not include the complete set of Eqs. (8). The information contained in this complete set is equivalent to that of the fundamental equation. In fact, applying Euler's relation to the first-order homogeneous equation (4.2), we obtain

$$U = \sum_i^{r+1} X_i \frac{\partial U}{\partial X_i} = \sum_i^{r+1} X_i P_i. \qquad (5.9)$$

However, the different equations of state are not independent from each other, but satisfy the so-called Maxwell relations

$$\frac{\partial P_i}{\partial X_k} = \frac{\partial P_k}{\partial x_i}, \qquad i, k = 1, 2, \cdots, \qquad (5.10)$$

since these quantities are the mixed second derivatives of the fundamental equation.

Differentiating (9) and subtracting (1), we obtain the Gibbs-Duhem (GD) relation

$$\sum_{i=1}^{r+1} X_i \, dP_i = 0. \qquad (5.11)$$

Dividing by the scale factor X_{r+1}, we obtain

$$dP_{r+1} = -\sum_{i=1}^{r} x_i \, dP_i \, . \tag{5.12}$$

We assume that Eqs. (8) can be solved for the densities:

$$x_i = x_i(P_1, P_2, \cdots, P_r), \qquad i = 1, 2, \cdots, r. \tag{5.13}$$

The condition of solubility is the nonvanishing of the Jacobian

$$D_r = \frac{\partial(P_1, P_2, \cdots, P_r)}{\partial(x_1, x_2, \cdots, x_r)} \neq 0. \tag{5.14}$$

Inserting (13) in (9), we obtain P_{r+1} as a function of the other intensities:

$$-P_{r+1} = \psi(P_1, P_2, \cdots, P_r). \tag{5.15}$$

where the minus sign is for convenience.

This function is an integral of the differential equation (12), which indicates that the integrability condition

$$\frac{\partial x_i}{\partial P_k} = \frac{\partial x_k}{\partial P_i}, \qquad i, k = 1, 2, \cdots \tag{5.16}$$

is satisfied. However, a direct integration of (12) does not yield (15) but involves undetermined integration constants.

Conversely we can regain Eq. (8) from (13), provided that

$$D_r^{-1} = \frac{\partial(x_1, x_2, \cdots, x_r)}{\partial(P_1, P_2, \cdots, P_r)} \neq 0. \tag{5.17}$$

Conditions (14) and (17) are significant. We shall see that (14) breaks down for heterogeneous and critical equilibria, and (17) near absolute zero. We shall refer to states in which these conditions are satisfied as *regular*.

We have from (8) and (5)

$$\psi(P) = \sum_{i=1}^{r} x_i P_i - u(x), \tag{5.18}$$

where x and P each stand for the entire set $x_1, x_2, \cdots, x_r$ and $P_1, P_2, \cdots, P_r$, respectively. Thus (15) is the Legendre transform of the fundamental equation $u = u(x)$ expressed in density variables. Obviously, the size of the system cannot be expressed in terms of intensities alone. In regular states the transformation (18) allows one to obtain $\psi = \psi(P)$ from $u = u(x)$ and vice versa, the Legendre transformation preserves the information contained in the fundamental equation.

The basic extremum principle of thermostatics is easily translated into the present context. We define the function

$$w(x'; P) \equiv u(x') + \psi(P) - \sum_{1}^{r} x_i' P_i , \tag{5.19}$$

where x' represents the set of densities $x_1', x_2', \cdots, x_r'$. In general the x' are distinct from the x that represent the densities of the system when in equilibrium with the reservoir; we say also that the x and P are "matched." It is easy to see that

$$w(x', P) \geqq 0 \tag{5.20}$$

In fact, according to (18) we have $w(x, P) = 0$ for matching variables. Otherwise

$$w(x'; P) = w(x'; P) - w(x; P) = u(x') - u(x) - \sum_{i} (x_i' - x) P_i . \tag{5.21}$$

This is the minimum work required to transfer the system from the state x to x' while it is in contact with a reservoir in equilibrium with the initial state. This quantity, of course, must be non-negative, hence (20) is proved. This relation can be considered as a minimum principle in which the x' are varied at constant P, or vice versa.

One often uses fundamental equations in terms of a mixed set of variables, say $P_1, P_2, \cdots, P_k, X_{k+1}, \cdots, X_{r+1}$. We shall use occasionally for the corresponding Legendre transform the short notation

$$-U[P_1, P_2, \cdots, P_k] = \sum_{i=1}^{k} X_i P_i - U(X_1, X_2, \cdots, X_{r+1}) \tag{5.22}$$

More specifically by using the variables made explicit in (3), we have

$$U = TS - pV + \sum_{j} \mu_j N_j . \tag{5.23}$$

Thus the Gibbs function is

$$G(T, p, N_1, N_2, \cdots, N_r) = U[T, p] = U - TS + pV. \tag{5.24}$$

We shall consider now some of the properties of the intensive variables that are connected with their specific physical interpretation.

In Section III temperature, or rather $1/T$, was seen to govern the level of thermal equilibrium. In the energy scheme the equality of temperature throughout a composite system is the condition of equilibrium with respect to the performance of work, while entropy is reversibly transferred between subsystems of definite temperatures. Such processes are called Carnot cycles. The temperature concept emerging from the analysis of this process is identical to the one established in Section III and we shall not pursue this line of thought any further.

We turn to the discussion of pressure which is the intensity governing mechanical equilibrium. The formal analogy between temperature and pressure is incomplete. Although the mechanical coupling of two fluid systems can often be formalized as an "exchange of volume" between subsystems, this is not necessarily the case for all types of coupling. Thus the volume of such systems could be varied by a separate piston of cross section A' and A'', while the pistons are rigidly connected with each other. The virtual process is now

$$\delta V'/A' + \delta V''/A'' = 0. \tag{5.25}$$

The corresponding equilibrium condition is easily shown to be

$$A'p' = A''p''. \tag{5.26}$$

Only for $A' = A''$ do we get the standard thermostatic situation.

The possibility of a volume coupling of the type (25) was first pointed out by Ehrenfest–Afanassjewa (*23*) in order to emphasize the contrast with the case of temperature, a parameter governing thermal equilibrium independently of the mechanism providing the energy coupling.

In case of a heterogeneous system in which one phase can expand at the expense of the other, we have $A' = A''$ and the general case $A' \neq A''$ is of no importance in the context of this paper. It is nevertheless interesting that the formalization of mechanical interaction, as an exchange of an additive invariant, has its limitations. We return to this point at the end of this section.

We note that the use of adiabatic pistons enables one to consider volume exchange as a one-variable process, provided that one is within the energy scheme. Thus we have

$$\delta V' + \delta V'' = 0, \qquad \delta S' = 0, \qquad \delta S'' = 0. \tag{5.27}$$

Within the entropy scheme one has to consider the two-variable process

$$\begin{aligned} \delta U' + \delta U'' &= 0, \\ \delta V' + \delta V'' &= 0. \end{aligned} \qquad \text{(b)} \quad (5.28)$$

since adiabatic walls do not ensure the constancy of the energies of the subsystems when work is performed through the exchange of volume. In order to ensure $\delta U' = \delta U'' = 0$, the energy lost or gained through work has to be compensated by the right amount of heat in each subsystem. This, of course, cannot be automatically achieved through a constraint.

Finally we consider equilibrium processes, in which there is a displacement of mass, i.e., of mole numbers. The conjugate intensity is the chemical potential. The analogy between chemical potential and temperature governing mass and energy flow, respectively, is quite far reaching. However the chemical potential involves a number of additional features which we shall briefly survey.

For a one-component system the chemical potential $\mu = P_{r+1}$ can be reduced to quantities introduced earlier. Choosing the mole number N as a scale factor, we obtain from (9) and (12)

$$\mu = u - Ts + pv \tag{5.29}$$

and

$$d\mu = -s\,dT + v\,dp. \tag{5.30}$$

Accordingly, the chemical potential, besides being an intensity, is also the Legendre transform of the molar energy, or, in standard terminology, it is identical to the molar Gibbs function. For multicomponent systems, the chemical potential can be identified with the partial molar Gibbs function.

An interesting distinction between energy and mass flow is connected with the availability of restrictive walls. While there exist walls that are restrictive of mass and nonrestrictive of energy and entropy, the converse is generally not true. If a wall is nonrestrictive of at least some chemical species (semipermeable membrane), it is nonrestrictive of entropy and energy also. Consequently, the virtual processes involve simultaneous independent variations of the mole numbers and of the entropy:

$$\begin{aligned} \delta N_i' + \delta N_i'' &= 0, && \text{(a)} \\ \delta S' + \delta S'' &= 0. && \text{(b)} \end{aligned} \tag{5.31}$$

The criterion of equilibrium consists of the simultaneous conditions

$$\begin{aligned} \mu_i' &= \mu_i'', && \text{(a)} \\ T' &= T''. && \text{(b)} \end{aligned} \tag{5.32}$$

Although these considerations are all but universally valid, there is at least one exception demonstrating the contingent character of our assumption on the availability of walls.

A narrow capillary acts in liquid helium II as an entropy filter, it inhibits the flow of the normal fluid that contains the entropy and transmits the superfluid, the entropy content of which is presumably zero (*24*). Thus the capillary is nonrestrictive of mass and restrictive of entropy flow. Accordingly, for two containers of helium II thus connected, $\Delta\mu = 0$, but we may have $\Delta T \neq 0$. Hence from (26) we obtain

$$\frac{\Delta p}{\Delta T} = \frac{s}{v} = s\rho. \tag{5.33}$$

This is the well known *H*-London relation experimentally verified in very narrow capillaries from the λ-point to 0.1°K (*25*).

Capillaries wide enough to allow the flow of entropy carried by the normal liquid no longer act as entropy filters, and (29) has to be replaced by a relation based on irreversible thermodynamics (*26*).

At this point, we are able to explain more fully the restrictive character of our definition of the independent variables as contained in *P a3*. Following the usual practice, we could supplement Eq. (11) by additional terms $P_i\,dX_i$, where the X_i are, say, the components of the electric and magnetic polarization, or of the elastic strain tensor. The corresponding P_i are the components of the electric and magnetic field intensity and of the elastic stress tensor. These additional variables satisfy the Maxwell relations (10) and (16). However the processes involving electric, magnetic, and elastic work cannot be formalized as exchanges of additive invariants, and the equilibrium condition (7) is not valid for the conjugate P_i. Thus in a heterogeneous system the magnetic intensity is not constant throughout the system. Moreover, the homogeneity of the system is impaired by the application of elastic stresses. For this reason, we denote these P_i as pseudo-intensities, the conjugate X_i as pseudo-thermodynamic variables, and we exclude them from present consideration.

VI. THE FIRST PHASE RULE

The basic problem of the thermostatic theory of phase equilibrium can be formulated as follows: We consider a thermodynamic system of given composition. We assume that the *possible* phases of this system are known and are specified in terms of its primitive fundamental equations (*P b1*). The problem is to predict the *actual* heterogeneous structure of a simple system specified in terms of its extensive variables.

This question is answered in principle in *P b2*. However, this answer is unwieldy, it involves, on equal terms, all possible phases of a system.

In order to arrive at more useful general results, we simplify the problem by using the concept of a generalized reservoir, which was introduced in the last section. We single out one modification in a heterogeneous system for special consideration, and study its equilibrium with the rest of the system which is assumed to be large enough to be taken as a reservoir.

The intrinsic properties of the modification are specified by the set of densities $x_1, x_2, \cdots, x_r$ that represent a point in x-space. The size of the system is given by the scale factor X_{r+1}.

The modification can be kept in equilibrium either by completely restrictive walls or by contact with a generalized reservoir of an appropriate set of intensities $P_1, P_2, \cdots, P_{r+1}$. These intensities are related to each other by the differential GD relation (5.12). The integral relation (5.15) associates an r-dimensional surface in P-space with the modifications of a single phase. This means that only r intensities are capable of independent variations while still providing the environment for the same phase in one of its modifications.

These results can be generalized at once to the case of heterogeneous system consisting of, say, m modifications. If each of these modifications is in equilibrium with the same reservoir, the modifications are also in equilibrium with each other. In usual parlance, the modifications are said to coexist.[15]

The equilibrium of the m modifications with the same reservoir is formally expressed in terms of the GD relations

$$dP_{r+1} = -\sum_{i=1}^{r} x_i^{(\kappa)}\, dP_i, \quad \kappa = 1, 2, \cdots, m, \tag{6.1}$$

where the intensities have the same value in all coexisting modifications but the densities $x_i^{(\kappa)}(P_1, P_2, \cdots, P_r)$ are characteristic of each modification.

Each of these equations can be used to eliminate one of the intensities, and therefore, we are left with $\delta_I = \mathrm{r} + 1 - m$ intensities capable of independent variations. The number δ_I is equal to the dimensions of the region of P-space that represents the heterogeneous equilibrium of the m distinct modifications. This conclusion rests on the tacit assumption that Eqs. (1) are linearly independent. Assuming more generally that only $\tilde{m}$ of the m equations (1) are linearly independent, we obtain

$$\delta_I = r + 1 - \tilde{m} = c + 2 - \tilde{m}, \tag{6.2}$$

in which we have made use of (1.7).

The non-negative number δ_I is called the variability of the heterogeneous system, or the number of its thermostatic degrees of freedom. We shall refer to (6.2) as the "first phase rule."

In deriving this rule Gibbs had assumed (p. 96 of Ref. *2*) that $\tilde{m} = m$, i.e., that the GD equation associated with different phases are linearly independent. This assumption is indeed plausible if it is taken in conjunction with the tacit assumption of the classical theory that two modifications are distinct if and only if they have different densities.

The situation has to be reconsidered from the point of view of the present definition of distinctness (*D d3*). We agree with the classical point of view in precluding an accidental linear dependence of GD relations of modifications having different densities. However, modifications related to each other by inversion or mirroring are energetically equivalent and have the same GD relation. Moreover, they are distinghishable by virtue of their symmetry. The last condition is not satisfied in the case of equivalence under a pure rotation.

We shall refer to such modifications or phases as modification doublets, or phase doublets. The distinction between these two concepts depends on the

---

[15] Note that we consider the phase surfaces "in the small." Thus coexisting liquid and vapor modifications are represented by separate GD relations that yield two surface elements upon integration. When they are investigated by methods "in the large" (Section IX), these surface elements turn out to be parts of the same surface that represent the fluid phase. However, this connection "in the large" is irrelevant in the present context.

analysis of the properties of the phase surface in the large and will be explained in Section IX.

The present version of the first phase rule differs from the classical formulation only to the extent that, when the system contains doublet modifications, or phase doublets only one of each pair is to be counted when the effective number of modifications m is being established.

We consider somewhat closer the properties of a system consisting of two coexisting modifications that differ in their densities. The complete description of the system is provided by the $r - 1$ independent intensities $P_1, P_2, \cdots, P_{r-1}$ and the scale factors X'_{r+1} and X''_{r+1}. Let us consider a phase transition, i.e., a process in which the boundary between the modifications is displaced, and

$$\Delta X'_{r+1} + \Delta X''_{r+1} = 0 \tag{6.3}$$

while the system is in contact with the same reservoir. In the course of this process the reservoir has to supply the quantities $\Delta X_1, \Delta X_2, \cdots, \Delta X_r$ which are in definite proportion to each other and define a direction in Gibbs space:

$$\Delta X_1 : \Delta X_2 : \cdots : \Delta X_r = (x_1'' - x_1') : (x_2'' - x_2') : \cdots : (x_r'' - x_r'), \tag{6.4}$$

where x_i', x_i'' are the densities of the coexisting modifications.

The differences $x_i'' - x_i'$ are characteristic properties of the phase transition: increments of volume, concentration and entropy. The latter is related to the latent heat l: $s'' - s' = l/T$.

Finally, we note that after subtracting the GD relation of a phase from that of a coexisting phase, we obtain the generalized Clausius–Clapeyron equation

$$\sum_{i=1}^{r} (x_i'' - x_i') \, dP_i = 0, \tag{6.5}$$

where the dP_i are intensity variations in the region of coexistence in P-space.

The first phase rule expresses an upper limit for the number of coexisting modifications. We turn now to the study of the intrinsic stability of homogeneous systems, in order to find out the conditions under which the heterogeneous equilibrium is realized rather than the homogeneous one.

VII. THE THERMOSTATIC FUNDAMENTAL FORMS. CONNECTION WITH EXPERIMENT.

The study of the intrinsic stability of a homogeneous phase involves the discussion of the quadratic expansion terms of the primitive fundamental equation. This quadratic form is also important in establishing the connection of the theory with experiment. Therefore we investigate its properties before turning to our central problem.

We consider a single homogeneous phase, and expand its primitive fundamental equation about a point $X_{10}, X_{20}, \cdots, X_{r+1,0}$ in Gibbs space. We vary only the first r variables, keep the scale factor X_{r+1} constant, and write the aforementioned expansion in terms of densities as

$$u(x_1, x_2, \cdots, x_r) = u_0 + \sum_1^r P_{i0}\xi_i + \frac{1}{2}\sum u_{ik}\xi_i\xi_k + \cdots, \tag{7.1}$$

where

$$\xi_i = \delta x_i = x_i - x_{i0}, \tag{7.2}$$

$$P_{i0} = (\partial u/\partial x_i)_0, \tag{7.3}$$

and

$$u_{ik} = (\partial^2 u/\partial x_i \partial x_k)_0 = (\partial P_i/\partial x_k)_0 = (\partial P_k/\partial x_i)_0. \tag{7.4}$$

We denote the quadratic form in (7.1) as the thermostatic *fundamental form*, or *stiffness form*. The corresponding matrix $\| u_{ik} \|$ will be called the *stiffness matrix*. Its physical meaning is evident from

$$\pi_i = \sum_1^r u_{ik}\xi_k, \tag{7.5}$$

where

$$\pi_i = \delta P_i = P_i - P_{i0}. \tag{7.6}$$

The determinant of the stiffness matrix is

$$\det | u_{ik} | = D_r, \tag{7.7}$$

where D_r is the Jacobian (5.14). We shall consider at first the regular case, for which $D_r \neq 0, \infty$. The cases $D_r = 0$ and $D_r = \infty$ are singular and will be considered below.[16]

In the regular case the fundamental equation has a Legendre transform (5.18) that can be expanded as follows:

$$\psi(P_1, P_2, \cdots, P_r) = \psi_0 + \sum_1^r x_{i0}\pi_i + \frac{1}{2}\sum_1^r \psi_{ik}\pi_i\pi_k + \cdots, \tag{7.8}$$

where

$$\psi_{ik} = (\partial^2\psi/\partial P_i \partial P_k)_0 = (\partial x_i/\partial P_k)_0 = (\partial x_k/\partial P_i)_0, \tag{7.9}$$

and

$$\xi_i = \sum_1^r \psi_{ik}\pi_k. \tag{7.10}$$

---

[16] In case the full set of $r + 1$ extensive quantities is subject to variation, the determinant of the expanded set of Eqs. (5) vanishes: $D_{r+1} = 0$. This follows from the fact that the GD equation (5.11) constitutes a linear relation among the $r + 1$ equations. This result is intuitively obvious: the scale factor X_{r+1} cannot be expressed solely in terms of size indepenent intensities.

The quadratic form in (8) we call the *compliance form.* We also refer to the stiffness and compliance forms jointly as the "fundamental forms." The matrix $\| \psi_{ik} \|$ is the *compliance matrix*. It is the inverse of the stiffness matrix and will be denoted also as

$$\| \psi_{ik} \| = \| u_{ik} \|^{-1} = \| u^{ik} \|. \tag{7.11}$$

In the next section we shall develop the general properties of the fundamental forms. Meanwhile, we consider their role in establishing contact between theory and experiment.

For the sake of simplicity, we confine the argument to one-component systems and interpret the variables as $x_1 = s$, $x_2 = v$, $P_1 = T$, $P_2 = -p$. We choose the scale factor as the mole number $X_3 = N$. Thus u, s, v are molar quantities. The fundamental equation is

$$u = u(s, v). \tag{7.12}$$

Its Legendre transform is

$$\psi(T, p) = -g(T, p) = Ts - pv - u, \tag{7.13}$$

where $g(T, p)$ is the molar Gibbs function. We shall refer to the formalism developed from the fundamental equation (14) as the T, p scheme. The equations of state in this scheme are

$$\frac{\partial \psi}{\partial(-p)} = \frac{\partial g}{\partial p} = v(p, T), \tag{7.14}$$

$$\frac{\partial \psi}{\partial T} = -\frac{\partial g}{\partial T} = s(p, T). \tag{7.15}$$

Equation (14) is the conventional thermal equation of state, and (15) is sometimes referred to as the caloric equation of state. The former is directly measurable,[17] and is tabulated for reference, but the latter is not, since caloric experiments yield only entropy differences.

The stiffness matrix is

$$\left\| \begin{matrix} \left(\dfrac{\partial T}{\partial s}\right)_v & \left(\dfrac{\partial T}{\partial v}\right)_s \\ -\left(\dfrac{\partial p}{\partial s}\right)_v & -\left(\dfrac{\partial p}{\partial v}\right)_s \end{matrix} \right\| = \left\| \begin{matrix} \dfrac{T}{c_v} & \left(\dfrac{\partial T}{\partial v}\right)_s \\ -\left(\dfrac{\partial p}{\partial s}\right)_v & \dfrac{1}{v} B_s \end{matrix} \right\|, \tag{7.16}$$

where B_s is the adiabatic bulk modulus.

---

[17] We assume here that the thermodynamic temperature is measurable. Actually, the measurement of T poses theoretical problems, which we shall briefly discuss later in this section.

The compliance matrix is

$$\left\| \begin{array}{cc} \left(\dfrac{\partial s}{\partial T}\right)_p & \left(\dfrac{\partial s}{\partial(-p)}\right)_T \\ \left(\dfrac{\partial v}{\partial T}\right)_p & \left(\dfrac{\partial v}{\partial(-p)}\right)_T \end{array} \right\| = \left\| \begin{array}{cc} \dfrac{c_p}{T} & v\alpha \\ v\alpha & v\kappa_T \end{array} \right\|, \tag{7.17}$$

where c_p, α, and κ_T are molar heat at constant pressure, expansion coefficient and isothermal compressibility, respectively.

The elements of the compliance matrix are directly measurable, and are also tabulated for reference. The caloric equation of state (15) and the fundamental equation (13) follow from the empirical quantities through direct integration. This procedure introduces two arbitrary integration constants into the fundamental equation. These are most conveniently taken as the entropy and enthalpy constants s_0 and $h_0 = g_0 + T_0 s_0$, where the subscript refers to a reference point T_0, p_0.

We recall that our discussion refers to the primitive fundamental equation of a single phase. The arbitrariness in the constants s_0, h_0 is reduced by the fact that the constants of different phases are related to each other by means of the latent heat:

$$l = T_0(s_0'' - s_0') = h_0'' - h_0'. \tag{7.18}$$

It is assumed that the two phases coexist at T_0. Similar relations hold for chemical species undergoing chemical reactions.

Within the context of classical molecular chemistry the foregoing procedure leaves the entropy and enthalpy constants of the chemical elements arbitrary. It would seem plausible to eliminate the arbitrariness of the entropy constant by using the strong formulation of the third law (*P c2*) and set $s_0 = 0$ for $T = 0$

Although quantum mechanics supports this procedure, there are some valid objections against it if one takes a purely empirical attitude. In the first place, absolute zero cannot be reached, and while the entropy of a system may practically vanish at attainable temperatures, there is no purely empirical method to verify this fact.

A second point is that there is no purely empirical method to establish whether or not a given system has actually reached equilibrium. This difficulty is particularly great at low temperatures, and would seem to render a fruitful application of *P c2* virtually impossible.

However, the problem of the entropy constant takes on a completely new aspect if one makes use of some simple results of quantum statistics.[18] We confine

---

[18] The injection of quantum statistical calculations seems to be against the traditional ground rules of the phenomenological theory. We believe that this objection is not serious. We shall return to this question in the final discussion.

ourselves to outlining the essential ideas involved. It is an important fact that in states of sufficiently low density, and not too low temperatures, practically every system breaks up into molecules that are, on the average, so far from each other that the system behaves as an ideal gas. Its equation of state is

$$\lim_{\rho\to 0} \frac{p}{T} \propto \rho. \tag{7.19}$$

Moreover, its entropy is the sum of the molecular entropies. Under these ideal conditions, the entropy can be computed from quantum statistics. With the exception of monatomic gases, this calculation is not based on first principles, but utilizes spectroscopic data, and its results are very reliable (*27*).

The calculated entropy contains no arbitrary constant, but it depends on the choice of independent components in the sense of *C a3*. With the entropy value pegged in the low-density limit, say at room temperature, the empirical data can be used to integrate to the lowest available temperatures. If the resulting entropy value is practically zero, we conclude that: (i) the system is in equilibrium and (ii) the temperature is low enough for the system to behave almost as it would at absolute zero. A finite positive entropy indicates that at least one of these conditions is not satisfied. It usually takes a combined experimental-theoretical study to separate the role of the two conditions (*15*).

To arrive at a negative entropy value is unusual; nevertheless, it is conceivable that the variables of the problem are being chosen in an inadequate fashion.[19]

The most interesting aspect of this result is that only through the detour over quantum statistics is the concept of thermodynamic equilibrium placed on a solid empirical basis.

Summing up we can say that the fundamental equation can be established from actual measurements supplemented by quantum statistical calculations carried out in the ideal gas limit.

The fundamental equations (12) and (13) are not the only ones available. Other equations are the Helmholtz free energy as a function of T, V, and the enthalpy as a function of S, p. Still more possibilities arise if the volume is chosen as a scale factor. Since all of these schemes are equivalent, we arrive necessarily at a huge number of thermodynamic identities. These identities have important uses. One of them is to express quantities of interest in terms of the compliance coefficients or other measurable quantities of the p, T scheme.

[19] Thus it is usual to ignore the contribution of nucear spins to the entropy, since in most cases this contribution does not diminish in the available temperature range. If further cooling leads to a freezing out of the spin entropy, one might arrive at negative entropies. This result is, of course, spurious and is eliminated by including the relevant type of entropy in the statistical calculation.

We do not enter into the derivation of these identities in this paper, but merely note that they allow us to short-circuit the explicit construction of the fundamental equations (14) and (13).

A second important use of the thermodynamic identities is the establishment of the thermodynamic temperature scale. From the empirical point of view, a temperature scale is based on the properties of a particular substance. Let the temperature defined in such a fashion be T^*, and the thermodynamic temperature calibrated on this empirical scale be given in terms of a function $T(T^*)$. Let us consider, now, a thermodynamic identity that involves only measurable quantities. Since thermodynamic identities are valid only in terms of the thermodynamic temperature, any identity of this sort constitutes a differential equation for $T(T^*)$. This equation can be integrated if a numerical value is assigned, by convention, to a fixed point.

According to the first phase rule, the triple point of a one component system is an isolated point in P-space ($c = 1$, $\tilde{m} = 3$, $\delta_I = 0$) and may serve as a fixed point. Recent international agreement has fixed the triple point of water as exactly 273.16°. The fact that different identities and different substances lead to the same temperature scales is a further verification of the theory.

VIII. LOCAL STABILITY

We proceed now to investigate the stability of homogeneous modifications by using the criteria of stability formulated in Section IV.

We mentally subdivide the system into two parts specified by the fixed scale factors X'_{r+1} and X''_{r+1}, and consider the virtual process

$$\delta X_i' + \delta X_i'' = 0, \qquad i = 1, 2, \cdots r. \tag{8.1}$$

The virtual work associated with this displacement is

$$\begin{aligned} \Delta W = \Delta U' + \Delta U'' &= \sum_1^r P_i' \delta X_i' + \frac{1}{2} \sum_1^r U'_{ik} \delta X_i' \delta X_k' \\ &+ \cdots + \sum_1^r P_i'' \delta X_i'' + \frac{1}{2} \sum_1^r U''_{ik} \delta X_i'' \delta X_k'' + \cdots . \end{aligned} \tag{8.2}$$

We have made use of the expansion (7.1), but rewritten it in terms of the extensive variables instead of the densities. The stiffness matrix is now

$$U'_{ik} = (\partial^2 U' / \partial X_i' \partial X_k')_0 \tag{8.3}$$

that is a homogeneous function of order (-1).

We let size of one of the systems tend to infinity:

$$X''_{r+1} / X'_{r+1} \rightarrow \infty . \tag{8.4}$$

Since $u'_{ik} = u''_{ik}$, we see from (1)–(4) and (7.4) that the quadratic term associated with the large system is negligible. This system can be considered as a reservoir that is adequately described solely in terms of its intensities $P_i'' = P_{i0}$.

We obtain from (2), after dividing by X_{r+1}', omitting the primes and using the notation (7.2):

$$\Delta w = \sum_1^r (P_i - P_{i0})\xi_i + \frac{1}{2}\sum_1^r u_{ik}\xi_i\xi_k + \cdots . \tag{8.5}$$

The condition of equilibrium provides $P_i = P_{i0}$, a result already obtained in Section V. According to *D f2* (Section IV), stability requires $\Delta w \geqq 0$.

The implications of this relation have to be discussed both for infinitesimal and finite displacements. The second question concerning *stability in the large* is taken up in the next section. Meanwhile, we turn to the discussion of infinitesimal displacements (local stability), determined primarily by the quadratic form (5).

It is shown in Appendix B (See also Ref. *11*), that this quadratic form can be brought to a diagonal form

$$\Delta w = \sum_{k=1}^r \lambda_k \eta_k^2, \tag{8.6}$$

where the η_k are linear combination of the ξ_k explicitly defined in Appendix C and

$$\lambda_k = \left(\frac{\partial P_k}{\partial x_k}\right)_{P_1,P_2,\cdots,P_{k-1},x_{k+1},\cdots,x_r} \equiv \left(\frac{\partial P_k}{\partial x_k}\right)_{P_1,P_2\cdots,P_{k-1}} . \tag{8.7}$$

The second expression is used as a short notation, in which only the intensities held constant are made explicit, while the constant densities are suppressed.

The diagonalization of the form (5) provided by (6) and (7) replaces the coupled processes $\delta X_1, \delta X_2, \cdots, \delta X_r$ by independent processes

$$\delta X_1,\ (\delta X_2)_{P_1},\ldots(\delta X_r)_{P_1,P_2,\cdots P_{r-1}} . \tag{8.8}$$

This interpretation of the diagonalization procedure is indeed suggested by (7); it is confirmed by a formal proof, since it follows from (C5), (C11), and (7.5) by a simple algebra, that η_i with $\eta_k = 0$ for $k \neq i$ is equivalent to $(\delta x_i)_{P_1P_2\cdots P_{i-1}}$.

Of course, the processes (8) involve implicitly a coupling of the δx_i. Thus for $(\delta x_2)_{P_1}$ we have

$$u_{11}\delta x_1 + u_{12}\delta x_2 = \delta P_1 = 0. \tag{8.9}$$

In each subsequent process one more extensive variable is involved. The substitution of independent basic processes for the coupled processes of the original problem is, of course, reminiscent of the diagonalization of quadratic forms by the eigenvalue method. However, it is evident from Appendix C that the two methods are substantially different.

Since the expansion (5) can be centered around any point of the primitive surface, we can classify the points of this surface as follows:

(a) Elliptic points: all $\lambda_k > 0$,

(b) Parabolic points: all $\lambda_k \geqq 0$, at least one $\lambda_k = 0$, (8.10)

(c) Hyperbolic points: at least one $\lambda_k < 0$.

The application of the definitions *D f1-4* (Section IV) allows us to draw the following conclusions. Elliptic points are stable with respect to small displacements (local stability), but may be either stable or instable with respect to finite processes (absolute stability and metastability, respectively).

Hyperbolic points represent states of essential instability. They are, in general, inaccessible both from the experimental and the theoretical points of view, and we shall always exclude them from consideration.[20]

The case of parabolic points is the most complex and also the most interesting. Whether or not parabolic points are stable depends on the large displacements which will be discussed in the next section.

While most parabolic points turn out to be instable, there are stable limiting situations. Yet, the stability of parabolic points is of a lower order than the one found in stable elliptic points. We refer to the latter as *normal*, to the former as *critical equilibrium* (*D f2*).

We shall compare now the properties of normal and critical equilibrium to the extent that this can be established from the properties of the fundamental quadratic form.

For the sake of simplicity, we consider the case of two independent variables, and elaborate on the foregoing conclusions. An elliptic point is characterized by the conditions

$$\lambda_1 = (\partial P_1/\partial x_1)_{x_2} > 0, \tag{8.11a}$$

and

$$\lambda_2 = (\partial P_2/\partial x_2)_{P_1} > 0. \tag{8.11b}$$

According to (*B*-15) we have

$$\lambda_2 = \frac{D_2}{D_1} = \left(\frac{\partial P_2}{\partial x_2}\right)_{x_1} - \frac{(\partial P_1/\partial x_2)^2_{x_1}}{(\partial P_1/\partial x_1)_{x_2}} > 0. \tag{8.12}$$

Hence

$$(\partial P_2/\partial x_2)_{x_1} \geqq (\partial P_2/\partial x_2)_{P_1} > 0, \tag{8.13}$$

---

[20] A widely used procedure involving the essentially instable states of a Van der Waals gas will be critically discussed in the next section. From the mathematical point of view, points with all of the $\lambda_k < 0$ have an elliptic character. In the present context these are essentially instable, as are hyperbolic points proper with positive and negative λ_k.

where the equality sign is relevant only if

$$(\partial P_1/\partial x_2)_{x_1} = 0. \tag{8.14}$$

Relations (13) are equivalent to (11a) and (11b). They are usually referred to as the *principle of Le Chatelier*. For a further discussion we refer to p. 63 of Ref. *20*.

For a system with the fundamental equation (7.12) we may choose $x_1 = v$, $x_2 = s$, or $x_1 = s$, $x_2 = v$. The relations (13) yield, in this case,

$$\frac{1}{c_v} \geqq \frac{1}{c_p} > 0, \tag{8.15}$$

and

$$B_s \geqq B_T > 0, \tag{8.16}$$

where B is the bulk modulus. The equivalence of (15) and (16) follows at once from the thermodynamic identities

$$\frac{B_s}{B_T} = \frac{c_p}{c_v}, c_p - c_v = Tv\left(\frac{\partial p}{\partial T}\right)^2_{v/B_T}. \tag{8.17}$$

The equality sign in (15) and (16) becomes relevant only if the expansion coefficient vanishes.

The foregoing results are easily generalized to several variables. In elliptic points we have the generalized Le Chatelier principle:

$$\frac{\partial P_k}{\partial x_k} \geqq \left(\frac{\partial P_k}{\partial x_k}\right)_{P_1} \geqq \left(\frac{\partial P_k}{\partial x_k}\right)_{P_1 P_2} \geqq \cdots \geqq \left(\frac{\partial P_k}{\partial x_k}\right)_{P_1, P_2, \cdots P_{k-1}} \geqq 0. \tag{8.18}$$

We express this relation in words as follows: A locally stable system is displaced from its equilibrium by the displacement dx_k. The system responds by changing its conjugate intensity by dP_k. This response is the largest if all the other x_i are fixed, and it decreases upon relaxation of each constraint that frees a variable x_i by coupling the system to a reservoir of intensity P_i.

We turn now to the discussion of critical equilibrium, when the stiffness matrix is singular and is of rank $(r-1)$[21]:

$$D_r = 0, \qquad \lambda_r = (\partial P_r/\partial x_r)_{P_1, P_1, \cdots P_{r-1}} = 0, \qquad \lambda_k \neq 0 \quad \text{for} \quad k \neq r. \tag{8.19}$$

The development of the thermostatic formalism made it repeatedly necessary to require that the condition $D_r \neq 0$ be satisfied. The breakdown of this condition in critical equilibrium indicates that in such states the thermostatic formalism reaches its limits of validity. An instance of this breakdown becomes evident from the discussion of Eqs. (7.5) which, in general are no longer soluble,

---

[21] This is the only case that has been found to be of physical interest, thus far.

while the associated homogeneous equation

$$\sum_{k=1}^{r} u_{ik}\delta x_k = 0 \tag{8.20}$$

has a nontrivial solution

$$\delta x_1 : \delta x_2 : \cdots \delta x_r = u^{1k} : u^{2k} : \cdots u^{rk}, \tag{8.21}$$

where the u^{ik} are proportional to the $(r - 1)$ dimensional minors of $\| u_{ik} \|$, and may be taken as the elements of the compliance matrix. Since the rank of this matrix is $(r - 1)$, the u^{ik} cannot all vanish and (21) defines a direction, the *critical displacement,* in Gibbs space (*D f2*).

The inhomogenous equations are soluble only if the condition

$$\sum_{i=1}^{r} u^{ik}\delta P_i = 0 \tag{8.22}$$

is satisfied. Hence, the transformation (7.5) maps an r-dimensional domain of x-space to an $(r - 1)$-dimensional domain of P-space. Consequently, the Legendre transformation becomes singular here. The implications of this fact will be discussed in the next section.

The most conspicuous symptom of critical equilibrium is that the compliance matrix becomes singular, its elements in general tend to infinity. In the system (7.12) this means that

$$c_p \longrightarrow \infty, \quad \kappa_T \longrightarrow \infty, \quad \alpha \longrightarrow \infty. \tag{8.23}$$

We note that c_v and κ_s are finite. The assumption $c_v \longrightarrow \infty$ would imply that the rank of the stiffness matrix is $(r - 2)$, that is $\lambda_r = \lambda_{r-1} = 0$.[21, 22] However, the parabolic character of critical equilibrium does not necessarily imply (23). If the minor $u^{ik} \approx 0$, the corresponding compliance coefficients will not tend to infinity. The displacive transitions discussed in the next section belong in this category (see Ref. *11*).

IX. STABILITY IN THE LARGE. SECOND PHASE RULE

We turn to the second part of our study of stability and consider processes in which the densities of a homogeneous system change by finite amounts.

One method of dealing with this problem is to consider the higher order terms of the expansion (8.5) of the fundamental equation.

A second approach, first developed by Gibbs, deals with the properties of the primitive surfaces in the large by geometrical methods.

---

[22] The conclusion that c_v is finite seems to be at variance with the exact theoretical findings concerning the critical point of the two-dimensional Ising model. This contradiction is only apparent; in the Ising model there is no distinction between c_v and c_p.

The first method alone is insufficient for our present problem. Usually, the two methods are used in combination (*2*, *4*, *28*). We shall show, however, that the purely geometrical method is by itself sufficient to establish the second phase rule, our main objective in this section.

While the expansion method has its use, it is unreliable in the most interesting case: in critical equilibrium.

If a simple system is prepared with such values of the extensive variables that correspond to an instable region of a phase surface, the system escapes unstability by breaking up into a heterogeneous system consisting of two or more modifications. These modifications may belong to the same phase surface or be on different phase surfaces. The vaporization and the crystallization of a liquid, furnish examples for the two situations.

We shall consider in detail the instability that arises within a single primitive surface. Of the two cases this one is richer in variants, and the results are easily extended to the instability involving several surfaces.

Heterogeneous equilibria of two modifications of the same phase occur under two kinds of conditions: (i) the coexisting modifications differ in their densities and, (ii) the modifications have equal densities, as ensured by symmetry relations, and are distinguished in terms of quasithermodynamic parameters η. We shall refer to the first case as a *condensation type equilibrium*, to the second as an *equilibrium of modification doublets* (or multiplets).[23]

We shall, at first, give a brief account of the Gibbs construction for condensation type equilibria (p. 44 of Ref. *2*) partly because this construction is not easily accessible (see, however, Chapters 2 and 3 of Ref. *4* and Chapter 12 of Ref. *28*), and partly because we wish to develop the analogy between the two cases.

The requirements of geometrical representation suggest that the simplest case with the smallest number of variables, be discussed first. The generalization to several variables offers no difficulties.

The simplest condensation type transition is the condensation of a one-component fluid with $c = 1$, $r = 2$. The primitive fundamental equation in molar quantities is

$$u = u(x_1, x_2) \tag{9.1}$$

where $x_1 = s$ and $x_2 = v$. We shall assume that the primitive surface contains a connected sheet of elliptic points bounded by a one-dimensional locus of para-

---

[23] Conforming to the time-independent character of thermostatics, we usually speak of phase equilibrium. The widely used terms condensation and phase transition have a temporal connotation: The extensive or intensive variables of a system are varied as a function of time and the system is assumed to run through a corresponding sequence of equilibrium states. The distinction between these concepts is of some importance, since the same heterogeneous equilibrium may be crossed in the course of different transitions produced by the variation of different variables.

bolic points.[24] Empirically, the phase surface of every fluid turns out to be of this sort.

In the elliptic domain the u surface is convex toward the x_1, x_2, plane. Outside, it is concave. According to the Gibbs construction the concave part of the surface is replaced by a ruled surface consisting of infinitesimal strips of tangent planes each of which has two common points with the u surface.

A geometrical representation is given in Fig. 2.

We single out one tangent plane with the tangent points A', A''. These so-called conjugate points represent distinct modifications of the same phase which have the same intensities, and hence can coexist. The line $A'A''$ is an isothermal, isobaric line. The projections of A', A'' on the x_1 - x_2 plane are $\tilde{A}'$, $\tilde{A}''$. Figure 2(a) represents the intersection with the plane x', u. The parabolic points contained in this plane are B', B''. Their projections are again marked by a tilde. The loci of the points A', A'' and B', B'', respectively, are called the *binodal* and the *spinodal* curve (Ref. *4*). The former marks the limit of absolute stability, the latter the limit of essential instability. The region between the two curves is that of metastability.

In constructing the derived fundamental equation, the portion $A'B'B''A''$ of the primitive surface is replaced by the double tangent $A'A''$. As A' and A'' move along the binodal curve, the tangent lines produce a ruled surface. The point E of this surface represents a heterogeneous state consisting of the modifications A' and A''. The relative amount of these modifications is $EA'':A'E$ (lever rule).

The heterogeneous equilibrium has a discontinuous character that manifests itself in discontinuities of the molar quantities $\Delta x = x_i'' - x_i'$. In Ehrenfest's terminology (*9*) we have a first order transition.

The discontinuous aspect of the equilibrium is particularly evident if it is considered as a transition in the sense explained in footnote 23, provided that the transition is performed by varying the intensities. In the P-space representation (i.e., in the p-T diagram for the system (1)), the entire segment $A'A''$, and hence the modifications A' and A'' are represented by a single point. The conjugate modifications are locally stable, and no intrinsic property foreshadows the fact that an infinitesimal variation of the intensities produces an abrupt transition into a modification, the densities of which differ by the finite amounts Δx_i from those of the original modification.

However, the basically discontinuous character of the heterogeneous equilibrium is masked by a spurious continuity if the transition is performed by a variation of the extensive variables at constant intensities. In such a process

---

[24] Phases consisting exclusively of elliptic points do not exhibit the intrinsic instabilities that will be discussed. Surfaces containing disjoint sheets of elliptic points can be considered as consisting of two primitive surfaces. This is the alternative mentioned, and set aside for later discussion, at the beginning of this section.

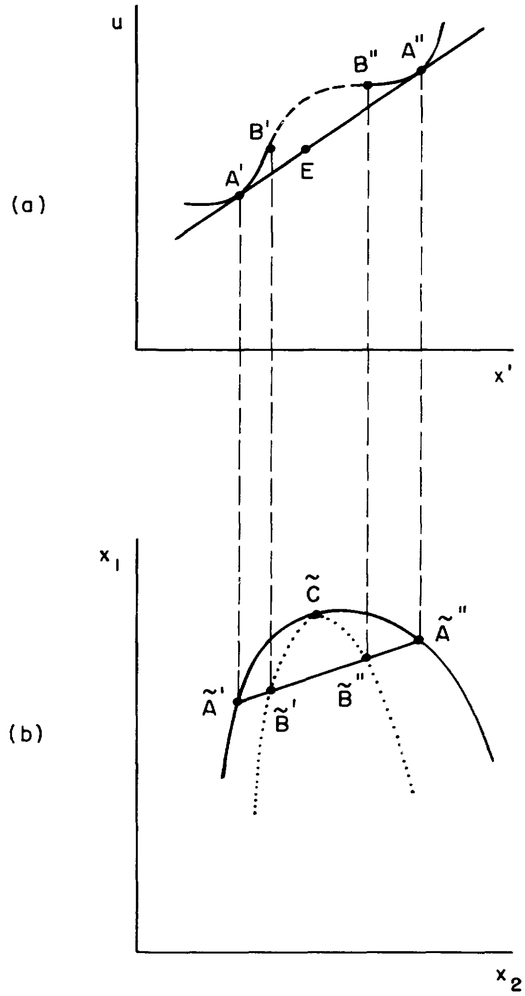


FIG. 2. Critical point of the first kind in terms of the fundamental equation $u = u(x_1, x_2)$; (a) intersection of the primitive surface with the $u - x'$ plane; the axis x' connects points A' and A''. (b) projection to the $x_1 - x_2$ plane; $\tilde{A}'\tilde{C}\tilde{A}''$ and $\tilde{B}'\tilde{C}\tilde{B}''$ are projections of the binodal and spinodal curves; $\tilde{C}$ is the projection of the critical point. The scales along x' in (a) and (b) are not the same.

the point E representing the system moves from A' to A'', while the extensive variables $X_i = N'x_i' + (N - N')x_i''$ change continuously as N' decreases from N to zero.[25]

---

[25] Thus the condensation of an ideal Bose-Einstein gas is a first-order transition. If it is cooled at a constant pressure $(dT)_p$ the system collapses and its volume vanishes at the

The region between the binodal and spinodal curves is metastable. Part of this region is observable: supersaturated vapor, superheated and supercooled liquid are well-known instances of this sort. However, in general, observation does not extend to the parabolic points that form the boundary between the metastable and the instable domains. We have seen in Section VIII that in parabolic points critical displacements are, in first approximation, not opposed by thermodynamic restoring forces. Such a displacement would move a point B' or B'' into the essentially instable region.[26]

This argument does not hold for parabolic points that are on the border between the absolutely stable and the instable domain, for example, point $\tilde{C}$ in Fig. 2(b). Such states are stable, and are indeed observed. Designated as critical points in the last section, they are seen to have two properties: (i) Critical points are stable parabolic points in which the determinant of the stiffness matrix vanishes and the elements of the compliance matrix are infinite. They are at the limit of the region of stability. (ii) Critical points are at the end of the coexistence region in P-space (coexistence line in the p-T diagram), where two modifications of different densities become identical.

We claim that, property (i) implies property (ii), since, passing through a stable parabolic point, a system can escape instability only by breaking up into coexisting modifications.[27]

We propose to establish an upper bound for the dimension of the set of stable parabolic points in P-space. According to the result obtained, we can substitute in this problem property (ii) for property (i). The problem transformed in such a way can be readily solved.

Since the coexistence region of a one-component, two-modification, system is one-dimensional and the end of this line an isolated point, we see that critical points can form only a zero-dimensional set (isolated points) in the p-T diagram.[28]

This result can be immediately generalized to systems with the fundamental

---

condensation temperature (see p. 52 of Ref. *24*). The same result is obtained for $(dp)_T$. The fact that the transition is continuous for $(dV)_T$ and $(dT)_v$ does not warrant the designation of the BE condensation as a third-order transition. By the same token we ought to consider the van der Waals type condensation as a continuous rather than as a first order transition.

[26] This conclusion concerning the instability of parabolic points is confirmed by the discussion of the higher terms in (8.5) (see p. 259 of Ref. *20*). A brief discussion is also given at the end of this section.

[27] It is usually assumed or deduced that the converse statement also holds, and property (ii) implies property (i). (see pp. 114 and 129 of Ref. *2* and pp. 79–81 of Ref. *20*). Although certainly correct for a wide variety of cases, we prefer to leave the limits of validity of this statement in abeyance. This question is irrelevant for the argument of this paper.

[28] No statement of such generality can be made about the points in x-space, since the contact between the binodal and the spinodal curves may extend over a one dimensional region in the space of densities (*29*).

equation $u = u(x_1, x_2, \cdots, x_r)$, where the variables are all densities. Homogeneous modifications of such a system can exist in an r-dimensional manifold. Along this surface the heterogeneous equilibrium has the same abrupt, discontinuous character as described above for the one-component system.[29]

The edge of the region of coexistence is $(r - 2)$-dimensional and it is to this edge that critical equilibrium is restricted. Thus, for the dimension of the manifold of critical points of the condensation type, we obtain

$$0 \leqq \delta_c \leqq r - 2 = c - 1. \tag{9.2}$$

This result was first obtained by Gibbs. We shall refer to (2) as an instance of the *second phase rule.*

We arrive at a more general form of this rule if we make use of the new definition of phase (*P d1*) and repeat the foregoing construction for the case of modification doublets, that are distinguished from each other only in terms of the quasi-thermodynamic variable η. The fundamental equation of a single-component system is now

$$u = u(x_1, x_2, \eta), \tag{9.3}$$

where $x_1 = s$, $x_2 = v$ and the number of particles is selected as a scale factor.

For the purposes of presentation it is convenient to start the discussion in terms of the well known Ising model of ferromagnetism (*30*). The external magnetic field is assumed to vanish. We shall see that the generality of our final results will not be affected by the special features of the model. One of these features is that the properties of the system are volume independent. Thus the fundamental equation reduces to

$$u = u(x_1, \eta). \tag{9.4}$$

The quasi-thermodynamic parameter η is discussed in some detail in Section X and Appendix *A*. Meanwhile, it is satisfactory to think of it as the parameter of long-range order, or as the magnetization of the system. The interactions arise only between nearest neighbors, and, among other things, the long-range dipole-dipole interactions are neglected. We assume, also, the absence of an external magnetic field. In contrast to a true ferromagnet, the Ising model satisfies our postulates with $r = 1$. According to (1.7) we should call this a "zero-component" system, an expression of the idealized character of the model. There is no objection against its use in thermodynamics; another zero-component system is, e.g., a volume filled with blackbody radiation.

---

[29] A fluid mixture of two components is an example of this sort. As explained above, the discontinuity of the transition becomes apparent in P-space, or in p, T, μ space in the special example considered. It is more usual to plot the states of the system in the p, T, c space, where c is the concentration. Thus a spurious continuity is introduced into the transition and the system boils, and freezes in extended ranges of temperatures.

We assume that the function (4) is defined also for nonequilibrium values of η. The equilibrium values of this parameter are determined from

$$\frac{\partial u}{\partial \eta} = 0, \qquad \frac{\partial^2 u}{\partial \eta^2} > 0. \tag{9.5}$$

In the Ising model, u is a function of η^2, hence the nonvanishing solutions of (5) appear in pairs: η and $-\eta$. This situation is indeed realized for sufficiently low entropy, or temperature.

Equation (4) is represented Fig. 3, the functions $\pm\eta(x_1)$ are the two solid lines $\tilde{C}\tilde{A}'$ and $\tilde{C}\tilde{A}''$ in Fig. 3. These, again, are binodal curves, the points A', A'' represent *doublet modifications* (see Section VI), referred to also as domains.

The $\tilde{B}'\tilde{C}\tilde{B}''$ curve is the locus of the parabolic points for which $\partial^2 u/\partial \eta^2 = 0$. The analogy between Figs. 2 and 3 is apparent. In general, the parabolic points are again instable, and, at best, the point C, and possibly another intersection with the $\eta = 0$ axis, will correspond to a stable order-disorder type of critical point.[30] This point is the limit between states which do and which do not exhibit a domain structure of doublet modifications.

It is evident that the construction of the critical point point depends only on the condition that the minimum problem (5) have at least two (or possibly more) branches of energetically equivalent solutions. Otherwise, it is entirely independent of the physical interpretation of the parameter η.

These results can be easily generalized to thermodynamic system of r degrees of freedom with the fundamental equation

$$u = u(x_1, x_2, \cdots, x_r, \eta). \tag{9.6}$$

This surface represents a modification that is stable in an r-dimensional domain of x-space, or alternatively, in an r-dimensional domain of P-space. We assume that this modification is characterized by a non-vanishing value of η. The modification coexists over its entire r-dimensional range of stability with the modification specified by $-\eta$, or with any other modification that is related to (6) by symmetry. We refer to the coexisting modifications as domains. The critical equilibria associated with the disappearance of the domain structure are at the $(r - 1)$-dimensional edge of the r-dimensional domain in P-space. Thus the dimensionality of order-disorder type critical point is

$$0 \leqq \delta_{0\text{-}d} \leqq r - 1 = c. \tag{9.7}$$

Equations (2) and (7) can be expressed in terms of a single equation as

$$0 \leqq \delta_{II} \leqq r - 2 + \sigma = c - 1 + \sigma, \tag{9.8}$$

---

[30] Note the logic of the argument. We do not attempt to predict that C has to be a critical point, although this may well be true. What turns out to be of importance is to set an upper limit to the set of stable parabolic points.

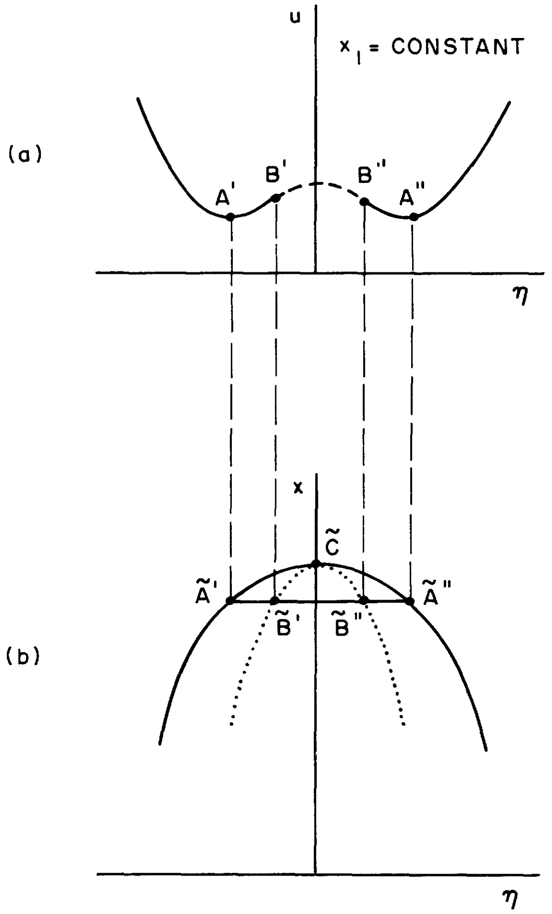


FIG. 3. Critical point of the second kind in terms of the fundamental equation $u = u(x_1, \eta)$; (a) intersection of the primitive surface with the $u - \eta$ plane. (b) projection to the $x - \eta$ plane; $\tilde{A}'\tilde{C}\tilde{A}''$ and $\tilde{B}'\tilde{C}\tilde{B}$ are projections of the binodal and spinodal curves, $\tilde{C}$ is the projection of the critical point.

where the "symmetry number"

$$\sigma = \begin{array}{ll} 0 \text{ if } & x' \neq x'', \\ 1 \text{ if } & \eta' = -\eta'' \end{array} \qquad (9.9)$$

Or, in words, σ is zero for the condensation type, and unity for the order-disorder type of transitions. We shall refer to these cases also as *critical points of the first and second kind*, respectively.

The meaning of the second phase rule (9) is not to predict critical behavior in a given instance, but to provide an upper bound for the dimension of the P-space manifold in which such singular behavior can arise.

In practice, critical equilibria are readily recognizable experimentally as λ-points. Thus the number δ_{II} will be observed experimentally, and (9) indicates the system of minimum complexity measured in terms of r and σ, in which the observed singularities can, indeed, occur.

It is evident from Figs. 2 and 3 that the geometrical representation of an isolated critical point ($\delta_{II} = 0$) calls for a three dimensional Gibbs space: the critical phenomenon arises out of the interplay of processes described in terms of two independent variables such as s and v, where one variable may be replaced by a quasi-thermodynamic parameter. The dimensionality of the space increases and the representation becomes impractical, for systems with more than one chemical component. This situation cannot be simplified by keeping certain variables x_i constant. Thus along the abscissa x' in Fig. 3a, in general, all the $x_1, x_2, \cdots, x_r$ vary.

Nevertheless, the diagonalization (8.6) of the fundamental form suggests a representation of critical equilibrium, in terms of a well-chosen one-variable process. This is indeed possible.

Consider the Legendre transform

$$-\varphi(P_1, P_2, \cdots, P_{r-1}, x_r) = \sum_{i=1}^{r-1} P_i x_i - u. \tag{9.10}$$

We have

$$\lambda_r = \frac{\partial^2 \varphi}{\partial x_r^2} = \left(\frac{\partial P_r}{\partial x_r}\right)_{P_1 P_2 \cdots P_{r-1}}. \tag{9.11}$$

Assuming that $\varphi(x_r)$ is sufficiently regular, to admit an expansion into a power series at the critical point, we obtain the following conditions for critical (stable) equilibrium:

$$\text{(a)} \qquad \frac{\partial^2 \varphi}{\partial x_r^2} = 0,$$

$$\text{(b)} \qquad \frac{\partial^3 \varphi}{\partial x_r^3} = 0, \tag{9.12}$$

and

$$\text{(c)} \qquad \frac{\partial^4 \varphi}{\partial x_r^4} > 0.$$

The situation is geometrically represented in Fig. 4, in terms of a one parameter family of curves $\varphi_{P_1}(x_r)$ where each curve corresponds to a different

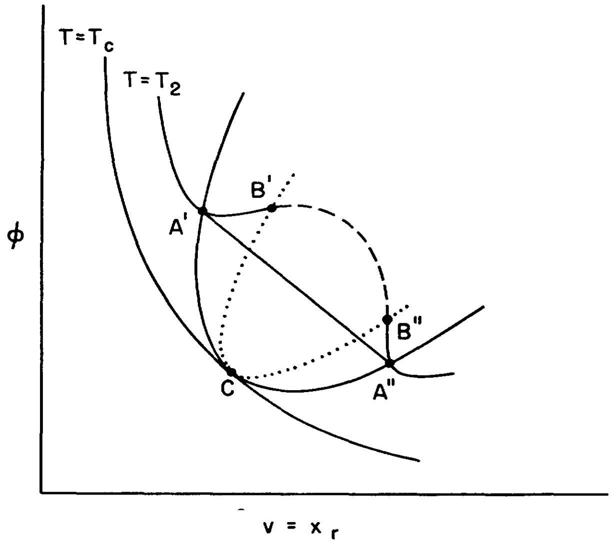


FIG. 4. Critical region in terms of the fundamental equation $\varphi = u[P_1, P_2, \cdots, P_{r-1}]$; curves of constant $P_1, P_2, \cdots, P_{r-1}$ plotted against x_r. $P_1 = T$, $P_2 = p$.

constant value of P_1, conveniently taken to be the temperature or pressure. We have chosen the slope of the curves negative, in order to apply it to the case of ordinary condensation with $x_r = v$, $P_r = -p$.

The presence of condition (12b) supports our earlier conclusion that the dimension of the set of critical points is by one less than that of the parabolic points defined by (12a) only.

For order-disorder type critical points the variable x_r in (10)–(12) replaced by η. In this case condition (12b) is satisfied by symmetry and the dimension of the set of critical points is increased by one, in accordance with (9).

However, it is worth noting that the regularity assumption inherent in the formulation of conditions (12) is not necessarily justified. Thus it has been suggested by Zimm (*29*) that at the ordinary condensation type critical point

$$(\partial^n p/\partial v^n)_T = 0 \tag{9.13}$$

for all values of n.

If this interpretation of the experimental facts were justified, conditions (12) would be inapplicable. However, there is no objection to the use of Fig. 4 in terms of the correct empirical functions, and the second phase rule is unaffected.

The geometrical representation of the condensation phenomenon that is most commonly used in thermodynamics texts arises through the differentiation of the curves of Fig. 4. The result is represented in Fig. 5. This method operates

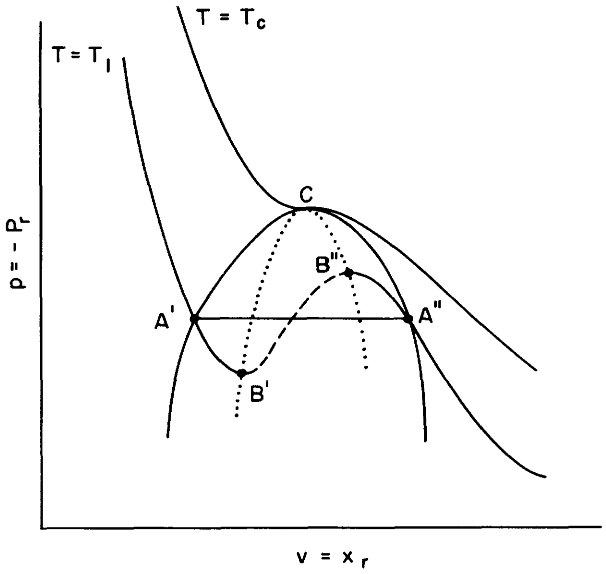


FIG. 5. Critical region in terms of the curves $-P_r = -P_r(x_r)$ at constant $P_1, P_2, \cdots, P_{r-1}$; $P_1 = T$.

in terms of the equation of state, and is ubiquitous in its application to the van der Waals gas.

The comparison of the two geometrical representations provides a graphic illustration for the superiority of the method of fundamental equations.

In Fig. 5, the conjugate points A' and A'' are located by means of the well known Maxwell rule of equal areas, the application of which involves an integration along the essentially instable branch $B'B''$ of the isothermal section of the equation of state. This procedure is objectionable because the instable region is not accessible to observation. Moreover, the instable branch of the isothermal cannot be considered as a legitimate theoretical construct, since a rigorous evaluation of the partition sum leads automatically to the absolutely stable heterogeneous states.

As we have seen, the Gibbs construction of the ruled surface, "boarding up" instable regions, relies only on the knowledge of the absolutely stable regions of the phase surface. The difference between the two cases stems from the fact that the equation of state contains less information than the fundamental equation. Thus a "van der Waals gas" does not constitute a fully defined thermodynamic system. A complete definition of the system would include, the specific heat as a function of, say, temperature and volume:

$$c_v = c_v(T, v). \tag{9.14}$$

With the aid of this information, the conjugate points A' and A'' can be identified by integrating around the critical point and always staying in the absolutely stable domain. In the concept of a "van der Waals gas" a spurious interpolation through the instable range is substituted for the missing information (14).

We turn now to the discussion of the equilibrium of modifications belonging to different phases. It is evident that the Gibbs construction of the ruled surface can also be carried out in this case to identify the conjugate points. The corresponding densities are, in general, different from each other. The modifications never become identical and we have no critical equilibrium.

We mention also the case of phase doublets. Their primitive surfaces are identical to each other, insofar as the dependence on densities is concerned, and they are distinguished only in terms of quasi-thermodynamic parameters. Phase doublets are related to each other by inversion or mirroring, since two systems equivalent under a pure rotation are considered to be identical to each other. A modification of such a phase automatically coexists with its mirrored modification. In contrast to the case of modification doublets, phase doublets cannot gradually become identical to each other, and do not give rise to critical points. An example will be discussed in the next section.

Up to this point we have considered critical equilibrium as a singularity of the behavior of a thermodynamic system. Actually, however, the critical phenomenon constitutes a singularity in the structure of the theory itself.

This is evident from the fact that the condition of validity $D_r \neq 0$ (see (5.14)) of the Legendre transformation leading from the variables $x_1, x_2, \cdots, x_r$ to $P_1, P_2, \cdots, P_r$ breaks down at the critical point. We may add that the condition of validity $D_r^{-1} \neq 0$ (see 5.17) of the inverse transformation fails in the vicinity of absolute zero.

While under normal conditions the x-space and the P-space representations of the fundamental equation provide an equivalent description of the thermodynamic system, the P-space representation contains less information at the critical point and the x-space representation provides less information near absolute zero.

The significance of this fact is best discussed in connection with the minimum principle (5.20). The conditions for an extremum are

$$\frac{\partial w(x'; P)}{\partial x_i'} \equiv \frac{\partial u(x')}{\partial x_i'} - P_i \equiv P_i' - P_i = 0, \tag{9.15}$$

$$\frac{\partial w(x'; P)}{\partial P_i} \equiv \frac{\partial \psi(P)}{\partial P_i} - x_i' \equiv x_i - x_i' = 0 \tag{9.16}$$

In such stable regular states in which the system is in a homogeneous modification these equations have a unique solution. The assumption that (15) is satisfied in distinct states, say x_i' and x_i'', implies that P_i' vanishes along the connecting line in x-space. Hence $D_r = 0$, a contradiction against the assumption of regularity.

Thus we have arrived at a result that, in a way, can be considered as the fundamental theorem of thermostatics. *If a system in a regular state is in equilibrium with its surroundings, the intensities of the surroundings uniquely determine the densities of the system and vice versa.* We propose to call this theorem the *principle of thermostatic determinism.* To motivate this terminology, we point out that the aforementioned extremum problems relate to two main types of thermodynamic investigations.

In the first type of investigation we predict and measure the properties (densities) of a system as a function of the intensities of the surroundings.

In the second kind of investigation we measure the system densities and infer from them the intensities of the surroundings. If the intensity is the temperature, the system plays the role of a thermometer.

The principle of thermostatic determinism assures us that in regular states both procedures lead to a unique answer.

We turn now to the discussion of the singular cases and consider in particular $D_r = 0$. According to Appendix B this condition implies that

$$(\partial P_r/\partial x_r)_{P_1,P_2,\cdots,P_{r-1}} = 0. \qquad (9.17)$$

The Legendre transformation maps an r-dimensional domain of x-space into an $(r - 1)$-dimensional domain of P-space and Eq. (15) cannot be uniquely solved for the x_i'.

In order to clarify the physical meaning of condition (17) we have to keep in mind that this equation applies to two essentially different situations: (i) A heterogeneous equilibrium of two or more modifications, each of which is described in terms of elliptic points of the primitive fundamental equation. (ii) A critical equilibrium corresponding to a stable parabolic point.

In case (i) the failure of thermostatic determinism is only apparent, and the principle can be "saved" by a minor improvement of its formulations. We consider, for simplicity, a one-component system that is, under regular conditions, adequately described in terms of the fundamental equation $G = G(T, p, N) = U[T, p]$. Given T, p and the scale factor N, the extensive variables X_i (i.e., S and V) and hence also the molar quantities s, v are uniquely determined. This is no longer true in the heterogeneous region of, say, two modifications. In this case it is sufficient to specify one of the intensities T or p, since the other is determined from the equation $p = p(T)$ of the coexistence region. However, for the molar quantities we obtain two sets of solutions: x_i' and x_i'' instead of one. Accordingly the extensive quantities range over

$$X_i = N[x_i' + (1 - \alpha)x_i''], \tag{9.18}$$

where α varies from 0 to 1.

However, this indeterminacy in the description of the system can be removed. According to Section VI, a complete description of the system is provided in terms of a single intensity, say T, and *two* scale factors, say N' and N''. In other words, given T, N', and N'', the variables p, V', V'', S', S'' are uniquely determined.

An appropriate fundamental equation in terms of the Helmholtz function $A = U[T]$ is

$$A(T, V, N) = N'a'(T, v') + N''a''(T, v''). \tag{9.19}$$

The generalization to several components and to more than two coexisting modifications is straightforward.

Case (ii), that of critical equilibrium, is essentially different from the one just considered. The uncertainty of the inference leading from the intensities P_i' to the densities x_i' is not just a shortcoming of the formalism, but has a basis in experiment. The molar quantities or densities x_i' exhibit anomalous large fluctuations in critical equilibrium. This fact is not accounted for in our postulational basis, and hence the latter needs revision.

The qualitative aspects of this revision can be easily explained. In *P b1* and *P d2* it was assumed that the virtual states of the system have no reality and serve only as comparison states for the entropy maximum principle, that assigns a single set of values of the extensive variables to the equilibrium state.[31] Fluctuation phenomena and, in particular, critical fluctuations, force us to assign physical reality to the virtual states. The state of equilibrium of thermostatics is actually not a well-defined state, but rather a statistical distribution over the virtual states. In elliptic points this distribution is sharply peaked and thermostatics, operating with the most probable values of these distributions, leads to quantitatively correct answers. Nevertheless the elimination of the statistical elements renders the theory from the qualitative, conceptual point of view inadequate. As we shall briefly outline in the final discussion, this difficulty is the point of departure for the development of statistical thermodynamics.

It is interesting to note that statistical mechanics predicts infinite fluctuations whenever (17) holds, that is in critical *and* in heterogeneous equilibrium. This property is sometimes used to provide a criterion for the heterogeneous state. While critical fluctuations in finite systems are finite (see e.g. Ref. *12*), they do assume anomalously large values. However, this is not the case in heterogeneous systems where the fluctuations are normal. The theoretical fluctuation is cal-

---

[31] Some qualifications of this uniqueness, irrelevant for the present purposes, will be given below.

culated from an ensemble that contains systems with arbitrary amounts of the two modifications. This corresponds to the incomplete description of the system in terms of the Gibbs function. Thus the "fluctuation" in question has a subjective character and is "transformed away" on transition to the Helmholtz function as explained above. Such an elimination of the objective critical fluctuations is of course impossible.

In the language of the statistical theory we may say that the ensemble of virtual states in homogeneous systems has the ergodic property, that is, the time average over the states of an individually given system is equal to the average over all virtual states; the ensemble corresponding to heterogeneous states is not ergodic and the results of statistical calculations are to be interpreted with care.

The distinction between the two situations is very clear in terms of the derived and primitive fundamental equations. At the derived surface

$$\mathrm{Det}\ | U_{ik} | = 0 \tag{9.20}$$

in the heterogeneous *and* in the critical equilibrium. In contrast, at the primitive surface

$$\mathrm{Det}\ | \mathfrak{U}_{ik} | = 0 \tag{9.21}$$

for the critical point and

$$\mathrm{Det}\ | \mathfrak{U}_{ik} | \neq 0 \tag{9.22}$$

for each of the coexisting modifications. The situation is obscured in the statistical mechanical procedure which deals only with absolutely stable states and fails to distinguish the primitive and the derived fundamental equations.

We turn now to the discussion of the case that the inverse Legendre transformation leading from the P_i to the x_i becomes singular because of $D_r^{-1} = 0$. In analogy with (17) we have now

$$(\partial x_r/\partial P_r)_{x_1, x_2, \cdots, x_{r+1}} = 0. \tag{9.23}$$

If x_r is interpreted as the specific entropy, this relation means that the specific heat vanishes. It follows indeed from *P c2* that this requirement is asymptotically satisfied near absolute zero. More precisely, we can associate with various substances characteristic temperatures θ in such a way that the specific heat tends to zero at least linearly in T/θ. As T falls significantly below θ, the inference of the temperature of the surroundings, obtained from the density of the system, becomes more and more uncertain. This is a well-known difficulty of low-temperature thermometry.

In conclusion, we indicate briefly that the principle of thermostatic determinism can be adapted to other situations. Thus in *P b1* and *P d2* we have tentatively stated that the maximization of the entropy function leads to an *almost unique* determination of the free variables of the problem. We are now in a position to make these statements more definite. However, we shall not fill in all the details of the simple proof.

It is easy to show that the heterogeneous phase structure of a simple system is uniquely determined, except for the unavoidable ambiguity connected with symmetry. Thus the relative amount of each modification doublet (or multiplet) remains undetermined.

Moreover, a homogeneous modification is distributed in a unique fashion over a composite system, the partitions of which are fixed, that is, if the volume of each subsystem is a fixed variable. On the other hand, ambiguities of the solution arise in systems partitioned by adiabatic movable pistons. If such a piston is released from an arbitrary initial position, generally a damped oscillation will result. The conservation laws are not sufficient to determine a unique asymptotic equilibrium position. The latter will depend on the mechanical details of the process that determine how the dissipated macroscopic kinetic energy is shared by the subsystems separated by the piston. Also, the heterogeneous equilibrium in a composite system exhibits considerable ambiguity, which, however, is of a rather trivial nature.

X. LAMBDA LINES AS LOCI OF CRITICAL EQUILIBRIUM

The replacement of the classical version of the second phase rule (9.2) has significant experimental implications. The simplest types of critical points predicted by the theory are summarized in Table I. The column with $\sigma = 0$ corresponds to the classical phase rule, the one with $\sigma = 1$ contains the categories predicted by the present theory. We shall survey some representative nonclassical cases and compare the theoretical predictions with experimental evidence.

TABLE I

VALUES OF δ_{II} WITH THE DESIGNATION OF THE CORRESPONDING TYPES OF PARABOLIC POINTS

σ r	0	1
1	0 none	0 critical point in Ising model
2	0 critical point in fluids	1 λ-line
3	1 critical line of mixing in two component fluids	2 λ-surface in two component systems

From the practical point of view by far the most important class is the one with $r = 2$, $\delta_{II} = 1$, $\sigma = 1$; more explicitly, the theory predicts the possibility of lines of parabolic points in the p - T diagram of one-component systems. We propose to consider this theoretical category as the interpretation of the well known λ-phenomenon[32] observed in many crystals (*31*, *32*) and in liquid helium (*24*, *25*). Plotting the compliance coefficients $\psi_{ik}(T)$ of certain one-component systems as functions of the temperature at constant pressure, it is found that these functions exhibit singularities of a shape reminiscent of the letter λ. The sharp maximum of this curve is called the λ-point: $T_\lambda(p)$. As the pressure varies, the λ-point traces in the $p - T$ plane the λ-line. We recall that ψ_{ik} stands for c_p, α, and κ_T.

In evaluating the merits of this interpretation we have to examine first, whether there is an empirical justification for considering λ-points as parabolic singularities, and second, whether the role assigned to symmetry is borne out by experiment.

Considering at first the matter of the singularity, we have to remember that the establishment of a mathematical singularity from experimental curves is always difficult. We cannot expect to observe actual infinities of the $\psi_{ik}(T)$ functions since the effects that limit the perfect homogeneity of a phase (see Section III) will tend to round off the ideally sharp singularity. Moreover, a smoothing is brought about by the experiment itself, which proceeds by discontinuous steps: a finite DU is measured against a finite DT.

In spite of these reservations, the appearance of most λ-points is sufficiently sharp to warrant the interpretation as a parabolic singularity. This conclusion has been confirmed most convincingly for helium. According to the precision measurements of Fairbank *et al.* (*33*) the specific heat of this substance has a logarithmic singularity at the λ-point.

Notwithstanding the fact that the correspondence between theory and experiment is most satisfactory, we have to dwell somewhat more on this question, because a point of view, that is substantially different from the one here advanced, has been widely accepted in the literature.

It is evident from the foregoing discussion that the classical phase rule (9.2) does not provide the possibility of critical lines in single-component systems.

The need for extending the classical phase theory was recognized by Keesom (*8*) and Ehrenfest (*9*). However, since the comparison of λ-*lines* with critical *points* must have appeared incongruous, λ-*lines* were compared with *lines* of ordinary heterogeneous equilibrium. These are characterized by the differences of the densities

---

[32] These are commonly called "second-order transitions." This terminology is based on an inadequate theoretical conception. At any rate, we prefer to designate the experimental findings by a theoretically noncommittal descriptive term.

$$\Delta x_i = x_i'' - x_i' = \Delta \frac{\partial \psi}{\partial P_i} = -\Delta \frac{\partial g}{\partial P_i}. \tag{10.1}$$

The analogy was conceived by postulating along the λ-line second order discontinuities:

$$\Delta \psi_{ik} = \lim_{t \to 0} [\psi_{ik}(T_0 + t) - \psi_{ik}(T_0 - t)] = \Delta \frac{\partial^2 \psi}{\partial P_i \, \partial P_k}. \tag{10.2}$$

Generally speaking, a transition of the *n*th order was supposed to have discontinuities in the *n*th derivative of the Gibbs function.

Ehrenfest also derived his well known analog to the Clausius–Clapeyron relations relating the slope of the λ-line to the $\Delta\psi_{ik}$.

Keesom and Ehrenfest advanced the concept of second-order transitions in an *ad hoc* manner to describe experimental facts. Subsequently this concept received theoretical support. The λ-phenomenon was given a microscopic interpretation in terms of order-disorder transitions and the approximate statistical calculations invariably led to discontinuities of the specific heat (*30*). Moreover Landau derived the Ehrenfest scheme from thermodynamic stability considerations (*34*). This theory seemed particularly significant since it did not appear to involve any specialized assumptions.

A radically new element entered the picture with Onsager's rigorous solutions of the two-dimensional Ising model (*35*), resulting in a logarithmic singularity of the specific heat. Although it was not immediately obvious to what an extent this result was applicable to actual systems, it was hard to avoid the inference that Landau's theory contained implicit assumptions that impaired its apparent generality.[33] It became necessary to reconsider the fundamental assumptions of the theory,[34] and also the method of analysis of the experimental results.

[33] The present theory agrees in some respects with that of Landau. Both theories derive the theory of phase equilibrium from principles of thermodynamic stability. Moreover, Landau was the first to recognize the importance of symmetry for the λ-phenomenon in crystals, a point to which we return shortly. In other respects there are deviations. In particular Landau's conclusions concerning the nature of the λ-singularity differ from those reached in the present paper. The origin of this discrepancy is that Landau expresses the fundamental equation of his system in terms of the Gibbs function. This presentation is equivalent with that in the energy scheme if (5.14) and (5.17) are satisfied. However, at critical points the Jacobian in (5.14) vanishes, and the compliance coefficients, that is, the second derivatives of the Gibbs function become infinite. Hence, the power series expansion of the Gibbs function does not have the regularity properties assumed by Landau (p. 434 of Ref. *20*).

[34] An attempt of this sort was advanced a few years ago by the author (Ref. *11*). Some of the ideas of the present paper can be found already in this earlier publication, in particular the identification of the λ-lines as singular, critical states. However, in that paper, the logical structure of the theory was left in a rudimentary state. Thus no distinction was made between the thermodynamic extensive variables listed in *P a3*, and the pseudo-

The numerous attempts to test the Ehrenfest relations were always inconclusive since it proved to be difficult, if not impossible to deduce definite $\Delta\psi_{ik}$ from the experimental curves. In particular, the $\Delta\psi_{ik}$ tend to increase as dT, the variation of the temperature between two consecutive measurements is decreased. Since the ideal thermostatic results can be expected to obtain only in the limit $dT \rightarrow 0$, the sharpening of the λ-anomaly in the course of this limiting procedure can be taken as a strong argument in favor of a parabolic singularity.[35] In view of this situation it is common usage at present to classify phase transitions in terms of categories that are more inclusive than those of Ehrenfest. Such categories have been advanced in particular by Pippard who developed also the analogs of Ehrenfest's relations adapted to λ-singularities rather than to second order discontinuities (*36*). In Appendix C relations practically identical to those of Pippard are derived within the present theory.

In spite of the usefulness of such *ad hoc* taxonomical considerations, we wish to emphasize that the most important results of thermodynamics are those in which taxonomical categories are derived from thermodynamic principles. The analysis of experimental results in such terms allows one to extract significant structural information concerning the systems involved. A case in point is the discussion of the role of symmetry in one component systems exhibiting singular λ-lines in the p, T plane.

The discussion of the symmetry of specific thermodynamic systems can be given only in a microscopic, structural sense. Accordingly the rest of this section as well as Section XI go beyond the limitations we have observed thus far. However, the discussion will be qualitative and conceptual throughout.

The connection between the macroscopic and the microscopic points of view is provided by the ensemble concept. In particular, an isolated thermodynamic system in equilibrium is described by a microcanonical ensemble Γ that has the ergodic property. By this we mean that there are incessant transitions among the microstates of constant energy that are associated with Γ, and that the

---

thermodynamic variables, such as the electric and magnetic moment and the elastic strain tensor. Accordingly, the number r defined in (2.7) differs from the definition given earlier. The present alternative is essential for the precise statement of the phase rules.

[35] The differences $\Delta\psi_{ik}$ are obviously not invariant with respect to the variation of the experimental conditions. In earlier attempts of testing the Ehrenfest relations, one often proceeded by prescribing special experimental conditions in order to arrive at definite $\Delta\psi_{ik}$. However, this procedure misses the point: the issue to be decided is not the validity of the Ehrenfest relations, since these are exact thermodynamic relations whenever the $\Delta\psi_{ik}$ are meaningful. The question is rather, whether or not these quantities are physically meaningful and whether the "order" of transitions is a relevant concept for their classification. The above mentioned difficulties indicate that, generally speaking, the answer to this question is in the negative.

average properties of the latter provide the macroscopic properties of the system.[36]

Let us consider now a symmetry operation R, such as the translation, rotation or inversion of the spatial coordinates, or the inversion of time. The ensemble Γ is said to be invariant under R, that is $R\Gamma = \Gamma$, if each microstate of the ensemble is transformed into a microstate that belongs to the same ensemble. In particular, some microstates may be themselves invariant under R.

The present theory predicts that the λ-lines of one-component systems are associated with at least one symmetry operation R in such a way that on one side of the λ-line in the p, T plane the ensemble Γ of the system is invariant under R, whereas on the other side we have modification doublets of the same energy, the ensembles of which satisfy the relations.[37]

$$R\Gamma^{+} = \Gamma^{-}, \qquad R\Gamma^{-} = \Gamma^{+}. \tag{10.3}$$

The ensembles Γ^{+} and Γ^{-} constitute a decomposition of the energy surface into two subsets, each of which has the ergodic property, while the entire surface is not ergodic. The sign of the quasi-thermodynamic parameter η serves to identify the subset relevant in an actual situation. The need for using such nonclassical parameters is connected with the limitation of validity of the classical ergodic hypothesis.

Assuming that the surrounding favors neither of the doublets over the other, from the intrinsic point of view the chances for their formation are equal, and in strict equilibrium one is realized with the exclusion of the other. In practice, of course, the expenditure of a small amount of surface energy produces a course grained mixture of both doublet forms. In crystallography this common phenomenon is called twinning. Assuming that the operation R is the inversion, the two forms Γ^{+} and Γ^{-} are called *enantiomorphic.*

In ferromagnetic and ferroelectric crystals the long range depolarizing forces

---

[36] We note that our postulation of the ergodic property serves merely to define the ensemble Γ: we confine ourselves to such subsets of the energy surface in phase space for which the ergodic property is valid. The mathematical ergodic theorem asserts the ergodic property for metrically transitive (indecomposable) parts of the energy surface. (See Chapter III of Ref. *37*.) The ergodic hypothesis of classical statistical mechanics goes beyond this statement and claims in effect, that the entire energy surface is metrically transitive. This conjecture may be plausible for chemically inert gases. However the occurrence of doublet modifications represents a physically significant counter example against its general validity.

[37] Strictly speaking, the thermodynamic argument requires only that there should be modification multiplets. Thus we may have a polar group pointing in any of the six cubic directions (1, 0, 0). This case could be considered as an accidental degeneracy of three doublet modifications. Empirically the modification always appears in pairs, but we have not been able to prove this result in sufficient generality.

compensate for the surface energy and one usually has a so-called domain structure of modification doublets even in strict equilibrium (*19*).

We shall occasionally refer to the modification Γ as the *high form* and to Γ^+, Γ^- or to their twinned mixture as the *low form*. These terms have a convenient double meaning. Strictly speaking they refer to forms of higher or lower symmetry, but, apart from a single known exception, they can be interpreted also as high and low temperature forms.[38] In order to understand the scope and limitation of this connection between symmetry and temperature, we have to discuss somewhat more specifically the structural properties of systems undergoing the high-low symmetry change.

We note that in case of crystals the specification of Γ ensembles that is adequate for the present purpose consists of three steps.

(i) The first step is the purely geometrical characterization of an ideal crystal in terms of a Bravais lattice and of atomic parameters determining the equilibrium positions of atoms in the primitive unit cell, the basis.

(ii) The sites of the ideal lattice are not actually occupied by atoms. Hence, as a second step one has to specify for the ensemble Γ the probabilities of finding an atom at a definite location with respect to the ideal lattice.

There are several mechanisms responsible for the deviation of the actual from the ideal configuration, such as small vibrations and crystal defects. The mechanism of particular interest in the present context is the ordering-disordering process arising in crystals in which the number of sites for atoms of a given kind is in excess over the number of atoms.

(iii) Certain crystals are not adequately characterized in terms of their configuration, but require also the specification of the distribution of momentum. This term is used here and in the following discussion in a generic sense to denote the linear momentum, or the orbital and spin angular momentum of the elec-

---

[38] The role of symmetry in the λ-phenomenon of crystals was first recognized by Landau (*34*), who showed that the continuous variation of the densities x of a system on crossing the λ-line is consistent with the discontinuous variation of symmetry if the symmetry of the low form is a subgroup of that of the high form. Although the abrupt change of the symmetry group defines a sharp λ-temperature $T_\lambda(p)$ at a constant pressure for any experimental run, the value of this temperature is subject to thermal hysteresis and, in contrast to temperatures of ordinary heterogeneous equilibrium, $T_\lambda(p)$ cannot be chosen as a reliable temperature fixed point. The reason for this situation was pointed out by Justi and Laue (*38*). A homogeneous phase can be supercooled below T_λ, just as it can be supercooled below the temperature of heterogeneous equilibrium. In the latter case the appearance of the low temperature phase releases enough latent heat to bring the system back to the equilibrium temperature, that is maintained as long as there are two phases present in equilibrium with each other. In the case of λ-points there is no latent heat, nor are two phases coexisting with each other in equilibrium. Hence there is no mechanism for assuring that the $T_\lambda(p)$ should be identical in different experimental runs, particularly for heating and cooling experiments.

trons, or the angular momentum arising from the degeneracy of the wave functions of the molecules or ionic groups. To describe such states we have to use quantum mechanical wave functions, that is probability amplitudes of states, rather than probabilities of geometrical configurations. The ensembles Γ are represented now in terms of density matrices.

Corresponding to the three levels in the description of the crystal we distinguish three classes of λ-transition.

(i) *Displacive transitions.* The change of the crystal symmetry is adequately described in terms of a slight distortion of the ideal lattice.

(ii) *Transitions involving configurational ordering.* The change of symmetry occurs because of the variation of the probability of occupancy of sites that are in excess over the number of atoms that occupy those sites. The lattice may become distorted in the course of the ordering process, that is in certain cases the ordering process is superposed over a displacive transition.

(iii) *Transitions involving momentum ordering.* The transition is brought about by a change in the degree of order of the distribution of momentum, understood in the generic sense explained above. This ordering of momentum may be in addition to a certain configurational ordering.

The majority of λ-lines studied so far belong into class ii. These cases of configurational ordering are also the best known and we confine ourselves to a sketchy discussion aiming mainly to establish the contrast and analogy with classes i and iii.

The prototype of a system with configurational ordering is the Ising model. We assume that each unit cell is capable of two geometrical configurations symbolically denoted by "+" and "−". For the purposes of statistical calculations it is usually assumed that the energy of the system depends only on the relative configuration of nearest neighbors and favors like neighbors. For the present qualitative discussion the law of interaction may be more general, so long as the interaction energy falls off with distance strongly enough to be consistent with the additivity of the macroscopic energy.

In each microcanonical ensemble there are definite probabilities p^+ and $p^- = 1 - p^+$ for finding the two cell configurations, respectively. Thus we can define the well-known long range order parameter

$$\eta = p^+ - p^- = 1 - 2p^-. \tag{10.4}$$

In the high form of the system $\eta = 0$, whereas in the low form $0 < |\eta| \leqq 1$ and the two signs of η correspond to the two doublet modifications.

We consider the operation R that interchanges the two configurations in each unit cell. Ensembles with $\eta = 0$ are invariant under R, while (3) holds for ensembles with $\eta \neq 0$. Thus the connection with the foregoing general discussion is established.

The Ising model admits a variety of concrete interpretations. We survey some of the most important ones in order to indicate the extent of experimental corroboration of the general theory.

First, we have *substitutional transitions* occurring in certain 50 atomic percent binary alloys, *AB*. The lattice is subdivided into two sublattices and the two configurations are obtained as an *A* atom occupies a site of the first or second lattice, respectively.

We denote another important subclass as *orientational transitions*. These arise in crystals containing molecules or ionic groups that admit two equilibrium orientations in each unit cell. Examples are the hydrogen halides and a number of ammonium salts (*32*).

An alternate theory of the transitions just mentioned was put forward by Pauling (*39*) and Fowler (*40*). This theory associates the transition temperature with a change from rotational oscillation of the molecules or ionic groups below the λ-point to a modification in which most of the molecules are freely rotating. This conception of "rotational" rather than "orientational" transitions found wide acceptance. Thus the review papers of Ref. *32* analyze the experiments from this point of view.

While the idea of orientational transitions is consistent with the present theory, that of rotational transitions is not, unless it is supplemented by additional assumptions guaranteeing the occurrence of modification doublets.

From the experimental point of view a decision between the two theories was until recently very difficult. The particles partaking in the reorientation or rotation are usually protons or deuterons, which are notoriously hard to locate experimentally.

This gap was filled to a large extent by the powerful method of nuclear magnetic resonances (*41*), that provides information on the dynamic behavior of protons in solids and liquids. The application of this method to the hydrogen halides, to NH_4Cl and to a number of other crystals (*42*) decided the question in favor of the orientational theory. It was found in particular that the λ-points are not associated with a marked change in the dynamic properties of the protons, and, conversely, whenever there is a change from a more to a less hindered rotation in a relatively narrow interval of temperature, this dynamic change does not manifest itself in caloric effects and certainly gives rise to no λ-transition.

While not every system can be analyzed in such an unambiguous way, there is no instance of a verified Pauling type transition.

Among the crystals investigated by the nuclear magnetic resonance techniques the case of methane requires special attention. The experimental results (*43*) were interpreted by Nagamiya (*44*) and Tomita (*45*) as follows. There is a marked transition from a state of coupled rotations to a relatively free rotation at about 65°K. There is no anomaly of the specific heat in this region.

The λ-transition at 20°K is associated with a change in ordering, but this is not of the purely orientational type, since the molecules are in a state of coupled rotation or resonance among equivalent states even at the lowest temperatures. These states are constructed by taking into account the degeneracy connected with the proton spin.

The author believes that methane is an example for a class iii λ-point involving momentum ordering, a case that will be discussed below. It is very likely that the Nagamiya model is consistent with our requirement that the low form of CH_4 should appear in doublet modifications. However, the problem needs more experimental and theoretical study.

We turn now to the class i λ-points denoted above as displacive transitions. These are conceptually simpler than order-disorder transitions, but they are less common and from the experimental point of view more elusive. The transitions can be described in geometrical terms. Thus in barium titanate the cubic crystal suddenly becomes tetragonal. The deviation from cubic symmetry is measured by the parameter

$$\eta = \frac{c}{a} - 1, \tag{10.4}$$

where a, c are the tetragonal lattice constants. The tetragonal axis is polar and we have $R\Gamma^+ = \Gamma^-$ for the low form and $R\Gamma = \Gamma$ in the high form. Thus the symmetry properties of the transition are the same as those of order-disorder transitions. In particular, η decreases continuously to zero as the λ-line is approached from below.

The displacive transitions known at present are all ferroelectric: Rochelle salt, barium titanate, and the ferroelectrics discovered by Mathias and his collaborators (*46*). The dielectric constant exhibits the same λ-singularity as, say, KH_2PO_3 in which the spontaneous polarization is due to the ordering of hydrogen bonds (*47*). The transitions are recognized to be displacive by the absence of λ-singularity in the specific heat. This is due to the fact that, in contrast to class ii transitions, the high and low forms do not differ substantially in their entropy.

A striking manifestation of this circumstance is the lower critical point T_l of Rochelle salt. Below T_l the crystal reverts to the nonpolarized high symmetry form that is stable above the upper Curie point T_u. Although there is no satisfactory theory for this lower critical point, the geometrical character of the symmetry change at least represents no conflict with thermodynamics. This is in contrast with order-disorder transitions, for which the high symmetry form is more disordered, has the higher entropy and always appears at the high-temperature side of the λ-point.

It is likely that there exist also nonferroelectric displacive transitions. However, because of the absence of anomalous physical properties these might easily escape detection. The change in symmetry should be observable by crystallographic means.

We conclude the discussion of displacive transitions by noting that this terminology was advanced originally by Buerger (*48*) to designate the high-low transition of quartz (*49*).

In its low form quartz is trigonal (D_3) and transforms at $t = 574°C$ into the hexagonal high form (D_6). The trigonal form exists in the equivalent "obverse" and "reverse" forms which are the modification doublets in the present terminology. These modifications are transformed into each other by atomic displacements that are small compared to the lattice constant. This is in contrast to the orientational and substitutional transitions in which the distance between equivalent positions of atoms partaking in the ordering process is comparable to the lattice constant. This is of course the basis for designating this transition as displacive. At the same time, however, the transition has a pronounced λ-singularity of the specific heat. In fact, this is presumably the first instance of a λ-point for which an infinite specific heat was reported on purely empirical grounds (*50*). Thus we have a clear indication of an order-disorder process among the slightly displaced equivalent positions. We find it advantageous to reserve the descriptive term "displacive" to the above mentioned class of purely geometrical transitions. The modification doublets of low quartz often occur simultaneously in the same crystal, they are called electrical, or Dauphiné twins. The twinning disappears at the λ-point.

Another property of quartz is that both the D_3 and the D_6 forms are optically active, and appear in right and left forms because of a helical arrangement of the constituent atoms. The optical twinning persists over the entire range of stability of the crystal without ending in a λ-point. In our terminology, the situation is to be described as a *phase doublet*.

The most interesting, but also the most problematic λ-transitions are those of class iii involving momentum ordering. Ferromagnetic and antiferromagnetic substances exhibiting definite patterns in the distribution of electron spin (*51*) belong in this class. The symmetry requirements of the present theory are evidently satisfied. In fact, we may interpret the symmetry operation R as the operation of time reversal $t \rightarrow -t$, that transforms every spin distribution into an energetically equivalent, yet distinct distribution.

The interpretation of R as time inversion may not be the only possibility. The same result could in some cases be achieved in terms of spatial operations. However, since time inversion is independent of the special geometrical properties of particular systems, it is presumably the appropriate operation for all class iii transitions.

The importance of including time inversion in the discussion of the symmetry of thermodynamic systems has been pointed out before (*52*). Also "magnetic space groups" and "magnetic symmetry classes" have been derived by joining the time inversion to the spatial rotation and inversion operations (p. 428 of Ref. *20*).

We have pointed out above that the λ-point of solid methane is likely to represent an instance of momentum ordering. The occurrence of this phenomenon is certainly not common, but depends on a delicate balance of the interaction forces tending to bring about configurational ordering with the effect of quantum mechanical zero point energy opposing such a configurational order. Thus the replacement of one fourth of the protons with deuterons (CH_3D), or the application of hydrostatic pressure is sufficient to upset this balance and produce another transition (see Clusius et al. *15*; *53*), below which the system is presumably in a state of configurational order (*54*).

We wind up the discussion of crystalline λ-lines with the conclusion that in all cases where the structural properties of the system are sufficiently understood, the connection between λ-singularity and symmetry is borne out by experiment. This conclusion forms the basis for the more significant but at the same time more problematic, inferences concerning liquid helium.

XI. SUPERFLUIDITY

The p, T diagram of helium exhibits a λ-line $T_\lambda(p)$ marked by an anomaly in the specific heat $C_p(T)$ and in the other compliance coefficients. This anomaly is similar to the one found at the critical gas-liquid temperature T_c, provided the heating is performed at the exact critical pressure p_c. Moreover, the entire λ-line is thermodynamically identical to the λ-lines that are otherwise found only in solid and liquid crystals.[39] In fact, it was the anomaly of the specific heat of helium that was first designated as a λ-point (*55*) and was used as the experimental basis for the concept of second order transitions (*8*, *9*). Finally, it is for helium, that the singularity of the specific heat was experimentally established in a most convincing fashion by the precision measurements of Fairbank *et al.* (*33*).

---

[39] The discussion of this section is based on the assumption that the λ singularities in helium form a one-dimensional manifold. Actually, the singularity of the function $C_p(p, T)$ has been experimentally established only at the saturation vapor pressure. In the absence of direct measurements of $C_p(T)$ at higher pressures, we can support the above assumptions by plausibility arguments. First, we conjecture that the saturation vapor pressure has no special significance for the intrinsic properties of the liquid, and these are unlikely to undergo a qualitative change upon application of a slight excess pressure. Second, the Pippard relations (see Appendix C) are derived from the assumption that the λ-singularity is only weakly influenced by pressure. According to personal communication of Chase and Maxwell, the relation (C9) is satisfied if their experimental values of the coefficient of expansion $\alpha(T)$ are plotted against the $C_p(T)$ of Ref. *33*.

Under such conditions, it is plausible to assume that the λ-line of helium belongs thermodynamically to the same category as the λ-lines of crystals and that the second phase rule of the present theory applies to it. This assumption leads to the conclusion that the λ-line consists of critical points of the second kind, the symmetry number is $\sigma = 1$, and the superfluid state exists in at least two energetically equivalent modifications that are related to each other by symmetry. Since this situation prevails even at absolute zero, we have to conclude that the ground state wave function of the superfluid state is at least doubly degenerate.

This conclusion is remarkable for more than one reason. We state in essence that the existence of a singular λ-line cannot be inferred from the density of low-lying excitations, but is closely connected with the structure and degeneracy of the ground-state wave function of the many body system. The prediction of a degenerate ground state for helium is at variance with the existing microscopic theories of this system and conflicts also with the widely accepted statement of Nernst's law, according to which the ground state wave function of a quantum mechanical system is nondegenerate. (See p. 66 of Ref. *20.*) However, the argument raised against the present ideas from the point of view of Nernst's law is easily dismissed, since it stems from an inadequate formulation of this law.

The degree of degeneracy of a quantum state with the angular momentum quantum number J is $2J + 1$. Since J cannot be zero for systems containing an odd number of electrons, we have obvious counter examples to the statement that the ground-state wave function is nondegenerate.

In order to arrive at a proper formulation of Nernst's law we consider for the sake of definiteness a crystal of N unit cells with a free spin per cell. The degree of degeneracy g_0 of the lowest ferromagnetic state is between the limits:

$$2 \leqq g_0 \leqq N + 1, \tag{11.1}$$

where the upper limit refers to the embedding of the angular momentum $N\hbar/2$ into an isotropic medium. The double degeneracy of the lower limit is a consequence of time reversal and persists in any anisotropic medium. The third law requires that the entropy per unit cell should tend to zero in the limit of an infinite crystal. This is indeed the case:

$$s_0/k = \lim_{N\to\infty} (\log g_0)/N = 0. \tag{11.2}$$

This result is to be compared with the entropy per unit cell in the high-temperature paramagnetic limit

$$s/k \approx \lim_{N\to\infty} (N \log 2)/N = \log 2. \tag{11.3}$$

In the high-temperature region the wave functions associated with the cells are joined with random phases, while in the region of validity of Nernst's law there are strict phase relations to be observed, and the degeneracy of microscopic units does not manifest itself in a zero point entropy.

Our thermodynamic conclusion that the ground state wave function of helium is degenerate and presumably a doublet, could be substantiated only by an explicit construction of such a wave function. This task is beyond the scope of the thermostatics developed in this paper. Nevertheless we are in a position to discuss a number of heuristic ideas which shed some light on the intuitive meaning of our general results, and which provide us also with clues as to the direction in which a quantum mechanical theory is to be sought.

Our starting point is the Ising model, the only system for which a parabolic singularity has been rigorously demonstrated (*35*). This model can be adapted to a variety of physical situations. In particular, Lee and Yang (*56*) have shown that the Ising model in a magnetic field is mathematically identical with that of a "lattice gas," in which each lattice point is either vacant or occupied by an atom. To each configuration of the lattice of spins there corresponds a configuration of the lattice gas in which a lattice point is vacant or occupied according to whether the corresponding spin is "+" or "−".

The density of the gas in appropriate units is formally identified with

$$\rho = (1 - \eta)/2, \tag{11.4}$$

where η is the long-range order parameter, or the intensity of magnetization. The pressure of the gas is

$$p = -F - H, \tag{11.5}$$

where F is the free energy per spin and H is the magnetic field.

Lee and Yang have shown that the lattice gas is, on the one hand, formally identical to the Ising model, and has, on the other hand, properties that are remarkably similar to those of a real fluid. In particular, the critical temperature of magnetization (Curie temperature) corresponds to the fluid critical point.

It is apparent that the correspondence established by Lee and Yang is consistent with the second phase rule as summarized in Table I. Both systems have isolated critical points ($\delta_{II} = 0$), but in the Ising model $r = 1$, $\sigma = 1$ while in the real fluid $r = 2$, $\sigma = 0$. The lattice gas provides the link between the systems. The pressure (11.5) is on the one hand analogous to the fluid pressure, but formally it has symmetry properties around the critical pressure that allow one to connect it with the magnetic field. The symmetry of the lattice gas, which is spurious, so far as the real fluid is concerned, manifests itself in the "law of constant diameters" according to which the sum of the conjugate densities of the coexisting gas and liquid is independent of the temperature. This law is a fairly good approximation to the "law of rectilinear diameters" of the real fluid, stating that the sum of the conjugate densities is a slowly varying linear function of the temperature (*57*). The residual discrepancy between the law of constant diameters and the experimental situation is inherent in the model. In contrast to the symmetry of opposite spins in the Ising model, the symmetry between particles and holes in the real fluid is only approximate. Therefore, the author cannot accept the contention of Lee and Yang that a real continuum gas can be considered as the limit of a lattice gas as the lattice constant becomes infinitesimally small.

The representation of the properties of real fluids in terms of the lattice gas poses another problem. Strictly speaking a lattice corresponds to a crystal. If the lattice gas is to represent a fluid, the term "lattice" is not to be taken too literally. In ambiguous situations it is useful to qualify that term and to distinguish "metric" and "topological" lattices. The term "metric lattice" should indicate a strict interpretation and imply the existence of long range correlations of the distances of lattice points which produce sharp x-ray diffraction lines. The topological lattice is merely a graph, in which the number of lines branching out of lattice points represent coordination numbers. The lattice of the Ising model can be given such a topological interpretation and coupling energies assigned to bonds without any reference to the distance of the spins.

The representation of fluids in terms of lattices is understood in the same topological sense. According to (11.4) the density is defined in combinatoric rather than metric terms. In addition, the lattice may have also approximate metric properties ranging over a few lattice constants.

Of course, this handling of the situation by means of a semantic refinement is only the formulation of the problem, rather than its solution. In a satisfactory theory of fluids the topological lattice should be given a rigorous mathematical expression.

Lee and Yang have shown under rather general assumptions concerning the coupling forces, that the lattice gas cannot have more than one critical point. In order to find a model exhibiting a one dimensional continuum of critical points which is to represent the helium λ-line, we have to utilize the Ising model in a different fashion.

We start from an idea suggested by Fröhlich (*58*), whose point of departure was a crystalline model of liquid helium advanced earlier by London. For a critical discussion of this theory we refer to Section B of Ref. *24*. The idea is briefly this: liquid helium is considered a binary alloy consisting of N atoms and N holes constituting congruent lattices. Above the λ-point the atoms and holes would be distributed at random over the 2 N sites, whereas below the λ temperature the atoms would statistically prefer one of the sublattices. We should have a class ii substitutional λ-transition.

In contrast to the isolated critical point T_c of the lattice gas, the Fröhlich model leads to a λ-line $T_\lambda(p)$. The crux of the difference is that: in the lattice gas, the magnetization, a *pseudo-thermodynamic* variable, is replaced by the density, a *thermodynamic* variable in the sense of *P a3*. In the Fröhlich model, the magnetization is turned into a long range order parameter, a *quasi-thermodynamic* variable. The existence of a λ-temperature is assured even at a constant value of the density, considered as an arbitrary parameter of the problem. A variation of the density, or pressure leads to a one dimensional continuum of λ-points $T_\lambda(p)$.

Either one of the two sublattices may be selected for occupancy and thus the model leads to the doublet modifications required by thermostatics. In spite of this success in explaining the λ-line, the Fröhlich model has serious shortcomings and was almost at once abandoned.

In the first place, no evidence of lattice structure was revealed by x-ray studies. Even more important, the energetic conditions in helium do not favor the postulated configurational ordering. If the sites of lattice A are all occupied, those of lattice B are all vacant, the transitions of atoms from lattice A to B lowers the energy of the system. This is, of course, in conflict with the idea of configurational ordering.

In view of this situation London rejected the idea of configurational order and suggested that helium might be the seat of an order in momentum space. Moreover, in rediscovering the condensation phenomenon of the Bose–Einstein gas, he provided a model exhibiting such an order. The theoretical work following up London's initiative is so extensive and so well known that a documentation is neither practical, nor necessary in the present context.

In spite of the success of the boson gas theories, we have to conclude from our thermostatic analysis that the structureless nondegenerate ground state of the ideal Bose–Einstein gas is inconsistent with the experimental λ-singularity. The strength of the thermodynamic argument leads us to formulate the conjecture, that under the coupling conditions prevailing in liquid helium, the correct solution of the many boson problem has a degenerate ground state. This degeneracy might be due to the time inversion symmetry similar to the well known case of the Kramers degeneracy (*59*).

The fact that the theories of the low density imperfect boson gas have not yet confirmed the above conclusions should not be taken as a serious argument against them. We believe that the ideal crystal and the ideal gas theories represent complementary limits of the correct theory. It should be extremely difficult to demonstrate subtle structural properties from power series expansions centering around the ideal gas limit, and it is an attractive possibility to bracket the correct theory from both ends. The exploration of the crystalline limit as a point of departure for improvements is so much the more promising, as the neglect of this approach is in curious contrast with the high degree of sophistication reached by the gas theories.

As an introduction to such a program, we reconsider the above mentioned two arguments levelled against the Fröhlich model. First, we note that the lattice should be interpreted in the weaker, topological sense and thus the conflict between fluidity and lattice structure is neither more, nor less serious than that arising in connection with the lattice gas. Second, we hold that London's argument for replacing configurational order by an order in momentum space is convincing. However, the Bose–Einstein condensation in momentum space provides only an incomplete picture of the class iii λ-transition involving momentun ordering, discussed in the last section.

The relation between momentum ordering and Bose–Einstein condensation is conveniently explained in terms of ferromagnetism. The low energy excitations of a ferromagnet can be represented as superposition of spin waves (*60*). Starting from a crystal exhibiting complete spin alignment, a spin wave corresponds to a single inverted spin traveling by means of resonance across the crystal. These waves have the character of bosons. Their condensation temperature corresponds to the Curie temperature.

The spin waves lack any spatial localization and condense in momentum space. Hence a theory that would conceive of ferromagnetism merely as an instance of Bose–Einstein condensation would not provide any information about the structure, symmetry and degeneracy of the ground state. Yet there is compelling experimental evidence for a spatial pattern of spin distribution both for ferromagnetic and antiferromagnetic substances (*51*).

We conjecture that the situation in liquid helium is analogous, although the structure, symmetry and degeneracy of the ground state are extremely elusive from the experimental point of view. Therefore, at the present juncture, the clarification of these ideas is primarily a task for theory. The central issue is establishing the nature of the localized "momentum" that in helium plays the role of that of the electron spin in ferromagnets. Closely connected with this problem is the nature of the order that is emerging in the λ-transition.

We wind up our discussion by formulating with due reserve an intuitive qualitative picture of liquid helium that is suggested by the foregoing considerations and is consistent with the experimental situation.

According to this tentative picture liquid helium is considered as a topological lattice in which each elementary cell is the seat of a quantized angular momentum. In the lowest state the moments are lined up parallel to each other ensuring that the large zero-point kinetic energy does not contribute to randomness and entropy. Time reversal would transform this state into another of opposite momentum. Thus we arrive at an interpretation of the doublet modifications of thermostatics. In the bulk liquid in rest there are equal amounts of both modifications forming a domain structure. An application of a torque would bring about an unbalance of the domains of opposite angular moments in analogy to the action of a magnetic field on a ferromagnet. The critical velocity (p. 198 of Ref.

25) might represent the transition from an orderly domain structure to a turbulent regime in which the domains are tangled in a complex fashion.

Finally the system will exhibit elementary excitations analogous to spin waves, in which a single cell has an elementary momentum with a sign opposite to that of the domain. Such excitations might be identified with Landau's rotons.

While the above picture is in qualitative agreement with experiment, we were not able so far to deduce quantitative experimental predictions. A satisfactory theory could be developed only in quantum mechanical terms.

Finally we digress to make a short remark on superconductivity. The well-known analogy with superfluidity would suggest the conjecture that here again we deal (a) with a λ-anomaly (b) with degenerate ground-state wave functions transformed into each other by time reversal.

However, this analogy is incomplete and superconductors seem to be the only examples for an Ehrenfest type discontinuity of the specific heat. Under such conditions the second phase rule cannot be invoked either to assert or to deny conclusion (b), although we conjecture that the conclusion is correct.

XII. DISCUSSION

Technically we should at this point evaluate our postulational basis by surveying the success of the theory in describing experimental facts. However, to the extent that the present theory covers traditional ground, we can confine ourselves to a mere listing of the types of predictions of thermostatics that are subject to verification. (i) identities among measurable quantities; (ii) the possibility of establishing an absolute temperature scale that is independent of the properties of any particular substance; and (iii) the behavior of systems in the space of intensities, in particular, under conditions of special interest, such as phase equilibrium, critical points, and the vicinity of absolute zero.

The experimental verification is overwhelming on all three counts and for all substances investigated. In particular, the present theoretical framework is wide enough to describe the so-called λ-phenomenon or second order transition which did not fit into the classical theory.

The discussion of the λ-phenomenon is a convenient point of departure for bringing out the characteristic new aspects of the present theory.

From the point of view of the classical phase theory of Gibbs, λ-points have contradictory properties. On the other hand, they are like critical points, the specific heat c_p and the other compliance coefficients exhibit characteristic singularities. On the other hand, instead of appearing in isolated points in the $p - T$ plane, the λ-points of one-component systems form λ-lines in that plane. In this respect, they are more like lines of heterogeneous equilibrium. Ehrenfest sought to solve this difficulty simply by creating new categories which he called phase transitions of higher order. Quite apart from the fact, whether or not these categories are complete enough, this procedure can be accepted as satisfactory only if one considers thermodynamics merely as a taxonomical discipline in which experimental facts are characterized according to their empirical features.

In the present theory we return to the point of view implicit in Gibbs' work according to which taxonomical conclusions are derived from fundamental principles. The failure of the classical theory to account for λ-lines can be traced to the fact that the specification of homogeneous phases or modifications in terms of extensive variables is incomplete, and has to be supplemented by nonclassical variables that account for their symmetry properties. The incorporation of the symmetry concept into the fundamental principles of thermostatics is among the most characteristic features of our generalization of thermostatics. We now proceed to assess the merits of this innovation.

In the first place, the new definition of phase leads to a generalized version of the phase rules. For one-component systems the theory predicts the possibility of λ-lines in the p, T plane, along which the compliance matrix is singular, provided the following situation prevails: the λ-line separates two regions of the p, T plane that correspond to what we call the high and low forms, respectively. The former has the higher symmetry, and its group contains at least one element R which is no longer in the symmetry group of the low form. Assuming that R^2 is an element of the symmetry group of the low form we conclude that the low form consists of doublet modifications which are transformed into each other by the operation R[37]. In the ideal case the low form appears as one or the other of the two modifications forming the doublet. Practically, however, one usually has twinning and the two modifications appear as different domains of the same sample.

The equivalence under the operation R assures the identity of the modifications from the energetic point of view. At the same time the distinctness of the two numbers of the doublet clearly requires the use of nonclassical variables that are not invariant under R. The value of these variables goes gradually to zero as the λ-line is approached.

This situation is to be compared with the classical gas-liquid critical points which are isolated singular points in the p, T plane and where two modifications become identical that differ from each in classical variables, rather than forming a doublet.

We draw attention to the unusual logic of the argument: we assume that the singularity of the compliance matrix is given from experiment and we infer the existence of modification doublets. In statistical mechanics the usual procedure is the opposite. We start from a hypothetical model; the symmetry properties, such as the occurrence of modification doublets, are known from the outset, and the problem is to establish the thermodynamic functions, in particular the singularity of the compliance matrix. This is only one of many features for which the thermodynamic and the mechanical procedures appear complementary to each other.

The comparison of our theoretical conclusions with experiment raises another methodological problem. Apart from limiting cases falling within the scope of classical crystallography, the symmetry properties of thermodynamic systems cannot be discussed in macroscopic terms. Thus the injection of symmetry into thermostatics calls at the same time for an extension of the theory in the microscopic direction.

In Sections X and XI we have made a start toward a microscopic discussion of symmetry. However, we went only as far as qualitative arguments would allow us to go. In such a way we arrived at a classification of the λ-transitions into

(i) displacive transitions,
(ii) transitions involving configurational ordering,
(iii) transitions involving momentum ordering.

"Momentum" is understood here in the generic sense and includes angular momentum and in particular spin. All of these dynamic quantities change sign on time reversal.

The characteristic symmetry operation R is defined for class i transitions in geometrical terms, for class ii in terms of probabilities of configurations, and for class iii in terms of quantum mechanical probability amplitudes, the operation R can be chosen in this case as the operation of time reversal.

We conclude from the survey of the experimental material that whenever the structural properties of a substance are sufficiently understood, the λ-phenomenon conforms to the conditions set by the generalized phase rule.

The case of liquid helium is in a special category. The theory leads to the inference that the ground state wave function of the superfluid is at least doubly degenerate. The operation R that transforms the two wave functions into each other is presumably the operation of time reversal.

This conclusion is at variance with the existing theories of superfluidity, and does not seem to be amenable to any direct experimental check. Under these conditions we are bound to ask: how reliable are the conclusions reached and what is the outlook for providing a quantum mechanical substantiation for them?

Concerning the first question we point out that the logic of MTE is considerably more rigorous than that of traditional thermostatics. In fact the cumulative effect of many apparently minor changes brings about a logical structure that forms the basis for a reconsideration of the traditional relation of thermodynamics and mechanics, an idea that will be developed elsewhere.

Among other features, we do not formulate the second law in the usual weak form, merely excluding the possibility of processes away from equilibrium. We rather postulate that equilibrium is reached, and the proper entropy is attained. The results are not universally valid for actual systems, but represent the *normal equilibrium behavior*, a practical criterion of equilibrium for actual systems.

Our strong formulation of the entropy maximum postulate yields a theorem which we have called the *principle of thermostatic determinism.* Suppose that an infinite reservoir is in equilibrium with a simple system for which it provides the surroundings. The principle of determinism states that the intensities of the surroundings uniquely determine the densities of the systems, provided condition (5.14) is satisfied. Conversely, the densities uniquely determine the intensities, if condition (5.17) holds. The inferences leading from the properties of the surroundings to that of the system and vice versa, correspond to basic experimental situations. Hence the special designation of the uniqueness theorem seems justified.

The principle of determinism has its limitations. Condition (5.14) breaks down at the critical point, and (5.17) fails near absolute zero. These limitations are real.

In a surroundings having the intensities of the critical point, the densities of the system are indeed subject to large uncertainties which originate in extremely large fluctuations.

Near absolute zero, where the entropy is practically zero, the system becomes insensitive to temperature variations. Physically, this effect comes from long-range space-time correlations in the system, which, within limits, renders it insensitive to outside influences.

The actual occurrence of these extreme situations makes it evident that a correct thermodynamics should include both the elements of randomness and of space-time correlations. The difficulties of developing such a theory are at present very great; however a step by step approximation seems feasible.

First, there is a purely statistical theory in which space-time correlations are assumed to be absent. A statistical theory of this sort, which is applicable to fluctuations around equilibrium, was developed in collaboration with Quay (*14*). This theory, the statistical thermodynamics of equilibrium (STE), will be published as the second paper of this series.

This theory appears as a very natural generalization of MTE. In MTE the virtual states of a composite system have only a formal meaning, they are comparison states for the extremum problem that determines equilibrium. In STE equilibrium is conceived as a statistical distribution over virtual states, and these assume physical reality as fluctuation states. The additive invariants now become additive random variables. Their distributions and other relevant statistical questions can be derived by using methods which had been advanced by Szilard (*61*) and Mandelbrot (*62*). By and large STE turns out to be an axiomalization of Gibbs' theory of ensembles (*63*).

The next step should be a quantum mechanical theory of space-time correlations. A satisfactory solution of this problem seems to be the prerequisite for substantiating our conclusions concerning the nature of the superfluid ground state.

APPENDIX A. QUASI-THERMODYNAMIC VARIABLES

We proceed to discuss a few ideas concerning the definition of quasi-thermodynamic variables.

We divide a thermodynamic system into M^3 volume elements (cells) of equal size and shape, where M is a large number. For crystals the cell may coincide with the unit cell. For fluids it is conveniently a cubic cell that contains a moderately large number of molecules.

The cells are numbered, as usual by vectors **n**, with integral components running from 0 to $M - 1$. Let the system be specified by a set of numbers $z_{\mathbf{n}}$, which denote the value of the quantity z in the cell **n**. The physical nature of the quantity z is left open at first. The fundamental domain of M^3 cells may be extended by means of periodic boundary conditions, and the $z_{\mathbf{n}}$ expressed in terms of normal coordinates.

Assuming, for simplicity a simple cubic lattice, we define

$$\xi_{\mathbf{k}} = \sum_{\mathbf{n}} z_{\mathbf{n}} e^{-2\pi i \mathbf{k}\cdot\mathbf{n}}, \tag{A1}$$

where the components of k are $0, 1/M, \cdots, M - 1/M$. The zero-wave vector normal coordinate

$$\xi_0 = \sum_{\mathbf{n}} z_{\mathbf{n}} \tag{A2}$$

is a translation invariant. Moreover

$$\xi_{1/2,1/2,1/2} = \sum_{\mathbf{n}} z_{\mathbf{n}}(-1)^{n_1+n_2+n_3} \tag{A3}$$

is an invariant with respect to the cubic group with a lattice constant of two units. We define the coordinates (2) and (3) as macrovariables, that are appropriate for the description of course-grained homogeneous phases. These conclusions were reached also by Landau and Lifshitz from more elaborate considerations (Chapter XIV of Ref. *20*).

The coordinates ξ_0 include the additive invariants of *P a3*. The normal coordinates $\xi_{\mathbf{k}}$ for which the components of the wave vector are not all zero or ½, are microvariables. They multiply by a phase factor $e^{2\pi i \mathbf{k}\cdot\Delta\mathbf{n}}$ as the lattice is displaced by $\Delta\mathbf{n}$. These variables are used for the description of instantaneous spatial fluctuations.

In addition to these thermodynamic variables there are translational invariants, and these are the ones that are admissible as quasi-thermodynamic variables η. If z is the coordinate of an atom in the unit cell, ξ_0 represents the ideal lattice of the average positions of the equivalent atoms. The corresponding microvariables are phonons.

If an atom or a group of atoms has two equivalent positions in a unit cell, we may assign the values $z = +1$ and -1 to cells in which one or the other position is actually occupied. In this case $\xi_0 = N^+ - N^-$, where N^+ and N^- are the number of atoms in the two positions, respectively. Since $N^+ + N^- = M^3$, the quantity

$$\frac{\xi_0}{M^3} = \frac{N^+ - N^-}{N^+ + N^-} \tag{A4}$$

is the well-known parameter of long-range order, while the variables with $k \neq 0$ describe the microdistribution or short-range order.

The nature of the λ-phenomenon depends essentially on the coordinate z, or ξ_0. In most cases, this parameter is of type (4). However, in some instances, as in Rochelle salt and barium titanate (*11*, *36*), the parameter η_0 relates to a small distortion of the crystal lattice from a higher to a lower symmetry form (orthorhombic to monoclinic in Rochelle salt). This case is sometimes called a displacive transition.

From the quantum mechanical point of view, in both of the cases mentioned above, ξ_0 corresponds to the absolute square of the wave function. However, it is very likely that in some cases, and particularly in helium, ξ_0 is to be defined in terms of the wave function itself.

It is apparent from the discussion of the text that not all parameters ξ_0 that are admissible on the basis of the foregoing considerations will be actually used as quasi-thermodynamic parameters η. The fundamental equation in, say, the energy scheme can be unconditionally minimized with respect to the $\xi_0 = \eta$. Thus

$$\partial u / \partial \xi_0 = 0. \tag{A5}$$

As a rule, this equation yields a unique solution for ξ_0. Only in those cases in which this solution is not single-valued will the parameter η be significant in accounting for a λ-phenomenon.

APPENDIX B. THE DIAGONALIZATION OF QUADRATIC FORMS

Consider the quadratic form

$$w_r = \tfrac{1}{2} \sum u_{ik} \xi_i \xi_k , \tag{B1}$$

where

$$u_{ik} = u_{ki} . \tag{B2}$$

We assume at first that the form is nonsingular, that is, its determinant

$$D_r = | u_{ik} | \neq 0. \tag{B3}$$

There are infinitely many linear (affine) transformations $\eta = T\xi$ that diagonalize the form (1). In particular we may obtain a diagonal form with the diagonal elements ±1 and 0. The restriction to length-preserving, orthogonal transformations leads to the well known eigenvalues which in this case turn out to be physically meaningless. The physically significant diagonal form is obtained from the requirement that the transformation be unimodular, of determinant unity, or, in geometrical language, volume preserving.

We assume $u_{11} \neq 0$ and proceed by "completing the square":

$$w_r = \frac{1}{2} u_{11} \left(\xi_1 + \frac{1}{u_{11}} \sum_2^r u_{ik} \xi_k \right)^2 - \frac{1}{2u_{11}} \left(\sum_2^r u_{ik} \xi_k \right)^2 + \frac{1}{2} \sum_2^r u_{ik} \xi_i \xi_k . \tag{B4}$$

Substituting

$$\eta_1 = \xi_1 + \frac{1}{u_{11}} \sum_2^r u_{1k} \xi_k , \tag{B5}$$

we obtain

$$w_r(\xi_1 , \xi_2 , \cdots , \xi_r) = \tfrac{1}{2} u_{11} \eta_1^2 + w_{r-1}(\xi_2 , \xi_3 , \cdots , \xi_r), \tag{B6}$$

where the residual form w_{r-1} is of only $r - 1$ variables. This procedure can be applied to w_{r-1} and then continued until the diagonalization is completed, provided that each residual form contains at least one nonvanishing diagonal element.

If all diagonal elements vanish, the matrix of the form has a segment such as

$$\begin{Vmatrix} 0 & u_{jl} \\ u_{lj} & 0 \end{Vmatrix} . \tag{B7}$$

In this case, we substitute

$$\xi_j = \tfrac{1}{2}(\eta_j - \eta_l) \tag{B8}$$

and

$$\xi_l = \tfrac{1}{2}(\eta_j + \eta_l),$$

which yields

$$2u_{jl}\xi_j\xi_l = u_{jl}(\eta_j^2 - \eta_l^2). \tag{B9}$$

The combination of substitutions such as (5) and (8) always allows us to complete the diagonalization:

$$w_r = \tfrac{1}{2}\sum_{1}^{r} \lambda_i \eta_i^2. \qquad \text{(B10)}$$

We note that transformation (8) yields diagonal elements of opposite sign (hyperbolic form). Such situations are of no physical interest in thermostatics, hence we confine ourselves to the case in which the transformation connecting the ξ_i and the η_i is of the "triangular" form:

$$\begin{aligned} \eta_i &= \xi_i + a_{12}\xi_2 + a_{13}\xi_3 + \cdots + a_{1r}\xi_r\,, \\ \eta_2 &= \qquad\quad \xi_2 + a_{23}\xi_3 + \cdots + a_{2r}\xi_r\,, \\ \eta_r &= \qquad\qquad\qquad\qquad\qquad\quad\ \xi_r\,. \end{aligned} \qquad \text{(B11)}$$

If the form (1) is singular and the corresponding matrix is of rank $(r - 1)$, the last diagonal element vanishes, and we have $\lambda_r = 0$. In general, more than one diagonal element may vanish, but, thus far, such cases have not been found of interest in thermostatics.

The coefficients λ_i can be calculated in a straightforward manner by a step-by-step construction of the transformation (11). However, it is easier to proceed as follows:

The transformation (11) is unimodular, hence the discriminant of (1) is invariant.

$$\lambda_1\lambda_2 \cdots \lambda_r = D_r\,. \qquad \text{(B12)}$$

This is seen also from the fact that (11) involves merely the subtraction of the multiple of one row of the determinant (3) from another while similar operations are performed on the columns.

Similar relations hold for the principal minors

$$\lambda_1\lambda_2 \cdots \lambda_k = D_k\,, \qquad \text{(B13)}$$

where

$$D_k = \begin{vmatrix} u_{11}\, u_{12} \cdots u_{1k} \\ u_{21}\, u_{22} \cdots u_{2k} \\ \cdots\cdots\cdots\cdots \\ u_{k1}\, u_{k2} \cdots u_{kk} \end{vmatrix} = \frac{\partial(P_1, P_2, \cdots, P_k)}{\partial(x_1, x_2, \cdots, x_k)}$$

$$= \frac{\partial(P_1, P_2, \cdots, P_k, x_{k+1} \cdots x_r)}{\partial(x_1, x_2, \cdots, x_r)}\,. \qquad \text{(B14)}$$

From (13), we have

$$\lambda_k = D_k/D_{k-1} \qquad \text{(B15)}$$

and from (14),

$$\lambda_k = \frac{\partial(P_1, P_2, \cdots, P_k, x_{k+1} \cdots x_r)}{\partial(P_1, P_2, \cdots, P_{k-1} \cdots x_r)} \tag{B16}$$

$$= (\partial P_k/\partial x_k)_{P_1P_2\cdots P_{k-1}x_{k+1}\cdots x_r} .$$

While different unimodular transformations lead to different sets of λ_i, the product (12) is a unimodular invariant.

APPENDIX C. DERIVATION OF THE PIPPARD RELATIONS FOR THE λ-LINES OF ONE-COMPONENT SYSTEMS.

We consider the implications of the working hypothesis that in the neighborhood of the λ-line the Gibbs function can be approximately expressed as

$$G(p, T) = f(t) + tg(p) + h(p), \tag{C1}$$

where

$$t = T - T_\lambda(p) \tag{C2}$$

and

$$\xi = (dp/dT)_\lambda = (dT_\lambda/dp)^{-1} \tag{C3}$$

is the slope of the λ-line and is assumed to be constant; $f(t)$ is a singular function with the properties:

$$\lim_{t\to 0} f(t) = 0, \qquad \lim_{t\to 0} f'(t) = \text{finite}, \qquad \lim_{t\to 0} f''(t) = -\infty,$$

and $g(p)$ and $h(p)$ are slowly varying functions of the pressure. The primes denote differentiation.

The meaning of this assumption is that the parabolic singularity does not vary appreciably as we move along the λ-line.

We obtain through differentiation

$$-(\partial G/\partial T)_p = S = -f' - g \tag{C4}$$

$$(\partial G/\partial p)_T = V = -f'\xi^{-1} - g\xi^{-1} + tg' + h' \tag{C5}$$

$$-(\partial^2 G/\partial T^2)_p = C_p/T = -f'' \tag{C6}$$

$$\partial^2 G/\partial T\partial p = V\alpha = -f''\xi^{-1} + g' \tag{C7}$$

$$-(\partial^2 G/\partial p^2)_T = V\kappa_T = -f''\xi^{-2} + 2\xi^{-1}g' - tg'' - h''. \tag{C8}$$

From here

$$\alpha = C_p/VT\xi + \alpha_0 , \tag{C9}$$

$$\kappa_T = \alpha\xi^{-1} + \kappa_0 . \tag{C10}$$

The intercepts in these relations are

$$V\alpha_0 = -(\partial S/\partial p)_\lambda = g', \tag{C11}$$

$$V\kappa_0 = -(\partial V/\partial p)_\lambda = \xi^{-1}g' - h'' - tg''. \tag{C12}$$

The determinant of the compliance matrix is

$$\begin{aligned} D^{-1} &= C_pV\kappa_{T/T} - (V\alpha)^2 \\ &= -C_pV\alpha_{0/T\xi} + C_pV\kappa_{0/T} - (V\alpha_0)^2, \end{aligned} \tag{C13}$$

where D is the stiffness matrix.

For the sound velocity μ_s we have

$$\begin{aligned} \mu_s^2 &= -V^2\left(\frac{\partial p}{\partial V}\right)_s = -\frac{V^2C_pD}{T} \\ &= V^2(\xi^{-1}V\alpha_0 + V\kappa_0 - T(V\alpha_0)^2/C_p)^{-1}. \end{aligned} \tag{C14}$$

This expression is finite for $C_p \rightarrow \infty$. The same is true of the specific heat at constant volume $C_V = T(V\kappa_T D)^{-1}$ which tends to a finite maximum at the λ-line:

$$C_{V\lambda} = T\xi V(\xi\kappa_0 - \alpha_0). \tag{C15}$$

Setting $T \approx T_\lambda$ and $g'' \approx 0$, we obtain the relations of Pippard (p. 143 of Ref. *10*) and those of Chase (*36*).

Received: February 26, 1960

REFERENCES

1. L. Onsager, *Phys. Rev.* **37,** 405 (1931); **38,** 2265 (1931).

2. J. W. Gibbs, "Collected Works," Vol. I, "*Thermodynamics.*" Yale Univ. Press, New Haven, 1948.

3. C. Carathéodory, *Math. Ann.* **67,** 355 (1909); *Sitzber. preuss. Akad. Wiss.* p. 39 (1925).

4. J. D. van der Waals, "Lehrbuch der Thermostatik." J. A. Barth, Leipzig, 1927.

5. P. T. Landsberg, *Revs. Modern Phys.* **28,** 363 (1956).

6. P. Ehrenfest. *Z. Phys. Chem.* **77,** 227 (1911).

7. W. Blaschke, "Vorlesungen über Differential Geometrie," Vol. II. "Affine Differential Geometrie," Chapter 4. Springer, Berlin, 1923.

8. W. H. Keesom, *Comm. Kamerlingh Onnes Lab. Leiden* Suppl. 75a and *Proc. Koninkl. Akad. Wetenschap. Amsterdam* **36,** 147 (1933).

9. P. Ehrenfest, *Comm. Kamerlingh Onnes Lab. Leiden* Suppl. 75b and *Proc. Koninkl. Akad. Wetenschap. Amsterdam* **36,** 153 (1933).

10. A. B. Pippard, "Elements of Classical Thermodynamics for Advanced Students of Physics, Cambridge Univ. Press, Cambridge, 1957.

11. L. Tisza, "Phase Transformations in Solids," Smoluchowski, Mayer, and Weyl, eds., Chapter 1. Wiley, New York, 1951.

12. M. J. Klein and L. Tisza, *Phys. Rev.* **76,** 1861 (1949).

13. L. Tisza, *Phys. Rev.* **110,** 587 (1958).
14. P. M. Quay and L. Tisza (to be published).
15. F. Simon, Ergeb. exakt. Naturw. **9,** 222 (1930);
L. Pauling, *J. Am. Chem. Soc.* **57,** 2680 (1935);
K. Clusius, L. Popp, and A. Frank, *Physica* **4,** 1105 (1937);
J. O. Clayton and W. F. Giauque, *J. Am. Chem. Soc.* **54,** 2610 (1932).
16. R. H. Fowler and E. A. Guggenheim, "Statistical Thermodynamics," pp. 535–39. Cambridge Univ. Press, London and New York, 1939.
17. W. F. Giauque and J. W. Stout, *J. Am. Chem. Soc.* **58,** 1144 (1936).
18. See, e.g., M. Planck, *Physica* **2,** 1029 (1935).
19. See C. Kittel, "Introduction to Solid State Physics," 2nd ed., p. 414. Wiley, New York, 1956.
20. L. D. Landau and E. M. Lifshitz, "Statistical Physics," pp. 33, 215. Pergamon Press, London and Paris, 1958.
21. N. F. Ramsey, *Phys. Rev.* **103,** 20 (1956).
22. M. J. Klein, *Phys. Rev.* **104,** 589 (1956).
23. T. Ehrenfest-Afanassjewa, *Z. Physik* **33,** 933, (1925); **34,** 638 (1925); Also "Die Grundlagen der Thermodynamik," p. 4. E. J. Brill, Leiden, 1956.
24. F. London, "Superfluids," Vol. II, p. 72. Wiley, New York, 1954.
25. K. R. Atkins, "*Liquid Helium.*" Cambridge Univ. Press, London and New York, 1959.
26. S. R. deGroot, "Thermodynamics of Irreversible Processes." North Holland Publishing Co., Amsterdam, 1952.
27. See J. E. Mayer and M. G. Mayer, "Statistical Mechanics." Wiley, New York, 1940.
28. J. A. Beattie, lecture notes (unpublished).
29. B. H. Zimm, *J. Chem. Phys.* **19,** 1019 (1951).
30. See G. H. Wannier, *Revs. Modern Phys.* **17,** 50 (1945).
31. F. Seitz, "The Modern Theory of Solids," pp. 37, 57, 511. McGraw-Hill, New York, 1940.
32. A. Eucken, *Z. Elektrochem.* **45,** 126 (1939) (review paper);
C. P. Smyth, *Chem. Rev.* **19,** 329 (1936) (review paper).
33. W. M. Fairbank, M. J. Buckingham, and C. F. Kellers, *in* "Proceedings of the Fifth International Conference on Low Temperature Physics and Chemistry, p. 50. Madison, Wisconsin, 1958.
34. L. D. Landau, *Physik. Z. Sowjetunion* **11,** 26 and 545 (1937) and Chapter XIV of Ref. *20*.
35. L. Onsager, *Phys. Rev.* **65,** 117 (1944);
G. F. Newell and E. W. Montroll, *Revs. Modern Phys.* **25,** 353 (1953).
36. A. B. Pippard, *Phil. Mag.* [8], **8,** 473 (1956) and Chap. 9 of Ref. *10*;
C. E. Chase, *Phys. Rev. Letters* **2,** 197 (1959).
37. A. I. Khinchin, "Mathematical Foundations of Statistical Mechanics," translated by G. Gamow. Dover, New York, 1949.
38. E. Justi and M. V. Laue, *Physik. Z.* **35,** 945 (1934).
39. L. Pauling, *Phys. Rev.* **36,** 430 (1930).
40. R. H. Fowler, *Statistical Mechanics*, 2nd ed. Cambridge Univ. Press, Cambridge, 1936.
41. N. Bloembergen, E. M. Purcell, and R. V. Pound, *Phys. Rev.* **73,** 679 (1948).
42. N. L. Alpert, *Phys. Rev.* **75,** 398 (1949).
43. J. T. Thomas, N. L. Alpert, and H. C. Torrey, *J. Chem. Phys.* **18,** 1511 (1950).
44. T. Nagamiya, *Progr. Theoret. Phys. Japan* **6,** 702 (1951).
45. K. Tomita, *Phys. Rev.* **89,** 429 (1953).
46. B. Matthias, Chapter XII of Ref. *11*.

47. J. C. Slater, *J. Chem. Phys.* **9,** 16 (1941).

48. M. J. Buerger, *J. Chem. Phys.* **15,** 1 (1947).

49. W. H. Bragg and R. E. Gibbs, *Proc. Roy. Soc.* **A109,** 405 (1925);
R. E. Gibbs, *Proc. Roy. Soc.* **A110,** 443 (1926).

50. H. Moser, *Physik. Z.* **37,** 737 (1936).

51. C. G. Shull and J. S. Smart, *Phys. Rev.* **76,** 1256 (1949);
C. G. Shull, W. A. Strauser, and E. O. Wollan, *Phys. Rev.* **83,** 333 (1951).
C. G. Shull, E. O. Wollan, and W. C. Koehler, *Phys. Rev.* **84,** 912 (1951).

52. E. P. Wigner, "Group Theory," Translated by J. J. Griffin, Chapter 26. Academic Press, New York and London, 1959;
H. Zocher and C. Torok, *Proc. Natl. Acad. Sci. U.S.* **39,** 681 (1953).

53. R. Stevenson, *J. Chem. Phys.* **27,** 656 (1957).

54. H. M. James and T. A. Keenan, *J. Chem. Phys.* **31,** 12 (1959).

55. W. H. Keesom and A. P. Keesom, *Proc. Koninkl. Akad. Wetenschap. Amsterdam* **35,** 786 (1932).

56. T. D. Lee and C. N. Yang, *Phys. Rev.* **87,** 910 (1952);
C. N. Yang and T. D. Lee, *Phys. Rev.* **87,** 404 (1952);
F. Cernuschi and H. Eyring, *J. Chem. Phys.* **1,** 574 (1939).

57. See J. O. Hirschfelder, C. F. Curtis, and R. B. Bird, "Molecular Theory of Gases and Liquids," p. 362. Wiley, New York, 1954.

58. H. Fröhlich, *Physica* **4,** 639 (1937).

59. H. A. Kramers, *Koninkl. Ned. Akad. Wetenschap, Proc.* **33,** 959 (1930).

60. A. Sommerfeld and H. Bethe, *Handbuch der Physik,* Vol. 24, Part 2, p. 607. Springer, Berlin, 1933.

61. L. Szilard, *Z. Physik* **32,** 735 (1925).

62. B. Mandelbrot, I.R.E. Transactions on Information Theory, IT-2, No. 3, September, 1956.

63. J. W. Gibbs, "Collected Works," Vol. II, "Elementary Principles of Statistical Mechanics." Yale Univ. Press, New Haven, 1948.