DOCUMENT RESUME

ED 045 349 SE 009 211

AUTHOR	Pearlman, Norman
TITLE	Heat and Motion.
INSTITUTION	Commission on Coll. Physics, College Park, Md.
SPONS AGENCY	National Science Foundation, Washington, D.C.
PUB DATE	66
NOTE	77p.; Monograph written for the Conference on the New Instructional Materials in Physics (University of Washington, Seattle, 1965)
EDRS PRICE	EDRS Price MF-$0.50 HC-$3.95
DESCRIPTORS	*College Science, *Instructional Materials, *Physics, *Quantum Mechanics, *Thermodynamics

ABSTRACT

Unlike many elementary presentations on heat, this monograph is not restricted to explaining thermal behavior in only macroscopic terms, but also developes the relationships between thermal properties and atomic behavior. "It relies at the start on intuition about heat at the macroscopic level. Familiarity with the particle model of mechanics, which is assumed from an earlier course, is used to develop an understanding of temperature and heat in macroscopic terms. It then relates these ideas to behavior of the internal degrees of freedom in macroscopic objects, and shows that these degrees of freedom behave differently from what would be expected if they simply mimicked large-scale motion. In this way, contact is made with discussion of quantum mechanics elsewhere." Topics are: Model-Building in Physics; The Particle Model for Motion; The Direction of the Flow of Time; Temperature and Thermal Equilibrium; Do Thermal Equilibrium States Exist?; and Internal Degrees of Freedom. (Author/PR)

ED0 45349

-009 211

GENERAL PREFACE

This monograph was written for the Conference on the New Instructional Materials in Physics, held at the University of Washington in the summer of 1965. The general purpose of the conference was to create effective ways of presenting physics to college students who are not preparing to become professional physicists. Such an audience might include prospective secondary school physics teachers, prospective practitioners of other sciences, and those who wish to learn physics as one component of a liberal education.

At the Conference some 40 physicists and 12 filmmakers and designers worked for periods ranging from four to nine weeks. The central task, certainly the one in which most physicists participated, was the writing of monographs.

Although there was no consensus on a single approach, many writers felt that their presentations ought to put more than the customary emphasis on physical insight and synthesis. Moreover, the treatment was to be "multi-level" --- that is, each monograph would consist of several sections arranged in increasing order of sophistication. Such papers, it was hoped, could be readily introduced into existing courses or provide the basis for new kinds of courses.

Monographs were written in four content areas: Forces and Fields, Quantum Mechanics, Thermal and Statistical Physics, and the Structure and Properties of Matter. Topic selections and general outlines were only loosely coordinated within each area in order to leave authors free to invent new approaches. In point of fact, however, a number of monographs do relate to others in complementary ways, a result of their authors' close, informal interaction.

Because of stringent time limitations, few of the monographs have been completed, and none has been extensively rewritten. Indeed, most writers feel that they are barely more than clean first drafts. Yet, because of the highly experimental nature of the undertaking, it is essential that these manuscripts be made available for careful review by other physicists and for trial use with students. Much effort, therefore, has gone into publishing them in a readable format intended to facilitate serious consideration.

So many people have contributed to the project that complete acknowledgement is not possible. The National Science Foundation supported the Conference. The staff of the Commission on College Physics, led by E. Leonard Jossem, and that of the University of Washington physics department, led by Ronald Geballe and Ernest M. Henley, carried the heavy burden of organization. Walter C. Michels, Lyman G. Parratt, and George M. Volkoff read and criticized manuscripts at a critical stage in the writing. Judith Bregman, Edward Gerjuoy, Ernest M. Henley, and Lawrence Wilets read manuscripts editorially. Martha Ellis and Margery Lang did the technical editing; Ann Widditsch supervised the initial typing and assembled the final drafts. James Grunbaum designed the format and, assisted in Seattle by Roselyn Pape, directed the art preparation. Richard A. Mould has helped in all phases of readying manuscripts for the printer. Finally, and crucially, Jay F. Wilson, of the D. Van Nostrand Company, served as Managing Editor. For the hard work and steadfast support of all these persons and many others, I am deeply grateful.

Edward D. Lambe
Chairman, Panel on the
New Instructional Materials
Commission on College Physics

HEAT AND MOTION

PREFACE

The title of this monograph, "Heat and Motion", immediately brings to mind the twin concepts: heat, connected with the random motion of particles that constitute matter; temperature, as a measure of the intensity of that motion. These ideas are certainly far from novel, but the approach to them in this monograph is somewhat unconventional, so a prospective reader might find it useful to contrast the usual treatments with that adopted here.

The most natural approach to these ideas might seem to be through atomic theory, but this turns out to be difficult at the elementary level. A cumbersome and sophisticated statistical apparatus is needed before even the simplest results can be proved rigorously. For this reason, and also because it is possible to derive a great many useful results while completely ignoring the atomic constitution of matter, it was the custom for a long time to present a topic called "Heat" at the elementary level. The relation of heat and temperature to molecular motion might be mentioned, but that relation was not at the focus of attention. That position was reserved for careful discussions of temperature scales, calorimetry, and such matters. The "Laws of Thermodynamics" were stated and derived in purely macroscopic terms. Since the atomic constitution of matter was largely ignored, the subject was pretty much divorced from the mainstream of Physics. As the undergraduate curriculum became crowded with other subjects that were regarded as more important, less room was left for what appeared to be peripheral and specialist concentration, and this kind of course became less popular.

Recently, much effort has been directed towards introducting quantum mechanics at an early stage of the college physics curriculum, but successful methods for doing this have so far been rather elusive. These attempts have compounded the problem of discussing thermal effects starting from the molecular level, since they add another hurdle to the statistical one mentioned above. In the absence of rigorous methods suitable for elementary discussions, the connection of heat and temperature with molecular motion has been presented as a sort of intuitive premise. But while intuition may serve as a reliable guide for macroscopic motion (although it may fail here, as anyone knows who has tried to discuss a gyroscope intuitively) it can fail completely when applied at the quantum level. At that level, in fact, it is necessarily misleading, since atoms and molecules are not Lilliputian versions of billiard balls. The source of much of the difficulty of presenting quantum mechanics at an elementary level is precisely the unjustified expectation that an electron, for instance, should behave like a Newtonian particle. And one of the grounds of this expectation, in the mind of a beginning student, is just the uncritical application of classical kinetic ideas to the motion of atomic systems, to which he may have been exposed in an attempt to relate the behavior of such systems to the thermal properties of macroscopic objects.

This monograph adopts an approach which attempts to bypass these difficulties by postponing the connection between atomic behavior and thermal properties, but by no means ignoring it. It relies at the start on intuition about heat at the macroscopic level. Familiarity with the particle model of mechanics, which is assumed from an earlier course, is used to develop an understanding of temperature and heat in macroscopic terms. It then relates these ideas to behavior of the internal degrees of freedom in macroscopic objects, and shows that these degrees of freedom behave differently from what would be expected if they simply mimicked large-scale motion. In this way, contact is made with discussion of quantum mechanics elsewhere.

One of the major threads in the discussion here is the contrast between the behavior of thermal systems, which move spontaneously towards the uniformity of equilibrium, and the ideal reversible behavior of classical particles. Besides providing a natural introduction to the investigation of the internal degrees of freedom, this theme serves as a link between classical and quantum behavior.

The argument is carried as far as possible, without becoming involved in detailed quantitative calculations. It assumes familiarity with simple differentiation and integration, in addition to a first course in elementary physics. Some of the mathematical details, and some topics which require treatment in greater detail that those in the text proper, are presented in Appendices.

The development sketched above can obviously be carried far beyond the point reached in this monograph. As it stands, it can be used as an introduction to a course which surveys thermodynamics and statistical mechanics. But it might also find a place in a less conventional course, one which sought to trace the development of quantum mechanics from classical ideas. It provides an example of the motivation of that development which is not as commonly stressed as others which are less directly related to "classical" macroscopic experiments.

A recent discussion of the different ways in which various theories in physics have developed included the following comment:

> "Two poles...are found: one attracts the minds that thirst for an explanation of the world; the other aggregate (attracts) those who look for order in the world, no matter what that order may mean. Thermodynamics was built by the second tribe, atomic theory by the first."

Granting this to be an accurate description of the historical origins of these two disciplines, it is nevertheless possible to believe that they no longer need be regarded as polar opposites, and this monograph is written with that conviction.

The apparatus in Fig. 6.1 was constructed by Professor H. Daw, who kindly provided the photograph, and also Figs. 6.2 and 6.3.

CONTENTS

1 MODEL-BUILDING IN PHYSICS

In studying heat and temperature we are dealing with ideas which are so deeply imbedded in common experience that we should expect them to pervade all of physics. But strangely, the student of mechanics hears hardly any mention of heat or temperature, yet mechanics is supposed to be the basic subject in physics. He likewise finds heat and temperature missing in his study of electricity, of magnetism, of optics. Let us try to see how this peculiar situation comes about. This effort will illuminate the interconnections among all of these subjects. It will also confirm the original expectation that heat and temperature are deeply involved in all natural phenomena, and help us to understand how this involvement is best expressed in precise physical terms.

Our puzzle is related to a question once posed to Albert Einstein. He was reminded that the Chinese had developed the compass, gunpowder, and printing by the fifteenth century and so were at that time far more advanced in science than Europe. Yet the Chinese proceeded very little further, while western civilization since the fifteenth century has been characterized by an enormous increase in the ability to understand and control nature. Why, Einstein was asked, did the Chinese fail to advance while the West, once so far behind, outstripped them?

Many answers to this question have been suggested, based on supposed differences in the basic outlook on life in the East as compared to the West. But Einstein's answer was simpler: it was that the failure of the Chinese was not surprising - the wonder was that progress was made anywhere, considering the complexity of natural phenomena.

In Europe, it happened that techniques of investigation were developed which facilitated this progress. This development, as we look back on it, was not due to a conscious analysis of what we now call the "scientific method." Some direct efforts of this kind were made, but they seem now to have been mostly sterile. Rather, progress was the result of trial and error; the "breakthrough," as newspapers would call it nowadays, was largely a matter of chance.

The techniques for investigating nature which turn out to be successful are based on a curious fact: It is profitable to approach the understanding of natural phenomena in what appears to be an indirect manner. Instead of trying to describe all aspects of a complicated phenomenon together, a simplified version is invented by isolating those factors which are of special importance to the particular problem under consideration. In this process of selection it is often convenient to consider the behavior of an invented abstract model rather than that of the real physical object involved in the phenomenon. The model can then be given such properties, and can be required to obey such physical laws, as result in behavior which reproduces the important aspects of the actual phenomenon. The criterion of "importance," of course, is decided in terms of each particular problem that is being investigated. The model and its properties, together with the physical laws, make up a theory of the phenomenon which can be tested by comparing its predictions with the actual observed behavior.[1]

---

[1] Sometimes the whole collection of concepts - abstract model (of the physical object), its properties and the physical laws are referred to collectively as a "model of the phenomenon." This can be confusing, and the confusion can be compounded by the use of the word "model" in still a different sense. For instance, a number of small balls connected by rods to a vertical shaft, so that when a crank is turned they rotate with appropriate speeds about a larger ball, is called a model of the solar system. This is clearly not the sense in which the word model was introduced above.

The whole roundabout procedure is somewhat similar to the behavior of a man who searches under a street lamp for the keys he lost, not because he dropped them there, but because there is no light elsewhere. The remarkable circumstance in the analogous enterprise of model-building is that it works: very often, the keys are found. Unfortunately, however, there is no magic formula for model-building. Its success depends on repeated attempts to isolate the significant factors or, to put it another way, to find the model appropriate to a particular problem. Every success is preceded by innumerable false starts which are found to lead up blind alleys.

2 THE PARTICLE MODEL FOR MOTION

2.1 AN EXAMPLE OF MODEL-BUILDING.

The development of understanding of motion of inanimate objects is a good example of the fitful way in which progress is made. The Greeks devoted much philosophical speculation to this problem but it was not until more than a millenium later, before the kind of attack described above began to be employed. The so-called Dark Ages were actually a period of intense activity on the problem of motion, and during this time concepts such as inertia and impulse were developed. These turned out to be fundamentally important in the model that was ultimately successful, so that Newton was not simply being modest when, centuries later, he attributed his seeing further than others to his standing on the shoulders of giants.

These efforts to understand motion culminated, in the hands of Galileo and Newton, in what is often called the "Newtonian model" of mechanics, which incorporates the classical laws of kinematics and dynamics. This model has withstood the test of time in the restricted range of application from which it was developed (macroscopic objects moving with speeds up to the order of those observed in the solar system).

Yet as we know, it contains no mention of heat or temperature. This omission is an example of one of the inherent limitations of model-building. Now it should come as no surprise to find that model-building is not a perfect instrument. The defect we have just discovered is one of several which must be understood if this technique is to be used properly. We will first examine some of the other defects of model-building. Then we will return to a discussion of the Newtonian model, which describes motion adequately (we will refer to its area of competence as "mechanical behavior"), and will see how it can be enlarged to incorporate the pervasive influence of heat and temperature on what we will call "thermal behavior."

The first of the other limitations of model-building that we consider is rather subtle, and was overlooked for a long time. It arises from the fact that a given model is usually based on a limited range of observation. Outside this range the model may fail to describe what is observed. For instance, Newtonian mechanics breaks down for very high particle speeds, that is, speeds approaching that of light. A new model (in this case, Einstein's special theory of relativity), must be devised to describe motions and interactions in the extreme range of speeds. But now, the existence of a satisfactory prior model imposes a condition on the new one. The new model must be compatible with the old in the original range of observation. That is, the predictions of the special theory of relativity must depend on speed, v, in such a way that as v approaches zero these predictions become indistinguishable from those of the Newtonian model.

Even within their appropriate range, models often have a limitation of another sort. We may say that we understand a phenomenon "in principle" when we know the laws obeyed by the model we have set up, and have satisfied ourselves that the model "works." We become convinced that it works as we find that calculations based on the model agree with observation, and also find that we never are faced with any discrepancies. But usually, even if the laws obeyed by the model seem to have a simple mathematical form, we find that we can perform the calculations only in a few very simple situations. Even cases which are only slightly more complex than the simplest may lead to formidable mathematical difficulties. Indeed, we may discover that, in general, actual solutions are very hard to obtain, or even impossible. In other words, the keys we find beneath the lamppost may not be the ones we are looking for.

The Newtonian model again presents a good example of this situation. When it is applied to the motion of objects which exert mutual attractive forces, it turns out that an explicit solution can be found only for the simplest case, namely two particles. For three or more particles, solutions can only be found for special configurations. In general, methods involving highly complex mathematics must be used even to find approximate solutions for more than two particles.

Fortunately, the solar system is a case in which solutions, accurate to any desired degree of accuracy, may be found by successive approximations. Such calculations lead to the highly successful picture we now possess of the planets and their moons moving in orbits determined by laws that are the same as those which determine terrestrial motion. But even here a practical difficulty enters. Over a century ago, Adams and Leverrier, working independently, each spent several years calculating the orbit of Uranus. From the departure of their calculations from the observed orbit, when only the influences of the known planets were used, they were led to predict the existence of a new planet beyond Uranus, and the existence of this new planet, later named Neptune, was soon confirmed. Such calculations must be carried out nowadays for the orbits of artificial satellites, but if we are interested in a rendezvous between two satellites the results are needed in minutes, not years. Therefore, even though the laws of mechanics have been known for three hundred years, such a rendezvous between two satellites becomes a feasible project only when high-speed computers become available.

Such technical questions, important though they may be to the satellite engineer, are not relevant when we ask, "Do we understand motions in the solar system?" The answer to this question is an almost unqualified "yes." The most important qualification is that the Newtonian theory is not adequate for motions under very strong gravitational forces. As a result, the motion of Mercury, the planet nearest the sun, cannot be calculated correctly using the Newtonian model. This is another example of the "range of observation" limitation discussed above. The extension required here is the General Theory of Relativity, first proposed by Einstein, and still a subject for active research.

We can now return to our original question, which was to explain the absence of the concepts of heat and temperature from the overwhelmingly successful Newtonian model for mechanics. Our answer at this stage might be summed up by the realization that this omission is no accident, but is rather an example of an inherent limitation of the model-building technique, as applied to the description of motion.

2.2 MECHANICAL STATES IN THE PARTICLE MODEL.

Let us now proceed to examine a related question: Is it possible to modify the mechanical model in such a way as to incorporate heat and temperature? As a first step in attacking this question it will be useful to recall some of the details of the simplest mechanical model, the "mass-point." In order to describe the motion of an object (without rotation), it can be represented by a "particle" or "mass-point" which is simply a geometrical point which has the mass of the object. If the motion of the object is all we are concerned with, its mass is its only significant attribute - its size, shape, color, etc., are all irrelevant. In the presence of special forces, e.g., electric fields, other parameters must be specified, in this case the electric charge, but this too is associated with the point particle. The motion of the object is described by giving the position of the particle as a function of time, which can be done most compactly by specifying the vector $\vec{r}(t)$ from an arbitrary origin to the position of the particle. (Note that we shall denote vector quantities by arrows.)

To say that the origin is arbitrary simply means that its choice is purely a matter of convenience in describing the motion of a given particle. Of course, different choices will result in different functions $\vec{r}(t)$, but the relation between any two such choices does not change with time, so long as the origins are fixed. For instance, Fig. 2.1 shows a particle which at time $t = t'$ is at position P. This position is described by two vectors, $\vec{r}_1(t)$ and $\vec{r}_2(t)$, drawn from the two origins O_1 and O_2, respectively. The vector $\vec{R}_{21}$, from O_2 to O_1 connects the two position vectors according to the relation

$$\vec{r}_2(t') = \vec{r}_1(t') + \vec{R}_{21}. \tag{2.1}$$

It should be clear that this relation holds for all times t, not merely $t = t'$, so long as the two origins are fixed, for then the vector $\vec{R}_{21}$ is a constant, independent of time. The vector $\vec{r}(t)$ can also be written

$$\vec{r}(t) = \vec{x}(t) + \vec{y}(t) + \vec{z}(t), \tag{2.2}$$

where the vectors $\vec{x}(t)$, $\vec{y}(t)$, $\vec{z}(t)$ are projections of $\vec{r}(t)$ along arbitrary Cartesian axes. Another way of writing this relation is

$$\vec{r}(t) = x(t)\,\vec{i} + y(t)\,\vec{j} + z(t)\,\vec{k}, \tag{2.3}$$

where now $x(t)$, $y(t)$, $z(t)$ are the scalar components of $\vec{r}(t)$ and $\vec{i}$, $\vec{j}$, $\vec{k}$, are unit vectors along the Cartesian axes.

A basic condition that must be met if an object is to be represented by a particle is that it can be distinguished from its surroundings, and, in principle, isolated from them. The separation of "object" from "surroundings" may sometimes be a highly idealized operation. For instance, it is sometimes useful to think of a small volume in a large mass of liquid as the object, the rest of the liquid then constituting the surroundings. Although this is an extreme example, a little reflection will show that the separation usually involves some degree of idealization. A billiard ball resting on a flat, smooth surface, for instance, might appear to be a case with no need for idealization, until one begins to inquire as to the exact nature of the boundary between ball and surface. Especially if they were made of the same material, it might be difficult to decide which atoms at the boundary belonged to "object" and which to "surroundings."

In applying the particle model, we imagine that such decisions have somehow been made. Appropriate decisions of this sort can be recognized, since conclusions drawn from them will not conflict with observation. The real significance of the separation, as far as the model is concerned, is that the motion of the object is determined by its interactions with the surroundings. As one extreme possibility, there are no such interactions at all; the particle is "isolated." According to Newton's First Law (which we shall refer to by the shorthand N-I), such a particle is in "equilibrium" and its motion is not arbitrary. The position vector $\vec{r}(t)$ must satisfy the condition that its time derivative is constant:

$$d\vec{r}(t)/dt = \text{const.} = \vec{v}(0) = \dot{\vec{r}}(0), \tag{2.4}$$

where the last two symbols, $\vec{v}(0)$ and $\dot{\vec{r}}(0)$, are simply two alternative ways of writing $d\vec{r}(t)/dt$, the velocity. Note that a dot over a vector will always mean the time derivative of that vector, and two dots will denote the second time derivative.

The notation $\vec{v}(0)$ or $\dot{\vec{r}}(0)$ is chosen to emphasize that the velocity of a particle in equilibrium does not change with time. It always maintains the value it has at $t = 0$, the (arbitrary) origin of the time scale.

A particle may be in equilibrium even if it is not isolated. The net effect of all its interactions with its surroundings may vanish, and once again Eq. (2.4) will hold. Whenever a particle is in equilibrium, that is, whenever its velocity is constant, its position vector satisfies a simple relation, which is obtained by integrating Eq. (2.4). This relation is

$$\vec{r}(t) = \vec{r}(0) + \dot{\vec{r}}(0)\ t, \qquad (2.5)$$

where $\vec{r}(0)$ is the position vector of the particle at $t = 0$ and $\dot{\vec{r}}(0)$ is its velocity vector at the same time. Hence specifying $\vec{r}(0)$ and $\dot{\vec{r}}(0)$ completely determines $\vec{r}(t)$ at any other time.

This pair of vectors, position and velocity at a given instant, are said to define the "mechanical state" of a particle at that instant. The statement at the end of the last paragraph can now be restated as follows: if we know the mechanical state of a particle in equilibrium at any instant, say $t = t'$, then we can determine its mechanical state at any other instant, say $t = t''$. For Eq. (2.5) can be rewritten

$$\vec{r}(t'') = \vec{r}(t') + \dot{\vec{r}}(t')(t'' - t'), \qquad (2.6)$$

and also,

$$\dot{\vec{r}}(t'') = \dot{\vec{r}}(t'). \qquad (2.7)$$

The left-hand sides of these equations give the mechanical state at $t = t''$, and these vectors are determined by the right-hand sides if we know $\vec{r}(t')$ and $\dot{\vec{r}}(t')$, the mechanical state at $t = t'$.

We can make the same assertion of a connection between the mechanical states of a particle at two different times, even if the interaction of a particle with its surroundings does not vanish (that is, even if it is not in equilibrium), provided that we know the nature of the interaction. This interaction is represented by the resultant force $\vec{F}$ (the vector sum of all the forces), acting on the particle. When the resultant force is not zero, the particle velocity is not constant. Its time rate of change is given by Newton's Second Law (which we will refer to as N-II), as

$$d\vec{v}(t)/dt = \vec{a}(t) = \vec{F}(t)/m \qquad (2.8)$$

where $\vec{a}$ is the acceleration vector and m is the mass of the particle. Integration gives

$$\vec{v}(t'') = \vec{v}(t') + \int_{t'}^{t''} \vec{a}(t)\ dt = \vec{v}(t') + \int_{t'}^{t''} \vec{F}(t)/m\ dt \qquad (2.9)$$

Since now $\vec{v}(t)$ is not constant, Eq. (2.4) now reads

$$d\vec{r}(t)/dt = \vec{v}(t),$$

which, when integrated, gives the gen-

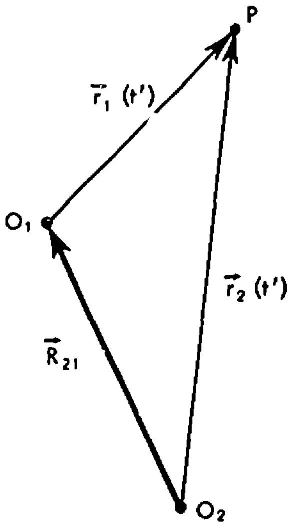

Fig. 2.1 The relation between position vector $\vec{r}_1$ (origin O_1) and position vector $\vec{r}_2$ (origin O_2) for a particle at P, and the vector $\vec{R}_{21}$ from O_2 to O_1.

eralization of Eq. (2.5) to the nonequilibrium case:

$$\vec{r}(t'') = \vec{r}(t') + \int_{t'}^{t''} \vec{v}(t)\, dt. \quad (2.10)$$

Hence, in the nonequilibrium case also, if we know the mechanical state at a certain time, say $t = t'$ (that is, if we know $\vec{r}(t')$ and $\dot{\vec{r}}(t')$), we can find the mechanical state at any other time, say $t = t''$ (that is, we can find $\vec{r}(t'')$ and $\dot{\vec{r}}(t'')$). In this case we must use Eqs. (2.9) and (2.10), which require that we know the interaction between the particle and its surroundings, that is, we must know the net force $\vec{F}(t)$ acting on the particle. When $\vec{F}(t) = 0$, the particle is in equilibrium and Eqs. (2.6) and (2.7) apply.

3 THE DIRECTION OF THE FLOW OF TIME

3.1 TIME REVERSAL IN THE PARTICLE MODEL.

For our present purpose, namely, discussing the connection between mechanical and thermal behavior, the most significant feature of these relations is that the acceleration, which is determined by the force, is the second time derivative of the position vector. As a consequence, for a given force the acceleration is unchanged if we reverse the direction of time, that is, if we differentiate twice with respect to $-t$ instead of t. One result of this property of N-II is that the mechanical state at $t = 0$ not only determines mechanical states at later times (t positive), but also at earlier times (t negative). But a more interesting consequence is that there is really no distinction between "later" and "earlier." We have already remarked that there is no special significance to the instant at which we set $t = 0$, that is, to the time at which we start our clock. Now we see that in addition, it makes no difference in which direction the clock hands move after they start.

This irrelevance of the direction of time in the Newtonian model certainly contradicts our fundamental experience of what we call "the passage of time." We shall see later that the ideas of heat and temperature are immediately related to our experience that time does indeed flow in just one direction. But first let us see in some detail how the Newtonian model accommodates itself to this peculiar insensitivity to the direction of the flow of time.

Let us consider a very simple case, that of an object falling freely and accelerating because of the gravitational force acting on it. No one who has sat through home movies taken at a swimming pool has been spared the joke of the operator who reverses the motor, so that the diver stops in midair and then gracefully rises to his original position on the board. This is recognized as a ludicrous scene, but it is by no means clear, while watching it, which of the segments is "forward" and which is "reverse." The "diver" might actually have been an acrobat jumping from an unseen trampoline, so the rise to the diving board might have been the "ordinary" motion, forward in time, while the "dive" was the same motion, projected in reverse. After all, the segments can be shown in either order by splicing the film. So there is as yet no clear-cut way to determine which is the "proper" direction of time, the direction of experience. Let us see if any further insight is gained by examining the problem analytically, using the relations we have stated above.

If we wish to follow the motion of a particle subject to certain forces, we have seen that the motion can be described as the passage through a succession of mechanical states, each described by the pair of vectors, $\vec{r}(t)$ and $\vec{v}(t)$. If at a certain instant, say $t = t_0$, we (somehow) reverse the direction of time, then the mechanical state suddenly changes, since $d\vec{r}/dt = -d\vec{r}/d(-t)$. That is, the velocity reverses its direction if time is reversed. Now according to N-II, changes in velocity are associated with forces. Hence we can tell that time is reversed at $t = t_0$ by observing that the velocity reverses without the application of a force. But the question remains, which of the two directions of time is "correct"?

Let us reduce the diver, or acrobat, in the movie, to a particle which is thrown upwards at $t = 0$ with a certain speed v_0, so that

$$\vec{v}(0) = v_0\,\vec{k},$$

where $\vec{k}$ is the unit vector in the vertical (z) direction. Then the force and acceleration are

$$\vec{F} = -mg\vec{k}, \quad \vec{a} = -g\vec{k}$$

where g is the gravitational acceleration. The particle will decelerate while rising until it momentarily comes to rest at the height h, say at time $t = t_h$, so its mechanical state is then

$$z(t_h) = h; \quad v(t_h) = 0,$$

(is this an equilibrium state?), and thereafter it descends. In the course of its descent it does not pass through the same mechanical states as during its ascent, since it passes through each position with a velocity that is just the negative of the velocity it had at the same position, when it was rising. The acceleration, however, is the same throughout the entire motion, descent as well as ascent. If now time had been reversed at the top of the path, that is, at $t = t_h$, the subsequent motion, now from t_h backward to $t = 0$, would have been exactly the same as the "actual" fall during the interval $t = t_h$ to $t = 2t_h$, since, as we have seen, velocities change sign when t is reversed. So our analysis still provides no way of knowing whether we are seeing the particle falling, or instead, watching its rise "in reverse."

Figure 3.1 is a plot of v(t) for both motions, with the forward motion represented by a solid line, and the reverse motion by a dashed line. These two motions are seen as distinct be-

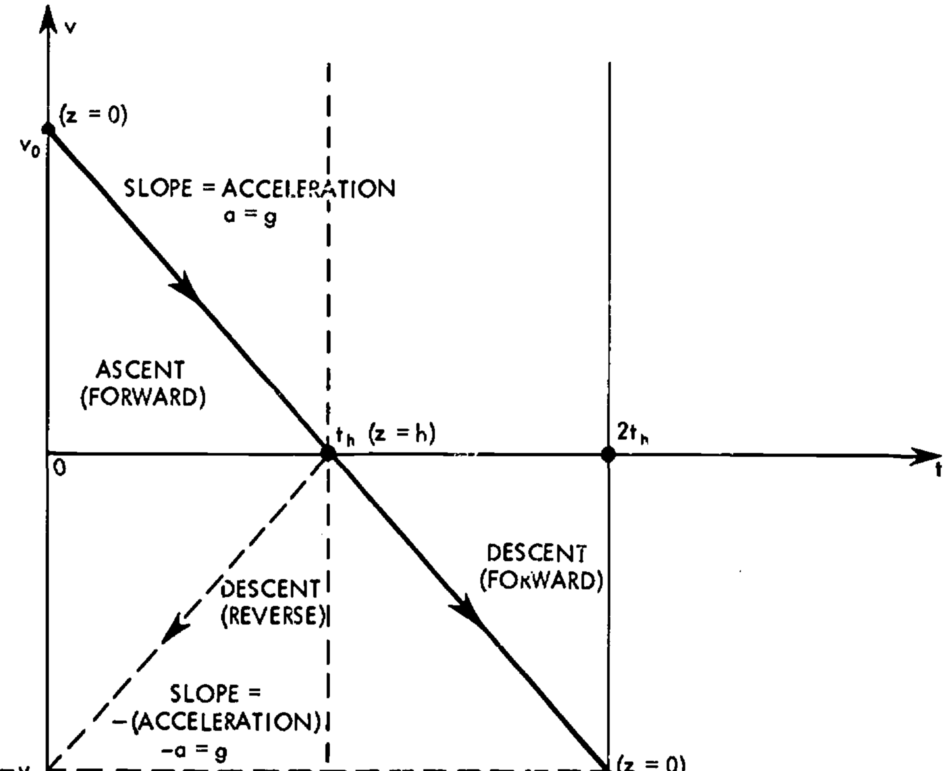

Fig. 3.1 v(t) for ball thrown upwards at $t = 0$ with initial speed $v = v_0$. Arrows indicate direction of time flow: forward on solid line; reversed at $t = t_h$ for heavy dashed line. Idealized case-gravitational force only.

cause they are both plotted along the same time axis, that corresponding to forward motion. If the positive direction along the axis were changed at $t = t_h$ to correspond to the reversal of time at that instant, this would result in a reflection of the dashed line in the vertical line through $t = t_h$, and the two "descent" lines would superimpose exactly. We have asked the question, "Which direction of time is 'correct'?" The particle model answers, unequivocally, "Both!"

This seemingly trivial example is worth discussing in such detail because the conclusion that can be drawn from it is in fact completely general. Stated another way, this conclusion is as follows: As far as the Newtonian particle model is concerned, any motion whatever is reversible, in the sense that if it is a possible motion when viewed with time flowing in one direction, it is also a possible motion when viewed with time flowing in the reverse direction. Now nothing could be further from experience. The "correct" direction for the flow of time - and as far as all our experience is concerned, the only possible direction - is the direction determined for us by the fact that all observed motions are clearly impossible if viewed with time flowing backwards. We can use the example under discussion as an illustration.

3.2 THE IRREVERSIBILITY OF NATURAL MOTIONS.

So far, the description of the motion of the particle thrown upwards has been an idealization which neglected all frictional effects. We can imagine how the particle would behave in this ideal case because we know that air friction diminishes as air pressure decreases. Hence we have been describing motion along a path in a perfect vacuum. But if the path is through the air or any other fluid (gas or liquid), an object which falls long enough will stop accelerating. Its speed approaches a limit, which we call the "terminal speed" and denote by v_T. The function $v(t)$ for this natural motion, as distinguished from the ideal vacuum case described above, is shown in Fig. 3.2. Here the particle starts falling from $z = 0$ at $t = 0$ (instead of from $z = h$ at $t = t_h$, as in Fig. 3.1) so that v is always negative and approaches $-v_T$ as t increases without limit. If time is reversed at $t = t_R$ (dashed vertical line in the figure) the speed becomes positive and decreases to zero at $t = 0$. Both paths are superimposed in the figure by a simple trick - the ordinate axis representing speed is reversed at $t = t_R$. Hence the left-hand ordinate axis, with speed increasing upward, is to be used for the forward motion, $t = 0$ to $t = t_R$, and the right-hand ordinate axis (dashed line), is to be used for the reverse motion, $t = t_R$ back to $t = 0$. The functional form of $v(t)$, which is given in the caption of the figure, is derived in Appendix 1.

Now the assertion was made above that all natural motions are impossible when reversed. In our present motion this would mean that the ascent from $z = z(t_R)$, which is negative, to $z = 0$ is an impossible motion. In order to see this impossibility clearly, it is helpful to extend the particle model somewhat. Before making this extension, let us discuss a related motion. Imagine that the object bounces from a horizontal surface, reversing its velocity at each bounce. The height to which it rises diminishes after each bounce, and ultimately it comes to rest on the surface; this is the "natural" motion. The reverse motion, in which an object at rest on the surface suddenly begins to bounce without any change in the forces acting on it, and rises higher and higher with each bounce, is clearly impossible.

The reverse motion in Fig. 3.2 can likewise be seen to be impossible if we extend the particle model by introducing the quantity "mechanical energy." As we know, this quantity, which we will denote E, contains a con-

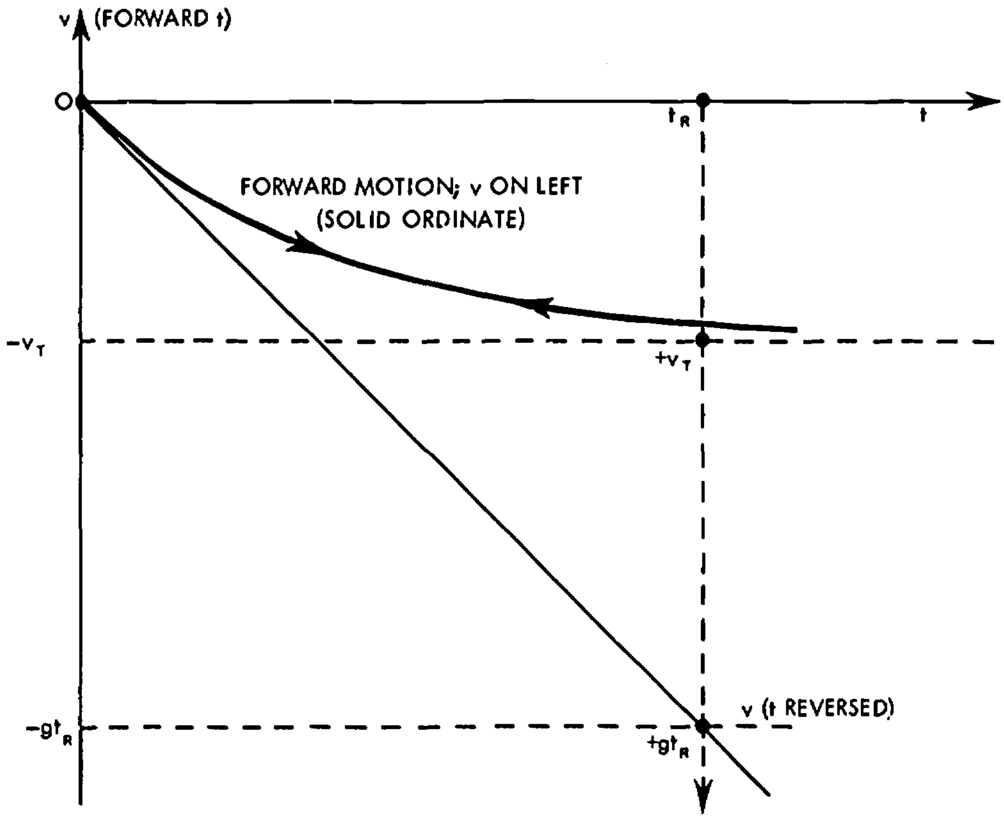

Fig. 3.2 Here, v(t) for free fall, starting at $t = 0$, with retarding force proportional to speed, is compared with $v = -gt$ (no retarding force). Time is reversed at $t = t_R$. Speed for forward motion is given on left (solid) ordinate; for reversed motion, on right (dashed) ordinate. The speed in the forward motion is given by

$$v(t) = v_T\left[\exp - \left(\frac{g}{v_T}\right) - 1\right].$$

tribution from the motion of the particle, namely the kinetic energy, denoted K, defined by

$$K = \tfrac{1}{2}mv^2 = \tfrac{1}{2}m(v_x{}^2 + v_y{}^2 + v_z{}^2). \quad (3.1)$$

If a conservative force, such as gravitation, acts on the particle, E also has a contribution called the potential energy, denoted V, which in general depends on position. Hence E is a function of the mechanical state of the particle.

We pointed out in connection with Fig. 2.1 that the choice of position vector for a particle was arbitrary, since the origin may be moved from one point to another by the addition of a constant vector. The value of potential energy V, will then differ by a constant for the two choices of origin. Correspondingly, an arbitrary constant value can be assigned to V for any choice of origin. For the freely falling particle, it is convenient to choose the point from which it is dropped, $z = 0$, as origin, and to assign the value $V(0) = 0$. Then as a function of z,

$$V(z) = mgz. \qquad (3.2)$$

We can also express K as a function of z, since in free fall

$$v = -gt; \qquad z = \tfrac{1}{2}gt^2, \qquad (3.3)$$

so that

$$K = \tfrac{1}{2}mg^2t^2 = mg(\tfrac{1}{2}gt^2) = -mgz = -V. \tag{3.4}$$

(Note that z is negative for the falling particle, so K is positive and V is negative.) Hence, in free fall $E = K + V = 0$ always. As we have seen, the numerical value of the mechanical energy is arbitrary; but this example shows that if only conservative forces act on a particle the mechanical energy is constant. For the mechanical energy to increase, a nonconservative force must act on the particle in the direction of its motion.

Now in Fig. 3.2, it is clear that v in the retarding medium is always smaller in magnitude than v for free fall. Hence at any position, K (retarded), will be less than K (free fall), and the difference between the two increases as the particle falls. Since V depends only on position, E (retarded), is also less than E (free fall), by an ever increasing amount. In other words, E decreases in the forward motion (natural motion), of the particle falling in a retarding medium. In the reverse motion, therefore, E would be increasing as the motion proceeds, although the only nonconservative force acting on the particle, the frictional force, acts in a direction opposite to its motion. Such an increase in E under these conditions is never observed in natural motions, so we can finally distinguish the "correct" direction for the flow of time. The decrease of mechanical energy due to the action of frictional forces (or "dissipation"), is described analytically in Appendix 2.

From our discussion so far, we must conclude that the possibility of distinguishing the natural direction for the flow of time in the Newtonian model rests on the universal occurrence of frictional forces in the motion of terrestial objects.[2] Now Galileo abstracted the law of equilibrium (N-I), from common experience, by the device of imagining the ideal limit in which frictional forces vanish. Hence it is not surprising that when dissipation is eliminated, the distinction of a unique direction for the flow of time also disappears.

One further question can be answered in a simple quantitative way: When is a particular motion adequately represented by the dissipationless, ideal approximation? Suppose the motion is observed during the interval $t'' - t' = \Delta t$. This might be the period of rotation of the moon, or the time between two bounces of a ball. During this time the mechanical energy changes by

$$E(t'') - E(t') = \Delta E,$$

and if the rate of dissipation of mechanical energy is dE/dt, then

$$\Delta E = (dE/dt)\,\Delta t,$$

which is negative (E decreasing), since dE/dt is negative. Now for the dissipationless approximation to be valid, it must be true that

$$|\Delta E| \ll E \quad \text{or} \quad |dE/dt| \ll E/\Delta t. \tag{3.5}$$

---

[2] Frictional forces also occur in astronomical motions (for instance, the tides), but in these cases, the rate at which mechanical energy is dissipated, compared to the total mechanical energy involved, is very much less than for ordinary terrestrial motions. For instance, billions of years were required for the dissipation, by tidal forces, of the mechanical energy that the moon once possessed when it rotated about its own axis. A bouncing ball, on the other hand, comes to rest in seconds.

4 TEMPERATURE AND THERMAL EQUILIBRIUM

4.1 THERMAL EQUILIBRIUM AND THE DEFINITION OF TEMPERATURE.

We are now in a position to give at least a partial answer to the question we posed at the outset of our discussion, namely, why the concepts of heat and temperature are absent from the theory of motion, or mechanics. Heat, we know, is produced in a great variety of ways: chemical transformations (as in burning fuel), radiation, etc. The manner of producing heat most directly associated with motion, however, is just the dissipation process we have been discussing. Hence the particle model of mechanics, which proceeds from the law of equilibrium, loses contact with thermal phenomena when it makes dissipation a secondary phenomenon. We have already seen how this same step also loses the distinction of a unique direction of time, and how this unique direction can be recognized by observing the effect of dissipation: It is the direction of time for which dE/dt due to frictional forces is negative. Later we will justify the remark made above, that a unique direction of time is directly associated with thermal phenomena.

We raised another question as well, namely how mechanical and thermal phenomena were to be connected by becoming related parts of an enlarged model. This problem is considerably more complex than the first, since each area developed almost independently of the other. Their major point of contact had the character we described earlier as "technical," rather than being a part of the logical structure of either mechanical or thermal phenomena.

The contact between the two kinds of phenomena is the term $\vec{F}$, the force, in N-II. The laws of motion, generalized to include rotation of extended objects, can be used for the analysis of complex machines as well as of planets. In this analysis, the value of $\vec{F}$, applied to the machine through a shaft, for instance, must be known, but the manner in which the force is produced by the prime mover which turns the shaft is immaterial. On the other hand, the first important prime movers of the Industrial Revolution were steam engines, which produced forces through the operation of thermal processes. The analysis of these engines was one of the most important motivations for understanding thermal phenomena, and all sorts of thermal engines are still of tremendous technical importance. As a result, the study of thermal phenomena often proceeds, even today, from a consideration of how heat can be turned into work. But clearly, this point of view is not likely to illuminate the connection between mechanical and thermal phenomena, which is of primary interest to us.

We therefore proceed in a different fashion. Our first step will be to refine the intuitive notion of temperature, which is a basic element in thermal phenomena, using a formulation related as closely as possible to mechanical phenomena. The beginning of the discussion is already contained in our earlier description of dissipation, because we know that an object falling in a retarding medium becomes "warmer" as it falls. There is, however, no simple way of incorporating this fact in the particle model, since the particle possesses only mass, and the mass does not change.

What we seek, therefore, is a way of describing the notion of temperature in terms of the particle model, and of making it quantitative. To do this we must investigate something more complicated than a mass-point; indeed, as we shall see later, the concept "temperature" has no meaning at all when applied to a single particle. In this investigation, we shall have to explore rather carefully some aspects of the behavior of matter which are so much a part of common experience that they are often taken for granted. When we fill a pot with water from the kitchen faucet, we usually find it colder than the air in the room. If it stands on the table for a while, it warms up to room temperature. In order to bring it to a still higher temperature, we must place it in contact with the heating element on the stove. But just because these phenomena are so "ordinary," the conclusions we shall be able to draw from them have a very wide range of applicability. Hence we can use simple examples in the discussions and still achieve results of wide generality.

Let us construct a simple pendulum with a bob supported by a thin metal wire, hung from the top of a transparent bell jar which can be evacuated, or, if we wish, filled with gas at greater than atmospheric pressure. If we start the pendulum swinging, we know that it will ultimately come to rest, and we expect that the rate at which the amplitude, A, decreases (namely, $dA/dt = \dot{A}$), will depend on the pressure of the gas in the bell jar. As the jar is evacuated, $\dot{A}$ decreases in magnitude, but it remains finite even at very small pressures, since even when gas friction has become negligible, the internal friction in the wire continues to dissipate mechanical energy as the bob swings. Under these circumstances, that is, the bob coming to rest in the evacuated bell jar, we can observe a definite mechanical effect associated with the dissipation of mechanical energy: The wire is longer after the mechanical energy is dissipated and the bob is at rest than it was when the motion was first started. It is also warmer, and so is the bob, so we suggest that the length of the wire be associated with the degree of warmth, or "temperature" of the bob.

Unfortunately, it is not a trivial matter to turn this simple observation into a quantitative association of wire length with temperature, and to establish an actual temperature scale as those in everyday use. For in departing from the particle model, we have opened a Pandora's box of new questions about our system. The mechanical state of a particle could be described by two vectors, $\vec{r}$ and $\vec{v}$, together with m, the mass of the particle.[3] But in describing our system, if we wish to use the length of the wire to specify the "state" of the system in some new sense which transcends that of the simple mechanical state, we must include much more information. Suppose we repeat our observation many times, each time with the original length of wire and the initial height of the bob. The same amount of mechanical energy will be dissipated each time, but we will find that the final length of the wire will depend on the particular metal used, and also on the material of the bob. Our new sort of state, which we will refer to as "thermal state" for the moment, would be clumsy indeed if this additional information always had to be included in its specification.

Fortunately, this turns out not to be the case. We shall see that it is possible to specify precisely what is meant by "thermal state" by introducing just one additional parameter, which, as suggested above, is the temperature. And, most important, the temperature can be given a quantitative meaning which is independent of the properties of specific materials. The price of this generality is that the definition is somewhat abstract, so we have to describe it with care. In doing so, we will quote the result of much careful observation, made especially during the eighteenth and nineteenth centuries, that preceded the final formulation of the definition.

To make the discussion more concrete, we will continue to use our sample system, the pendulum in a bell jar,

---

[3] In fact, the mass can be absorbed into the specification of the mechanical state, so just two vectors suffice. Since momentum $\vec{p}$ is given by $\vec{p} = m\vec{v}$, the two vectors $\vec{r}$, $\vec{p}$ completely describe the mechanical state of a particle of mass m.

as shown in Fig. 4.1. One new feature has been added: a plate mounted atop the jar. The suspension wire is connected to the lower part of the plate, which passes through the top of the jar. With this arrangement, we can conceive of making a variety of observations on the bob or on the wire, either from outside the bell jar, or, if necessary, by putting additional apparatus inside. For instance, we can measure the elongation of the wire with a telescope outside the jar; in the same way we can measure the size of the bob and thus determine its density. We can measure various optical properties such as reflectivity or index of refraction (if the bob is of transparent material), also from the outside. By inserting suitable instruments in the jar we could measure the bob's electrical resistivity or its magnetic moment. We could even determine its resistance to mechanical deformation (compressibility), and other mechanical properties. We shall forbear from extending this list of measurable properties of the system at this point, but clearly not because we have exhausted all possibilities. These properties are among the contents of the Pandora's box that we mentioned earlier. Let us indicate all such properties by the symbol P_i, where the index i distinguishes among the properties. In the short list we have given i runs from 1 to 6; we shall allow it to run to n, where the actual value of n is limited only by our imagination and ingenuity. We can also imagine making a set of measurements of all the properties at about the same time; we indicate such a set of simultaneous measurements by the symbol (P_i). Thus the symbol P_i denotes a particular property, as, for instance, the density; but we shall also speak of the value P_i, by which we shall mean the result of a measurement of the property P_i.

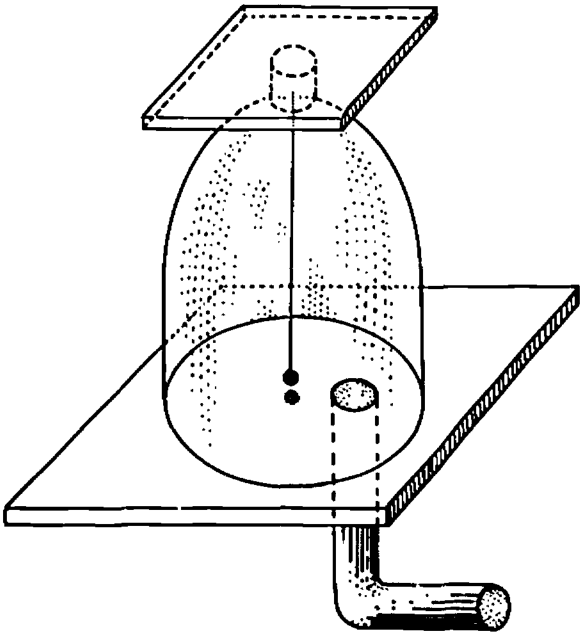

Fig. 4.1 Bell jar with pendulum bob and supporting wire inside, and external plate attached to upper end of wire. Jar can be evacuated or filled with gas through hole in plate and connecting tube.

Now there are two particular characteristics of the values P_i that are of interest to us here. The first is the way in which they change with time, that is, the values $dP_i/dt = \dot{P}_i$. The second is the way in which the P_i depend on what goes on outside the bell jar. This latter characteristic can be modified at will, between extreme limits. If we make the plate of wood, or better, of asbestos, or better yet, if we remove it entirely; and if we exhaust as much air as possible from within the jar, we then find that about the only way we can influence the values P_i from outside the jar is to shine light through it. The changes in the values P_i induced in this manner are not arbitrary. They are in the same direction as observed when the bob comes to rest after having been given an initial push. They also depend on the intensity and duration of the light in the same way as they depend on the magnitude of the push. If, in addition, we coat the outside of the jar with some highly reflecting material, we find that (P_i) is essentially independent of whatever happens outside the jar. In this case, we can speak of any system inside the jar as being isolated.

On the other hand, with a transparent jar and air or some other gas within it, and a metal plate on top, we find that the values (P_i) will change measurably if the surroundings change. Furthermore, they can now change in different directions; if we denote by ΔP_i a change in a value P_i, the changes ΔP_i observed when we put some ice on the metal plate will in general have signs opposite from the ΔP_i that occur when we put a container of boiling water on the plate. Under these circumstances, the system inside the jar is clearly no longer isolated from its surroundings. But we have in no case done mechanical work on the system; we have not, for instance, started it swinging in order to bring about changes in the values (P_i). To emphasize this point, we shall speak of "thermal contact" of the system with its surroundings when the changes in the values (P_i) do not depend on changes in the mechanical state of the system or of any part of its surroundings.

Now that we have carefully described the possible thermal relations between a system and its surroundings - from perfect thermal isolation (values (P_i) completely independent of surroundings), to varying degrees of thermal contact (values (P_i) depending on surroundings to a lesser or greater extent) - we can consider the other characteristics of the values that we mentioned, namely, their time rates of change, $(\dot{P}_i)$. The formal statement of the behavior of (P_i) is quite simple, but some discussion is required to appreciate its significance. So first we give the formal statement:

> Under certain special circumstances, each of the values P_i (in the set (P_i) of all the values of properties of a system measured at a certain time), is constant; that is, $\dot{P}_i = 0$ for all values.

First of all, we should recognize that this statement is not true in general, that is, for a system chosen arbitrarily, under unspecified conditions. In general, we may find that the set of values (P_i) measured at one time will bear no special relation to the set measured at another time. Hence the formal statement above is not trivial; the circumstances under which it holds must indeed be "special."

We now have to inquire: What are the special circumstances? What is the significance of this statement for the thermal behavior of the system? The answer to the first question depends on whether the system we are considering is thermally isolated or in thermal contact with its surroundings. If we take any system (one which may initially be in some sort of thermal contact with its surroundings), and isolate it thermally, we find that if we wait long enough, the statement above will apply. That is, the values of the properties will change at rates $(\dot{P}_i)$ which diminish to zero, and finally reach a set of values (P_i) which thereafter remain constant. Hence one of the special circumstances referred to in the statement is simply thermal isolation. Alternatively, if we have a system in thermal contact with its surroundings and arrange as carefully as we can that the surroundings do not change, then again the statement will apply. From this point on, let us give a name to the condition specified by the statement, that all values P_i are constant in time, or all $\dot{P}_i = 0$; we will call this condition "thermal equilibrium" and when it is satisfied, we will say that the system is in a "thermal equilibrium state." We can differentiate among different thermal equilibrium states by the set (P_i) of constant values of the properties measured for each such state.

This definition of a thermal equilibrium state is quite precise, but vastly more complicated than the definition of a mechanical equilibrium state. There, just the two vectors $\vec{r}$ and $\vec{p}$ sufficed; here, we need the set of values (P_i), $i = 1, 2, \ldots, n$, where there is no obvious limit to the size of n. The definition of thermal equilibrium could clearly have no useful role in a theory of thermal processes if it had to remain in this form. But remarkably enough, it is possible to abstract from this definition just one new parameter, which implicitly contains the same information regarding thermal equilibrium as does the entire set of values (P_i). Let us see how this can be done, and why this new parameter is associated with the intuitive notion of temperature.

Our discussion rests on an empirical relation among various thermal equilibrium states, which can be described in the following way. Suppose we have ascertained that two different systems, initially quite independent of each other, are each in thermal equilibrium. The systems might each be a pendulum in a bell jar as in our earlier discussion, or they might be of entirely different types, not necessarily the same for each. Now let us put the two systems in thermal contact with each other. If the systems each were of the type sketched in Fig. 4.1, for instance, this could be done by placing the bottom plate of one (system A), on the top plate of the other (system B). Before being placed in thermal contact, they each had their respective sets of values (P_{iA}), (P_{iB}). Common experience tells us that after being placed in thermal contact, they will come to mutual thermal equilibrium and will then have respective sets of values, (P'_{iA}), (P'_{iB}), which are in general different from the original sets of values, before they were placed in thermal contact. Now that they are in mutual thermal equilibrium, their sets of values have a special property. If we now isolate them we know that their respective thermal equilibrium states will not change; while in isolation, the properties retain their respective values (P'_{iA}), (P'_{iB}). Therefore, if we return them to mutual thermal equilibrium, we observe a special kind of behavior, different from what we find in general: These separated systems, when brought into thermal contact, retain the sets of values (P'_{iA}), (P'_{iB}), which they had before thermal contact.

In other words, the values (P'_{iA}), (P'_{iB}), are invariant with respect to mutual thermal contact of systems A and B. Now these "other words" are not only more elegant than the original description, they are also more useful. For whenever we find an aspect of a physical system which is invariant in some context, we can define a new physical quantity in terms of that invariance. Mechanics abounds with examples: indeed mass, energy, momentum, etc., are all useful in describing mechanical phenomena precisely because each is related to some sort of invariant behavior of a mechanical system.

So we have finally arrived (almost), at the end of our tortuous path leading to the definition of temperature: Whenever two separated systems are each in a thermal equilibrium state that has the property of remaining unchanged after being brought into thermal contact, we say that these two states have the same temperature. It is easy to see that this definition can be expanded beyond two states, to include any number whatever. This can also be done without bringing all the states into mutual thermal equilibrium, for it is also a fact of experience that if system A and system B are in mutual thermal equilibrium (have the same temperature), and A is also separately in thermal equilibrium with system C, then B will be found to be in thermal equilibrium with C if they are placed in mutual thermal contact. Thus A, B, C all have the same temperature, so this can be established for any number of systems by observing their behavior in mutual thermal equilibrium, pair by pair.

We will describe in the next section how this property of systems in mutual thermal equilibrium, their common temperature, can be given a quantitative form. That is, we will show how to assign a number, which we will denote T, to represent the temperature of a thermal equilibrium state. Assuming for the moment that we have already done this, we now have another value to add to the set (P_i) associated with each thermal equilibrium state. But this number T is not on an equal footing with all the rest in the set (P_i), for it completely specifies the thermal equilibrium state, in a way which is not shared by any of the other values.

In order to see the meaning of this statement, that the temperature has a special significance which distinguishes it from the other properties of a system, let us return to the operation we considered in defining the temperature. We brought two systems, initially separated and each in thermal equilibrium, into thermal contact, and used the outcome of this operation to define temperature. Now let us consider instead the converse question: We know the initial values of the properties in thermal equilibrium, (P_{iA}), (P_{iB}), and we would like to *predict* the outcome. But our Pandora's box contained a large number of properties that can be defined for any particular system, and an enormous number of different conceivable systems. As a result, making even this simple prediction becomes a major task.

In special cases, the answer is simple. For instance, if the two systems are the same, and if just one of the properties, say the density, has different initial values, then they will not be in mutual thermal equilibrium when brought into thermal contact; their values of all their properties will change and ultimately they will each arrive at new thermal equilibrium states. But if the two systems are different, or if two systems of the same kind have different initial values for several of their properties, we cannot make a prediction so easily. Either result is possible, depending on the particular set of initial values (P_{iA}), (P_{iB}). We could make the prediction if we had made the identical observation before, or if we had made a whole series of observations, so that we could relate all the different values of each P_i which are found in different thermal equilibrium states.

But all this complexity is reduced to trivial simplicity if we remember the definition of temperature. No two distinct thermal equilibrium states have the same temperature, even though they may share common values of other properties. For instance, we may find the same density in states corresponding to different temperatures. Conversely, we may find *different* densities in states corresponding to the same temperature. So our prediction can be made solely on the basis of whether or not $T_A = T_B$ before the systems are brought into thermal contact. We need not inquire at all into the values of any of the other properties. It is in this sense that temperature, T, defines a thermal equilibrium state completely, and implicitly conveys information also contained in the entire set of values (P_i).

4.2 TEMPERATURE SCALES.

In the last section, we defined "temperature" abstractly, as the common property of all systems in mutual thermal equilibrium. But this definition does not suffice to establish temperature scale. In order to define a temperature scale, we need a definite procedure for assigning, to each set of thermal equilibrium states in mutual thermal equilibrium, a number to represent T. Such a quantitative scale is needed in all applications of the notion of temperature, for instance, in the mundane task of reporting the weather. (What is the "system" in this case?) However, our interest in discussing how to establish a temperature scale goes beyond a concern with practical applications. We recall that we set ourselves the task of bridging the gap between mechanical and thermal phenomena by incorporating both in a consistent and coherent model of nature. We were led to our abstract definition of temperature while working at this task. The problem we face now is a typical one, which arises whenever an abstract definition must be given a concrete form. In the case of temperature, the problem is especially complex, since our definition was couched in a particularly abstract and mathematical form.

We might ask, however, was this approach, from the abstract to the concrete, really wise? Might it not have been better to go in the opposite direction, from concrete examples of temperature scales to an abstract definition of temperature based on these concrete examples? Actually, in proceeding as we did, we were relying on an intuitive feeling for the idea of temperature. The abstract definition could scarcely have any meaning at all for a reader with no experience with thermometers, or with the sensations of "hot" and "cold." What we would like to avoid in our discussion is exclusive reliance on sensation in deciding what we mean by "hotter" and "colder." It is well known how unreliable such judgments can be. For instance, if a person places one hand in very hot water and the other in ice water, the sensations in the two hands will be very different when both are put into a pan of lukewarm water. But even if we can avoid such confusion, we would like to be able to describe our model of nature in a way that is independent of human physiology.

However, we cannot simply sidestep the notions "hotter" and "colder," since they are central to the problem of establishing a temperature scale. In fact, the abstract definition of temperature given in the last section is really incomplete, precisely because it has no reference at all to these notions. What we must do is to give meaning to these ideas on the basis of physical phenomena, without invoking human sensations. In doing so, we will not only see how to establish temperature scales, but also to make our abstract definition of temperature more complete.

For guidance in how to proceed, let us examine how another scale, say that of mass, may be set up. The mass of a certain object is arbitrarily taken as the unit, and of course different choices of the unit lead to different scales. For any specific choice of a unit, the mass of any other object can be compared with that of the unit by placing them on opposite pans of an equal-arm balance. The descending arm of the balance then contains the "heavier," or more massive, object. The choice of associating "descending arm" (rather than "ascending arm"), with larger mass is not really arbitrary, nor does it depend on physiology, although it would agree with the choice made by hefting the two objects. It depends on the statement that two identical objects, taken together, have twice the mass of one of them alone.

The exactly analogous procedure is not possible with temperature, unfortunately, since there is no procedure for adding two temperatures that corresponds to the simple addition of two masses, and consequently, the choice of a particular object as a "unit" is impossible. Instead, we can choose a particular thermal equilibrium state of some convenient system as a "standard state"; we will call this system a "thermometer." We can then use the value of a specified property of the thermometer in the standard state (we will call this the "thermometric property"), to fix the one point of the temperature scale. The values of the thermometric property when the system is in other thermal equilibrium states, compared with its value for the standard state, then give numbers that represent the temperature in these other states. Now we can put the thermometer in thermal contact with any other system, say our pendulum bob. When the two systems have come into mutual thermal equilibrium, we can assign to the equilibrium state of the bob the temperature associated with the corresponding equilibrium state of the thermometer.

Now we can ask whether the new thermal equilibrium state is hotter or colder than the standard state. We can give an unequivocal answer by observing what happens when mechanical energy is dissipated in a system: if its temperature changes at all, it always gets hotter, never colder. This statement holds for all systems, regardless of the way in which mechanical energy is dissipated: rubbing friction in solids, stirring (viscous friction), in fluids, dissipation of electrical energy when currents flow; it is one of the widest generalizations that can be made from experience. It can also be put into analytical form, whereupon it becomes the basis for a powerful tool in discussing thermal phenomena, namely the second law of thermodynamics. Here, however, we use it only to connect the notions "hotter" and "colder" to temperature scales. To make this connection explicit, let us denote the temperature of the standard state by the number T_s, and that of some other state by T'. Then of the two transitions between these two states, T_s to T' or T' to T_s, it is possible to achieve only one by dissipating mechanical energy in the system. The one that is possible defines the beginning temperature as "colder," the final temperature as the "hotter" of the two. Suppose we find in this way that, say, T_s is colder than a particular temperature T_1, and hotter than another temperature T_2. Then we will always find that it is impossible to go from T_1 to T_2 by dissipating mechanical energy; that is, we will always find that T_1 is hotter than T_2.

Notice that we have said nothing yet about a relation between the magnitudes, or algebraic values, of the numbers we have assigned to represent the temperatures of various thermal equilibrium states. Although we have decided unequivocally that the state labeled T_1 is hotter than the state labeled T_2, this does not automatically imply that the number T_1 is larger than the number T_2. In fact, one of the earliest temperature scales assigned the number 0 to the equilibrium state of boiling water, and the number 100 to that of ice. The opposite convention is now universal; all temperature scales we use assign numbers in such a way that they increase algebraically as we go from colder thermal equilibrium states to hotter. Hence all states with numbers smaller than that assigned to the standard state are colder than any state with a number larger than that of the standard state. If the standard state is assigned to the number zero, this means that all states with negative temperatures are colder than any state with a positive temperature. But this statement also is merely a matter of convention, since the choice of standard state is arbitrary. One of the temperature scales we discuss below, the "gas-thermometer" scale, removes this arbitrariness to some extent, since it assigns the temperature "zero" in a way that appears to be natural and inevitable.[4]

Now that we have given the meaning of "hotter" and "colder" a physical basis, we can return to our description of how a temperature scale is set up. We recall that we start with a particular system to serve as a thermometer, this choice corresponding to the selection of a particular object as the unit of mass, in the preparation of a scale for mass. Just as the choice of a particular object for the unit of mass is essentially arbitrary, being limited only by questions of technical convenience, so also is the choice of a particular thermometer. Different applications impose different technical considerations, so there is a great variety of thermometers in actual use. One of the most common is the liquid-in-glass type. A small mass of liquid, commonly mercury or colored alcohol, is contained in a bulb attached to a capillary tube. This system is the thermometer, and the density of the liquid (or its volume, inversely

---

[4] This aspect of the gas-thermometer scale, which leads to interesting and useful significance for the label $T_g = 0$, and even for negative temperatures, is discussed in Appendix 3. However, reading this appendix should be deferred until the present section is completed, since the appendix assumes familiarity with the description of the gas-thermometer scale which is given below.

related to the density), is the thermometric property. Fahrenheit, in the eighteenth century, was one of the first to use a thermometer of this sort, and he chose a freezing mixture of salt and water as the standard state. The relative volumes of bulb and capillary tube are chosen so that when such a thermometer is in equilibrium with the standard state, most of the liquid will be in the bulb, but some will be in the capillary tube. The origin of the temperature scale is fixed by assigning the number zero as the temperature of all systems in thermal equilibrium with this particular state. In other thermal equilibrium states, since the density of the liquid will be larger or smaller, smaller or larger amounts of the liquid will be in the capillary tube. When the thermometer is in equilibrium with the body of an "average" adult, the density of the liquid is lower, its volume larger, and so the length of capillary occupied will be larger. Fahrenheit assigned the number 100 to the temperature of this new equilibrium state, and any other state can then be assigned a number corresponding to the length of capillary occupied by liquid when the thermometer is in equilibrium with the chosen state. When the thermometer is in equilibrium with a mixture of water and ice which itself is in thermal equilibrium, 32% of the length between the zero and 100 marks is filled, so the temperature is 32 "degrees Fahrenheit" (°F). In a similar manner, the temperature of water in equilibrium with steam at atmospheric pressure is found to be 212°F. Clearly, this thermometer cannot be used at temperatures so low that the liquid freezes, or so high that it vaporizes.

The temperature scale determined by such a thermometer will depend on the particular liquid used, and also on the material of the bulb and capillary, since densities of different materials vary in different ways over the range of temperature in which the thermometer is used. A system in which the thermometric property did not depend on the particular substance making up the system would clearly provide the possibility of constructing a superior thermometer. But whether or not such a system exists is not a matter of definition or logic. It is purely an empirical question, to be decided by experiment. However, the results of experiment, namely that a thermometer does exist which is largely independent of the particular substance used, has interesting consequences for our model of thermal phenomena. We shall explore some of these consequences in later sections, but first we will describe one such thermometer, the constant-volume gas thermometer.

If a mass of gas, which we denote m, is enclosed in a fixed volume V_c, its pressure, P, is found to be different in different thermal equilibrium states. The pressure can then be used as a thermometric property, and the numerical value of temperature, which we will denote T_c, can be assigned according to a rule analogous to that used with the liquid-in-glass thermometer. There the numerical value of temperature was associated with the length of liquid in the capillary, as compared with the length when the thermometer was in the standard state. Here we use a somewhat different standard state, namely that in which water, water vapor, and ice are in mutual thermal equilibrium. The pressure of the gas is found to have a reproducible value, which we denote P_0, when its container is in thermal equilibrium with the standard state, and the temperature of the standard state is defined as $T_c = 0$. The temperature, T_c, in any other thermal equilibrium state is then defined in terms of the ratio $P(T)/P_0$ by the equation

$$P(T_c)/P_0 = 1 + aT_c = 1 + T_c/T_0, \quad (4.1)$$

where in the last equality we have simply written $1/T_0$ for the constant, a. This constant determines the size of the unit temperature difference, or "degree." In the common Celsius scale (formerly called the centigrade scale), it is fixed by setting the temperature of the "steam point" (water in equilibrium with its vapor at standard atmospheric pressure), as equal to 100 degrees Celsius (°C). Therefore, if the measured pressure at the steam point is P(100),

$$P(100)/P_0 = 1 + 100\,a,$$

$$a = [P(100)/P_0 - 1]/100 = 1/T_0 \qquad (4.2)$$

Therefore, Eq. (4.1) can be written

$$T_c/T_0 = P(T_c)/P_0 - 1. \qquad (4.3)$$

This equation is not quite symmetrical, because of the −1 on the right-hand side. A simple algebraic trick can convert the equation into a more symmetrical form, in terms of a new temperature scale which differs from that in Eq. (4.3) only by changing the origin, that is, the state which is denoted "zero degrees," but leaving the size of the degree unchanged. If the term −1 is transposed,

$$\frac{P(T_c)}{P_0} = \frac{T_c}{T_0} + 1 = \frac{(T_c + T_0)}{T_0} = \frac{T_g}{T_0};$$

$$T_g = T_c + T_0. \qquad (4.4)$$

We will call the temperature scale denoted by T_g, and defined by the last equality in Eq. (4.4), the "gas thermometer scale." This is not a universal scale, since it does depend to some extent on the particular kind of gas in the volume V_c. Suppose we place several such thermometers, each containing a different kind of gas, in mutual thermal equilibrium, after having prepared each with the <u>same</u> pressure P_0 in the standard state. We will find, in general, that each thermometer will have a slightly different pressure when they are all in a thermal equilibrium state which is different from the standard state. That is, if we write Eq. (4.4) as

$$T_g = [P(T_g)/P_0]T_0, \qquad (4.5)$$

we find that $P(T_g)$ is slightly different for each gas. Since according to our definition of temperature, T_g is the same for all thermometers, this means that T_0 must have a slightly different value for each gas. Now we can imagine repeating the same procedure, each time starting with a lower value of P_0 in the standard state (either by using less gas in the same volume V_c, or larger V_c with the same m, or both), and then returning to the state identified by T_g. As P_0 is decreased, the value of $P(T_g)$ is naturally found to decrease as well. But we will also find the important result that the differences in $P(T_g)$ for the different gases become smaller and smaller. That is, as P_0 decreases, the ratio $P(T_g)/P_0$ approaches a constant value which is independent of the gas in the thermometer. If the same procedure is carried out for other thermal equilibrium states, that is, with different values of T_g, the same effect is observed. The ratio $P(T_g)/P_0$ again approaches a limiting value (as P_0 approaches zero), which is again independent of the gas used. This limiting value depends on temperature T_g, but in a very simple fashion: it is proportional to T_g. Hence under these conditions (limit of low pressure), T_0 is a constant which is the same for all gases, and at all temperatures. We have been quoting the results of extensive experimental investigation carried out at the end of the seventeenth century by Boyle, Mariotte, Charles, Gay-Lussac and others, and considerably refined since then. The best current value of T_0, determined by the procedure described above, is

$$T_0 = 273.16 \text{ degrees}.$$

Thus in the limit of low pressure, Eq. (4.5) becomes the definition of a temperature scale which is independent of the substance used, although it still refers to a particular system, an enclosure filled with gas. We will refer to this scale as the "ideal gas" scale.

We will not digress to discuss in general the question of how temperatures are measured in practice when high precision and accuracy are required. One aspect of this problem, the intercomparison of temperatures scales associated with different kinds of thermometers, is discussed briefly in Appendix 4. In later sections, we will return to a discussion of the ideal gas temperature scale which will give us further insight into the connection between mechanical and thermal phenomena. At this point, we can already see one such relation. The ideal gas scale distinguishes among different thermal states, using only the purely mechanical procedure of measuring pressures, and without requiring any information about any properties of specific substances.

5 DO THERMAL EQUILIBRIUM STATES EXIST?

From the point of view of our main line of discussion, our description of temperature scales in the last section was a long digression, but it was intended to serve a definite purpose. In the previous section, we stated the definition of thermal equilibrium in a purposely abstract form in order to make it clear that this definition is generally valid, without reference to specific properties of particular substances. But a price must be paid for such an abstract treatment: It makes the concept described remote from experience, and it becomes difficult to connect the abstract idea with concrete reality. The discussion in the last section was intended to close this gap by showing that thermal equilibrium, and the concept of temperature which it defines, have a perfectly definite meaning. Furthermore, we saw that this meaning could be described in terms of simple operations carried out on real objects. Having completed that discussion, we can return now to a further examination of thermal equilibrium with a clearer idea of the physical processes involved.

Notice that thermal equilibrium differs from mechanical equilibrium in a rather important respect. A system in a thermal equilibrium state remains in that state as long as it is isolated, or, if it is in contact with its surroundings, as long as its surroundings do not change. In the mechanical model, there is nothing strictly comparable to this property. Mechanical equilibrium requires only that the velocity $\vec{v}$ = const, so that a particle in equilibrium passes through an infinite number of positions in succession, and hence through an infinity of different mechanical states. One special mechanical equilibrium state, that for which $\vec{v} = 0$, and therefore $\vec{r}$ = const, (particle at rest), might seem analogous to the thermal equilibrium state, since a particle at rest does not pass through other mechanical equilibrium states in the course of time. However, the particular mechanical equilibrium state corresponding to a particle at rest does not really have any special significance. As a matter of fact, this lack of significance for the "rest" state is no accident, but a fundamental feature of mechanics. It arises from the fact that all frames of reference, which differ only by virtue of a constant relative velocity between their origins, are precisely equivalent in the mechanical model. Therefore, if a particle is at rest in one reference frame, there exists an infinity of completely equivalent reference frames, each moving with respect to the first with a constant velocity, with respect to which the particle is still in equilibrium, but no longer at rest.

We can make this relation between equivalent reference frames more explicit by referring again to Fig. 2.1. The two reference frames in the figure are distinguished from each other by the two origins, O_1 and O_2. In our earlier discussion, the vector joining the two origins was a constant, $\vec{R}_{21}$. In our present discussion, this vector is a function of time, since the two reference frames are moving with respect to each other. Since their relative velocity is a constant vector, which we can denote $\vec{V}$, we can write

$$\vec{R}_{21}(t) = R_{21}(0) + \vec{V}t, \qquad (5.1)$$

where $\vec{R}_{21}(0)$ is also a constant vector. Now suppose O_1 is the origin of the reference frame in which the particle is at rest. Then $\vec{v}_1 = 0$, and, taking the time derivative of both sides of Eq. (2.1), with Eq. (5.1) for $\vec{R}_{21}$, we have

$$\vec{v}_2 = \vec{V}.$$

Hence, while the particle is not at rest in the reference frame whose origin is O_2, since $\vec{v}_2$ is not zero, it is still in equilibrium in that frame, since $\vec{V}$ is a constant vector.

It is worth recognizing that these two different kinds of specification for equilibrium (a stable state with fixed parameters in the thermal case, a succession of different states, related by Eqs. (2.6) and (2.7) in the mechanical case), both come from the same kind of abstraction from natural phenomena. The principle of abstraction which is common to both is the recognition, in each case, of the fundamental characteristic of a disturbance. For a particle, a disturbance does not involve merely a change in position, but rather a change in velocity; this is the basic statement of N-II. For a thermal system, on the other hand, any change of its parameters whatever is caused by a disturbance, so thermal equilibrium is represented by a stable state.

This stability of thermal equilibrium states, in the absence of disturbance, leads immediately to the specification of a unique direction of time without any reference to mechanical effects. We recall our description of mutual thermal equilibrium: two isolated thermal systems, each in thermal equilibrium, but at different temperatures, are placed in thermal contact with each other, while still isolated from any other surroundings. In time (at a rate which depends on their degree of thermal contact), they each reach a new equilibrium state, and these two final states have the same temperature, which is between the two initial temperatures. Now if we watched this process in reverse, we would see an isolated system, with two parts in mutual thermal equilibrium, as the initial state. This is a different initial condition than we considered previously, and we recall that when we discussed time reversal in mechanical systems, we saw that changing the direction of time also changed the initial mechanical state, since the velocity was reversed. But in the case of the motion of particles, the behavior after time reversal was still a possible motion, even though it was not exactly the same as that which preceded time reversal. In the present case, however, when we consider the behavior of our thermal systems, we see that this is no longer the case. The process as observed with time reversed would show us the two systems, originally in mutual thermal equilibrium, changing their temperatures in opposite directions, one getting cooler and one warmer, while thermal contact was still maintained, although they had no contact, mechanical or thermal, with any surroundings but each other. This process, which would be observed after time reversal, is never observed in nature. Two systems, once in mutual thermal equilibrium, are always observed to maintain that condition as long as they are isolated from any other surroundings. Hence the direction of time, in which the "disequilibrium" process would be described, is the wrong direction, and is so recognized with no reference to any mechanical effects such as dissipation of mechanical energy, which we used earlier to specify the "correct" direction of time flow.

5.1 FLUCTUATIONS IN THERMAL PROPERTIES.

If this were an entirely accurate description of what can and cannot be observed when we examine the approach to thermal equilibrium, nature would be very different in certain fundamental respects. Just what these differences might be, we will discuss later. At the moment, we can remark that the description given above of the approach to thermal equilibrium is fairly accurate, and can serve as a first approximation to the true situation. However, the details which it ignores are just as important to a complete understanding of the process as those which were included. Let us go beyond this glib description and examine some of these missing details.

When we look at the approach to thermal equilibrium in further detail, we will be led to the conclusion that, strictly speaking, thermal equilibrium can never be observed. In section 4.1, thermal equilibrium was defined by the condition ($\dot{P}_i = 0$) for all measured values of all properties of a system. As we shall see, this condition can never be satisfied precisely, but for two different kinds of reasons. We will now examine these two reasons in turn.

In general, if we examine some thermal system at a particular instant, at which we arbitrarily set $t = 0$, it will not be in a thermal equilibrium state. If at that instant we thermally isolate the system from its surroundings, or, alternatively, if we fix the conditions of the surroundings so they remain the same as they were at the initial instant, $t = 0$, the system will approach some final thermal equilibrium state. We expect this approach to be smooth, or continuous, in the mathematical sense. We also expect that the rates of change ($\dot{P}_i$) will decrease in magnitude as the (P_i) approach their final values, which we will denote ($P_i(\infty)$). Both of these conditions can be satisfied in the following way. We define a quantity $\Delta P_i(t)$ as the difference between the final value of P_i and the value of P_i at time t:

$$\Delta P_i(t) = P_i(\infty) - P_i(t). \qquad (5.2)$$

If this quantity is positive, then at time t, P_i is smaller than its final value, so at time t, P_i must be increasing; hence $\dot{P}_i(t)$ is also positive. Conversely, if $\Delta P_i(t)$ is negative, then $\dot{P}_i$ is also negative. The simplest way of ensuring that $\dot{P}_i(t)$ decreases in magnitude as $P_i(t)$ approaches its equilibrium value is to set this time derivative proportional to the difference remaining at time t, that is, to $\Delta P_i(t)$. If we choose a proportionality constant which has the dimensions of time, and denote it τ, we can write

$$\dot{P}_i(t) = \Delta P_i(t)/\tau. \qquad (5.3)$$

We should point out that we have not "proved" that the (P_i) and their time derivatives must satisfy this equation; we have simply shown that if they do, they will satisfy the two expectations stated at the beginning of this paragraph. Therefore, Eq. (5.3) should provide at least an approximate description of the actual approach to thermal equilibrium.

Equation (5.3) has a form which is found to apply to a great variety of physical problems, since the conditions suggested at the beginning of the last paragraph, which led us to postulate this equation (not "derive" it), are reasonable expectations in many different physical situations. Radioactive decay is one of these situations; it also governs the approach to terminal speed of a particle falling in a resistive medium, in certain circumstances. In this last context, it was discussed in Appendix 1, where it was shown how a solution can be derived. Here we will simply state the solution, and show that the statement is correct by calculating the time derivative, and comparing the result with Eq. (5.3). We assert

$$\Delta P_i(t) = \Delta P_i(0)\ e^{-t/\tau}, \qquad (5.4)$$

where e is the base of natural logarithms. Taking the time derivative of both sides,

$$-\dot{P}_i(t) = -\frac{1}{\tau}\ \Delta P_i(0)\ e^{-t/\tau} = \frac{-\Delta P_i(t)}{\tau} \qquad (5.5)$$

where we have used Eqs. (5.3) and (5.2) in making substitutions. Hence, we see that the time derivative of Eq. (5.4) does agree with Eq. (5.3), so we have indeed found a solution to the latter equation. We will call this solution the equation of "exponential relaxation" because of the exponential factor, and refer to τ as the "relaxation time."

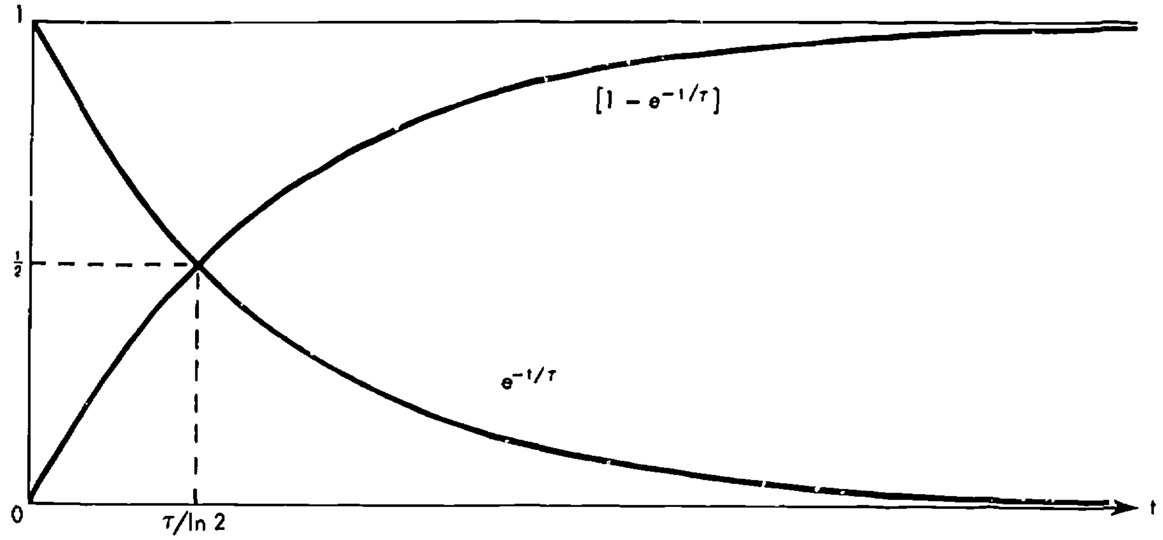

Fig. 5.1 The quantities $e^{-t/\tau}$ and $1 - e^{-t/\tau}$ as functions of t.

The measured quantities in Eq. (5.4) are not the values $\Delta P_i(t)$, since these involve $P_i(\infty)$, which are unknown quantities before equilibrium is reached. The quantities measured directly are values of $P_i(t)$ at successive times, including of course $P_i(0)$. A convenient way of calculating the initially unknown quantities τ and $P(\infty)$ can be demonstrated by rearranging Eq. (5.4) in the following way:

$$P_i(t) = P_i(0)\, e^{-t/\tau} + P_i(\infty)(1 - e^{t/\tau}). \tag{5.6}$$

Both the exponential, $e^{-t/\tau}$, and the quantity in the bracket in Eq. (5.6), are plotted in Fig. 5.1. It is easy to show that the two curves cross at the common value ½, and that this value is shared at the time $t = \tau/\ln 2$, where ln is the natural logarithm. Thus for $t << \tau/\ln 2$, the first term in Eq. (5.6) is much larger than the second, and τ can be found directly from $P_i(t)$, by plotting $\ln P_i(t)$ as a function of t. If the second term in Eq. (5.6) can be neglected, such a plot is a straight line with a negative slope, the magnitude of which is the reciprocal of τ. When τ has been determined in this way, the quantity $1 - e^{-t/\tau}$ can be calculated as a function of time. For $t >> \tau/\ln 2$, the first term in Eq. (5.6) is negligible compared to the second, and $P_i(\infty)$ can be determined. Equations (5.4) and (5.6) are plotted in Fig. 5.2, for the last two cases, $P(0) > P(\infty)$ and $P(0) < P(\infty)$.

We can now state the first, and less significant reason, why $\dot{p}_i(t) = 0$ is never observed. It is simply that, according to the first equality in Eq. (5.5), its magnitude is always different from zero, for all finite times. Nevertheless, the limiting values $P_i(\infty)$ ca[illegible] determined, as we just saw, so [illegible] ntities $\Delta P_i(t)$ can be calculat[illegible] l times. Furthermore, the $\Delta P_i($[illegible] ach zero as a limit, so they [illegible] aller than any preassigned q[illegible] if we wait long enough. [illegible] re is a finite limit to the p[illegible] of any apparatus that might b[illegible] measure $P_i(t)$, this means th[illegible] ely the $\Delta P_i(t)$ will be so s[illegible] heir difference from zero wi[illegible] n by the "random error" [illegible] suring apparatus. At this sta[illegible] ld assert that thermal equi[illegible] as been reached, since we [illegible] onger measure a value

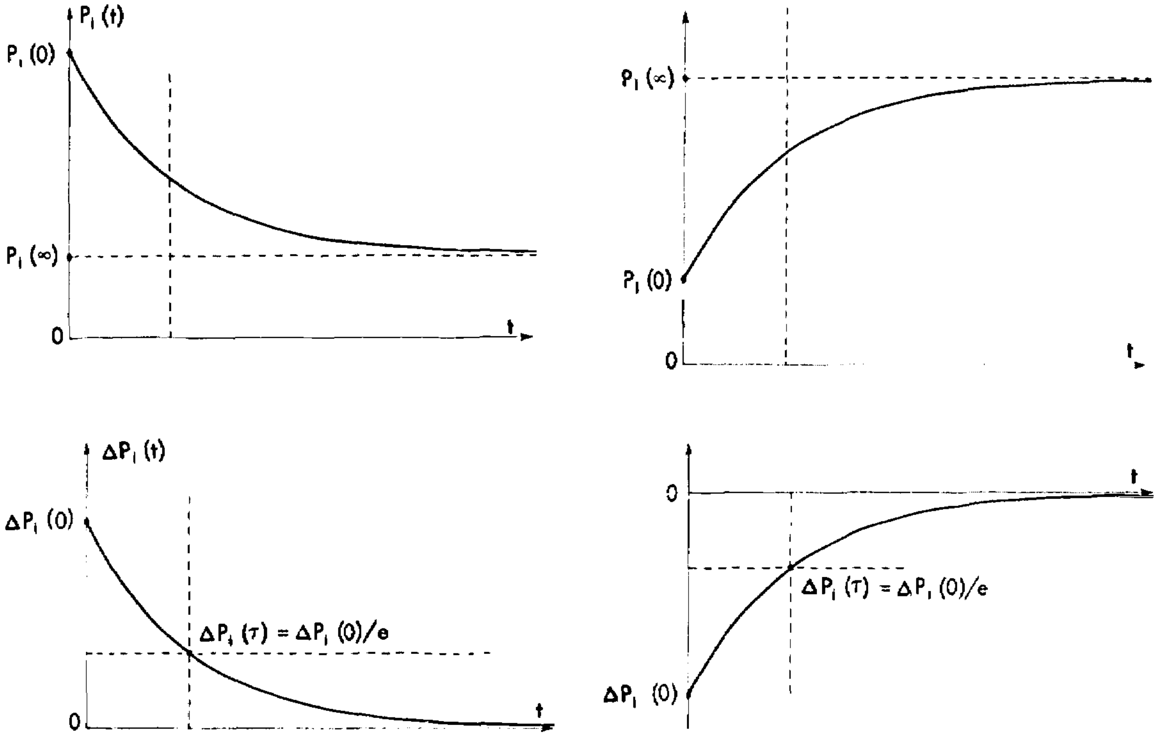

Fig. 5.2 $P_i(t)$ according to Eq. (5.5)(upper curves), and $\Delta P_i(t)$ according to Eq. (5.4) (lower curves), for two cases: $P_i(0) < P_i(\infty)$ (right-hand curves), and $P_i(0) > P_i(\infty)$(left-hand curves).

of $\Delta P_i(t)$ (or of $\dot{P}_i(t)$), different from zero.

It therefore appears that specification of thermal equilibrium involves the characteristics of the apparatus that has been used in measuring the $P_i(t)$. A state described as a thermal equilibrium state on the basis of measurements with one apparatus might be found to be still changing, when measurements are made with instruments of greater precision. But as instruments of higher and higher precision are used, a new feature appears, which can be described in the following way. According to Eq. (5.4), $\Delta P_i(t)$ always has the same sign during the approach to equilibrium, for any single property (that is, for any specified i). However, for $t >> \tau/\ln 2$, the magnitude of $\Delta P_i(t)$ becomes so small that the limit of precision of the apparatus is reached, and the values of $\Delta P_i(t)$ begin to have both signs, positive and negative; their average value, which we will write $<\Delta P_i(t)>$, becomes zero. When this sort of behavior is observed, the corresponding average value of $P_i(t)$ is an estimate of $P_i(\infty)$, according to Eq. (5.2). In order to compare the precision of different kinds of apparatus, we should compare the magnitude of these limiting values of $\Delta P_i(t)$: A convenient way of doing this is to square each $\Delta P_i(t)$, and then sum the squares. Both negative and positive values give contributions to the sum, so the average of the squares is a measure of the magnitudes of $\Delta P_i(t)$. This average is written $<(\Delta P_i(t))^2>$ and is usually called the "mean-square deviation." Clearly, the mean-square deviation has dimensions which are the square of the dimensions of the corresponding P_i. If we take the square root of the mean square deviation, we obtain a quantity which is called the "root-mean-square deviation," which we will denote $(\Delta P_i)_{rms}$ or $\sigma(\Delta P_i)$. The root-mean-square (or "rms"), deviation is a measure of the magnitudes $\Delta P_i(t)$ which, as we see, has the same dimensions as P_i itself. For later reference, we write the formal definition:

$$\sigma(\Delta P_i) = (\Delta P_i)_{rms} = \sqrt{<(\Delta P_i)^2>}$$

$$= \left[\frac{1}{N}\sum_{j=1}^{N}(\Delta P_i)_j^2\right]^{\frac{1}{2}} \qquad (5.7)$$

where, in the last equation, we have written out the average explicitly: N different measurements of $P_i(t)$ have been included, each giving a value of $(\Delta P_i)^2$; these values have been labeled with the index j, which runs from 1 to N. From our discussion above, we see that $\sigma(\Delta P_i)$ may serve as a measure of the precision of the instrument used in the measuring $P_i(t)$; the greater the precision, the smaller we would expect $\sigma(\Delta P_i)$ to be.

Although the result of this long discussion, leading to the definition of the rms deviation $\sigma(\Delta P_i)$, allows us to compare the precision of different kinds of measuring apparatus, this is not the primary reason for defining $\sigma(\Delta P_i)$. Our main motivation was that certain very important characteristics of a series of measurements of $P_i(t)$, made for $t >> \tau/\ln 2$, can be stated succinctly in terms of that definition. These characteristics are:

(a) If measurements of greater and greater precision are made, it is found that $\sigma(\Delta P_i)$ does not decrease indefinitely; a lower limit to $\sigma(\Delta P_i)$ is reached, and the rms deviation does not decrease below this lower limit, regardless of the precision of the measuring apparatus.

(b) This lower limit to $\sigma(\Delta P_i)$, while it is independent of the measuring apparatus, does depend on the system whose values of $P_i(t)$ are being measured, and in two ways:

(i) The (noninstrumental) lower limit to $\sigma(\Delta P_i)$ in general increases as the size of the system decreases;

(ii) It also increases with higher temperature.

We shall refer to deviations $\Delta P_i(t)$ which are observed when this lower limit to $\sigma(\Delta P_i)$ has been reached, so that properties of the measuring apparatus are irrelevant, as fluctuations of the system. What we have stated above, then, are properties of fluctuations observed, at least in principle, in all thermal systems.

The qualification, "in principle," must be made because of the statement labeled (b)(i) above. For sufficiently large systems, the fluctuations become so small that the noninstrumental limit to $\sigma(\Delta P_i)$ is never reached. Hence the deviations $\Delta P_i(t)$ that are observed, are related to the measuring apparatus, and are not a property of the system, and therefore are not fluctuations as defined above. This circumstance makes it possible to avoid fluctuations in practice, simply by increasing the size of the system, or of the surroundings with which the system is in thermal contact. For instance, if we wished to avoid temperature fluctuations of a pendulum bob in an arrangement such as that illustrated in Fig. 4.1, we would place a very massive object, which itself was in thermal equilibrium, on the upper plate. In actual practice, the bob would be placed directly in a large mass of liquid, since the best thermal contact would be achieved this way. Boiling an egg is a familiar example of this procedure. Because of the wide range of temperatures over which different liquids condense, "temperature baths" are not limited to the ordinary temperatures we commonly experience. The lowest temperature that can be reached in this way is about 0.3°K above the absolute zero, the liquid for this bath being the rare isotope of helium, He^3, boiling under reduced pressure.

We can now answer the question posed earlier, "Do thermal equilibrium states exist?" According to the definition in section 4.1, the answer must be "No," because of the presence of fluctuations. But it is much more reasonable to base definitions on physical grounds, so a preferable answer is "Yes," with the understanding that the conditions $\dot{P}_i = 0$, or $\Delta P_i = 0$, are not satisfied exactly, because of the existence of fluctuations. The possibility of observing fluctuations is thus incorporated into an expanded definition of thermal equilibrium. The conclusions that we can draw from this extended definition, about the internal constitution of matter, are explored in the next section.

Before proceeding to this discussion, we should consider two points which have so far been neglected, whose clarification will help us later on. In the first place, even with our definition of thermal equilibrium extended so as to include fluctuations, there is still a sense in which thermal equilibrium states do not exist. Suppose we consider a system in thermal isolation, which has come to thermal equilibrium in the extended sense, so that the P_i are exhibiting fluctuations about their average values, $P_i(\infty)$. Now perfect thermal isolation is an idealization from experience, and can never be achieved in practice. Therefore, if the temperature $T(\infty)$ of the isolated system is different from that of the surroundings, say T', there will be a slow variation of $T(\infty)$ toward T'. This variation may be approximated by Eqs. (5.4) or (5.6), but with a relaxation time τ' which is long compared to the relaxation time τ, which gave the rate at which the isolated system approached $T(\infty)$. Hence in order to observe the equilibrium of the isolated system, we must wait until $t \gg \tau/\ln 2$, but we must not wait so long that $t \approx \tau'/\ln 2$.

The same sort of "superposition" of thermal equilibrium states, with relaxation times of very different size, is found in most physical systems. For instance, objects that are made of metal that has been deformed by forging, stamping, etc., during manufacture, will exhibit very slow changes in size and shape in the course of time. On a different scale of size, galaxies evolve at rates which are very slow on a human time scale. A small portion of a galaxy may therefore appear to be in equilibrium at any instant, despite its slow change to a new state.

In general, therefore, thermal equilibrium states as observed in physical systems are not likely to be truly stable; such stability would imply that fluctuations about values $P_i(\infty)$ persist indefinitely, in the absence of externally imposed disturbances. Rather, they are most likely to be "metastable," and the apparent equilibrium can only be observed in an interval of time which satisfies the condition

$$\tau/\ln 2 \ll t \ll \tau'/\ln 2, \qquad (5.8)$$

where τ is the relaxation time for the metastable state that is being observed, and τ' is the relaxation time for the transition to some other state, presumably also metastable. In speaking about thermal equilibrium, therefore, we will limit ourselves to times of observation which satisfy the condition expressed by Eq. (5.8). Thus, even though we cannot expect to observe a system in a "true" thermal equilibrium state, we can expect that the metastable states we do observe will be a good approximation to the idealization of "true" equilibrium.[5]

The other comment about fluctuations that we will find useful in the

---

[5] There has been much speculation, largely metaphysical, concerning the possibility that the universe as a whole is approaching some sort of final equilibrium state. But the relaxation rate for this process, if it is in fact taking place, must be so enormous on our ordinary time scale that it is hardly reasonable to expect to get much information about it from measurements which are accessible to us.

next section is that fluctuations are more easily observed in the values of some of the properties than in others. We spoke above about fluctuations $\Delta P_i(t)$, without discriminating at all among the various values of i, that is, among the various properties that can be measured. But the possibility of observing fluctuations demands that we possess measuring apparatus of sufficiently high precision. If we denote the root-mean-square deviation of instrument readings, which is related entirely to the precision of the instrument, by $\sigma_I(\Delta P_i)$, then in order to observe fluctuations in the values of the property P_i, it must be true that

$$\sigma_I(\Delta P_i) \ll \sigma(\Delta P_i), \qquad (5.9)$$

where $\sigma(\Delta P_i)$ is the root-mean-square deviation associated with the fluctuations. This condition puts a rather severe limitation on any effort to translate our remarks about fluctuations in general, in any property of a system, into observations on values of a particular property. It turns out that only those properties involving mechanical or electrical measurements are likely to display fluctuations which can be observed with instruments satisfying Eq. (5.4). One example of such a measurement is the determination of position, which, as we saw in section 3, is associated with mechanical equilibrium.

If we reexamine our earlier discussion of the laws of mechanics, we see that there is a closer connection between mechanical and thermal equilibrium than we have pointed out so far. In fact, it is not hard to see that a thermal equilibrium state implies mechanical equilibrium as well. This is obvious for systems in thermal isolation, and can also be seen to be true for systems in thermal contact with their surroundings. For if a system in thermal contact with its surroundings is not in mechanical equilibrium, a finite net force must be acting on it. But it is a fact of experience that the action of such a force is accompanied by the dissipation of mechanical energy, and we have seen that, in general, such dissipation raises the temperature of the system. Indeed, we can go further and observe that if the system is in mechanical equilibrium, but moving with constant velocity with respect to its surroundings, then again dissipation will occur, since the relative motion will give rise to frictional forces. Hence, in general, a system in thermal equilibrium is not only in mechanical equilibrium, but also at rest relative to the surroundings with which it is in thermal contact. Thus with respect to the coordinate system in which the surroundings are at rest, the kinetic energy of the system is $K = 0$. We recall from the beginning of this section, however, that this particular mechanical equilibrium state, in which the system is at rest, has no special mechanical significance. The laws of mechanics are exactly the same in the reference frame associated with this "rest" state, as in any other reference frame moving with respect to it, at constant velocity.

We therefore conclude that if we examine the motion of an object which is in thermal equilibrium, using the mechanical model to represent the object by a mass-point, then the mass-point is in mechanical equilibrium, and in particular, at rest with respect to the thermal surroundings of the object. Hence the extended definition of thermal equilibrium, which includes fluctuations, implies that fluctuations should be associated with the mechanical equilibrium state as well. This is a requirement that is not associated with the mechanical model itself, but is imposed on it by the relation we have been discussing, between thermal and mechanical equilibrium.

Although we have come to this conclusion, that fluctuations are to be expected in mechanical equilibrium states from their connection with thermal equilibrium states, the actual historical development of these ideas took place in almost exactly the opposite direction. As remarked, such fluctuations should be relatively easy to observe since they require measurement of position, for which instruments whose precision satisfies Eq. (5.9) are available. As a result, such mechanical position fluctuations were observed long before their connection with thermal equilibrium was realized, and the reason for their occurrence remained a puzzle for a long time. We will describe some of these observations in the next section, which, for the reasons just discussed, will give most attention to mechanical fluctuations.

5.2 MECHANICAL FLUCTUATIONS

Since now we wish to devote our attention to fluctuations of mechanical properties, we must examine the opposite extreme in size from that discussed at the end of the last section, namely the extreme of very small systems. But we do not go to the ultimate extreme of molecules, atoms, electrons, and nuclei. Even though we will be dealing with small systems, they will always be macroscopic. That is, they are at least large enough to be visible with an ordinary microscope, using visible light, and they can be manipulated, by direct or indirect means. The significance of this distinction, between macroscopic size (though small), and atomic and subatomic sizes, is by no means merely verbal. On the macroscopic level, the ordinary laws of dynamics, as expressed in Newton's laws, describe the mechanical behavior of matter. All the systems we will discuss display mechanical behavior which is adequately described by N-I and N-II. At the atomic level, on the other hand, this is no longer true; the behavior of systems at this size level must be described by the laws of quantum mechanics. However, the behavior of fluctuations which we shall describe is a consequence of purely "classical" or nonquantum-mechanical dynamics.

These remarks should not be taken to imply that purely quantum-mechanical fluctuations do not exist. In fact, the laws of quantum mechanics have the consequence that fluctuations occur in all atomic systems. However, the effects of these quantum-mechanical fluctuations on the behavior of macroscopic systems are generally so small that they cannot be observed. That is, they produce values of $\sigma(\Delta F_1)$ which are smaller than the lower limit described in statements (a) and (b) in section 5.1. There are special circumstances, which we shall not consider at all, in which the effects of quantum-mechanical fluctuations can be observed in measurements on macroscopic systems. Ordinarily, however, purely quantum-mechanical fluctuations are even smaller than those due to the limited precision of measuring apparatus.

We inquire, then, about the origin of "classical," or macroscopic fluctuations. Are they associated with the basic laws of classical mechanics, that is, N-I and N-II, in a manner analogous to the connection between quantum-mechanical fluctuations and the basic laws of quantum mechanics? These are deep questions, and clearly a complete answer would require studying the laws of quantum mechanics themselves, so we will not be able to answer them exhaustively. Instead of pursuing them further here, we turn to an examination of the kinds of mechanical fluctuations that have been observed.

What was probably the first example of a mechanical fluctuation, in the sense we have been using this term, was reported by Brown in 1827. Brown was a botanist, not a physicist, and the system he was observing was a liquid suspension of pollen grains. Each particle in such a suspension is falling in a retarding medium, so the analysis we gave in section 5 and Appendix 1 should apply. As we saw, the downward velocity of the grains should approach a constant magnitude, the terminal speed v_T. We would thus expect to see the grains settle to the bottom of the liquid, and it is common experience that coarse grains settle faster than fine grains. This observation is contained in our analysis, for the approach to terminal speed should be roughly exponential. It can therefore be characterized by a relaxation time τ, as we discussed in the last section. In Appendix 1, it is shown that

$$\tau = v_T/g, \qquad (5.10)$$

(compare this with the exponent in the expression for v(t) which is given in the caption to Fig. 3.2.). Hence high terminal speed, which corresponds to fast settling, is associated with large τ, and small τ corresponds to low v_T. At the end of Appendix 1, it is shown that the relaxation time τ decreases as the square of the radius for spherical particles. Therefore v_T, and also the rate of settling, decreases rapidly as the size of particles in a suspension decreases.[6]

Before describing Brown's observations in detail, it will help to establish a more quantitative criterion for what we mean by "fast settling" and "slow settling." If the settling time is of the order of seconds, we can reasonably call it "fast," while if it is of the order of hours (say, 10^4 sec or longer), we can call it slow. If the suspension is in a container of height H, the settling time t_s is of the order H/v_T. Now the radius of the particles, R, will necessarily be much smaller than H, so for slow settling, (assuming $R/H \approx 10^{-4}$)

$$t_s \approx H/v_T \approx 10^4 \text{ sec}; \quad R/v_T \approx 1 \text{ sec}. \qquad (5.11)$$

That is, for slow settling, a particle will move about its own radius, or less, in a second. Such settling will thus be almost imperceptible to the naked eye, unless one watches for hours. In order to observe the settling of his pollen grains, which had $R \approx 1$ micron (10^{-4} cm), Brown used a microscope.

In view of our discussion, Brown's observations were startling. As observed with the microscope, the grains were in incessant random motion. There was no apparent preference for downward motion, either in the motion of all the particles in the field of view at a particular instant, or in the motion of any one particle followed in the course of time. The motion of the grains was so rapid that it was not possible to determine their speeds directly. Later observations, however, used an indirect procedure to arrive at an estimate of the average speed. If the motion is followed for a period of, say, 10 seconds, it will appear roughly as sketched in Fig. 5.3. The average velocity thus has a magnitude, $|v_{ave}|$, which is the ratio of the magnitude of the displacement $|AB|$ (which we will denote d), to the time interval. This turns out to be of the order of the last expression in Eq. (5.11); that is, in 10 seconds, d is about 10R, so $|v_{ave}| \approx R$ cm/sec. But clearly the path is much larger than the displacement, so the average speed, $\langle v \rangle$, is much larger than R cm/sec.

This result is clearly paradoxical; the average speed is of the order of that describing fast settling, as discussed in connection with Eq. (5.11), but the actual settling is very slow. The reason, of course, is

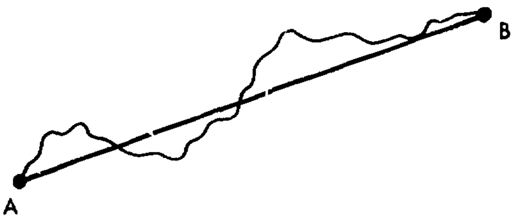

Fig. 5.3 Displacement (line AB) and path (much longer than displacement) in Brownian motion.

---

[6] Since v_T is proportional to g, according to Eq. (5.10), while according to Appendix 1, τ is independent of g, the rate of settling can be increased by artificially increasing g, as in a centrifuge.

that the motion of the grains is randomly directed, rather than being limited to the downward direction. It was natural for Brown, at first, to connect this random motion with the fact that the pollen grains were living matter. But it soon became apparent that there was no such connection; the crucial requirement for observing this sort of mechanical behavior, which is called "Brownian motion," is simply size. Sufficiently small particles, be they dust, smoke, or whatever, exhibit Brownian motion when suspended in a fluid (liquid or gas).

As we saw at the beginning of this section, fluctuations in the mechanical properties are to be expected for a system in thermal equilibrium, and presumably Brownian motion is an example of this sort of behavior. In the last section, we discussed fluctuations in a system in terms of all its properties, which we labeled P_i. If now we wish to be more explicit, we could choose, as specific examples of mechanical properties, the instantaneous velocity components $v_x(t)$, $v_y(t)$, $v_z(t)$. Of course, the average value of each of these components is zero, since otherwise the entire group of particles would have a mass motion in some direction. The mean squares of these components, $\langle v_x^2(t)\rangle$, $\langle v_y^2(t)\rangle$, $\langle v_z^2(t)\rangle$, do not vanish. Unfortunately, we have just seen that these components cannot be measured directly, so there is no direct way of determining their mean squares. However, there is a simple relation between these mean squares, and the mean squares of the components of the average velocity, $\vec{v}_{ave}$, which was defined in connection with Fig. 5.3. In order to calculate these mean squares, namely $\langle(v_{ave})_x^2(t)\rangle$, $\langle(v_{ave})_y^2(t)\rangle$, $\langle(v_{ave})_z^2(t)\rangle$ we must record displacements d_1, d_2, d_3, . . . d_i, . . . We then take components of these displacements, $(d_i)_x$, $(d_i)_y$, $(d_i)_z$. To get the corresponding average-velocity components $(v_{ave})_x$, $(v_{ave})_y$, $(v_{ave})_z$ (as distinguished from the unmeasurable instantaneous velocity components, v_x, v_y, v_z), we must divide each of the displacement components by the time interval during which the particle made the displacement. Let us choose the <u>same</u> time interval, say Δt, for every displacement.[7] Then clearly

$$\langle(v_{ave})_x^2\rangle = \langle d_x^2\rangle/(\Delta t)^2, \qquad (5.12)$$

and similarly for the y and z components. Now, when the same time interval is used for each displacement, it can be shown that

$$\langle v_x^2(t)\rangle = N\langle(v_{ave})_x^2\rangle \qquad (5.13)$$

(and similarly for the other two components), where N is a constant that depends only on the choice of the time interval Δt. In fact, N is just the ratio of the time interval Δt to the average time required, call it $\langle t_c\rangle$ for the particle to cover one of the small straight-line segments which together make up the actual path, shown as the jagged line in Fig. 5.3. Of course, since we cannot follow the actual path in detail, we cannot really measure the value of $\langle t_c\rangle$, so we cannot determine the value of the constant N. However, we would expect that in thermal equilibrium, $\langle t_c\rangle$ will have a constant value which is independent of time, so we can conclude that N is also a constant (since Δt was chosen to be constant), even though we cannot determine its actual value.

Now that we have shown how the mean square of each instantaneous velocity component can be determined, even though the instantaneous velocity cannot be measured directly, we can state several remarkable properties of these quantities. These properties have been determined from many careful experiments, some of the most important of which were performed by the French

---

[7] Of course, the average of each component of the displacement is zero, that is $\langle d_x\rangle = \langle d_y\rangle = \langle d_z\rangle = 0$. Therefore, the average of each component of the average velocity also vanishes, since $\langle(v_{ave})_x\rangle = \langle d_x\rangle/\Delta t$, and similarly for the other components.

physicist Perrin at the end of the nineteenth century, carried out on a variety of suspensions of small solid particles. In the first place, in a suspension in which all particles have the same mass, these mean squares are all equal for each component of every particle. That is,

$$\langle v_x{}^2(t)\rangle = \langle v_y{}^2(t)\rangle = \langle v_z{}^2(t)\rangle, \qquad (5.14)$$

for any particle, for any arbitrary choice of Cartesian axes. We will therefore denote each of the terms in Eq. (5.14) by the expression $\langle v_u{}^2(t)\rangle$, where the subscript u refers to a component in any arbitrarily chosen direction. We can now state the next result, which is

$$\langle v_u{}^2(t)\rangle \propto T_g. \qquad (5.15)$$

That is, the mean square of any Brownian motion velocity component is proportional to temperature as measured on the gas-thermometer (or ideal gas), scale. Finally, from experiments somewhat different from those just described, it is possible to conclude that, at constant temperature, the mean-square velocity component is inversely proportional to the mass of the particle; that is,

$$\langle v_u{}^2(t)\rangle \propto 1/m. \qquad (5.16)$$

If we now combine Eqs. (5.15) and (5.16), we can write

$$\langle K_u\rangle \propto T_g, \qquad (5.17)$$

where $\langle K_u\rangle$ is the average kinetic energy associated with motion in the (arbitrary) direction denoted by u.

Thus the essentially random Brownian motion has, on the average, certain definite regularities, among which is another connection between mechanical properties and temperature. A little later, we will discuss the deep-lying significance of this appearance of order out of chaos. But first, we will rephrase the result expressed by Eq. (5.17) in a form that reflects, more directly, the kinematics of the motion we have been describing.

For a single particle in a suspension, the total average kinetic associated with its Brownian motion is just three times $\langle K_u\rangle$, since although u denotes an arbitrary direction, there are only three such directions that can be chosen independently at one time, in three-dimensional space. We can therefore say that the motion of the particle in a suspension has three "degrees of freedom." Stated formally, the number of degrees of freedom, which we will denote N_{df}, is just the number of independent terms required in the sum that gives the kinetic energy K (see Eq. 3.1). For instance, if the motion of the grains were constrained to lie in a surface (perhaps by suspending them in a very thin liquid film), then $N_{df} = 2$. For a particle representing a bead sliding on a wire, $N_{df} = 1$. So in general, we can write for the average of the total kinetic energy,

$$\langle K_t\rangle = N_{df}\langle K_u\rangle. \qquad (5.18)$$

This equation is actually more general than we have indicated so far. Until now, we have been speaking of particles, so we have been limited to discussing only translational motion. But rigid bodies can perform not only translational motion, described in terms of a particle located at the center of mass, but also rotational motion about an axis through the center of mass. Again, in three-dimensional space, there are three independent axes about which rotation can take place, so for a freely rotating rigid body there are three additional terms in the kinetic energy, each having the form

$$(K_u)_{rot} = \tfrac{1}{2}I_u(d\phi/dt)^2, \qquad (5.19)$$

where I_u is the moment of inertia about the axis denoted by the subscript u, and $d\phi/dt$ is the angular speed (measured in radians/sec), about that axis. In our discussion henceforth, we shall not distinguish between rotational and translational kinetic energy unless we are describing a specific physical arrangement of a particle or rigid body. No such distinction is necessary in a general discussion because

$$\langle (K_u)_{rot} \rangle = \langle (K_u)_{trans} \rangle, \qquad (5.20)$$

where $\langle (K_u)_{trans} \rangle$ is the average kinetic energy in translational Brownian motion, which earlier we denoted simply by $\langle K_u \rangle$. Hence Eq. (5.18) remains valid in general, so long as N_{df} is taken to mean the total number of degrees of freedom, translational as well as rotational.

So far, we have been considering free motion, with no potential energy contribution, V, to the total mechanical energy.[8] But in oscillating motion since (see section 5.3 following) there is a contribution to V from the restoring force (or restoring torque), and $\langle V_u \rangle = \langle K_u \rangle$. Hence for a degree of freedom in which oscillating motion is taking place, the average of the total mechanical energy is

$$\langle E_u \rangle = \langle K_u \rangle + \langle V_u \rangle = 2\langle K_u \rangle. \qquad (5.21)$$

We shall describe an example of such motion in section 5.3. Then, in section 6.1, we embark on a discussion of the significance of the relation expressed by Eq. (5.18).

5.3 BOLTZMANN'S CONSTANT

We describe in this section a measurement of mechanical fluctuations which allows the constant, in the proportion expressed by Eq. (5.17), to be determined directly. We saw above, in connection with Eq. (5.13), that the constant of proportionality could not be found from measurements on Brownian motion of suspended particles because the quantities measured in such observations are not the instantaneous velocity components, v_u, but the components, $(v_{ave})_u$ of the average velocity. If we wrote Eq. (5.17) as an equation, we would introduce a constant for the ratio K_u/T_g, and this constant, call it C_1, should be the same for all degrees of freedom. Then the ratio $\langle v_u{}^2 \rangle/T_g$ is another constant, say C_2, and $C_2 = 2C_1/m$, where m is the mass of the particle. But we cannot measure $\langle v_u{}^2 \rangle$, since the path cannot be followed in detail, so we must replace it by the measured quantity $\langle (v_{ave})_u{}^2 \rangle$. The ratio $\langle (v_{ave})_u{}^2 \rangle/T_g$ is still another constant,

$$C_3 = C_2/N = C_2\langle t_c \rangle/\Delta t = (2\langle t_c \rangle/m\Delta t)\cdot C_1 \qquad (5.22)$$

Hence although C_3 can be measured for any particular suspension, we cannot thereby determine the value of the constant C_1, which is really of interest because it is the same for all degrees of freedom. The reason is, as we see from Eq. (5.22), that the ratio C_3/C_1 contains the undetermined quantity $\langle t_c \rangle$.

Fortunately, Eq. (5.21) provides a means for avoiding this difficulty. In an oscillating system, the energy can be expressed in terms either of the kinetic or potential energy. Thus Eq. (5.21) could equally well be written

$$\langle E_u \rangle = 2\langle V_u \rangle. \qquad (5.23)$$

Now V_u is a function of the displacement, so $\langle V_u \rangle$ can be found by measuring the mean-square displacement of fluctuations.

The oscillating system used in the experiment we will describe[9] was a

---

[8] The particles in a suspension have gravitational potential energy, V_g; but since the downward velocity $v_T \approx 0$, V_g is constant, and independent of time to a good approximation for any particle. This is in contrast to K_u, which does vary with time because $v_u{}^2$ is varying with time. Only the averages $\langle v_u{}^2(t) \rangle$ and $\langle K_u(t) \rangle$ are constant in time.

[9] This experiment was reported by E. Kappler, in the journal "Annalen der Physik," Volume 11, page 233, published in 1931.

torsional pendulum consisting of a small mirror ($m \approx 5$ milligram), suspended on a very fine quartz fiber (diameter about 0.1 micron). The system has just one degree of freedom, corresponding to rotation of the mirror about the axis of the fiber. If the fiber is twisted by an initial angle ϕ_0, it exerts a restoring torque, M, which is proportional to the angle of twist, and so can be written

$$M = -\kappa\phi \qquad (5.24)$$

where κ is the restoring torque constant. Hence the mirror executes sinusoidal torsional oscillations of angular frequency.

$$\omega = (\kappa/I)^{\frac{1}{2}} \qquad (5.25)$$

where I is the moment of inertia of the mirror. Thus κ can be determined from Eq. (5.25) by measuring ω and I. The potential energy V_ϕ is just

$$V_\phi = \tfrac{1}{2}\kappa\phi^2 \qquad (5.26)$$

and the kinetic energy is (see Eq. (5.19))

$$K_\phi = \tfrac{1}{2}I\dot{\phi}^2. \qquad (5.27)$$

Since

$$\phi(t) = \phi_0 \cos \omega t, \quad \dot{\phi}(t) = -\phi_0 \cos \omega t \qquad (5.28)$$

we have

$$V_\phi(t) = \tfrac{1}{2}\kappa\phi_0{}^2 \cos^2 \omega t,$$

$$K_\phi(t) = \tfrac{1}{2}I\omega^2\phi_0 \sin^2 \omega t \qquad (5.29)$$

Now the average, over one period, of both $\sin^2 \omega t$ and $\cos^2 \omega t$ is equal to $\frac{1}{2}$, so we see from Eqs. (5.29) and (5.25) that $V_\phi(t)$ and $K_\phi(t)$, averaged over a period, are equal, as stated in connection with Eq. (5.21).

As the oscillations proceed, the initial energy $E_0 = \frac{1}{2}\kappa\phi_0{}^2$ is dissipated, and ultimately the system reaches an equilibrium state, in which the only remaining motion is fluctuations about $\phi = 0$. From Eq. (5.23) we can write

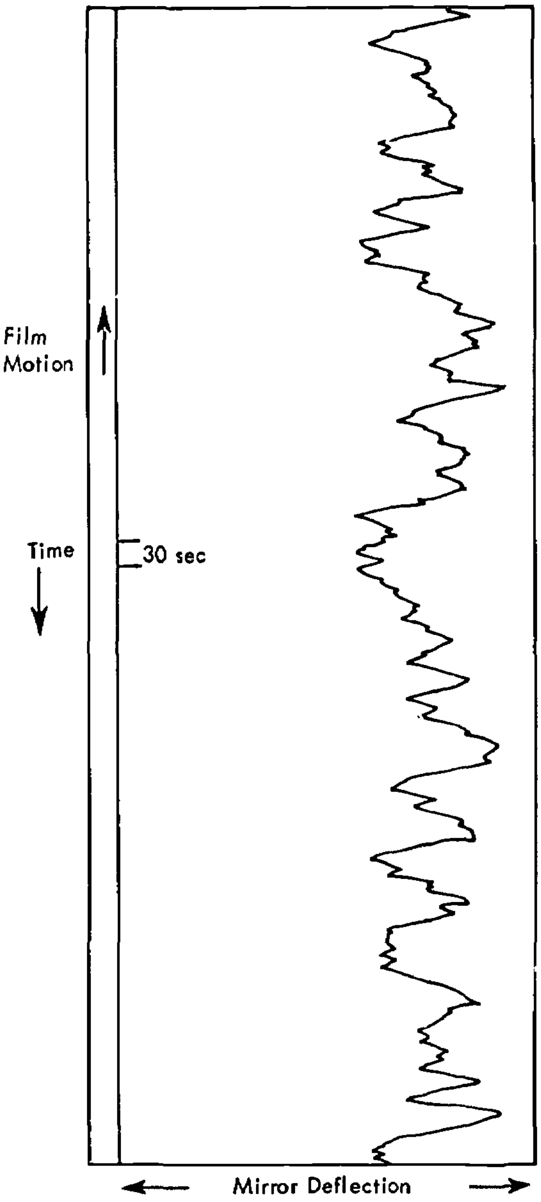

Fig. 5.4 Fluctuations in angular position of the mirror attached to the quartz fiber in Kappler's experiment, as recorded by reflecting a beam of light from the mirror to a photographic film moving slowly in the vertical direction. The distance between the two short lines represents an interval of 30 seconds.

$$\langle E_u \rangle = 2(\tfrac{1}{2}\kappa)\langle \phi^2 \rangle = k_B T_g, \qquad (5.30)$$

where k_B is the proportionality constant we have been seeking, between the average energy in mechanical fluctuations, and T_g (note that this constant is just twice the constant C_1, which was introduced in the discussion of Eq. (5.22)). The constant k_B is called Boltzmann's constant, after the physicist who postulated the relation in Eq. (5.17). Its value can be determined from Eq. (5.30), once κ has been found from Eq. (5.25), by measuring $\langle \phi^2 \rangle$, the fluctuations in the position of the mirror as it makes random oscillations about the fiber axis, in equilibrium. Kappler recorded the position of the mirror by reflecting a point light source from it onto a photographic film, which was moved slowly in the vertical direction. An example of one of these records, showing fluctuations over a period of about half an hour, is shown in Fig. 5.4.

Since $\langle E_u \rangle$ has the dimensions of energy, the dimensions of k_B are energy/degree, so its value depends on the size of the degree for the temperature scale used. With the degree appropriate to the T_g scale, which we denoted °K[10], the best current value is[11]

$$k_B = 1.38054 \times 10^{-16} \text{ erg/°K or}$$

$$k_B = 1.38054 \times 10^{-23} \text{ J/°K},$$

where J is the Joule, the unit of energy on the MKS scale, (1 J = 10^7 erg).

The small value of k_B in cgs or MKS units explains why mechanical fluctuations are not commonly observed, but require special circumstances, such as suspensions of small particles, in order to be seen. From Eq. (5.30) we can see that

$$\langle v_u^2(t) \rangle = (k_B/m)T_g, \qquad (5.31)$$

so that for a mass as small as a microgram (say a cube of ice about 0.1 mm. on a side), and a temperature in the neighborhood of room temperature (about 3×10^2 °K), $\langle v_u^2(t) \rangle \approx 4 \times 10^{-8}$ cm²/sec². The observed speeds will be of the order of $\sigma(v_u)$, or the square root of $\langle v_u^2(t) \rangle$, or $\approx 2 \times 10^{-4}$ cm/sec = 2 micron/sec, and thus would be imperceptible without magnification. For a mass of a gram, $\sigma(v_u)$ will be 10^3 times smaller yet.

---

[10] This degree is the same as that on the common metric Celsius scale (°C), since the two scales differ only by the displacement of their origin, that is, $T_g(°K) = T_c(°C) + T_0$ (see Eq. (4.4)) where $T_0 = 273.16$°K.

[11] Boltzmann's constant can be determined in a variety of different ways besides fluctuation measurements; in fact, fluctuations do not give the most precise value.

6 INTERNAL DEGREES OF FREEDOM

6.1 REGULARITY IN RANDOM EVENTS.

The proportionality of Eq. (5.18), which, in view of Eqs. (5.21) and (5.30) can be written

$$\langle K_u\rangle = \tfrac{1}{2}k_BT_g, \qquad (6.1)$$

expresses a remarkable regularity. For it asserts that this average is a constant, in equilibrium, which has the same value for every degree of freedom, rotational as well as translational. By virtue of Eq. (5.21) the average of every potential energy contribution in an oscillating system also has the same value. But of course, $K_u(t)$ fluctuates randomly as the velocity fluctuates, so Eq. (6.1) means that we have found order in the midst of chaos. It was pointed out at the end of the last section that $\langle K_u\rangle$ is a very small energy on a macroscopic scale of energies, but regardless of its size, its very existence poses a problem which we now investigate.

Actually, this sort of regularity in averages associated with random events is commonly observed, provided that the number of random events is large enough. For instance, if a fair coin is tossed N times, the order in which heads and tails appear is random. However, as N increases, the ratio of tails to heads, t/h, approaches unity. But in the same limit, that is, ever-increasing N, the average of the magnitude of the difference between the number of heads and the number of tails, $\langle|h - t|\rangle = \langle|\Delta N|\rangle$, does not decrease. In fact, this average increases without limit as N increases. Hence the sequence of tails and heads remains random, no matter how long the game is played.

At first sight, it might appear that the statement, "Average of $|\Delta N|$ increases, as N increases" is inconsistent with the statement, "The ratio of tails to heads [illegible], approaches unity as N increases [illegible] we can see that the two statements are compatible, in the following way.

The average value of the ratio of tails to heads is

$$\langle t/h\rangle = \langle (h \mp |\Delta N|)/h\rangle = 1 \mp \langle |\Delta N|/h\rangle \qquad (6.2)$$

(Since $|\Delta N|$ is always positive, the positive sign must be used in Eq. (6.2) if ΔN itself is negative.) As N increases, h, the number of heads, approaches N/2, so Eq. (6.2) can be written:

$$\langle t/h\rangle \approx 1 \mp 2\,\langle |\Delta N|\rangle/N \qquad (6.3)$$

Now as stated above, the average of $|\Delta N|$ increases with N, but it does not increase as fast as N. For large N, $\langle|\Delta N|\rangle$ is proportional to the square-root of N, that is, to $N^{\frac{1}{2}}$, so that as N increases, $\langle t/h\rangle$ approaches

$$\langle t/h\rangle \approx 1 \mp 2/N^{\frac{1}{2}}, \qquad (6.4)$$

and therefore the ratio $\langle t/h\rangle$ has the limit unity, as N increases without limit.

This discussion suggests that the constant averages $\langle v_u^2(t)\rangle$ and $\langle K_u(t)\rangle$ must arise as limits like that in Eq. (6.4). That is, they represent averages of some sort of repeated random process, the number of whose repetitions is very large. We must now ask, what is the nature of the random process involved? That is, what is the physical origin of the observed result, that successive values of $v_u(t)$ differ from each other?

If we are to retain the validity of Newton's Second Law, we must answer that $\Delta v_u = v_u(t_2) - v_u(t_1)$ can only arise from the action of a component of force, F_u, acting during the time interval $\Delta t = t_2 - t_1$. We have, from N-II,

$$\Delta v_u = \frac{1}{m} \int_{t_1}^{t_2} F_u(t)\, dt = \frac{1}{m} J_u \quad (6.5)$$

where J_u is the impulse. Now the systems we have been describing, namely particles in a fluid suspension, are mechanically isolated from everything but the fluid itself. The suspending fluid is therefore the only possible source of the required impulse J_u. And the origin of this impulse must be sought in degrees of freedom associated with the internal structure of the fluid. We have so far neglected to mention these internal degrees of freedom, for reasons which are discussed in Appendix 5.

Much of twentieth-century physics has been devoted to an investigation of the internal structure of matter. The modern atomic theory, based on quantum mechanics, which has resulted from this effort, gives us a largely satisfactory picture of this internal structure. In quantum mechanics, the distinction between the macroscopic and atomic levels is a fundamental one. It is basic to quantum mechanics that the degrees of freedom associated with the internal structure of matter, which we shall henceforth refer to as "internal degrees of freedom," do not, in general, behave like small copies of those we are familiar with on the macroscopic level. For instance, we can never expect to see an electron, no matter how refined our techniques become, in the same way that we can see pollen grains which are invisible to the naked eye, by using a powerful microscope.

Nevertheless, there are some respects in which the internal degrees of freedom do behave much like those of "classical" Newtonian particles. As a result of this similarity, many relations between thermal and mechanical properties can be understood without invoking quantum mechanics. The rest of this monograph will be devoted to examining some of these relations on the basis of the behavior of the internal degrees of freedom.

In the first place, from the discussion of Eq. (6.5), we see that if the averages $\langle K_u(t) \rangle$ and $\langle v_u^2(t) \rangle$ for a particle in a suspension are constant because the impulses J_u arise from random interactions with the internal degrees of freedom, then the number of those degrees of freedom, $(N_{dI})_I$, must be large. From the definition of N_{dI} in the last section, we would expect $(N_{dI})_I$ to be of the order of the number of atomic particles making up the internal structure of the suspension. If there are N_a of these atomic particles (atoms or molecules), in the suspension and they are all moving freely in three dimensions, then

$$(N_{dI})_I = 3N_a. \quad (6.6)$$

Secondly, we now have a very simple way of understanding the dissipation of mechanical energy. Suppose we drop a pollen grain of mass m, which has the temperature T_g' into a liquid at the same temperature. We give the grain an initial downward speed v_0. It will decelerate and quickly reach terminal velocity v_T, which, as we have seen, is very small. The initial mechanical energy $(m/2)v_0^2$ has all been dissipated. Now at the outset, this mechanical energy was associated with just one degree of freedom, namely that of the center of mass of the pollen grain, in the downward direction. The dissipation process consists simply of the sharing of this energy among all the internal degrees of freedom, numbering $(N_{dI})_I$, of both the pollen grain and the suspension. Before the pollen grain was fired into the suspension, each of the internal degrees of freedom had $\langle K_u \rangle = \frac{1}{2}kT_g'$ (see Eq. (6.1)). Since the atomic particles in the pollen grain are not moving freely, but also possess potential energy due to the elastic forces that bind them to each other, the total energy of each of their degrees of freedom is given by Eq. (5.21). Therefore, at the outset, the total mechanical energy of the system is

$$E_t = \tfrac{1}{2}mv_0^2 + \{2[(N_{df})_i]_{\text{grain}} + [(N_{df})_i]_{\text{suspension}}\}(\tfrac{1}{2}kT_g'), \quad (6.7)$$

and from Eq. (6.6), we have

$$E_t = \tfrac{1}{2}mv_0^2 + \{(\tfrac{1}{2}kT_g')[6(N_a)_{\text{grain}} + 3(N_a)_{\text{suspension}}]\}. \quad (6.8)$$

If the total mechanical energy is conserved, then after the pollen grain has come essentially to rest (since $v_T \approx 0$),

$$E_t = (\tfrac{1}{2}kT_g'')[6(N_a)_{\text{grain}} + 3(N_a)_{\text{suspension}}]. \quad (6.9)$$

The final equilibrium temperature $T_g'' > T_g'$, since we know that dissipation increases the temperature. The quantity in brackets in Eqs. (6.7) and (6.8) is called the internal energy, since it is associated with the internal degrees of freedom, in contrast to the term $\tfrac{1}{2}mv_0^2$, which is a property of the center of mass of the pollen grain. We will denote the internal energy by U, and we see from Eqs. (6.7) and (6.8) that it is a function of temperature. We can combine these two equations in the form:

$$U(T_g'') - U(T_g') = \Delta U = \tfrac{1}{2}mv_0^2, \quad (6.10)$$

or, in words, dissipation of mechanical energy increases the internal energy.

We will discuss the internal energy U further in the next section. Before doing so, we consider one further question connected with our discussion of dissipation. According to our description of dissipation, energy which is initially concentrated in one degree of freedom, namely the amount $\tfrac{1}{2}mv_0^2$ initially possessed by the z component of the motion of the pollen grain, is shared, during the approach to equilibrium, by all the degrees of freedom. Equilibrium is reached when this sharing is complete. We now inquire as to the mechanism by which this sharing takes place. In other words, how is macroscopic mechanical energy, $(K_z)_{\text{center of mass}}$, converted to internal energy U?

The answer to this question does not require the introduction of any new assumptions, at least for the Brownian motion of particles in a suspension, or for the molecules or atoms in a gas. Each particle in the internal structure possesses, on the average, an amount of energy equal to $\langle K_u \rangle$ for each degree of freedom. Hence each of these particles is incessantly moving, and therefore colliding with its neighbors. These collisions produce the impulses J_u. Since the collisions are random, their result is the uniform sharing of energy, on the average, by each degree of freedom. In fact, we can write an equation analogous to Eq. (6.3), replacing t/h by the ratio of K for any two degrees of freedom, say K_i and K_j. Then, as in Eq. (6.4) we have

$$(K_i/K_j) \approx 1 \mp (2/N^{\frac{1}{2}}) \quad (6.11)$$

where now N is the (enormous, number of successive collisions.

The same argument can be extended to the internal structure of a solid. In this case, since the atoms or molecules are bound by elastic forces, the average energy is $\langle E_u \rangle = 2\langle K_u \rangle$ per degree of freedom. Again, the continual motion results in interactions which provide the impulses J_u. Once more, the result of the random impulses is that each degree of freedom has an equal share of the total internal energy.

Since we have been treating all degrees of freedom, those associated with the macroscopic model as well as those of the internal structure, on an equal footing, it should be possible to construct a macroscopic model[12] of this equalizing process, which we will

---

[12] Notice that here, the word model is taken in the literal sense, like that of the ball and rod model of the solar system referred to in footnote 1.

refer to as "equipartition of energy." The game of billiards is an example of a model for the internal structure of a gas. Initially only the cue ball possesses K, but after it hits the other balls, which were initially at rest, they all share in the motion. The process is somewhat obscured by friction between the balls and the table. A better model can be constructed by placing smooth disks on a smooth table with many small holes through which air is forced, so that the disks are supported on a cushion of air. A photograph of such a model is shown in Fig. 6.1. Friction is very much smaller on such an air table than on a pool table, so that the approach to equilibrium can be followed for a much longer time. It is found that the ratio in Eq. (6.11), which here can be measured directly by using stroboscopic illumination to measure $v_u(t)$, as shown in Fig. 6.2, rapidly approaches unity. Further, by using disks of different mass, as shown in Fig. 6.3, the proportionality stated in Eq. (5.16) is also confirmed.

To the degree that the behavior of the macroscopic degrees of freedom associated with the motion of the center of mass of the disks (for each

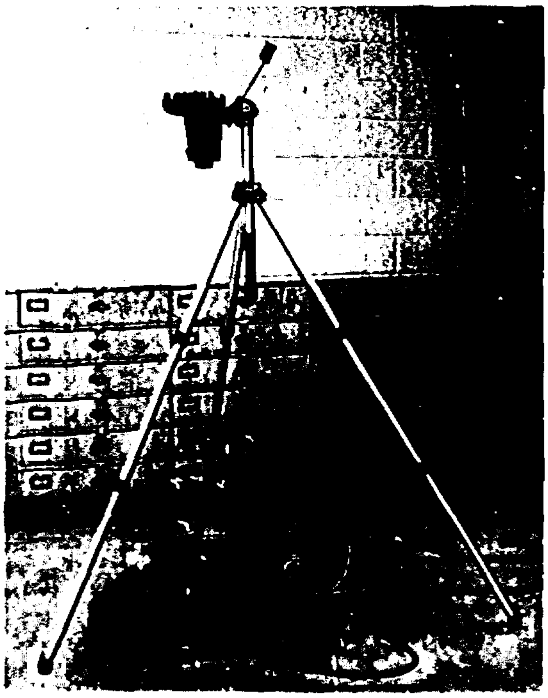

Fig. 6.1 General arrangement for observing two-dimensional collisions with negligible friction. The camera at the top of the tripod is focused on the air table on the floor. A wire strung along the edge of the air table serves as a rough approximation to a (nonuniform) temperature bath when the motor in the foreground rocks the table about a vertical axis. The pump which forces air through the hole in the table can be seen in the right background.

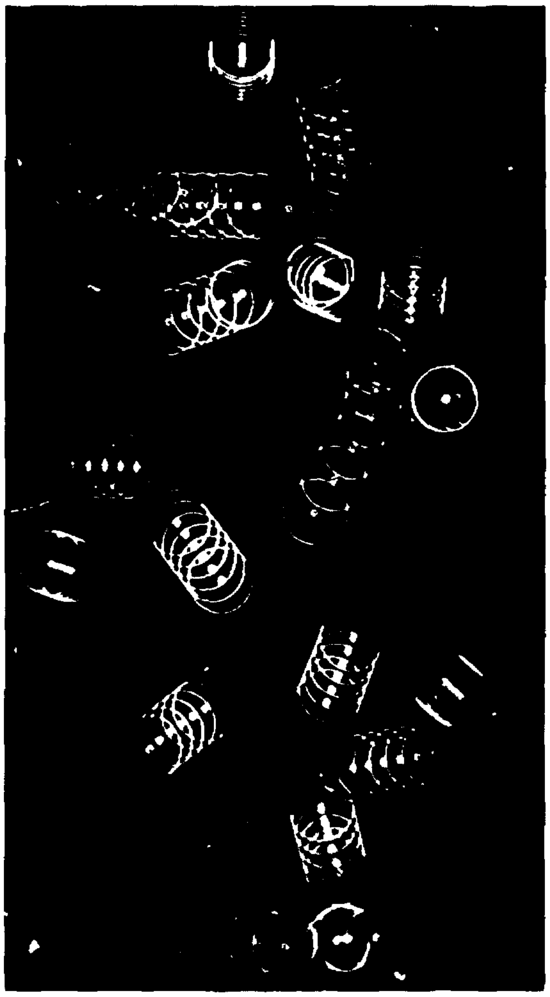

Fig. 6.2 A multiple exposure with stroboscopic illumination, using identical disks in the arrangement of Fig. 6.1. A collision took place near the bottom of the figure.

disk, $N_{df} = 2$ if rotation is neglected, $N_{df} = 3$ if rotation is considered), is similar to that of the internal degrees of freedom, this model provides a convenient way of examining the approach to equilibrium in detail.

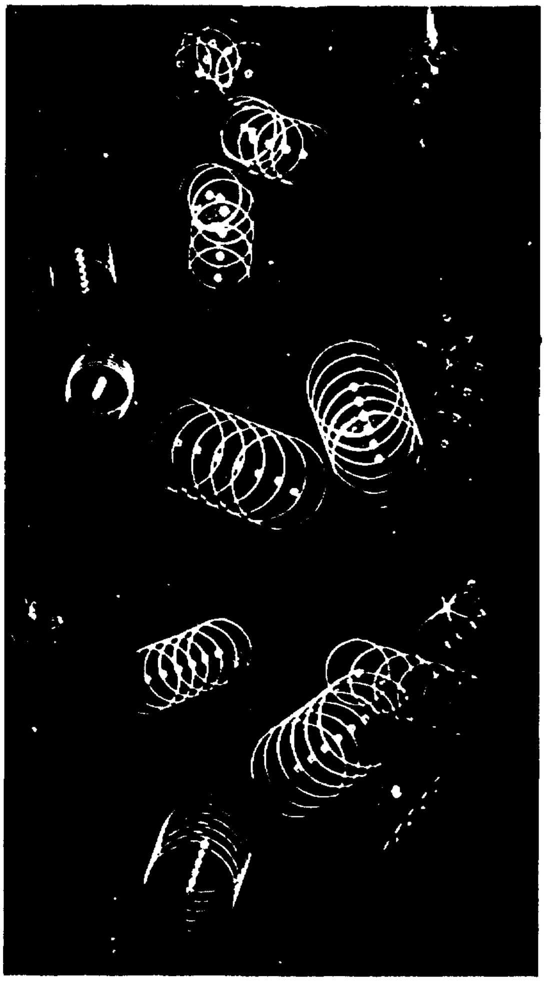

Fig. 6.3 The same as Fig. 6.2 with large and small disks.

6.2 INTERNAL ENERGY

Our definition of internal energy of a system as the total mechanical energy associated with all the degrees of freedom of its internal structure (which we have called "internal degrees of freedom") was deceptively simple.[13] We approached the definition through the examination of fluctuations in free macroscopic particles, although we have also referred to oscillating systems in passing. For a free macroscopic particle, we have asserted that $\langle K_u \rangle$ is related to T_g by Eq. (6.1) and for a macroscopic oscillator, Eq. (5.21) relates $\langle E_u \rangle$ to T_g. We will refer to these as "classical" relations, since they are obeyed by macroscopic degrees of freedom, and we will call any degree of freedom, macroscopic or internal, for which these hold, a "classical" degree of freedom. Now both of these results are examples of the "lamppost" technique described in section 1. We have stressed in various places, and especially in Appendix 5, that in general, internal and macroscopic degrees of freedom cannot be expected to behave alike. That is, it is most unlikely that internal degrees of freedom will behave "classically." Hence we may wonder how Eqs. (6.1) and (5.21) can be used in discussing internal energy, or even if they can be used for that purpose at all. There is just one physical system that is a reasonable approximation to a collection of free particles on the atomic level - a gas at low pressure. The suspension of fine particles we have been discussing is also a reasonable approximation, but on a macroscopic level. And our analysis showed (see Eq. (5.13)), that in this system the relation between measurable speeds and T_g contains the constant N, whose value cannot be determined. In section 5.3, we discussed a macroscopic system whose fluctuations do give a value for k_B. But as stressed in Appendix 5, the assumption that all internal

---

[13] Notice that we exclude from the internal energy any mechanical energy that an internal degree of freedom, in a rigid body, possesses by virtue of the motion of the center of mass of the body. That is, the motion corresponding to internal energy is motion relative to the center of mass. It is therefore described in a coordinate system in which the center of mass of the rigid body is at rest.

degrees of freedom behave like macroscopic ones, is in general unwarranted.

In the next section, we shall see how departures from classical behavior by the internal degrees of freedom can be observed in purely macroscopic measurements. The rest of this section is devoted to two other matters. First, we make more explicit how to distinguish between mechanical energy, which is our name for energy associated with degrees of freedom associated with the motion of the center of mass or of rotation about the center of mass (we shall refer to these as "center of mass degrees of freedom"), and internal energy, belonging to the degrees of freedom of the internal structure. Then we discuss how the internal energy can be related to the macroscopic equilibrium properties, the (P_i), which we introduced in section 4.1.

6.2.1 Mechanical Energy and Internal Energy

The first problem, of distinguishing between mechanical energy and internal energy, is important because we based our introduction of internal degrees of freedom in the last section on complete equivalence, so far as sharing energy is concerned, between the degrees of freedom of the center of mass and those of the internal structure. If there were no qualifications to this equivalence, then the particle model could not possibly describe motion adequately, even for macroscopic objects. For the particle model ignores the internal degrees of freedom completely. In fact, a useful definition of a "mechanical system" (in contrast to a "thermal system"), is one in which only changes in mechanical energy due to external forces need be considered. In such an idealized system, the internal energy is constant, unaffected by the motion of the object or by the forces acting on it. Hence, mechanical energy is conserved, and there is no sharing of energy with the internal degrees of freedom (no dissipation). We must therefore inquire how it can be that conversion of mechanical energy into internal energy occurs in all real systems, while the ideal mechanical system is actually a useful abstraction, despite the fact that it ignores all internal structure.

Of course, every macroscopic object has internal structure, and internal degrees of freedom associated with it. This structure is very complex; it comprises molecules, whose constituent atoms themselves are made of electrons and nuclei, which in turn consist of neutrons and protons. It turns out, as we shall see in the next section, that equipartition of energy does not apply to all the internal degrees of freedom. However, at least some of them behave "classically," as defined earlier in this section, and the following remarks apply only to these.

Let us examine the consequences, for these classical internal degrees of freedom to which equipartition of energy does apply, of giving a velocity, $\vec{V}$, to the center of mass of an object which was initially at rest. The mechanical energy, of course, is increased by $\frac{1}{2}mV^2$. When the object was at rest, the components of the internal structure were subject to interaction forces which we denote $(F_u)_0$ (see Eq. (6.5)). These forces are responsible, as we have seen, for the attainment of equipartition of energy among the classical degrees of freedom. When the center of mass is moving with velocity $\vec{V}$, we denote the interaction forces $(F_u)_V$. We now ask, what is the relation between $(F_u)_0$ and $(F_u)_V$?

We recalled above that the particle model is an ideal abstraction in which dissipation does not occur, and that dissipation is simply the conversion of mechanical energy into internal energy. Now when our object is at rest, we know that the interaction forces $(F_u)_0$ do not change the mechanical energy. An external force is required to set the object in motion. In referring to an ideal situation in which dissipation does not occur, the particle model embodies the assumption that the interaction forces $(F_u)_V$, which act when the object is moving with velocity $\vec{V}$, do not change the mechanical energy, $\frac{1}{2}mV^2$, either. Now if this were not a good assumption, the particle model would not be a useful abstraction at all. In fact, no reasonable model could be made, which, like the particle model, ignores the internal structure.

This assumption about the interaction forces $(F_u)_V$ could, of course, be justified by the success of the particle model. But the remarkable feature about this property of the interaction forces is that it does not require a special assumption at all. For, as we discussed at the beginning of section 4, forces are the same in all reference frames moving with respect to each other with constant velocity.[14] (Such frames are all "inertial frames" with respect to each other.) Therefore, when we give a macroscopic object a constant velocity $\vec{V}$, we do not change the internal interaction forces at all. Hence, the forces $(F_u)_V$, like the forces $(F_u)_0$, do not affect the mechanical energy, because $(F_u)_V = (F_u)_0$. As a consequence, the energy associated with the macroscopic motion is not shared with the internal degrees of freedom, unless there is some kind of new force exerted on the body as a result of the motion. Such forces are provided by friction, when the object moves in a resistive medium, or by the impulses occurring when it strikes an obstacle. These forces will then produce dissipation, and ultimately result in the sharing of the macroscopic energy among all the internal degrees of freedom, thus increasing the internal energy.

6.2.2 Macroscopic Properties and Internal Energy

We now turn to the relation between internal energy and the macroscopic properties, (P_i), of an object in thermal equilibrium. The expression given in Eq. (6.10) for the increase in internal energy resulting from the dissipation of mechanical energy, is a little misleading. In that equation, U was written as a function of temperature only. This is certainly true for the classical degrees of freedom of a free particle. But in general, if we wish to express the internal energy of a system whose internal degrees of freedom are more complicated, we must expect that U will depend on other properties of a thermal equilibrium state besides T_g. Thus we may write

$$U = U(T_g, P_i).$$

If we do this, we might appear to be back at the point at which we were led to introduce the concept of temperature. The internal energy has a unique value in each thermal equilibrium state (we will describe this property of the internal energy by calling it a "function of state"), but unfortunately, other properties besides the temperature, and conceivably many other properties, might be required in order to specify the state and hence the internal energy. But again, the fact that the number of internal degrees of freedom is enormous comes to our rescue, as it did in establishing Eq. (6.1).

All of the properties P_i of a macroscopic thermal system in equilibrium are related to averages of corresponding properties of components of the internal structure. For instance, the volume V_s, is the product of the number N_a of atoms or molecules, and the average volume $\langle V_a \rangle$ associated with each of the atoms or molecules. Because of the fluctuations in the posi-

---

[14] When an object is rotating about its center of mass, the reference frame in which it is at rest is not an inertial frame with respect to the reference frame in which the object is rotating. Hence the internal forces are affected by rotation, and the argument given here does not hold. Thus, in order for the particle model to be valid for rotation, the special assumption is required, that the changes in interaction forces due to rotation do not, by themselves, lead to appreciable dissipation.

tions of the individual atoms of molecules, V_a will vary from one to the next, but because of the enormous number of atoms or molecules, $\langle V_a \rangle$ will not, in general, have detectable fluctuations, so there will be no observable fluctuations in V_s either. The same sort of argument holds for other properties P_i.

Now we have seen that for condensed matter the elements of internal structure are not free. The interaction forces among them will in general depend on distance, since they are basically of electrical origin. Hence the potential energy of the internal degrees of freedom will depend on V_a, and therefore the average potential energy $\langle V_u \rangle$ will be a function of $\langle V_a \rangle$. Therefore, we may expect in general that the internal energy, U, in a thermal equilibrium state is a function of V_s, the volume of the system, as well as a function of T_g: $U = U(T_g, V_s)$. All the other P_i will also depend on either T_g (through $\langle K_u \rangle$), or on V_s, so we would expect that just the two macroscopic variables, T_g and V_s, would completely determine all the P_i in a thermal equilibrium state.

This expectation is borne out, provided that $\langle V_u \rangle$ depends only on the average separation of the atoms or molecules, as expressed by $\langle V_a \rangle$. If the atoms have a net magnetic moment, however, then $\langle V_u \rangle$ will also depend on the external magnetic field H or the magnetic induction B, and $U = U(T_g, V_s, H)$. Or if dipole moments are also present, U will depend on the electric field, E, so $U = U(T_g, V_s, H, E)$ in this case. Finally, the dependence of $\langle V_u \rangle$ on $\langle V_a \rangle$ may be different in different chemical species, so if a system is chemically heterogeneous, that is, if it contains chemical species $M_1, M_2, \ldots$, then U will depend on the relative amounts of each, $m_1(M_1), m_2(M_2) \ldots$, $U = U(T_g, V_s, m_1, m_2 \ldots)$, if no magnetic moments or electric dipoles are present.

We can conclude from this discussion that since the internal energy is a function of the thermal equilibrium state in the sense defined, the thermal equilibrium states must be uniquely defined by just the limited number of macroscopic variables that appear in the various parentheses as arguments of the function U. Hence relations among different thermal equilibrium states can be expressed purely in terms of macroscopic variables, with no need at all to mention the variables on the atomic level to which these macroscopic variables are related. The remark "no need at all" requires some qualification, however. There <u>is</u> a need to inquire about the internal structure, if we are interested in the <u>source</u> of the order that is observed on the macroscopic level. Nevertheless, it remains true that for certain purposes such inquiry about the atomic level is a luxury. The study of relations among thermal equilibrium states, which is the subject of "classical thermodynamics," is a very powerful tool for investigating a large class of natural phenomena. Much of the study of chemical reactions, for instance, can be treated entirely by classical thermodynamics.

6.3 HEAT AND HEAT CAPACITY

We see from the above discussion that if, as a result of some operation performed on a system in a thermal equilibrium state, which we will denote S_1, the internal energy U of the system is changed, then the system must have undergone a transition to a new state, say S_2. Now the application of a force to an object which is free to move is <u>not</u> such an operation; the center of mass will acquire kinetic energy equal to

$$K = \int_{\vec{r}_1}^{\vec{r}_2} \vec{F}(\vec{r}) \cdot d\vec{r}, \qquad (6.12)$$

where $\vec{r}_1$ is its initial position vector, $\vec{r}_2$ is the position vector at which the force is removed, and $\vec{F}(\vec{r})$ is the net external force applied so that the integral in Eq. (6.12) is the work done on the object. As we have seen, the internal energy of a system has a value which is independent of the mechanical energy of its center of mass, although as a result of naturally occurring process this mechanical energy is sooner or later dissipated into internal energy.

It is possible, however, for a force to do work which changes the internal energy of a system. For instance, if the force is applied to a rigid body which is constrained so that it can't move, the work done by the force is stored as elastic internal energy in the rigid body.[15] The elastic energy is internal energy because, in this case, the force deforms the body. So V_s is changed, and therefore so is $\langle V_s \rangle$; and we have seen that $\langle V_u \rangle$, and therefore U, depends on $\langle V_s \rangle$. (Does this argument hold for shear deformation, in which only the shape, but not the size, of the object is changed?) In both of these cases, that is, a rigid body (a) whose center of mass has been given kinetic energy (before this has been dissipated into internal energy and the body has come to rest), or (b) which has been given elastic potential energy, the work which has been done can all be recovered as mechanical energy. For instance, the object with kinetic energy, K, can be brought to rest by having it lift a weight, so that K is converted into gravitational potential energy, V_g.

In general, however, work which has been converted into internal energy cannot be completely recovered. A complete discussion of the conversion of internal energy into work is well beyond the scope of this monograph, so we will conclude with a further discussion of the reverse process, the performance of work on a system so as to change its internal energy. We will be interested in two of its aspects in particular: first, to describe in more detail how it is related to measurements of internal energy; and second, to show briefly how some very important properties of the internal degrees of freedom can be deduced from measurements of internal energy.

We have already given brief attention to the first question, measurement of internal energy, when we wrote Eq. (6.10). We will rewrite the relation stated there as

$$U(S_f) - U(S_i) = \Delta U = \Delta W_A. \qquad (6.13)$$

In this relation, S_i and S_f refer to the initial and final states respectively, as identified by their respective values of T_g, V_s, and whatever other macroscopic variables are needed to specify them completely (for instance, electric field, E; magnetic field, H, etc.).[16] The subscript A is attached to the expression for the work done, to emphasize that this relation holds only under special circumstances, namely, when work is performed on the system, with the system thermally isolated from its surroundings. Such a process is called an "adiabatic process," and the subscript A was chosen to refer to the term adiabatic.

The reason for attaching this condition of thermal isolation, that is, requiring that the process be adiabatic, is that the internal energy can also be changed without any work being done on the system. We have stressed the fact that the temperature of a system will change if it is placed in thermal contact with another system at

---

[15] This description is, to some extent, a contradiction in terms. The ideal rigid body, introduced as an abstraction in classical mechanics, has a fixed volume V_s, which is independent of external forces. In order to discuss elastic forces, as we have just done, a different model, the "elastic body" is introduced. However, the theory of elasticity, which has been extensively developed, starting early in the nineteenth century, treats the elastic body as a mathematical continuum, without any discrete internal structure. It regards such a body as made up of geometrical points, much in the same way as we regard space, and does not introduce the notions of atoms or molecules.

[16] Note that if the system does mechanical work at the expense of its internal energy, ΔU will be negative, and so will ΔW_A.

a different temperature. Since U is a function of T_g (as well as other variables), a temperature change corresponds to a change in U (as indicated in Eq. (6.10)), provided that the other variables which define the state S, remain constant. In what follows, we shall, for simplicity, consider only systems for which $U = U(T_g, V_g)$ and no other variables. For such a system, then, if its volume, V_g, remains essentially constant, a change in T_g implies a change in U. Hence we have asserted that changes in U occur in systems with constant V_g, when they are placed in thermal contact with a system at a different temperature.

This process, exchange of internal energy between systems originally in thermal equilibrium at different temperatures after they are placed in thermal contact, is usually referred to as "heat transfer." The term "heat" means simply internal energy which is exchanged between systems in thermal contact, in the course of their approach to mutual thermal equilibrium. If this exchange is not accompanied by the performance of any work, that is, if it is a purely thermal process, then there is no other change in U.

Suppose we have a system labeled "1," initially in thermal equilibrium state S_{1i} at temperature T_{1i} (henceforth the subscript g on T will not be written, although it is still implied that we are using the gas-thermometer scale), placed in thermal contact with a system labeled "2," initially in thermal equilibrium in state S_{2i} at temperature T_{2i}. When placed in thermal contact, they reach states S_{1f} and S_{2f}, which have the same temperature, T_f. If the two systems are in thermal isolation from all surroundings but each other, and if no work has been done on either system, then

$$U(S_{1i}) + U(S_{2i}) = U(S_{1f}) + U(S_{2f}), \qquad (6.14)$$

since there has been no change in the total internal energy of the two systems taken together. If we transpose terms, we have

$$U(S_{1f}) - U(S_{1i}) = -[U(S_{2f}) - U(S_{2i})] \qquad (6.15)$$

or

$$\Delta U_1 = -\Delta U_2 . \qquad (6.16)$$

Each of the terms in Eq. (6.16) represents an internal energy change due to heat transfer. Such internal energy changes are usually denoted in a special way, by the letter Q:

$$\Delta U_1 = \Delta Q_1 ; \quad \Delta U_2 = \Delta Q_2 \qquad (6.17)$$

and we see from Eq. (6.16) that

$$\Delta Q_2 = -\Delta Q_1 . \qquad (6.18)$$

In common parlance, ΔQ is often referred to as "an amount of heat," but we see from our discussion that such an expression can be misleading. The word "amount" is usually associated with a substance that can be obtained in isolation; we speak of "an amount of money," for example. But "heat" is not a substance in this sense at all. It merely represents a change in internal energy, U, occurring in a certain specified way, namely, as the result of thermal contact alone.

If we consider a general process, in which U changes both by thermal contact and by the performance of mechanical work as well, we can combine Eqs. (6.17) and (6.13), and write

$$\Delta U = \Delta W + \Delta Q. \qquad (6.19)$$

(Note that we have omitted the subscript A on ΔW, since now the process is no longer adiabatic.) This relation is called the First Law of Thermodynamics. We see that it expresses the conservation of internal energy, and thus is a generalization of the conservation of mechanical energy of macroscopic degrees of freedom. The actual historical development which led to its assertion was very different from the way in which we arrived at its statement.

That development took place before any but the vaguest notions had been established, concerning internal degrees of freedom. As a result, Eq. (6.19) was first stated as a generalization from purely macroscopic measurements, especially those of Joule. He performed mechanical work on various fluid systems, such as baths of water and mercury. The center of mass of the liquid always remained at rest, so none of the work ΔW_A became macroscopic kinetic energy; it was all dissipated into internal energy. The fluids were in thermal isolation from their surroundings. Joule detected the change in internal energy, ΔU, by measuring the resulting change in temperature, ΔT. He found that for a given fluid system, exactly the same ΔT was observed for a given expenditure of work ΔW_A regardless of the manner in which the work was dissipated; he used frictional processes of different sorts, and electrical dissipation as well. (The heating of a wire through which an electrical current passes is now called "Joule heating.") He argued, as we have done, that since T_i and T_f were the same in each case, that the different processes he used always took the system from the same initial thermal equilibrium state to the same final equilibrium state (changes in volume V_s were negligible). Hence he was able to conclude that ΔU was the same in each case. Since the value of ΔU did not depend on the particular process he employed, that is, on the manner in which ΔW_A was dissipated, or on the particular fluid system in which ΔW_A was dissipated, he concluded, as we did earlier, that U, the internal energy, is a "state-function" in the sense that we defined. Thus Joule was able to assert the validity of Eq. (6.19) without making any assumptions whatever concerning the internal structure of matter, or the internal degrees of freedom associated with that structure. Not only was Joule's argument a success, so were further deductions concerning the connection between macroscopic thermal and mechanical behavior that could be shown to follow from it. Those successes actually inhibited acceptance of atomic theory to a certain extent toward the end of the nineteenth century, as discussed in Appendix 5.

Equation (6.19) describes only <u>changes</u> in internal energy, and not any actual values of the internal energy in a particular state, which we have been denoting $U(S_1)$, $U(S_2)$, etc. So macroscopic measurements can only allow us to determine such differences ΔU. Still, these differences, measured macroscopically, can give information about the internal degrees of freedom, whose behavior they represent.

First, let us describe how ΔU may be determined in practice. For the simple systems we are describing, in which the state S is determined by T and V_s alone, if V_s is held constant, $\Delta W = 0$, so from Eq. (6.19)

$$\Delta U = \Delta Q. \tag{6.20}$$

Now ΔQ can be measured by dissipating mechanical energy completely; a convenient way of doing this is by passing an electrical current though a wire. If the electrical energy is dissipated at a constant rate, $\dot{E}$, for a time Δt,

$$\Delta Q = \dot{E}\Delta t = \Delta U. \tag{6.21}$$

If the system remains homogeneous, that is, if it does not transform from solid to liquid, or liquid to vapor, this dissipation is accompanied by a temperature change ΔT. The ratio

$$\Delta Q/\Delta T = C \tag{6.22}$$

is called the "heat capacity." Under the conditions we have been describing, that is, constant volume so that $\Delta W = 0$, the ratio in Eq. (6.22) is the heat capacity at constant volume, denoted C_v, and we see from Eq. (6.20) that

$$C_v = \Delta U/\Delta T. \tag{6.23}$$

Therefore, measurements of heat capacity at constant volume give directly the way in which U changes with temperature, and indirectly, corresponding information about $\langle E_a \rangle$, the average mechanical energy per atom.

We shall discuss just one example of this sort of investigation. Although it deals with a very simple system, it still displays the inadequacy of classical mechanics as applied to internal degrees of freedom.

The system we investigate is a mass of gas. We have already remarked that all gases behave alike in conforming to Eq. (4.5) at sufficiently low pressure. This behavior is simplest to understand on the atomic model, in which a gas is just a collection of particles which are independent except for collisions. The collisions change K_u for the pair of colliding particles, but as we saw at the end of section 5.2, they bring about thermal equilibrium, in which $\langle K_u \rangle$ is a constant, and the same for each degree of freedom. Now in dense gases, at high pressure, the particles are close enough to exert mutual forces even when they are not colliding, and these will also make a contribution to $\langle E_a \rangle$. But when the pressure becomes very small, these forces, which depend on the average distance between particles, also become very small, so $\langle E_a \rangle \approx \langle K_a \rangle$. Therefore

$$U = (N_{df})_I (\tfrac{1}{2}kT) \qquad (6.24)$$

according to Eqs. (5.18) and (6.1). Hence from Eq. (6.23),

$$C_v = (N_{df})_I (\tfrac{1}{2}k). \qquad (6.25)$$

A system such as this, in which U = U(T) only, that is, with no dependence on V_g or any other variable such as E or H, is called an "ideal thermal system." As we have seen from Eq. (6.25), such a system has a heat capacity which is independent of T.

While no such ideal thermal system, such as the ideal gas we have been describing, exists in nature, a gas at low pressure is a reasonably good approximation. But strangely, in hydrogen, for which measurements of C_v are shown in Fig. 6.4, the expectation of a constant value which is independent of T, is not confirmed. Between about 300°K and 80°K, C_v decreases by the ratio 5 to 3. Now hydrogen gas is known from chemical evidence to form diatomic molecules, H_2, in the gas phase. The observed decrease can only be interpreted by assuming that $(N_{df})_I$

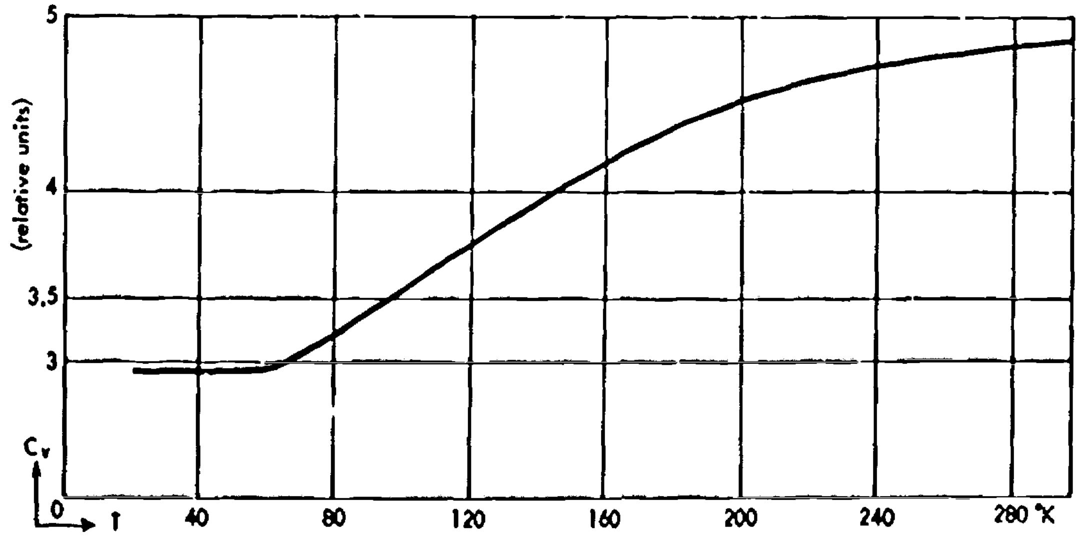

Fig. 6.4 The heat capacity at constant volume of one mole of hydrogen gas, as a function of temperature.

has decreased by the ratio 5/3 in this temperature interval. The denominator of the fraction, 3, is just the number of degrees of freedom of a free particle; the numerator, 5, suggests that when the temperature reaches about 300°K, the H_2 molecules have somehow achieved two additional degrees of freedom. From other evidence, these extra degrees of freedom represent rotation about two axes perpendicular to the axis of the molecule itself. At still higher temperatures, C_v increases further, in the ratio 7/5 to the value at about 300°K. Thus, two further degrees of freedom contribute to $\langle E_a \rangle$ at high temperatures; these are probably associated with vibrations along the axis between atoms in the molecule. Thus the classical model adequately explains the constant portions of the curve $C_v(T)$; but it has no way at all of accounting for the observation that $(N_{df})_I$ is also apparently a function of T. This result turns out to be a consequence of a fundamental feature of quantum mechanics as applied to the internal degrees of freedom. The expression for $\langle E_u \rangle$, as calculated by quantum mechanics, is considerably more complicated than the classical one, Eq. (5.30), in which $\langle E_u \rangle$ is simply proportional to T.

We cannot describe this quantum-mechanical result in detail, but a qualitative discussion can show how it leads to an apparent dependence of $(N_{df})_I$ on T. As we just saw, below about 80°K, the only degrees of freedom of the H_2 molecule whose energy varies with T are those of the three-dimensional translational motion of the center of mass of the molecule. These degrees of freedom behave classically; there is equipartition of energy among them, each one's share being $\frac{1}{2}k_BT$. All the other internal degrees of freedom of the molecule can be described as "frozen."[17] Their energy is not a function of T, so they do not contribute to the heat capacity.

Now the possibility of a frozen degree of freedom, one whose energy does not depend on T, can arise in the particle model of classical mechanics, if the motion in that degree of freedom is somehow constrained. For instance, as long as the disks described at the end of section 5.3 must move on a horizontal plane surface, they have only two degrees of freedom; the third, corresponding to vertical motion, might be described as frozen. In the case of the hydrogen molecule, the clearest indication of nonclassical behavior is the gradual "thawing" of two degrees of freedom, that is apparent in Fig. 6.4 between about 80°K and about 300°K. As mentioned above, there is evidence, which we cannot discuss here, that these two degrees of freedom correspond to rotation of the molecule about two axes, perpendicular to the line joining the two H atoms which make up the H_2 molecule. Above about 300°K, these two degrees of freedom are unfrozen, and behave classically.

As a result of the quantum-mechanical phenomenon of "thawing" in the rotational degrees of freedom, C_V increases, between 80°K and 300°K, from the constant value corresponding to three classical degrees of freedom, to the constant value corresponding to five classical degrees of freedom. Therefore, in a temperature region where some of the degrees of freedom are classical, while all the rest are frozen, Eq. (6.25) correctly accounts for the heat capacity, provided that the factor $(N_{df})_I$ is interpreted as the number of classical degrees of freedom.

The temperature region, in which a degree of freedom makes its transition from being frozen to behaving classically, occurs at different temperatures for different degrees of freedom. We have seen that this transition temperature region is between 80° and 300°K for the two rotational degrees of freedom. Above 300°K, the heat capacity of H_2 remains constant

---

[17] At one time, the word "ankylosed" was proposed to describe such degrees of freedom, but fortunately, it never gained much currency.

until about 600°K, when another degree of freedom begins to thaw, that corresponding to vibration of the two H atoms which make up the molecule, joining them along the line. The heat capacity then begins to increase again, but this degree of freedom does not reach its full classical contribution of k_B (why not $\frac{1}{2}k_B$?), until a temperature of several thousand degrees is reached. By that temperature, the internal structure of the H atoms themselves begins to contribute to the heat capacity, as the atoms begin to be ionized. They lose their electrons, and another aspect of the internal structure of the H_2 molecule becomes apparent. Thus, we see how successive thawing of different degrees of freedom leads to a dependence of $(N_{df})_I$, interpreted as the number of classical degrees of freedom, on T.

The freezing of degrees of freedom at low temperature, and their thawing as the temperature is raised, is a consequence of the general property which follows from quantum mechanics, that the energy associated with every degree of freedom is "quantized." This quantum-mechanical property is best appreciated by contrasting it with the corresponding situation in classical mechanics, in which a degree of freedom can have any energy whatever. For instance, according to Eq. (5.27), the kinetic energy associated with rotation is

$$K_\phi = \tfrac{1}{2}I\dot{\phi}^2 = \tfrac{1}{2}I\omega^2. \qquad (6.26)$$

Since the angular momentum, L, is

$$L = I\omega, \qquad (6.27)$$

we can write Eq. (6.26) as

$$K_\phi = L^2/2I \qquad (6.28)$$

In classical mechanics, there is no restriction on the angular momentum; L can have any value, and, correspondingly, the rotating system can have any value of K_ϕ. According to quantum mechanics, however, the angular momentum of a rotating system can take on only certain discrete values. For a H_2 molecule at low temperature, with the vibrational motion frozen so that the distance between the two H atoms is constant, this restriction takes the form

$$L^2 = \hbar^2\, r(r + 1), \qquad (6.29)$$

where r is an even integer, r = 0,2,4, . . . and $\hbar$ is Planck's constant,

$$\hbar = 1.05 \times 10^{-27}\ \text{gm cm}^2/\text{sec}.$$

(Note that since the unit of energy, the erg, has dimensions gm cm^2/sec^2, the dimensions of $\hbar$ can also be stated as erg sec.) Hence the only possible values of K_ϕ are those in the sequence

$$(K_\phi)_r = (\hbar^2/2I)\, r(r + 1)\ \text{erg}. \qquad (6.30)$$

Thus, in the lowest possible energy state for rotation, the "ground state," r = 0, and $(K_\phi)_0 = 0$, while in the first excited state for rotation, r = 2, and $(K_\phi)_2 = 6(\hbar^2/2I)$.

The quantum-mechanical characteristic, of frozen degrees of freedom at low temperature which thaw as the temperature increases, arises because of the finite difference in energy between the ground state and the first excited state, which occurs in all quantum-mechanical systems. As long as T is so low that the energy k_BT is very small compared to the difference in energy between the ground state and the first excited state, which we denote $(\Delta E)_e$, the degree of freedom remains frozen; when k_BT begins to be of the order of magnitude of $(\Delta E)_e$, the degree of freedom begins to thaw; and when k_BT has become about the same size as $(\Delta E)_e$, thawing is complete, and the degree of freedom behaves classically, and shares in equipartition of energy. In the case of the H_2 molecule, the moment of inertia can be calculated, since the mass of each H atom is $m_H = 1.67 \times 10^{-24}$ gm, and their separation, R, is $R = 0.74 \times 10^{-8}$ cm, so $I = 0.46 \times 10^{-40}$ gm cm^2.

Hence, from Eq. (6.30), $(\Delta E)_e$ is about 8×10^{-14} erg. We have seen from Fig. 6.4 that thawing of the rotational degrees of freedom is essentially complete when T = 300°K, which corresponds to $k_B T$ about 4×10^{-14} erg, or about one-half $(\Delta E)_e$.

The same sort of situation exists with respect to the vibrational motion of the H atoms with respect to each other, along the line joining them. For such vibration, the allowed values of the energy are, according to quantum mechanics,

$$(E_u)_n = \hbar\omega\,(n + \tfrac{1}{2}) \qquad (6.31)$$

where n is an integer, n = 0,1,2, . . . and ω is the angular frequency of vibration, determined by the mass of the H atoms and the restoring force constant (see Eq. (5.25) for an analogous expression applying to torsional oscillation). Hence in this case, $(\Delta E)_e$ is just $(\Delta E)_e = \hbar\omega$. The angular frequency for vibration of the H_2 molecule is about 8×10^{14} rad/sec, so $(\Delta E)_e$ is about 80×10^{-14} erg, or about ten times larger than that for the rotational degrees of freedom. Hence the thawing of the vibrational motion does not occur until the temperature is about ten times as large as that required for thawing the rotational motion, as described above.

Apart from this quantitative difference, the rotational and vibrational degrees of freedom also differ in an interesting qualitative respect. We have seen that for rotational motion, $(K_\omega)_0 = 0$, but in the vibrational ground state, according to Eq. (6.31), $(E_u)_0 = \frac{1}{2}\hbar\omega \neq 0$. Hence, while it is true, that as T approaches zero, all internal degrees of freedom tend to occupy their ground states, it is not true, as is sometimes incorrectly stated, that "at T = 0, all molecular motion ceases." As we have just seen, rotation does cease, but because $(E_u)_0 \neq 0$ for vibration, vibration continues, even at T = 0.

However, in all cases, the energy in the ground state is not a function of T, so an internal degree of freedom in its ground state makes no contribution to the heat capacity. Hence, when the ground states of the internal degrees of freedom become occupied as the temperature approaches zero, C_V approaches zero for all macroscopic systems. This property of C_V is closely related to the third law of thermodynamics, which is discussed briefly in Appendix 3.

That our discussion is concluded at this point should by no means be taken to imply that all of the details of our subject have been exhausted. Almost exactly the opposite is true; we have barely scratched the surface. But our goal, as stated at the outset, was rather limited, and we have given at least the beginnings of an answer to the question posed there: How can thermal phenomena be incorporated, in a coherent way, in the mechanical model of motion?

Appendix 1 TERMINAL SPEED

We choose a coordinate system as described in the text, with the positive z axis (and therefore the k unit vector), pointing up. Since we are concerned only with motion along the z axis, we need no subscripts; v will denote the speed along this axis.

For a particle falling freely the acceleration is negative (downward), and has magnitude g, so (N-II) gives

$$m(dv/dt) = -mg, \qquad \text{(A1.1)}$$

where $-mg$ is the gravitational force on the particle. In a resistive medium, there is an additional force which we will denote f, that retards the motion and is therefore directed opposite to v. This force does not depend on v in a simple way over a wide range of speeds. We shall therefore apply the "lamppost principle" discussed in the text, and assume a simple form for the speed-dependence of f, namely, f proportional to v. Because it is a retarding force we write

$$f = -kv, \qquad \text{(A1.2)}$$

where k is a constant of proportionality. (What are its dimensions?) Although this form for f is not generally valid, it has the virtue that the resulting equation of motion is somewhat easier to solve than a more realistic version, $f \propto v^2$.

Equations (A1.1) and (A1.2) then give for a particle falling in a resistive medium,

$$dv/dt = -g - (k/m)v. \qquad \text{(A1.3)}$$

This equation is inhomogeneous, since it contains the constant term g, while the other two terms involve the variables v and t. One particular solution can be found immediately, corresponding to equilibrium motion at constant terminal speed, $v = -v_T$. In this case $dv/dt = 0$ so

$$v = -mg/k; \quad v_T = mg/k. \qquad \text{(A1.4)}$$

From the intuitive discussion in the text, this solution corresponds to the situation after the particle has been falling a long time, that is, for $t \to \infty$. It is not correct when the particle is just beginning to fall. We need, in addition, a solution which is valid for finite t, and which reduces to the ideal value $v = -gt$ for $t \approx 0$. Such a solution, combined with (A1.4), will describe the motion at all times.

A solution valid for finite t can be found from the homogeneous equation obtained by deleting the constant term $-g$ from (A1.3):

$$dv/dt = -(k/m)v. \qquad \text{(A1.5)}$$

This equation states that the rate of decrease of v is proportional to v. It is thus identical with the equation for radioactive decay, except that in the latter, v is replaced by N, the number of decaying particles. The solution is obtained by cross multiplying and then integrating:

$$dv/v = -(k/m)dt \qquad \text{(A1.6)}$$

$$\ln v = -(k/m)t + \ln v_0, \qquad \text{(A1.7)}$$

where ln denotes the natural logarithm (base e) and $\ln v_0$ is a constant of integration. Raising both sides to powers of e, (A1.7) becomes

$$v(t) = v_0 \, e^{-(k/m)t}. \qquad \text{(A1.8)}$$

If Eq. (A1.8) is combined with Eq. (A1.4), the solution for $t \to \infty$, we have

$$v(t) = v_0 \, e^{-(k/m)t} - v_T. \qquad \text{(A1.9)}$$

This equation contains the still unspecified constant v_0. However, we have not yet satisfied the boundary condition at $t = 0$, namely that

$$v(0) = 0. \qquad (A1.10)$$

This condition, in conjunction with Eq. (A1.9) determines the constant $v_0 = v_T$. Thus we have

$$v(t) = v_T\, e^{-(k/m)t} - v_T, \qquad (A1.11)$$

and this is the curve which was drawn in Fig. 3.2 of the text.

We can check to see if Eq. (A1.11) is indeed the correct solution of Eq. (A1.3) by comparing the value it gives for the acceleration $a(t)$ with that specified by Eq. (A1.3). Since $k/m = g/v_T$, we have

$$v(t) = v_T\left[\exp -\left(\frac{g}{v_T}\right)t - 1\right], \qquad (A1.12)$$

so that $a(t) = dv/dt$ is

$$\begin{aligned} a(t) &= v_T\ (-g/v_T)\exp -\left(\frac{g}{v_T}\right)t \\ &= -g \exp -\left(\frac{g}{v_T}\right)t \qquad (A1.13) \end{aligned}$$

But Eq. (A1.3) can be written

$$\begin{aligned} a(t) &= -g - (g/v_T)v(t) \\ &= -g - (g/v_T)v_T\left[\exp -\left(\frac{g}{v_T}\right) - 1\right] \end{aligned}$$

$$(A1.14)$$

using Eq. (A1.12), or

$$\begin{aligned} a(t) &= -g - g \exp -\left(\frac{g}{v_T}\right)t + g \\ &= -g \exp -\left(\frac{g}{v_T}\right)t \qquad (A1.15) \end{aligned}$$

which agrees with Eq. (A1.13), as calculated from Eq. (A1.12).

Equations (A1.11) and (A1.12) can easily be made more general, so as to encompass cases in which the particle has an initial speed $v_i = v(0)$ different from zero. It is easy to see that this initial condition is satisfied by choosing $v_0 = v_i + v_T$ in (A1.9); the case considered above corresponds to $v_i = 0$. Then we have for the general equations

$$v(t) = (v_i + v_T)e^{-(k/m)t} - v_T \quad (A1.16)$$

$$v(t) = v_i\ e^{-(g/v_T)t} + v_T\left[e^{-(g/v_T)t} - 1\right]$$

$$(A1.17)$$

from which it is clear that $v(0) = v_i + v_T - v_T = v_i$ (since the exponential $= 1$ at $t = 0$), and $v(\infty) = -v_T$, since the exponential approaches zero as t increases without limit.

If we recalculate the acceleration, we find

$$a(t) = -\left(\frac{v_i + v_T}{v_T}\right)g \exp -\left(\frac{g}{v_T}\right)t \quad (A1.18)$$

so that

$$a(0) = -\left(\frac{v_i + v_T}{v_T}\right) g \qquad (A1.19)$$

Hence $a(0)$ differs from the value $a(0) = -g$ it has when the particle is dropped freely, by the factor $(v_i + v_T)/v_T$. If the particle is thrown up, $v_i > 0$ so this factor > 1 and the deceleration is more rapid than in free fall. If it is thrown down, $v_i < 0$, and the downward acceleration is smaller than in free fall. If $v_i = -v_T$, of course, $a = 0$ and the speed $v(t) = -v_T$ always. If the magnitude of the initial speed is larger than v_T, the acceleration is upward; the particle slows down from $v(0) = -v_i$ to $v(\infty) = -v_T$.

We can use these equations to contrast the behavior shown in Fig. 3.2, which corresponds to time reversal, with that corresponding to bouncing, in which the velocity changes sign because of the contact force between a ball, say, and a surface. We pointed out that the time-reversed motion in Fig. 3.2 is impossible, since it corresponds to $dE/dt > 0$ in the presence of frictional forces. The rise of the ball after bouncing is a possible motion, and we can see from

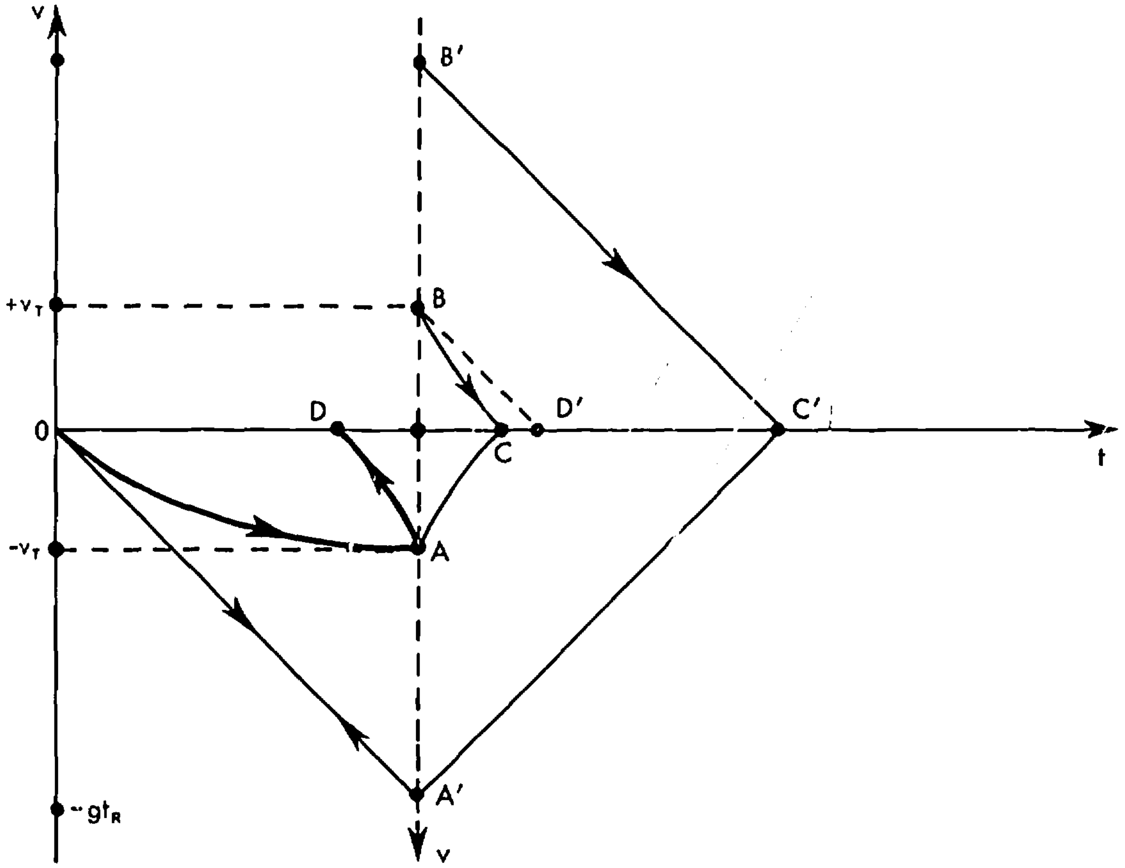

Fig. A1.1 Bouncing ball in retarding medium, with bounce at $t = t_R$. Falling ball: curve OA. Rise after bounce: curve BC. Same rise, viewed on reversed velocity scale (dashed ordinate), with time reversed at $t = t_R$: curve AD. In vacuum (no retardation), corresponding curves are: OA', B'C', and A'O, which retraces OA'.

Fig. A1.1 that it is very different from the impossible motion in Fig. 3.2.

In Fig. A1.1 the curve OA corresponds to the fall preceding the bounce, and is thus the same motion as shown in Fig. 3.2. The bounce occurs at $t = t_R$; we assume that it is perfect, so the velocity goes from $-v_T$ to $+v_T$. The succeeding rise is shown in forward time by curve BC. Notice that the initial (negative), slope of curve BC is twice that of the line BD', which corresponds to a ball rising with initial speed v_T, but in a vacuum (no retarding force). This corresponds to the factor $(v_i + v)/v_T$ mentioned above. Here, this factor equals two, since $v_i = v_T$. To represent this motion, viewed with time reversed at $t = t_R$, as in Fig. 3.2, where both velocity and time axes are reversed at $t = t_R$, the curve BC is first reflected in the horizontal axis, to give curve AC (this corresponds to reversing the velocity axis), and then AC is reflected in the vertical axis, finally giving curve AD (this corresponds to reversing the time axis). Thus the rising motion, after the bounce, viewed with time reversed at $t = t_R$, certainly does not superimpose upon the falling motion, curve OA.

In the absence of frictional dissipation, however, the superposition is perfect. The fall is represented by line OA', with velocity $-gt_R$ becoming

$+gt_R$ after the bounce. Line B'C' is the rise after the bounce with time flowing in the same direction as before the bounce. The reversal of velocity and time at $t = t_R$ now gives C'A' and A'O, respectively, and A'O is an exact superposition on OA'.

These equations for the exponential variation of v(t) and a(t) have several interesting aspects. Notice that according to Eq. (A1.13) the acceleration, which at $t = 0$ is $a = -g$, the free-fall value, decreases exponentially to zero. Hence the terminal speed, v_T, is reached only after an infinite time. However, exponential decrease is quite rapid. For instance, $e^{-3} \approx 0.05$, so according to Eq. (A1.12), when $(g/v_T) = 3$, v will already have reached 95% of v_T; and when $(g/v_T)t = 9$, v will be within about a quarter of one percent of v_T. These statements can be used to give a quantitative meaning to the statement in the text that the particle reaches the terminal speed if it falls "long enough." We notice that g/v_T has the dimension $\sec^{-1}$; so v_T/g has the dimension of time, and can be set equal to a "decay time," denoted τ.

$$\tau \equiv v_T/g. \qquad (A1.20)$$

Now we can repeat the remarks above about the rate at which v approaches v_T, by saying, for instance, that $v/v_T = 0.95$ when $t = 3\tau$.

From Eq. (A1.4), we see that

$$\tau = mg/kg = m/k, \qquad (A1.21)$$

so that the decay rate depends on the mass of the particle and on the coefficient of proportionality, k, between retarding force and speed, but not on the free-fall acceleration. The larger τ is, the slower v approaches v_T. Equation (A1.17) thus shows that the approach to v_T becomes slower (also, of course, v_T increases, as shown by Eq. (A1.4)), as the retarding force coefficient, k, decreases. This would be expected intuitively, but it might not have been so obvious that τ also increases for more massive particles. This effect, and the associated relation, $v_T \propto m$, would not have been at all surprising in the Aristotelian view of motion, according to which heavy objects fall faster than light ones. We can see from our discussion precisely how Galileo's refutation of the Aristotelian point of view depends on idealization to the situation in which resistance due to the medium is absent.

We know that the resistance offered by the medium depends on the size and shape of the object, so we expect that k has this dependence also. Hence we are dealing with an example in which the particle model cannot be expected to be complete. Thus, k should also depend on the medium. The retardation effect of a medium is usually expressed in terms of its <u>viscosity</u>. At this point we need concern ourselves only with its dimensions, namely dyne sec/cm². So if we denote viscosity by η, we have the dimensional equation

$$\eta \doteq \text{dyne sec/cm}^2. \qquad (A1.22)$$

But from Eq. (A1.2)

$$k \doteq \text{force/speed} \doteq \text{dyne sec/cm}. \qquad (A1.23)$$

Hence, comparing Eqs. (A1.22) and (A1.23), we can write

$$k = \eta \ell \qquad (A1.24)$$

where ℓ is a length associated with the object. Inserting this in Equation (A1.21), we see

$$\tau = m/\eta\ell. \qquad (A1.25)$$

Thus the retardation effect, expressed now in terms of τ, has been separated into three factors: m and ℓ, which relate to the object, and η, which depends only on the medium. We note, finally, that the remark that "ℓ is a length associated with the object" cannot be made more explicit without specifying the detailed size and shape of the object. For a particular object, a complicated calculation would be required to find exactly how it is retarded by the medium. The result of the calculation would then give a number with the dimension of length. This number would not necessarily be simply related to the size of the object. For instance, we would expect intuitively that a thin square plate of side a would have a shorter τ (and lower v_T), than a sphere of diameter a, if both had the same mass. That is, the plate has a lower terminal speed, and reaches it sooner, than the sphere. Hence ℓ_{plate} should turn out to be larger than ℓ_{sphere} , for the same a and m.

On the other hand, we can also compare τ for spheres of the same material, but different radius r. We see that since $m = \rho V$, where ρ is the density and V the volume, $V = (4\pi/3)r^3$, we have

$$\tau \propto \rho r^3/\eta \ell_{sphere} \text{ or } \tau \propto \rho r^2/\eta \qquad (A1.26)$$

where we have assumed that ℓ_{sphere} is proportional to r. Hence τ increases rapidly with the size of the sphere.

One final aspect of these equations should now be mentioned. In going from Eq. (A1.1) to Eq. (A1.3), a term was neglected which should be included. The term $-mg$ in Eq. (A1.1) is the weight of the particle in a vacuum; in a medium its apparent weight will be less, due to the buoyant force, so in Eqs. (A1.3) and all those that follow, m should be replaced by m', where

$$m' = m - m_M, \quad m_M = \rho_M V, \qquad (A1.27)$$

where m_M is the mass of the amount of medium displaced by the particle; ρ_M is thus the density of the medium. Hence, in Eq. (A1.26), the density of the particle should be replaced by $\rho' = \rho - \rho_M$.

Appendix 2 FRICTIONAL FORCES AND MECHANICAL ENERGY

We would like to calculate the rate, dE/dt, at which mechanical energy changes as a result of forces applied to a particle, and to show that the rate is always negative for frictional forces. If we write Eq. (3.1) in terms of Cartesian components of the speed, we have

$$K = \tfrac{1}{2}m(v_x{}^2 + v_y{}^2 + v_z{}^2). \qquad (A2.1)$$

This can be written more compactly, with the summation notation, if we introduce numerical indices: $v_x = v_1$, $v_y = v_2$, $v_z = v_3$. Then (A2.1) becomes

$$K = \tfrac{1}{2}m \sum_{i=1}^{3} v_i{}^2 = K(v_i). \qquad (A2.2)$$

We can also write K in terms of the momentum, $\vec{p}$, defined by

$$\vec{p} = m\,\vec{v},$$

so that

$$p^2 = \sum_{i=1}^{3} p_i{}^2 = m^2 \sum_{i=1}^{3} v_i{}^2, \qquad (A2.3)$$

and therefore

$$K(p_i) = (1/2m) \sum_{i=1}^{3} p_i{}^2. \qquad (A2.4)$$

If we also use numerical indices for components of the position vector, or $x = r_1$, $y = r_2$, $z = r_3$, then we can write the dependence of the potential energy V on the position vector $\vec{r}$ as $V(\vec{r}) = V(x,y,z) = V(r_1,r_2,r_3) = V(r_i)$. So we can write for the mechanical energy,

$$E(p_i,r_i) = K(p_i) + V(r_i). \qquad (A2.5)$$

Now E does not depend on t explicitly, so in order to calculate dE/dt we must use the partial derivative chain rule, which takes into account the implicit dependence of E on t through the time dependence of p_i and r_i. We therefore have

$$\frac{dE}{dt} = \sum_{i=1}^{3} \left[\left(\frac{\partial E}{\partial p_i}\right)\left(\frac{dp_i}{dt}\right) + \left(\frac{\partial E}{\partial r_i}\right)\left(\frac{dr_i}{dt}\right)\right]. \qquad (A2.6)$$

According to (A2.5), E depends on p_i only through K, and on r_i only through V; so

$$\partial E/\partial p_i = \partial K/\partial p_i; \quad \partial E/\partial r_i = \partial V/\partial r_i. \qquad (A2.7)$$

From (A2.4), we have

$$\partial K/\partial p_i = (1/2m)\, 2p_i = p_i/m = v_i. \qquad (A2.8)$$

Also, the potential V is related to a conservative force $\vec{F}^{(c)}$ by

$$F_i{}^{(c)} = -\partial V/\partial r_i. \qquad (A2.9)$$

Finally, the term dr_i/dt is simply v_i. We can therefore factor out v_i from each of the terms in A2.6, with the result

$$\frac{dE}{dt} = \sum_{i=1}^{3} v_i \left(\frac{dp_i}{dt} - F_i{}^{(c)}\right) \qquad (A2.10)$$

The total force $\vec{F}$ on the particle may have a nonconservative part $\vec{F}'$ as well as a conservative part; that is,

$$\vec{F} = \vec{F}^{(c)} + \vec{F}' \quad \text{or} \quad \vec{F}^{(c)} = \vec{F} - \vec{F}'$$

But according to N-II, Eq. (A2.8),

$$\vec{F} = m\, d\vec{v}/dt = d\vec{p}/dt, \quad \text{so}$$

$$dp_i/dt - F_i{}^{(c)} = F_i{}'. \qquad (A2.11)$$

Therefore, from (A2.10),

$$dE/dt = \sum_{i=1}^{3} v_i F_i' = \vec{v} \cdot \vec{F}' \qquad (A2.12)$$

where the last equality expresses the result using the vector dot-product. Hence the mechanical energy is constant unless nonconservative forces act on the particle. Frictional forces are nonconservative, and also, are always directed opposite to $\vec{v}$. If we denote frictional forces by $\vec{f}$, as in the text, then

$$dE/dt = \vec{v} \cdot \vec{f} < 0. \qquad (A2.13)$$

Note that this is a result for the total mechanical energy. It is not necessarily true that $dK/dt < 0$ when frictional forces act; that is, frictional forces need not slow down the particles. In orbital motion, for instance, as with artificial earth satellites, when frictional forces act, $dK/dt > 0$ and the particle speeds up. But the orbit corresponding to a higher speed has a smaller radius, so dV/dt decreases. The magnitude of dV/dt is larger than that of dK/dt, so stated by (A2.13), dE/dt is negative.

Appendix 3 ZERO TEMPERATURE ON THE GAS-THERMOMETER SCALE

Despite its modest title, this appendix has a rather ambitious aim. We start with the question: What is the significance of a temperature denoted by $T_g = 0$? Using the analysis of this question, we proceed to the statement of a very general result, the unattainability of "absolute zero," which can be given a precise meaning in terms of our abstract definition of temperature, without necessarily referring to any specific temperature scale. Up to this point, we have been considering macroscopic systems exclusively. Now, using this result, we are able to examine the thermal behavior of certain remarkable atomic systems. We find that these atomic systems possess thermal properties not shared by any macroscopic system. Finally, we are able to show how the concept of temperature can be generalized in a surprising way, so as to describe the unusual thermal behavior of these special atomic systems.

A.1 ABSOLUTE ZERO

The same standard state, namely mutual thermal equilibrium among ice, water, and water vapor, is used in defining the Celsius scale, the gas thermometer scale, and the ideal gas scale. On the Celsius scale, defined by Eq. (4.1), the standard state is assigned the temperature zero, that is, $T_s = 0°$ C. Both the other scales are defined by Eq. (4.5), and since $P(T_s) = P_0$, this gives $T_s = T_0$ °K. In the gas-thermometer scale, the constant T_0 varies with the gas used,[18] while in the ideal gas scale, T_0 has a universal value, independent of the gas used (in the limit of low pressures), and also independent of temperature. In both the gas-thermometer and ideal gas scales, the state for which $T_g = 0$, would correspond to zero measured pressure, in a gas thermometer which originally was filled to pressure P_0 in the standard state.

We will see later that the thermal equilibrium state identified by $T_g = 0$ does have a special significance. However, we cannot deduce anything about this state from the properties of a gas-thermometer scale. The reason is simply that all real gases liquefy when the temperature is low enough, at a finite pressure, $P > 0$. When this happens, the system (homogeneous mass of gas), changes its form, and separates into two phases, a liquid and its vapor. Once the liquid has formed, Eq. (4.5) no longer relates pressure to temperature, since it describes only the properties of a homogeneous mass of gas, and not the behavior of a vapor in equilibrium with its liquid phase.

Now the ideal gas scale does not refer to any real gas, but nevertheless, it is just as impossible to deduce any properties of the state $T_g = 0$ from this scale, as from the gas-thermometer scale. The definition of the ideal scale requires a sequence of gas thermometers filled with a real gas, each using a smaller pressure P_0 in the standard state. But when the temperature is so low that all gases have liquefied, we cannot carry out this limiting process, and so cannot define the ideal gas scale. Hence, there is no meaningful way to assign the temperature $T_g = 0$ to any thermal equilibrium state. Similarly, there is no way to assign a negative value of T_g to any thermal equilibrium state, since we cannot conceive of measuring a negative pressure in a gas thermometer.

These arguments lead to the fol-

---

[18] In general, T_0 also varies with temperature, so the term "constant" might appear to be a misnomer. However, the variation is not large, over temperature ranges which are not extremely wide. Because of this slow variation of T_0 with temperature, it may be regarded as at least approximately constant, especially if high accuracy is not demanded of gas thermometer temperature measurements.

lowing trap, which we must be careful to avoid. We have established a temperature scale on which negative temperatures, $T_g < 0$, are meaningless. On this scale, the state with $T_g = 0$ is therefore colder than any other, since all other possible states have higher temperature. But we have also shown that it is meaningless to assign the temperature $T_g = 0$ to any thermal equilibrium state. We might therefore be tempted to conclude that no such "coldest" thermal equilibrium state can exist. Such an argument would be incorrect, because our argument rests on the properties of the ideal gas scale. Conceivably, there could be other temperature scales, with different properties, for which the argument above breaks down.

Unfortunately, at this point we cannot discuss this question in general, that is, without an explicit connection with a specific temperature scale, such as the gas-thermometer or the ideal gas scale. However, it is possible to state the results of such a discussion. In the first place, we can define what we mean by a "coldest" state, without reference to any particular temperature scale. A coldest state would have the property, that <u>any</u> other state could be reached from it by the dissipation of mechanical energy. It is possible to conclude, from observations which we cannot describe here, that:

> Starting with any thermal equilibrium state of any system, it is impossible to reach such a coldest state, by any means whatever.

This statement is usually called the Third Law of Thermodynamics.

From this statement we can conclude that there is no lower limit to the temperature which can be assigned to attainable states, whatever the temperature scale employed. This fact can be incorporated into any temperature scale, by adjusting its origin so that all attainable states have positive temperature, and the unattainable coldest state has zero temperature. We will call temperature scales which have been adjusted in this way, Third-Law temperature scales. We see, for instance, that T_g is such a scale, while T_c is not. If we wish to discuss temperature on a Third-Law scale without specifying whether it is T_g or some other such scale, we will use the notation T_{III}. Since the identification of the coldest state, $T_{III} = 0$, does not refer to any specific temperature scale, but rather to all scales on which attainable states have positive temperature, the coldest state is often referred to as "absolute zero." The Third Law of Thermodynamics, correspondingly, may be described as asserting that "absolute zero is unattainable."

A.2 NEGATIVE TEMPERATURE

We see that if we had fallen into the trap discussed above, we would have reached the correct conclusion, but for the wrong reasons. We should now examine our earlier conclusion that negative temperatures, $T_g < 0$, are meaningless. If we examine the way in which we defined Third-Law scales, T_{III}, it is easy to so see that states for which $T_{III} < 0$ cannot be colder than the absolute zero state, since "colder than coldest" is self-contradictory. But this does not necessarily mean that no states exist, for which $T_{III} < 0$. It does mean, that like all other states, such states must be <u>hotter</u> than the state at absolute zero, $T_{III} = 0$.

Now it is easy to see that if such "negative T_{III}" states do exist, they cannot be found in every thermal system. In particular, they could not be attained by any system which can be put into thermal equilibrium with a gas thermometer. For if this were possible, the temperature could be measured on the T_g scale, and negative temperatures are certainly impossible on that temperature scale. We are thus led to consider whether thermal systems can exist, such that because of their physical nature they cannot be put into thermal equilibrium with a gas thermometer.

It might seem obvious that the answer to this question is "no." This would indeed be the correct answer if we restricted ourselves to macroscopic systems, that is, to systems which are large enough to be manipulated, or, in other words, systems which can be picked up and moved about. In principle, such manipulation could include arranging thermal contact with a gas thermometer, or at least with some other thermometer which could be compared with a gas thermometer in a second operation. But this sort of manipulation is impossible, even in principle, for systems on the scale of atoms and nuclei. Hence it is on that scale of size that we must seek for thermal systems whose temperature cannot be measured on the T_g scale.

It is by no means obvious that such states, for which T_g cannot be defined, can be found even on the atomic scale. After all, every macroscopic object which we put into thermal equilibrium with a gas thermometer is itself an assemblage of atoms, and in this sense, an "atomic system." It was therefore a remarkable event when, about 1950, Purcell, Ramsey, and Pound reported the existence of an atomic system that had a measurable temperature which could not be directly related to the T_g scale. Their report was based on observations they had made on the interaction of lithium fluoride (LiF) crystals with magnetic fields. This interaction occurs because the nuclei in LiF crystals have magnetic moments. The magnetic field can supply energy directly to the system of nuclear magnetic moments, in much the same way that mechanical energy can be supplied to the crystal as a whole.

The first conclusion that could be drawn from the experiments of Purcell, Ramsey, and Pound, was that the nuclear magnetic moments could be regarded as a thermal system that was in rather poor thermal contact with the crystal as a whole. The temperature of the system of nuclear magnetic moments (we will refer to this system by the symbol N), could be measured magnetically, on a scale we will denote T_m. The temperature of the crystal itself (we will refer to this system by the letter L), could of course be measured on T_g, since L is a macroscopic system. Thermal contact between systems N and L was investigated by experiments which can be described very roughly as follows: In the first stage, a crystal of LiF is in thermal isolation from its surroundings, with no magnetic field applied. According to our basic definition of thermal equilibrium, systems N and L must arrive at a state of mutual thermal equilibrium. This state will be recognized by the fact that $T_m(N_e)$ and $T_g(L_e)$, which are respectively the temperature of the nuclear magnetic moment systems, measured on the magnetic scale, and the temperature of the crystal, measured on the gas-thermometer scale, are both constant. The subscripts e on the symbols for the system emphasize the fact that this is an equilibrium state. Now the magnetic field is applied, and the systems are momentarily in the second stage of the experiment. The temperature of the system N immediately jumps to a new value, $T_m(N_2)$, while the temperature of system L is unchanged, since no mechanical energy was dissipated in that system. The momentary second stage is followed by the third stage, which takes several minutes. During this stage, the temperatures of both systems change slowly, showing that they are in thermal contact; but since the changes take place slowly, we recognize that the thermal contact is poor. Finally, the fourth and last stage of the experiment is reached, in which both systems have reached new constant temperature, $T_m(N_e')$ and $T_g(L_e')$, showing that they are once again in mutual thermal equilibrium. However, as we would expect, the new equilibrium states, denoted N_e' and L_e', are different from those in the first stage, N_e and L_e.

The series of stages in this experiment is exactly analogous to observations that could be made with the arrangement of Fig. 4.1, using macroscopic systems exclusively. If the plate on top of the bell jar is made of wood, a system on the plate would be in poor thermal contact with the pendulum bob inside the jar. Applying the magnetic field to the crystal would correspond to suddenly changing the temperature of the system on the wooden plate. Because of the poor thermal contact there would be no corresponding jump in the temperature of the pendulum bob, which is the analog of the crystal in the LiF experiment. The third and fourth steps of the experiment would proceed just as above, with the temperatures of the system on the plate, and of the pendulum bob, both changing and ultimately reaching a new thermal equilibrium.

Now in either case, the LiF experiment or its completely macroscopic analog, we can easily deduce which of the initial systems was hotter in the momentary stage two, by simply comparing the initial and final equilibrium temperatures of just one of the systems, namely the crystal in the LiF experiment, or the pendulum bob in the macroscopic analog. If $T_g(L_e') > T_g(L_e)$, for instance, then the crystal was warmed by its contact with the system of nuclear magnetic moments, so system N must have been heated by the application of the magnetic field, and was therefore hotter than L during the momentary stage two and thereafter, until the final equilibrium was reached. Now with the macroscopic analog experiment, this sort of discussion would be unnecessary, since temperatures of both the system on the plate and of the pendulum bob can be measured on the scale $T = T_g$ and thus compared directly. In the LiF experiment, on the other hand, the temperature of system N can be measured directly only on scale T_m. Now in stages one and four of the experiment, systems N and L are in thermal equilibrium, so $T_m(N_e)$ is equal to $T_g(L_e)$ and $T_m(N_e')$ is equal to $T_g(L_e')$.[19] However, in stage two of the LiF experiment, system N is not in thermal equilibrium with system L, so the comparison between $T_g(L_e)$ and $T_g(L_e')$ is the only way of determining whether $T_m(N_2)$ is hotter or colder than $T_g(L_e)$.

Now as long as we are dealing with macroscopic systems, we would expect to find that however large the initial increase of temperature of the system on the plate in stage two of the experiment, we can always prepare another system that is even hotter. For instance, if the system on the plate has temperature, say, $(T_g)_2$ in stage two, we can prepare another system with temperature $(T_g)_2 + 1$. This other system, since it is hotter than the system on the plate, will have its temperature lowered if it is brought into contact with the system on the plate. That is, there is no "hottest" temperature on the T_g scale.

In remarkable contrast to this situation, it was found in experiments on LiF crystals that, by suitable interactions with magnetic fields, it was possible to prepare system N of nuclear magnetic moments in states which were hotter than *any* state of a macroscopic system. That is, it could be shown that special states, which we will denote $N_2{}^*$, could be prepared which had the property that *any* macroscopic system, whatever its thermal equilibrium state, would be warmed on being brought into thermal contact with that state. Of course, this conclusion could only be reached indirectly, since it is clearly impossible to carry out experiments on an unlimited number of macroscopic systems in

---

[19] The equality of these two pairs of temperatures does not necessarily mean that they have the same numerical value, only that they are assigned to thermal equilibrium states in mutual equilibrium. In the same sense, the temperature 32°F is "equal" to the temperature 0°C, although the numerical values on the two scales are different. The reason is that both temperatures refer to the same thermal equilibrium state, namely the common standard state on the Celsius, gas-thermometer, and ideal gas scales.

a finite length of time. Nevertheless, this was an inescapable conclusion from the experiments which were performed.

Another way of stating this conclusion is to say that such a special state N_2* is hotter than any thermal equilibrium state of any macroscopic system. The question arises, whether it is possible to assign to such a state a temperature $T_{III}(N_2*)$ on a Third-Law temperature scale. Since $T_{III} = 0$ defines the unattainable coldest state, and positive T_{III} corresponds to ordinary thermal equilibrium states of macroscopic systems, we have available only negative temperatures, $T_{III} < 0$. But now there is no bar to using negative temperatures, as there was on the T_g scale, since as we have seen, a state like N_2* is not in thermal equilibrium with any thermal equilibrium state that has a temperature on the T_g scale. So we have finally arrived at a meaning for negative temperatures on a Third-Law scale, at least for certain special atomic systems like that of the nuclear magnetic moments in a LiF crystal. They correspond to "superhot" states which are hotter than any states with positive temperature on a T_{III} scale.

One question remains - how shall we assign numerical values to the temperatures of two different "superhot" states? A consistent convention is to associate "hotter" with larger algebraic value of the temperature, rather than larger magnitude. Thus, for instance, a state with $T_{III} = -100$°K is hotter than the state with $T_{III} = -1000$°K. We can now arrange all states, both "normal," corresponding to macroscopic systems, and "superhot," in one consistent order. Starting with the (unattainable) coldest state with $T_{III} = 0$, increasingly hotter states have larger and larger temperature values, which approach the limit $T_{III} = +\infty$. States hotter than this limit are "superhot"; their temperatures start from the limit $T_{III} = -\infty$, and increasingly hotter "superhot" states have temperatures which increase in algebraic value, approaching $T_{III} = 0$, but from below. This limit is the "absolutely hottest" state; like the "absolutely coldest" state, it is unattainable. There are now two states corresponding to "absolute zero," depending on whether the approach is from positive or negative temperatures; the Third Law states that neither of these states is attainable.

The convention we have described for handling negative Third-Law temperatures is the only one that leaves ordinary temperatures T_g unchanged. It is rather clumsy, however, since the same state is represented by either $T_{III} = +\infty$ or $T_{III} = -\infty$, and also $T_{III} = 0$ represents two vastly different states, depending on the direction in which it is approached. A simple modification of the scale produces a drastic change in conventional temperature scales, but it eliminates these difficulties. We introduce a scale denoted R_{III}, defined by the negative reciprocal of T_{III},

$$R_{III} = -(1/T_{III}).$$

The relation between R_{III} and T_{III} is shown in Fig. A3.1. The curves in the figure are hyperbolas, according to the definition of R_{III}, and the arrows on the hyperbolas indicate the direc-

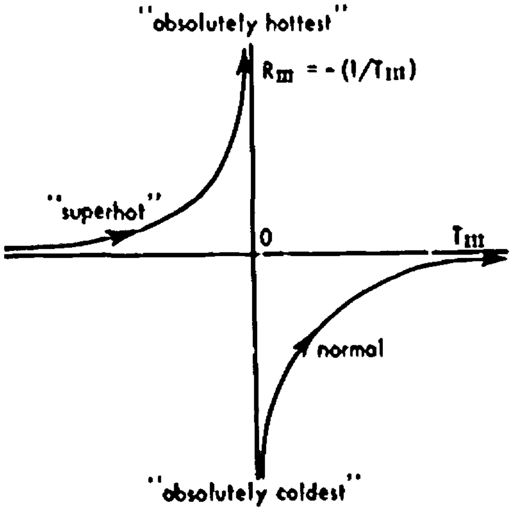

Fig A3.1 Temperatures on the scale R_{III} plotted as a function of temperatures on the scale T_{III}.

tion corresponding to "hotter." On the scale R_{III}, the "absolutely coldest" state has $R_{III} = -\infty$, and the hottest normal state is $R_{III} = 0$, approached from below, so all normal temperatures, appropriate to macroscopic systems, are negative. All "superhot" temperatures are positive, the coldest of these being $R_{III} = 0$, approached from above. The "absolutely hottest" state is $R_{III} = +\infty$.

It is not likely that the R_{III} scale, which clarifies the relation between normal and "superhot" temperatures, will ever be commonly used. However, since the original experiments on LiF were performed, many other atomic systems have been found which can also attain "superhot" states. Furthermore, such systems are no longer merely laboratory curiosities. Their "superhot," or, as they are commonly called, "negative temperature" (on a T_{III} scale), states make possible the operation of masers for amplification of minute microwave signals, and also lasers, which produce intense beams of coherent light.

Appendix 4 INTERCOMPARISON OF TEMPERATURE SCALES

As we have pointed out, every system which is chosen to serve as a thermometer, and indeed, every thermometric property of such a system establishes its own temperature scale. There are a great many thermometric properties in general use, besides the liquid-in-glass thermometers we have described. A common type is based on the properties of a coiled strip, made from strips of two different metals welded together. The two metals are chosen so that they expand at different rates when they are heated. Then a coil made from such a bimetallic strip will rotate about a fixed point to which one end is attached so that it cannot move. The angle of rotation of a pointer attached to the other end of the strip then serves as a thermometric property. Oven thermometers are usually of this kind.

The electrical resistivity of metals, alloys, and semiconductors varies with temperature, and so can be used as a thermometric property. If, in a loop made of wires of two different metals or alloys, the two junctions are at different temperatures, an electrical voltage is developed in the loop, and the magnitude of this voltage can serve as a thermometric property.

Now this profusion of temperature scales leads to difficulties, whenever two temperatures are to be compared, which have been measured using different thermometers, say A and B. If the scale appropriate to each is used, then it can happen that two thermal equilibrium states can have the same numerical value of the temperature, $T_A = T_B$, but still not be in mutual thermal equilibrium. Or it can happen that one thermal equilibrium state can have two different numerical values of temperature, $T_A \neq T_B$, when measured with two different thermometers.

These situations can arise as follows: Suppose, to take a concrete case, we consider a mercury-in-glass thermometer (A), and a platinum resistance thermometer (B). When both are in equilibrium in the standard state of the Celsius scale (0°C), the length of mercury in the capillary is L(0°C) cm, and the resistance of the platinum wire is R(0°C) ohm. When both are in equilibrium with water boiling at standard atmospheric pressure (P = 760 mm of Hg), the corresponding values are L(100°C) cm and R(100°C) ohm. Now suppose that, in equilibrium with a third state, we observe that the length of mercury in the capillary has expanded to 0.40 of the distance from L(0°C) to L(100°C). That is

$$L(T_A) = L(0°C) + 0.40[L(100°C) - L(0°C)] \qquad (A4.1)$$

We would be entitled to denote this temperature as $T_A = 40°(A)$, since it has been determined from thermometer (A), mercury-in-glass. Now in equilibrium with the *same* thermal equilibrium state, we might find that

$$R(T_B) = R(0°C) + 0.41[R(100°C) - R(0°C)], \qquad (A4.2)$$

so that we would assign to this state, $T_B = 41°(B)$, corresponding to thermometer (B), platinum resistance. The reason for the discrepancy is that the density of mercury between 0°C and 100°C does not vary in exactly the same way as the resistance of platinum, between the same two temperatures.

In order to avoid the confusion about numerical values of temperature expressed in Eqs. (A4.1) and (A4.2), and its further compounding when other thermometers are considered as well, careful measurements with gas-thermometers in standard laboratories all over the world are used to establish a scale which is adopted as standard by international agreement. In terms of this international Celsius scale, measurements with any chosen thermometer can be compared with those of any other. For instance, it might be found that the state described above, for which $T_A = 40°(A)$ and $T_B = 41°(B)$, has the temperature $T_C = 40.5°C$ on the international Celsius scale. The point on the capillary, which is described by Eq. (A4.1) would not be marked "40," but rather, "40.5," and we would say that this thermometer has been calibrated against the international Celsius scale. After this calibration procedure has been carried out for any particular thermometer, that thermometer can be used to measure temperature on the Celsius scale, even though this standard scale, and that defined by the particular thermometer chosen, will in general be different.

Appendix 5 INTERNAL DEGREES OF FREEDOM AND THE ATOMIC MODEL

The amount of detail that we have gone through in this discussion leading to the conclusion that matter possesses internal degrees of freedom may appear strange to the average reader, who has undoubtedly been raised to accept the existence of atoms and molecules as an article of faith. He or she is therefore perfectly well aware that the force F_u introduced in section 6.1 "obviously" comes from random collisions of the particle with the atoms or molecules of the fluid in which it is in suspension. We certainly do not wish to challenge this faith in the existence of atoms and molecules, but we do wish to examine it rather more closely than is sometimes done. Our argument was designed to show that the necessity for postulating an internal structure follows directly from observations on purely macroscopic behavior. For such a conclusion to be convincing, the discussion on which it was based could not rely on any assumptions about internal structure; otherwise the reasoning would be circular. Furthermore, the characteristics of the internal structure can also be deduced from macroscopic observations. Some aspects of this deduction are described in sections 6.1 and 6.2.

After all, belief in the existence of atoms is not exactly a novelty. It dates back well over two millenia; it had its adherents among the classic Greeks and Romans. But during this ancient period (which lasted, in fact, well into the eighteenth century), this belief had no real foundation in the known behavior of nature. It was in the realm of metaphysics rather than physics. And it does not leave that realm simply by becoming absorbed into everyday thinking if there is little conception of how it is related to the understanding of natural phenomena.

As a matter of fact, the naive version of atomic theory that was developed during the eighteenth and nineteenth centuries, in which the atoms and molecules are simply Newtonian particles, is inadequate for understanding many kinds of phenomena which it should be able to explain. The essential distinction between the macroscopic and atomic levels of size was made at the beginning of section 5.2. It was pointed out there that the macroscopic degrees of freedom are always adequately described by Newtonian mechanics, while the degrees of freedom associated with the internal structure of matter must be described by quantum mechanics. The nature of this distinction was clearly realized only about 1925 when quantum mechanics was developed in two independent but equivalent formulations. Before then, attempts to apply Newtonian mechanics to phenomena on the level of atomic size required various revisions of classical mechanics for different problems such as atomic spectra, atomic collisions, nuclear transmutations, electrical properties of solids, and these revisions sometimes appeared to have little connection with one another. Many physicists who had been trained in the nineteenth century, before much attention had been paid to these problems, therefore had little faith in the validity of the naive atomic theory. Its long history as essentially speculative philosophy, largely divorced from a relation to natural phenomena, tended to reduce what little faith they had. Their attitude towards attempts to give it a firm connection with observation was that a model was being devised which was being diverted from the essential behavior of matter by trivialities. They could point, as an example, to Aristotle's model for motion which had been led astray by regarding friction as the essential feature of motion, rather than equilibrium, with constant velocity. This attitude was well expressed by Mach, who wrote (in 1915), "I do not consider the Newtonian Principles as a completed and perfected thing, yet in my old age, I can accept the theory of relativity just as little as I can accept the existence of atoms and other such dogma." In a certain sense, Mach's attitude was correct, since the naive atomic theory of which he was writing was certainly wrong. But in another sense, the judgment of Mach and the others who shared his point of view, many of whom had made great contributions to physics, was bad. What they regarded as one of the blind alleys we mentioned in section 1, turned out to be a broad avenue whose exploration led to the most coherent and encompassing picture of nature ever devised. But in one ironical sense, the doubters were right, after all. The difference between the naive atomic theory, which Mach rejected, and quantum mechanics, is certainly not trivial. The quantum-mechanical behavior of the internal degrees of freedom of matter can be very unlike that of Newtonian mass-points. For instance, we may inquire as to the location of an electron in an atom. The model of which Mach was complaining would answer in terms of a point charge rotating about a nucleus. In contrast, quantum mechanics will speak of probabilities, rather than a well-defined location. And the probabilities are given in terms of wave functions, which have finite values everywhere, not just at a geometrical point. So quantum mechanics agrees with Mach, in rejecting the mass-point as a suitable model, in general, for the internal degrees of freedom.

As a result, it is not really accurate to speak of the ultimate triumph of the atomic theory, more than two thousand years after its first pronouncement. The long dispute between those who asserted that matter was made up of a continuum, and those who asserted that it had an internal structure of indivisible parts separated by vacuum, seems to have been resolved in favor of the latter point of view. But the atoms of the structure we now recognize bear little resemblance to those described by Lucretius: "Bodies of absolute and everlasting solidity . . . absolutely solid and unalloyed." And the void between atoms is far from featureless: it is the seat of gravitational, electric, and magnetic fields. Therefore, while quantum mechanics is a descendant of primitive atomic theory, it is better described as a remarkably successful fusion of the atomic and continuum points of view.

It should not be imagined, however, that quantum mechanics is competent to answer all questions about the internal degrees of freedom of matter. The structure of nuclei is still an unsolved problem, and even some phenomena on the atomic level, which can be described quite simply, have not been accounted for in detail. One of the outstanding examples is the existence of the different phases of matter. A collection of atoms or molecules can transform from a gas, which fills the volume of its container, to a liquid, which has a definite volume at a fixed temperature, but whose shape is still determined by its container, to a solid, which has a definite size and shape, in which the atoms are arranged in a regular three-dimensional pattern. The processes by which these transformations occur are still imperfectly understood.