Special Topics in Transport Phenomena

by Giovanni Astarita, Raffaella Ocone

- ISBN: 0444829970
- Publisher: Elsevier Science & Technology Books
- Pub. Date: January 2002

Preface

This monograph comes out of a long history of two things: on the one side, teaching a course on transport phenomena, which one of us (GA) has done for a longer time than he cares to confess; on the other side, working together, which the other one of us probably thinks has gone on for too long already, though the time involved is significantly less than what was being talked about before.

Teaching transport phenomena is a strange experience. There is so much conceptual content in the subject that one has no hope whatsoever of covering any reasonable fraction of it in a two-semester course; and yet, that's exactly what one is called upon to do. There is a redeeming feature, however: a textbook which is so obviously a classic, its contents so obviously what one is expected to teach to the students, that in a way the task is made easy by telling the students, on the very first day of class, that they will eventually be expected to have mastered Bird, Stewart and Lightfoot — BSL. At least, they know exactly what is expected of them. This does make things easier, or at least not so overwhelmingly difficult that one would give up in despair, but still, let's face it: who has ever covered BSL in two semesters?

However, one goes on and tries to teach something. The content of BSL is just too much, so one tries to extract concepts and concentrate on those, the details being left to BSL. One of us has done that for too long, and has found out after a while that he wasn't teaching BSL at all — he was teaching concepts which the students found challenging, though they had difficulty tracking them down in BSL. And yet, one still expects students to have grasped BSL, doesn't one? And on the other hand, one becomes quite fascinated with the concepts — or the special topics, or the ideas — one has been concentrating upon, and one doesn't know exactly what is going on. Unless one is lucky (as one of us has been) in having among his students the other one of us — a student who seems to have no problem mastering BSL on her own, and finds the discussion of special topics, concepts, or whatever, a challenging experience in the classroom.

Well, how can all this be translated into a monograph, if at all? When one writes a book, one has an audience in mind. But this turns out to be the ideal situation if one considers a monograph whose audience is intended to be the population of graduate students in chemical engineering. This is just the audience one wants: an audience who has mastered BSL already, so that one can concentrate on what one thinks, perhaps wrongly, to be rather interesting special topics, challenging conceptual issues, things which are, let's face it, just shear fun. Fun to teach, fun to think about, fun to write about - without (thank God for BSL) having to worry about all the nitty gritty details. Or, for that matter, about the concepts at the first level of conceptual development, or even the second or the third one. Isn't it rather nice to be able to concentrate on very special cases? Particularly if one can try to show that they are not that special after all?

Now when we teach, not abstractly, but concretely in the classroom, we have a few problems of a down to earth nature. When we need to do some algebra, we have two choices. One, we get it all done on transparencies so there are no sign mistakes and everything runs smoothly. Well, we can do that, and by the time our collection of transparencies is large enough, we think of writing a book which is supposed to be in competition with BSL. People have done that, and it has been a mistake — you can't beat BSL at their own game. The other choice is to be cavalier about algebra, miss the signs, and ask students to work it out themselves at home whenever a problem arises — that works beautifully, but one doesn't get a collection of slides to stimulate one to write a book.

There is another thing about actual teaching in the classroom: one interacts with students. One says something and a student objects to it. Now 90% of what students say in class is just meant to make themselves look like conscientious students and that is entirely irrelevant. But the remaining 10% — that's what makes teaching in the classroom worthwhile. If it weren't for that, one might as well bury BSL in a computer data bank, require students to access it and work out all the problems, and be done with it. Should anyone still think there's some use in an instructor actually walking up to the lectern and saying something, one could choose the best one to do so, videocassette him or her, and put the videocassette in the computer data bank as well. The whole point about teaching is the interaction with students; and so we instructors, short of declaring bankruptcy, have to maintain that such interaction is absolutely crucial. Now we have already admitted that such interaction may have any meaning only in 10% of the cases — can we say anything more?

Perhaps yes, and we have tried to do so in this monograph by introducing three characters: Sue, Ralph, and Bob. Bob is an engineer, a no-nonsense type, one who knows life and isn't going to be taken in by any too-sophisticated argument. He keeps our feet on the ground; his observations are frustrating at times, particularly when one is trying to discuss a subtle conceptual point, but he compels us to always point out the practical significance of what we are discussing, and that is healthy. Ralph is the typical good student. He never makes a mistake in algebra, he knows all the formulas in the books, he remembers everything. He is perhaps the most frustrating of the useful students — getting him to accept a new viewpoint is very, very hard indeed. He knows BSL by heart, and he enjoys those problems which are marked with a subscript of 5, very difficult. He knows he can work those out much better than anybody else in the class. Ralph keeps us from being sloppy — or at least he tries very hard to do so. Our third student, Sue, is — let's face it — the one we prefer, though we would never tell her. She has just slipped by her undergraduate work, because she didn't work very hard. Should she do any algebra, she is guaranteed to make a terrible mess of it. She is very fuzzy about the content of the books she has studied in her undergraduate days — just the opposite of Ralph. That makes it easy for her to accept new viewpoints: she hasn't mastered the old ones well enough to cherish them. And she's smart. She is, in fact, a totally imaginary character: we all hope that a student will ask the question we want to be asked, and it never happens. Sue is our dream of a student, who always asks that question. If one only could get a Sue in one's class in real life, everything would just be absolutely wonderful.

Now why Bob, Ralph and Sue? Because about a third of the students in chemical engineering nowadays are female. Half of the authors of this monograph is female, but with only two authors it's hard to match the one-third ratio exactly, and half is the closest approximation one can obtain. We do, however, have a great advantage: we don't need to worry about our grammar being considered sexist. We use "he" and "she" interchangeably, without worrying about percentages; we never feel the need to use him/her, which we find annoying in the extreme.

We have talked about BSL earlier. We assume our reader has studied at least some parts of it. What else do we assume our reader has studied? Well, let's first answer an easier question: what else do we hope our reader has, well, not studied, but at least leafed through. We hope he has read a classic which unfortunately isn't read much at all, "Dimensional Analysis" by P.W. Bridgman, Yale University Press, 1922. She would also have done well reading

"Diffusion and Heat Exchange in Chemical Kinetics" by D.A. Frank-Kamenetskii, Princeton University Press, 1955 (we do not even hope our reader knows Russian), and "Physicochemical Hydrodynamics" by B. Levich, Prentice-Hall, Englewood Cliffs 1962. "Process Fluid Mechanics", same Publisher, 1980, and "Process Modeling", Longman 1986, by M.M. Denn would also be on our list of favourites (Mort Denn is great — he slips in the word "process" no matter what he is writing a book about. We haven't been able to find a reasonable way of doing the same ourselves). And, since we are dreaming, "Rational Thermodynamics", Springer-Verlag 1984, by C.A. Truesdell wouldn't hurt by any means. Now that's what we would hope our reader has read, but not what we'll assume she has in fact read — except BSL. That, dear reader, you should have studied, and if you haven't, well, it's just tough luck.

In a preface, one is supposed to acknowledge help received, isn't one? Help has been received, mainly by students. Clever ones, who asked challenging questions; less clever ones, who had the courage to say they hadn't quite understood what was going on; some students who worked on some special problem for their senior thesis in Naples, and were able to prove that the simple problem intended for a senior thesis was a difficult one worth a higher level thesis; some students who worked for a PhD thesis, and were able to show that the sophisticated problem given to them was really trivial, and managed to get a PhD because they thought of their own problem afterwards; students of the Bob type, who kept our feet on the ground; of the Ralph type, who obliged us to do our algebra correctly; and wouldn't it be nice in real life to be able to thank also a student of the Sue type?

There is another category of acknowledgements we need to make. We started on this project because courses titled something like "Special Topics in Transport Phenomena" are commonplace in many chemical engineering departments. So, at the very beginning, we wrote to friends in a large number of such departments asking for their advice on what should go into a book with such a title. We received a large number of very thoughtful replies, and these have been taken into account. The number of people who replied is too large for a list to be given here, but our sincere thanks go to all entries in this non-existent list.

Naples, Italy, and Nottingham, UK

Foreword

The publication of this monograph has taken longer than expected, even though it was almost in its final version a couple of years ago. Many events have happened in the meantime, and there was a time when we almost gave up on its publication. It has been only after the untimely decease of Gianni that I started to think again about it.

As the reader can see from the Preface, Gianni taught transport phenomena for many years and — as he himself used to admit — he was teaching concepts instead of details; students could find the details in the Bird, Stewart and Lightfoot book — very hard work indeed. Having been a student of his myself a long time ago, I found Gianni's lectures extremely difficult, but amazingly challenging: I always wanted to attend his lectures again after graduation, just to enjoy them without the fear of eventually having to pass the exam. Of course I never managed to do that, and when a few years later I was team-teaching with him, I appreciated that his teaching habits were still unchanged — and the feelings of the students too. This is why I was eventually convinced that this monograph deserves to be published: it collects somehow Gianni's way of teaching and some of the concepts he tried to transfer to students.

Chapter 6 was re-written later. In our original project we meant to furnish very general concepts about the hydrodynamics of granular materials. At that time the field was not yet so popular, and basic concepts were still in the development stage. When I reconsidered the possibility of publishing this monograph, I realised that many very good books on the subject had come out in the meantime (such as, for instance, the book by J.P.K. Seville, U. Tüzün and R. Clift, "Processing of Particulate Solids", Blackie Academic & Professional, 1997; and the one by L.-S. Fan and C. Zhu, "Principles of Gas–Solid Flow", Cambridge University Press, 1998). Therefore, there was no reason to repeat in one chapter what was already available and published in a much more complete way. Chapter 6 has been completely re-written with the help of Tommaso Astarita, and the reader will find it somewhat different in style from the other chapters. Chapter 6 now deals with new results obtained by ourselves on compressible flow of granular materials, a very challenging topic which is still at its exordium, and which looks very promising in new developments. A flavour of "granular" thermodynamics is given too.

A number of people must be acknowledged for their help in writing Chapter 6. Many thanks go to Tommaso for contributing most of the results on compressible flow of granular materials, and for sketching most of the figures. I would like to thank Renee Boerefijn for providing useful comments, and Yvonne Campbell for typing most of Chapter 6 quickly and professionally.

Many thanks go to Samir Khan for sketching the majority of the graphs.

Finally, a special thank you goes to Ari Kummer for his invaluable help and support during the latest part of the work, which is always the most difficult and demanding.

Raffaella Ocone
Edinburgh, 2001

Table of Contents

Chapter 1

Introduction to Methodology

1.1 INTRODUCTION

On the first reading of this chapter, the reader will probably get the impression that we do not follow an organized line of thought—we wander here, there, and everywhere. Well, in a sense we do, but there is a hidden line of thought, which hopefully will become apparent to our readers on the second reading. This is meant to be a provocative chapter, one which hopefully will provide ample food for thought, and we hope our readers get quite mad with us no later than halfway through. We also hope they'll forgive us by the end.

What is the scope of the chapter? We analyze several apparently unrelated problems, and all of them are formulated in the simplest possible form (don't worry, there are complexities aplenty), in fact in such a simple form that in most cases an analytic solution to the governing equations can be obtained. The purpose is to extract from these simple problems some lessons of presumably general applicability: the reader is asked not to skip the remarks, which may mostly appear trivial in connection with the specific problems considered, but will later be seen to be anything but trivial. We also try, on more than one occasion, to find out something about the solution of a problem without actually solving it, or to find an approximate solution. This may appear futile when an exact analytical solution is available, but it is meant to pave the way for doing that kind of thing when an exact solution is not available. For the problems at hand, the powerfulness of these techniques is reinforced by comparison with the exact solution.

A very important aim of this Chapter is to convince the reader that the subjects of Transport Phenomena and of Thermodynamics are not mutually exclusive ones; in fact, they are strongly intertwined, much more so than is usually thought. The usual attitude in this regard is to think that Thermodynamics tells us what the equilibrium conditions are; if the actual conditions are not equilibrium ones, knowledge of thermodynamics allows us to establish what the driving force for a transport process may be, and that is all the coupling between the two subjects that there might be. Well, the situation is significantly more complex than that. This is an important conceptual point, and it becomes more clear if the analysis is uncluttered by irrelevant complexities. This is the main reason for formulating all the problems in the chapter in their simplest possible form which leaves the conceptual content still there.

Now, is there some kind of simplification which we might do once and for all? Indeed there is: a geometrical one. It so happens that the space we live in, which we may perhaps be willing to regard as an Euclidean one (an assumption which certainly makes life easier) is, however, no doubt endowed with an embarrassingly large number of dimensions: three. Following the old rule of thumb that there really are only three numbers, 0, 1 and ∞ (perhaps only 0 and ∞ in fluid mechanics, we consider high Reynolds number flows, low Reynolds number flows, and an instructor who deals with the case where the Reynolds number is about unity is regarded as fussy in the extreme — whoever remembers what Oseen contributed to fluid mechanics?) we come to the conclusion that our space is awkwardly close to having infinitely many dimensions. Do we really need to bother about this as soon as we begin? Perhaps we may avoid the issue. Geometry has a nasty tendency to make things complex even when they are not; or, if things are conceptually complex, geometry tends to mask this interesting kind of complexity with the trivial one of having three dimensions to worry about: just as the complex game of chess. Perhaps we can stick to the deceptively simple game of checkers, without, as Edgar Allan Poe rightly observed, losing any of the conceptual subtleties, in fact keeping them in the sharp relief they deserve: we may stick, for the time being, to spatially one-dimensional problems, where all quantities of interest are functions of at most one spatial variable, say X. However, we do not want to degenerate to the silly simplicity of tick-tack-toe, and so we will not make the steady state assumption, so that time t is, in addition to X, an independent variable.

All problems in engineering science are formulated on the basis of two types of equations: *balance* equations and *constitutive* equations. A balance equation can be written either for a quantity for which a general principle of conservation exists (such as mass, linear momentum, angular momentum, energy, etc.), or for a quantity for which no such principle exists (like entropy, or the mass of one particular component in a reacting mixture), provided its rate of generation is included in the balance equation. We begin by considering the former case.

Let $F(X,t)$ be the flux of the quantity considered, i.e., the amount of it crossing the surface orthogonal to X at X per unit time, and let $C(X,t)$ be the concentration, i.e., the amount of the quantity considered per unit volume. The unsteady state balance equation takes the form:

$$\partial F/\partial X + \partial C/\partial t = 0 \tag{1.1.1}$$

Now there are several subtleties with Eq. (1.1.1) and some of these will be discussed later on. However, for the time being we are happy with it as it stands, and we ask ourselves the following question.

Suppose the system is initially at equilibrium, with $C = 0$ everywhere (if it has some value other than 0, it can always be set to zero by normalization).

Remark 1.1.1

Normalizing C to an initial value of zero is more than simply trying to make the algebra a little less cumbersome. When a problem is essentially linear, we want to keep it that way; now, mathematically a problem is linear if boundary conditions are homogeneous (a linear combination of the dependent variable and its derivatives is zero). This can be accomplished, for sufficiently simple problems, by appropriate normalization.

Correspondingly, $F=0$ initially:

$$t < 0, \ \ X > 0, \ \ F = C = 0 \tag{1.1.2}$$

At time zero a jump of C is imposed at $X=0$, say:

$$t > 0, \ \ X = 0, \ \ C = J \tag{1.1.3}$$

where J is the imposed jump. The question which we ask is: what are the functions $C(X,t)$ and $F(X,t)$ at $X > 0$, $t > 0$?

Remark 1.1.2

This shows that we are focusing on the propagation of an imposed jump. It will be seen that this is in fact a very interesting problem, and indeed one can immediately ask oneself an interesting question: under what conditions will the imposed jump propagate as such (i.e., staying a jump?).

And under which conditions will it, if it does stay a jump, decay in amplitude?

It is perhaps obvious that Eqs. (1.1.1–3) do not give enough information to answer this question. What is needed is a *constitutive equation*: an equation which assigns the value of $F(X,t)$ in terms of $C(X,t)$, so that the problem becomes a mathematically well posed one.

However, before discussing constitutive equations, it is useful to define a dimensionless concentration $c = C/J$, so that Eqs. (1.1.1–3) become:

$$\partial F/\partial X + J\partial c/\partial t = 0 \tag{1.1.4}$$

$$t < 0,\ X > 0,\ F = c = 0 \tag{1.1.5}$$

$$t > 0,\ X = 0,\ c = 1 \tag{1.1.6}$$

Remark 1.1.3

Since initially both F and C are zero, why did we choose to consider the case where a jump of C is imposed, rather than a jump of F? Mainly because of tradition. Indeed, it will often be useful to consider what may be called the dual problem, where a jump Q of the flux F is imposed. In that case, a dimensionless flux $f = F/Q$ comes to mind straightaway, but there is no immediate concentration scale available to define a dimensionless concentration.

1.2 THE CLASSICAL PLUG FLOW REACTOR

Suppose we have a plug flow reactor (PFR) without knowing it is a PFR (we are allowed to be silly that early in the game), and we wish to determine its Residence Time Distribution by measuring the response to some forcing function on the feed. We choose a step forcing function. Specifically, the PFR is fed with a steady state mass flow-rate of pure water, and at time zero we switch the feed to one of water containing a concentration J of ink. We monitor the transparency of the exit stream as a function of time; we are willing to assume that the transparency, when appropriately normalized, is proportional to $1–c(L,t)$, where L is the axial length of the PFR. (Of course, this problem is so degenerately simple that we know the solution even before beginning; the response is simply a step function, delayed by an amount equal to the residence time L/U, where U is the axial velocity.)

We first model our system by neglecting diffusion in the axial direction (this is what we mean by the adjective "classical"). It follows that F is the convective flux, say:

$$F = UC = JUc \tag{1.2.1}$$

Equation (1.2.1) is, quite obviously, a *constitutive* equation. In writing it, we do not in any sense claim its general validity: we are simply writing the simplest possible form of constitutive equation for the flux which doesn't leave us any problem to deal with. We are neglecting diffusion, as well as other phenomena which may very well play a role.

It is now natural to define a dimensionless flux f:

$$f = F/JU = c \tag{1.2.2}$$

so that Eq. (1.1.4) becomes:

$$U\partial f/\partial X + \partial c/\partial t = 0 \tag{1.2.3}$$

Remark 1.2.1

> The temptation to substitute $f = c$ into Eq. (1.2.3) is overwhelming, but before giving in to it we notice that the system of balance equation (in the present case, Eq. (1.1.4)) and constitutive equation (in the present case, Eq. (1.2.1)) is generally in what is called "canonical form". Now, in general one may eliminate the flux by substitution of the latter in the former (or, in more complex cases, we may eliminate the flux between the two equations by a procedure more complex than direct substitution). This produces a differential equation which in general is not in canonical form. There is a variety of problems for which qualitative analysis, approximate analytical solutions, and numerical solution are easier if the canonical form is retained.

Having made this point, we now do substitute (1.2.2) into (1.2.3) to obtain:

$$U\partial c/\partial X + \partial c/\partial t = 0 \tag{1.2.4}$$

Now this is a hyperbolic (more about this later) differential equation for c, subject to boundary conditions (1.1.5) and (1.1.6), which constitutes a well-posed mathematical problem.

Remark 1.2.2

With a constitutive equation such as (1.2.2), the second part of the initial condition (1.1.5) ($c = 0$) guarantees that also the first part ($f = 0$) is satisfied. This is often, but not invariably the case. Elimination of the flux from the canonical set may sometimes be misleading in that it obscures the fact that an initial condition on the flux has to be satisfied.

Since we are interested in $c(L,t)$, it is natural to define a dimensionless position $x = X/L$, and a dimensionless time $\tau = Ut/L$ (so that $\tau = 1$ is a time equal to the residence time in the PFR). With this, the problem becomes:

$$\partial c/\partial x + \partial c/\partial \tau = 0 \tag{1.2.5}$$

$$\tau = 0,\ x > 0,\ \ c = 0 \tag{1.2.6}$$

$$\tau > 0,\ x = 0,\ \ c = 1 \tag{1.2.7}$$

Remark 1.2.3

L is an “external” length scale, i.e., a length scale which is not intrinsic to the differential equation. One should be careful with such length scales, since in some cases they may in fact not influence the solution of the problem, and their inclusion may invalidate the dimensional analysis of the problem at hand. For instance, for the present problem the concentration distribution $C(X,t)$ is in fact not expected to depend on L, since L does not appear in either the differential equation or the boundary conditions.

It is now useful to define curves (actually straight lines) in the X-τ plane as follows:

$$X = \tau + K \tag{1.2.8}$$

where K is an arbitrary parameter. The substantial derivative of c along any one of these curves is:

$$Dc/D\tau = \partial c/\partial\tau + \partial c/\partial x \qquad (1.2.9)$$

and hence:

$$Dc/D\tau = 0 \qquad (1.2.10)$$

which guarantees that c is constant along any of the lines described by Eq. (1.2.8).

Remark 1.2.4

Hyperbolic problems are best analyzed by determining the “characteristic curves”. The technique appears trivial here, but it is very useful for more complex problems. Notice that the D/Dτ derivative is one along the characteristic curve; in this case it happens to coincide with the intrinsic substantial derivative $D/Dt = \partial/\partial t + U\partial/\partial X$, but this is not necessarily the case in all hyperbolic problems.

In view of Eq. (1.2.10), one concludes that:

$$c = c(\tau - X) = c(\phi) \qquad (1.2.11)$$

is the general solution to Eq. (1.2.5). Boundary conditions (1.2.6) and (1.2.7) become:

$$\phi < 0,\ c = 0 \qquad (1.2.12)$$

$$\phi > 0,\ c = 1 \qquad (1.2.13)$$

which, together with Eq. (1.2.11), furnish the solution to our problem. The imposed jump simply travels undisturbed (i.e., its magnitude stays unity at all times) with velocity U. A transparency of unity will be observed at the exit up to $\tau = 1$, and a transparency of zero after that. The solution is sketched in Fig. 1.2.1.

This is, of course, a trivial result which could have been obtained almost by inspection. However, there are lessons which can be learned from it, and some of them are summarized below. Much more will be discussed about each one of the points below in later sections of this and other chapters.

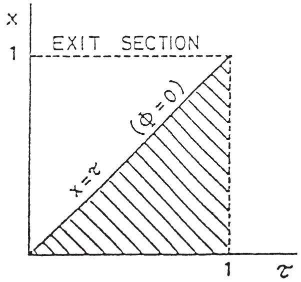

Fig. 1.2.1. Solution of the problem in the x–τ plane. The hatched triangular domain $\tau > x > 0$ is the one considered in the reformulation of the problem based on Eq. (1.2.27).

1. When the constitutive equation is of the very simple type where the flux is proportional to the concentration, hyperbolic equations arise.
2. Hyperbolic equations admit discontinuous solutions, and hence any imposed jump travels at some finite speed through the system.
3. The speed of propagation of the jump is, for the conditions of the problem considered, constant, and so is the amplitude of the jump.

The two parts of conclusion 3 could have been obtained by a procedure which, for the very simple problem considered, appears as overly sophisticated — mathematical overkill. However, it is not going to be that in more complex problems. The procedure is the Kotchine one, and is discussed below.

Suppose that discontinuities may exist in the solution of a problem described by Eq. (1.2.3), which is rewritten below in dimensionless form:

$$\partial f/\partial x + \partial c/\partial \tau = 0 \tag{1.2.14}$$

Let the position of the discontinuity be $y = s(\tau)$, with $ds/d\tau = V$ being the speed of propagation of the discontinuity. The following change of variables is useful:

$$y = s - x \tag{1.2.15}$$

so that:

$$\partial/\partial x = -\,\delta/\delta y \tag{1.2.16}$$

$$\partial/\partial\tau = \delta/\delta\tau + V\delta/\delta y \tag{1.2.17}$$

where δ indicates the partial derivative in the y–τ system of independent variables. Equation (1.2.14) becomes:

$$-\delta f/\delta y + V\delta c/\delta y + \delta c/\delta\tau = 0 \tag{1.2.18}$$

Now both f and c are continuous and differentiable both to the right and to the left of $x = s$ (i.e., of $y = 0$), and hence limits for x approaching s from the right and from the left, identified with subscripts R and L, are defined. Let Γ be any variable which is continuous and differentiable on both sides of the discontinuity, and let $[\Gamma]$ be $\Gamma_R - \Gamma_L$. Take the integral of Eq. (1.2.14) between $y = -\varepsilon$ and $y = +\varepsilon$, and then the limit as ε approaches zero. Terms of the type $\delta\Gamma/\delta t$ will yield a value of zero; terms of the type $\delta\Gamma/\delta y$ will yield $[\Gamma]$. Hence one obtains:

$$[f] = V[c] \tag{1.2.19}$$

Remark 1.2.5

> The Kotchine procedure has been applied to the balance equation. In general, it is useful only for the problem formulated in canonical form (say to equations containing only first derivatives).

Equation (1.2.19) is nothing else but a balance across the discontinuity, and it could be obtained essentially by inspection: if there is a finite discontinuity of concentration ($[c]$ different from zero), there must be a finite discontinuity of flux in order to carry away the excess liberated as the discontinuity propagates with speed V (see Fig. 1.2.2).

A more subtle application of Kotchine's procedure is to consider it in connection with the constitutive equation, not the balance one. For the problem at hand, this is so trivial that in fact no Kotchine procedure is needed, since Eq. (1.2.2) implies that:

$$[f] = [c] \tag{1.2.20}$$

which, when combined with Eq. (1.2.10), yields:

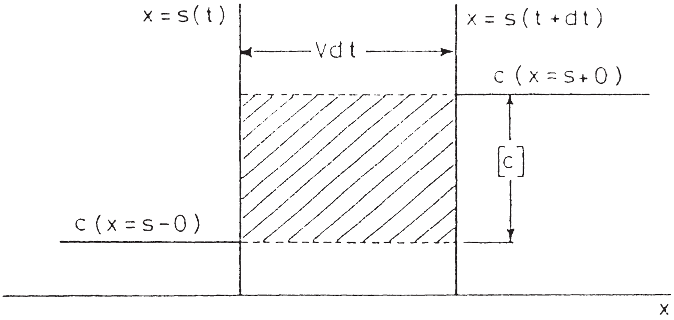

Fig. 1.2.2. Balance across a travelling discontinuity. An amount $[C]V$ must be supplied per unit cross-sectional area and unit time; this is provided by the flux difference $[F]$.

$$V = 1 \tag{1.2.21}$$

i.e., the imposed jump propagates with unit speed.

Remark 1.2.6

The first part of conclusion 3 has been obtained without actually obtaining the solution of the problem. It is often good practice to obtain as much information as possible on the solution of a problem before actually trying to find the solution.

Equation (1.2.5) is a partial differential equation which admits discontinuous solutions. Now the problem considered is so simple that an exact analytical solution has been obtained; however, in more complex cases one would need to use numerical techniques of integration, and the existence of a travelling discontinuity forces very cumbersome numerical techniques. This can often be avoided by a reformulation of the problem, which is sketched below for the problem at hand.

Since the jump imposed at $X = \tau = 0$ travels with unit speed, the solution to the right of the discontinuity ($x > \tau$) is simply $f = c = 0$. It follows that, to the left of the discontinuity ($x = \tau - 0$), $f_L = -[f]$ and $c_L = -[c]$. The substantial derivative along the travelling discontinuity is (for the problem at hand $V = 1$):

$$D/D\tau = \partial/\partial\tau + V\partial/\partial x \tag{1.2.22}$$

Remark 1.2.7

> This is now a substantial derivative along the travelling discontinuity, which (for the case at hand) happens to coincide with the intrinsic substantial derivative.

Thus one has:

$$\mathrm{D}c_\mathrm{L}/\mathrm{D}\tau = \partial c/\partial x + \partial c/\partial \tau \tag{1.2.23}$$

$$\mathrm{D}f_\mathrm{L}/\mathrm{D}\tau = \partial f/\partial x + \partial f/\partial \tau \tag{1.2.24}$$

Equation (1.2.20) now yields:

$$2\mathrm{D}c_\mathrm{L}/\mathrm{D}\tau = \partial(f+c)/\partial x + \partial(f+c)/\partial \tau \tag{1.2.25}$$

or, in view of Eqs. (1.2.20) and (1.2.5):

$$\mathrm{D}c_\mathrm{L}/\mathrm{D}\tau = 0 \tag{1.2.26}$$

Remark 1.2.8

> The second part of conclusion 3 is implied by Eq. (1.2.26). Notice that again it has been obtained without actually solving the problem.

Since $c_\mathrm{L}(0) = 1$, Eq. (1.2.26) integrates to:

$$c_\mathrm{L} = 1 \tag{1.2.27}$$

Now the problem described by Eqs. (1.2.5–7), the solution of which was sought in the whole positive y–τ quadrant, reduces to a problem described by Eqs. (1.2.5–7), the solution of which is sought in the triangular domain $\tau > x > 0$ (see again Fig. 1.2.1). Equation (1.2.27) is the boundary condition at $x = \tau$ which substitutes Eq. (1.2.6). The essential point to be noticed is that, in the new domain of integration, there are no discontinuities, and hence in the new formulation numerical solution is likely to be very simple. For the problem at hand, the solution is obviously $c = 1$ in the triangular domain (and $c = 0$ at $x > \tau$).

Remark 1.2.9

> We have learned a new lesson. Given a hyperbolic problem which presents significant difficulties of numerical integration, it may sometimes be possible, by judicious application of Kotchine's procedure, to transform it to a form where there are no discontinuities in the domain of integration.

The problem considered so far is extremely simple, and it may appear to be trivial. However, it does not take much of a variation to add quite a bit of spice to it. Let us consider the case where parameters are not constant; in our problem, the only intrinsic parameter is the velocity U, and so let us consider the case where U is a function of C. Now if the reader wonders what is a realistic problem for which that might be true, it is a highway traffic problem. F is the flux of cars, C their concentration, and U their velocity. Say U has a maximum U^* (55 mph, or whatever), but it goes to zero when the concentration C reaches the bumper-to-bumper value. Hence U depends on C.

Since $F = U(C)C$, one has:

$$\partial F/\partial X = (U + C\mathrm{d}U/\mathrm{d}C)\partial C/\partial X \tag{1.2.28}$$

and the balance equation becomes:

$$(U + C\mathrm{d}U/\mathrm{d}C)\partial C/\partial X + \partial C/\partial t = 0 \tag{1.2.29}$$

which is still hyperbolic, but is not linear any more. When mathematicians meet a nonlinear problem they know what to do with, they call it quasilinear, so (1.2.29) enjoys that denomination. Let $C(X,0)$ be the initial concentration distribution, and let $C_0 = C(X_0,0)$ be a particular initial concentration at point X_0. Consider the characteristic line $X = f(t), f(0) = X_0$ such that:

$$\mathrm{d}f/\mathrm{d}t = U(C) + C_0U'(C_0) \tag{1.2.30}$$

where the $'$ denotes the ordinary derivative. The substantial derivative along $f(t)$ is equal to the left hand side (LHS) of (1.2.29), and is thus zero: C stays at C_0 along $X = f(t)$, so that the latter is indeed a characteristic curve. The characteristic curves are all straight lines, but their slope depends on the point of the $t = 0$ axis through which they go.

It may be helpful, in order to understand characteristic lines, to think of police cars along the highway. They can move any way they wish (by using the emergency lane): forwards, backwards, and at any speed they wish. They are given the order to regulate their speed in such a way as to see a constant concentration of cars; they start moving at time zero. In this interpretation, d*f*/d*t* is the police cars' speed, and the substantial derivative along $f(t)$ is the time derivative as observed by a policeman, so that the substantial derivative of concentration is, by definition, zero.

Let's try this out on a somewhat realistic problem. Let's say that $U = U^*(1 - C/J)$, where U^* is 55 mph (or whatever), and J is the bumper-to-bumper concentration. A traffic light has been red for quite some time (they do have a frustrating tendency to do so), and a lot of cars are stopped behind it when it turns green at $t = 0$; none is in front, since, although the light was red, nobody came in through the other side (traffic lights do that too, of course; drivers behind the light are getting dangerously mad). So the initial concentration distribution is:

$$C(X,0) = JH(-X) \tag{1.2.31}$$

where $H()$ is the Heaviside step function. Now we can calculate d*f*/d*t* from Eq. (1.2.30), to find out that:

$$X_0 > 0, \mathrm{d}f/\mathrm{d}t = U^* \tag{1.2.32}$$

$$X_0 < 0, \mathrm{d}f/\mathrm{d}t = -U^* \tag{1.2.33}$$

(police cars behind the traffic light will start going in reverse with speed U^* at $t = 0$; police cars in front of it will move forward at the same speed).

So in front of the light the concentration is still zero at $X > U^*t$. Behind the light, it is still J (and hence U is still zero) at $X < -U^*t$ (that is when the horns start to blow—you see the green light, and yet you still can't move. It is not the fault of the drivers in front of you, it is the quasilinear hyperbolic problem's fault. Coupled, of course, with the council's fault of placing a traffic light where nobody comes from the other side anyway). The situation is sketched in Fig. 1.2.3.

Now our analysis tells us what happens at $X > U^*t$, and at $X < -U^*t$. What happens in between (where, you may notice, there are no police cars whatsoever)? The concentration there is $C = (J/2)(1 - X/(U^*t))$, as can be verified by substitution into Eq. (1.2.29) (assuming that cars do not use the absence of police cars to break any traffic rules).

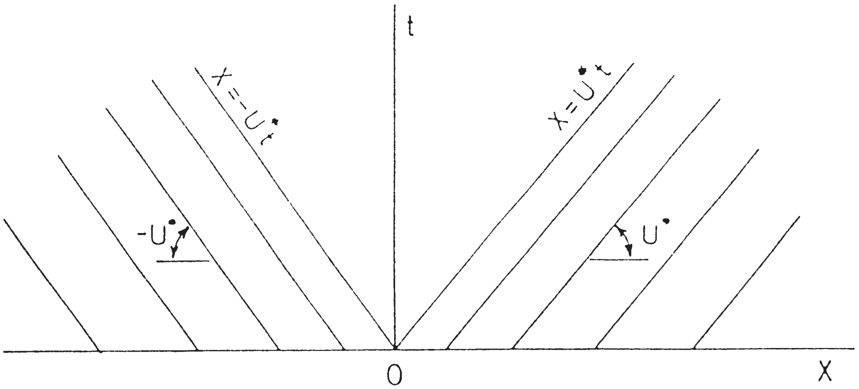

Fig. 1.2.3. Sketch of the characteristic lines for the red traffic light problem.

Now let us try out a case where the traffic jam is ahead, rather than behind. Say that:

$$C(X,0) = \alpha J + (1-\alpha)JH(X); 0 < \alpha < 0.5 \tag{1.2.34}$$

so that it is bumper to bumper at positive X. Now we can go through the same analysis again to obtain:

$$X_0 < 0;\ \mathrm{d}f/\mathrm{d}t = U^*(1-2\alpha) \tag{1.2.35}$$

$$X_0 > 0;\ \mathrm{d}f/\mathrm{d}t = -U^* \tag{1.2.36}$$

Now the situation is as sketched in Fig. 1.2.4. At $X > U^* t(1-2\alpha)$, i.e., in the rightmost region, every point lies on one and only one characteristic line originating at $X_0 > 0$, and hence the traffic is still bumper to bumper: that does not surprise us, since nothing has happened in front to ease up the situation. In the leftmost region, $X < -U^* t$, every point lies on one and only one characteristic line originating at $X_0 < 0$, and hence the concentration is still αJ and the velocity $U^*(1-\alpha)$. But what about the triangular region in between, where every point lies on two different characteristic lines, one originating at $X_0 > 0$ and one at $X_0 < 0$? This is a region where police cars are quite abundant, some of them coming forward and some coming backwards. But what did the police car sitting at $X = 0$ decide to do? If the driver was looking

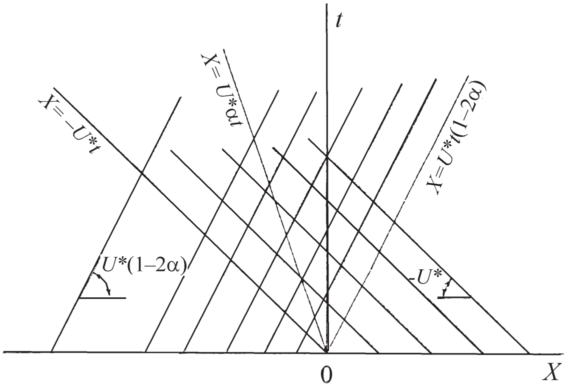

Fig. 1.2.4. Sketch of the characteristic lines for the traffic jam ahead problem. The speed of propagation of the shock wave is obtained from Kotchine's procedure.

ahead, it started backwards with speed U^*, and if he was looking behind it started forward with speed $U^*(1 - 2\alpha)$. It may of course, statistically, have done anything in between, and we need to establish what it did on the average—how does the discontinuity propagate. Here Kotchine's procedure helps us. The balance equation still yields $[F] = V[C]$, where V is the speed of propagation of any discontinuity, and the constitutive equation tells us that $[F] = [UC]$; it follows that:

$$V = [UC]/[C] \tag{1.2.37}$$

For the problem at hand, there is an initial discontinuity at $X = 0$ for which $[UC] = -U^* J\alpha(1 - \alpha)$ and $[C] = J(1 - \alpha)$, hence $V = -U^*\alpha$. Thus a shock wave propagates upstream with velocity $U^*\alpha$; this can easily be checked by a mass balance. (The restriction $\alpha < 0.5$ was imposed so that the "left" characteristics would have a positive slope; however, the same argument applies also at $\alpha > 0.5$, with the middle triangular region lying wholly in the left quadrant). At $X > -U^*\alpha\mathrm{t}$, $C = J$ and the traffic is jammed; at $X < -U^*\alpha t$, the traffic is still moving.

Now let us look at the problem from another viewpoint. The flux F equals $U^*C(1 - C/J)$, and hence it has a maximum of $U^*J/4$ at $C = J/2$, see Fig. 1.2.5. Now suppose $\partial C/\partial X \geq 0$ (there is risk of a traffic jam ahead). As long as $C < J/2$, $\partial F/\partial X \geq 0$, and hence $\partial C/\partial t \leq 0$: the traffic is thinning out, no problem. But if $C > J/2$, the traffic will thicken, and a traffic jam will occur eventually. A shock

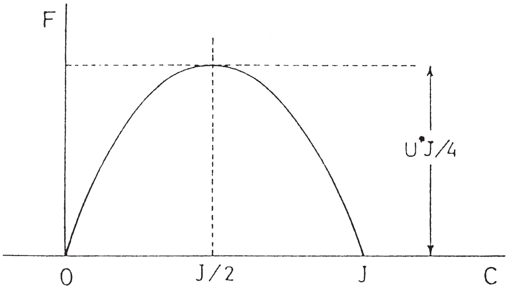

Fig. 1.2.5. Flux of cars as a function of concentration. At $C = 0$ the flux is zero because there are no cars. At $C = J$ there is no flux because a traffic jam has developed.

wave will develop without need of it being imposed to start with. Situations of this type are quite common in Gas Dynamics (to be treated at an elementary level is Section 1.5), where shocks may develop in consequence of the nonlinearities of the governing equations.

1.3 THE PLUG FLOW REACTOR WITH DIFFUSION

In this section, we introduce a somewhat more complex constitutive equation than Eq. (1.2.1), by considering the possibility of axial diffusion in a PFR. Thus we write:

$$F = UC - D\partial C/\partial X \qquad (1.3.1)$$

where D is the axial diffusivity.

Remark 1.3.1

Again, the constitutive equation is such that the initial condition $c = 0$ guarantees that also $F = 0$ (in fact, any initial distribution $C(X,0)$ determines the initial flux distribution $F(X,0)$ through Eq. (1.3.1). This will no longer be true at the next level of complexity of the constitutive equation, where relaxation phenomena are taken into account.

It is still useful (within the qualifications noted earlier) to define dimensionless quantities $f = F/JU$, $x = X/L$, $c = C/J$, so that the constitutive equation becomes:

$$f = c - Pe\partial c/\partial x \tag{1.3.2}$$

where $Pe = D/UL$ is the inverse of the Peclet number. With $\tau = Ut/L$, the balance equation is:

$$\partial f/\partial x + \partial c/\partial \tau = 0 \tag{1.3.3}$$

and thus the Kotchine procedure as applied to the balance equation yields:

$$[f] = V[c] \tag{1.3.4}$$

However, we may now apply Kotchine's procedure (in a form which is not absolutely trivial) to the constitutive equation, Eq. (1.3.2). Since the last term on the right hand side (RHS) is the only one which does not go to zero identically, one obtains:

$$[c] = 0 \tag{1.3.5}$$

which, together with Eq. (1.3.4), guarantees that there are in fact no discontinuities whatsoever: the imposed jump is smoothed out immediately.

Remark 1.3.2

> Inclusion of a diffusive term in the constitutive equation has resulted in the impossibility of discontinuous solutions. This, as will be seen, is a result of rather general validity.

Remark 1.3.3

> The result in Eq. (1.3.5) is based on the implicit assumption that the flux does not become infinitely large at the surface of discontinuity.

Having used the canonical form for all it can tell us, we now substitute the constitutive equation into the balance equation to obtain:

$$\partial c/\partial x - Pe\partial^2 c/\partial x^2 + \partial c/\partial \tau = 0 \tag{1.3.6}$$

A subtle question about the boundary conditions arises at this stage. One could, of course, simply still use Eqs. (1.2.6–7); however, the physics of the problem should be revisited. As long as diffusion was not considered, there was no way that events forced at some value of x could produce results uphill, i.e., at lower values of x: the only mechanism for the propagation of effects was convection, which is effective only in the downstream direction (F has necessarily the same sign as U). Diffusion, however, is a mechanism which acts in both directions, and therefore water in the reactor may in fact diffuse upstream into the feed tube. Hence Eq. (1.2.7) does not necessarily describe correctly the physical conditions at the inlet.

Remark 1.3.4

> The whole point of writing boundary conditions is that one hopes them to subsume all that happens outside of the system specifically isolated for analysis and is relevant to its behaviour. This implies that for any given problem the boundary conditions to be written may depend on what the equations governing the behaviour of our system are.

For the case at hand, two different possibilities arise: the first one is to assume that diffusion in the feed tube cannot take place, so that perhaps one can still use Eq. (1.2.7) (but see Appendix 1.3 in this regard) as a boundary condition; the other one is to assume that diffusion takes place in the feed tube just as effectively as it does in the PFR, and that at time zero we abruptly substitute for the feed tube which was full of water one which is full of an ink solution. Thus we write (see Remark 1.3.7 below):

$$\tau = 0, x < 0, \ c = 1 \tag{1.3.7}$$

Let's first examine the problem as described by this last choice (an easy analytical solution is obtainable). In particular, let us consider the case where the Peclet number is significantly smaller than unity, say axial diffusion is expected to at most result in a minor correction of the result obtained in the previous section. Physically, one expects that a "quasi-discontinuity", or wave, propagates through the system, perhaps at the same unit speed as in the case of no diffusion; the presence of the diffusive term being felt in that the wave becomes progressively less sharp as it travels downstream, see Fig. 1.3.1.

Remark 1.3.5

> It is always good practice to think of a qualitative sketch of the solution of a problem before tackling the actual analysis of it.

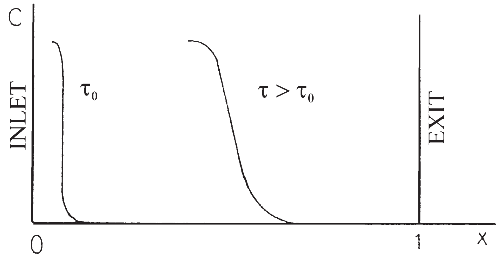

Fig. 1.3.1. Sketch of the solution. The jump imposed travels downstream and gets spread out in doing so.

This physical picture suggests that perhaps the problem is best described in a coordinate system which moves with the wave, so that we introduce the following new variable y:

$$y = x - \tau \tag{1.3.8}$$

Let δ be the partial derivatives in the y–τ system of independent variables. One obtains:

$$\partial/\partial x = \delta/\delta y \tag{1.3.9}$$

$$\partial^2/\partial x^2 = \delta^2/\delta y^2 \tag{1.3.10}$$

$$\partial/\partial\tau = \delta/\delta\tau - \delta/\delta y \tag{1.3.11}$$

and indeed our hope has been fulfilled, since Eq. (1.3.6) reduces to:

$$Pe\delta^2 c/\delta y^2 = \delta c/\delta\tau \tag{1.3.12}$$

which is parabolic (parabolic equations do not admit discontinuous solutions). Furthermore, we have found an old friend, the one-dimensional unsteady conduction equation.

Boundary conditions (1.2.6) and (1.3.7) reduce to:

$$\tau = 0,\ y > 0,\ \ c = 0 \tag{1.3.13}$$

$$\tau = 0,\ y < 0,\ \ c = 1 \tag{1.3.14}$$

Now the Neuman variable $z = y/2\sqrt{(\tau Pe)}$ is known to reduce the diffusion equation (1.3.12) to an ordinary differential equation ($d^2c/dz^2 + 2z dc/dz = 0$). The BCs in Eqs. (1.3.13) and (1.3.14) can be expressed in terms of z as follows:

$$z = +\infty,\ c = 0 \tag{1.3.15}$$

$$z = -\infty,\ c = 1 \tag{1.3.16}$$

and thus an analytical solution $c = c(z)$ can be obtained:

$$2c = 1 - erfz \tag{1.3.17}$$

Remark 1.3.6

> The fact that the Neuman variable reduces the partial differential equation to an ordinary one, and that the boundary conditions can be written in terms of it, means that the solution is a “similar” solution: at any given τ, the shape of the $c(y)$ curve is the same, and all such curves can be made to coincide by appropriate rescaling of the y variable (specifically, the scaling has to be inversely proportional to $\sqrt{\tau}$). The possibility of obtaining a similarity solution is often more easily predicted by keeping the problem in dimensional form. For the case at hand, L does not appear in either the differential equation or the boundary conditions (when formulated in dimensional form), and there is no intrinsic length scale; hence a similarity solution is guaranteed to exist. Notice that one can have a similarity solution even if the differential equation does not reduce to an ordinary one: any solution obtained by separation of variables, of the type $c = f(x)g(\tau)$, is a similarity solution.

The importance of Remark 1.3.6 cannot be over-emphasized. Suppose one had chosen a different normalization for the problem at hand, based on the concept that, since no boundary conditions are imposed at $X = L$, the parameter L does not really play any role. Thus one could have chosen a different (starred) set of independent variables:

$$x^* = XU/D = x/Pe \tag{1.3.18}$$

$$\tau^* = tU^2/D = \tau/Pe \tag{1.3.19}$$

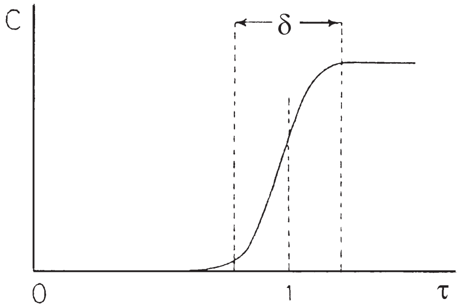

Fig. 1.3.2. Response at the exit of the reactor.

The dimensionless speed of propagation of the wave would still be unity. However, the differential equation would now be:

$$\partial c/\partial x^* - \partial^2 c/\partial x^{*2} + \partial c/\partial \tau^* = 0 \tag{1.3.20}$$

which contains no parameters, in contrast to Eq. (1.3.6). Correspondingly, one would have $z^* = (x^* - \tau^*)/2\sqrt{\tau^*}$ as the Neuman variable, and Eq. (1.3.17) would become $2c = 1 - erfz^*$, again with no parameters. Where has the Peclet number disappeared in this alternate formulation?

In order to resolve this apparent paradox, it is useful to first analyze the solution as obtained in the unstarred variables. The response at $X = L$ ($x = 1$) would be as sketched in Fig. 1.3.2, with the "spread" δ of order $2\sqrt{Pe}$. If Pe is a small number, this means the spread is small, as compared to the reactor length. In the starred variable alternative, the spread would be of order $2\sqrt{\tau^*_0}$, where τ^*_0 is the value of τ^* at which the wave midpoint reaches the exit. Now the speed of propagation is still unity, but the exit is located at $x^* = 1/Pe$, so that $\tau^*_0 = 1/Pe$. Thus the ratio of the spread to the length of the reactor is again $2\sqrt{Pe}$, as we expected. The point is that we are not interested in the spread *per se*, but only as compared to the actual length of the reactor, and in this sense L **is** a relevant external length scale.

At any given position x (≤ 1 necessarily) the spread is of order $2\sqrt{xPe}$. Thus the spread is smaller near the entrance of the reactor (as we expected), and is negligibly small at the inlet section itself. Indeed, except at τ values smaller than $2\sqrt{Pe}$, the value of c at $x = 0$ is practically unity. This is satisfactory, since the question of whether (1.2.7) or (1.3.7) is the correct initial condition becomes presumably irrelevant.

Remark 1.3.7

> It is often useful to solve a problem for boundary conditions which, though perhaps not very realistic, make the solution easier. If then the solution "almost" satisfies those BC's which we think are more realistic, one can hope that one is doing well, though of course there is no guarantee that such is invariably the case.

We now turn attention to an alternate approach to the solution of Eq. (1.3.6). Suppose we were not smart enough to solve it exactly, and we therefore are willing to look for an approximate solution. In particular, we want an approximate solution valid only for small values of *Pe* — our reactor is almost an ideal PFR.

We now observe that, if we are willing to set *Pe* to 0 in Eq. (1.3.6), we recover Eq. (1.2.5), for which we have a solution. Well, isn't that good enough?

It may seem it is, but closer inspection of the situation generates some doubts. The solution to the $Pe = 0$ problem is $c = 0$ at $x > \tau$, $c = 1$ at $x < \tau$. Notice that this not only satisfies the *reduced* differential equation (Eq. 1.2.5) as well as all the boundary conditions, but also the *complete* differential equation (1.3.6) everywhere — except at $x = \tau$, where the term $\partial^2 c/\partial x^2$ becomes very nastily singular, and certainly very different from zero. No matter how small the actual value of *Pe* may be, there is certainly some difficulty in neglecting this term at $x = \tau$, and we should do some more thinking.

Remark 1.3.8

> The situation encountered here is a typical one. The coefficient of the highest order derivative in a differential equation approaches zero: this is always a singular limit. In general, the singularity is related to the fact that some of the BCs cannot be imposed on the problem obtained by setting the coefficient to zero, since the differential equation becomes of a lower order. However, even in those cases (as the present one) when all boundary conditions can also be imposed on the lower order problem, the limit is still singular.

Remark 1.3.9

> Here we have also approached dangerous ground from another viewpoint. We are seeking an approximate solution to some equation (in the present case, Eq. (1.3.6). What we are trying to do is to drop some terms in the equation (say the diffusive term in this case), and then solve the resulting

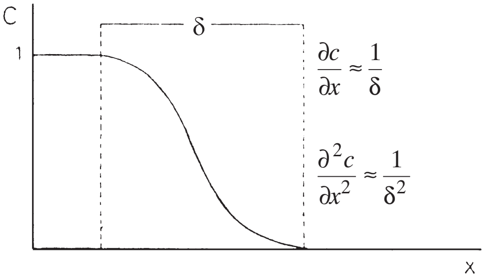

Fig. 1.3.3. Sketch for the evaluation of the order of magnitude of terms in the differential equation.

> problem. What will be obtained by this procedure is an exact solution of an approximate equation, not an approximate solution of an exact equation. We generally hope that is good enough, but we should always be careful.

Now we may reason as follows. The problem arises at $y = 1$; perhaps, at y values sufficiently different from unity, the solution we have for the reduced problem is reasonable; say we tentatively assume that:

$$x > \tau, c = 0; \; x < \tau, \; c = 1 \tag{1.3.21}$$

Now how much larger than τ should x be for c to be 0 within some reasonable approximation? We may reason as follows. When x and τ are not very different, the diffusive term cannot be neglected; hence it must be of the same order of magnitude of some other term in the differential equation. Let δ be the "spread", i.e., the range of x values around τ where the solution to the lower order problem is not acceptable. (This is usually called a "boundary layer"). Now within the boundary layer $\partial c/\partial x$ is of order $1/\delta$, and $\delta^2 c/\delta x^2$ is of order $1/\delta^2$ (see Fig. 1.3.3). Since the wave propagates with finite speed, $\partial c/\partial \tau$ is also of order $1/\delta$. Hence for the diffusive term to be of the same order of magnitude as the other terms, δ must be of order $\sqrt{Pe}$.

Are we happy? Partially so. At the exit of the reactor, we know our answer is correct. But it isn't correct near the inlet. The problem is that the initial conditions force an initial boundary layer thickness of zero, and thus we cannot simply look at the problem the way we have done. The two terms $\partial c/\partial x$ and $\partial c/\partial \tau$ have the same order of magnitude, but their algebraic sum may well be of a much smaller order of magnitude if they have opposite signs, as indeed is the case here, since $\partial c/\partial x < 0$ but $\partial c/\partial \tau > 0$. We need to establish the order of magnitude of $\partial c/\partial x + \partial c/\partial \tau$, and then require $Pe\partial^2 c/\partial x^2$ to be of the same order of magnitude.

This now suggests all too naturally the introduction of the y variable, since $\partial c/\partial x + \partial c/\partial \tau = \delta c/\delta y$, see Eqs. (1.3.9–11). Thus the order of magnitude analysis must be applied to Eq. (1.3.12), and now we obtain $\delta = \sqrt{(Pe\tau)}$, which we know is the right result.

Remark 1.3.10

The boundary layer thickness has been obtained without actually solving the boundary layer equations (discussed below). The procedure is of general validity: require that the term to be dropped in the lower order problem has the same order of magnitude as the sum of all other terms in the differential equation. The procedure may at times, as we have seen, be rather tricky.

Having established the boundary layer thickness, we may now proceed to the rescaling of our problem by introducing a "stretched coordinate". We want the diffusive term to be of the same order of magnitude as the sum of all other ones, and hence we define a variable z as follows:

$$z = (x - \tau)/2\delta = (x - \tau)/2\sqrt{(Pe\tau)} \tag{1.3.22}$$

(Wise as we are in this case, since we possess the exact solution already, we have added the factor 2 in the denominator because we know that is going to make the algebra slightly less cumbersome). With this, the differential equation within the boundary layer becomes:

$$\mathrm{d}^2c/\mathrm{d}z^2 + 2z\mathrm{d}c/\mathrm{d}z = 0 \tag{1.3.23}$$

Remark 1.3.11

In this particular case, the differential equation within the boundary layer reduces *exactly* to an ordinary one. This is by no means true in general; however, often the differential equation reduces to an ordinary one to within an approximation of order δ.

We now need BCs on Eq. (1.3.23), and we write them in the form of Eqs. (1.3.15) and (1.3.16).

Remark 1.3.12

BCs for the boundary layer differential equations are written *in general* by requiring the limit of the solution $c(z)$ within the boundary layer (the inner solution) when the stretched coordinate approaches ∞ (i.e., far enough outside of the boundary layer) to coincide with the limit of the solution $c_0(x)$ of the lower order problem when the unstretched one approaches zero (zero being the position of the boundary layer). For the case at hand, this corresponds to:

$$\lim_{z=+\infty} c(z) = \lim_{x=\tau+0} c_0(x,\tau) = 0$$

and the analogous one for $-\infty$, $\tau - 0$.

Having come so far, we have essentially recovered the exact solution of the full problem which we had found before — except that now the irrelevance of the initial condition (1.2.7 or 1.3.7) becomes more obvious.

Appendix 1.3

The case where *Pe* becomes very large is an entirely different matter. Let's first examine what we expect when it becomes infinitely large. If the diffusivity is infinitely large, no concentration gradient may exist; hence our reactor has become a CSTR. The balance equation for a CSTR is:

$$UJ = UC + L\mathrm{d}C/\mathrm{d}t;\ C(0) = 0 \tag{1.3.24}$$

and we believe something like BC (1.2.7), i.e., that the feed stream has a concentration J at all positive times. In dimensionless form:

$$1 = c + \mathrm{d}c/\mathrm{d}\tau;\ c(0) = 0 \tag{1.3.25}$$

which obviously integrates to:

$$c = 1 - \exp(-\tau) \tag{1.3.26}$$

Since there are no gradients in the reactor, Eq. (1.3.26) also describes the response measured at the outlet.

The important point to notice is that Eq. (1.3.26) does not satisfy BC (1.2.7). What is going on here? We need to revisit the BCs again. If the diffusivity is zero in the feed tube, what we need to do is to write a balance across the feed section, which is indeed a surface of discontinuity (since D undergoes a discontinuity there). Kotchine's procedure is of help again, and it yields (since this time $V = 0$):

$$[F] = 0 \tag{1.3.27}$$

Now F is given by Eq. (1.3.1), and hence one obtains:

$$UJ = UC - D\partial C/\partial X \tag{1.3.28}$$

as the appropriate BC at the inlet. In dimensionless form:

$$1 = c - Pe\partial c/\partial x \tag{1.3.29}$$

This is, in some vague sense, satisfied by Eq. (1.3.26), because $\partial c/\partial x$ is zero but Pe is infinitely large and hence c may be different from unity. The important point is that, at finite values of Pe, Eq. (1.3.29) should (perhaps) be used as the BC. One now also needs a BC at the outlet, and the same reasoning (with diffusivity zero in the outlet tube) yields

$$x = 1, \partial c/\partial x = 0 \tag{1.3.30}$$

Remark 1.3.13

BCs (1.3.29) and (1.3.30) are called the Danckwerts boundary conditions for a tubular reactor with axial diffusion.

It turns out that an axial diffusivity model is not a very good one when the reactor is "almost" a CSTR. The equations can be integrated to obtain a series solution, which degenerates properly to both the PFR and the CSTR solutions. If only the correction which is first order in $1/Pe$ is retained, one obtains:

$$c = 1 - (1 + 3/8Pe)\exp(-\tau(1 + 1/4Pe)) \tag{1.3.31}$$

for the response measured at the exit section.

1.4 PURE DIFFUSION

In this section, we investigate the situation where there is no velocity component in the X direction, and hence no convective flux. If the constitutive equation for the flux is written in the usual form, we expect something like:

$$F = -D\partial C/\partial X \tag{1.4.1}$$

but, as will be seen, there are some subtleties.

First consider the classical Stokes–Rayleigh problem. A semi-infinite body of an incompressible Newtonian fluid bounded by a flat plate located at $X = 0$ is initially at rest. At time zero, the plate is suddenly set in motion with velocity J/Φ in the Y direction (orthogonal to X). Φ is the density of the fluid, so that J is indeed the jump of the concentration of Y component of linear momentum imposed at $X = t = 0$. We make the hypothesis that, at all X and all t, the only nonzero component of velocity, $U(X,t)$, is in the Y direction, and that pressure is constant in both space and time (the constant value can conveniently be set to zero, since for incompressible fluids with no free surfaces only the pressure gradient plays any role, not pressure itself). These assumptions can be seen to satisfy the equation of continuity and the momentum balance in the X and Z directions. The momentum balance in the Y direction becomes:

$$\partial F/\partial X + \partial(\Phi U)/\partial t = 0 \tag{1.4.2}$$

where F is the X component of the flux of the Y component of momentum, i.e., the XY tangential stress. With the assumed kinematics, this is indeed the only nonzero component of the stress in excess of pressure. Equation (1.4.2) is, as expected, of the form of (1.1.1), since $C = \Phi U$.

The constitutive equation is Newton's law of friction:

$$F = -\mu\partial U/\partial X = -D\Phi\partial U/\partial X \tag{1.4.3}$$

which is indeed in the form of Eq. (1.4.1) when the ratio of the viscosity μ to the density Φ is interpreted as the diffusivity of momentum, $D = \mu/\Phi$. U can be regarded (more about this later) as the "potential" for momentum transfer.

Kotchine's procedure as applied to the balance equation (1.4.2) gives:

$$[F] = V[\Phi U] \tag{1.4.4}$$

which by itself does not negate the possibility that velocity may suffer a discontinuity.

Remark 1.4.1

> The balance of momentum by itself does not negate the possibility of discontinuities of velocity. Perhaps that does not come as a surprise to people who have some familiarity with gas dynamics. However, it paves the way to the idea that discontinuities of the potential are not inherently impossible.

Kotchine's procedure as applied to the constitutive equation (1.4.3) yields however $[U] = 0$. For a compressible fluid this would imply $[F] = VU[\Phi]$; but for an incompressible fluid $[\Phi] = 0$, and hence no discontinuities of any type are possible.

Remark 1.4.2

> Notice that this time the impossibility of discontinuities is a consequence of two constitutive assumptions: that Newton's law (1.4.3) applies, and that density is a constant. In other words, the distinction between the concentration ΦU and the potential U is a relevant one unless Φ is constant.

Now consider another problem: a semi-infinite solid body bounded by an exposed surface located at $X = 0$, initially at some uniform temperature which can be normalized to zero (temperatures U will be measured by subtracting the initial value). At time zero, the exposed surface is suddenly brought at some temperature J/Φ, where Φ this time is the constant pressure specific heat per unit volume, so that J is indeed the jump of the enthalpy concentration imposed at $X = t = 0$. If F is the heat flux in the X direction, the constitutive equation is written in the form of Fourier's law:

$$F = -k\partial U/\partial X = -\mathrm{D}\Phi\partial U/\partial X \tag{1.4.5}$$

which again can be identified with Eq. (1.4.1) if $k/\Phi = D$ is identified with the diffusivity of heat. The balance equation is again Eq. (1.4.2). If Φ is constant, no discontinuities are possible.

Now consider a semi-infinite body of fluid initially at rest, and containing a zero concentration of some solute. The concentration is brought to some value J at $X = t = 0$. Now the constitutive equation is (1.4.1) (Fick's law), and the balance equation is (1.4.2) (with $\Phi = 1$, and hence certainly constant), and coherently U is the concentration.

Remark 1.4.3

> Mass transfer is the only case where the diffusive constitutive equation is usually written in terms of the concentration, with no distinction between C and the potential U.

The boundary conditions for all three problems are:

$$t = 0,\ x > 0,\ U = F = 0 \tag{1.4.6}$$

$$t > 0,\ x = 0,\ U = J/\Phi \tag{1.4.7}$$

Remark 1.4.4

> The three problems are described by exactly the same equations, and hence the solution of one of them is also the solution of the other two. This may appear trivial today. However, the heat transfer problem was (partly) solved by Fourier in the early 1820s; the momentum transfer problem was formulated by Stokes and solved by Rayleigh in the early second half of the 19th century; and the mass transfer problem was formulated and solved by Higbie in 1935. The analogy must not have been so obvious if it took more than 100 years to go from the solution of the first to the solution of the third problem.

Remark 1.4.5

> Why did we write BC (1.4.7)? After all, we are imposing a velocity J/Φ on the plate, not on the fluid itself. We are imposing the "no-slip" boundary condition. What about hydrodynamics of ideal fluids, where the no-slip boundary condition is not imposed? Well, the answer would be that, no matter what we do to our plate (as long as we limit ourselves to tangential motions of it), the fluid couldn't care less, and it would just stay there sitting still forever. Notice that even with a finite viscosity the differential equations would be satisfied by $U = 0$, the only thing which would not be satisfied is the no-slip boundary condition. We are really sticking our necks out here about this condition. Now in the heat transfer case, we are imposing a jump of the external temperature at $X = 0–$ (assuming we can do that); why do we assume that the same jump is also imposed at $X = 0+$? Analogous considerations apply to the mass transfer case (where at best we can impose a jump of the chemical potential at $X = 0–$). This has all to do with the Fundamental Interface Assumption (FIA), to be discussed in Section 1.9.

Elimination of the flux yields:

$$D\partial^2U/\partial X^2 = \partial U/\partial t \tag{1.4.8}$$

and since everything is linear, it is natural to define a dimensionless potential $\theta = \Phi U/J$ (but more about this later), which leaves the differential equation unaltered, but eliminates the parameter J/Φ from BC (1.4.7). For the time being we need not worry about F, because the first part of (1.4.6) guarantees that the second part is satisfied as well. The problem is reduced to:

$$\mathrm{D}\partial^2\theta / \partial X^2 = \partial\theta / \partial t \tag{1.4.9}$$

$$\theta(X,0) = 0, \ \theta(0,t) = 1 \tag{1.4.10}$$

Since θ is dimensionless, it can only depend on dimensionless quantities. But we notice that we are short of parameters to make both X and t dimensionless: we have only D. Hence the fact that θ must be $\theta(X,t,D)$ implies that $\theta = \theta(X/\sqrt{(Dt)})$ — the Neuman variable emerges naturally from the dimensional analysis (read Remark 1.3.6 again now). The solution is:

$$\theta = 1 - erf(X/2\sqrt{(Dt)}) \tag{1.4.11}$$

This is all very well known, but the result is worth some detailed consideration. First of all, the fact that there are no discontinuities means that, at all $t > 0$ (no matter how small), $\theta > 0$ for all X's. Consider in particular the mass transfer case: the concentration is different from zero at all X's for any positive time, no matter how small. Let's make a small calculation. At $X = 4.6\sqrt{(Dt)}$, Eq. (1.4.11) gives $\theta = 0.022$. So an appreciable number of molecules has reached this position from the exposed surface — which is the only source for them. These molecules have travelled at an average speed $X/t = 340\sqrt{(D/t)}$. At a D value of 10^{-2} cm^2/s (reasonable for a gas), this corresponds to an average speed, in cm/s, of $34/\sqrt{t}$ if t is measured in seconds. At $t = 10^{-20}$ s (an admittedly extremely short time) this corresponds to an average speed about ten times larger than the speed of light — a somewhat disturbing conclusion. Maybe something isn't quite right with our assumptions if we consider times as short as this.

Remark 1.4.6

For the mass transfer case, this paradox is solved when one considers the correct formulation of a linear constitutive equation for the flux, see the

> chapter on thermodynamics. However, the same paradox also arises of course for momentum and heat transfer (nothing can propagate faster than light), and this will be solved by considering relaxation.

Now suppose we wish to calculate the flux $F(X,t)$. In terms of θ, the flux can be written as:

$$F = -DJ\partial\theta/\partial X \tag{1.4.12}$$

and hence one obtains from Eq. (1.4.14):

$$F = 2J\sqrt{(D / \pi t)}\,\exp(-X^2/4Dt) \tag{1.4.13}$$

Now there is one disturbing feature about Eq. (1.4.13): the flux at $X = t = 0$ is infinitely large. For the momentum transport case, this would correspond to an infinite force at the bounding plate, which one could never impose in reality. Furthermore, if we reread Remark 1.3.3 we feel a bit awkward — we are developing a theory which in some sense assumes that F never becomes infinitely large, and we get out from the solution of a problem that on occasion it does.

Remark 1.4.7

> Now of course this conclusion does not imply that one is unwilling to regard the Navier–Stokes equations as satisfactory for the description of most of classical fluid mechanics. They are satisfactory. The point we are making is simply concerned with the limit behaviour of their solutions for problems where discontinuities are to be considered.

There is another strange result about Eq. (1.4.13): we have obtained a similarity solution for θ, but the solution for F is not a similarity one. What's going on here?

Well, suppose that we had decided to eliminate U rather than F from the canonical set. After all, why should we always want to eliminate the flux rather than the concentration? By cross-differentiation one obtains:

$$D\partial^2F/\partial X^2 = \partial F/\partial t \tag{1.4.14}$$

which has the same form as Eq. (1.4.9) — the Neuman variable should work equally well. However, the boundary conditions are now the second one of

(1.4.6) (which guarantees the validity of the first one), and what is the other one? It isn't easy to transform Eq. (1.4.7) into a BC for F.

However, let us revisit the physics a bit. Suppose that, in the momentum transfer problem, at time zero we start pushing the plate with a constant tangential force Q; that in the heat (mass) transfer problem, at time zero we start imposing on the exposed surface a constant heat (mass) flux Q. This certainly disposes of the problem of having an infinite flux at $X = t = 0$, though of course a different physical problem is being described: we are considering the "dual" problem discussed in Section 1.1. Now the second boundary condition becomes:

$$t > 0, X = 0, F = Q \tag{1.4.15}$$

and the solution of Eq. (1.4.14) is (we know that by inspection by now):

$$F/Q = 1 - erf(X/2\sqrt{(Dt)}) \tag{1.4.16}$$

so that now a similarity solution has been obtained for F/Q, but of course not one for U. The point is that in this second problem there is an external flux scale Q, just as in the first version there was an external potential scale J.

Remark 1.4.8

What kind of similarity solution one may obtain (if any), depends not only on the dimensionality of the problem (absence of an external length scale), but on what kind of jump is imposed. The similarity solution is obtained (if at all) for that dependent variable for which there is an external scale.

1.5 SHOCK WAVES IN GASES

In this section, we analyze the elementary theory of shock waves in gases. It will soon become evident why the theory as presented is an elementary one. The physical problem which we consider is again one-dimensional in space, see Figure 1.5.1. A surface of discontinuity is located at $X = 0$; we have chosen the frame of reference as one which moves with the shock, so that the shock is stationary by definition. Furthermore, steady state prevails on both sides of the shock, so that in all balance equations the time derivative of concentration is zero (we are obviously restricting attention to steady shocks).

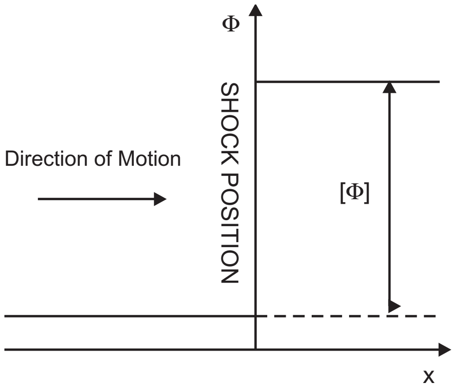

Fig. 1.5.1. Coordinate system for analysis of steady shocks. The coordinate systems has been chosen so that the shock is stationary.

First consider the balance of mass. The concentration is simply the density Φ, while the flux is $\Phi U = W$, where U is the velocity and W is the mass flow-rate per unit cross sectional area. The balance equation is:

$$\partial(\Phi \mathrm{U}) / \partial X = 0 \tag{1.5.1}$$

and hence Kotchine's procedure yields:

$$[\Phi U] = 0 \tag{1.5.2}$$

Remark 1.5.1

When the quantity considered is the total mass, it appears at first sight that one writes down only a balance equation, and no constitutive equation. However, notice that Eq. (1.5.1) contains two variables, Φ and U, and thus by itself cannot determine either one.

Equation (1.5.2) simply states the obvious fact that the mass flow-rate is the same on the two sides of the shock.

We now move to linear momentum. The concentration is ΦU, and we need to establish what the flux may be. There is certainly a term ΦU^2, which represents the convective flux (velocity U times concentration ΦU), but there is also a term representing the force orthogonal to X per unit area, i.e., the XX component of the internal stress. The elementary theory is based on the assumption that this component of stress is simply pressure, and therefore the flux of momentum is taken to be $p + \Phi U^2$.

Remark 1.5.2

Pressure needs of course not be zero at equilibrium. This sets momentum apart from mass and heat: at equilibrium, the heat flux and the mass flux are zero, but the momentum flux is not — though it is uniform in space (in the absence of gravity). In the presence of gravity, pressure has at equilibrium the hydrostatic distribution, and hence the momentum flux is nonzero and nonuniform. In solids at equilibrium, the momentum flux (or stress tensor) not only does not need to be zero, it does not even need to be isotropic.

Remark 1.5.3

Notice that the assumption that the XX component of stress is pressure is a constitutive assumption, i.e., in this case a constitutive equation is being used.

Remark 1.5.4

We are writing fluxes in a form which contains no equivalent of a diffusive flux, i.e., without gradients. The problem is thus going to be hyperbolic, and that is what makes discontinuities possible.

With this, the momentum balance becomes:

$$\partial(p + \Phi U^2)/\partial X = 0 \tag{1.5.3}$$

and Kotchine's procedure yields:

$$[p + \Phi U^2] = 0 = [p] + [\Phi U^2] \tag{1.5.4}$$

Remark 1.5.5

The operator [] is linear, and hence it commutes with all other linear operators, in particular with the sum.

Now notice that $\Phi U^2 = W^2/\Phi$, and that we have established that $[W] = 0$. Hence Eq. (1.5.4) can be rewritten as follows:

$$[p] = -[\mathrm{W}^2/\Phi] = -W^2[1/\Phi] \tag{1.5.5}$$

This implies that $[p]$ and $[\Phi]$ always have the same sign: either both pressure and density increase through the shock, or they both decrease. In the first case, the phenomenon is called a detonation; in the second case, a deflagration.

Now we may regard the gas on the left of the shock as unperturbed, and the shock is moving leftwards through it with a speed $V = U_L = W/\Phi_L$. A little algebra thus gives:

$$\Phi_L[1/\Phi^2]V^2 = -[p] \tag{1.5.6}$$

Suppose both $[p]$ and $[\Phi]$ are differentially small. Equation (1.5.6) yields:

$$V^2 = [p]/[\Phi] = \mathrm{d}p/\mathrm{d}\Phi \tag{1.5.7}$$

This result was obtained in the late 17th century by Newton, who was trying to establish the speed of sound, and argued that sound is nothing else but a small pressure discontinuity travelling through an unperturbed gas. In an ideal gas, $p = \Phi RT/M$, where M is the molecular weight (Newton did not know the ideal gas equation, but he knew how to calculate the isothermal pressure-temperature relationship for air). Thus he calculated $V = \sqrt{(RT/M)}$, which for air at ambient conditions is 261 m/s. Not too bad for his times, but in fact 23% lower than the experimentally determined value.

Newton was, perhaps, endowed with some engineering mentality, and 23% didn't seem too bad. In the second edition of the *Principia* he fudged the problem by adding vague ideas about the "crassitude of air particles", and he got the right result by such fudging. If we know what the right result is, fudging the theory so as to yield it is generally easy. It is not, however, logically satisfactory; 23% may not be that much, but it is enough to be significant, and we need to understand what is going on here.

In the early 19th century, Laplace reasoned as follows. At any given temperature, p is proportional to Φ for air at ambient conditions. However, if air is compressed very rapidly, temperature is not constant, and in fact one observes experimentally that:

$$\mathrm{dln}p = 1.67\,\mathrm{dln}\Phi \tag{1.5.8}$$

Now as the gas crosses the shock, its pressure certainly changes very rapidly, and hence one should use Eq. (1.5.8), thus yielding $V = \sqrt{(5RT/3M)}$. This yields 337 m/s, which is right within experimental accuracy for the velocity of sound in air at ambient conditions. Not too bad for the times.

Remark 1.5.6

> Both Newton and Laplace had no thermodynamic theory whatsoever available to them: the first and second laws of thermodynamics were formulated in the second half of the 19th century. This should make it abundantly clear that the results obtained so far have no thermodynamic content whatsoever: they are purely mechanical results, with the actual velocity of sound coming out of the experimentally determined p–Φ relationship corresponding to rapid compression or expansion.

Today, however, we do have thermodynamics available (though it still appears as a rather mysterious science), and so we may proceed to consider energy. The concentration of energy is $\Phi E + \frac{1}{2}\Phi U^2$, where the first term is the concentration of internal energy (E is internal energy per unit mass, not per mole), and the second one is the concentration of kinetic energy. The convective flux is thus $\Phi U(E + \frac{1}{2}U^2)$. What about the conductive flux? That would be the heat flux q, which is zero at equilibrium. But let us be consistent with ourselves: for momentum flux, we have assumed that the conductive flux is p, i.e., it is the same as at equilibrium. Well, let us make the same assumption for the conductive flux of energy, i.e., $q = 0$.

Remark 1.5.7

> The assumptions that the XX component of stress is p, and that $q = 0$, are very delicate ones, and will need to be discussed in detail. The first one follows, in the Maxwellian theory of gases, from the assumed kinematics, since the gas undergoes a pure compression (or dilation) when it crosses the shock, with no shear. The second one is a moot point even in the Maxwellian theory.

However, the balance of energy is a bit more tricky, since we have to take into account a generation term: work done on the system. Work is force times displacement, and thus rate of work is force times velocity. If expressed per unit area, work done at X is thus pU. Net rate of work done on the element between X and $X + \mathrm{d}X$ is therefore $-\partial(pU)/\partial X$. Having digested this somewhat confusing bit of information, we write the energy balance as follows:

$$\partial(\Phi U(E + \tfrac{1}{2}U^2 + p/\Phi))/\partial X = 0 \qquad (1.5.9)$$

But we have already established that $\partial(\Phi U)/\partial X = 0$, and we are so familiar with thermodynamics that we know that enthalpy H is defined as $H = E + p/\Phi$. Thus Eq. (1.5.9) reduces to:

$$\partial(H + \tfrac{1}{2}U^2)/\partial X = 0 \tag{1.5.10}$$

and Kotchine's procedure yields:

$$[H + \tfrac{1}{2}U^2] = 0 \tag{1.5.11}$$

Now a bit of algebra yields the result:

$$[H] = [p][1/\Phi^2]/2[1/\Phi] \tag{1.5.12}$$

We may now revisit the speed of sound from the vantage point of our understanding of thermodynamics. All jumps are infinitesimal, and hence Eq. (1.5.12) reduces to:

$$\mathrm{d}H = \mathrm{d}p/\Phi \tag{1.5.13}$$

But we usually delude ourselves into believing that we know not only the first, but also the second law of thermodynamics, and if we believe what thermodynamicists pull out of it, we believe the so called Maxwell relations (Maxwell would be horrified at the thought that they are ascribed to him). One of these relations is:

$$\mathrm{d}H = \mathrm{d}P/\Phi + T\mathrm{d}S \tag{1.5.14}$$

where S is that extremely mysterious quantity, entropy. Whatever entropy may be, if Eqs. (1.5.13) and (1.5.14) are both right, it has the same value on the two sides of the shock, i.e., $\mathrm{d}S = 0$. Thus one obtains:

$$V^2 = (\partial p/\partial\Phi)_S \tag{1.5.15}$$

Remark 1.5.8

Once thermodynamics enters the picture, partial derivatives bear subscripts, which they do not in any other branch of science. Why that should be so is not our problem here, and fortunately so.

Remark 1.5.9

There is a somewhat heuristic argument to convince ourselves that the derivative $\mathrm{d}p/\mathrm{d}\Phi$ is to be taken at constant entropy. Since we have set the heat flux to zero, and we have included no viscosity, we have pushed out of the problem all mechanisms of dissipation. Hence we are describing a "reversible" process, and thus anything adiabatic is isoentropic. There must be at least some readers who have no problem when they read an equation such as $\mathrm{d}S = \delta q_{\mathrm{REV}}/T$, and those readers are expected to find the argument offered above at least mildly convincing.

Now we go to our thermodynamics textbook and find out that the isoentropic derivative of pressure with respect to density can be expressed as:

$$(\partial \ln p/\partial \ln \Phi)_{\mathrm{S}} = c_{\mathrm{P}}/c_{\mathrm{V}} \tag{1.5.16}$$

where c_{P} and c_{V} are the constant pressure and the constant volume specific heats. We then may indulge in a bit of statistical mechanics to find out that, for a Maxwellian gas (a collection of molecules which are endowed with mass but no volume), $c_{\mathrm{P}}/c_{\mathrm{V}} = 5/3$, and, with mild surprise, we find out that Laplace was right after all. Maybe the concept of entropy may really serve some purpose, even when one is considering processes which not only are not at equilibrium, but are not even quasistatic in any mildly meaningful sense of the word: after all, our gas, when it crosses the shock, undergoes discontinuities of pressure, density and temperature, and if going through a discontinuity is quasistatic what is not?

Much as we might enjoy music, the speed of sound is not the only phenomenon we are interested in, and perhaps we should move to the case where $[p]$ and $[\Phi]$ are not infinitesimal. However, we have been as cavalier as possible about constitutive equations so far that we may as well keep being that way, and our Maxwellian gas theory tells us that:

$$H = 5p/2\Phi \tag{1.5.17}$$

Now we need a little algebra (again left to the reader) to reduce Eq. (1.5.12) to:

$$\Phi = (1 + 4\beta)/(4 + \beta) \tag{1.5.18}$$

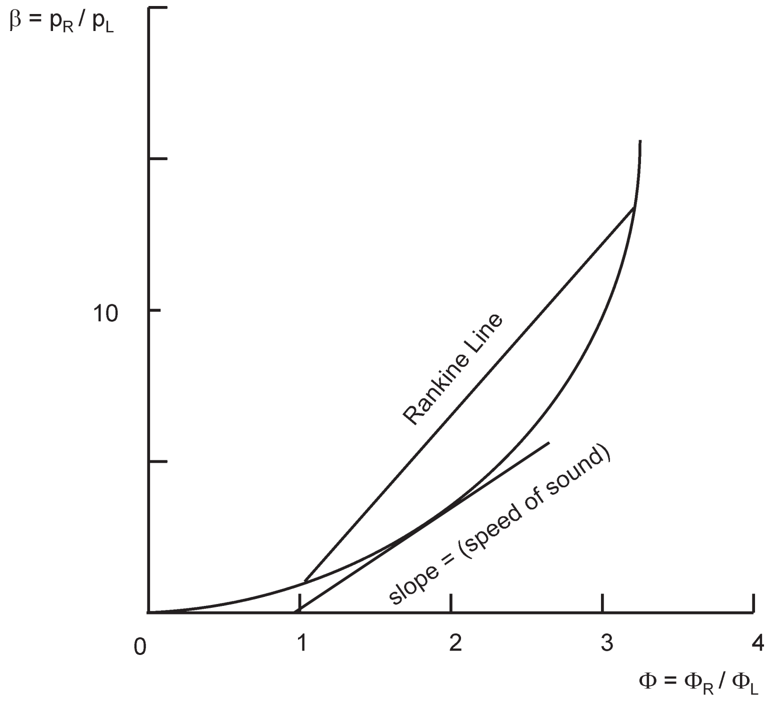

Fig. 1.5.2. Hugoniot curve for a Maxwellian gas undergoing no chemical reaction. The speed of propagation is the square root of the slope of the Rankine line.

where $\beta = p_R/p_L$ and $\phi = \Phi_R/\Phi_L$. Equation (1.5.18) is called a Hugoniot curve in the Φ–β plane. It goes through $\Phi=1$, $\beta = 1$, where it is tangent to an isoentropic curve (as we knew from the previous analysis); see Fig. 1.5.2.

Remark 1.5.10

Hugoniot obtained this result in the late 19th century. Scientists were not as afraid of discontinuities in those days as they are now; consequently, gas dynamics is a science which even today can deal with discontinuities without being regarded as awkwardly peculiar. Discontinuities of course arise with equal frequency in other branches of engineering science, but textbooks usually camouflage them in a variety of ways.

One may notice that the point β– which represents conditions to the right of the shock may be either to the right or to the left of 1–1. In the former case, one has a detonation, and the speed of propagation (the square root of the slope of the "Rankine line") is larger than the speed of sound; while in the latter case it is lower. Detonations propagate supersonically, but deflagrations propagate subsonically (if we believe this theory).

However, do we really believe Eq. (1.5.17)? We are still willing to be as cavalier as possible with regard to constitutive equations; but suppose a chemical reaction is taking place as the gas crosses the shock? (This is the problem of flames). The gas may be ideal, in fact Maxwellian, but still, if combustion goes to completion through the shock, it has a different composition on the right than it has on the left, and Eq. (1.5.17) does not account for that. Well, let Qp_L/Φ_L be the heat of reaction; now H on the left is given by Eq. (1.5.17), but H on the right is given by:

$$H = 5p/2\Phi + Qp_L/\Phi_L \tag{1.5.19}$$

We now need some more algebra to obtain:

$$\Phi = (1 + 4\beta)/(4 + Q + \beta) \tag{1.5.20}$$

which is again a Hugoniot curve. However, it does not go through 1–1 (see Fig. 1.5.3).

Now notice that if $\beta < 1$ everything is all right, because $\Phi < 1$, and we have a deflagration (which propagates subsonically). If $\beta > \dfrac{(3+Q)}{3} > 1$, everything is all right again because $\Phi > 1$ and we have a detonation (which propagates supersonically). But if $1 < \beta < (3 + Q)/3$, everything is by no means all right, because $\Phi < 1$, and that would give an imaginary speed of propagation. So now there is an unattainable region of the Hugoniot curve.

Remark 1.5.22

In actual fact, one is still at a loss about what Φ and β may be. The classical theory of flames relies on the Jouguet analysis which implies that the steady state solution is the "Jouguet line", i.e., a line going through 1–1 and tangential to the Hugoniot curve. It is not our purpose here to discuss combustion theory, and we leave this point, if of interest to the reader, to the literature on combustion.

All the results given above are standard ones, which can be found in any book on Gas Dynamics, though often much more algebra is used to obtain

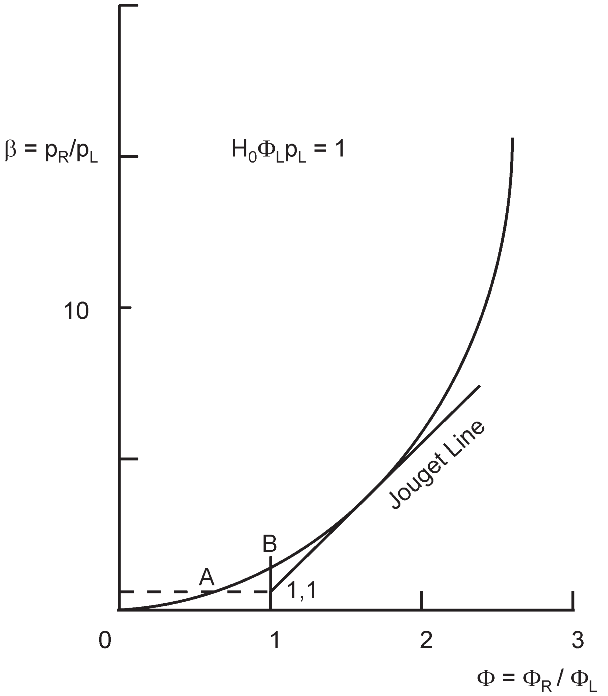

Fig. 1.5.3. Hugoniot curve for a flame. The part of the curve between A and B is unattainable. $Q = 3$ for the curve drawn. The actual values of β and Φ are obtained from the Jouguet line, which is tangent to the Hugoniot curve and goes through 1–1.

them. The point we wish to address here is the following one. Consider the gas on both sides of the shock. It has constant pressure, constant temperature, constant density, and constant composition. However, all four quantities have different values on the two sides of the shock. Would it be in any way repugnant to our understanding to regard the gases on the two sides of the shock as two different phases? After all, what more can be different between two phases, in addition to density, pressure, temperature, and composition? Indeed, in the classical phase equilibrium theory pressure and temperature are regarded as equal in the two phases, and the interface is a surface of discontinuity only for density and composition. If anything, our gases on the two sides of the shock are more different than two phases at equilibrium are.

And yet we have difficulty in accepting this idea — two gas phases, that just does not sound right. The problem is that, at equilibrium, one can never have two gas phases — but here we are considering a process which is about as far away from equilibrium as one can conceivably be. We may, therefore, at least tentatively, accept the following conclusion: "There may be instances where the very existence of two separate phases is sustained by a transport phenomenon".

Remark 1.5.12

Hopefully, the reader is by now convinced that the theories of transport phenomena and of thermodynamics are not mutually exclusive. They are, if anything, two aspects of the same theory.

Remark 1.5.13

At equilibrium, some quantity (temperature, velocity, chemical potential) does not undergo a discontinuity at the interface. It is generally assumed that neither does it undergo a discontinuity in the presence of transport phenomena. However, in the problem considered here velocity, temperature and chemical potentials are all discontinuous at the shock.

A final observation is in order. The equations of gas dynamics are nonlinear enough that not only an imposed shock will be sustained and propagate through the system, but shocks may develop without any being imposed. We are not going to develop the theory in any detail, but simply to point out the essential physics. Since the velocity of sound is given by Eq. (1.5.15), propagation of a pressure wave results in heating of the gas (pressure is more than proportional to density). But at higher temperature the speed of sound is larger — thus subsequent pressure changes travel more rapidly, and thus they catch up eventually with the preceding ones. This means of course that a shock may develop even if the imposed pressure at some boundary is continuous in time.

1.6 PURE DIFFUSION REVISITED

In this section, we reconsider the problem of Section 1.4 with a slight twist on it. The body considered is not semi-infinite, but it has thickness $2L$, and the same things are happening on both exposed surfaces. We therefore assume that the problem is symmetrical, and limit our attention to the region $0 \leq X \leq L$, with the hypothesis that, at $X = L$, the flux is zero (why should it be either positive or negative? Positive would mean pointing right, and negative pointing left, and if right and left are undistinguishable, it cannot point anywhere).

Remark 1.6.1

Problems may be such that the boundary conditions enjoy some kind of symmetry. We usually assume that, if such is the case, the solution to the problem enjoys the same kind of symmetry. One should, however, be careful with such an assumption, since sometimes it may backfire on us.

Since we are daring, however, we will assume symmetry of the solution, and thus we write the second boundary condition as:

$$X = L,\ \partial\theta/\partial X = 0 \tag{1.6.1}$$

in addition to Eq. (1.4.10).

Remark 1.6.2

In Section 1.4, the solution was sought in the positive $X - t$ quadrant, and BCs were written on the (proper) boundaries of that domain, i.e., $X = 0$ and $t = 0$. Here the solution is sought over the strip $t > 0$, $0 < X < L$, and hence a new proper boundary comes into the picture at $X = L$. (By "proper" we mean a boundary which is not at ∞.)

Now we are all familiar with integration by separation of variables, Fourier transforms, and the like, so (in principle) we have no problem in finding the solution. Those of us who are graduate students and have to take a closed book exam will sweat for hours, miss the right sign a couple of times, be driven to tears, and hope for the best at the end. Those of us who, by sheer luck, are beyond that stage, look up the solution in some book and hope that there are no typos, to find out that:

$$\theta = 1 - 2\Sigma_N(\exp(-A_N 2\tau)\sin(A_N y)/A_N) \tag{1.6.2}$$

where:

$$A_N = \pi(N + ½);\ y = X/L;\ \tau = tD/L^2 \tag{1.6.3}$$

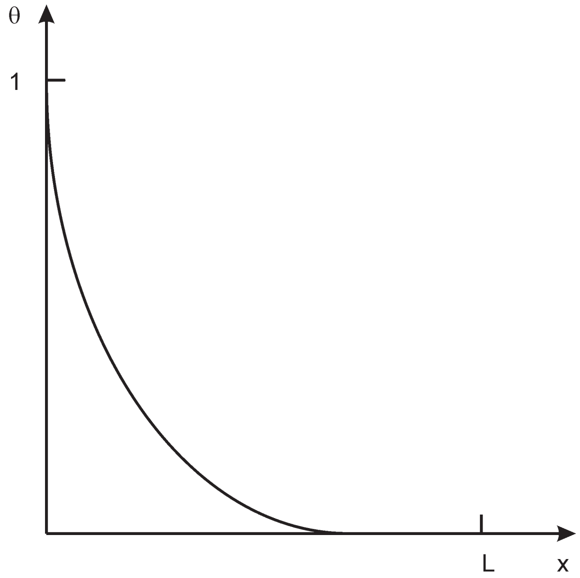

Fig. 1.6.1. Sketch of the solution at low values of τ.

The series in Eq. (1.6.2) always converges, or so we are told by mathematicians, and that makes those of us who worry about such things mildly happy. We then try to calculate θ at, say, $\tau = 0.001$ and $y = 0.5$, and we find out that there is something rotten in the state of Denmark. Actually, those of us who don't have to take closed book exams don't have to actually do the calculation either, so the same graduate student as before is going to sweat this out, but the point of the matter is that at $\tau = 0.001$ (or any such small value the professor may cunningly have thought of) it turns out that the series in Eq. (1.6.2) does converge, but with such an excruciating slowness that, to all practical purposes, it might as well not converge at all. Something must be done about this, and let us try to see what can be done.

What about Remark 1.3.5? Let's try to sketch out the solution of our problem. As τ approaches ∞, there is no doubt that θ will be close to unity, but that is not our problem. Our problem is: what does the solution look like at τ values which are much smaller than unity? Let's sketch it; see Fig. 1.6.1. There will be a θ wave travelling through our body, but, since $\tau << 1$, it has hardly had time to reach anything close to the midplane $x = 1$. θ will be different from zero only in the surface layer, and practically zero at any x which comes even close to being of order unity. We sketch that in Fig. 1.6.1, and we convince ourselves that in actual fact BC (1.6.1) is (albeit only approximately) satisfied at X values much smaller than L, and from then on it is satisfied for all values of X. But then, what about Remark 1.3.7? The solution to the problem discussed in Section 1.4, after all, approximately satisfies BC (1.6.1) at $X >> 2\sqrt{(Dt)}$, say at $x >> 2\sqrt{\tau}$, so why don't we use it? Don't tell the professor, of course. Tell him that you have developed a numerical code which works no matter how small τ is, and nobody will bother to check that you have actually done so.

If you have followed our advice, you have a lot of time left to do some thinking (after all, you didn't go to the computer at all), without your advisor knowing about that. That's a very useful time, since graduate students are usually so busy they can never think. First think of the fact that, although the differential equation is the same as in Section 1.4, this time the analytical solution (Eq. 1.6.1) is not a similarity one. Well, that does not surprise us at all, since we have had time to cogitate on Remark 1.3.6, and thus we know that this time indeed there is an external length scale. Fine, so far; but then we are going to use, at small values of τ, the solution obtained in Section 1.4, which is a similarity solution. What's going on here? A bit of thinking will help: as long as BC (1.6.1) is (approximately) satisfied at X values well below L, parameter L doesn't really play any role, and thus a similarity solution is indeed expected. We suggest this to be good enough grounds to go back to your advisor and give him the solution. You'll need the computer for some fancy graphics, but the whole logic of the problem is in your head (which helps concentrating on the fancy graphics while you are on the computer).

Should you find out that you still have some time for thinking, you may ask yourself a few intelligent questions. First, D may depend on θ, $D = D(\theta)$. Well, by now we are expert at making things dimensionless, and we quickly recognize that τ has to be defined as $tD(1)/L^2$. We now define a dimensionless diffusivity $d(\theta) = D(\theta)/D(1)$, do our algebra right, and obtain the following differential equation:

$$\partial(d(\theta)\partial\theta\,/\partial x)/\partial x = \partial\theta/\partial\tau \tag{1.6.4}$$

subject to:

$$\theta(0,\tau) = 1; \theta(x,0) = 0 \tag{1.6.5}$$

$$x = 1, \partial\theta/\partial x = 0 \tag{1.6.6}$$

Now, since we remember Remark 1.2.6, we ask ourselves whether we can say something about the solution in general, independently of what the $d(\theta)$ function may be. Let's give it a try. First, let's integrate Eq. (1.6.4) between $x = 0$ and $x = 1$ to obtain:

$$-\partial\theta / \partial x|_{x=0} = \frac{\mathrm{d}}{\mathrm{d}\tau}\int_0^1 \theta \mathrm{d}x \tag{1.6.7}$$

This result is beautiful, since we have got rid of the function $d(\theta)$. At $x = 0$, $\theta = 1$, and hence $d(\theta) = 1$. At $x = 1$, $\partial\theta/\partial x = 0$, and we don't care what $d(\theta)$ may be. That's why we thought about integrating over x. (Well, we like to think that's why we thought of it). Now does Eq. (1.6.7) make sense? Indeed it does. The instantaneous flux at the surface equals the instantaneous total rate of accumulation in our finite body. So far, so good. Now let's integrate over time, from $\tau = 0$ to τ (after all, we were lucky when we decided to integrate over space, so let's try our luck again). We obtain:

$$w = \int_0^1 \theta \mathrm{d}x = -\int_0^\tau (\partial\theta / \partial x)|_{x=0}\, \mathrm{d}\tau \tag{1.6.8}$$

This also makes sense. w is the total amount which has entered the body at time τ, and it is equal to the integral of the flux at the exposed surface over time. Now the solution of Eqs. (1.6.4)–(1.6.6) must be of the form $\theta = \theta(x, \tau)$, and hence Eq. (1.6.8) implies that $w = w(\tau)$. But this is something we can check by experiment: w is measurable, and if we plot it vs. t/L^2 curves for different sample thicknesses L should all superimpose. Well, it turns out that, in the case of sorption of low molecular weight solvents in polymers, curves of w vs. t/L^2 do not superimpose. We are intelligent enough by now to know that we cannot explain that sort of result by assuming strange forms of the $d(\theta)$ function: the curves should superimpose no matter what the $d(\theta)$ function may be. We'll have to do something about our basic assumptions — more about this later.

Remark 1.6.3

The result that $w = w(\tau)$ has been obtained without actually solving Eq. (1.6.4). Thus the experimental check can be done before trying to find such

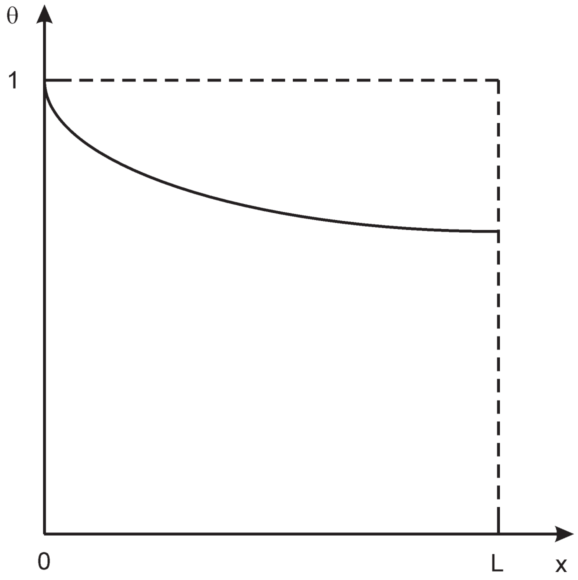

Fig. 1.6.2. Sketch of the solution at large values of τ.

> a solution. If the experiments do not check, the model is wrong, and what would be the point of finding a solution to the equations of a wrong model?

Now let's look at another facet of the problem. Let's ask ourselves what the solution of the problem might be at large values of τ (forget about Eq. (1.6.2) for the time being). Again we sketch what we expect in Fig. 1.6.2. The wave has now had ample time to reach the midplane, and in fact θ(1) is significantly larger than zero. The shape of the curve is perhaps the same at all times, but for a rescaling of the θ axis? Wait a moment: what makes some sense is that 1–θ as a function of x has the same shape at all times. Let's try out this idea; we write tentatively:

$$\theta = 1 - f(x)g(\tau) \tag{1.6.9}$$

Remark 1.6.4

> The reader has certainly understood already that we are trying a solution by separation of variables. Yes, we are, and the form of the separation has been suggested by our sketch. Should Eq. (1.6.9) turn out to be right, we would have obtained a similarity solution for 1–θ, not for θ.

Now the initial condition has no hope of being satisfied, since we are expecting (1.6.9) to hold, if at all, only at large times. We substitute into the differential equation and do the usual separation of variables algebra to obtain:

$$f''/f = g'/g = -K \tag{1.6.10}$$

where K is a constant to be determined (we have put in a minus sign because we know that $g' < 0$). We have boundary conditions for f, i.e., $f(0)=0, f'(1) = 0$, but we do not have any for g. We first deal with $f''/f = -K$, and (having carefully forgotten about higher harmonics), we conclude that $K = \Pi^2/4$. Thus we obtain:

$$\theta = 1 - B\sin(\Pi x/2)\exp(-\Pi^2\tau/4) \tag{1.6.11}$$

with B an undetermined constant, since we lack a boundary condition for g. But this is the leading term in Eq. (1.6.2), and the whole picture starts making a bit more sense than blind application of Fourier's transform technique would provide. In fact this is the point where we realize that we have forgotten about higher harmonics, and that in fact any solution of the following form would do:

$$\theta = 1 - B_N\sin(A_N x)\exp(-A_N{}^2\tau) \tag{1.6.12}$$

with the A_Ns given by Eq. (1.6.3) (each one of these solutions, except the one for $N = 0$, has a very peculiar shape for the problem at hand — we weren't so silly when we forgot about higher harmonics). Now we observe that the differential equation (Eq. 1.4.9) is linear, and hence any linear combination of solutions is also a solution, and what is Eq. (1.6.2) but a linear combination of solutions of the type of Eq. (1.6.12)? Equation (1.6.2) tells us that we have to choose the constants B_N as $2/A_N$ in order to satisfy the initial condition. But at $\tau = 0$ the sum in Eq. (1.6.2) must have a value of unity —and it takes an awful lot of harmonics to represent unity as a series of sines. So now we know why trying to use Eq. (1.6.2) at $\tau = 0.001$ was an almost hopeless task.

Remark 1.6.5

When the solution of a problem is in terms of an infinite series, it is generally quite useful for one of the asymptotic conditions (in the case at hand, $\tau >> 1$). In that region, often one does reasonably well with just the

> leading term in the series. A slight improvement is obtained with the second term, but a law of diminishing returns rapidly sets in. When there is no exact solution available, an asymptotic one may be obtainable. Trying more than a first order correction is not very useful, and the other asymptote is not going to be reached this way.

1.7 THE STEFAN PROBLEM

In 1891, Stefan got interested in the rate of formation of ice layers in polar seas. He reasoned as follows. Suppose the water has a uniform temperature of 0°C, but no ice has formed yet. Suppose now the ambient air suddenly gets to some subfreezing temperature J/Φ (with $J < 0$), and stays at that temperature thereafter. Let's forget, for the sake of simplicity, about heat transfer resistance in the ambient air, so that a jump J/Φ is in fact imposed on the exposed surface. An ice layer will start forming, with thickness $S(t)$ — the latter function being the unknown of the problem. The physics are as follows: as the ice layer thickens, the latent heat of crystallization is developed at the freeze surface; this must be carried away by unsteady heat conduction through the ice layer.

Stefan assumed that the temperature at the freezing surface was 0°C at all times. Now this assumption is open to some question. 0°C is the equilibrium temperature of the ice–water transition. Is it also the temperature at the freeze surface? Is no undercooling needed to drive the freezing phenomenon at some finite rate? All these questions will need to be addressed in due time, but for the time being we go ahead with the classical formulation of the Stefan problem. The assumption discussed above implies that the residual body of water stays at 0°C at all times, so we may forget about it.

Remark 1.7.1

> When beginning with a problem, one often is willing to assume that the constitutive equations which hold at equilibrium also hold under non-equilibrium conditions — unless by making such an assumption one loses the problem altogether.

Actually, there is one more subtlety to note. That temperature on the water side of the freezing surface stays at 0°C is perhaps reasonable; but does it stay at 0°C also on the ice side of the same surface? Well, temperature doesn't usually exhibit discontinuities, or so we have been led to believe. Though perhaps now the material discussed in Section 1.5 induces some doubts.

Remark 1.7.

The freeze surface is obviously an interface between two phases. At equilibrium, temperature would be uniform in each one of the two phases, and it would be the same in the two phases. The assumption here is that, although it would not be uniform under nonequilibrium conditions, it would still be continuous — though its gradient is allowed to undergo a discontinuity at the interface. We will refer to this in the following as the Fundamental Interface Assumption (FIA).

We are now in a position to formulate the problem mathematically. The differential equation for temperature U is:

$$k\partial^2 U/\partial X^2 = \Phi \partial U/\partial t \tag{1.7.1}$$

where k is the thermal conductivity and Φ is the specific heat per unit volume; k/Φ is of course the diffusivity of heat. At time zero, $S = 0$, and hence the solution is sought in the domain $S(t) \geq X \geq 0$. Thus we write the boundary conditions:

$$X = 0, t > 0;\ \ U = J/\Phi < 0 \tag{1.7.2}$$

$$X = S(t), t > 0;\ U = 0 \tag{1.7.3}$$

This would all be nice, if we knew what $S(t)$ is, but we don't. We need to write additional equations so that the function $S(t)$ is determined. So let's write a balance at $X = S(t)$. If Γ is the change of enthalpy at freezing per unit volume, so that $-\Gamma$ is the latent heat per unit volume, and the freeze surface advances at a rate $\mathrm{d}S/\mathrm{d}t$, $-\Gamma \mathrm{d}S/\mathrm{d}t$ is the rate of latent heat generation per unit area of freeze surface. This must be carried away by the local heat flux $k\partial U/\partial x$, and hence:

$$X = S(t),\ k\partial U/\partial X = -\Gamma \mathrm{d}S/\mathrm{d}t \tag{1.7.4}$$

(recall that Γ is negative, and $\partial U/\partial X > 0$, so that F is leftward). This, together with the boundary condition $S(0) = 0$, constitutes a well-posed mathematical problem, belonging to a class which mathematicians call free boundary

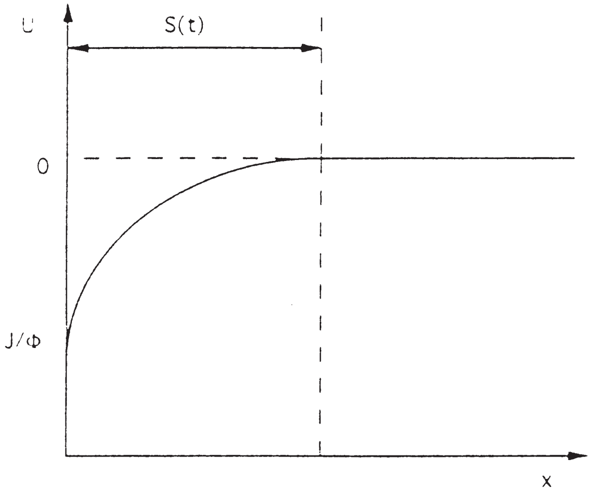

Fig. 1.7.1. Sketch of temperature profile for the classical formulation of the Stefan problem.

problems — problems where one of the boundaries is "free" in the sense that it is not known *a priori*, and has to emerge from the solution of the problem. The classical formulation of the Stefan problem is sketched in Fig. 1.7.1.

Before proceeding any further, it is useful to observe that $X = S(t)$ is clearly a surface of discontinuity, and we would have expected Eq. (1.7.4) to follow from Kotchine's procedure. Well, Kotchine's procedure would yield $[F] = [C]\mathrm{d}S/\mathrm{d}t$. Now F on the right is zero, and F on the left is $-k\partial U/\partial X$, so the left hand side of (1.7.4) is indeed $[F]$. So we would identify Γ with $-[C]$ — does that make sense? Indeed it does. C would be the concentration of enthalpy. Temperature is the same on the two sides of the interface, but there is water on the right and ice on the left, and the enthalpy of water exceeds that of ice by $-\Gamma$.

Remark 1.7.3

The question of applying Kotchine's procedure to Stefan type problems is quite a bit subtler than this, but for the time being what has been said above is enough.

Now let's try to make everything dimensionless. $\theta = U\Phi/J$ is fine, since that way $\theta = 1$ at the exposed surface and zero at the freeze surface. Let's multiply Eqs. (1.7.1) and (1.7.4) by J/Φ. Now (1.7.4) produces a dimensionless group

$\Gamma/J\Phi$ which it is all too natural to call the Stefan number, **St** (both Γ and J are negative, so **St** is nicely positive). The problem has reduced to:

$$D\partial^2\theta/\partial X^2 = \partial\theta/\partial t \tag{1.7.5}$$

$$\theta(0,t) = 1;\ \ \theta(S(t),t) = 0 \tag{1.7.6}$$

$$X = S(t), -D\partial\theta/\partial X = \mathbf{St}\,\mathrm{d}S/\mathrm{d}t;\ S(0)=0 \tag{1.7.7}$$

We now wish to make t and X dimensionless, but we are short one parameter, and thus we find out that a similarity solution is sure to come out. The Neuman variable $z = X/2\sqrt{(Dt)}$ is going to work again. But this means that S itself must be proportional to $\sqrt{(Dt)}$, say:

$$S = 2K\sqrt{(Dt)} \tag{1.7.8}$$

with K a constant to be determined.

Remark 1.7.4

Notice that the fact that S is proportional to $\sqrt{(Dt)}$ has been obtained without finding the solution of the problem.

The problem has reduced to:

$$\mathrm{d}^2\theta/\mathrm{d}z^2 + 2z\mathrm{d}\theta/\mathrm{d}z = 0 \tag{1.7.9}$$

$$\theta(0) = 1;\ \theta(K) = 0 \tag{1.7.10}$$

$$\theta'(K) = -2K\mathbf{St} \tag{1.7.11}$$

This is a second order differential equation subject to three BCs, and hence the value of K can be extracted. The solution is obtained from (1.7.9) and (1.7.10):

$$\theta = 1 - erfz/erfK \tag{1.7.12}$$

and when this is substituted in (1.7.11) the value of K is determined:

$$\exp(-K^2) = \mathbf{St}\sqrt{\Pi}K erfK \tag{1.7.13}$$

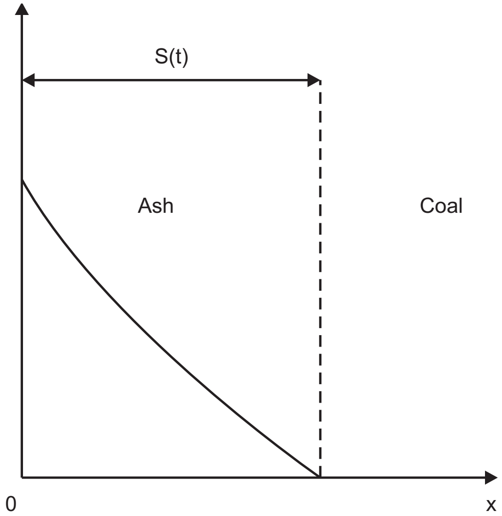

Fig. 1.7.2. Combustion of a slab of coal. The region at $X < S(t)$ is ash; the region at $X > S(t)$ is unburned coal. The curve is a sketch of the oxygen concentration profile.

Now let us consider an entirely different problem: combustion of a coal particle. We do not wish to tackle spherical coordinate systems so far, so we are going to consider the combustion of a coal slab. We assume that the combustion reaction is irreversible and instantaneous, which means that oxygen and coal can never coexist. So an ash layer is formed, through which oxygen must diffuse to the inside coal layer; the ash layer has thickness $S(t)$, with $S(0) = 0$ (see Fig. 1.7.2).

At the combustion surface $X = S(t)$ we assume that the oxygen concentration is zero — does that make sense? Well, by our assumptions oxygen and coal cannot coexists, and hence the oxygen concentration is zero to the right of $S(t)$, and so is the oxygen chemical potential. At equilibrium, chemical potential is the same in all phases — just as temperature. But we did assume that temperature, under nonequilibrium conditions, does not undergo discontinuities, so we make the same assumption for the chemical potential (this is the FIA again, see Remark 1.7.2). So the chemical potential (and thus also the concentration) of oxygen is zero to the left of $X = S(t)$, and it has a constant value J at $X = 0$. We normalize concentration with J, and thus we write again Eq. (1.7.5) for the unsteady diffusion of oxygen through the ash layer. BCs (1.7.6) also apply again. What about (1.7.7)?

Remark 1.7.5

> We have now encountered two instances of the FIA. For heat transfer, temperature is the same on the two sides of an interface — we don't even think about that, we just call it the temperature at the interface. For mass transfer, chemical potential is the same on the two sides of an interface — we do think about that, because we work with concentrations and concentration *is not* the same on the two sides of an interface. What about momentum transfer? The velocity is the same on the two sides of an interface — we always use the no-slip condition as a boundary condition in momentum transfer problems.

The flux of oxygen at $X = S(t)$ towards the combustion surface is $-DJ\partial\theta/\partial X$. This must be equal to the amount of coal per unit surface and time which is being burned. If Γ is the molar density of coal, the rate of combustion per unit surface is $\Gamma \mathrm{d}S/\mathrm{d}t$. We now give in to the temptation of defining **St** as Γ/J, and, lo and behold, Eq. (1.7.7) comes out again (with, this time, both Γ and J being positive). This may be a combustion problem, but it is described by exactly the same equations as the Stefan problem, and it thus has the same solution. (In the formulation discussed above, the literature on noncatalytic gas solid reactions refers to this model as the sharp interface one).

Let's try a third one. When low molecular weight solutes are absorbed in glassy polymers, often swelling occurs. Solubility in the glassy polymer is almost zero, and we'll take it to be zero; at sufficiently low activities of the solute in the external phase, practically nothing gets absorbed. As the activity is increased, at some point swelling is observed (at equilibrium), and at that point the solubility in the swollen phase is, say, Γ. Now suppose we expose a slab of polymer to some activity in excess of the threshold value, so that the solubility is now $\Gamma + J$. We normalize concentration as follows:

$$\theta = (C - \Gamma)/J \tag{1.7.14}$$

so that $\theta = 1$ at the exposed surface. Now presumably there might be a surface located at $X = S(t)$ where the swelling occurs, and presumably the concentration at that surface is Γ — after all, that is the solubility corresponding to the threshold activity for swelling. This makes θ at $X = S(t)$ equal to zero. So far, so good; we have recovered Eqs. (1.7.5–6) again. This time there is no doubt that $\Gamma = -[C]$, and thus $[F] = -\Gamma \mathrm{d}S/\mathrm{d}t$ (by now Kotchine's procedure has no more secrets for us). But there is no flux in the glassy part, and thus $[F] =$

$-F_L = D\partial C/\partial X = (D/J)\partial\theta/\partial X$. We are experts by now, and we define **St** as being Γ/Jagain — and here Eq. (1.7.7) springs up again. Stefan problems are coming out here, there, and everywhere (there will be a sizable prize for anyone who can come up with a physically not entirely absurd Stefan problem in momentum transfer. The problem should be describable exactly by Eqs. (1.7.5–7).

Now let's first ask ourselves whether all this makes any sense at all. Ice layers do indeed form at a rate which is inversely proportional to the square root of time, and under some appropriate conditions the same can be said about combustion of coal particles (in the appropriate spherical geometry version of the problem). We are not so lucky with polymers, where $S(t)$ has been measured, and has often been observed to be of the type $S = Kt^n$, with $1 \geq n \geq \frac{1}{2}$. By now we know that polymers are strange animals, so we do not worry too much about this. Instead, we try to understand more deeply the similarity between the combustion and the freezing versions of the problem.

In the combustion problem, we needed to make the assumption that the reaction is both instantaneous and irreversible, but we had no such need in the freezing problem. What's going on here?

In the case of a chemical reaction, one may define an "extent of reaction", w. Let's say that $w = 0$ corresponds to coal, and $w = 1$ corresponds to ash. What does the statement that the reaction is irreversible mean? It means that, at equilibrium, w will be unity as long as there is any oxygen left. That could well not be the case; for a reversible reaction, the equilibrium value w^* could well be different from unity.

Remark 1.7.6

The extent of reaction w is what is called an internal state variable in thermodynamics. The state depends on w, but we cannot impose the value of w arbitrarily and independently of the values of the external state variables (such as volume and temperature) which we choose to impose. Hence the possibility of defining an equilibrium value of w, w^*, which is the one corresponding to the situation where $\partial w/\partial t = 0$ if things like volume and temperature are held constant.

Now let us give some consideration to kinetics. In actual fact, at any instant in time w may be different from w^*. Then the reaction would proceed at some rate which, other things being equal, would depend on how far from equilibrium we happen to be, say:

$$\partial w/\partial t = R(w^* - w);\ R(0) = 0;\ R' > 0 \tag{1.7.15}$$

Chemical kinetics are not so accommodating as transport phenomena, and the $R()$ function is seldom, if ever, linear; but let's not worry about this for the time being. Let's rather consider that $R()$ has dimensions of a frequency, and we can thus write $R() = r()/\beta$, require $r(w^*)$ to be unity, and identify β with the intrinsic time scale of the reaction at $w = 0$. Now we can attach some meaning to the statement that the reaction is instantaneous: in some sense to be made rigorous later, we are saying that β is almost zero, say it is much smaller than any other time scale which may be relevant in our problem. If that is the case, it is likely that, in some sort of zero order approximation, we can approximate w with w^*. We will need to discuss all this in much more detail, of course, but let us first ask ourselves what the equivalent of w and β may be in the freezing problem.

Remark 1.7.7

> We see already that singular limits are going to creep up on us. β approaching zero — it is easy to envisage cases where the coefficient of the highest order derivative of some differential equation is proportional to β.

In fact, let us look at the combustion problem again, to realize that our assumptions were essentially that $w = 0$ to the right of the travelling discontinuity, and $w = 1$ to the left of it — the reaction has not taken place at all in the unburned coal core, and it has gone to completion in the ash layer. This is rather reminiscent of the last part discussed in Section 1.5 — a chemical reaction goes to completion within an extremely thin layer which we model as a surface of discontinuity. In Section 1.5, the fact that the gases on the two sides of the discontinuity could be regarded as two distinct phases was offered as a suggestion, and one which is rather repugnant to our preconceived ideas; here, that coal and ash are two different phases seems a much more reasonable hypothesis.

Remark 1.7.8

> The problem we are facing is: what do we mean by two different phases? We will need an appropriate definition, applicable to cases where transport phenomena are taking place, not only to equilibrium.

Now what could the equivalent of w be in the freezing problem? Well, it seems obvious that we should consider a degree of crystallinity — which is unity in the ice layer, and 0 in the unfrozen water core. What is the analog of the assumption of irreversible chemical reaction? That w^* is either zero (at $\theta \leq 0$) or 1 (at $\theta \geq 0$). In the combustion problem, w^* is zero only at zero oxygen concentration (θ cannot become negative), unity at nonzero oxygen concentrations. For those of us who like formality, we might write:

$$w^* = H(\theta) \tag{1.7.16}$$

where $H()$ is the Heaviside step function. This applies to both the combustion and the freezing problem.

Remark 1.7.9

> It will turn out that writing Eq. (1.7.16) is more than a formality. The derivative of the Heaviside function is Dirac's delta function, and that will play a crucial role in the generalized formulation of Kotchine's "balance across discontinuities" equation.

Now let us move up one step and ask ourselves the following question: what is the analog of the assumption of instantaneous reaction? It is, of course, the assumption that as soon as the actual value of θ exceeds zero, even by a differential amount, the water gets entirely frozen — there is no kinetics of the crystallization phenomenon in the Stefan problem. Indeed, one now recognizes that the assumption that temperature is 0°C at the freeze surface is essentially an assumption of "instantaneous reaction". And, in the freezing problem, there is no doubt whatsoever that the travelling discontinuity is indeed the interface between two phases. We have a strong reticent to regard two gases as two phases, mild awkwardness at regarding two solids (coal and ash) as two phases, but water and ice — that's familiar ground.

Well, we have got ample food for thought so far. In the next section, we are going to do some of the more complex thinking. Right now, we simply go back to our analysis in Sections 1.4 and 1.6 to make the following observation. The finite-body problem considered in Section 1.6 had a small time asymptotic solution given by the semi-infinite body solution. The smaller the time, the better was the approximation of using the Stokes–Rayleigh problem solution — but it was never absolutely rigorous. Here we have a different situation: if we consider the finite-body Stefan problem, the solution of the semi-infinite case is *exactly* right until the travelling discontinuity reaches the half-thickness of the sample. However, had we started directly by considering the finite-body problem, we would never have guessed the existence of a similarity solution. This consideration also offers food for thought.

1.8 KINETICS

The analysis in the previous section has suggested that one needs to consider kinetics not only in the case of chemical reactions, but possibly in a variety of other cases, and in fact whenever an internal state variable w can be identified. Furthermore, we have seen that surfaces of discontinuity may exist across which the potential U is continuous, but the concentration C is not: in the freezing case, the temperature was the same on the two sides of the interface, but enthalpy was not. In the coal combustion problem, the oxygen concentration was zero on both sides of the interface, but something funny was still happening, since its flux was finite on the left of the interface, but it was zero on the right. Not quite the same thing as in the former case, it would appear; so let us look at the situation in a little more detail.

First of all, since the flux of oxygen is finite on the left and zero on the right, where does the oxygen reaching the combustion surface end up? It ends up, of course, consumed by the chemical reaction. This observation points the way for formulating this type of problems in general. Since oxygen can be consumed by a chemical reaction, there is no principle of conservation available for oxygen itself. So we should be careful with the balance equation. Indeed, let us consider a chemical reaction of the type:

$$\mathbf{B} + \mathbf{C} = \mathbf{A} \tag{1.8.1}$$

Remark 1.8.1

Of course, in general there will be stoichiometric coefficients other than unity. Those would simply clutter the algebra at this stage.

Reaction (1.8.1) is a typical "academic" reaction (reactants and products don't have any chemical formula) but there is a somewhat unusual arrangement of letters: this is due to the fact that, in the back of our minds, **C** is coal, **A** is ash, and — by default — **B** is oxygen. Now when the chemical reaction takes place in either direction, the number of moles of any one of the three components is not conserved, but some appropriate sum of number of moles of reactants and products will. In the case of Eq. (1.8.1), for example, the sum **B** + **A** stays constant. So we can write a balance equation for the sum of the concentrations of **B** and **A**, which we will indicate with C — the concentration which appears in the balance equation. C is the real counterpart to the concentration of enthalpy in the freezing problem, since it is a quantity for which a conservation principle is available. However, let's be careful. Since we are dealing with a mass transfer problem (albeit one which occurs in the presence of a chemical reaction), the concentration of the diffusing species is the potential U for diffusion, where U is not C, it is the concentration of the diffusing component (oxygen in the coal combustion case). Well, let's go ahead, and let C_A be the concentration of the reaction product **A**.

Now the case considered in the preceding section, coal combustion, is one where only **B** can diffuse, so that the flux of **B** + **A** and the flux of **B** coincide with each other. Let's focus on this special case.

Remark 1.8.2

> The case is not so special as it may appear at first sight. For instance, oxidative corrosion of a metal falls into the same category, with **C** being, say, iron, **A** being rust, and **B** being again oxygen. Instances of the same type occur even in cases where reactants and products may all diffuse.

The nondiffusing reactant, **C**, has initially some concentration Γ. If $C_A = 0$, the reaction has not taken place at all; if $C_A = \Gamma$, the reaction has gone to completion. It is therefore natural to define an extent of reaction w as follows:

$$w = c_A/\Gamma \tag{1.8.2}$$

and the total concentration C is:

$$C = U + w\Gamma = J(\theta + w\mathbf{St}) \tag{1.8.3}$$

Now let's look at the freezing problem. The quantity which is diffusing is heat, and the potential for diffusion is temperature U. The system is at constant pressure, and hence the quantity for which a balance equation can legitimately be written is enthalpy, the concentration of which is indicated again with C. As long as we stay in one particular phase (either ice or water, but not both), one has:

$$C = \Phi U \tag{1.8.4}$$

where Φ is the specific heat per unit volume.

Remark 1.8.3

> Of course, there is no reason why Φ should be a constant, and even less that it should have the same value in the two phases. At this stage, we make as

many simplifying assumptions as possible, and thus we are willing to regard Φ as constant and equal in the two phases. We are trying to look at the essential features of the problem, after all.

Remark 1.8.4

The preceding remark suggests a more general consideration. In the initial stages of modelling, all those simplifying assumptions which (a), are not patently absurd, and (b) don't result in losing the problem altogether, should be made. The essential features of the model will thus be more clearly elucidated.

However, the enthalpy of water is larger than that of ice by an amount equal to the latent heat $-\Gamma$. Now we want the initial concentration of the quantity for which we are going to write a balance to be zero, so we normalize enthalpy in such a way that water at 0°C has zero enthalpy. The enthalpy concentrations in the two phases are:

$$C(\text{ice}) = \Phi U + \Gamma;\ C(\text{water}) = \Phi U \tag{1.8.5}$$

(recall that $\Gamma < 0$). Now if we want to carry the analogy with the combustion case a bit further, we observe that in our case the "reaction" is something like water + (–heat) = ice, and it has not taken place at all in the water phase, while it has gone to completion in the ice phase. It is therefore natural to define the extent of reaction w as the fraction of ice, or degree of crystallinity. This may seem a bit artificial, but for instance in the case of polymers the concept of a degree of crystallinity is a commonly used one; with H_2O, admittedly, we are accustomed to think that it is either water or ice, and no wishy-washy middle case is legitimate. Still, let's go ahead — nothing wrong that the variable w is expected to be able to take up only the values 0 and 1. After all, in the coal combustion case we did make the same assumption, that the solid is either coal or ash.

Remark 1.8.5

This paves the way for the discussion in the next section. It is often useful to consider a variable w which identifies the phase one is considering. This will have values of 0 in one phase, 1 in the other phase, and values in between may conceptually be used to describe a possibly very thin "interface layer".

With this, the concentration of enthalpy in both phases (as well as, conceptually, for conditions where $0 < w < 1$) can be written as:

$$C = \Phi U + w\Gamma = J(\theta + w\mathbf{St}) \tag{1.8.6}$$

which is exactly in the same form as Eq. (1.8.3): the analogy between the two problems is becoming more transparent.

Let's proceed to a bit of generalization. Contrary to the problems of pure diffusion, where C was a function of only U, here it is a function of two variables, U and w:

$$C = C(U,w) \tag{1.8.7}$$

Well, we aren't going to be afraid of moving from one to two variables, are we? So let's define:

$$\Phi = \partial C/\partial U;\ \Gamma = \partial C/\partial w \tag{1.8.8}$$

and for the time being let us assume that Φ and Γ are constants. We then recover Eq. (1.8.6). (For the freezing problem, C decreases as w increases, and hence $\Gamma < 0$).

When all this is substituted in Eq. (1.1.1), one obtains:

$$\partial F/\partial X + J\partial\theta/\partial t = -J\mathbf{St}\partial w/\partial t \tag{1.8.9}$$

and if we keep assuming that something like Fourier or Fick's law applies:

$$F = -DJ\partial\theta/\partial X \tag{1.8.10}$$

Remark 1.8.6

Equation (1.8.9) can be interpreted in two different ways. The first one is the one we have discussed, where it is the balance of a quantity for which a conservation principle is available (enthalpy, the sum of a reactant and a product concentrations). The alternate interpretation is that it is the balance equation for a quantity for which there is no conservation principle available (heat, reactant **B**). In this second interpretation, the RHS is the rate of generation. In the freezing case, $J < 0$, and indeed one has a positive generation of heat. In the combustion case, $J > 0$, and one has a negative generation of oxygen.

Remark 1.8.7

> In the general case, D, Γ and Φ (and hence **St** as well) are going to depend on both θ and w. But we are still in the spirit of Remarks 1.8.3 and 1.8.4, and we will regard them as constant.

Now however we have one more dependent variable than we had before (w), and thus we need one more equation. This is, of course, the constitutive equation for the kinetics, say:

$$\partial w/\partial t = r(w, \theta)/\beta \tag{1.8.11}$$

Now the function $r()$ is subject to some constraints. First, it must be such that its value becomes zero when w has the equilibrium value $w^*(\theta)$:

$$r(w^*(\theta),\theta) = 0 \tag{1.8.12}$$

which introduces an additional constitutive function, $w^*(\theta)$, in our problem — one needs to know the equilibrium value of the internal state variable as a function of the potential (in the coal combustion problem, one needs to know the degree of conversion of coal to ash as a function of temperature; in the water freezing problem, $w^* = 1$ when temperature is less than 0°C, and $w^* = 0$ otherwise). Hopefully, $w^*(\theta)$ should be obtainable from classical thermodynamics. Next, we consider a problem where at time zero $w = 0$ (or, alternately, we normalize w so that it is zero at time zero). We then choose the time scale β so that:

$$r(0,1) = 1 \tag{1.8.13}$$

which guarantees that β is the intrinsic time scale of the kinetic phenomenon at the exposed surface at time zero.

Remark 1.8.8

> Notice that an intrinsic time scale has been defined without assuming that the function $r()$ enjoys any kind of linearity (the reaction does not need to be a “first order” one). Of course, the value of β will depend on the initial condition (since w has been normalized to be zero at time zero) as well as on the imposed jump J/Φ.

Finally, the boundary conditions: since $\partial w/\partial t$ appears in our equations, we need, in addition to the usual ones, an initial condition for w, which we have just finished discussing. Thus we write:

$$F(X,0) = \theta(X,0) = w(0) = 0 \tag{1.8.14}$$

$$\theta(0,t) = 1 \tag{1.8.15}$$

Now we are ready to apply Kotchine's procedure. Equations (1.8.11–13) yield:

$$[F] = V^*([\theta] + \mathbf{St}[w]) \tag{1.8.16}$$

$$[\theta] = 0 \tag{1.8.17}$$

$$[w] = 0 \tag{1.8.18}$$

so that everything is smooth and nice, no discontinuities of anything, nothing of even mild interest — at first sight. If the Stefan problem is to emerge in any asymptotic sense, we can only hope in some singular limit (or hope that there is none, as the tastes may be).

We have been careful so far in making things dimensionless at a very slow pace, and in fact we still have the dimensional quantities F, X, t, D, V^*, and β in our equations. We now want to introduce a dimensionless time and a dimensionless distance, but we are going to be very careful in doing so.

Let's first try a brutal choice. We have an intrinsic timescale β available to us, so $\tau = t/\beta$ appears as a very reasonable choice (but perhaps we foresee some trouble if β is going to approach zero, as the astute reader has by now understood it will). Let's plunge even deeper, and let's consider the finite body problem, with $\partial\theta/\partial X = 0$ at $X = L$ as the additional BC, so that $x = X/L$ is also a very natural choice. Now we have a choice on how to normalize flux, and we choose as follows:

$$f = F\beta/LJ \tag{1.8.19}$$

(Notice that f has the same sign as F/J. If $J < 0$, $f > 0$ means that F is leftwards). We further define the Juttner number **Ju** as:

$$\mathbf{Ju} = \beta D/L^2 \tag{1.8.20}$$

Remark 1.8.9

If $1/\beta$ is interpreted as a pseudo first order kinetic constant for a chemical reaction, the Juttner number is identified with the inverse of the square of the Thiele modulus arising in the theory of diffusion in porous catalysts. In actual fact, the latter problem was considered, and essentially solved, by Juttner much earlier [1909, 1910] than by Thiele [1939].

With this, the canonical set of our problem has become:

$$\partial f/\partial x + \partial\theta/\partial\tau = -\mathbf{St}\partial w/\partial\tau \tag{1.8.21}$$

$$f = -\mathbf{Ju}\partial\theta/\partial x \tag{1.8.22}$$

$$\partial w/\partial\tau = r(w, \theta) \tag{1.8.23}$$

with BCs given by:

$$f(x,0) = \theta(x,0) = w(0) = 0 \tag{1.8.24}$$

$$\theta(1,\tau) = 1 \tag{1.8.25}$$

$$x = 1,\ \ \partial\theta/\partial x = 0 \tag{1.8.26}$$

We have two dimensionless groups, **St** and **Ju**. If both are of order unity, everything is smooth and uninteresting; should the reader enjoy such things, now is the time to go to the computer and solve the problem numerically, for several different forms of the functions $w^*(\theta)$ and $r(w,\theta)$. Astarita and Kenny [1987] have published some numerical results for the case where $w^*(\theta)$ is given by Eq. (1.7.16), $r = \theta(1-\theta)(1-w)$, and $\theta(0,\tau) = 0.2$. (This particular choice has some merit in a problem of polymer crystallization). But let's move to some more interesting cases, when either **Ju** or **St** or some combination thereof approaches infinity or zero.

Remark 1.8.10

One wishes to move to special cases not for the sake of singular limits (without which our life would be much happier), but for two important reasons. First, we want to understand the logical status of the Stefan problems considered in the previous section. Second, as long as the

> solution is going to be nice and smooth, we foresee no difficulty of numerical integration; but in a singular limit we do foresee it, and thus we want a "feel" for what's going to happen there. Indeed, if readers have tried a computer calculations for a small **Ju** value, something like 0.001, they are likely to have run into some trouble.

The case where **St** approaches zero is trivial, since in the limit one recovers the pure diffusion problem in Section 1.6, with $w(\tau)$ determined by the solution of (1.8.23) with θ given by (1.6.2). An interesting case is the one where **Ju** approaches zero: substitution of Eq. (1.8.22) into (1.8.21) shows that this is a singular limit, since the coefficient of the highest order derivative approaches zero. This case is discussed below.

But first of all, let us establish what the solution looks like as τ approaches ∞, and in fact at how large values of τ will this situation be approached. Well, the physics of our problem tell us that, after a sufficiently long time, θ will be unity everywhere (the sample has become saturated: in the combustion case, we'll have ash saturated with oxygen; in the freezing case, ice at the temperature of the exposed surface). Furthermore, w will be $w^*(1)$ everywhere (ash, ice everywhere). In fact, we can normalize the definition of w so that $w^*(1) = 1$. So now we know what we are going to approach, and the question is, how long will it take?

Let us go back to dimensional quantities. The transformation of the internal state variable w from its initial value of zero to its final value of unity takes a time of order β. Should there be no kinetic phenomenon, diffusion by itself would take a time of order L^2/D. This would be the time required to absorb into our sample an amount JL per unit exposed surface area, under a driving force J/Φ: the final concentration would exceed the initial one by an amount J because U would have changed by an amount J/Φ, see Eq. (1.8.8). But in the presence of the kinetic phenomenon the final concentration would exceed the initial one by an amount $J + \Gamma$, and hence an amount $L(J + \Gamma)$ needs to be absorbed under the same driving force J/Φ. Hence the time needed for saturation is $(J + \Gamma)L^2/DJ = (1 + \mathbf{St})L^2/D$. Thus the time needed to reach final equilibrium is of the order of the largest one of the following three:

$$\beta,\ L^2/D,\ \mathbf{St}L^2/D \tag{1.8.27}$$

or, in the dimensionless time scale τ, the largest one of:

$$1,\ 1/\mathbf{Ju},\ \mathbf{St}/\mathbf{Ju} \tag{1.8.28}$$

Now we can realize even more fully how nasty the limit **Ju** = 0 is: the second two time scales approach ∞ in the τ time scale. So perhaps our scaling was not the best one, at least not in this case. Let's make the following change of variables:

$$\sigma = tD/L^2 = \mathbf{Ju}\tau \tag{1.8.29}$$

$$g = FL/JD = f/\mathbf{Ju} \tag{1.8.30}$$

so that the canonical set becomes:

$$\partial g/\partial x + \partial\theta/\partial\sigma = -\mathbf{St}\ \partial w/\partial\sigma \tag{1.8.31}$$

$$\mathrm{g} = -\partial\theta/\partial\mathrm{x} \tag{1.8.32}$$

$$\partial w/\partial\sigma = r(w,\theta)/\mathbf{Ju} \tag{1.8.33}$$

Changes of variables are not expected to produce wonders, and all that we have obtained has been to eliminate **Ju** from the balance and the constitutive equation — at the cost of having it appear in the kinetic equation. But we have obtained another advantage: in the σ time scale, the three times of interest are:

$$\mathbf{Ju},\ 1,\ \ \mathbf{St} \tag{1.8.34}$$

and of these we can forget about **Ju**, since it is negligibly small, and we have a decent time scale for equilibration of the order of unity or **St**.

What does Eq. (1.8.33) mean? $\partial w/\partial\sigma$ cannot be extremely large at all times, after all. We are trying to get first a zero order approximation to our problem, and so we are willing to make some rather drastic approximations. Now suppose that $w = w^*(\theta) + \varepsilon$, with ε a very small number. In other words, the internal state variable is always very close to its equilibrium value — that's what we expect from a very rapid chemical reaction, and β approaching zero is the case of a very fast reaction. Let's assume that $r()$ is differentiable with respect to w at θ, $w^*(\theta)$. Then, from Eq. (1.8.33):

$$\partial w/\partial\sigma = rW(w^*(\theta), \theta)\varepsilon/\mathbf{Ju} \tag{1.8.35}$$

What does this mean? It means that, with ε of order **Ju** (i.e., a very small displacement from equilibrium), we can have a finite rate $\partial w/\partial\sigma$, of order say unity. Thus to within $O(\mathbf{Ju})$ we may assume that $w = w^*(\theta)$ everywhere. Now if $w^*(\theta)$ is given by Eq. (1.7.16), this means that w is either 0 or unity — a discontinuity of w emerges from our smooth model. Let us try to follow this line of thought.

Remark 1.8.11

> Notice that, by making the zero order approximation, $w = w^*(\theta)$ everywhere, i.e., equilibrium prevails everywhere. In spite of this, the rate of reaction is finite. This is typical of "instantaneous reaction" theory, and should be thoroughly understood.

Near the exposed surface, there must be a region where θ is positive — this is forced by BC (1.8.25). Following our argument, this is a region where $w = 1$ (ice, ash). Let $s(\sigma) < 1$ be the thickness of this region ($s(0) = 0$). But at sufficiently short times, and sufficiently far away from the exposed surface, $\theta = 0$ — it starts from that value, after all. So let's try this. At $x > s(\sigma)$, $\theta = w = 0$. This satisfies all differential equations, and all boundary conditions except (1.8.25); we call this the right solution. At $x < s(\sigma)$, $w = 1$, $\partial w/\partial\sigma = 0$, and consequently θ is the solution of:

$$\partial^2\theta/\partial x^2 = \partial\theta/\partial\sigma \tag{1.8.36}$$

(the "left" solution). Now Eq. (1.8.36) is subject to BC (1.8.26) and to ...?

If we still believe Eq. (1.8.17) (but more about this in the next section), $\theta(s(\sigma),\sigma) = 0$. This starts to look like the Stefan problem, doesn't it? Now let's look at Eq. (1.8.16), which in our set of dimensionless variables becomes:

$$[\partial\theta/\partial x] = -\mathbf{St}[w]\mathrm{d}s/\mathrm{d}\sigma \tag{1.8.37}$$

This time $[w]$ is not zero, if we haven't made any logical mistake so far: $[w] = -1$. $\partial\theta/\partial x$ is zero to the right of $s(\sigma)$, and thus we obtain:

$$x = s(\sigma), -\partial\theta/\partial x = \mathbf{St}\mathrm{d}s/\mathrm{d}\sigma \tag{1.8.38}$$

and the problem has reduced exactly to the Stefan problem — the "left" solution is the solution of the Stefan problem, and the "right" solution is what

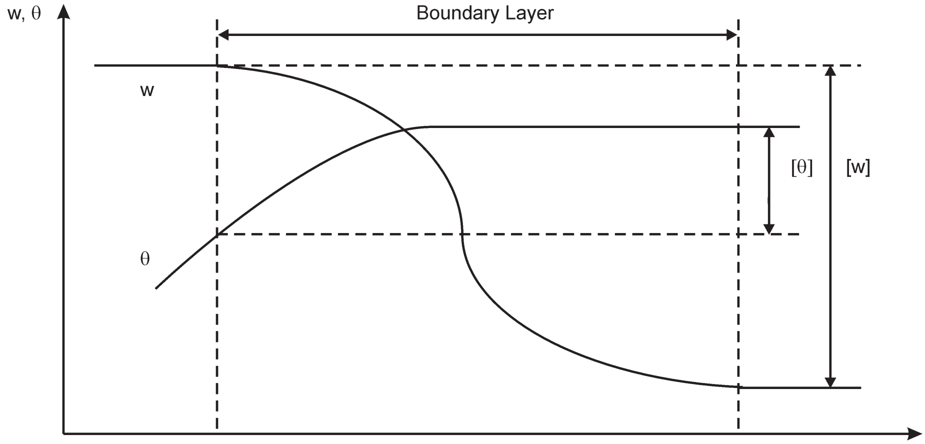

Fig. 1.8.1. Sketch of w and θ distributions within the boundary layer for order of magnitude estimations.

was assumed in the Stefan problem to be the solution to the right of the freeze surface. Since we know that the Stefan formulation makes sense, we are quite happy with this result. However, we have been very cavalier indeed; we have retained (1.8.17) to the full impact of its implications, but we have thrown (1.8.18) to the wind entirely. Furthermore, the solution to the Stefan problem has $\partial w/\partial\sigma = 0$ both to the right and to the left of $s(\sigma)$, but it is very large indeed at the surface of discontinuity itself. We should, perhaps, be a little more careful. This is a singular limit problem, and a boundary layer analysis is definitely in order.

So let us go back to the formulation where the singularity of the limit is most evident; this is obtained by substituting (1.8.22) into (1.8.21):

$$\mathbf{Ju}\,\partial^2\theta/\partial x^2 = \partial\theta/\partial\tau + \mathbf{St}\,\partial w/\partial\tau \tag{1.8.39}$$

This time we must be very careful indeed. We have the following difficulty. We know that w changes by an amount of the order of unity across the boundary layer, but we do not know by how much θ may change — after all, we are performing the whole analysis partly because we are not entirely convinced that we can retain (1.8.17) and yet forget about (1.8.18). We expect the distributions of w and θ within the boundary layer to be as sketched in Fig. 1.8.1, with $[\theta]$ a yet to be determined quantity (we are using []

to identify total changes through the boundary layer, since we expect those to degenerate asymptotically into actual discontinuities in the limit where the boundary layer thickness shrinks to zero).

Remark 1.8.12

One should not give in to the temptation of invariably regarding the order of magnitude of changes of dimensionless dependent variables to be of order unity.

Let's first examine the order of magnitude of $\partial\theta/\partial\tau$. The boundary layer moves by an amount equal to its thickness ε in a time $\varepsilon/(\mathrm{d}s/\mathrm{d}\tau) = \varepsilon/(\mathbf{Ju}\mathrm{d}s/\mathrm{d}\sigma)$. In the σ time scale, $\mathrm{d}s/\mathrm{d}\sigma$ is of order unity, hence the estimate becomes $\varepsilon/\mathbf{Ju}$. Over such a time interval, θ changes by an amount $[\theta]$. Hence:

$$\partial\theta/\partial\tau \approx [\theta]\mathrm{Ju}/\varepsilon \tag{1.8.40}$$

The term $\mathbf{Ju}\partial^2\theta/\partial x^2$ is of course of order $\mathbf{Ju}[\theta]/\varepsilon^2$, and the term $\mathbf{St}\partial w/\partial\tau$ of order **St**. So we get something like:

$$\mathbf{Ju}[\theta]/\varepsilon^2 \approx \mathbf{Ju}[\theta]/\varepsilon + \mathbf{St} \tag{1.8.41}$$

Now it takes a bit of logical thinking to convince oneself that, unless **St** is much less than unity, the only way to have the LHS of the same order of magnitude as the RHS is to have:

$$\varepsilon = \sqrt{(\mathbf{Ju}[\theta])} \tag{1.8.42}$$

where in fact, for the sake of simplicity, we have assumed **St** to be of order unity.

In fact, should ε be much smaller than that, the LHS would exceed the RHS; and should it be much larger, the LHS would be negligible as compared to the last term. But this leads to another conclusion: $\partial\theta/\partial\tau$ is negligible within the boundary layer, since it is of order ε with respect to the other terms. Thus the differential equation for θ within the boundary layer will be, to within order ε, an ordinary one.

Remark 1.8.13

Now reread Remark 1.3.11 and draw your own conclusions.

The stretched coordinate z is defined in the usual way:

$$z = (x - s(\tau))/\varepsilon \tag{1.8.43}$$

Let, as usual, δ indicate partial derivatives in the z–τ system of independent variables. One obtains:

$$\partial/\partial x = (1/\varepsilon)\delta/\delta z \tag{1.8.44}$$

$$\begin{aligned}\partial/\partial\tau &= \delta/\delta\tau - (1/\varepsilon)(\mathrm{d}s/\mathrm{d}\tau)\delta/\delta z = \\ &= \delta/\delta\tau - (\mathbf{Ju}/\varepsilon)(\mathrm{d}s/\mathrm{d}\sigma)\delta/\delta z\end{aligned} \tag{1.8.45}$$

Now let us apply Eq. (1.8.45) to w itself. The LHS is, by definition, of order unity, and so must the RHS be. This implies that:

$$\varepsilon = \mathbf{Ju} \tag{1.8.46}$$

and thus the thickness of the boundary layer has been established. When this is compared with Eq. (1.8.42), one concludes that $[\theta]$ is of order **Ju**, i.e., that within the zero order approximation:

$$\theta_L = 0 \tag{1.8.47}$$

which is indeed one of the BCs assumed in the Stefan model.

Remark 1.8.13

Notice that we are not here invoking the usual FIA assumption — we are proving it, if in only a limited sense. The point will be discussed in detail in the next section.

Now we have established that the differential equation for θ in the boundary layer is an ordinary one, and thus by substitution into (1.8.39) we obtain:

$$\mathrm{d}^2\theta/\mathrm{d}z^2 = -\,\mathbf{JuSt}(\mathrm{d}s/\mathrm{d}\sigma)\mathrm{d}w/\mathrm{d}z \tag{1.8.48}$$

Rather than trying to find the solution of (1.8.48), we simply integrate it between $-\infty$ and $+\infty$, noting that:

$$z = +\infty;\, w = 0,\, \mathrm{d}\theta/\mathrm{d}z = 0 \tag{1.8.49}$$

$$z = -\infty;\ w = 1,\ \mathrm{d}\theta/\mathrm{d}z = -\mathbf{Ju}(\partial\theta/\partial x)_{\mathrm{L}} \tag{1.8.50}$$

The result is:

$$-(\partial\theta/\partial x)_{\mathrm{L}} = \mathbf{St}(\mathrm{d}s/\mathrm{d}\sigma) \tag{1.8.51}$$

which is of course the other BC of the Stefan problem formulation.

Remark 1.8.14

We have only integrated the boundary layer equation to obtain the boundary conditions of the zero order problem, i.e., to show that the singular limit is approached regularly. Should we actually be interested in the fine details of the w and θ distributions within the boundary layer (as sometimes may well be the case), one would need to obtain the solution to Eq. (1.8.49) subject to the appropriate matching conditions. This would of course require to assign a specific form to the constitutive functions $r()$ and $w^*()$.

Remark 1.8.15

In the initial stages of modelling, one wants on the one side to follow the advice of Remark 1.8.4, and on the other a somewhat contradictory requirement — make as few assumptions as possible, so that the results are of as general validity as feasible. The choice of the appropriate assumptions to make or not to make is really the art of modelling. But there is one piece of advice which is generally useful. If one wants to make some kind of assumption (for instance, that Fourier's law applies), one may as well go all the way, and take conductivity to be constant. Don't be wishy washy, make a definite choice instead: be either very general, or very specific.

What we have shown in the analysis above is that, when **Ju** approaches zero, the solution of the complete problem degenerates regularly into the solution of the corresponding Stefan problem. This may look as not much of a result, but the reader is asked to think about the following point. In the classical formulation of the Stefan problem, the existence of a discontinuity (of degree of crystallinity, of the temperature gradient, whatever) is *assumed* to exist. In the complete formulation, where kinetics are taken into account, the discontinuity is shown to arise, in a well defined asymptotic sense, from a model which by itself is essentially a continuous one. Thus one concludes that discontinuous and continuous models are not mutually exclusive — the former may arise as proper asymptotes of the latter ones.

1.9 THE FUNDAMENTAL INTERFACE ASSUMPTION

In the preceding section, we concluded that the jump of potential U is zero even though the jump of extent of reaction w may be nonzero. Here we ask ourselves the following question: how general is the conclusion that the potential U does not suffer discontinuities? The assumption that the potential does not suffer discontinuities is what we have called the fundamental interface assumption (FIA).

First, let us be absolutely clear about what the FIA is. In heat transfer, it is the assumption that, at the interface between two different phases, temperature is the same on the two sides of the interface — in fact, we usually speak of *the* temperature of the interface. In momentum transfer, it is the assumption that velocity is continuous across interfaces — the so-called no-slip condition. Though this is also universally used, it is somehow less intuitive than the analogous assumption for temperature. Indeed, in the classical theory of the hydrodynamics of ideal fluids, this assumption is relaxed — the solution of the Euler equations of motion for flow past a submerged object does not satisfy the no-slip condition, and the tangential velocity on the surface of a fixed solid boundary is nonzero.

Remark 1.9.1

> This point will be discussed in more detail later. Here we only wish to point out that, as compared to the Navier–Stokes equations, the Euler equations are of lower order; thus the solution of them is a lower order solution which cannot satisfy all the boundary conditions. Since one would lose the problem altogether if one gave up the BC that the orthogonal component of velocity is zero at a solid boundary, one is forced to give up the one concerning the tangential component, thus violating the no-slip condition.

The case of mass transfer is even less liable to be waved away by “physical intuition”. We write equations in terms of concentrations, and even at equilibrium concentration suffers a discontinuity at the interface. So we do not assume that concentration does not suffer a discontinuity, we assume that chemical potential does not: the concentrations on the two sides of the interface are in the “equilibrium” relationship with each other.

Remark 1.9.2

> We are again approaching the muddy waters of thermodynamics with these considerations. Most of us would of course rather like to avoid this,

> but we are forced into it — see Remark 1.5.12, which now appears more threatening than it did the first time.

The above considerations tell us that we have to make one more distinction than we have done so far. Indeed, so far we have distinguished between the concentration of the diffusing quantity (ΦU) and the quantity the gradient of which appears in the constitutive equation for the flux (U); we have called the latter the potential. Now in the case of heat transfer ΦU is enthalpy per unit volume, and U is temperature; in the case of momentum transfer ΦU is the concentration of momentum (Φ is density), and U is the velocity; in the case of mass transfer Φ is unity, and U is concentration. However, the quantity which is the same in both phases at equilibrium, and about which we are asking ourselves questions about continuity under nonequilibrium conditions, is U for momentum and heat transfer, but not for mass transfer. We need to give such a quantity a name, and "potential" (which has, unfortunately, been used already), cannot do.

The still unnamed quantity we are dealing with is identified by the fact that, at equilibrium, it has the same value in both phases. Thus it is identified by an equilibrium, or thermodynamic, characteristic, and it comes natural to call it the thermodynamic potential, $\mathbf{G}$. The symbol $\mathbf{G}$ is justified by the fact that in the one case where the distinction between the potential and the thermodynamic potential is relevant (mass transfer), $\mathbf{G}$ is indeed the partial molar free enthalpy, or chemical potential.

We need, however, to discuss the thermodynamic meaning of $\mathbf{G}$ a little more carefully. In a two phase system at equilibrium, $\mathbf{G}$ is required to be spatially uniform within each phase ($\text{grad}\mathbf{G} = 0$), as well as equal in both phases. Under non-equilibrium conditions (and under appropriate qualifying assumptions), the scalar product of the corresponding flux with the gradient of $\mathbf{G}$ is required to be nonpositive. For the one-dimensional problems we are considering in this chapter, this reduces to the following requirement:

$$F\partial\mathbf{G}/\partial X \leq 0 \tag{1.9.1}$$

Remark 1.9.3

> The qualifications on Eq. (1.9.1) are discussed in detail in the next chapter. Suffice it here to say that for the type of constitutive equations considered so far, (1.9.1) needs always to be satisfied.

Remark 1.9.4

> Under the qualifications discussed in the preceding remark, the LHS of Eq. (1.9.1) is minus the local rate of energy dissipation, and if divided by temperature T it is minus the local rate of entropy production.

Notice, however, that the requirements discussed above do not identify **G** entirely. Indeed, any unique and invertible function of **G** would do equally well in satisfying the equilibrium requirement, and any monotonously increasing function of it would also satisfy the nonequilibrium requirement in Eq. (1.9.1). Consider, for instance, heat transfer by conduction, and let us tentatively identify F with the heat flux q, and **G** with temperature. Temperature is certainly uniform in space at equilibrium, equal in all phases at equilibrium, and Eq. (1.9.1) makes sense, since we expect heat to flow by conduction in the direction of decreasing temperature. So far, so good, and we write:

$$q\partial T/\partial X \leq 0 \tag{1.9.2}$$

However, suppose one considers the quantity $\alpha = -1/T$. This is also spatially uniform, and equal in both phases at equilibrium. If we substitute this in Eq. (1.9.2), we obtain:

$$(q/\alpha^2)\partial\alpha/\partial X \leq 0 \tag{1.9.3}$$

and who is to tell us that α is not the thermodynamic potential, and q/α^2 the appropriate flux?

Be that as it may (see the chapter on thermodynamics), let us assume that we have satisfied ourselves about what **G** is, and what F is. After all, we are so satisfied in the three transport phenomena we have so far considered. Furthermore, our aim in this section is to establish the logical status of the FIA, which is the statement that [**G**] is zero across an interface, and such a statement would also apply to any quantity such as the quantity α discussed above. Now we know that at equilibrium F is either zero, or has some value F_0 which can be calculated from an equilibrium analysis (see Remark 1.5.2). Suppose we want to write the simplest possible constitutive equation for F under nonequilibrium conditions. The zero order equation is:

$$F = F_0 \tag{1.9.4}$$

This may seem so degenerately simple that we are losing our problem altogether. However, notice that this is exactly what we have done in Section 1.5. Furthermore, this is the assumption that is made in the theory of the hydrodynamics of ideal fluids, i.e., that the only internal stress in a fluid in flow (i.e., a fluid which is not at equilibrium) is an isotropic pressure. This shows at least that Eq. (1.9.4) does not necessarily imply loss of any significance in momentum transfer. However, it is certainly degenerately simple for heat and mass transfer (where $F_0 = 0$), so let us proceed to the next step and write:

$$F = F_0 - L\partial \mathbf{G}/\partial X \tag{1.9.5}$$

where $L \geq 0$ is called the mobility. For momentum transfer, with **G** identified as the velocity, L would be the viscosity, and (1.9.5) would be Newton's linear law of friction. Momentum transfer is too complicated for thermodynamicists, and so they do not have strong opinions about this. For heat transfer, with **G** identified with temperature, L would be the thermal conductivity, and (1.9.5) is Fourier's law. Here thermodynamicists do have strong opinions, and they disagree with us, but we are not going to let them deter us. In mass transfer, with **G** the chemical potential, thermodynamicists would agree with (1.9.5), but we are uneasy with it, because we prefer Fick's law, and *L is not* the diffusivity. Well, we do not care that much — we can recover Fick's law by identifying the diffusivity with $\mathrm{L}\partial \mathbf{G}/\partial U$ and hope for the best, that $\partial \mathbf{G}/\partial U$ is well behaved and thus is always positive, so that the requirement that $L \geq 0$ implies that $D \geq 0$. However, this time hoping for the best isn't going to work.

We could ask ourselves, at this point, what is all the fuss about. Let us simplify matters by regarding F as really being $F - F_0$, so that at equilibrium it is zero. Now what difference does it make whether we write $F = -L\partial \mathbf{G}/\partial X$ or $F = -D\partial U/\partial X$? In both cases, we are going to allow D to depend on U, or L on **G**, so why bother? Well, the reason for bothering is that in general D is more nearly constant than L. However, while L is necessarily nonnegative, we cannot say the same for D — there is no guarantee that $\partial \mathbf{G}/\partial U$ is always positive.

With that much background, let us move to the analysis of the FIA. We have two phases, a transport phenomenon is taking place, X is orthogonal to the interface. We do not yet know exactly what we mean by having two phases, but let's not worry about that for the time being, we have an intuitive feeling for what we mean, and there will be a chapter on thermodynamics later on. The discussion in Sections 1.5 and 1.9 suggests that we can identify the two phases with an additional internal state variable w, $w = 0$ being one phase, and

$w = 1$ the other one. We now go to some textbook where surface phenomena are discussed, and we convince ourselves that what we call an interface (and we describe mathematically as a surface of discontinuity) is really a very thin layer (of perhaps close to molecular dimensions) through which quantities have very steep, but finite gradients. Well, that doesn't bother us; we picture in our minds an interface layer where w changes steeply, but not discontinuously, from the value 0 to the value 1. We write our constitutive equation as follows:

$$F = -L(w,\mathbf{G})\partial\mathbf{G}/\partial X \tag{1.9.6}$$

Why have we bothered with allowing L to depend on w and $\mathbf{G}$? Well, we have to. Consider for instance heat transfer (in our version): L would be the thermal conductivity k, and we know that k is not going to have the same value in both phases, so we need to let it be a function of w. Now Remark 1.8.15 tells us that, once we are forced into regarding L as dependent on w, we should allow it to depend on $\mathbf{G}$ as well.

Remark 1.9.6

Notice that in the analysis in the previous section D was taken to be constant — for the freezing case, that would correspond to the thermal conductivity of water and ice being equal. Not too bad an approximation, but not a very good one either.

Remark 1.9.7

We are going to assume that Eq. (1.9.6) holds through the interface "layer", i.e., at w values ranging between zero and unity. Is that a good assumption? Well, it's the best one we can make at this stage.

Let the interface layer have thickness 2ε, which we are willing to regard as negligibly small. We choose X so that $X = 0$ is at the midplane of the interface layer. Since we will eventually describe the interface as a surface of discontinuity, we identify $[F]$ with:

$$[F] = \int_{-\varepsilon}^{\varepsilon} \frac{\partial F}{\partial X} = \frac{\mathrm{d}}{\mathrm{d}t}\int_{-\varepsilon}^{\varepsilon} C\mathrm{d}X = 2\varepsilon\,\frac{\mathrm{d}C_{\mathrm{AVG}}}{\mathrm{d}t} \tag{1.9.7}$$

where $C = \Phi U$ is the concentration of the quantity being transported, and the subscript AVG identifies the average value within the interface layer. Since we are trying to develop a zero order theory, all terms of order ε are neglected, and thus we conclude that:

$$[F] = 0 \tag{1.9.8}$$

Remark 1.9.8

> In transport phenomena theory, Eq. (1.9.8) is always used across interfaces, both at steady and at unsteady state. The analysis above, which is an unsteady state one, gives support to this assumption. Notice that the result is independent of the constitutive assumption in Eq. (1.9.6).

Now both w and $\mathbf{G}$ may have any distribution within the interface layer; consequently, L will have some arbitrary distribution. Let L_{MAX} and L_{MIN} be the largest and the smallest value of L within the interface layer; the second law of thermodynamics requires both (and hence in particular L_{MIN}) to be nonnegative. We are going to make one additional, but mild, hypothesis: L_{MIN} is nonzero, i.e., *the mobility never vanishes*. With that, we can write:

$$L_{\mathrm{MAX}} \geq L_{\mathrm{MIN}} > 0 \tag{1.9.9}$$

$$-F \geq L_{\mathrm{MIN}} \partial \mathbf{G}/\partial X \tag{1.9.10}$$

$$-F \leq L_{\mathrm{MAX}} \partial \mathbf{G}/\partial X \tag{1.9.11}$$

Now we integrate Eqs. (1.9.10) and (1.9.11) from $-\varepsilon$ to $+\varepsilon$ to obtain:

$$L_{\mathrm{MAX}}[\mathbf{G}] \geq 0 \geq L_{\mathrm{MIN}}[\mathbf{G}] \tag{1.9.12}$$

We are now home. The only way of satisfying both (1.9.9) and (1.9.12) is:

$$[\mathbf{G}] = 0 \tag{1.9.13}$$

which is of course the FIA. We have thus proved the following result:

> *If Eq. (1.9.6) holds throughout the interface layer, with $L>0$, $[\mathbf{G}] = 0$, i.e., the FIA holds.*

Remark 1.9.9

Notice that, should L_{MIN} be allowed to be zero, Eq. (1.9.12) would not imply Eq. (1.9.13).

It is now apparent why the distinction between U and $\mathbf{G}$ is a relevant one. The result which has been obtained is crucially related to the sign of L, and the analogous result does not hold for U, i.e., $[U]$ may well be nonzero. Indeed, in mass transfer $[U]$ is nonzero even at equilibrium. If one regards $\mathbf{G}$ as a function of w and U, at equilibrium it is known that $\mathbf{G}(0,U_0) = \mathbf{G}(1,U_1)$, where U_0 and U_1 are the concentrations in the two phases. Hence $\mathbf{G}$ as a function of U is not invertible, and nothing can be said about the sign of the diffusivity within the interface layer.

It is perhaps useful at this stage to do some more thinking. The result we have obtained is satisfactory in several respects. First, it gives to the FIA a stronger status than that of an independent assumption: as long as we use constitutive equations of the type of (1.9.6), we might as well use the FIA, since the former imply the latter under mild additional assumptions. Second, it somehow tells us that the result in the previous section was to be expected: there might be a discontinuity of w, but not one of $\mathbf{G}$. So far, so good. However, what about the analysis in Section 1.5, where discontinuities of $\mathbf{G}$ propped up here, there, and everywhere? Granted, in that section we used the zero order constitutive equation (Eq. (1.9.4)), and thus the mathematics are alright; but what about the physics? A gas may have a small viscosity, a small thermal conductivity, whatever; but they are nonzero — and as soon as we put that in, it would appear that the FIA should hold, even if discontinuities of w are not forbidden any more. Perhaps something like the analysis in Section 1.3 would do? Perhaps. But we certainly need to do some hard thinking on all that.

It is also satisfactory to realize that the hydrodynamics of ideal fluids is a selfconsistent theory. The constitutive assumption is Eq. (1.9.4), which does not imply the FIA. Consistently, the no-slip condition is not used as a boundary condition in the solution of the Euler equations of motion. In this regard, it is useful to reread Remark 1.9.1. We give up the no-slip condition for consistency (as well as because if we let go of any other boundary condition we would lose the problem altogether).

One final observation is in order. The whole analysis in this section, and in all preceding ones, is based on constitutive equations such as (1.9.6), or even simpler ones such as (1.9.4). What about the possibility of more complex constitutive equations? This point will be discussed in the next section.

1.10 RELAXATION

In this section, we introduce the concept of relaxation as applied to the constitutive equation for the flux. We develop the theory in its most elementary form, focusing on the conceptual issues which are related to the idea of relaxation.

First of all, let's take a look at the physics involved. Consider the classical constitutive equation for the flux:

$$F = -k\partial U/\partial X \tag{1.10.1}$$

Suppose the validity of Eq. (1.10.1) has been established by running experiments at steady state. For instance, in the heat transfer case, one has run experiments where a slab of the material considered is held between two plates at different temperatures, and the heat flux has been measured. (If the material is fluid, one has had care of having the two plates rather close to each other, horizontal, with the high temperature one on top, so as to minimize the risk of free convection). For momentum transfer, one has placed the material between two parallel plates, one fixed and one moving with constant velocity, and one has measured the tangential force needed to keep the fixed plate from moving. What has been established from such experiments? It has been established that Eq. (1.10.1) holds at steady state. Any constitutive equation which at steady state degenerates into (1.10.1) would do equally well in describing our data.

Remark 1.10.1

> The two plates would need to be infinitely long in both directions orthogonal to X, which would require a rather inconveniently large laboratory. Well, we can substitute the two plates with two coaxial cylinders separated by a very small gap, so that curvature could be neglected. We have just invented the Couette viscometer.

Well, can't we do unsteady state experiments? Of course the answer in principle is yes; in practice, it means one has to face important experimental problems. In the momentum transfer case, we'll need to clean out of our data the effect of inertia, and that is not easy. However, let's suppose that we can do that; for instance, rather than keeping the forcing plate moving at a constant velocity, we oscillate it sinusoidally, so that the position Z of some mark on it is given by:

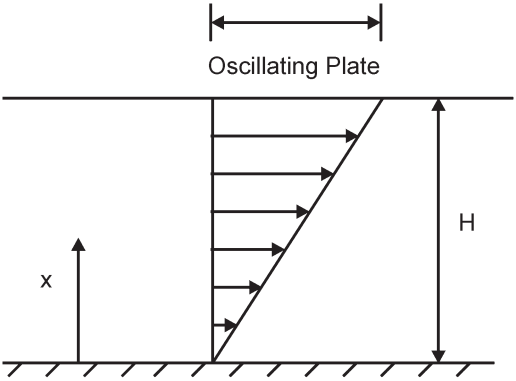

Fig. 1.10.1. An experiment of oscillatory shear. F is the tangential force which the fluid exerts on the upper plate. As long as inertia is neglected, the velocity profile is triangular at all times.

$$Z = A\sin(\Omega t) \tag{1.10.2}$$

where A is the amplitude of the oscillation and Ω is the frequency. Correspondingly, the velocity of the forcing plate is:

$$v = A\Omega\cos(\Omega t) \tag{1.10.3}$$

Suppose somehow we have set up things in such a way that inertia can be neglected, so that the velocity profile at any one time is triangular (see Fig. 1.10.1). If H is the distance between the two plates, the velocity gradient is:

$$\partial U/\partial X = v/H = (A\Omega/H)\cos(\Omega t) \tag{1.10.4}$$

and, if Eq. (1.10.1) holds, the tangential stress should be in phase with v, i.e., it should be:

$$F = -(Ak\Omega/H)\cos(\Omega t) \tag{1.10.5}$$

Well, experiments of this kind have been performed, and if the fluid between the two plates is a polymeric one (something like a polymer melt, to be extreme; but also something like sinovial fluid, ketchup, shampoo, or liquid soap) one observes that:

$$F = -(Ak\Omega/H)\cos(\Omega t + \phi) \qquad (1.10.6)$$

Furthermore, ϕ depends on Ω, with ϕ approaching $\pi/2$ as Ω approaches ∞, and 0 as Ω approaches zero. What are we to do with this?

Remark 1.10.2

The behaviour as Ω approaches infinity is reminiscent of that of a solid: F is in phase with Z, i.e., with the strain. That gives us food for thought.

Let's try the following variation. Let's write the constitutive equation as:

$$F + \beta\partial F/\partial t = -k\,\partial U/\partial X \qquad (1.10.7)$$

where β is the "relaxation time". Let us ask ourselves whether Eq. (1.10.7) can describe the observed behaviour

Remark 1.10.3

Equation (1.10.7) reduces to (1.10.1) at steady state. Thus no steady state experiment can tell us which one of the two should be used.

Those readers who have no problems with things like $\exp(i\Omega t)$ will work through this problem in no time at all. There may be a few readers who feel a bit uncomfortable with $\exp(i\Omega t)$. Let's try to find something out without actually solving the system of Eqs. (1.10.4) and (1.10.7). It is perhaps immediately obvious that the solution will be such that F is periodic with frequency Ω, say indeed of the form of Eq. (1.10.6). This is enough to conclude that, of the two terms on the LHS of Eq. (1.10.7), the second one is of the order of $\beta\Omega$ times the first one. But this, in fact, is all we need. If $\beta\Omega << 1$, we can neglect the second term, recover Eq. (1.10.1), and conclude that $\phi = 0$ — as indeed is observed experimentally. If $\beta\Omega >> 1$, we can neglect the first term on the LHS of Eq. (1.10.7), and hence $\phi = \pi/2$, again in agreement with experiment. It has just worked like a breeze, almost too good to be true

Remark 1.10.4

It *is* too good to be true. It works for the extremely simple geometry considered, but as soon as one moves to more complex geometries there are severe headaches in store, severe enough that it takes a long time to convince oneself again that, for this simple case, indeed the analysis as presented is correct.

Well, that makes us rather happy with (1.10.7), doesn't it? Somebody in the back of the room now raises her hand to point out that (1.10.7) does not guarantee that (1.9.1) is satisfied. Very good point indeed, Sue. If $\phi$is different from zero, there is a portion of the cycle during which F and $\partial U/\partial X$ have the same sign. The *integral* of $F\partial U/\partial X$ over a cycle is negative (except in the limit $\Omega = \infty$ when it is zero), but the instantaneous value is, at times, positive. But let's read Remark 1.9.3 again — (1.9.1) is supposed to hold at steady state only, and here we are most definitely not at steady state.

Remark 1.10.5

Why are we interested in the integral of $F\partial U/\partial X$? Well, the instantaneous rate of work done on the system is $-Fv$, and hence has the sign of $-F\partial U/\partial X$. Thus the integral of $-F\partial U/\partial X$ over a cycle has the same sign as the total work done on the system, and we expect that to be nonnegative. Notice that F is a momentum flux, and is thus the tangential force exerted by the fluid below on the fluid above, or by the fluid on the plate; thus $-F$ is the force exerted by the plate on the fluid.

Remark 1.10.6

With the internal stress interpreted as the flux of momentum, its sign is fixed, and a compressive force is positive. This is the usual sign convention used in fluid mechanics. As a force per unit area, the sign can be chosen arbitrarily. People working in solid mechanics like to regard tensions as positive (structures are more likely to collapse under tensile loads than under compressive ones), and so they generally use the opposite convention, so that internal stress is minus the momentum flux.

Another hand is raised. Now the argument has been accepted for momentum transfer (who understands thermodynamics as applied to momentum transfer anyhow?), but heat transfer, that's another matter. Heat flowing, if for only a portion of some cycle, in the direction of increasing temperature — that is not, by any means, acceptable. Good point again, of course. That's the advantage of writing a book — one gets asked only the questions one wants to be asked.

However, air conditioners and refrigerators do work, don't they? So heat does flow towards higher temperatures on occasion. Yes, but at the expense of some cost in electricity (our interlocutor is just perfect, isn't she?). That's right — but the integral over a cycle of $F\partial U/\partial X$ is negative, so we are expending work after all. We get a little bit of it back on occasion, but on the whole we are still spending more than we are getting back, right? As Ω approaches ∞ we aren't spending anything — but we aren't getting anything for free either. If we believe that the second law of thermodynamics can be formulated somehow like "there is no such thing as a free lunch", well, we aren't getting any free lunches.

Remark 1.10.7

> Again thermodynamics is mudding the water. Well, we'll soon get to it. A little anticipation: with elastic solids we expect $\phi = \pi/2$, so we aren't spending anything. Are elastic solids perhaps only capable of reversible transformations?

Now of course all hands in the class are raised, in fact there are more hands raised than people in the class, but we are not going to allow anybody to ask another question, we want to get on with our algebra. So we'll use (1.10.7) after all, and do some thinking with it.

First of all, we use Kotchine's procedure, it has always served us fine so far. Let V^* be the (dimensional) speed of propagation of the discontinuity. We are going to regard all parameters as constant, and thus we write the balance and constitutive equations as follows:

$$\partial F/\partial X + \partial C/\partial t = 0 \tag{1.10.8}$$

$$\beta\partial F/\partial t + D\partial C/\partial X = -F \tag{1.10.9}$$

and the Kotchine procedure gives:

$$[F] = V^*[C] \tag{1.10.10}$$

$$\beta V^*[F] = D[C] \tag{1.10.11}$$

Equation (1.10.10) is our old friend, the balance across the discontinuity; Eq. (1.10.11) is novel, and perhaps we should cogitate on it a bit. It does allow discontinuities of the flux, provided there is a discontinuity of the concentration, and a finite speed of propagation. This is of course related to the fact that the canonical set is hyperbolic.

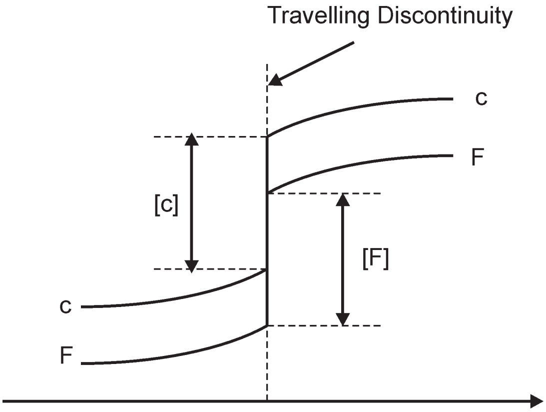

Fig. 1.10.2. Sketch for understanding Eq. (1.10.11). Both $\partial C/\partial X$ and $\partial F/\partial t$ are ∞ at the travelling discontinuity.

Now if $[C]$ is nonzero at the travelling discontinuity, $\partial C/\partial X$ is infinitely large (see Fig. 1.10.2). In the absence of the first term on the LHS of (1.10.9), this would result in an infinitely large flux, and we refrain from considering this possible. But as (1.10.9) stands, an infinitely large value of $\partial C/\partial X$ can be compensated by an infinitely large value of $\partial F/\partial t$, with finite F. If the discontinuity propagates at some finite speed, an infinite value of $\partial F/\partial t$ corresponds to a jump in F, as indeed Eq. (1.10.11) tells us.

Remark 1.10.8

The considerations in the paragraph above are hand-waving ones. They are convincing only because we have Eq. (1.10.11) already available. While Eq. (1.10.10) can in fact be obtained directly from a balance across the discontinuity, (1.10.11) can convincingly be obtained only from Kotchine's procedure.

Combination of Eqs. (1.10.10) and (1.10.11) gives us the value of V^*:

$$V^{*2} = D/\beta \tag{1.10.12}$$

which tells us that the discontinuity may propagate both ways with equal speed. There is a faint analogy with the speed of sound, isn't there? But let's not pursue this analogy at this stage. Rather, we observe that, as β approaches zero (so that we would recover the classical constitutive equation), V^{*2} approaches infinity — and indeed with the classical constitutive equation the imposed jump is smoothed out immediately, which means in essence that it has travelled with infinite speed. We see now the solution of the paradox discussed immediately before Remark 1.4.5. Let's be specific: for the speed of propagation to be less than the speed of light, we need D/β to be less than 10^{21} cm^2/s^2. Well, one needs a very small value of β indeed to satisfy that constraint. In fact, even if we only wish (for reasons which are not very clear) V^* to be less than the speed of sound, D/β has to be less than about 10^9 cm^2/s^2, and with a diffusivity of 10^{-2} cm^2/s (a gas) this requires a β value in excess of 10^{-11} s — not a very stringent requirement.

Remark 1.10.9

A relaxation time for an ideal gas can be estimated from the classical Maxwellian theory; it turns out to be of the order of the ratio of viscosity to pressure. For air at ambient conditions, this gives $\beta \approx 10^{-10}$ s.

Remark 1.10.10

For the oscillating experiment described above, β is clearly of the order of $1/\Omega$ when we measure a ϕ value significantly different from zero. With polymeric materials, β values of the order of several seconds are not uncommon. The diffusivity of momentum (or kinematic viscosity) of a polymer melt may well be of the order of 100 cm^2/s, and so V^* may be of the order of a few centimetres per second, well within experimental detectability.

Addition of β in our list of parameters has entirely changed the dimensionality of our problem. In solving something like the Stokes–Rayleigh problem, rather than contemplating the possibility of a similarity solution, we now have as many intrinsic scales as we may wish:

1. time scale: β. Hence $\tau = t/\beta$
2. length scale: $\sqrt{(D\beta)}$. Hence $x = X/\sqrt{(D\beta)}$
3. flux scale: $J\sqrt{(D/\beta)}$. Hence $f = F/J\sqrt{(D/\beta)}$
4. velocity scale: V^*. Hence $V = 1$

Remark 1.10.11

The existence of an intrinsic velocity scale guarantees that in momentum transfer problems a sort of modified Mach number (the ratio of an imposed velocity to the velocity scale) will emerge.

Remark 1.10.12

Notice that all internal scales become degenerate when $\beta = 0$. This guarantees that β approaching zero is going to be a nastily singular limit.

Our dimensionless canonical set becomes:

$$\partial f/\partial x + \partial\theta/\partial\tau = 0 \tag{1.10.13}$$

$$\partial f/\partial\tau + \partial\theta/\partial x = -f \tag{1.10.14}$$

subject to:

$$f(x,0) = \theta(x,0) = 0 \tag{1.10.15}$$

$$\theta(0,\tau) = 1,\ f(0,0+) = 1 \tag{1.10.16}$$

Several observations are in order concerning the BCs. This time the second one of (1.10.15) does not imply the first one, and hence one needs to take both parts into account separately (see Remarks 1.2.2 and 1.3.1). The second one of (1.10.16) is obtained from Kotchine's procedure, which in dimensionless form yields $[f] = V[c]$, $V[f] = [c]$, or equivalently $V = 1$, $[f] = [c]$. The initial unit jump of c corresponds to an unit jump of f.

Now one gets tempted to eliminate the flux between Eqs. (1.10.13) and (1.10.14) to obtain a differential equation for θ. However, this time direct substitution is not feasible, and we have to use cross-differentiation to obtain:

$$\partial^2\theta/\partial x^2 = \partial^2\theta/\partial\tau^2 + \partial\theta/\partial\tau \tag{1.10.17}$$

Equations (1.10.13) and (1.10.15) furnish the additional BC needed, i.e.:

$$\tau = 0, x > 0;\ \partial\theta/\partial\tau = 0 \tag{1.10.18}$$

Now the problem as formulated above can be solved analytically by Laplace transforms. The solution is:

$$\theta = H(\tau - x)\exp(-\tau/2)F(x,\tau) \qquad (1.10.19)$$

where $H()$ is the Heaviside step function, and $F(x,\tau)$ is in the form of an integral of Bessel functions with $F(x,x) = 1$. The Heaviside function guarantees that $c = 0$ to the right of the discontinuity. From Eq. (1.10.19), one obtains $-[\theta] = \theta_L = \exp(-\tau/2)$, i.e., the jump decays exponentially as it propagates through the system.

Remark 1.10.13

> We now have a dissipative discontinuity. What do we mean by that? Inclusion of a classical diffusive term in the constitutive equation gives rise to dissipation (the process is thermodynamically irreversible), but, as seen in Sections 1.3 and 1.4, it smooths out discontinuities. Without a diffusive term, we have discontinuities but no dissipation: in Section 1.2, the imposed jump propagates undisturbed. Here we have a propagating discontinuity, but the jump decays in time.

This is all very nice, but we aren't going to be that lucky often, and sometimes we'll need a numerical solution. The reader is asked to consider the difficulties involved with reproducing numerically the result in Eq. (1.10.19): in the positive x–τ quadrant where the solution is sought, the discontinuity is sure to give us some severe headaches.

So let's try to do some simplifications. First of all, we know from the start that the discontinuity travels at unit speed, and that to its right $f = \theta = 0$. Hence we are really only interested in the solution in the triangular region $\tau > x > 0$ (see Fig. 1.10.3). Let's try to work with the canonical set, rather than with Eq. (1.10.17); numerically, having only first derivatives to worry about is going to make life easier, and the price to be paid, keeping track of both f and θ rather than only of θ, is a minor one. Since the discontinuity travels with unit speed, the substantial derivative along it is:

$$(D/D\tau)_L = (\partial/\partial\tau + \partial/\partial x)_L \qquad (1.10.20)$$

Furthermore, Kotchine's procedure tells us that $f_L = \theta_L$. Now we apply (1.10.20) to both f and θ to obtain:

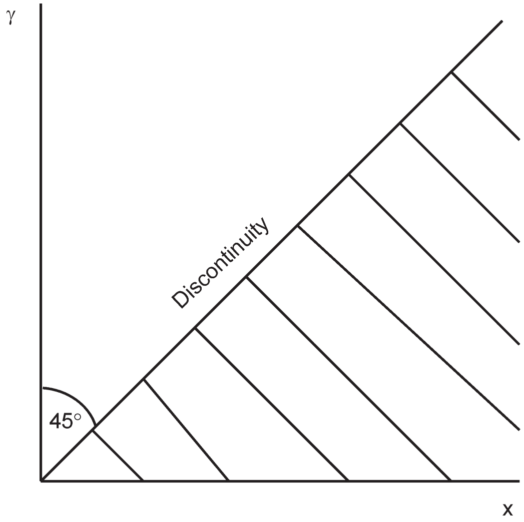

Fig. 1.10.3. Domain of integration for Eq. (1.10.17) (the whole positive quadrant). In the simplified version, the domain of integration is the triangular hatched one.

$$Df_L/D\tau = (\partial f/\partial\tau + \partial f/\partial x)_L \tag{1.10.21}$$

$$D\theta_L/D\tau = (\partial\theta/\partial\tau + \partial\theta/\partial x)_L \tag{1.10.22}$$

Summing the last two and substituting the canonical set one obtains:

$$2D\theta_L/D\tau = 2Df_L/D\tau = -f_L = -\theta_L \tag{1.10.23}$$

which integrates to:

$$\theta_L = f_L = \exp(-\tau/2) \tag{1.10.24}$$

Our way is paved now. We'll integrate numerically the canonical set, forward in time τ, in the triangular domain, using (1.10.24) and the first part of (1.10.16) as BC. No difficulties are expected, and in fact the numerical technique is very stable indeed. We only need a starting point, and this is obtained by finding a linearized solution valid for $\tau = \varepsilon$. This is, quite obviously:

$$f = 1; \;\; \theta = 1 - x \tag{1.10.25}$$

which satisfies both differential equations, and the first part of (1.10.16), exactly, and Eq. (1.10.24) to within $O(\varepsilon)$.

Remark 1.10.14

The technique sketched above is a slightly less trivial version of the one discussed at the end of Section 1.2. It is useful to reread Remark 1.2.9 at this stage.

Now notice that, if we had considered the finite-body problem, we could have fooled ourselves into believing that there is an external length scale. The external length scale L only enters the picture when the travelling discontinuity has reached the position $X = L$, i.e., after a time L/V^*. The solution to the semi-infinite body problem is the exact solution of the finite body one up to that time.

1.11 KOTCHINE'S PROCEDURE REVISITED

In this section, we deal with two issues. First, we analyze the case where there might be, at least in some asymptotic sense, a finite rate of generation per unit surface, rather than per unit volume. Second, we deal with the essential aspects of the case when both relaxation and kinetics are of importance.

We begin by considering again Eq. (1.8.9), which is reported below:

$$\partial F/\partial X + J\partial\theta/\partial t = -J\mathbf{St}\partial w/\partial t \tag{1.11.1}$$

If Kotchine's procedure is applied to (1.11.1), without any consideration of what the constitutive equations for the flux and for $\partial w/\partial t$ may be, one obtains:

$$[F] = JV^*([\theta] + \mathbf{St}[w]) \tag{1.11.2}$$

where V^* is the dimensional speed of propagation of the discontinuity.

Let's look at the physical meaning of this equation. The left hand side is the difference between the fluxes on the two sides of the discontinuity, i.e., it represents a net flux away from the discontinuity. Why is there such a net flux? There are two reasons, which are most easily understood by considering the freezing problem. First, if there is a finite value of $[\theta]$, a finite amount of sensible heat per unit area is liberated as the discontinuity propagates; this is measured by the term $JV^*[\theta]$. Second, there is a finite amount of latent heat liberated per unit interface area; this is measured by the term $JV^*\mathbf{St}[w]$.

Now suppose the FIA holds (as we have seen it will if the constitutive equation for the flux is (1.8.12), so that $[\theta] = 0$. Now we have $[F] = JV^*\mathbf{St}[w]$, and this seems, at least superficially, to be in contrast with Eq. (1.9.8). We need to understand this point thoroughly.

Equation (1.9.8) was obtained by considering that any difference in flux at the two extremes of the interface layer would need a finite rate of accumulation in the interface layer, and that the latter was negligible (of order of the thickness of the interface layer, ε). But now we are saying that there might be a finite generation per unit area —which means that the generation per unit volume is infinitely large, and hence the fact that the interface layer has negligible thickness does not imply that $[F] = 0$. We need to ask ourselves, how realistic is the assumption of a finite rate of generation per unit area?

Mathematically, it is easy to describe a finite rate per unit area R_S, on some surface located at $X = S(t)$, by writing:

$$\partial w/\partial t = R + R_S\delta(X-S) \tag{1.11.3}$$

where $\delta()$ is the Dirac delta function. Notice that the delta function has dimensions of the inverse of its argument, and hence R_S has dimensions of a velocity, while R has dimensions of a frequency.

Remark 1.11.1

Kinetic models with a finite rate per unit volume invariably produce an intrinsic kinetic time scale, as was seen in Section 1.8. Kinetic models with a finite rate per unit area produce instead an intrinsic velocity scale.

If (1.11.3) is blindly substituted into (1.11.1), $[\theta]$ is taken to be zero (in deference to the FIA), and Kotchine's procedure is used, one obtains:

$$[F] = V^*\mathbf{St}R_S \tag{1.11.4}$$

which simply states that the flux jump must compensate the rate of generation per unit area. Correct, but not very illuminating, is it?

And yet, if we think of it, we find out that we use finite rates of generation per unit area here, there, and everywhere. Suffice it to consider the surface of a catalyst where a reaction is taking place at some finite rate per unit area.

Let's first look at the problem from the viewpoint of the Dirac delta function appearing in Eq. (1.11.3). What does it mean that the value of the delta function has dimensions of an inverse length? Perhaps what it means is as follows. The rate of generation per unit volume is everywhere finite, but it becomes very large indeed in a thin layer near $X = S$. The layer is so thin that we are willing to describe it as a surface, so that we substituted for something like a very sharp Gaussian near $X = S$ (a Gaussian with spread ε equal to the thickness of the layer) the delta function. Well, how thin has the layer to be for this to be legitimate within some prescribed level of approximation? Clearly we cannot answer this question unless we have some other length scale with which we may compare ε. But in Section 1.8 we developed the kinetic problem which yields asymptotically the Stefan problem by considering the finite body case — so $\varepsilon << L$ is our obvious criterion. A finite rate per unit area has meaning insofar as we have an external length scale, at least insofar as classical kinetics of the type discussed in Section 1.8 are concerned (but see below).

But let's pursue the implications of the kinetic analysis in Section 1.8 a bit further. We established that, when **Ju** becomes vanishingly small (notice that **Ju** does contain an external length scale), we could assume that $w = w^*(\theta)$ everywhere (to within order **Ju**). This means that, under such conditions, we can rewrite Eq. (1.11.2) as follows:

$$[F] = JV^*([\theta] + \mathbf{St}[w^*]) \tag{1.11.5}$$

Now if the FIA holds, $[\theta] = 0$. But if $w^*(\theta)$ is continuous, $[\theta] = 0$ implies that $[w^*] = 0$, and we get Eq. (1.9.8) again. What's going on here? What is going on is very simple: for the problems we have analyzed, $w^*(\theta)$ is given by Eq. (1.7.16), and is therefore not continuous, so $[\theta] = 0$ does not imply that $[w^*] = 0$. Indeed, we found out in Section 1.8 that:

$$[\theta] \approx -\mathbf{Ju} \tag{1.11.6}$$

and hence, if (1.7.16) holds:

$$[w^*] = -1 \tag{1.11.7}$$

This clears up the paradox, but we are not very comfortable with it, since it seems that we need a rather strange equilibrium function such as (1.7.16) to get any Stefan type problem at all. Chemical reactions would need to be irreversible, and we know from thermodynamics (do we?) that no real reaction is truly irreversible. Some sort of singularity of the limit of irreversible reactions looms on the horizon.

However, (1.7.16) turns out not to be such a strange animal after all when we consider the freezing problem — indeed, at equilibrium, H_2O is water at temperatures above 0°C, and ice below, so (1.7.16) is indeed appropriate. The same conclusion would be reached in a variety of other problems when w is used to identify two different phases.

Remark 1.11.2

The theory of phase equilibria is indeed a theory of discontinuities. This point is discussed in Appendix 1.11.

What, however, if the FIA does not hold — as may be the case if relaxation phenomena play a role? This is a legitimate question, but in order to attack this problem we need to have kinetics in the picture as well — otherwise we lack the w variable to play with. What this means is that we end up with two intrinsic time scales: the relaxation time (which we'll still call β) and the intrinsic time scale of the kinetic phenomenon, which we'll call $\varepsilon\beta$.

Remark 1.11.3

The astute reader has no doubt already foreseen that we are going to consider the case where ε approaches zero.

In the spirit of the preceding remark, we tentatively assume that, to within $O(\varepsilon)$, we can indeed write $w = w^*$ everywhere. Notice that this time no external length scale is needed — we have an internal time scale (the relaxation time β) to compare our kinetic time scale with, and therefore we don't need the diffusion time L^2/D which contains an external length scale. So the balance equation still gives Eq. (1.11.5), and we have no FIA to make everything degenerate. The constitutive equation for the flux is:

$$F + \beta\partial F/\partial t = -DJ\,\partial\theta/\partial X \tag{1.11.8}$$

and Kotchine's procedure now yields:

$$V^*\beta[F] = D[\theta] \tag{1.11.9}$$

When this is combined with Eq. (1.11.5) we get:

$$V^{*2} = (D/\beta)([\theta]/([\theta] + \mathbf{St}[w^*])) \tag{1.11.10}$$

(Notice that we have chosen definitions in such a way that both $[w^*]$ and $[\theta]$ are negative, while **St** is positive. So the whole RHS is positive, as it should, of course, be).

This should be discussed in some detail. First of all, notice that if ε is finite $[w]$ is not equal to $[w^*]$, it is in fact zero, and one recovers again the result $V^{*2} = D/\beta$. However, as ε approaches zero, (1.11.9) holds, and this time we have a nonconstant speed of propagation, in spite of the fact that all parameters are constant. In fact we have a lower speed of propagation than we would have with a finite ε. Finally, this time $[\theta]$ is not zero, and thus we do not need a discontinuous equilibrium function in order to have a finite value of $[w^*]$: given any smooth function $w^*(\theta)$, a finite $[\theta]$ would in general result in a finite $[w^*]$.

We have, however, been very cavalier about the whole analysis, and we need to be a bit more careful. Let's use all the intrinsic scales discussed in Section 1.10, so that the dimensionless formulation of our problem is:

$$\partial f/\partial x + \partial\theta/\partial\tau = -\mathbf{St}\partial w/\partial\tau \tag{1.11.11}$$

$$\partial f/\partial\tau + \partial\theta/\partial x = -f \tag{1.11.12}$$

$$\varepsilon\partial w/\partial\tau = r(w, \theta) \tag{1.11.13}$$

A "primary" discontinuity of θ still propagates at unit speed. Now we perform an analysis similar to the one at the end of Section 1.10, and, observing that, to the left of the primary discontinuity, $w = 0$, and hence $\partial w/\partial\tau = r(0,\theta_L)/\varepsilon$ to obtain:

$$2D\theta_L/D\tau = -\theta - (\mathbf{St}/\varepsilon)r(0,\theta_L) \tag{1.11.14}$$

Now, if ε approaches zero, the second term on the RHS predominates, and if the $r()$ function is even moderately well behaved (notice that $\partial r/\partial\theta > 0$) θ_L will decay in time proportionally to something like $\exp(-\mathbf{St}\tau/\varepsilon)$, i.e., very rapidly indeed. In fact so rapidly that, to the left of the primary discontinuity, $\partial w/\partial\tau$ will be negligible in spite of the fact that ε approaches zero — it will be proportional to something like $(1/\varepsilon)\exp(-\mathbf{St}\tau/\varepsilon)$, and exponentials predomi-

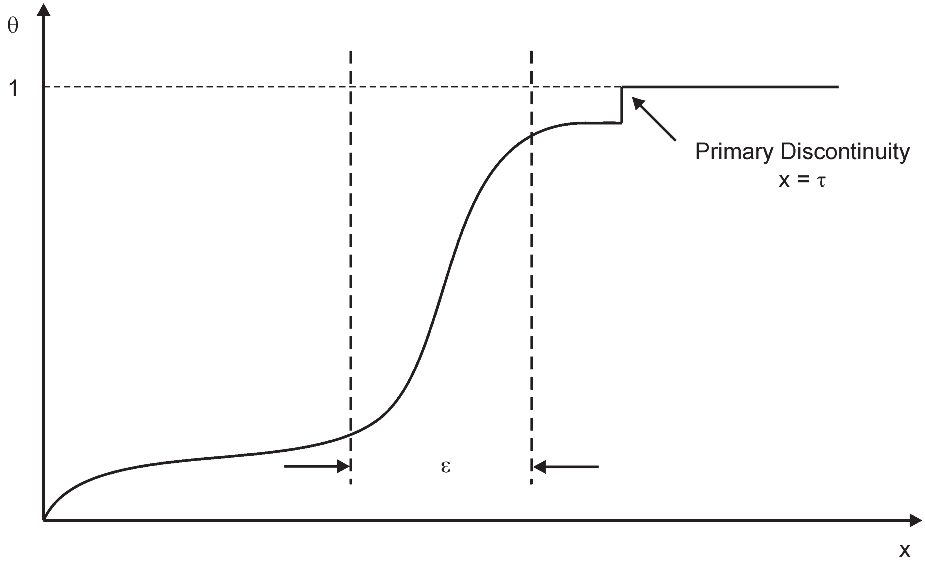

Fig. 1.11.1. Sketch of the expected profile of θ when ε approaches zero.

nate on any power, don't they? So to all practical purposes we may neglect the primary discontinuity, and focus on the secondary one, the propagation of which is governed by Eq. (1.11.10). If the reader is not thoroughly convinced by this argument (as he would undoubtedly be justified to be), it is carried out in detail by ourselves [1987] elsewhere. For a qualitative sketch (see Fig. 1.11.1).

Since on the right of the secondary discontinuity $w \approx 0$, $\theta \approx 0$, the numerical integration of the $\varepsilon = 0$ problem can be carried out as for a standard free boundary problem, with the motion of the boundary governed by Eq. (1.11.10). One cannot, however, duplicate the procedure which gives explicitly θ on the left of the discontinuity, though it can easily be done numerically. This is also discussed in our paper.

Now notice that Eq. (1.11.10) gives, at $\tau = 0$:

$$\tau = 0;\ \ V^{*2} = (D/\beta)(1/(1 + \mathbf{St})) \tag{1.11.15}$$

i.e., the surface of discontinuity propagates with a finite initial speed — contrary to the Stefan problem case where it propagates at an initial speed of infinity (V^* is proportional to $1/\sqrt{t}$. It is a bit cumbersome, but straightforward, to convince oneself that, at very large values of τ, the equations degenerate to those of the Stefan problem. Thus S is proportional to t at low times, and to $\sqrt{t}$ at large times — perhaps the experimental observation with polymers that S is proportional to t^n with $1 \geq n \geq ½$ (discussed in Section 1.7) can be explained this way.

Appendix 1.11

In this appendix, we pave the way for some points to be discussed in the next Chapter. As will be seen, we obtain some results which are usually regarded as pertaining to thermodynamics, by analyzing the mathematics of discontinuities of the type discussed in this chapter. In particular, we point out the fact that the theory of equilibrium phase transitions is essentially a theory of discontinuities.

Consider a function of two variables $F(x,y)$. Let $x = X(y)$ be a curve in the x–y plane across which F is continuous, $[F] = 0$, but its derivatives F_X and F_Y suffer a discontinuity. Let D/Dy be the substantial derivative along the $X(y)$ curve, say for any quantity Ω:

$$D\Omega/Dy = \Omega Y + \Omega_X \mathrm{d}X/\mathrm{d}y \tag{1.11.16}$$

The operators [] and D/Dy are both linear, and hence they commute:

$$D[]/Dy = [D/Dy] \tag{1.11.17}$$

By definition:

$$0 = D[F]/Dy = [DF/Dy] = [F_Y] + [F_X]\mathrm{d}X/\mathrm{d}y \tag{1.11.18}$$

and hence:

$$\mathrm{d}X/\mathrm{d}y = -[F_Y]/[F_X] \tag{1.11.19}$$

We are not cheating with the notation. [] does describe a jump, and if x is interpreted as position and y as time, $\mathrm{d}X/\mathrm{d}y$ does describe the "propagation" of the discontinuity.

Now let $X'(y)$ be a curve in the x–y plane across which F, F_X and F_Y are continuous, but F_{XX}, F_{YY} and F_{XY} are not. By definition:

$$0 = D[F_X]/Dy = [F_{XY}] + [F_{XX}]\mathrm{d}X'/\mathrm{d}y \tag{1.11.20}$$

$$0 = D[F_Y]/Dy = [F_{YY}] + [\mathrm{F}_{\mathrm{XY}}]\mathrm{d}X'/\mathrm{d}y \tag{1.11.21}$$

and hence:

$$\mathrm{d}X'/\mathrm{d}y = -[F_{YY}]/[F_{XY}] = -[F_{YY}]/[F_{XY}] \tag{1.11.22}$$

It follows that:

$$\pi = [F_{YY}][F_{XX}]/[F_{XY}]^2 = 1 \tag{1.11.23}$$

These results are trivially deduced from the definitions, and quite obviously have no thermodynamic content whatsoever. The only assumption on which the results above are based is that F is indeed an unique function of x and y.

Now, however, let's suppose that F is free enthalpy (which at equilibrium is continuous at a phase transition, since it has the same value in both phases at the equilibrium point), and let's interpret x and y as temperature and pressure, so that $-F_X$ is entropy, and F_X is volume. In an ordinary phase transition, the jump of entropy is finite (there is a finite latent heat) and so is the jump of volume (the two phases have different densities). So the first part of the analysis above holds for this case, and (1.11.19) is now seen to be nothing else than the Clausius Clapeyron equation, which governs the variation of the equilibrium pressure transition with temperature.

Remark 1.11.3

> The reader is asked to recall how he first learned the Clausius Clapeyron equation. If he's anything like we'd like our readers to be, he has carefully forgotten it, of course, since what can one do with arguments about heat being added reversibly and the like? Is it of any comfort to find out that the equation can be derived from a trivial application of the definition of substantial derivative?

But let us look at the second part of our simple minded analysis, and let us ask ourselves what the second derivatives appearing here, there and everywhere physically represent. $-F_{XX}$ is now constant pressure specific heat divided by temperature (i.e., $(\partial S/\partial T)_P$, to use the thermodynamicists' derivatives with subscripts), $-F_{YY}$ is the isothermal compressibility, and F_{XY} is both $-(\partial S/\partial p)_T$ and $(\partial V/\partial T)_P$ (the latter is proportional to the coefficient of cubic expansion. There will, again, be a sizable prize for anybody who can suggest a way of measuring the first one, except by measuring the coefficient of cubic expansion). Equations (1.11.22) are now recognized as the Ehrenfest equations [1933] holding for "secondary" phase transitions (at which there is no discontinuity of volume and a zero latent heat), and Eq. (1.11.23) as the statement that the Prigogine–Defay ratio π is unity for such transitions. It is perhaps obvious that it would be easy now to start playing with third derivatives, and this has in fact been done; but, unfortunately, even those transitions which were believed to be secondary ones turn out not to obey the Ehrenfest equations. And not, of course, because any of the algebra above is wrong; but because the basic assumption (i.e., that everything depends only on temperature and pressure) is not valid for such transitions.

PROBLEMS

1.1 The highway speed of cars has been determined to be given, rather exactly, by $U = U^*[1 - (C/J)^2]$. Rework the Red Traffic Light and the Traffic Jam Ahead problems for this case.

1.2 Suppose the forcing function on a PFR is not a step function, say it is given by $c(0,t) = c_0[1 - \exp(-kt)]$. Determine the response at the exit, possibly also for the case where there is axial diffusion but the Peclet number is large. Does anything of interest happen when kL/U approaches zero?

1.3 Is the velocity of sound in liquids larger or smaller than in gases?

1.4 Suppose the slab considered in Section 1.6 is exposed to ambient air, and is thus subject to the annual cycle of temperature. How would you calculate the temperature at the centre of the slab during the year?

1.5 Based on the model in Section 1.7, estimate the combustion time for a coal particle with an initial diameter = 1 cm.

1.6 Combustion of methane ($CH_4 + 3O_2 = CO_2 + 2H_2O$) has been determined to have a rate which is zero order in methane and of order 1/2 in oxygen; the "kinetic constant" is 10 gmoles$^{1/2}$ cm$^{-3/2}$ s^{-1}. The reaction is very nearly irreversible. Estimate an appropriate intrinsic time scale for an initially equimolar mixture at 1 atm pressure.

1.7 Calculate the total work done over a length of time *T* (containing very many cycles) on the system discussed in Section 1.10. How does it depend on frequency, other things being equal? Explain.

SUGGESTED READING

1.1 The standard textbook on Transport Phenomena is of course the one by Bird, Stewart and Lightfoot [1960], with which we assume our reader is familiar. A useful companion book is the one by Denn [1986], which deals with the general problem of the mathematical modelling of physical problems.

1.2 The highway traffic problem in Section 1.2 is taken from the book by Beltrami [1987] which is a useful and easily readable source for mathematical techniques of interest. The original paper by Kotchine was published in 1926; Kotchine's theorem is discussed in detail by Truesdell and Toupin [1960] and by Foraboschi [1974]; the latter is a book which we wholeheartedly recommend to any reader who is willing to tackle with the Italian language.

1.3 Peclet number corrections for axial diffusion in a tubular reactor are discussed thoroughly by Froment and Bischoff [1979].

1.4 The Stokes–Rayleigh problem was formulated first for heat transfer by Fourier [1822]; for momentum transfer, it was formulated by Stokes [1848] and the solution was given by Rayleigh [1911]; the mass transfer version was given by Higbie only in 1935. The solution of the infinite speed paradox for the mass transfer case was first given by Mueller and Villaggio [1976].

1.5 Shock waves in gases are discussed in most textbooks on gas dynamics and on combustion. A good introductory book is the one by Barnard and Bradley [1985]. The original papers are by Hugoniot [1887, 1888] and Rankine [1870]. The velocity of sound was first studied by Newton [1687]; the Laplace [1846] analysis goes back to 1816.

1.6 The classical textbook for the unsteady diffusion equation is the one by Crank [1975]. Anomalies in the diffusion behaviour in polymers are reviewed by Astarita [1987b].

1.7 The original paper by Stefan was published in 1891. Stefan-like problems are the subject of an ample mathematical literature; a review can be found in the book edited by Fasano and Primicerio [1983]. The coal combustion version is discussed by Ramachandran and Doraiswamy [1982]. The polymer diffusion version is discussed by Astarita [1983].

1.8 The emergence of Stefan problems as asymptotes of kinetic ones is discussed for the heat transfer case by Astarita and Kenny [1987]; for the coal combustion case by Chakraborty and Astarita [1987], and for a more complex heterogeneous reaction problem by Astarita [1987a]. The validity of the $w = w^*$ approximation (with a finite rate) in the limit of instantaneous reactions is discussed in detail by Goddard et al. [1970].

1.9 The material in Section 1.9 is of course mainly concerned with thermodynamics; the recent textbook by Astarita [1989] is recommended for a deeper understanding of this whole area.

1.10 The first explicit mention of Eq. (1.209) in the context of unsteady heat transfer is due to Cattaneo [1948]. It can, however, be traced back to Maxwell himself [1890]; relaxation in a Maxwellian gas is discussed by Grad [1958], Ikenberry and Truesdell [1956] and Truesdell [1956]. Dissipative discontinuities in systems with relaxation are discussed in great detail in the book by Coleman et al. [1965].

1.11 The material in Section 1.11 is largely drawn from our paper [1987]. The point concerning Appendix 1.11 is discussed in detail by Astarita, Wissinger and Paulaitis [1989].

REFERENCES

Astarita, G., in: Free Boundary Problems: Theory and Applications, A. Fasano and M. Primicerio (Eds.), Pitman, Houston, 1983.
Astarita, G., Chem. Eng. Sci., 42 (1987a) 413.
Astarita, G., Adv. Transp. Proc., 5 (1987b) 339.
Astarita, G., Thermodynamics. An Advanced Textbook for Chemical Engineers. Plenum Press, New York, 1989.

Astarita, G. and J.M. Kenny, Chem. Eng. Comm., 53 (1987) 69.

Astarita, G., R. Wissinger and M.E. Paulaitis, J. Polym. Sci., B, Polym. Phys., 27 (1989) 2105.

Barnard, J.A. and J.N. Bradley, Flame and Combustion. Chapman and Hall, London, 1985.

Beltrami, E., Mathematics for Dynamic Modelling. Academic Press, Boston, 1987.

Bird, R.B., W.E. Stewart and E.N. Lightfoot, Transport Phenomena. Wiley, New York, 1960.

Cattaneo, C., Atti Sem. Mat. Fis. Univ. Modena, 3 (1948) 83.

Chakraborty, A.K. and G. Astarita, Ind. Eng. Chem. Fund., 26, (1987) 972.

Coleman, B.D., M.E. Gurtin, I. Herrera and C. Truesdell, Wave Propagation in Dissipative Materials. Springer-Verlag, Berlin, 1965.

Crank, J., Mathematics of Diffusion. Clarendon Press, Oxford, 1975.

Denn, M.M., Process Modelling. Longman, New York, 1986.

Fasano, A. and M. Primicerio (Eds.), Free Boundary Problems: Theory and Applications. Pitman, Houston, 1983.

Foraboschi, F.P., Principii di Ingegneria Chimica. UTET, Torino, 1974.

Froment, G. and K.B. Bischoff, Chemical Reactor Analysis and Design. Wiley, New York, 1979. 2nd edn. 1989.

Fourier, J., Oeuvres, Vol. I: Théorie Analytique de la Chaleur. Paris, 1822.

Goddard, J.D., J.S. Schultz and R.J. Basset, Chem. Eng. Sci., 25 (1970) 665.

Grad, H., Principles of the Kinetic Theory of Gases. Enc. of Phys., Vol. XII. Springer-Verlag, Berlin, 1958.

Higbie, R., Trans. Am. Inst. Chem. Engrs., 31 (1935) 365.

Hugoniot, H., J. Ecole Polyt., 57 (1887) 3; 58 (1888) 1; J. Math. Pures Appl., 3 (1887) 477; 4 (1887) 153.

Ikenberry, E. and C. Truesdell, J. Ratl. Mech. Anal., 5 (1956) 1.

Juttner, F., Z. Phys. Chem., 65 (1909) 595; Z. Math. Phys., 58 (1910) 385.

Kotchine, N.E., Rend. Circ. Matem. Palermo, 50 (1926) 305.

Laplace, P.S. Oeuvres. Imprimerie Royale, Paris, 1846. The work on the speed of sound was originally published in 1816.

Maxwell, J.C. Scientific Papers. Cambridge University Press, 1890.

Muller, I. and P. Villaggio, Meccanica, 11 (1976) 191.

Newton, I., Philosophiae Naturalis Principia Mathematica. Streeter, London, 1678.

Ocone, R. and G. Astarita, A.I.Ch.E.J., 33 (1987) 423.

Ramachandran, P.A. and L.K. Doraiswamy, A.I.Ch.E.J., 28 (1982) 881.

Rankine, W.J.M., Phil. Trans. Roy. Soc. London, 160 (1870) 277.

Lord Rayleigh, Phil. Mag., 6–21 (1911) 697.

Stefan, J., Ann. Phys. und Chem., N.F., 42 (1891) 269.

Stokes, G.G., Cambr. Dubl. Math. J., 3 (1848) 209; Phil. Mag., 23 (1848) 349.

Thiele, E.W., Ind. Eng. Chem., 31 (1939) 916.

Truesdell, C.A., J. Ratl. Mech. Anal., 5 (1956) 55.

Truesdell, C. and R. Toupin, The Classical Field Theories. Enc. of Physics, Vol. III/1, Springer-Verlag, Berlin, 1960.

Chapter 2

Thermodynamics

2.1 INTRODUCTION

Although of course there is no motivation strong enough to make the study of thermodynamics attractive, we hope we have offered enough of it in the first chapter to make it at least not entirely insufferable. In this chapter, we deal with the thermodynamics of heat, momentum and mass transfer, as well as with the essentials of the thermodynamics of systems with internal state variables (say, typically but not universally, thermodynamics of chemical reaction kinetics). Of course we are dealing with transport phenomena, which, by definition, take place only in systems which are not at equilibrium, and hence we need to approach thermodynamics from the viewpoint of non-equilibrium theory.

We do not discuss in the detail they deserve the fundamental concepts of thermodynamics; the reader is referred to the book by Astarita [1989], and in particular to Chapters 1, 2 and 3 of it, for those. Rather, we concentrate on the conceptual aspects of thermodynamics as applied to the analysis of transport phenomena — which, by being non-equilibrium processes, have very much to do with the second law of thermodynamics. The first law is a balance of energy, and balance equations are commonplace in the theory of transport phenomena; the first law, by virtue of expressing the principle of conservation of energy, just lines up together with the principles of conservation of mass, linear momentum, angular momentum. It is quite different with the second law, which is not a balance equation.

In this first section, we limit ourselves to a formulation of the first and second laws of thermodynamics which can be used in cases where state variables may have nonzero gradients, or, in the case of internal state variables, may have nonequilibrium values. We assume the reader is familiar with the general field formulation of the principles of conservation of mass, linear momentum, and angular momentum. We also assume our reader has knowledge of such things as vectors, tensors, and functional analysis; but just in case these assumptions might be too optimistic, we give some literature references and we include some fundamental material in the last section.

Let us try to write down a reasonably general formulation of the first law of thermodynamics. It has to be applicable to systems which are not at equilibrium, and hence under conditions where, e.g., temperature may well not be constant in space or in time. We therefore choose to write it in the form of a field equation — a balance of total energy applied to a system which is the neighbourhood of a material point.

Remark 2.1.1

> The reader is asked to contrast this with the usual development of "Thermodynamics of Irreversible Processes", where a rather unconvincing attempt is generally made at writing equations directly for macroscopic systems.

We assume that such a system does not exchange matter with its surroundings (in the terminology of thermodynamics, it is a closed system); this, as will be seen, excludes diffusive phenomena from consideration, since when there is diffusion there is no way of identifying a "closed" system, not even a differential one (as will be discussed in more detail in the section on diffusion).

Now the total energy per unit mass in the neighbourhood of a material point is $E + K$, where E is internal energy per unit mass, and K is kinetic energy per unit mass, i.e.:

$$K = \tfrac{1}{2}\mathbf{v}\cdot\mathbf{v} = \tfrac{1}{2}v^2 \tag{2.1.1}$$

where v is the modulus of the velocity vector $\mathbf{v}$. Now let P be the rate at which work is being done on the system considered per unit volume, q the rate at which heat is being supplied to the element considered per unit volume, and Φ the local density. We suggest that the first law can be written as:

$$\Phi \mathrm{D}(E + K)/\mathrm{D}t = P + q \tag{2.1.2}$$

where $\mathrm{D}/\mathrm{D}t$ is the intrinsic substantial derivative ($\mathrm{D}/\mathrm{D}t = \partial/\partial t + \mathbf{v} \cdot \mathrm{grad}$). Equation (2.1.2) is, after all, a balance equation; the rate of accumulation of total energy equals the rate at which energy is supplied to our system. Let's assume that we are convinced (2.1.2) is correct. Now what we need is to write equations for both P and q.

Remark 2.1.2

The appearance of the intrinsic substantial derivative is of course related to the fact that our system is the neighbourhood of a material point, and thus it experiences rates of change which in the laboratory frame are described by $\mathrm{D}/\mathrm{D}t$.

As far as P is concerned, we restrict attention to the case where the only work done is mechanical work. There are two kinds of forces acting: a body force per unit mass $\mathbf{g}$ (usually gravity), and contact forces $\mathrm{d}\mathbf{f}$ acting through an elementary surface $\mathrm{d}\mathbf{s}$ bounding the element considered. The internal stress tensor $\mathbf{T}$ is defined as:

$$\mathrm{d}\mathbf{f} = \mathbf{T} \cdot \mathrm{d}\mathbf{s} \tag{2.1.3}$$

(with some uncertainty about the sign still left unresolved, see Remark 1.10.6). Now how can we resolve this uncertainty? Let's look at Fig. 2.1.1. The area vector $\mathrm{d}\mathbf{s}$ may be oriented either way, and so may the force vector $\mathrm{d}\mathbf{f}$. Let's say we have oriented them both towards the interior part of the system we are considering. Then $\mathbf{T}$ is indeed a flux of momentum — it represents the influx of momentum into our system. That makes sense, and it makes compressive forces positive —which is not the convention used by people working in solid mechanics. Now we can calculate the total work done by internal stresses on

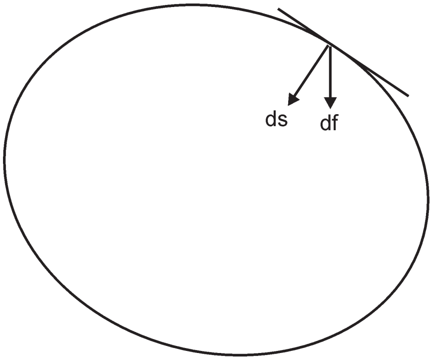

Fig. 2.1.1. Sign convention for the stress tensor.

the finite system considered in Fig. 2.1.1 as the integral of d**f** over the whole bounding surface. We then apply Gauss' theorem to conclude that the net rate of work per unit volume done on an element by contact forces, P_C, is given by:

$$P_C = -\operatorname{div}(\mathbf{T}\cdot\mathbf{v}) \tag{2.1.4}$$

Here, use has been made of the fact that the stress tensor **T** is symmetric. The rate of work done per unit volume by body forces is of course $\Phi\mathbf{g}$, where Φ is density, and hence one obtains:

$$P = -\operatorname{div}(\mathbf{T}\cdot\mathbf{v}) + \Phi\mathbf{v}\cdot\mathbf{g} \tag{2.1.5}$$

The balance of momentum, with no constitutive assumption whatsoever, can be written as:

$$\Phi \mathrm{D}\mathbf{v}/\mathrm{D}t = -\operatorname{div}\mathbf{T} + \Phi\mathbf{g} \tag{2.1.6}$$

where the continuity equation has been used (with no assumption of incompressibility). If Eq. (2.1.6) is multiplied scalarly by **v**, the result is:

$$\Phi \mathrm{D}K/\mathrm{D}t = -\mathbf{v}\cdot\operatorname{div}\mathbf{T} + \Phi\mathbf{v}\cdot g \tag{2.1.7}$$

(If one here makes the constitutive assumption that internal stress is an isotropic pressure, $\mathbf{T} = p\mathbf{I}$, Eq. (2.1.7) yields the Bernoulli equation).

Remark 2.1.3

For steady state, and with **g** expressible as $-g\operatorname{grad}h$, the Bernoulli equation becomes $\mathbf{v}\cdot \operatorname{grad}(\tfrac{1}{2}v^2 + p/\Phi + gh) = 0$, which implies that the term in brackets is constant along any streamline. This does, unfortunately, look like some statement about energy being conserved, which it absolutely is not —Bernoulli did his work with no thermodynamics available, and he did not possess the concept of energy.

One can now subtract Eq. (2.1.7) from (2.1.5) to obtain:

$$W = P - \Phi \mathrm{D}K/\mathrm{D}t = -\mathbf{T}:\operatorname{grad}\mathbf{v} \tag{2.1.8}$$

where : indicates the scalar product of two tensors (i.e., the sum of the products of homologous elements of the matrices of the Cartesian components of the two tensors). W represents the *net* rate of work done on the element, i.e., that part of the total rate of work which does not result in the accumulation of kinetic energy. Notice that the body force **g** does not contribute anything to W, independently of whether it can or cannot be expressed as the gradient of a scalar.

The velocity gradient grad**v** can be decomposed in the sum of its symmetric part **D** (the rate of strain) and its skew part **W** (the rotation rate):

$$\mathbf{D} = \tfrac{1}{2}(\mathrm{grad}\mathbf{v} + \mathrm{grad}\mathbf{v}^{\mathrm{T}}) \tag{2.1.9}$$

$$\mathbf{W} = \tfrac{1}{2}(\mathrm{grad}\mathbf{v} - \mathrm{grad}\mathbf{v}^{\mathrm{T}}) \tag{2.1.10}$$

and, since **T** is symmetric, **W** contributes nothing to the net rate of work W ($\mathbf{T}\!:\!\mathbf{W} = 0$, since it is the scalar product of a symmetric and a skew tensor), and one thus has:

$$W = -\,\mathbf{T}\!:\!\mathbf{D} \tag{2.1.11}$$

Let's interpret this result. If **D** is zero, $\mathrm{grad}\mathbf{v} = \mathbf{W}$, and the motion is a rigid-body one (that is why **D** is the rate of strain: it is that part of grad**v** which results in deformation of the material). Thus one concludes that *rigid-body motions give a zero net rate of work*. The work term appearing in thermodynamics is zero unless the material considered undergoes deformation.

The balance of energy reduces to:

$$\Phi \mathrm{D}E/\mathrm{D}t = W + q \tag{2.1.12}$$

which shows how degenerate the rigid-body case is: for a rigid-body motion, the only way to increase internal energy is by heating, and all mechanical work results in accumulation of kinetic energy: there is no mechanism for interchange between the two forms of energy, and no thermodynamics is left.

Remark 2.1.4

Notice that the statement that a body is a rigid body is a constitutive assumption. This begins to show that thermodynamics tells us what are the consequences of constitutive assumptions we may wish to make, but it does not tell us whether those constitutive assumptions are or are not realistic.

We now come to one of the great mysteries of thermodynamics. When dealing with mechanics, as the discussion above has shown, calculation of the work done per unit time requires careful algebra with vectors — work is the scalar product of force and displacement vectors, after all. No such complication is usually encountered in thermodynamics, where work is written something like $-p\mathrm{d}V$, with V being volume. What's going on here? Well, as we have seen, the net rate of work appearing in (2.1.12) is nonzero only if the body considered is being deformed, and the description of deformation is mathematically complex. Thermodynamicists avoid mathematical complexities as much as possible, and that is a commendable attitude. Thus they observe that one can restrict attention to only the very simplest form of deformation, i.e., a change in density. If that is the case, the rate of strain tensor takes a particularly simple form, say:

$$3\mathbf{D} = (\mathrm{div}\mathbf{v})\mathbf{I} = -(1/\Phi)\mathbf{I}\mathrm{D}\Phi/\mathrm{D}t = \Phi\mathbf{I}\mathrm{D}V/\mathrm{D}t \tag{2.1.13}$$

where the equation of continuity has been used to express $\mathrm{div}\mathbf{v}$ in terms of $\mathrm{D}\Phi/\mathrm{D}t$, and $V = 1/\Phi$ is the specific volume. The matrix of the Cartesian components of **D** is diagonal for this case, and all three elements on the main diagonal are $(1/3)\mathrm{D}\ln V/\mathrm{D}t$.

With Eq. (2.1.13), we have expressed the kinematics in terms of $\mathrm{D}V/\mathrm{D}t$. If we wish to recover, at least in some special case, the classical result that $\mathrm{d}w = -p\mathrm{d}V$, we need to know what pressure p is — the stress tensor could well not be isotropic, and thus we need a *definition* of pressure.

Now this seems preposterous; we all know what pressure is. Do we? In a fluid at rest, the force acting through any elementary surface is orthogonal to the surface itself, proportional to its area, and the proportionality constant, which we call pressure, is independent of the orientation of the surface. Now in a material undergoing deformation none of these three things needs be true; so what is pressure? We have a student who comes to the rescue this time. His name is Bob, and he is a no-nonsense type, with a lot of engineering sense. He suggests that pressure is whatever a pressure gauge would read. Is it? Is the reading of the pressure gauge independent of its orientation? So if we have pressure gauges at different orientations and they don't give the same reading, what is pressure? "Some kind of average" is Bob's suggestion. Well, what about the average of three mutually orthogonal pressure gauges? That does seem reasonable, and it corresponds to the following definition of pressure:

$$p = \mathrm{tr}\mathbf{T}/3 \tag{2.1.14}$$

Remark 2.1.5

There are subtleties with Eq. (2.1.14), and some of these will be discussed in Section 2.3. There is, however, one subtlety which can be disposed of right away: provided the three gauges are orthogonal to each other, the average of their readings is independent of their individual orientations, and thus (2.1.14) is a legitimate definition. The fact that the average is independent of the orientation of the three gauges is what mathematicians express by saying that the trace of a tensor is an invariant.

If this is substituted in (2.1.11), one obtains:

$$W/\Phi = -p\mathrm{D}V/\mathrm{D}t \tag{2.1.15}$$

If one now defines dw as $(W/\Phi)\mathrm{d}t$ (i.e., dw is the work done per unit mass on the element considered over the differential time interval dt), and dV as $(\mathrm{D}V/\mathrm{D}t)\mathrm{d}t$ (i.e., dV is the change of specific volume in the same time interval), one obtains $\mathrm{d}w = -p\mathrm{d}V$, and thermodynamicists are happy and can tell us that they had known that all the time and all our algebra is just an exercise in futility. Well, maybe it is, but there are of course infinitely many cases where the rate of strain tensor is not given by (2.1.13) (the motion is not a pure compression or expansion), and in such cases W needs to be calculated from (2.1.11), not from (2.1.15).

Remark 2.1.6

Now thermodynamicists would want us to believe that dw is, yes, a differential, but somehow it is not "exact", whatever that may mean. They sometimes reinforce this idea by writing it as δw rather than dw. Then on occasion thermodynamicists don't want dw (or δw, as the case may be) to be quite so inexact as they have made it to be, and for the occasion they write δw_{REV}. We won't worry with all this; W is a finite quantity, and if we do our algebra right it will presumably be exact.

We now need to write an equation for q, and we do that by recognizing that there are again two mechanisms for heat transfer: conduction, as described by a heat flux vector $\mathbf{q}$, and radiation, as described by a rate of heat supply per unit mass Q. (We exclude convection from consideration, since we are dealing with a closed system, so that heat cannot be supplied by convection). The first mechanism is analogous to a contact force — heat flows by conduction only between parts which are in direct contact with each other. Q is analogous to **g** — it acts directly on interior parts of a body. Thus we write:

$$q = -\operatorname{div}\mathbf{q} + \Phi Q \qquad (2.1.16)$$

Remark 2.1.7

Thermodynamicists do the same kind of thing with q as they do with W. First, they call $q\mathrm{d}t$ the heat differential $\mathrm{d}q$ (they never want time to appear in their equations). Next, they again say that $\mathrm{d}q$ is perhaps a differential, but not an exact one; so they write it as δq. Then they proceed to consider the case where it may perhaps be exact, and they write δq_{REV}. Finally, they define entropy through $\mathrm{d}S = \delta q_{\mathrm{REV}}/T$. Well, should any one of our readers like that, Remark 2.1.11 below becomes irrelevant.

So far, so good; we have not made any constitutive assumption, (though we know that the equations which have been developed will need the help of constitutive equations to be of any pragmatical use). Two conceptual points deserve attention. First, temperature has not yet appeared in our equations. Second, and perhaps even more surprisingly, there are no differentials in any one of the equations, so that luckily the obscure question about whether they are or are not exact does not arise (see Remarks 2.1.6–7).

Remark 2.1.8

Temperature has not appeared yet, but if it had we would have been perfectly comfortable with it. And so we are comfortable with E and q — and yet, before writing the first law, what did we know about what E and q mean? They are, let's admit it, undefined, primitive concepts. If one uses words from everyday life for undefined concepts (such as energy and heating), the fact that the concepts are undefined slips by unnoticed. (In the everyday use of the word energy, one is accustomed to a phrase like "save energy". And yet if we believe the first law, energy cannot be destructed and it cannot be created, so what does "save energy" mean?)

Now we need to do the nasty thing — write the second law of thermodynamics. What is it? It is a statement that not only there are no free lunches, but in fact that one is likely to have to pay for more for a lunch than it is worth. Possibly as little as it is worth, if we are lucky, but not less. (This may seem a bit on the joking side to the reader, but in fact "there is no such thing as a free lunch" is at least on a par with other statements of the second law, such as "heat flows in the direction of increasing temperature", or "there cannot be a perpetual motion machine of either the first or the second kind". In fact, it has the advantage that, while there may be some people who do not know what heat is, and even more people who do not know what a perpetual motion machine of any kind is, there are very few who do not know what a lunch is, let alone a free one). So what we need to do is to write an equation of the type of a balance equation, but with a ≥ sign instead of an = sign. A "balance" of what? Well, entropy, of course, whatever it might be. (It must be something like the cost of a lunch — it keeps going up). Let j be the rate of supply of entropy per unit volume (something like the actual value of the lunch), and let S be the entropy per unit mass. In order to express our pessimism about the actual cost of lunches, we write:

$$\Phi \mathrm{D}S/\mathrm{D}t = j \tag{2.1.17}$$

Remark 2.1.9

This time the fact that entropy is a primitive undefined concept has not slipped by unnoticed, in spite of our witty remarks about lunches. Entropy is not a word from everyday language.

Remark 2.1.10

Let's keep the analogy with free lunches alive a bit longer. So we often pay for a lunch more than it is worth, that's the second law. But then what is the first law? Here it is: when we do pay for a lunch more than it is worth, somebody pockets the difference. It doesn't get lost, but we can't get it back.

Now we need an equation for j, and we postulate that again there are two mechanisms, a conductive and a radiative one, so that we write:

$$\begin{aligned} j &= -\,\mathrm{div}(\mathbf{q}/T) + \Phi Q/T = \Phi Q/T - (\mathrm{div}\mathbf{q})/T + (\mathbf{q}\cdot\mathrm{grad}T)/T^2 = \\ &= q/T + (\mathbf{q}\cdot\mathrm{grad}T)/T^2 \end{aligned} \tag{2.1.18}$$

Remark 2.1.11

We are not saying, in any sense, that **q** and Q in Eq. (2.1.18) should bear a REV subscript. They are the real-life heat flux and heat radiation, and things in real life have a tendency not to be reversible.

Remark 2.1.12

> Temperature, of course, is another primitive undefined concept. We have some instinctive idea of what it means — but how are we going to measure it? It's not such an idle question, since Europeans and Americans seem to be unable to agree on the answer. They both take the melting of ice and the boiling of water as the reference points, but while Europeans choose to assign the values 0 and 100 to those points, Americans choose the much more sensible values of 32 and 212. They then both divide the interval in equal parts — though how to do that is a mystery. There can, of course, be at most one temperature scale for which (2.1.18) holds — that's the "absolute temperature".

We can now eliminate q between Eqs. (2.1.12) and (2.1.18), introduce free energy $A = E - TS$, do a bit of algebra, to obtain:

$$0 = \mathrm{D}A/\mathrm{D}t + S\mathrm{D}T/\mathrm{D}t - W/\Phi + (\mathbf{q}\cdot\mathrm{grad}T)/\Phi T \tag{2.1.19}$$

Let's try our luck with a very simple case: steady state heat conduction through a stationary body. All D/Dt's are zero, $W = 0$ because the body is stationary, ΦT is certainly positive, and thus we obtain:

$$\mathbf{q}\cdot\mathrm{grad}T \leq 0 \tag{2.1.20}$$

Hence, heat flows in the direction of decreasing temperature. Well, that seems natural, and we may proceed.

Remark 2.1.13

> Actually, what Eq. (2.1.20) tells us is that the steady state conductive heat flux in a stationary body has a positive component in the direction of decreasing temperature, which is a significantly more qualified statement.

Minus the RHS of Eq. (2.1.19) may be called a rate of energy dissipation Z (so that $Z = 0$). But that is not good enough for the thermodynamicists, they prefer a rate of entropy production, and they want it per unit volume — the reasons for this preference are best left uninvestigated. Well, let's make them happy. We multiply the RHS by Φ/T and call the result the rate of entropy production per unit volume $\Gamma = \Phi Z/T = 0$. So we obtain, for steady state heat conduction through a stationary body:

$$\Gamma = -(\mathbf{q}\cdot\text{grad}T)/T^2 = -\mathbf{q}\cdot\text{grad}(-1/T) \tag{2.1.21}$$

Well, thermodynamicists like it that way, so they call $-1/T$ the potential, and never mind about Fourier, he didn't know about thermodynamics anyhow, since he was working in the prehistoric times of the early 19th century. So the discussion in Section 1.8, with the introduction of the rather awkward quantity $\alpha = -1/T$, was not so strange after all. This and other points will need to be discussed thoroughly.

Finally, we consider the application of Eq. (2.1.19) to another simple problem: any isothermal deformation (so that $\mathrm{D}T/\mathrm{D}t$ and grad$T$ are zero). The second law reduces to:

$$W/\Phi \leq \mathrm{D}A/\mathrm{D}t \tag{2.1.22}$$

i.e., in isothermal deformations the best we can hope is that all the work done results in accumulation of free energy — but generally some of it will be "lost", in the sense that A will not increase quite as much as we are doing work on the system.

2.2 HEAT TRANSFER

In this section, we are going to discuss the thermodynamics of heat conduction. The temperature gradient grad$T$ is going to play a major role, of course, so let's simplify notation by using τ for it. For the time being, we are going to neglect mechanical effects, except that perhaps we want to keep into account the fact that, as temperature changes, so will density — or will not, as the case may be. Since we are dealing with thermodynamics, we may as well use specific volume $V = 1/\Phi$ rather than density — after all, that's what thermodynamicists always do. We are going to forget about changes of shape, so that the only deformation of possible interest is a change of specific volume. Hence we rewrite the second law in the following form:

$$\mathrm{D}A/\mathrm{D}t + S\mathrm{D}T/\mathrm{D}t + p\mathrm{D}V/\mathrm{D}t + V\mathbf{q}\cdot\tau/T \leq 0 \tag{2.2.1}$$

Well, so far so good. Now we expect something like Fourier's law to be possibly of importance, say we expect the heat flux $\mathbf{q}$ to depend on the temperature gradient τ. Let's not take position yet as to how it may depend on it, but let's only consider the fact that, at equal τ, $\mathbf{q}$ may well depend on temperature itself, as well as on density (or volume, or pressure, or whatever the appropriate variable may be — though it will turn out that we can't be quite so cavalier about this). Let's tentatively say that:

$$\mathbf{q} = \mathbf{f}(V,T,\tau) \tag{2.2.2}$$

Now let's think about Eq. (2.2.2) a bit. It is of course a constitutive equation, though the presence of a so far unspecified function $\mathbf{f}()$ would rather tell us that it represents a class of possible constitutive equations. Can we interpret it to mean that $\mathbf{q}$ is a function of state, as thermodynamicists like to say? Well, if we think that thermodynamics is a science of equilibrium, this question is nonsense: at equilibrium both τ and $\mathbf{q}$ are zero, and the question does not arise. But under nonequilibrium conditions? We want to deal with heat conduction, after all, and allowing for finite values of τ is the very least we can do. We have the following possibility: let's regard (2.2.2) as the constitutive equation for $\mathbf{q}$, which we will call a function of state, at least tentatively (rereading Section 1.5 at this point is useful). $\mathbf{q}$ is a function of state which has the property of taking the value of zero at equilibrium — nothing intrinsically wrong with that; for instance, we know that free energy (under appropriate, and perhaps somewhat mysterious qualifying conditions) has a minimum at equilibrium, and hence its derivative (with respect to what is so far another mystery) is zero at equilibrium, so the idea of some function of state being zero at equilibrium is not so repugnant after all. Let's pursue this idea.

Remark 2.2.1

This will be discussed in the section on kinetics. The derivative is with respect to the internal state variable(s) w, and is called the affinity.

If we do admit that $\mathbf{q}$ is a function of state, and if we accept the idea that (2.2.2) is not entirely unreasonable, then we have to admit that the state is perhaps the ordered set V,T,τ. Temperature gradient being what determines the state? Well, what's wrong with that? Why is temperature itself OK, and its gradient should not? Let's go ahead; the state is V,T,τ. But then all "functions of state" are given by an equation such as (2.2.2), say:

$$A = a(V,T,\tau) \tag{2.2.3}$$

$$S = s(V,T,\tau) \tag{2.2.4}$$

$$p = f(V,T,\tau) \tag{2.2.5}$$

Remark 2.2.2

> What was meant by developing a field theory of thermodynamics is perhaps more clear at this point. V,T,τ is a *local* state: it is defined at each point, and may well be different from one point to another; indeed, it is guaranteed to be different at different points unless $\tau = 0$.

We are really allowing ourselves an orgy of unconventionality: not only are we allowing τ to be, in addition to the time-honoured variables V and T, in the list which determines the state; we are even using different symbols for a function and for its value (notice the lower case $a()$ in (2.2.3)). Well, let's see what comes out of all this. Let's just substitute (2.2.2–5) into (2.2.1) and see what happens:

$$\begin{aligned}&[\partial a/\partial V + p]\,\mathrm{D}V/\mathrm{D}t + [\partial a/\partial T + S]\,\mathrm{D}T/\mathrm{D}t \\ &+ [\partial a/\partial \tau]\cdot \mathrm{D}\tau/\mathrm{D}t + V[q\cdot\tau]/T \leq 0\end{aligned} \tag{2.2.6}$$

We have been using more brackets than needed, and the reason will soon become clear. If indeed (2.2.3–5) hold, all quantities in brackets depend on, and only on, V, T and τ. That is our constitutive assumption, and we have to be brave enough to stick by it — if we start changing our mind halfway through, we'll never get anywhere. But suppose we assign values of V, T and τ, so that the value of all terms in brackets is fixed. We are still free to assign $\mathrm{D}\tau/\mathrm{D}t$ any way we wish, aren't we? Large, small, pointing this way or that way. How can (2.2.6) still be satisfied no matter what $\mathrm{D}\tau/\mathrm{D}t$ we choose? Well, it's hard to admit it, but there is only one possibility: that $\partial a/\partial\tau = 0$. Not too bad a conclusion, after all; did we really expect free energy to depend on the temperature gradient? In fact, we foresee a possible generalization: as long as only gradients (of what? Let's leave the answer to that open for the time being) appear in our constitutive equations, perhaps free energy will behave and be independent of them, as we have always, more or less implicitly, assumed it to do. So now the unpleasant third term in Eq. (2.2.6) has disappeared, since we have concluded that:

$$\partial a/\partial\tau = 0;\ A = a(V,T) \tag{2.2.7}$$

Now let's proceed. V, T and τ are fixed (and so is the value of all brackets). Aren't we free to assign DV/Dt any way we wish? Well, yes, though people are starting to feel uncomfortable with the whole argument. Be that as it may, let's proceed, and conclude that the only way of satisfying (2.2.6) is. .. yes, our reader has guessed right. The only way is that:

$$p = -\partial a/\partial V = (-\partial A/\partial V)_{\mathrm{T}} \tag{2.2.8}$$

Remark 2.2.3

> This may appear as a very strange way of deducing consequences of the second law, but it is a powerful one. The second law is an inequality, and by standard algebra one can only derive other inequalities from it. However, algebra is an algorithm for shortcutting logical thinking, and if we conclude something by thinking we should trust it at least as much as we do algebra.

Now, particularly in the form with subscripts, this looks familiar — let's go back to our undergraduate thermodynamics textbook, and lo and behold, pressure is minus the partial derivative of free energy with respect to volume at constant temperature. Maybe this isn't as crazy a method of doing things as it appeared at first sight. Perhaps our reader is even beginning to see why subscripts appear in thermodynamic partial derivatives — but more about this later. For the time being, we note simply that (2.2.7) and (2.2.8) together, if taken seriously, mean that:

$$\partial p/\partial \tau = 0;\ p = f(V,T) \tag{2.2.9}$$

which also leaves us rather satisfied — we didn't expect pressure to depend on the temperature gradient, did we?

It's a breeze now to go one step further, and conclude that, since at fixed V, T and τ we may still choose DT/Dt any way we wish to conclude that:

$$S = -\partial a/\partial T = (-\partial A/\partial T)_{\mathrm{V}} \tag{2.2.10}$$

which our thermodynamics textbook tells us is right. Whether that fact should strengthen our belief in what is written in our thermodynamics textbook, or in the powerfulness of the method we are discussing here, or in both, or perhaps in neither — that's all open to question. But at least we haven't met any patently absurd conclusion so far, so let's proceed.

Remark 2.2.4

We have obtained (2.2.8) and (2.2.10) without ever invoking anything like δq_{REV}. Hopefully our readers will have some satisfaction with that.

Everything in (2.2.1), except the very last term, has dropped out, and we are left with:

$$\mathbf{q} \cdot \tau \leq 0 \tag{2.2.11}$$

i.e., heat does flow in the direction of decreasing temperature after all. This may look a bit like a magician pulling a rabbit out of his hat, but it is a real rabbit, and let's give the magician the benefit of the doubt.

The hand of our preferred student is now raised high in the back of the class (good students never sit in front). Let's introduce her to our reader. Her name is Sue, she has done well but not exceptionally well as an undergraduate. In fact she has just gone by because she's smart, and she has learned so little by heart that she's perfectly willing to unlearn it. She has got some basic concepts right, though: "In the section on relaxation, you just tried to convince me that, at least occasionally, heat may flow in the direction of increasing temperature. I didn't believe it, of course; but that's what you said. Now you come up all proud to have proved, by very devious means, that that can never happen. You can't eat your cake and have it too".

Good question, of course; Sue does have a point there. Let's discuss it. In the section on relaxation, we wrote something for the heat flux which, in a three-dimensional formulation, would look something like:

$$\mathbf{q} + \beta\partial\mathbf{q}/\partial t = -k\tau \tag{2.2.12}$$

There are lots and lots of subtleties about (2.2.12), but let's forget about them for the time being. The point is, (2.2.12) is in contrast with (2.2.2), isn't it? Equation (2.2.12) allows $\mathbf{q}$ to take values which, at fixed V and T, are not uniquely determined by τ, in contrast with (2.2.2). So we are not guilty of selfcontradiction after all: *if* $\mathbf{q}$ depends uniquely on τ at fixed V and T, *then* heat cannot flow in the direction of increasing temperature, not even occasionally or exceptionally. That's what we have shown, and there is no contradiction.

Sue likes this, but she's only mildly convinced. "OK, there is no contradiction; but can't you make up your mind? Can, or cannot heat flow in the direction of increasing temperature, if only exceptionally?" Our answer is that it depends on what the assumed form of the constitutive equation is — if we believe (2.2.2), it cannot; if we believe (2.2.12), it sometimes can. "This is absolutely not acceptable. Thermodynamics is supposed to tell us what can happen and what cannot, and the answer should better be independent of our constitutive assumptions".

Well, Sue is a good student, and she certainly asks the right questions, but she's an optimist — we all have been when young. Thermodynamics isn't that powerful. Thermodynamics doesn't tell us what can and what cannot happen *in general*, except for the highest possible efficiency of heat engines. It tells us what can happen *if* we make certain constitutive assumptions — it tells us something about constitutive equations. Even with those, it doesn't tell us what they are, but only what they cannot be. Thermodynamics doesn't tell us whether air at ambient conditions is or is not an ideal gas; it only tells us that, should it be an ideal gas, then the difference between its specific heat at constant pressure and at constant volume is going to be equal to the gas constant. Whether we are willing to regard air at ambient conditions as an ideal gas is something which thermodynamics leaves to our freedom of choice.

Sue is still not entirely satisfied with all this, but we are going to proceed anyhow. We'll assume that indeed (2.2.2) holds, and therefore (2.2.11) as well — not just most of the time, but invariably. We can now try to specify the **f**() function a bit more, and we'll make the usual assumption we always make when beginning to play with a constitutive equation —we'll assume it is *linear*. Let's just begin as easily as we can. We write:

$$\mathbf{q} = -k\tau \tag{2.2.13}$$

which is of course Fourier's law for an isotropic material as we know it, and k is the thermal conductivity. We substitute this in (2.2.11), and conclude that $k = 0$ — that's why we put the minus sign in (2.2.13) in the first place, isn't it? We dislike, as everybody does, to work with negative quantities. Well, this is not very exciting, we didn't need all this algebra just to conclude that the thermal conductivity is non-negative.

Still, we haven't met any inconsistency so far, so let's see what happens next. Why did we assume a linear form for **f**()? Well, we know that at equilibrium both τ and $\mathbf{q}$ are zero; so we expanded in a Taylor series, truncated at the first term — something like that. But, as a student in the front row now points out, that leads to an inconsistency. Let's introduce him as well. His name is Ralph; he did extremely well as an undergraduate, never misses a sign in an equation (and is unfortunately dedicated to pointing it out when we do), and knows by heart all the equations he has been taught as an undergraduate; moreover, he never questions the validity of any one of them. He is more than mildly annoyed by our tendency to question them. "Why did you expand in terms of the gradient of T rather than the gradient of $\alpha = -1/T$? That's what Onsager says we should do, after all". Well, does anybody have anything to say about that?

Sue, who is not only smart but lucky as well, missed the class on the Onsager reciprocal relations because she had something much better to do on that day, and so she has no preference for $-1/T$ over T; in fact she finds $-1/T$ rather funny. She has done some thinking in the meantime. "It doesn't matter. As long as we are expanding in a Taylor series, and truncating at the first term, grad$T$ and gradα are proportional to each other, and thus it doesn't matter". Sue would probably make a horrible mess with the algebra if she had to actually do a Taylor series expansion, but obviously she has understood what such an expansion is. What she has just stated is certainly true, but again there is some optimism involved. Let's see why. We can have as small a value of grad$T$ as we wish, and still, if we have a sufficiently large body, temperature differences may be as large as we wish, and therefore what are we going to choose, α or T? It goes back to the question of whether k in (2.2.13) is more nearly independent of temperature than the L which would appear in the thermodynamicists' equation is, or the other way around. We are dealing with transport phenomena, and we know that (2.2.13) is better than what the thermodynamicists prefer if we want to approximate k with a constant. Fourier preferred (2.2.13) — in fact he was lucky because thermodynamics hadn't been invented (or discovered, as the case may be) yet, and so he didn't have to contend with the silly suggestion that one should use α. Mercury thermometers go up with T, at least approximately, don't they?

Now let's go back to heat conduction in a stationary body, but let's allow conduction to be unsteady. Since there is no work term, the balance equation is:

$$\mathrm{div}\mathbf{q} + \Phi\partial E/\partial t = 0 \tag{2.2.14}$$

This guarantees that div$\mathbf{q}$ and $\partial E/\partial t$ cannot have the same sign, so that we can write:

$$0 = \mathrm{div}\mathbf{q}\partial E/\partial t = \mathrm{div}(\mathbf{q}\partial E/\partial t) - \mathbf{q}\cdot \mathrm{grad}(\partial E/\partial t) \qquad (2.2.15)$$

Now suppose that the BCs are steady, so that we are considering a temperature distribution which is slowly approaching some steady state value. We can integrate (2.2.15) over the whole volume of our system, apply Gauss' theorem, and notice that the integral of the first term vanishes because $\partial E/\partial t = 0$ on the boundaries. We are thus left with:

$$\int \mathbf{q}\cdot\nabla(\partial E/\partial t)\mathrm{d}V \geq 0 \qquad (2.2.16)$$

Now suppose we can find a quantity Ω such that:

$$\partial\Omega/\partial t = -\mathbf{q}\cdot \mathrm{grad}(\partial E/\partial t) \qquad (2.2.17)$$

In this case, (2.2.16) guarantees that the volume integral of Ω, $<\Omega>$, decreases as steady state conditions are approached, no matter what the initial conditions may be, and hence that $<\Omega>$ has a minimum at steady state: a variational principle has been established. It is perhaps obvious that the second law has not entered this derivation at all.

Remark 2.2.5

The second law has nothing to do with this result, but for thermodynamicists the temptation to mix up with it is overwhelming. Something is minimum at steady state? But free energy is minimum at equilibrium, so that something must somehow be related to free energy. The fact that $<\Omega>$ is minimum at steady state, not at equilibrium, doesn't deter them. Steady state, equilibrium, they are about the same — what's all the fuss about? See the paragraph following (2.2.19).

Now let's consider a simple case: $\mathbf{q}$ is given by Eq. (2.2.13) with k a constant, and $E = cT$, with the specific heat per unit volume c also a constant. Then:

$$\mathrm{grad}(\partial E/\partial t) = c\partial\tau/\partial t \qquad (2.2.18)$$

$$\Omega = kc\tau^2/2 \qquad (2.2.19)$$

where $\tau^2 = \tau \cdot \tau$. The result makes a lot of sense, as can easily be checked by looking at simple steady solutions such as heat conduction through a slab. Now however thermodynamicists would want us to believe that the quantity to be minimized is the volume integral of the rate of entropy production, $<\Gamma>$; but $<\Gamma>$ and $<\Omega>$ are not proportional to each other, are they? (In a body of sufficiently large size, $<\Gamma>$ and $<\Omega>$ are not proportional to each other even if τ^2 is very small indeed). It is of course possible to doctor the constitutive equations for **q** and E (by introducing the potential $-1/T$, among other things) in such a way that $<\Omega>$ comes out to be proportional to $<\Gamma>$, but what's the advantage of doing that?

Now let's consider another point. Equation (2.2.13) not only assumes that **q** depends linearly on τ, but in fact that it depends on it isotropically. In a solid, it could certainly be possible for the thermal conductivity to be different in different directions, and hence, if we want to retain linearity, we should write:

$$\mathbf{q} = -\mathbf{K} \cdot \tau \tag{2.2.20}$$

where **K** is the conductivity tensor. **K** can be decomposed into its symmetric part $\mathbf{K}^+$ and antisymmetric part $\mathbf{K}^-$:

$$\mathbf{K} = \mathbf{K}^+ + \mathbf{K}^- \tag{2.2.21}$$

Now the second law requirement is:

$$\tau \cdot \mathbf{K}^+ \cdot \tau + \tau \cdot \mathbf{K}^- \cdot \tau = 0 \tag{2.2.22}$$

The second term on the LHS is intrinsically zero, since $\mathbf{K}^-$ is skew (given a skew tensor **A**, and *any* vector **a**, the scalar $\mathbf{a} \cdot A \cdot a$ is invariably zero, as Ralph knows very well, though he dislikes the idea of vector algebra being used when one is discussing thermodynamics), and thus (2.2.22) requires the symmetric part $\mathbf{K}^+$ to be positive definite, and poses no restrictions whatsoever on the skew part $\mathbf{K}^-$. Indeed, the $\mathbf{K}^- \cdot \tau$ component of the heat flux vector is orthogonal to τ, so that as far as it is concerned heat flows along constant temperature lines, and why should that be forbidden by the second law? Yet thermodynamicists want us to believe that $\mathbf{K}^- = 0$ no matter what — that is indeed the fundamental Onsager reciprocity relation.

2.3 MOMENTUM TRANSFER

In this section, we approach the difficult problem of the thermodynamics of momentum transfer. Momentum is a vector, and its flux is a tensor: there is no way of avoiding geometrical complexities when dealing with momentum transfer. In order to simplify matters at least in part, in this section we limit attention to isothermal momentum transfer, so that temperature does not enter our list of relevant variables.

One of the major difficulties with the thermodynamics of momentum transfer is the following one. As has been discussed in Section 2.1, thermodynamics usually restricts attention to work done as a result of changes in volume. On the other side, in fluid mechanics the assumption of incompressibility is a common, and very powerful one. Of course, if one takes incompressibility seriously, there can be no change of volume, and hence also V disappears from our list of relevant variables. It would seem therefore that very little is left to say about thermodynamics, but that is not the case by far.

We do not make the assumption of incompressibility in this section. Hence the systems we consider, when flowing, may undergo two types of deformation: changes of density, and changes of shape. Now it is useful to keep these two types distinct, and this is easily accomplished as follows. Any tensor can be decomposed into the sum of an isotropic and a deviatoric tensor, say:

$$\mathbf{A} = (\mathrm{tr}\mathbf{A}/3)\mathbf{I} + \mathbf{A}^*; \ \mathrm{tr}\mathbf{A}^* = 0 \tag{2.3.1}$$

When this is applied to the rate of strain tensor $\mathbf{D}$, the isotropic part is recognized as the part which measures changes of density, see Eq. (2.1.13), and the deviatoric part $\mathbf{D}^*$ measures changes in shape, since it is zero in a pure compression or dilation:

$$\mathbf{D} = \Phi\mathbf{I}\mathrm{D}V/\mathrm{D}t + \mathbf{D}^*; \ \mathrm{tr}\mathbf{D}^* = 0 \tag{2.3.2}$$

The same decomposition can be applied to the stress tensor $\mathbf{T}$ to yield:

$$\mathbf{T} = p\mathbf{I} + \mathbf{T}^*; \ \mathrm{tr}\mathbf{T}^* = 0 \tag{2.3.3}$$

where p is as defined in Eq. (2.1.14). For reasons which will soon become clear, we call p as defined in (2.1.14) the *mechanical* pressure; it is the average of the readings of three pressure gauges mounted at right angles to each other.

With this much background, we can now express the net rate of work per unit mass W/Φ as:

$$W/\Phi = -p\mathrm{D}V/\mathrm{D}t - \mathbf{T}^*:\mathbf{D}^*/\Phi \qquad (2.3.4)$$

which has the advantage of containing a familiar first term on the RHS. Equation (2.3.4) tells us that, when $\mathbf{D}^*$ is not zero (as in general it will not be), the net rate of work contains a term in addition to the one usually encountered in thermodynamics.

Now let's begin with a very simple constitutive equation, the one on which the theory of the hydrodynamics of ideal fluids is based: *the stress in a flowing fluid is an isotropic pressure*. This is known to be true at equilibrium, and it is the simplest possible assumption one can make in order to develop a non-empty theory of fluid mechanics. The constitutive assumption is essentially the statement that $\mathbf{T}^* = 0$, and hence Eq. (2.3.4) yields the result $W/\Phi = -p\mathrm{D}V/\mathrm{D}t$. *In the hydrodynamics of ideal fluids*, the net rate of work is expressible as in classical elementary thermodynamics. In the hydrodynamics of incompressible ideal fluids, the net rate of work is simply zero.

Remark 2.3.1

> We are taking the position that $\mathbf{T}^* = 0$ is the constitutive equation of classical hydrodynamics. One could also take the more stringent position that the constitutive assumption is that both the shear and the bulk viscosity are zero. With the latter, there can never be energy dissipation in hydrodynamics.

Is there any thermodynamics left in isothermal hydrodynamics of incompressible ideal fluids? Well, very little indeed. Both V and T are constant, and every term on the RHS of Eq. (2.1.19) becomes zero — not very illuminating, except that the equation is always satisfied as an equality, so that there is no energy dissipation, no matter how the fluid may be flowing. The isothermal hydrodynamics of incompressible ideal fluids is a theory which describes only thermodynamically reversible phenomena: no energy is ever dissipated. This has important consequences, as the following two examples show.

First consider the case of an airfoil, say the wing of a plane flying subsonically on a horizontal course with velocity U. We can use hydrodynamics of ideal fluids: after all, the Reynolds number is extremely large, suction of air through the wing will avoid detachment of the boundary layer, and the plane

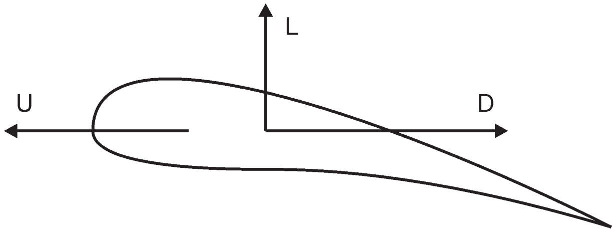

Fig. 2.3.1. Lift and drag on an airfoil.

flies subsonically, so that we may forget about the compressibility of air. So we go ahead and we calculate the force **F** exerted by air on the wing, and our plan is as follows. We'll decompose **F** into a vertical component which we'll call the lift L, and a horizontal component which we'll call the drag D (see Fig. 2.3.1).

L had better be large enough to sustain the weight of the plane, and we'll calculate therefore the value of U needed to keep the plane from falling down. We'll then have D, and we'll design the engines accordingly — it makes sense, doesn't it?

Well, we go through the calculation, and we get indeed a value of L and the corresponding value of U; if the wings are designed based on this calculation, the plane will indeed avoid falling to the ground. However, we calculate a D value of zero — the plane, as far as hydrodynamics is concerned, may well turn off the engines once it has reached a steady horizontal course. What's going on here? Well, it's easy: L is vertical, the plane flies horizontally, and hence L does no work; it is a "conservative" force. There is no dissipation connected with the existence of a finite L. But D, that's an entirely dissipative force: DU is in fact the total rate of energy dissipation in a steady horizontal course. Since hydrodynamics cannot describe dissipation, we should have known from the beginning that the calculation would yield $D = 0$.

Remark 2.3.2

In general, the drag exerted on any submerged object is calculated to be zero from hydrodynamics. This is what used to be called D'Alembert's paradox at the beginning of this century.

The second example is steady flow through a pipe, with an average velocity U. The no-slip boundary condition cannot be imposed on the equations of hydrodynamics, and hence one finds out by inspection that $\mathbf{v} = U\mathbf{k}$ (with $\mathbf{k}$ the unit vector in the axial direction), p = constant is a solution to the equations of hydrodynamics which satisfies all the boundary conditions. So we conclude that there is no pressure drop. Well, the pressure drop is a dissipative force, and we were silly to try to calculate it from the equations of hydrodynamics. Notice that in this case we do not even need the assumption of incompressibility. If we really believe that air is an ideal gas ($pV = RT/M$), and an ideal fluid ($\mathbf{T}^* = 0$), there is no pressure drop in the flow of air through a tube. Doesn't thermodynamics maintain that all transformations of an ideal gas are reversible? Well, maybe somebody is happy with that, but engineers who design air conditioning equipment would regard that conclusion as nonsense, much as they'd love to save the cost of circulation pumps.

But even hydrodynamics of ideal fluids can include some dissipation if one includes compressibility. The constitutive assumption is that $\mathbf{T}^* = 0$, but apart from that p is left open to be whatever will come out of the equations. So let's see what happens. At equilibrium, p is a function of V and T, and since we are restricting attention to isothermal conditions, p depends only on volume — and volume does not change in time at equilibrium. Now if we are including compressibility it must be because density may change over rather short time scales, so that we want to avoid any argument of the "quasistatic" variety. That makes Sue happy, since she doesn't really understand what quasistatic means; it makes however Ralph very unhappy, since he dislikes the idea of thermodynamics having to do with anything but quasistatic processes. Sue has heard the word "quasistatic" (many were the classes she missed, for a variety of commendable reasons, but unfortunately for her she didn't miss "quasistatic": it was discussed on a very rainy day), and she thinks that, if it means anything, it must mean something like "slow enough". Well, let's say that's what it means. Now if density changes very slowly we would forget about compressibility anyhow, and we don't want to do that. So all that we really know is that *if* $\mathrm{D}V/\mathrm{D}t = 0$ (or if it is in some sense small enough), then $p = f^*(V)$ — the reason for the asterisk will soon become clear. Now it seems legitimate to consider the possibility that p may be expressed as:

$$p = f(V, \mathrm{D}V/\mathrm{D}t);\ f^*(V) = f(V, 0) \tag{2.3.5}$$

Sue has two questions this time. First an easy one, why did we include $\mathrm{D}V/\mathrm{D}t$? That's really easy: so that we can see whether there is any difference between equilibrium conditions and conditions where density may perhaps change very rapidly. Now the real question: to pursue such an objective, why didn't we include **D**, rather than only DV/Dt? Well, because $\mathbf{D}^*$ isn't going to contribute anything to W, see Eq. (2.3.4), and, had we included $\mathbf{D}^*$, we would simply have made the algebra more cumbersome to reach the same conclusions. Try it out afterwards at home, will you, Sue? In the meantime Ralph is shaking his head with a very disconsolate expression — he doesn't believe any of this.

OK, let's proceed. Since our system is isothermal, the second and fourth term in (2.1.19) drop out, and we have only to worry about free energy A. So we write:

$$A = a(V, \mathrm{D}V/\mathrm{D}t) \tag{2.3.6}$$

We then perform the usual algebra to obtain:

$$(\partial a/\partial V + p)\,\mathrm{D}V/\mathrm{D}t + (\partial a/\partial(\mathrm{D}V/\mathrm{D}T))\mathrm{D}^2V/\mathrm{D}t^2 \leq 0 \tag{2.3.7}$$

Now if we assign V and DV/Dt we can still choose $\mathrm{D}^2V/\mathrm{D}t^2$ any way we wish, can't we? After all, we are assuming compressibility, and let's go all the way with that. So we conclude that — does anybody have the answer? Yes, Sue, you're right — we conclude that:

$$\partial a/\partial(\mathrm{D}V/\mathrm{D}t) = 0 \tag{2.3.8}$$

No, Ralph, that does *not* mean that whatever we add to V and T in the list of variables on which A depends, it will inevitably drop out. It does this time, but not inevitably. Now we are left with:

$$(\mathrm{d}a/\mathrm{d}V + p)\mathrm{D}V/\mathrm{D}t \leq 0 \tag{2.3.9}$$

What can we say about that? No, Ralph, we cannot conclude that $p = -\mathrm{d}a/\mathrm{d}V$: p does depend on DV/Dt, that's our hypothesis, and so we can't keep the value of the bracket fixed and still assign DV/Dt arbitrarily. "But you just showed that A does not depend on DV/Dt — you even wrote the derivative as a total one". Yes, A does not depend on DV/Dt; but we don't know about p yet. All we can conclude from (2.3.9) is that:

$$\text{if } \mathrm{D}V/\mathrm{D}t > 0,\ p \leq -\mathrm{d}a/\mathrm{d}V \tag{2.3.10}$$

$$\text{if } \mathrm{D}V/\mathrm{D}t < 0,\ p \geq -\mathrm{d}a/\mathrm{d}V \tag{2.3.11}$$

which, under some assumption of smoothness which is likely to be a mild one, implies that when $\mathrm{D}V/\mathrm{D}t = 0$, $p = -\mathrm{d}a/\mathrm{d}V$. So what do we conclude? That *at equilibrium* ($\mathrm{D}V/\mathrm{D}t = 0$) we will have:

$$p = p^* = f^*(V) = -\mathrm{d}a/\mathrm{d}V = -(\partial a/\partial V)_\mathrm{T} \tag{2.3.12}$$

which makes us very happy, since that's what our thermodynamics textbook tells us. So we'll call p^* the *thermodynamic* pressure. But under non-equilibrium conditions, the mechanical pressure will be less than the thermodynamic pressure in expansions, and more than it in compressions. This of course now describes irreversibility — the work we get back when the gas expands is less than the work we have done to compress it.

But let's move to something more interesting. Let's say we still consider fluids, so that at equilibrium $\mathbf{T} = p^*\mathbf{I}$, but under nonequilibrium conditions we want at the very least to include ordinary Newtonian fluids, so that $\mathbf{T}^*$ needs not be zero. What are we going to write down as constitutive equations? If we keep compressibility in the picture, we have to write something like:

$$\mathbf{T} = \mathbf{F}(V, \mathbf{D}) = p(V,\mathbf{D})\mathbf{I} + \mathbf{F}^*(V,\mathbf{D}) \tag{2.3.13}$$

$$A = a(V, \mathbf{D}) \tag{2.3.14}$$

where $\mathbf{T}^* = \mathbf{F}^*(V,\mathbf{D})$. We do the usual algebra to obtain (the second and fourth term on the RHS of (2.1.19) are still zero):

$$(\partial a/\partial V + p)\mathrm{D}V/\mathrm{D}t + \mathbf{T}^*\mathbf{:}\mathbf{D}^*/\Phi + (\partial a/\partial \mathbf{D}):\mathrm{D}\mathbf{D}/\mathrm{D}t \leq 0 \tag{2.3.15}$$

We are experts by now, and tensors don't scare us in the least, so we see by inspection that the last term drops out; free energy again depends only on volume. No, Ralph, it still isn't quite inevitable; it happens frequently, but not always. We are left with:

$$(\mathrm{d}a/\mathrm{d}V + p)\mathrm{D}V/\mathrm{D}t + \mathbf{T}^*:\mathbf{D}^*/\Phi \leq 0 \tag{2.3.16}$$

Now in a situation of pure compression or expansion ($\mathbf{D}^* = 0$), we obtain the same conclusions as before, see Eqs. (2.3.10–12). For an incompressible fluid, or for a compressible one which undergoes an isochoric motion ($\mathrm{D}V/\mathrm{D}t = 0$), we obtain:

$$\mathbf{T}^*\!:\!\mathbf{D}^* \leq 0 \tag{2.3.17}$$

Let's try this all out for the simple case of a Newtonian fluid, i.e., one for which the function **F**() is linear in **D**. This can be written as:

$$\mathbf{T} = p^*\mathbf{I} - 2\mu\mathbf{D} - \mu'(\mathrm{tr}\mathbf{D})\mathbf{I} \tag{2.3.18}$$

where μ is the ordinary shear viscosity (the only one appearing in the case of incompressible fluids), and μ' is the bulk viscosity. The first term on the RHS guarantees that we'll recover the correct result at equilibrium ($\mathbf{D} = 0$). Thus the mechanical pressure is:

$$p = p^* - (2\mu + 3\mu')\Phi \mathrm{D}V/\mathrm{D}t \tag{2.3.19}$$

Since in pure compressions or expansions ($\mathbf{D}^* = 0$) (2.3.10–12) must hold, we conclude that:

$$2\mu + 3\mu' = 0 \tag{2.3.20}$$

On the other side, for an isochoric motion $\mathbf{T}^* = -2\mu\mathbf{D}$, and thus we conclude that:

$$\mu = 0 \tag{2.3.21}$$

This time Ralph is really mad. "All this algebra, which I couldn't follow anyhow, since you never bother to spell it out in detail. Deviatoric tensors, bulk viscosities, all these paraphernalia — and free energy still depends only on volume, and the viscosity cannot be negative — I knew all that from the beginning". Well, Ralph, would you have preferred all this algebra to lead to the conclusion that viscosity is negative? And you knew of course about (2.3.21), though God knows why you believed it; but what about (2.3.20)? Is it so obvious that the bulk viscosity could be negative, provided its absolute value is less than 2/3 of the shear viscosity?

Indeed, Stokes laid down as a hypothesis that $2\mu + 3\mu' = 0$, and if that is true pure expansions and compressions are nondissipative. The Stokes hypothesis comes out to be right for the theory of Maxwellian gases — which means that the assumption about the momentum flux in Section 1.5 is self-consistent. But the Maxwellian theory does not predict μ to be zero, in fact the value of μ can be calculated out of it (under certain simplifying assumptions). Thus the Maxwellian theory of ideal gases predicts a negative bulk viscosity — is that obvious as well?

But let's move to a slightly more complicated case, that of an incompressible non-Newtonian fluid of the Reiner–Rivlin variety, for which **T** is given by:

$$\mathbf{T} = \alpha\mathbf{I} - 2\mu\mathbf{D} - \mu''\mathbf{D}^2 \tag{2.3.22}$$

where α is an undetermined scalar, and μ and μ'' are the viscosity coefficients — which may depend on the invariants of **D**.

Remark 2.3.4

> If the fluid is incompressible, the constitutive equation determines the stress to within an arbitrary additive isotropic tensor. This is the analog of the case of rigid bodies, for which the constitutive equation leaves the stress tensor entirely arbitrary. An incompressible fluid is rigid to changes of density, and hence that part of the stress which does work in such changes is undetermined by the constitutive equation. This is worded in the literature on continuum mechanics as the principle of internal constraints.

Now if we want to make things simple, we are tempted to assume that μ and μ'' are constant. Well, if we do that, and if we perform the algebra correctly, we come to the conclusion that the only way of satisfying (2.3.17) is to have $\mu'' = 0$ — in which case we get the incompressible Newtonian fluid again. Equation (2.3.22) is legitimate with μ'' different from zero only if μ and μ'' are not constant, in which case they need to satisfy an inequality which includes the invariants of **D**. That is by no means an obvious result, as witnessed by the fact that many papers have been published where solutions of boundary value problems are obtained for a constitutive equations such as (2.3.22) with μ and μ'' both nonzero, and both constant. It would be nice to have such a fluid available — it would contradict the second law, and all our energy conservation problems would be solved automatically. Those papers would teach us how we could get lots and lots of free lunches.

So far we have considered ideal fluids, Newtonian fluids, and a somewhat restricted class of non-Newtonian fluids. Let's jump to the other extreme, let's consider solids. We'll be a bit cavalier about the algebra, but the results are essentially correct. Let's assume that we know what we mean when we speak of an "undistorted" configuration of a solid as that configuration which corresponds to zero internal stresses (it's undistorted, and in a vacuum). Every material point in the body considered occupies a position **Y** in the undistorted configuration; **Y** may be used as a label for the material point considered. In actual fact, the solid may be in some other configuration, which may change in time; thus the position **X** of any material point **Y** at time *t* is given by:

$$\mathbf{X} = \Omega(\mathbf{Y},t) \tag{2.3.23}$$

The deformation gradient tensor **F** is defined as:

$$\mathbf{F} = \partial\Omega/\partial\mathbf{Y} \tag{2.3.24}$$

and it could well be nonunity even if the body considered has simply been translated and rotated from its undisturbed configuration to some other one by a rigid body motion. Hence **F** is not an appropriate measure of the deformation of the body at time *t*. However, one can define a strain tensor **G** as follows:

$$\mathbf{G} = \mathbf{F}^{\mathrm{T}} \cdot \mathbf{F} - \mathbf{I} \tag{2.3.25}$$

which has the desired properties. In fact, one needs little algebra to conclude that:

$$\mathrm{d}\mathbf{Y}^1 \cdot (\mathbf{G} + \mathbf{I}) \cdot \mathrm{d}\mathbf{Y}^2 = \mathrm{d}\mathbf{X}^1 \cdot \mathrm{d}\mathbf{X}^2 \tag{2.3.26}$$

which guarantees that, when $\mathbf{G} = 0$, the transformation from **Y** to **X** preserves distances (as can be seen by considering the case where $\mathrm{d}\mathbf{X}^1 = \mathrm{d}\mathbf{X}^2$) and angles, and hence it is a rigid body motion. Thus **G** indeed measures the strain.

Now velocity **v** is $\partial\Omega/\partial t$, and hence:

$$\mathrm{grad}\mathbf{v} = \partial\mathbf{v}/\partial\mathbf{X} = (\partial\mathbf{v}/\partial\mathbf{Y}) \cdot (\partial\mathbf{Y}/\partial\mathbf{X}) = (\mathrm{D}\mathbf{F}/\mathrm{D}t) \cdot \mathbf{F}^{-1} \tag{2.3.27}$$

and thus one obtains after some algebra:

$$\mathrm{D}\mathbf{G}/\mathrm{D}t = 2\mathbf{D} \tag{2.3.28}$$

Now under isothermal conditions one would write, for an elastic solid in which the stress depends only on strain:

$$\mathbf{T} = \mathbf{T}(\mathbf{G});\ \mathbf{T}(0) = 0 \tag{2.3.29}$$

$$A = \tfrac{1}{2}a(\mathbf{G}) \tag{2.3.30}$$

where the factor ½ has been put in for reasons which will soon become clear. One now goes through the same algebra concerning Eq. (2.1.19) to obtain:

$$(\partial a/\partial \mathbf{G} + \mathbf{T}):\mathbf{D} \leq 0 \tag{2.3.31}$$

But the term in brackets does not depend on **D**, and hence:

$$\mathbf{T} = -\,\partial a/\partial \mathbf{G} \tag{2.3.32}$$

Equation (2.3.32) has two very important implications. First, stress in an elastic solid can be expressed as the derivative of a scalar with respect to strain — it admits a potential (which is in fact free energy). This is the basis of a lot of theorems and methods of stress analysis in the classical theory of elasticity, where A is called the strain energy function. Second, it also guarantees that in the theory of elasticity there is no dissipation — it is, in this respect, analogous to the hydrodynamic theory. Of course dissipation does occur in real solids, just as it does in real fluids; but that only means that the classical theory of elasticity is not good enough to describe it in solids, just as hydrodynamics isn't good enough to describe it in fluids.

It is now useful to recap the results we have obtained. We have restricted attention to isothermal momentum transfer, so that temperature does not enter any of the equations, and the second law reduces to:

$$W/\Phi = \mathrm{D}A/\mathrm{D}t \tag{2.3.33}$$

Those processes for which $\mathrm{W}/\Phi = \mathrm{D}A/\mathrm{D}t$ are reversible; the rate of energy dissipation is $\mathrm{W}/\Phi - \mathrm{D}A/\mathrm{D}t$.

As far as free energy is concerned, we have concluded that it depends only on V for all the fluid materials which have been considered, and only on **G** for elastic solids. (It may be worthwhile to point out that the value of **G** determines also the value of V). For incompressible fluids of the type we have considered, $\mathrm{D}A/\mathrm{D}t$ is simply zero, while for compressible fluids $\mathrm{D}A/\mathrm{D}t = -p^*\Phi \mathrm{D}V/\mathrm{D}t$.

The term W/Φ is best expressed as in Eq. (2.3.4), say as the sum of a "compressive" part W_C/Φ and a "distortive" part W_D/Φ:

$$W_C/\Phi = -p\mathrm{D}V/\mathrm{D}t \tag{2.3.34}$$

$$W_D/\Phi = -\mathbf{T}^*{:}\mathbf{D}^*/\Phi \tag{2.3.35}$$

with of course the compressive part being zero in all incompressible materials.

The results we have obtained are as follows.

1. *Rigid bodies*. $W = 0$, and all motions are thermodynamically degenerate.
2. *Incompressible ideal fluid*. $W = 0$, and all motions are reversible, with both sides of (2.3.33) being zero.
3. *Compressible ideal fluids*. There is some room for choice in the definition. $\mathbf{T}^* = 0$, and hence $W_D = 0$ — there is no doubt about that; as a limiting case of Newtonian fluids, that corresponds to $\mu = 0$. It is possible to complete the definition by requiring μ' to be zero as well; in that case $p = p^*$, $W_C/\Phi = \mathrm{D}A/\mathrm{D}t$, and all processes are reversible. If this more restrictive definition is relaxed by allowing μ' to be nonzero (and hence, in view of (2.3.20), $\mu' > 0$), one gets $W_C/\Phi = \mathrm{D}A/\mathrm{D}t + 3\mu'(\mathrm{D}V/\mathrm{D}t)^2$, i.e., part of the compressive rate of work is dissipative. Notice that up to now the distortive rate of work has always turned out to be zero.
4. *Incompressible Newtonian fluids*. The compressive rate of work is zero, and so is $\mathrm{D}A/\mathrm{D}t$. The distortive rate of work is $W_D = 2\mu\mathbf{D}{:}\mathbf{D}$, which is intrinsically positive and entirely dissipative.
5. *Compressible Newtonian fluids*. The distortive rate of work is $2\mu\mathbf{D}^*{:}\mathbf{D}^*$, and is again intrinsically positive and entirely dissipative. The compressive rate of work is $W_C/\Phi = \mathrm{D}A/\mathrm{D}t + (2\mu + 3\mu')(\mathrm{D}V/\mathrm{D}t)^2$. If the Stokes hypothesis ($2\mu + 3\mu' = 0$) holds, the compressive work is reversible; if it does not hold, the dissipative part is $(2\mu + 3\mu')(\mathrm{D}V/\mathrm{D}t)^2$ which is intrinsically positive.
6. *Non-Newtonian viscous fluids* (the constitutive equation for the stress being Eq. (2.3.13)). The distortive rate of work is entirely dissipative, the compressive rate of work may be larger than $\mathrm{D}A/\mathrm{D}t$, and hence it may contain a dissipative term. Notice that for all fluids which have been considered the distortive rate of work is either zero or entirely dissipative.
7. *Elastic solids*. $W/\Phi = \mathrm{D}A/\mathrm{D}t$: all processes are reversible, and both the compressive and the distortive rate of work are nondissipative.

There are two important points to be considered at this stage. The first one is as follows. Equation (2.3.33) allows us to interpret A as stored elastic energy in isothermal deformations; this is indeed the usual interpretation in the theory of elasticity. Now consider the special case of an ideal gas, for which internal energy E is a unique function of temperature, and is thus a constant in isothermal processes. If one assumes that all processes are reversible, the compressive rate of work results in a finite $\mathrm{D}A/\mathrm{D}t$, say in accumulation of elastic energy. Now $A = E - TS$, and the only way for A to increase (in compression) is through a decrease of entropy (indeed, for ideal gases S is proportional to $\ln V$, and hence an isothermal compression results in a decrease of entropy). There is a more general conclusion to be drawn: elastic energy can in general be accumulated both by an increase of internal energy and by a decrease of entropy, and the reversible part of the net rate of work may result in an entropy decrease.

The other important point is the following one. Our list of materials (1 to 7) clearly does not exhaust all possibilities. It is conceivable that a material may exist for which the distortive rate of work is not zero, not entirely dissipative, and not entirely reversible, and this case does not arise for any one of the materials considered. Indeed, consider again the oscillatory experiment discussed in Section 1.10. The rate of work per unit area of the bounding planes is $-Fv$, and the rate of work per unit volume is thus $-Fv/H$; it is clearly a distortive rate of work, since volume is constant in time. From Eqs. (1.10.3) and (1.10.6) one gets:

$$\mathrm{W_D} = \mu(A\Omega/\mathrm{H})^2 \cos(\Omega t)\cos(\Omega t + \phi) \tag{2.3.36}$$

(the constant k in Section 1.10 is of course the shear viscosity μ). With $\phi = 0$ (Newtonian fluid), W_D is always positive, and since $\mathrm{D}A/\mathrm{D}t = 0$, W_D is entirely dissipative as expected. With ϕ different from zero, we know the fluid is not a Newtonian one. But we also see that W_D is negative for a portion of the cycle, which forces us to admit that free energy does not depend only on volume and is therefore not constant: if it were constant, during that portion of the cycle where W_D is negative, (2.3.34) would be violated. This means that the fluid considered is not only non-Newtonian, but of a type which cannot be described by Eq. (2.3.13): the internal stress *is not*, at constant V, a unique function of $\mathbf{D}^*$. Indeed, Eq. (2.3.13) would predict that, no matter what the $\mathbf{F}^*()$ function, the stress in the oscillatory experiment should always be in phase with the velocity, say $\phi = 0$.

During that part of the cycle where W_D is negative, accumulated elastic energy is being released ($\mathrm{D}A/\mathrm{D}t < 0$), since this is the only way that (2.3.34) can be satisfied; this in turn implies that elastic energy has been accumulated during the other part of the cycle, and hence that W_D is not entirely dissipative. But if we calculate the integral of W_D over a cycle and divide it by the duration of the cycle (so that we obtain the average net rate of work) we obtain $\mu(A/H)^2\cos\phi$. With $0 < \phi < \pi/2$, as observed experimentally, this is always positive and hence W_D is on the average dissipative. Thus one concludes that indeed W_D is partly reversible and partly dissipative.

This leads to the consideration of viscoelasticity, which will be briefly discussed in Chapter 4; we observe here that viscoelasticity is intrinsically connected with the concept of relaxation, which will be discussed in Section 2.6.

2.4 DIFFUSION

The thermodynamics of diffusion is significantly more complex than that of momentum or heat transfer. One can immediately see one of the reasons: work and heat enter directly into the formulation of the second law, and hence it is not surprising that one can rather directly deduce restrictions that the second law imposes on the constitutive equations for heat and momentum transfer. However, there is no term in Eq. (2.1.19) which is obviously related to diffusion, and therefore deduction of the restrictions imposed by the second law is expected to be significantly more roundabout.

Another important point is the following one. In the previous sections, we have avoided the issue of coupling between different transport phenomena: in Section 2.2, we have often restricted attention to stationary bodies, so that we only had to worry about heat conduction; in Section 2.3, we have restricted attention to isothermal systems, so that we only had to worry about momentum transfer. Now how can one do something of the same kind with diffusion? One would like to be able to consider a stationary body — but what do we mean by that? Consider a very simple example (see Fig. 2.4.1). A long cylinder sealed at both ends contains a mixture of two components which, to simplify matters, we'll assume behaves as an ideal gas. Temperature is uniform, and let's say that pressure is uniform as well (though we'll see that that leads to a contradiction unless one neglects the viscosity of the gas). Diffusion is taking place, so that the composition of the gas is not uniform along the axial direction X. The question which we ask ourselves is the following deceivingly simple one: is the system stationary or not?

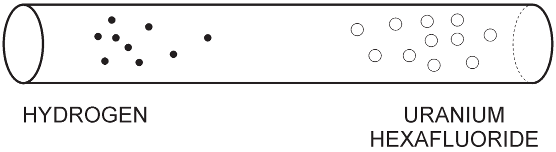

Fig. 2.4.1. Equimolar counterdiffusion sketch.

Ralph and Bob both think it is. Bob's reasoning is that the cylinder is stationary and sealed at both ends, so the system is stationary; no engineer in the real world would even consider the question as worth being asked. Ralph has a more sophisticated argument, which runs as follows. In an ideal gas at constant temperature and pressure, the total molar density c (total number of moles per unit volume) is independent of composition. It follows that the molar fluxes of the two gases are equal in magnitude and opposite in sign, and the total molar flux is thus zero. QED. It is an obvious case of equimolar counterdiffusion, which has been worked out extensively in his undergraduate class, and Ralph is sure that this time he must be right. Well, let's make it a bit more concrete: the two components are hydrogen (molecular weight 2) and uranium hexafluoride (molecular weight 289). Is that information in any way relevant?

Sue has something to say. The total molar flux is zero, she trusts Ralph's algebra about this. But if uranium hexafluoride is diffusing leftwards, and hydrogen rightwards, there is a significant mass flux leftwards. She does not confess the fact that she'd get all mixed up if she tried to actually calculate how large it is, but she's sure it's there. So perhaps the system isn't stationary after all. Any objections? Ralph finds this rather amusing. He has been taught that one can define both a mass average velocity $\mathbf{v}$, and a mole average one $\mathbf{v}^*$, and they are different from each other. It's all a question of definitions. But $\mathbf{v}^* = 0$ for the case considered, in mass transfer one always works with moles and not with grams, and so the system is stationary.

Well, at the very least we have established that there is room for discussion. Since everybody agrees that the mass flux is not zero, will pressure be constant along the axial direction? Don't we usually think that when there is a nonzero mass flux in a pipe, there is a nonzero pressure drop as well? When we consider mass transfer in a flowing fluid, we usually assume that the motion of the system is regulated by the Navier–Stokes equations (this is the equivalent of the "isothermal" heat transfer approximation). Well, the Navier–Stokes equations contain a density Φ, and a velocity $\mathbf{v}$. The density is the mass density, isn't it? And what is the velocity — **v** or $\mathbf{v}^*$? It does make a difference; if it is $\mathbf{v}^*$, all terms in the Navier–Stokes equations become zero for our example; if it is **v**, they do not. It may all be a matter of definitions, but it seems to make a difference. And $\Phi \mathrm{D}\mathbf{v}/\mathrm{D}t$ *is not* equal to $c\mathrm{D}\mathbf{v}^*/\mathrm{D}t$, so we have to make our choice. Well, as far as the Navier–Stokes equations are concerned, we need to choose Φ and **v**, that seems natural. As far as diffusion is concerned, we may perhaps buy Ralph's argument that it doesn't make any difference, so why don't we choose the mass based formalism anyhow?

Ralph has something to say this time. We want to deal with diffusion, so chemical potentials are going to play some role, in spite of our derogatory remarks about Onsager. The chemical potentials are the partial *molar* Gibbs free energies (Ralph hates our choice of "free enthalpy" for the Gibbs free energy), and hence we've got to work with molar units. How are we going to get out of this one?

Well, let's give it a try. Consider any quantity L which is an absolutely additive function of mass — the sort of thing which thermodynamicists call an extensive quantity. Something like internal energy E, entropy S, free energy $A = U - TS$, and yes, enthalpy $H = E + pV$ and free enthalpy $G = H - TS$ (though the latter two present some problem unless we know what we mean by pressure — is it the average of the readings of three mutually orthogonal pressure gauges, p, or the thermodynamic pressure p^*?). Well, let's forget about that problem, let's assume that p and p^* are the same thing — after all they sometimes are. Let such a quantity be indicated with L. What do we mean by saying that L is extensive? We mean that, given two systems in exactly the same conditions, one of which has a mass α times that of the other one, L for the latter will be α times what it is for the other one. Now let's express this concept mathematically for a mixture at constant p and T. One of our mixtures has m_K grams of component $K = 1, 2, \dots N$, and thus L is given by an equation of the following form:

$$L = l(\mathbf{m}) \tag{2.4.1}$$

where m is an N-dimensional vector with components m_K. The other mixture has αm_K grams of component K — it has exactly the same composition, only there is a different amount of it. Now the fact that L is extensive means that:

$$l(\alpha\mathbf{m}) = \alpha l(\mathbf{m}) \tag{2.4.2}$$

Everybody agrees so far? Well, let's proceed. We rewrite (2.4.2) as follows:

$$(\alpha - 1)l(\mathbf{m}) = l(\alpha\mathbf{m}) - l(\mathbf{m}) \tag{2.4.3}$$

and we now go to our book of calculus I to write:

$$(\alpha - 1)l(\mathbf{m}) = [(\alpha - 1)\partial l/\partial\mathbf{m}] \cdot \mathbf{m} + O((\alpha - 1)^2) \tag{2.4.4}$$

We now divide throughout by $(\alpha - 1)$, take the limit as α approaches unity, define $\mathbf{L}^*$ as $\partial l/\partial\mathbf{m}$, and we obtain:

$$l(\mathbf{m}) = \mathbf{L}^* \cdot \mathbf{m} = \Sigma_K L^*_K m_K \tag{2.4.5}$$

Now what are the L_K*'s? Well, we have a mathematical definition for them, and so all we need is a name — we'll call them the partial mass values of L. With free enthalpy, G_K^* will be the partial mass free enthalpy. Ralph is unhappy, but we proceed anyhow. In fact we consider the case where the mixture has a slightly different composition, say $\mathbf{m}' = \mathbf{m} + \mathrm{d}\mathbf{m}$; correspondingly, $L' = L + \mathrm{d}L$, and $\mathbf{L}^{*\prime} = \mathbf{L}^* + \mathrm{d}\mathbf{L}^*$. From (2.4.1), we calculate:

$$\mathrm{d}L = \mathbf{L}^* \cdot \mathrm{d}\mathbf{m} \tag{2.4.6}$$

while from (2.4.5) we calculate:

$$\mathrm{d}L = \mathbf{L}^* \cdot \mathrm{d}\mathbf{m} + \mathbf{m} \cdot \mathrm{d}\mathbf{L}^* \tag{2.4.7}$$

and by comparison:

$$\mathbf{m} \cdot \mathrm{d}\mathbf{L}^* = 0 \tag{2.4.8}$$

which we'll call, for lack of a better name, the Gibbs–Duhem equation.

Ralph knows there's something fishy in all this. "You can't call it the Gibbs–Duhem equation. The Gibbs–Duhem equation is a thermodynamic result, and your whole derivation has nothing to do with thermodynamics". You are perfectly right, Ralph; it has nothing to do with thermodynamics. It is a simple consequence of the fact that L is an absolutely additive function of mass; the derivation would also apply to such things as the electrical capacity of a mixture. And yet, had we done it with the number of moles vector **n** instead of the number of grams vector **m**, we would have obtained what is usually considered to be the Gibbs–Duhem equation. So perhaps we have to make up our mind: either the derivation above is such that we can legitimately call (2.4.8) the Gibbs–Duhem equation, or the classical Gibbs–Duhem equation has nothing to do with thermodynamics, or perhaps both. Whatever the case may be, we may ask ourselves why we did all this.

Our purpose was to have a definition of chemical potential in terms of mass units, and we have got that: the mass based chemical potential μ is the partial mass free enthalpy. The molar based chemical potential μ_M is the partial molar free enthalpy, and the two are related to each other by:

$$\mu_\mathrm{M} = M\mu \tag{2.4.9}$$

where M is the molecular weight of the component considered.

Now we can finally move to thermodynamics. We wish to write the first and second laws for a system which is the neighbourhood of a material point, *where the material point is regarded as moving with the mass average velocity* $\mathbf{v}$. Correspondingly, the intrinsic substantial derivative $\mathrm{D}/\mathrm{D}t$ is $\partial/\partial t + \mathbf{v}\cdot\mathrm{grad}$. However, such a neighbourhood still exchanges matter with its surroundings: there is no net mass flow into (or out of) it, but there may well be net mass flows of individual components. Let $\mathbf{Q}_K$ be the mass flow of component K with respect to the material point considered; the $\mathbf{Q}_K$'s sum up to zero, but they are individually nonzero. Correspondingly, should one wish to do so, one can consider molar fluxes $\mathbf{N}_K = \mathbf{Q}_K/M_K$, but these do not sum up to zero. Let $\mathbf{u}_K$ be the diffusive velocity of component K, say $\mathbf{Q}_K = \Phi_K\mathbf{u}_K$, where Φ_K is the mass concentration of component K. We write the first law as follows (all terms being per unit volume):

$$\Phi D(E + K)/\mathrm{D}t = -\,\mathrm{div}\mathbf{q} + W + \Phi Q - \mathrm{div}(\Sigma_K \mathbf{Q}_K(E_K{}^* + \tfrac{1}{2}u_K{}^2)) \tag{2.4.10}$$

In Eq. (2.4.10), W is the net rate of work per unit volume, to be calculated below, and the last term represents the net convective influx of internal and kinetic energy, which is nonzero in spite of the fact that the net influx of total mass is zero. We now need to calculate the work rate W, and this requires some care.

First consider the work done by gravity forces. We could go through something like the derivation of Eq. (2.1.8) again, but we choose not to do that, for reasons which will soon become clear. Rather, we keep the $\mathrm{D}K/\mathrm{D}t$ term in our equations, instead of cancelling it out with the work of gravity forces, and we keep those ones in as well. In fact life is made easier by using the concept of potential energy with h height over some reference horizontal plane, since we recognize that $\mathbf{Q}_K gh$ is the convective flux of potential energy carried by component K. Thus the net convective influx of potential energy (or work rate done by gravity forces) per unit volume is:

$$W_\mathrm{G} = -\mathrm{div}(\Sigma_K \mathbf{Q}_K gh) \tag{2.4.11}$$

As for contact forces, we'll restrict attention to only pressure (coherently with our having considered H and G legitimate quantities). If ϕ_K is the volume fraction of component K, it is also the area fraction over which it acts with a pressure p, and hence the rate of work done by component K per unit area is $p\phi_K \mathbf{u}_K = p\phi_K \mathbf{Q}_K/\Phi_K$. But ϕ_K/Φ_K is the partial mass volume of component K, V^*_K, and hence we obtain for the net rate of work of contact forces:

$$W_\mathrm{C} = -\mathrm{div}(\Sigma_K \mathbf{Q}_K pV^*_K) \tag{2.4.12}$$

We now substitute (2.4.11–12) into (2.4.10), observe that $H^*_K = E^*_K + pV^*_K$, and obtain:

$$\Phi D(E + K)/\mathrm{D}t = -\mathrm{div}\mathbf{q} + \Phi Q - \mathrm{div}[\Sigma_K \mathbf{Q}_K(H^*_K + gh + \tfrac{1}{2}u_K^2)] \tag{2.4.13}$$

Now the second law. We are going to write again:

$$\Phi \mathrm{D}S/\mathrm{D}t = j \tag{2.4.14}$$

where j is the net influx of entropy. However, we need to recognize that there is a convective contribution to j, and hence we write:

$$j = -\mathrm{div}(\mathbf{q}/T) + \Phi Q/T - \mathrm{div}(\Sigma_K \mathbf{Q}_K S^*_K) \tag{2.4.15}$$

To simplify matters, let's assume temperature is constant. With this, elimination of $\mathbf{q}$ and Q between the first and second laws is easy, and we obtain after some algebra:

$$\Phi D(A + K)/\mathrm{D}t + \mathrm{div}(\Sigma_K \mathbf{Q}_K \phi^*_K) \leq 0 \tag{2.4.16}$$

where:

$$\phi^*_K = \mu_K + gh + \tfrac{1}{2}u_K^2 \tag{2.4.17}$$

We can also expand the divergence in Eq. (2.4.16), and use the balance of component K to obtain:

$$\Phi D(A+K)/\mathrm{D}t + \Sigma_K \phi^*_K \mathrm{D}\Phi_\mathrm{K}/\mathrm{D}t + \Sigma_K \mathbf{Q}_K \cdot \mathrm{grad}\phi^*_K \leq 0 \tag{2.4.18}$$

All hands in the class are raised at this point — everybody wants to know what the ϕ^*_K's are. A mass based chemical potential μ_K, that was hard enough to swallow; but ϕ^*_K —that's even harder. Well, ϕ^*_K is defined by (2.4.17), so what we really need is to interpret what it means. Well, let's first of all consider what happens at (Lagrangian) steady state, say when all D/Dt's are zero. We obtain:

$$\Sigma_K \mathbf{Q}_K \cdot \mathrm{grad}\phi^*_K \leq 0 \tag{2.4.19}$$

which somehow would tell us that ϕ^* is the thermodynamic potential for diffusion. Now this is absolutely preposterous, as Ralph is quick to point out: the thermodynamic potential for diffusion is the chemical potential, be it mass based or molar based, not ϕ^*. At equilibrium μ is constant in space, not ϕ^*. Well, this is a serious objection, and we need to consider it seriously. At equilibrium. presumably, there are no diffusive fluxes, and hence $u_K^2 = 0$. So the question reduces to the following one: at equilibrium, is μ or $\phi^* = \mu + gh$ uniform in space?

Everybody thinks that μ is the right answer (although Sue keeps an open mind so far), but let's just discuss the point. Let's say we have a one-component system, at equilibrium. Pressure has the hydrostatic distribution, presumably, doesn't it? So pressure is not constant in space. Yes, Bob, it's generally going to be a minor difference, but it is the difference we are discussing. So at equilibrium p has the following distribution:

$$\mathrm{grad}p = -\Phi gh \tag{2.4.20}$$

Now we go to our thermodynamics textbook, and find out that:

$$(\partial\mu/\partial p)_\mathrm{T} = V^* = 1/\Phi \tag{2.4.21}$$

(both μ and V^* are per unit mass). So we combine (2.4.20) and (2.4.21) to obtain (everybody agreeing that T *is* constant in space at equilibrium):

$$\mathrm{grad}\mu = -gh \tag{2.4.22}$$

independently of what the local density of our one-component system may be. But with u^2 being zero, (2.4.22) is equivalent to:

$$\mathrm{grad}\phi = 0 \tag{2.4.23}$$

and so what is uniform in space at equilibrium, μ or ϕ? It all looks like a rabbit out of the hat, but it works, and we'll accept it.

OK, people are (mildly) satisfied about the equilibrium argument. But what about the $\frac{1}{2}u^2$ term in (2.4.17)? That's zero at equilibrium, but it's not when diffusion is taking place. This time Bob is exactly right with his observation that that term is going to be negligible in any practical case — u is, after all the diffusive velocity! That term does, however, have an important conceptual content: its presence makes in principle the diffusion equations hyperbolic, and hence the paradox discussed in remark (1.4.6) is resolved. But let's follow Bob's advice and drop the $\frac{1}{2}u^2$ term from (2.4.17). And perhaps even the gh term, which will also generally be negligible. Now (2.4.19) reduces to:

$$\Sigma_K \mathbf{Q}_K \cdot \mathrm{grad}\mu_K \leq 0 \tag{2.4.24}$$

which is in the expected form, and tells us that indeed diffusion is a dissipative phenomenon, with minus the LHS of (2.4.24) being the rate of energy dissipation in steady diffusion. In general, the rate of energy dissipation for isothermal diffusion is given by minus the LHS of (2.4.16).

Now of course thermodynamicists would want us to write a constitutive equation where $\mathbf{Q}_K$ is linearly related to $\mathrm{grad}\mu_K$, say something like:

$$\mathbf{Q}_K = -\Sigma_J L_{JK} \mathrm{grad}\mu_J \tag{2.4.25}$$

or whatever the molar based equivalent of (2.4.25) may be; if we give in and do that, they'll insist that the matrix L_{JK} is symmetrical, but we know better than being attracted in such a trap - in fact we know that the only restriction that the second law places on the matrix is that it should be positive definite. But we aren't going to write the constitutive equation in the form of (2.4.25) anyhow; we have as much confidence in Fick as we have in Fourier, and since concentration and chemical potential are monotonously related to each other as long as we don't cross phase boundaries, we prefer to write the constitutive equation for the flux in the form of Fick's law.

2.5 KINETICS

So far, we have been discussing cases where the nonequilibrium condition of the system considered is related to the fact that some quantity which is uniform in space at equilibrium exhibits in fact finite gradients. Now this is quite obviously not the only possibility, since if we consider a chemical reaction the system may not be at equilibrium simply because the reaction is not at equilibrium, even though there are no gradients of anything in our system. Let's begin with a very simple example: a 2:1 molar mixture of CO and O_2 at, say, room temperature and pressure in some sort of closed vessel. If we go to our Thermodynamics textbook and do the appropriate calculations, we conclude that at equilibrium there would be quite a bit of CO_2 in our system, and hence that our mixture is not at equilibrium. And yet there are no gradients of anything, are there? So we need a new line of reasoning.

Let's say that we define w as the ratio of the actual number of moles of CO_2 to the initial number of moles of CO in our system, so that for our mixture $w = 0$. Should the oxidation reaction proceed to completion, w would take a value of unity, so w does look like a convenient parameter measuring how far the reaction has gone; in fact, w is a dimensionless *extent of reaction*. Our thermodynamic calculation would tell us what the *equilibrium* value of w (let's call it w^*) is, and since our system is characterized by $w = 0$ while w^* is finite, we would conclude that our system is not at equilibrium. Now let's keep things as simple as we can, and so let's say that our system is an ideal gas. We'll take a basis of one mole of oxygen in our starting mixture, so that the initial number of moles n_0 is 3. The total number of moles $n(w)$ corresponding to some value w is $3 - w$ (as can easily be seen by considering that when $w = 1$, $n = 2$), and hence pressure is $(3 - w)RT/V$. Now what this tells us is that pressure depends on volume V, temperature T, and extent of reaction w. So the state is V,T,w, isn't it? Ralph has something to say here. He maintains that the state is V,T, and that at equilibrium $w = w^*$ and therefore it is not a state variable — thermodynamics is a science of equilibrium. That may be the case, Ralph, but then why is it called thermo*dynamics*? Doesn't the word dynamics convey the idea of something which changes in time? In any event, the important point is that it is conceivable that w is not equal to w^*, and that therefore we may have a nonequilibrium situation connected with the fact that something like w doesn't have the equilibrium value, rather than to the fact that something else exhibits a finite gradient. You have studied your Onsager, haven't you, Ralph? Wouldn't you be willing to concede that $w^* - w$ is perhaps the appropriate force, and $\mathrm{d}w/\mathrm{d}t$ the corresponding flux? Or do you think that forces have got to be gradients, and fluxes actual transports of something in space?

Sue says that while $w^* - w$ may be a "force" — a driving force, she thinks — $\mathrm{d}w/\mathrm{d}t$ should be called by a name other than flux, since it isn't the transfer of anything in space, it is the reaction rate. Bob agrees with this. So do we, and we won't call it a flux, we'll call it the rate of change of w — we'll be daring and call it the rate of change of the internal state variable w. Ralph says that we are going to find out that $\mathrm{d}w/\mathrm{d}t(w^* - w)$ has got to be positive and that therefore the two terms are conjugate flux and force. Bob has an unexpected surge of imagination here, and says that $(\mathrm{d}w/\mathrm{d}t)^3(w^* - w)^5$ is going to have the same sign as $\mathrm{d}w/\mathrm{d}t(w^* - w)$, and hence why shouldn't $(\mathrm{d}w/\mathrm{d}t)^3$ and $(w^* - w)^5$ be the conjugate force and flux? Sue says that she doesn't understand what conjugate forces and fluxes are anyhow, and she'd enjoy it being explained what all the fuss is about. Bob agrees, but Ralph doesn't. There is big turmoil in class, and we need to assert our role as teachers to proceed.

Let's say that V,T,w is the state, so that every function of state depends on them — after all, nobody has objected to the fact that p does depend on them. However, who is going to tell us what w may be? We begin with $w = 0$, granted; but what after that? Bob says that we need to know the kinetics of the reaction, say some function $r = \mathrm{d}w/\mathrm{d}t$. Nice indeed. What is the argument of the $r()$ function? Sue has a suggestion: w itself, and temperature. Why not volume as well? After all, if we decrease the volume we increase the initial pressure, and don't we know that the rate of a non-equimolar reaction depends on pressure? The point is granted. Hence $\mathrm{d}w/\mathrm{d}t = r(V,T,w)$. Since V,T,w is the state, the rate of evolution of the internal state variable is a function of state. By all means no, says Ralph; the rate of reaction is not a function of state, it is simply zero at equilibrium, and functions of state are only defined at equilibrium. "You can't pull out a trick again like you did in Section 2.2 — there must be some limit to everything". But Sue and Bob want to proceed, so we are going to say that r is a function of state by a majority vote if not unanimously. So we write:

$$A = a(V,T,w) \tag{2.5.1}$$

$$\mathrm{d}w/\mathrm{d}t = r(V,T,w) \tag{2.5.2}$$

as well as analogous things for, say, entropy or pressure, which we don't write down explicitly in order not to embarrass Ralph too much. Now we make the assumption that V and T are constant, do a bit of algebra, and conclude that Eq. (2.1.19) reduces to:

$$0 = (\partial a/\partial w)r = \theta r \tag{2.5.3}$$

where $\theta = \partial a/\partial w$ is going to be called the affinity, $\theta = \theta(V,T,w)$.

Now since θ and r are both functions of state (a statement accepted only under protest by Ralph), under rather mild assumptions of smoothness (to be however discussed in some detail later) one concludes that when $r = 0$, $\theta = 0$. But $r = 0$ is the equilibrium condition, so one concludes that the affinity is zero at equilibrium. It turns out that $\partial a/\partial w = (\partial A/\partial w)_{V,T} = (\partial G/\partial w)_{P,T}$, and so we recover the classical result that the free enthalpy G is minimized at equilibrium at constant temperature and pressure conditions. This is mildly satisfying to the audience: Sue and Bob were sure we would pull at least this rabbit out of our hat, and Ralph knew it already anyhow. But let's be a bit less cavalier about our assumptions of smoothness so that we may see whether there is anything interesting in them. Equation (2.5.3) implies that θr is at a minimum at equilibrium since it cannot be negative. It follows that $\mathrm{d}(\theta r)$ is zero at equilibrium, and since $r = 0$ by definition at equilibrium, one concludes that $\theta^*(\mathrm{d}r/\mathrm{d}w)^* = 0$. Now if $(\mathrm{d}r/\mathrm{d}w)^*$ is nonzero, this implies that $\theta^* = 0$ as we expected it to be; but if $(\mathrm{d}r/\mathrm{d}w)^*$ is zero one is in a bit of trouble, right? Bob comes to our rescue here: he says that derivatives are nonzero but for very exceptional cases, and hence he is willing to bet that $\theta^* = 0$. We'll buy his argument. However, it does mean that the $r(w)$ function is invertible at the equilibrium point, and let's not forget about that.

But now let's consider the case where there are two reactions to keep track of, say two internal state variables w_1 and w_2 (or, for that matter, M independent chemical reactions, with $\mathbf{w}$ the M-dimensional vector of extents of reaction). Now Eq. (2.5.3) becomes:

$$0 \geq (\partial a/\partial \mathbf{w}) \cdot \mathbf{r} = \theta \cdot \mathbf{r} \tag{2.5.4}$$

We can again reason that $\theta \cdot \mathbf{r}$ has a minimum at equilibrium and so on, and therefore conclude that:

$$(\mathrm{d}\mathbf{r}/\mathrm{d}\mathbf{w})^{\mathrm{T}*} \cdot \theta^* = 0 \tag{2.5.5}$$

This time it is Ralph who is of great help: Eq. (2.5.5) can easily be satisfied by a nonzero θ^* provided it lies in the kernel of $(\mathrm{d}\mathbf{r}/\mathrm{d}\mathbf{w})^{\mathrm{T}*}$. Isn't it nice to have a student who knows what the kernel of a linear transformation is? Ralph, please come up to the blackboard tomorrow and give a 30 minutes class on kernels of linear transforms. In the meantime, can you tell us whether you agree that the kernel only contains the zero vector provided the Jacobian of the transform is nonzero? (This is just to show that we aren't losing the one-upmanship game with Ralph, it's a Jacobian vs. a kernel — it's a draw so far). It's simply a question of the determinant of the $(\partial \mathbf{r}/\partial \mathbf{w})^*$ matrix — information which is swallowed with patient superiority by Ralph, poorly hidden indifference by Bob, and utter amazement by Sue, who knows as little about determinants as she does about Jacobians anyhow. So let's move to a simple example. Suppose that, in the neighbourhood of the equilibrium point $\mathbf{w}^*$, the rates of reaction can be expressed as linear combinations of the displacements from equilibrium — yes, Bob, we are going to assume differentiability, we aren't going to be so cosmopolitan as to consider the possibility of nonanalytic functions. So we write, in a neighbourhood of equilibrium:

$$r_1 = k(w_1^* - w_1) + \alpha k(w_2^* - w_2) \tag{2.5.6}$$

$$r_2 = \tau k(w_1^* - w_1) + \alpha\tau k(w_2^* - \mathrm{w}_2) \tag{2.5.7}$$

Today is Ralph's good day — he spots immediately our trick. There should be four independent arbitrary constants, but we have only three: α, τ and k. Very good indeed. And that's what makes the Jacobian zero, since:

$$\begin{vmatrix} k & \alpha k \\ \tau k & \alpha\tau k \end{vmatrix} = k^2(\alpha\tau - \alpha\tau) = 0 \tag{2.5.8}$$

Sue has some protest to file here: anything as obvious as $\alpha\tau - \alpha\tau = 0$ should not be worded in such an obscure fashion as saying that the Jacobian of the transform is zero. Bob has a protest as well: the chance of the four constants being exactly in the ratio we have assumed is vanishingly small, perhaps even less likely than finding a nonanalytic function in nature. We rather unsuccessfully try to hide the fact that we wholeheartedly agree with both of them. However, it is true that, should the situation be as described by (2.5.6–7), (2.5.5) would be satisfied by any θ^* of the form $\Gamma, -\Gamma/\tau$, with no need of Γ being zero — vectors of that type would constitute the kernel of the transformation, as Ralph has known for quite some time. We are going to take Bob's position and assume that the cases where the $\mathbf{r}(\mathbf{w})$ transformation is not invertible near equilibrium are exceptional ones, and that therefore the affinity vector is indeed zero at equilibrium. Incidentally, Eq. (2.5.4) imposes on the matrix in

Eq. (2.5.8) the usual requirement of being positive definite, and not the one of being symmetric; a positive definite matrix may well not be symmetric.

Now let's move to the more interesting questions, which concern the situation where one is not at equilibrium. Equation (2.5.4) shows that $-\theta \cdot \mathbf{r}$ is the (nonnegative) rate of energy dissipation. When $\mathbf{w}$ is not close to $\mathbf{w}^*$, the rate of energy dissipation will be finite — we cannot make it small by slowing down the process, since the reaction takes place at whatever rate the kinetics dictate. There is just no way that we can impose to a chemically reacting system to undergo "quasistatic" processes, unless we start from a condition of equilibrium and arrange things so that the system never departs significantly from equilibrium.

2.6 RELAXATION

In this section, we analyze the thermodynamics of systems which exhibit relaxation phenomena. For momentum transfer, this corresponds to the analysis of viscoelastic materials, which do not fall into any one of the categories discussed in Section 2.3.

It is important to understand the physics behind relaxation phenomena, and in order to do so we begin by considering a very simple system: a constant temperature box equipped with a piston (not a frictionless one: a real piston as we encounter them in the laboratory), so that the volume V can be imposed arbitrarily at any one time. The box is filled at time zero with 2 moles of NO_2. We are willing to assume that whatever is in the box behaves at all times as an ideal gas, say we never impose values of V which would result in pressures high enough to give deviations from ideal gas behaviour. Given the (constant) temperature T and the volume vs. time function $V(t)$, is anybody in the class willing to suggest what the pressure vs. time behaviour will be? The question looks so incredibly simple that nobody volunteers the answer, so we are going to suggest ourselves that it might be:

$$p(t) = (?)\, 2RT/V(t) \tag{2.6.1}$$

We have, however, placed a question mark, since of course the problem has a trick, as the whole class had suspected. Suppose we perform an actual experiment, and in fact the simplest possible one: we simply keep the volume constant at some value V_0 (we refrain from moving the piston), and we measure pressure. Take our word that the measured pressure will slowly decrease in time, approaching asymptotically some value p^* which is *smaller* than $2RT/V_0$. What is going on here?

Bob has seen the trick by this time. “You said that ‘whatever is in the box’ behaves as an ideal gas. That led me to think that ‘whatever is in the box’ may not be NO_2 at all times — indeed NO_2 may dimerize to N_2O_4. But if ‘whatever is in the box’ behaves as an ideal gas, then $p = nRT/V$, with n the total number of moles. If pressure decreases in a constant volume isothermal experiment, it must mean that n decreases. But that’s exactly what will happen if dimerization occurs, and pressure will decrease until chemical equilibrium is reached at a total number of moles less than 2". Very well, Bob, that is exactly what is going on. But suppose — just suppose — that we did not know that NO_2 may dimerize, and neither did we know that ‘whatever is in the box’ behaves as an ideal gas. How would we have described the results of our experiment? We might have tried the hypothesis that p depends on both V and dV/dt, but that wouldn’t have worked — V is constant, and dV/dt is constant as well, in particular at a value of zero. So we would have been forced into admitting that the system exhibits relaxation, say that pressure is given by a constitutive equation of the following general form:

$$p + \beta dp/dt = f(V) \tag{2.6.2}$$

Equation (2.6.2) can, at least qualitatively, describe the results of our experiment: p^* would be $f(V_0)$, and if the initial pressure was larger than p^* (as indeed we know it was), its value would decay in time, over a time scale of order β, towards its equilibrium value p^*.

Bob has some objection here. He says we would be very silly indeed going this way: he has seen the trick right away, and he is sure Ralph could calculate exactly $p(t)$ if he is given the kinetics of the dimerization reaction. Good point; however, there may be instances where something is happening which is somewhat more mysterious than a dimerization reaction, and in such cases one may be forced into something like Eq. (2.6.2). So what we will do is as follows. On the one side, we actually do the calculation of $p(t)$ following Bob’s line of thought; on the other side, we’ll use the results of this calculation to infer what the description of the thermodynamics of the system would be had we not understood so well the underlying physics.

Let $w(t)$ be the extent of reaction, defined as the number of moles of N_2O_4. Correspondingly, the number of moles of NO_2 is $2 - 2w$, and the total number of moles is $2 - w$. The definition is such that $w(0) = 0$, and $0 \leq w \leq 1$. The pressure is given by:

$$p(t) = [2 - w(t)]RT/V(t) \tag{2.6.3}$$

Now we need the kinetics of the reaction, but before that we need to establish the equilibrium condition. It's all ideal gas chemical thermodynamics, and so if K is the equilibrium constant expressed in terms of partial pressures, the equilibrium value of w, w^*, is delivered by the solution of the following equation:

$$KRT/V = w^*/4(1 - w^*)^2 \tag{2.6.4}$$

It makes sense: w^* decreases as V increases, since the dimerization reaction is favoured by compression. Now we write the kinetics in the following very simple form:

$$\mathrm{d}w/\mathrm{d}t = (w^* - w)/\beta \tag{2.6.5}$$

Equation (2.6.5) is the simplest form which makes sense: it guarantees that w always evolves towards the equilibrium value w^*, and that the time scale of evolution is β. But we know that $p = (2 - w)RT/V$, and correspondingly $p^* = (2 - w^*)RT/V$. We substitute these into (2.6.5), do a bit of algebra, and we obtain:

$$pV + \beta\mathrm{d}(pV)/\mathrm{d}t = p^*V \tag{2.6.6}$$

which has a form which is qualitatively similar to Eq. (2.6.2).

Now let's go back to our program as sketched by Bob. How do we proceed from here? Ralph has no problem. Equation (2.6.4) gives w^* as a function of V, say $w^* = w^*(t)$, and we may substitute this into (2.6.5), which admits the obvious integrating factor $\exp(t/\beta)$, and, since the algebra is not performed by Sue, we get rather quickly the result (Ralph has been good enough to generalize to the case where the initial value of w, w_0, may be different from zero):

$$w = w_0 \exp(-t/\beta) + \int_0^t \exp[(t' - t)/\beta]\, w^*(t')\mathrm{d}t'/\beta \tag{2.6.7}$$

Now suppose we limit attention to situations where $t >> \beta$, say we start giving importance to the results of our experiment only after it has been performed for a time large as compared to the *relaxation time* β. The first term on the RHS of (2.6.7) can be dropped, right? The system has a *fading memory*: after a time sufficiently in excess of β, the response becomes independent of the initial condition.

Now let's indulge in some definitions. We are considering our system at some time t, and the integral in (2.6.7) has a kernel $\exp[-(t - t')/\beta]$. Let s be defined as $(t - t')/\beta$. What does s mean? The quantity $t - t'$ represents "how long ago" time t' occurred. Then s is a "how long ago" variable, measured in the time scale β. $s >> 1$ means "long ago enough that it doesn't matter any more" (since $\exp(-s)$ is negligibly small). Now the function $w^*(t')$ also appears in the integral in (2.6.7). We define a new function $w^{*t}(s)$ as follows:

$$w^{*t}(s) = w^*(t') = w^*(t - \beta s) \tag{2.6.8}$$

We call $w^{*t}(s)$ the *history* of w^* at time t: it gives w^* as a function of "how long ago" it has been imposed on the system.

Remark 2.6.1

Given any quantity $\Gamma(t)$, one can always define a function $\Gamma^t(s) = \Gamma(t - \beta s)$, which is called the "history of Γ at time t". After all, what is history, if not the description of how long ago things happened, rather than at what particular time they did happen? And what is ancient history? Something which has happened at $s >> 1$, if β is interpreted as our intrinsic time span of memory, our average lifetime. Anything which happened a very large number of lifespans ago is ancient history, and (perhaps) it doesn't matter any more.

One may now substitute the definitions in (2.6.7), do a bit of algebra, observe that the integral is dominated by the values of the integrand for small s's, and finally obtain:

$$w = \int_0^\infty \exp(-s) w^{*t}(s) \mathrm{d}s \tag{2.6.9}$$

Ralph has a protest to file. He maintains that our kinetic equation (2.6.5) is not realistic. If the system behaves as an ideal gas and all that, the kinetics will presumably be of the mass action type. Good point indeed. Let's follow the implications. We want to write a constitutive equation for $\mathrm{d}w/\mathrm{d}t$ in terms of w, w^*, and β. There is some turmoil in the class while trying to do the algebra correctly, but an agreement is finally reached on the following equation:

$$\mathrm{d}w/\mathrm{d}t = [w^*(1-w)^2 - w(1-w^*)^2]/\beta \tag{2.6.10}$$

Now we need, in principle, to integrate (2.6.10), subject to $w(0) = 0$, and this is a bit hard to do for an arbitrary $w^*(t)$, since no obvious integrating factor is available this time. Ralph maintains he sees no problem, since numerical integration is going to be a breeze. It surely is, but what if the kinetics are not quite of the mass action type? Since we have given up on the simplest possible form (Eq. 2.6.5), we have to ask ourselves what happens *in general* (see Remark 1.8.15), say in the case where $\mathrm{d}w/\mathrm{d}t = f(w)/\beta$, with $f(w^*) = 0$ and $(w^* - w)f(w) = 0$. All we can say in the general case is that w depends on $w^{*t}(s)$, or equivalently on $V^t(s)$, with the property of fading memory — say the initial volume ceases to have any influence at times well in excess of β. Since $p = (2 - w)RT/V$, we conclude that pressure depends on the instantaneous value of V, as well as on its history $V^t(s)$, with the dependence on $V^t(s)$ enjoying the fading memory property:

$$p = f_{\mathrm{P}}[V^t(s);V] \tag{2.6.11}$$

$f_{\mathrm{P}}[\,]$ is a *functional* of the history of volume which also depends parametrically on the instantaneous value of volume. Sue, don't panic, please. Functionals are not such a terrible thing — some definitions and basic concepts are discussed in Appendix 2.6. A definite integral is a very simple example of functional — the value of the definite integral of $f(x)$ depends on what the function $f()$ might be, not on the value of the dummy variable x, do you remember that? That's where the concept of a dummy variable comes in. Well, an integral is a linear operator, isn't it? Now if we have some kernel $k(x)$, the definite integral of $k(x)f(x)$ still depends only on what the function $f(x)$ might be, and it is still a linear operator — the definite integral of $2k(x)f(x)$ is twice that of $k(x)f(x)$. That's the case of (2.6.9) — $\exp(-s)$ is the kernel. Well, a functional is something the value of which depends only on what $f()$ might be, but not necessarily linearly. That's all there is, really, Sue. One does need *some* complex mathematics to describe a complex phenomenon, after all.

So let's suppose we know what a functional is. Now if we believe (2.6.11) (remember, we are still only considering the case where temperature is the same at all times), what should we write for free energy A? Sue is the first one to recover from the shock of having met functionals, since they don't look to her any more mysterious than ordinary functions, and so she suggests that we should write:

$$A = f_A[V^t(s);V] \tag{2.6.12}$$

Very well indeed, Sue. Let's proceed. What about the second law? We'll need to use (2.1.19), and fortunately the second and fourth term are zero since temperature is constant in both space and time. Furthermore, it's all ideal gases, and so the net rate of work W is expressible the way thermodynamicists like it to be. There is no fluid mechanics to worry about, so that substantial time derivatives are just ordinary ones, and we end up with:

$$\mathrm{d}A/\mathrm{d}t + p\mathrm{d}V/\mathrm{d}t \le 0 \tag{2.6.13}$$

Ralph has a happy expression on his face now. He doesn't believe anything of what we are doing, since he has been taught that when dealing with thermodynamics the most of mathematics one may need are partial derivatives (with subscripts), but he knows we are going to be in trouble with Sue now: we need to take the time derivative of free energy in order to make any use of (2.6.13), and this will lead us into some real stuff in functional analysis and Sue will be lost forever. Well, let's give it a try. We write on the blackboard something which we reserve to discuss after it has been written: namely, from Eq. (2.6.12) we deduce that:

$$\mathrm{d}A/\mathrm{d}t = (\partial f_A/\partial V)\mathrm{d}V/\mathrm{d}t + D \tag{2.6.14}$$

Ralph has his hand up even before we finish writing it down. "What is D?". Well, let's leave its mathematical structure somewhat mysterious so far, and let's see what it means physically. At some time we stop the piston and keep the volume constant. Equation (2.6.14) tells us that the rate of change of free energy equals D, and (2.6.13) that D is going to be nonpositive. For our specific example, D represents the rate of energy dissipation should the volume be kept constant, say the rate of energy dissipation connected with the evolution of the chemical reaction. Ralph has another ace up his sleeve. "And what is $\partial f_A/\partial \mathrm{V}$?". That's a good one. It is the partial derivative of f_A with respect to V, i.e., the derivative when $V^t(s)$ is kept constant. Thermodynamicists would write it with a $V^t(s)$ subscript, but that's too difficult typographically, isn't it, Ralph? Ralph's objection is a more serious one: how can we keep $V^t(s)$ constant and still consider a variation, albeit a differentially small one, of V? Isn't V nothing else than $V^t(0)$? Very good point indeed; we need to agree that (2.6.12) has to be understood in the sense that $V^t(s)$ over $(0,\infty)$ is the argument function in (2.6.12), not $V^t(s)$ over $(0,\infty)$. Yes, Bob, it does make a difference on occasion — those mathematicians who taught you calculus once were not entirely out of touch with reality. Go back to the physics of our problem. Suppose at some time t, whatever w may be, one changes the volume V very quickly (i.e., over a time interval which is negligibly small as compared to β, so that during our change of volume the reaction has in practice no chance to proceed). Pressure is $(2 - w)RT/V$, and hence there will be an abrupt change in pressure, and presumably one of free energy. $\partial f_P/\partial V$ would be simply $-(2-w)RT/V^2$, which of course still depends on V and on $V^t(s)$. *In an experiment done on a time scale much less than the relaxation time, the response (pressure) would appear as uniquely determined by the forcing function (volume).* Should β be very large indeed, so that all our experiments would be performed over a time scale very small with respect to it, pressure would appear to be a unique function of volume. Physically, a very large β value means a chemical reaction which is so slow as not to proceed at all to all practical purposes. A mixture of oxygen and hydrogen at room temperature would behave just as an ordinary mixture with no chemical reaction involved, because the kinetics of water formation at room temperature are extremely slow.

Under more than mild protest from Ralph, and to the utter amazement of Sue, who is willing to listen nonetheless, we proceed to substitute (2.6.14) into (2.6.13) to obtain:

$$[\partial f_A/\partial V + f_P]\mathrm{d}V/\mathrm{d}t + D \leq 0 \tag{2.6.15}$$

Now, Sue, can you deduce anything from (2.6.15)? The whole thing in square brackets depends only on V and $V^t(s)$, and hence we can still impose any $\mathrm{d}V/\mathrm{d}t$ we wish. Ralph is very mad indeed at this point. No we can't, he maintains; if we change $\mathrm{d}V/\mathrm{d}t$ we change $V^t(s)$ as well. A long discussion on functional analysis, topology of the space of histories, and other mysterious subjects ensues, at the end of which the point is reluctantly granted and the following conclusion is reached:

$$p = -\partial f_A/\partial V = -\partial A/\partial V|_{V^t(s)} \tag{2.6.16}$$

where the rightmost expression has been put in to satisfy Ralph's hunger for partial derivatives with subscripts in thermodynamics. Notice that neither p nor A have their equilibrium value, but they are still related to each other by an equation analogous to the classical one — except that a partial derivative occurs, rather than a total one (temperature is constant throughout).

Everybody in the class now agrees that, whatever D might be, (2.6.15) can only be satisfied if it is intrinsically nonpositive, say if:

$$D \leq 0 \tag{2.6.17}$$

There are mixed opinions in the class at this point. Ralph maintains that (2.6.17) does not mean anything since we have not given the appropriate mathematical definition of D; he says that, should we try to do that, it will lead us into Frechet differentials, functional derivatives, and all that, and he suspects we don't want to go into that area because we have not, as he did, taken a course in functional analysis. Bob says that D is some kind of derivative, and he doesn't want to know all the mathematical details about it — it's a derivative of a functional representing something in real life, and hence it is going to exist anyhow, so let's forget about the whole thing. Sue keeps somewhat silent, trying to hide the fact that she is still somewhat hazy about ordinary derivatives, and hence the definition of D is only going to enlarge the already wide horizon of her area of haziness. However, she suggests we should go back to our discussion of heat transfer, and we should better let her understand why, when there is relaxation, heat may, if only occasionally, flow in the direction of increasing temperature. The physics of it, please, not the abstruse mathematics.

Let's give it a try. Let's go back to our case where there is relaxation, say Eq. (2.2.12) is the constitutive equation for the heat flux. It is however a somewhat unusual one — it is a differential equation for **q**. We aren't afraid of differential equations, of course (Sue is, but let's pretend she isn't). We go ahead and integrate (2.2.12), and let's stay as general as possible, so that τ on the RHS is some arbitrary function of time $\tau(t)$. Ralph? "It's all going to be the same as before". Right. Within the same simplifying assumptions one has:

$$\mathbf{q}(t) = -k \int_0^\infty \exp(-s)\tau^t(s)\mathrm{d}s \tag{2.6.18}$$

So what should we write instead of (2.6.11)? The body is stationary, so we don't need to worry about p and V. For the heat flux we write:

$$\mathbf{q} = f_{\mathrm{q}}[\tau^{t}(s); T] \tag{2.6.19}$$

Sue has an objection. In (2.6.11), the argument of the functional is not only the history of volume, but also the instantaneous volume itself. Why don't we have τ as a parametric argument in (2.6.19)? Well, in the case of (2.6.11) we were forced into including V because we know that, at any given extent of reaction, pressure still depends explicitly on volume. In the present case, we have no reason to suppose that, at any given $\tau^{t}(s)$, $\mathbf{q}$ still depends explicitly on τ. This is not an exercise in making things as complex as one can; we are trying to describe real life phenomena as they actually take place in nature. We have included T so as to be able at least to describe the usual Fourier law with a temperature dependent thermal conductivity; we have not included the history of temperature $T^{t}(s)$ because we were not forced into doing so by some physical consideration. Equation (2.6.18) is the simplest possible class of constitutive equations capable of describing relaxation in heat conduction. The class is mildly satisfied with this, and we may proceed. Consistently with (2.6.19), we write:

$$A = f_{\mathrm{A}}[\tau^{t}(s); T] \tag{2.6.20}$$

$$S = f_{\mathrm{S}}[\tau^{t}(s); T] \tag{2.6.21}$$

Now we go through the same kind of algebra with the second law, and we obtain, after some discussion:

$$f_{\mathrm{S}} = -\,\partial f_{\mathrm{A}}/\partial T \tag{2.6.22}$$

$$D + (\mathbf{q} \cdot \tau)/\Phi T \le 0 \tag{2.6.23}$$

and this now does not require $\mathbf{q} \cdot \tau$ to be invariably nonpositive: heat can indeed, on occasion and under rather extreme conditions of unsteadiness, flow in the direction of increasing temperature. What does it mean, physically? Very simple, after all. If free energy is given by (2.6.20), it means that free energy can in fact be accumulated by some appropriate choice of the temperature gradient history $\tau^{t}(s)$. Then on occasion this free energy may be released to drive the heat flux against a temperature gradient — D being large enough and negative to allow, in spite of (2.6.23), a nonnegative $\mathbf{q} \cdot \tau$. In the

Maxwellian theory of gases, the heat flux is related to the actual movement of molecules; should the temperature gradient be reversed very quickly, the inertia of molecules would keep them moving in the same direction for at least some very short time, and hence the heat flux would be in the direction of increasing temperature.

Bob has a philosophical problem at this stage. For the case of the chemical reaction going on in a box, he reiterates the idea that we would be aware of it, and we would therefore describe the system with an internal state variable w, or maybe a vector of internal state variables $\mathbf{w}$, rather than through Eq. (2.6.11). For the heat transfer case, he can see that the whole analysis is self-consistent, but he has not been presented with any hard evidence that anything more complex than Fourier's law is in fact needed. So when is this whole type of analysis going to be both necessary and useful? Well, the area where it is absolutely needed is that of viscoelasticity. Remember the experiment discussed in Section 1.10? Well, Eq. (1.10.7) has the same form as Eq. (2.2.12), and hence one would arrive at similar results.

The problem, however, is that we would need to develop the analysis for the properly invariant three-dimensional formulation of Eq. (1.10.7), and the appearance of a time derivative on the LHS creates some major problems. With heat transfer in a stationary body we have avoided those problems, because in a stationary body the partial derivative with respect to time is properly invariant. But in momentum transfer the body considered is not stationary by definition, and therefore the mathematics become significantly more complex, and we will refrain from working them out. Equations (2.6.11–12) are a nice scalar, isothermal formulation of a system exhibiting relaxation in its mechanical response — though the only kind of flow considered is one of pure expansion or compression. So far, the whole class has hardly digested the idea of having to do with functionals; should this be compounded with all the tensors appearing in Section 2.3, things would tend to get out of hand.

However, let's discuss Bob's point from an another angle. In our example of dimerization, we understood the underlying physics, and hence we have introduced an internal state variable w. We have also written a constitutive equation for $\mathrm{d}w/\mathrm{d}t$. We then have calculated pressure because we knew what the pressure would be at any value of w. The logical path has been as follows:

1. Identification of w.
2. Constitutive equation for $\mathrm{d}w/\mathrm{d}t$, which contains $V(t)$.

3. Solution of the kinetic equation to obtain w as a functional of $w^t(s)$.
4. Calculation of pressure as a function of w.
5. (Optional). Conceptual elimination of w from the equations so as to obtain p as a functional of $V^t(s)$.

Now suppose we consider a polymeric fluid, which is known to exhibit viscoelasticity. We could, in principle, follow the same route. Step 1 would consist in identifying the instantaneous morphology of the polymer as the internal state variable; this however is not easy, since we are not too sure how we should describe the morphology. Step 2 would consist in writing a differential equation for the rate of evolution of the morphology in response to some assigned kinematics of motion; again, this would require modelling the relevant micromechanics. Step 3 would now give the morphology as a functional of its history. Step 4 would require calculating the stress corresponding to any given morphology, which would thus depend on the history of deformation. Step 5 may be impossible to perform formally, but is again conceptually possible. This route has been followed in the literature, and it is usually referred to as molecular modelling (sometimes kinetic theory) of polymeric materials. One may also skip the whole exercise and write down directly stress as a functional of the history of deformation, and then proceed to a thermodynamic analysis along the lines discussed in this section.

Ralph raises an objection here. He says that we have established that, provided f_A is such that (2.6.16) and (2.6.17) are satisfied, we have a theory which does not contradict the second law. He is willing to forget about the fact that we have not discussed the mathematical structure of D; but he has a problem nonetheless. Thermodynamics may, perhaps, apply to nonequilibrium conditions, though he still does not quite believe it; but should it not also apply to equilibrium? What is an appropriate equilibrium theory based on (2.6.11–12)?

Indeed. So let's ask ourselves how we would conduct an equilibrium experiment. We would keep the volume constant at some value V_0 for some very long time βT (ideally, T should be well in excess of 1), and we would measure pressure. Now at the beginning of our experiment we don't know what the previous history of volume of our sample might be (say we don't know the value of w_0); let's say it is $V^0(s)$,an unknown history function. So at time T the history $V^T(s)$ is:

$$\begin{aligned} V^T(s) &= V_0, \qquad s \le T \\ &= V^0(s-T), \quad s > T \end{aligned} \tag{2.6.24}$$

Now, as T becomes very large, the property of fading memory (if properly formulated mathematically), will guarantee that the response to $V^T(s)$ approaches the response to $V^*(s)$, which is defined below:

$$V^*(s) = V_0 \text{ for all } s \tag{2.6.25}$$

$V^*(s)$ is the history one would have if the sample had always been kept at a volume V_0, and hence the response to it is the equilibrium response. Therefore the equilibrium value of free energy is:

$$A^* = f_A[V^*(s);V_0] = f_A{}^*(V_0) \tag{2.6.26}$$

But since (2.6.16) must apply for every history of volume, it must apply to $V^*(s)$ as well, and hence we obtain for the equilibrium pressure:

$$p^* = -\mathrm{d}f_A{}^*/\mathrm{d}V_0 \tag{2.6.27}$$

Equation (2.6.27) is the classical equation of equilibrium thermodynamics for an isothermal system. Hence we conclude that the classical equilibrium theory is recovered correctly. Another example is given at the end of Appendix 2.6.

Appendix 2.6

We want to present some basic material on the differentiation of functionals. Let's begin by considering an ordinary scalar function of a scalar argument, $f(y)$, and let $f'(y)$ be the ordinary derivative of $f(y)$. Although the equation below is not the usual definition of the derivative, it can easily be seen that it coincides with it:

$$f(y+x) = f(y) + f'(y)x + R \tag{2.6.a.1}$$

where the residual R enjoys the following property:

$$\lim_{|x|\to 0} R/|x| = 0 \tag{2.6.a.2}$$

The term $f'(y)x$ depends on both x and y, and it has the property of being linear in x.

The advantage of the form chosen in (2.6.a.1) for the definition of the derivative $f'(y)$ is that it can easily be generalized to a functional $\boldsymbol{f}[f(s)]$. In fact we can write the following definition of the Frechet differential $\delta\boldsymbol{f}$

$$\boldsymbol{f}[f(s) + g(s)] = \boldsymbol{f}[f(s)] + \delta\boldsymbol{f}[f(s) \mid g(s)] + \mathbf{R} \tag{2.6.a.3}$$

with the requirement that $\delta\boldsymbol{f}$ should be linear in its second argument function $g(s)$. However, we need to deal with the residual term $\mathbf{R}$, and write the analog of (2.6.a.2). In order to do so, we need to establish what we mean by taking the limit where the difference function $g(s)$ (which is the analog of the difference value x in (2.6.a.1)) approaches zero. In other words, we need to establish what we mean by saying that $f(s)$ and $f(s) + g(s)$ are close to each other. Let's regard the argument functions $f(s)$ as points in the space over which the functional $\boldsymbol{f}$ is defined: we need to understand when we regard two such points as close to each other, but we do not need to establish what the distance between two such points might be when it is not infinitesimal. This is usually worded by saying that we need to know the topology of the space of argument functions, but not its metric.

Now one way to determine a topology is by assigning a metric: if we know how to measure distances between points, we of course also know how to identify the case where the distance is vanishingly small. It is therefore useful to actually assign a metric, i.e., a way of measuring the distance which we will indicate as $\|g(s)\|$. (We do not use the symbol $|g(s)|$, because that symbol identifies the absolute value of the value of $g(s)$ corresponding to some assigned value of s). Now distances between points must satisfy the triangular rule (sometimes called Schwartz rule), which can be written as:

$$\|g(s)\| = \|-g(s)\| \geq 0;$$

$$\|g(s) + h(s)\| \leq \|g(s)\| + \|h(s)\| \tag{2.6.a.4}$$

$\|g(s)\|$ is called the norm of $g(s)$. Of course, different norms could induce the same topology, but we need not worry about that here. Suppose we have chosen a norm and hence a metric and a topology. Now we require the residual $\mathbf{R}$ in (2.6.a.3) to have the following property:

$$\lim_{\|g(s)\|\to 0} \mathbf{R} / \|\mathbf{g(s)}\| = 0 \qquad (2.6.a.5)$$

and we are done with the definition of the Frechet differential. However, it is important to bear in mind that a Frechet differential may or may not exist according to which topology we have chosen, and that therefore we can choose the topology in such a way as to represent some physical ideas about what we consider as two close argument functions, and afterwards assume that the functionals of interest are Frechet differentiable with respect to that topology. This is, quite clearly, a sophisticated version of Bob's belief that everything in nature is always smooth and differentiable.

There are three important types of topologies which are used. The first one is called uniform, and it simply requires the difference between the values of two functions to be small for all values of s:

$$\|g(s)\| = \max|g(s)| \qquad (2.6.a.6)$$

The second topology is called Euclidean, and it is one where the two functions are regarded as being very close if they have close values almost everywhere, even if they have quite different values over some sufficiently small set of s values (from now on we will indicate with $<>$ integrals over s ranging from 0 to ∞):

$$\|g(s)\| = <|g(s)|> \qquad (2.6.a.7)$$

Finally, there is the weighted topology, where the two functions are regarded as close when their values are almost the same over some range of s values, but are allowed to be quite different for some other range of s values. This is obtained by introducing a weighting function $h(s)$:

$$\|g(s)\| = <h(s)|g(s)|> \qquad (2.6.a.8)$$

The weighted topology is the one which allows us to formalize the concept of fading memory. This is accomplished by taking an $h(s)$ which is strongly decreasing with increasing s, say e.g:

$$\|g(s)\| = <\exp(-ks)|g(s)|> \qquad (2.6.a.9)$$

One would then assume that the functional f is Frechet differentiable with respect to such a topology, which would ensure that the response is the same provided the forcing function $f(s)$ has been the same in the recent past, irrespective of what it might have been in the distant past.

Since the Frechet differential is linear in its second argument, it can be expressed as a linear transform of $g(s)$, say:

$$\delta f[f(s)|g(s)] = \langle l(s)g(s)\rangle \tag{2.6.a.10}$$

The kernel of the transform, $l(s)$, is called the functional derivative of ***f***, and it depends on $f(s)$, just as the ordinary derivative f' depends on y.

A simple example may perhaps clarify several issues. Consider the following functional:

$$f[f(s)] = \langle f(s)^2\rangle \tag{2.6.a.11}$$

One immediately obtains the following result:

$$\langle [f(s) + g(s)]^2\rangle = \langle f(s)^2\rangle + \langle 2f(s)g(s)\rangle + \langle g(s)^2\rangle \tag{2.6.a.12}$$

which identifies $\langle 2f(s)g(s)\rangle$ as the Frechet differential, $2f(s)$ as the functional derivative, and $\langle g(s)^2\rangle$ as the residual. With an Euclidean topology, (2.6.a.5) is satisfied, and hence the Frechet differential exists with respect to that topology. However, had one assigned a more stringent condition of closeness, say $\|g(s)\| = \langle g(s)^2\rangle$, (2.6.a.5) would not be satisfied, and hence the functional is not Frechet differentiable with respect to this alternate topology.

Finally, we come to the exact definition of the quantity D which has bothered Ralph for so long. Consider the history of volume at time $t + \mathrm{d}t$. It is given by:

$$V^{t+\mathrm{d}t}(s) = V^t(s - \mathrm{d}t) = V^t(s) - V^{t\prime}(s)\mathrm{d}t \tag{2.6.a.13}$$

where $V^{t\prime}(s)$ is the ordinary derivative of $V^t(s)$. It is now simply a matter of substituting into the definitions to obtain:

$$D = -\delta f_{\mathrm{A}}[V^t|V^{t\prime}(s); V] \tag{2.6.a.14}$$

If the Frechet differential of a functional exists, it can be calculated as follows:

$$\delta f[f(s)|g(s)] = (\mathrm{d}/\mathrm{d}\varepsilon)|_{\varepsilon=0}\, f[f(s) + \varepsilon g(s)] \tag{2.6.a.15}$$

For instance, applying (2.6.a.15) to (2.6.a.11) gives the Frechet differential as $<2f(s)g(s)>$. Notice however that the converse is not true: as the example shows, the derivative in (2.6.a.15) may exist even if the Frechet differential does not. Bob, one can't be quite so cavalier with Frechet differentials as you are accustomed to be with ordinary derivatives.

We can now discuss in some detail what the condition that $D \leq 0$ means in the specific case of a chemical reaction.If volume is held constant, the functional $\boldsymbol{f}_A$ is equivalent to the following chain of equations:

$$A = a(w) \tag{2.6.a.16}$$

$$w = \boldsymbol{f}_W[w^{t*}(s)] \tag{2.6.a.17}$$

$$w^{t*}(s) = h(V^t(s)) \tag{2.6.a.18}$$

In the specific example considered, $\boldsymbol{f}_W$ is the (linear) functional in Eq. (2.6.8), and function $g()$ is delivered by the solution of Eq. (2.6.4); the function $a(w)$ can easily be calculated from the theory of ideal gas mixtures. The derivative called for in (2.6.a.15) is obtained from the usual chain rule, which, for the case where $g(s)$ is $-V^{t\prime}(s)$, furnishes after easy algebra:

$$D = (\mathrm{d}a/\mathrm{d}w)(\mathrm{d}w/\mathrm{d}t) \leq 0 \tag{2.6.a.19}$$

Equation (2.6.a.19) is of course the requirement that, at constant volume, free energy can only decrease as a result of the chemical reaction. If, over the range $0 < w < 1$, there is only one minimum of free energy (as the equilibrium condition (2.6.4) guarantees to be the case), $\mathrm{sgn}(\mathrm{d}a/\mathrm{d}w) = -\mathrm{sgn}(w^* - w)$, and hence (2.6.a.19) is satisfied by the assumed constitutive equation (2.6.5).

2.7 PHASES

Contrary to common belief, the thermodynamic theory of multiphase systems, even in its restricted scope of a theory of phase equilibria, is conceptually much more complex than that of chemical reactions. The reason for this is that the theory of multiphase systems is basically a theory of stability, and this is known to be conceptually hard to develop. In this section, we only present a glimpse of the problems involved.

Let's begin with a very simple example: a one-component system which may undergo an ordinary phase transition. Let's say H_2O at its boiling point, to make things very simple indeed, at least apparently. We go to our book on thermodynamics, and we find out the we may regard the state as uniquely determined by temperature T and pressure p — do you agree so far, Ralph? Good. So would you care to tell us what the specific volume of H_2O at 100°C and 1 atm is? "Steam or water, Ma'm?". We don't know, Ralph. Pure H_2O at 1 atm and 100°C: if the statement that the state is entirely identified by p and T, that should be enough. Ralph finds the question a mean one. Anybody else have a suggestion? Could we perhaps do an experiment? Bob thinks this suggestion is preposterous: describe the experiment, and he is confident he can predict the outcome. OK, Sue, describe the experiment, please. But no frictionless pistons, Bob does not believe in their existence.

Sue gives it a try. She is perplexed, because this time, much to her dislike, she thinks Ralph is right and Bob is wrong. "No frictionless piston. I'll have a variable volume box, with the piston driven by appropriate electronics triggered by a pressure transducer, so that pressure is kept at 1 atm. I'll wait long enough to be sure that equilibrium is reached, and I'll read the volume". Good. Particularly the part about waiting long enough. How do we start the experiment, Sue? "At some value V_0 of the volume, it doesn't matter — you just showed us the initial condition does not matter if one waits long enough". Now, Ralph, it's all quasistatic, isn't it? Can we assume equilibrium prevails at all times? Something like 'a sequence of equilibrium states', doesn't that make you happy? Ralph knows we have a trick up our sleeve, but he's forced into conceding the point. Then, whatever V_0 may be, provided it is between the steam and the water specific volumes, pressure will never deviate from 1 atm, the pressure transducer will never trigger anything, and the volume will just stay where it was at the beginning, so that we are left with $V = V_0$, and that is rather disappointing, isn't it? We do the same experiment starting with different initial volumes, and we get different answers, no matter how long we wait. Bob has got it by this time. "The real independent variable is volume, not pressure. If we know volume, and it's anywhere between the steam and water specific volumes, than pressure is 1 atm. Pressure p is some function $f(V)$ at any given temperature, but it happens not to be invertible at 100°C in some range of V values". Bob, congratulations. We have finally found something which is not invertible, and hence is not differentiable either, or at least there is some such problem — and it's just H_2O at 100°C, under equilibrium conditions.

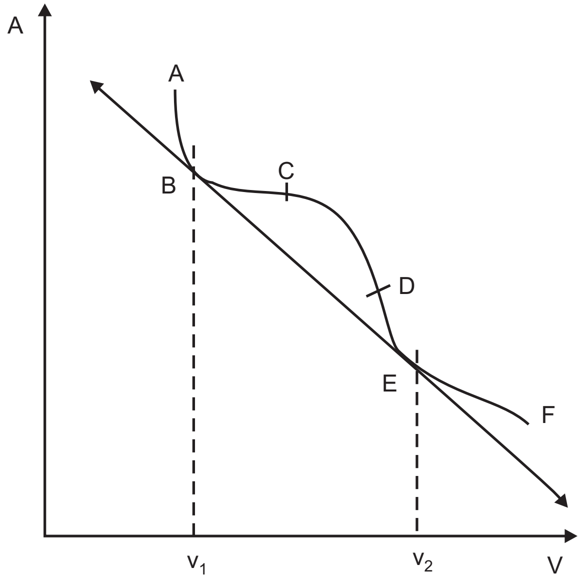

Fig. 2.7.1. The $a(V)$ curve for an ordinary phase transition.

Well, since we have established that even for such a simple system there is room for thought, let's see what thoughts we might produce. Should we work with free energy, A? Temperature is held constant, and so $A = a(V)$ — we have agreed that the pair V,T identifies the state. Ralph, do you agree that pressure is $-\mathrm{d}a/\mathrm{d}V$? The point is granted. Hence $a(V)$ is a monotonously decreasing function, isn't it? Otherwise we would need to admit the possibility of negative pressures; well, let's do that. Materials exhibiting a finite tensile stress are known to exist, after all, and we are lucky they do exist, otherwise civil engineering could not have progressed very far beyond Gothic cathedrals. Now let's say the $a(V)$ curve has the shape sketched in Fig. 2.7.1, where we have been conservative and have kept $a()$ a monotonously decreasing function. The corresponding $f(V)$ function is sketched in Fig. 2.7.2, and the whole class sees the end of the tunnel now. "The usual stuff on phase equilibria after all, with the Van der Waals constitutive equation". Ralph has just muttered this remark, but we heard it nonetheless. He's right, of course, as he always is. But we are going to make Ralph happy this time, because we are going to make use of a bit of mathematical overkill, much to Bob's distaste.

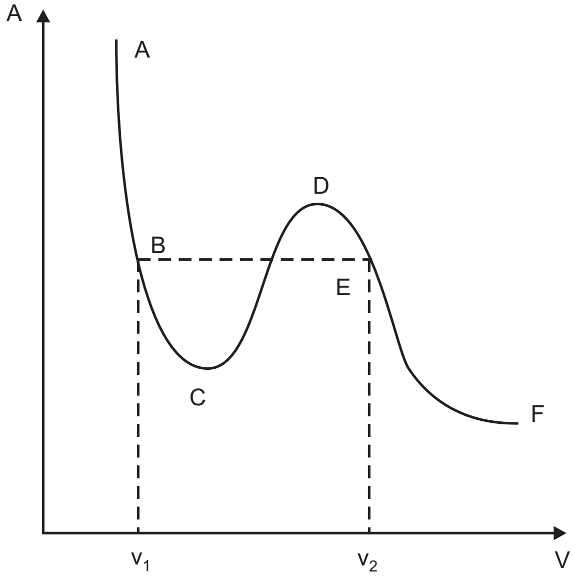

Fig. 2.7.2. The $p(V)$ curve corresponding to Fig. 2.7.1.

Suppose we consider our system at some volme V_0, and correspondingly with some free energy $A_0 = a(V_0)$. We now consider the possibility that it might, by Brownian motion, or by any mechanism whatsoever, split into a (possibly very large) number of regions, each one of which has some specific volume V_I which is differentially close to V_0:

$$V_I = V_0 + \mathrm{d}V_I \tag{2.7.1}$$

Let β_I be the fraction of the original mass having specific volume V_I. The following two conditions need of course to be satisfied:

$$\Sigma\beta_I = 1;\ \beta_\mathrm{I} > 0 \tag{2.7.2}$$

$$\Sigma\beta_I \mathrm{d}V_I = 0 \tag{2.7.3}$$

Now let's calculate the free energy of the system split in such a way, which is simply the sum of the free energies of all its parts:

$$A = \Sigma\beta_I a(V_I) \tag{2.7.4}$$

Now, Bob, under which conditions will the system tend to go back to the uniform state? Agreement is reached that it will happen provided the free energy of the uniform state is less than the value given by Eq. (2.7.4), for any choice of β_I and $\mathrm{d}V_I$ which satisfies (2.7.2–3). Good, we will call this a locally convex point: it is stable to every infinitesimal nonuniformity compatible with mass balance and the imposed constraints (in our case, that the total specific volume should be V_0). Now we can expand (2.7.4) in a Taylor series using (2.7.1). The zero order term reproduces V_0 because of (2.7.2); the first order term is zero because of (2.7.3); and thus the stated condition reduces to:

$$a''(V_0)\Sigma\beta_I(\mathrm{d}V_I)^2 > 0 \tag{2.7.5}$$

Since the β_I's are positive, this is equivalent to the requirement that $a''(V_0)$, the second derivative of $a(V)$ at V_0, should be positive. In more complex cases the condition of local convexity may not be quite so trivial, of course. Now, since $p = -\mathrm{d}a/\mathrm{d}V$, the condition of local convexity is satisfied whenever $\mathrm{d}p/\mathrm{d}V < 0$ — which, if you remember the “usual stuff” correctly, corresponds to the one-phase equilibrium branches of the plot in Fig. 2.7.1 (portions AB and EF), but also to some nonequilibrium conditions: portion BC, which is superheated liquid water, and portion DE, which is supercooled steam. Now does that make sense?

Sue has emerged from some private thoughts just in time to suggest that yes, it makes sense. Supercooled vapours and superheated liquids are known to be encountered on occasion, and they wouldn’t if they were unstable to infinitesimal perturbations. She remembers, if somewhat vaguely, that conditions which are unstable to infinitesimal disturbances will never be encountered in practice. Good. That’s exactly the point —and it does show that phase separation is a problem in stability. Ralph has something to say here. A locally convex point may be stable to infinitesimal perturbations, but that does not guarantee it is stable to finite ones. Ralph is right — when isn’t he? So let’s proceed a bit with our mathematical overkill. Let’s see what the condition is for a point to be absolutely stable, even to finite disturbances. That should lead us into some nice nonlinear stability analysis, shouldn’t it, Ralph? But first the definition. Suppose our system unmixes to several regions (mass fractions β_I) endowed with specific volumes V_I which, this time, are not required to be infinitesimally close to V_0. The following condition needs to be satisfied in addition to (2.7.2):

$$\Sigma\beta_I V_I = V_0 \tag{2.7.6}$$

Now the free energy of this multiphase system is given again by (2.7.4), with β_I satisfying (2.7.2) and V_I satisfying (2.7.6), and the point V_0 may be called globally convex if, given any β_I's and V_I's satisfying the required constraints, the free energy as given by (2.7.4) is always larger than $a(V_0)$. Does that satisfy everybody? It is a nonlinear stability analysis problem, and, much to Sue's agreement, we won't even try to solve the mathematics of it. However, we will observe that, given enough ingenuity, we can convince ourselves that portions AB and EF in Figs. 2.7.1 and 2.7.2 correspond to globally convex points. Of course, any globally convex point is also locally convex, but the converse is not true, and hence one may find points which are locally convex but not globally convex (portions BC and DE in our specific case). Such points may well be called metastable — they are stable to infinitesimal perturbations but not to finite ones.

In our specific example, what do the metastable points represent, physically? In what sense are they not stable to finite disturbances? Bob has the answer. A superheated liquid will not split into two phases just by the action of Brownian motion, since it is stable to infinitesimal perturbations. It may split into two phases by the phenomenon of nucleation, which is a finite disturbance: some (possibly small) amount of a phase with a significantly different specific volume needs to form, perhaps on a seed. Very good indeed. So we identify three possible types of points. First, globally convex ones, where, no matter what, our system will be homogeneous at equilibrium: portions AB and EF. Second, metastable points, where at local equilibrium the system may exist as a single phase, but at global equilibrium it will be split into two phases: portions BC and DE. Finally, spinodal points, which are not even locally convex, and will thus never be observed in nature, since homogeneous nucleation (due to e.g. Brownian motion) will take care of splitting them into separate phases.

Now let's assume everything is indeed at global equilibrium. Any value of V_0 between V_1 and V_2 will result in two phases, with specific volumes V_1 and V_2, in a relative amount such as to result in the average value V_0. Pressure in the two phases is the same (due to the zeroth law of mechanics, if one forgets about the possibility of capillary effects at curved interfaces), and it takes little algebra to convince oneself that $A_2 - A_1 = -p(V_2 - V_1)$, and hence that free enthalpy is the same in the two equilibrium phases, $G_2 = G_1$. This means that, while one cannot regard p and T as the independent variables in general, as far

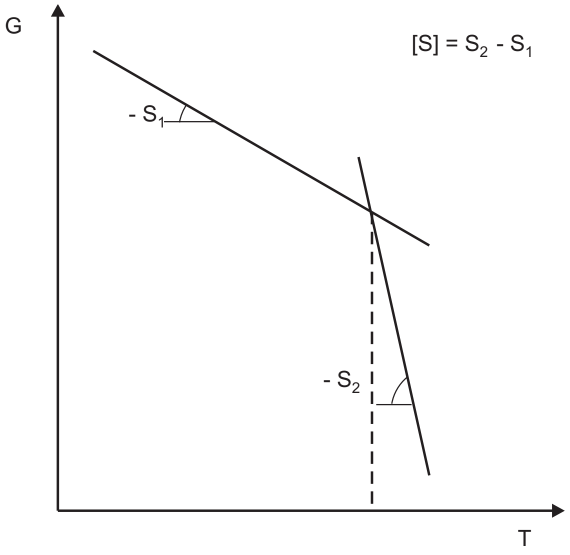

Fig. 2.7.3. Free enthalpy vs. temperature plot for an ordinary phase transition.

as G is concerned one may in fact do so. Hence we consider the plot in Fig. 2.7.3, where G is plotted vs. T at constant p, and only equilibrium values are considered. $(\partial G/\partial T)_P = -S$ (aren't you happy, Ralph?), and hence the curves have negative slope. Sue, entropy has got to be positive even if you have never understood the third law of thermodynamics — we haven't understood it either, but $S > 0$ we do believe. (We rather unsuccessfully try to hide the fact that we are proud that Sue could see that $S > 0$ has, somewhat vaguely and ununderstandably, to do with the third law). The two curves must be in the relative position shown, say the phase which at equilibrium will exist at the higher temperature will also have a larger entropy. Hence we agree to indicate with $[Q]$ the difference between the values of Q in the high and low temperature phase at equilibrium, and, given that at thermal equilibrium $[T] = 0$, we conclude that:

$$[S] > 0 \tag{2.7.7}$$

$$[G] = 0 \tag{2.7.8}$$

$$[A] = -p[V] \tag{2.7.9}$$

$$[H] = T[S] > 0 \tag{2.7.10}$$

$$[E] = T[S] - p[V] \tag{2.7.11}$$

Now we can go back to the Appendix 1.11 (Chapter 1), and obtain the Clausius Clapeyron equation — as long as the system is at equilibrium, G is a unique function of T and p, which is continuous across the phase transition, though its derivatives $-S$ and V are not; thus the argument there applies to the case at hand. Now what is the sign of $\mathrm{d}p/\mathrm{d}T$ at the phase transition? $[S]$ is positive by definition, but $[V]$ may have any sign, and one does not need to look far for a case where it is negative. Liquid water at 0°C has a smaller specific volume than ice does — and fortunately so, otherwise ice would sink in the sea, new ice would form, and we would all be in very serious trouble.

PROBLEMS

2.1 Consider an elevator moving vertically, and let $h(t)$ be its vertical position as a function of time. There are three forces acting on the elevator: the weight W, the pull of the rope P, and the drag on the shaft D. First write a momentum balance establishing a relationship between P, D, W and $h(t)$. Next write an energy balance, assuming q to be zero (the elevator is not being heated). Let $\phi = Wh$ be the potential energy of the elevator. Explain why ϕ does not appear in the equations in Section 2.1.

2.2 An ideal gas is flowing through a tube isothermally. The pressure at the exit is of course less than the one at the inlet. Is work being done on the gas by the tube walls, or the other way around? Is heat being subtracted from the gas to keep it isothermal, or *vice versa*? Is energy being dissipated? Do the answers to the second two questions, when compared with each other, make sense? Is the solution of this problem in any way connected with the analysis of two phase flow given by De Nevers, AIChEJ, 2 (1967) 387?

2.3 Consider steady one-dimensional heat conduction through slabs identified by their thickness L. The modulus of the temperature gradient τ is the same for all such slabs, and so is the cold face temperature T_0. The thermal conductivity is independent of temperature. Determine what condition L should satisfy for $<\Gamma>$ and $<\Omega>$ to be regarded as being at least approximately proportional to each other.

2.4 Although known to be impossible, suppose Eq. (2.3.22) holds true with constant values of μ and μ''. Consider the case where $D_{11} = -\tau$, $D_{22} = D_{33} = \tau/2$ (an uniaxial compression). Show that the second law can be violated for some value of τ, and determine the minimum value of τ for which it is violated.

2.5 Ralph has criticized Bob for his suggestion that the conjugate flux and force could be $(\mathrm{d}w/\mathrm{d}t)^3$ and $(w^* - w)^5$ (see the beginning of Section 2.5), maintaining that the product of those two quantities is not equal to the rate of entropy production, and hence the two quantities cannot be conjugate flux and force. Bob has countered by keeping $(w^* - w)^5$ as the force, and giving the conjugate flux that would satisfy Ralph's requirement. What is Bob's suggestion for the conjugate flux?

2.6 Consider a mixture of two optical isomers. Density is independent of composition, and since the isomerization reaction is equimolar, volume has no effect on the equilibrium composition. However, temperature does influence the value of the equilibrium constant, as well as the value of the time constant β. Some optical property Q (e.g., polarizibility) depends on the relative amount of the two isomers. Experiments are made where Q is measured, with the system subjected to some arbitrary temperature vs. time curve, $T(t)$. Develop a relaxation type theory describing how Q may depend on temperature and on its history.

2.7 Consider the Stefan problem, as discussed in Section 1.7, in its extension to a secondary phase transition. Set up the equations, and discuss the physical interpretation of any paradox you may seem to reach.

SUGGESTED READING

2.1 The book by Astarita (1989) is a good source for the general approach taken in the whole Chapter. A more fundamental and rigorous presentation (and for that reason also a harder one to read) is the book by Truesdell (1984).

2.2 The question of the skew part of the conductivity tensor was discussed by Carlslaw and Jaeger (1959, pp. 43–46) with the objectivity of heat transfer specialists who do not bother about thermodynamics. It is also discussed in Chapter 7 and Appendix 7A of the book by Truesdell (1984). However, by 1984 the Onsager-type of thermodynamics could not be disregarded any more, and the latter exposition is couched in a polemic attitude towards the so called 'thermodynamics of irreversible processes', as typified, as far as heat conduction is concerned, by the work of Casimir (1945). In this section, as well as, in fact, throughout this book, we take the attitude that most of 'thermodynamics of irreversible processes', as usually discussed in the literature, is nonsense. The papers by Jaynes (1980), Wei (1966), Coleman and Truesdell (1960), and Astarita and Sarti (1975) are useful in this regard.

2.3 An elementary exposition of the general thermodynamic theory of momentum transfer, which includes consideration of the theory of viscoelasticity, is given by Astarita (1975). A brief, rigorous, and elegant presentation of the tensor algebra needed in momentum transfer theory is in the Appendix of the book by Coleman, Markowitz and Noll (1966) and in the Appendix of the book by Ottino (1989).

2.4 The fundamental analysis of the thermodynamics of diffusion is in Chapter 5 of the book by Truesdell (1984). The symmetry of the mobility matrix was deduced from the momentum balance by Stefan (1871) for the case of binary mixtures; he then asserted it as a hypothesis for mixtures with more than two components (where it does not follow from the momentum balance). The first experimental tests in multicomponent mixtures were performed by Dunlop and Gosting (1955), who conclude that the asserted symmetry does not hold for at least two of the eleven mixtures tested. The fact that inclusion of the kinetic energy term in the potential for diffusion makes the equations hyperbolic, and hence disposes of the paradox discussed in Section 1.4, is due to Mueller and Villaggio (1976). An elegant and simple geometrical analysis, leading to the inclusion of the gravitational term in the chemical potential, has been presented by Jackson (1989).

2.5 The rational analysis of chemical reactions in mixtures has been developed by Jouguet (1921), De Donder (1931), Wei (1962), Aris (1966), and Bowen (1968a); see also Chapter 6 of the book by Truesdell (1984). A useful paper of introductory character is the one by Coleman and Gurtin, (1967); the approach is then generalized by Bowen (1968b).

2.6 The fundamental equivalence of a description based on internal state variables and one based on relaxation is discussed by Astarita and Sarti (1976) and by Sarti and Astarita (1975).

REFERENCES

Aris, R., Arch. Ratl. Mech. Anal., 29 (1966) 81.

Astarita, G., An Introduction to Nonlinear Continuum Thermodynamics. SpA Ed. di Chimica, Milan, 1975.

Astarita, G., Thermodynamics. An Advanced Textbook for Chemical Engineers. Plenum Press, 1989.

Astarita, G. and G.C. Sarti, Chim. Ind. (Milan), 57 (1975) 680, 749.

Astarita, G. and G.C. Sarti, Polym. Engng. Sci., 16 (1976) 490.

Bowen, R.M., Arch. Ratl. Mech. Anal., 29 (1968a) 114.

Bowen, R.M., J. Chem. Phys., 49 (1968b) 1625.

Carslaw, H.S. and J.C. Jaeger, Conduction of Heat in Solids, 2nd Edn., Oxford University Press, 1959.

Casimir, Revs. Mod. Phys., 17 (1945) 343.

Coleman, B.D., H. Markowitz and W. Noll, Viscometric Flows of Non-Newtonian Fluids. Springer Verlag, Berlin, 1966.

Coleman, B.D. and M.E. Gurtin, J. Chem. Phys., 47 (1967) 597.

Coleman, B.D. and C. Truesdell, J. Chem. Phys., 33 (1960) 28.

De Donder, Th., L'affinite', II Partie, Gauthiers Villars, Paris, 1931.

Dunlop, P.J. and L.J. Gosting, J. Am. Chem. Soc., 77 (1955) 5238.

Jackson, R., Chem. Eng. Sci., 44 (1989) 1871.

Jaynes, E.T., Ann. Revs. Phys. Chem., 31 (1980) 579.

Jouguet, E., J. de l'Ecole Polyt., II Ser., 21 (1921) 61, 181.

Mueller, I. and P. Villaggio, Meccanica, 11 (1976) 191.

Ottino, J.M., The Kinematics of Mixing: Stretching, Chaos, and Transport. Cambridge University Press, Cambridge, 1989.

Sarti, G.C. and G. Astarita, Trans. Soc. Rheol., 19 (1975) 215.

Stefan, J., Sitzungber. k.u.k. Akad. Wiss., 63 (1871) 63.

Truesdell, C., Rational Thermodynamics. Springer Verlag, New York, 1984.

Wei, J., J. Chem. Phys., 36 (1962) 1578.

Wei, J., Ind. Eng. Chem., 58 (1966) 55.

Chapter 3

Dimensional Analysis, Scaling, and Orders of Magnitude

3.1 INTRODUCTION

The title of this chapter may come as a bit of a surprise to the reader. Dimensional analysis — that has to do with Buckingham's theorem and the like, and it is perhaps related to scaleup, but the word scaling doesn't mean the same thing as scaleup, does it? And what have orders of magnitude to do with all this?

Well, let's begin by agreeing on what the terms mean, with perhaps the exception of 'dimensional analysis' which needs no clarification. Dimensional analysis is a very powerful tool for the chemical engineer, but one which should be used with care and intelligence. In its classical formulation, it has to do with the identification of the dimensionless groups which will appear in the solution of any given problem. It is discussed (one way or another) in practically every book on chemical engineering, and yet none of the presentations is entirely convincing. For any given problem, dimensional analysis is applied, and the results turn out to be both useful and correct. However, the student is left with the uncomfortable feeling that for every new problem there is a new twist in the technique of applying dimensional analysis, and one worries that, given a new problem for which there are no guidelines in the textbooks, one would not feel confident about using dimensional analysis. There is, however, a book which is a jewel in this regard, so much so that we don't feel up to improving upon it, in spite of the fact that it was published more than 70 years ago: "Dimensional Analysis" by P.W. Bridgman, Yale University Press, New Haven, 1922. We urge our readers to study this book — it's not a long one and it is easy to read, and once one has reached its end lots of things one had learned by heart but didn't really understand have become crystal clear. So we suppose we have agreed on what dimensional analysis is — the mysterious stuff connected with the Buckingham (or π) theorem which tells us what the relevant dimensionless groups might be.

What about scaling? Suppose one considers a problem the mathematical description of which contains some (dimensional) variable y. Now the values of y may have any order of magnitude, and furthermore their order of magnitude doesn't mean anything, since we may make it as large or as small as we wish by choosing an appropriate system of units. For instance, let's take distance from the centre-line in a cylindrical tube as our y variable. Is a value of $y = 1$ cm large, small, or about average? The fact that it is *one* centimetre is no guideline: had we chosen kilometres as our unit of length the number would be 10^{-5}, which seems rather small; and had we chosen microns it would be 10^4, which seems to be quite large. Now what we do know about y is that it has a maximum value equal to the radius of the tube, let's call Y the latter. Now the new variable $y' = y/Y$ is not only dimensionless, it is known to be of order of unity: dividing by Y is the appropriate *scaling* of the y variable. The examples of stretched coordinates discussed in Chapter 1 are all cases where a rather detailed analysis of the problem at hand is aimed essentially at the determination of the appropriate scaling factor Y.

The discussion about scaling in the preceding paragraph has already established the connection with the concept of order of magnitude. This is possibly the most difficult concept to grasp for an engineering student; and the more computers, pocket calculators and the like become available, the more difficult it becomes to grasp it. In the old days of the slide rule, the student soon learned the hard way that trying to keep track of digits she couldn't read on the slide rule was hopeless. That made it clear that if the first nonzero digit is 1, the third one may have some importance — perhaps — and the second one is likely to be of importance; but if the first nonzero digit is 9, the third one couldn't possibly be read and it therefore had to be assumed irrelevant. Today we get eight digits as an absolute minimum, and an instructor who tries to point out that there really are only three numbers, 0, 1 and ∞, is looked upon as one who is trying to be funny. And yet, given two phenomena, if we learn to compare them in the appropriate way, what possibilities do we foresee? Either the first one can be dismissed as irrelevant (their relative importance is of order zero), or both are relevant (of order 1), or the second one can be neglected (of order ∞). Stokes law for the friction on a sphere is quite good up to $\mathbf{Re} \approx 0.8$, after all, and yet it is based on assuming $\mathbf{Re} = 0$. So what difference does it make whether $\mathbf{Re}$ is, say, 0.2 or 0.6? And yet today's engineering student is likely to come up with the statement that he can use Stokes' law because **Re** = 0.35467032 which is less than unity (it is indeed).

Now the three techniques of dimensional analysis, appropriate scaling, and intelligent estimation of orders of magnitude, if used jointly, constitute an extremely powerful tool for the chemical engineer. Very often more than nine tenths of what one can ever hope to know about a problem can be obtained from this tool, without actually solving the problem; the remaining one tenth requires painstaking algebra and lots of computer time, it adds very little to our understanding of the problem, and if we haven't done the first part right, all that the algebra and the computer will produce will be a lot of nonsense. Of course when nonsense comes out of a computer people have a lot of respect for it, and that's exactly the problem.

Some sort of Murphy's law of course applies here, as it always does. The tool we are discussing is very powerful, and for that very reason it is difficult to use, and it cannot be taught as an algorithm. Algorithms are never very powerful; they only solve problems within the class for which the algorithm has been developed, and one doesn't develop an algorithm unless the corresponding class of problems is so well understood as to be conceptually trivial. Rather than trying to hide this unpleasant fact under a coat which might superficially look like an algorithm, we choose in this chapter the only pedagogical tool available when there is no algorithm — we try to teach by examples.

3.2 FUNDAMENTAL EXAMPLES

3.2.1 Flat plate with suction

Let's discuss an example where the question of using dimensional analysis as a tool for estimating orders of magnitude is hopefully particularly evident. We have a flat plate at zero incidence (located at $y = 0, x > 0$) which is invested by a fluid stream which, at $x = -\infty$, has only one nonzero velocity component, $v_X = U$. It all looks quite familiar, doesn't it? It sounds like the classical boundary layer example which is discussed in all standard textbooks on transport phenomena. Well, we are going to add a little twist to it: at the flat plate, we exert a suction, so that for $y = 0$ the y component of velocity is bound to be $v_Y = -V$, with — to make life easier — V we wish to investigate is the boundary layer.

Bob comes to our rescue here. He remembers that when there is suction the boundary layer thickness will be constant. Well, that can't quite be true as stated, but let's try it out this way: perhaps, as x approaches ∞, the boundary layer thickness δ will approach some constant value δ_∞. Is that possible? And, if yes, what is the value of δ_∞? Let's give it a try. The value of U is unlikely to play any role: after all, should V be zero (no suction), there would be no value of δ_∞. There is no characteristic length — the plate is infinitely long in the x direction. So there are only three parameters: viscosity μ, density Φ, and the suction velocity V. There is only one way to combine these to have the dimensions of a length, say $\mu/\Phi V$. Thus we tentatively conclude that:

$$\delta_\infty = \mu/\Phi V \tag{3.2.1.1}$$

Well, that's nice — Sue likes it a lot — but we aren't quite sure of it, are we? So let's proceed to a bit of algebra, so that Ralph is happy as well. We decide that our estimate is probably right, and thus we define a dimensionless coordinate orthogonal to the plate as:

$$Y = y\Phi V/\mu \tag{3.2.1.2}$$

which is nothing else but a Reynolds number based on distance from the plate and suction velocity. Now if indeed δ is constant along x, so is v_X, say $\partial v_X/\partial x = 0$. But the continuity equation now implies that $\partial v_Y/\partial y = 0$, and the boundary conditions ($v_Y = -V$ at both $Y = 0$ and $Y = \infty$) can be satisfied with $v_Y = -V$ everywhere. So far so good, and let's proceed. Let's say that $v_X = f(y)$. The equation of motion in the x direction becomes:

$$\Phi(-V\mathrm{d}f/\mathrm{d}y) = \mu \mathrm{d}^2 f/\mathrm{d}y^2 \tag{3.2.1.3}$$

which is *exact*: $\partial v_X/\partial x = 0$ according to our hypotheses, $\mathrm{d}P_0/\mathrm{d}x = 0$ as we said before, and $\partial^2 v_X/\partial x^2 = 0$ again according to our hypotheses. Equation (3.2.1.3) is subject to BCs as follows: $y = 0, f = 0; y = \infty, f = 1$. But let's transform to the Y variable; we get:

$$\mathrm{d}f/\mathrm{d}Y + \mathrm{d}^2 f/\mathrm{d}Y^2 = 0 \tag{3.2.1.4}$$

subject to the same BCs. There are no more parameters, everything is of order unity, so our choice of normalization couldn't be wrong, could it? The steady boundary layer thickness we guessed must be right. Indeed, (3.2.1.4) integrates to:

$$f = 1 - \exp(-Y) \qquad (3.2.1.5)$$

and, since we've got an exact solution with a constant boundary layer thickness, we know our initial tentative estimate was right.

Bob knew this all the time, and Sue finds it convincing if somewhat unexpected, but Ralph needs some convincing. Having done with that, we proceed to ask ourselves another question. How far from $x = 0$ do we need to go before δ is approximately δ_∞? Bob doesn't like this question; he knows it is much less than the length of an aeroplane wing. Well, there are things other than aeroplane wings in this world. (One needs to keep the cabin pressurized at 1000 meters, or so do the hostesses tell us, and since we need to suck in air anyhow, why not through the wings, where suction will also help delaying, or perhaps entirely avoiding, detachment of the boundary layer, and thus drag will be decreased — these aeronautical engineers are smart, aren't they?). So let's ask the question anyhow. How do we get an answer? First let's look at dimensional analysis. This time U is sure to play a role — so we need to conclude that the required length (let's call it L) is something like $L = \delta_\infty f(U/V)$. And we expect $f()$ to be much larger than unity, don't we? Well, that's interesting, but not very illuminating. Let's try some of the tricks which Sue likes so much. First of all, at $x = 0$ the momentum flux for Y between 0 and 1 is $\Phi U^2\delta_\infty$. At $x = L$, the momentum flux is less than that — so the difference is of order $\Phi U^2\delta_\infty$. Now this time Ralph is really mad. If something is, say, 100, and something else is known to be less than 100, how can we maintain the difference is of order 100? 99.5 is less than 100, but the difference, 0.5, is certainly not of order 100. Good point, Ralph. The question is: the momentum flux at $x = L$ is not expected to be almost equal to the one at $x = 0$, is it? Why would one expect it be to be 99.5% of it? One would expect it to be perhaps 60% of it, or some such number: then the difference is 40, and 40 and 100 do have the same order of magnitude, do they? Ralph is not convinced, and one needs to ask him to integrate appropriately Eq. (3.2.1.5) in order to convince himself. This however defeats the purpose of the whole exercise, since what one is trying to do is to avoid having to actually solve the problem. Well, let's proceed. The momentum flux difference is what we have estimated. What balances this difference? The shear stress at $y = 0$, τ, between $x = 0$ and $x = L$. So $\tau L = \Phi U^2\delta_\infty$. Therefore:

$$L \approx \Phi U^2 \delta_\infty / \tau \tag{3.2.1.6}$$

You're right, Ralph, τ is not constant along x. Still, it must have some sort of average value, and that's what we are considering here — it's an order of magnitude analysis, so we only need an *estimate* of τ. But at sufficiently large x the shear stress must be of order $\mu U/\delta_\infty$, and isn't that an estimate? Ralph doesn't like it at all, but we'll tentatively accept it and conclude that:

$$L \approx \Phi U \delta_\infty{}^2/\mu = \mu U/\Phi V^2 = \delta_\infty(U/V) \tag{3.2.1.7}$$

Now this estimate of L has the form predicted by dimensional analysis, and it is indeed much larger than δ_∞ if, as assumed, $U >> V$. Ralph, we all know it's not very convincing, but at least it isn't selfcontradictory, is it?

Ralph dislikes this with all his soul, so we try another one. Near $x = 0$, suction can't have much of an effect, so δ will grow the way we know when there is no suction — $\delta^2 = \mu x/\Phi U$. We now substitute δ_∞ for δ and L for x, and, lo and behold, we get the same answer. There is turmoil in the class this time. Sue maintains that we have cheated — our two methods of estimating L are really only one. Ralph agrees with that, and since he didn't believe the first trick, he doesn't believe the second one either. Bob is entirely on our side, since he thinks that estimating L is an exercise in futility anyhow — it's going to be much less than the length of an airfoil, he reiterates. They are, of course, right all three of them, but one of them, Ralph, is unhappy and unconvinced.

So let's try to be more convincing anyhow. If we haven't made anything silly so far, the appropriate length scale in the y direction is $\mu/\Phi V$, and the appropriate one in the x direction is $\mu U/\Phi V^2$. Correspondingly, let's define Y as done before, and X as x/L. Let's see what happens, shall we? We also normalize v_X by defining $u = v_X/U$, and v_Y by defining $w = v_Y/V$, and the whole class agrees with those choices. We now do a bit of algebra to find out that the equations of continuity and motion become:

$$\partial u/\partial X + \partial w/\partial Y = 0 \tag{3.2.1.8}$$

$$u\partial u/\partial X + w\partial u/\partial Y = \partial^2 u/\partial Y^2 \tag{3.2.1.9}$$

to within the usual boundary layer theory approximations. The BCs are:

$$X = 0, u = 1, w = -1 \tag{3.2.1.10}$$

$$Y = 0, u = 0, w = -1 \tag{3.2.1.11}$$

$$Y = \infty, u = 1, w = -1 \tag{3.2.1.12}$$

Now even Ralph has difficulty not being convinced. There are no parameters in the differential equations, the BCs are all in terms of 1's and zero's, so our normalization must be right. We have indeed guessed right both δ_∞ and L, haven't we?

Lessons to be learned

1. In the estimation of orders of magnitude, the cruder the better. Factors of say 5 can safely be neglected; numerical coefficients shouldn't clutter the algebra.
2. Always try to define dimensionless variables in such a way that there are as few parameters in the differential equations as possible, and boundary conditions present only 0's and 1's. If this is obtainable, one knows for sure that all scaling factors have been chosen appropriately.

3.2.2 Free convection

In the classical theory of heat transfer (and of course also of mass transfer) in a flowing fluid the standard hypothesis which is made is that the flow field is the same as would be observed in the absence of an imposed temperature difference. This is usually called the "isothermal" approximation. Its obvious consequence is that the energy balance equation becomes linear (if the effect of frictional heating is neglected), and hence that the heat transfer coefficient h is independent of the applied temperature difference δT. Indeed, one obtains the general result that $\mathbf{Nu} = f(\mathbf{Re},\mathbf{Pr})$, where δT does not appear on the right hand side. Now the "isothermal" assumption is obviously untenable in the free convection case, where the applied temperature difference is the very reason why the fluid is set in motion, so that the isothermal approximation would simply result in the conclusion that the fluid does not move at all. It follows that, in analyzing free convection, one should in principle solve the motion and energy equations simultaneously, an almost entirely hopeless task.

Let's begin by looking at the essential features of the flow. There are three forces acting: buoyancy forces G, viscous forces V, and inertia forces I. From the classical analysis of the Navier–Stokes equations we know that $V \approx \mu U/L^2$

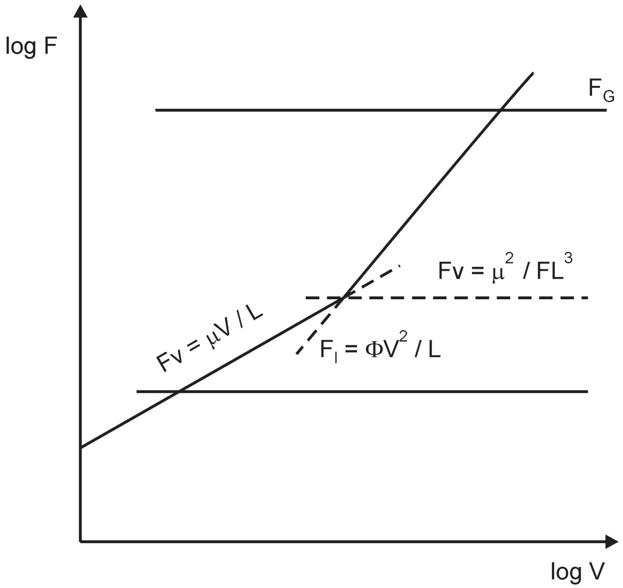

Fig. 3.2.2.1. Plot of the three forces vs. velocity.

and $I \approx \Phi U^2/L$, where U is the appropriate scaling factor for velocity, and L the one for geometrical distances. The problem we have in this case is that the scaling factor U is not known in advance — we expect it to depend on the imposed temperature difference. Before proceeding, we need an estimate of the buoyant forces order of magnitude, and this is easily identified with $\Phi gb\delta T$, where g is gravity and β is the coefficient of cubic expansion. The physics of the problem are as follows: the system will develop a flow field such that the combined effect of viscous and inertia forces balances the buoyant forces.

Now consider Fig. 3.2.2.1, where viscous and inertia forces are plotted vs. the velocity U which the system chooses. The two lines cross each other at a value $U = \mu/\Phi L$ corresponding to a Reynolds number of unity, where they are both given by $\mu^2/\Phi L^3$ (as can be calculated from the expression for either force by imposing the condition $\mathbf{Re} = 1$). Now we remember the lesson that there are really only two numbers, 0 and ∞, to conclude that at U values larger than $\mu/\Phi L$ we might as well forget about viscous forces, and *vice versa*. It follows that, if the buoyant forces are much larger than $\mu^2/\Phi L^3$, they need to be balanced essentially by inertia forces (upper horizontal line). The condition is:

$$\Phi g\beta\delta T >> \mu^2/\Phi L^3\text{: } \mathbf{Gr} = \Phi^2 L^3 g\beta\delta T/\mu^2 >> 1 \qquad (3.2.2.1)$$

The Grashof number **Gr** has emerged naturally from the analysis as the ratio of the imposed buoyancy force and the force arising when viscous and inertia forces are of the same order of magnitude. Since condition (3.2.2.1) implies that buoyancy forces are balanced by inertia forces, $\Phi g\beta\delta T \approx \Phi U^2/L$, say the characteristic velocity is $U \approx \sqrt{(g\beta\delta TL)}$. The Reynolds number is thus obtained, and it is seen to be the square root of the Grashof number:

$$\mathrm{Gr} >> 1,\ \mathbf{Re} \approx \sqrt{Gr} \qquad (3.2.2.2)$$

Now consider the converse situation where Gr

$$\mathbf{Gr} << 1 \qquad (3.2.2.3)$$

So far, we have convinced ourselves that the Grashof number is the dimensionless group which determines the type of flow field which will develop, since its value determines the value of the Reynolds number of the flow. Now let's ask ourselves how the Nusselt number may depend on the Grashof number. We know that the energy balance equation adds a new dimensionless group, the Prandtl number, and hence that we'll have a relationship of the type **Nu** = *f*(**Gr**, **Pr**). One could proceed a bit further than that, but this will be postponed to Section 3.2.6.

Now suppose one had started on this problem by straightforward dimensional analysis. Suppose one begins by listing the parameters on which h may depend, say:

$$h = h(\Phi, \mu, g, \beta, \delta T, L, k, c) \qquad (3.2.2.4)$$

A first important observation is that identification of the appropriate list of parameters is by no means a trivial exercise. With hindsight we know that the list in (3.2.2.4) is the correct one; but if this were a problem we had never heard about, would we have come to the same list? For instance, accustomed as we are to the idea that h does not depend on δT, would we have included the latter in the list? But let's assume we have been smart and we have indeed written down (3.2.2.4). Now we would count parameters, and there are 9 of them. We then would count fundamental units, and the smart ones among us would come up perhaps with four — mass, length, time, and temperature — so that they would conclude there are $9 - 4 = 5$ dimensionless groups, and we know that is wrong, there are only three. (Smart people born in the US would have added g_C in the list of parameters, and force in the list of fundamental dimensions, so that they would again count $10 - 5 = 5$). Bob at this point suggests we should add heat to the list of fundamental dimensions — one always does in heat transfer theory. Ralph does not agree, since heat is energy, and hence it has dimensions ML^2/T^2. Sue expresses the opinion that Bob must be right, but she doesn't know exactly why. Well, the question is that we have been neglecting frictional heating, so that we are dismissing any conversion of mechanical to thermal energy, and hence indeed we can use calories. Had the Griffith (or Brinkman) number $\mu U^2/k\delta T$ not been negligibly small we couldn't have used calories as an independent dimension. Notice that, should we wish to calculate the value of the Griffith number, we would need to express the dimensions of k in terms of the mechanical units of energy.

No general consensus is reached on this point, it looks a bit like a magician's trick. However, for the sake of discussion the point is granted, and so now we have 9 parameters and five dimensions — we still aren't home, since we know the right answer is three, not four dimensionless groups. Indeed, $\beta\delta T$ is by itself dimensionless, and so is $\Phi^2 L^3 g/\mu^2$. So how could we find out that only the product of these two dimensionless groups is relevant? A new magician's trick is performed at this stage. β enters the problem only insofar as it determines the $\delta\Phi$ corresponding to the imposed δT. But $\delta\Phi$ in turn is relevant only because there is gravity — hence β and g can only enter through their product, and we are finally home. Well, would the reader ever have thought of that without knowing the answer beforehand? Furthermore, suppose one actually reduces the list of parameters to eight by including βg instead of β and g separately, and one uses the five fundamental dimensions of mass, length, time, temperature, and heat. One now proceeds to use the constructive part of Buckingham's theorem to construct the three dimensionless groups. How many probabilities are there that one would come up with **Nu**, **Gr** and **Pr**, rather than with some other triplet with no recognizable physical meaning?

Lessons to be learned

1. The fundamental dimensions to be used are not a sacred cow. They can be chosen in different ways depending on what the problem under consideration is, and which approximations one is in principle willing to make (in the case at hand, neglect of frictional heating).

2. It is best to find dimensionless groups by looking at the order of magnitude of different terms in the balance equations which are relevant to the problem, and then taking ratios of the latter. The resulting dimensionless groups will then be identified with ratios of physically significant quantities. This step can often be performed without actually writing down the appropriate differential equations.

3.2.3 Transport between two fluid phases

Consider two fluid phases in contact with each other, the phase of interest being much more viscous than the other one. Let v be the velocity component in the interface plane, and let x be the normal to the interface pointing into the phase of interest (see Fig. 3.2.3.1). Since the tangential stress at the interface must be equal in value and opposite in sign in the two phases, the values of $\partial v/\partial x$ at the interface in the two phases are in inverse proportion to their viscosities, and hence $\partial v/\partial x$ in the more viscous phase can be set approximately to zero. It follows that, near the interface, the more viscous phase moves as a rigid body. Furthermore, except in very special cases, there is no velocity

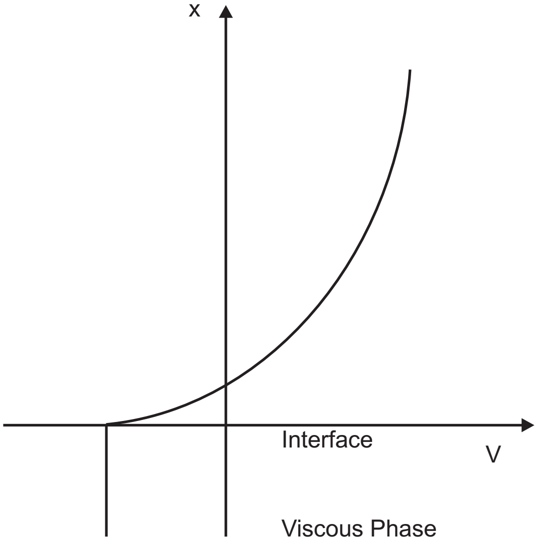

Fig. 3.2.3.1. Tangential velocity profile near the interface.

component in the x direction (the exceptions being cases where the interface is changing its area in time at a rather fast rate), and hence transport orthogonal to the interface in the more viscous phase takes place essentially by unsteady conduction. Let's focus on heat transfer, though the mass transfer case would be entirely analogous. The differential equation for transport is:

$$\partial T/\partial t = \alpha \partial^2 T/\partial x^2 \tag{3.2.3.2}$$

subject to $T(x,0) = T_0$, $T(0,t) = T_1$ (α is the heat diffusivity). Now of course the solution of this problem is very well known, but let's assume we don't know about it, and we therefore proceed to dimensional analysis. We first normalize temperature by defining $\Theta = (T - T_0)/(T_1 - T_0)$ — why? because that way Θ ranges between 0 and 1 — so that the equation becomes:

$$\partial \Theta/\partial t = \alpha \partial^2 \Theta/\partial x^2; \Theta(x,0) = 0, \Theta(0,t) = 0 \tag{3.2.3.3}$$

Now we notice that we do not have a characteristic length, and hence there must be a similarity variable. There is only one possible dimensionless combination of x, t and α, which is $\phi = x/\sqrt{(\alpha t)}$, and hence we convince ourselves that $\Theta = \Theta(\phi)$. Of course one could now try out the algebra to verify that indeed Eq. (3.2.3.3) reduces to an ordinary differential equation for $\Theta(\phi)$. More interestingly, the BC at $x = 0$ transforms to one at $\phi = 0$, which however corresponds also to $t = \infty$. Does that make sense? Indeed it does: we do expect temperature at large distances from the interface to be still T_0.

We may now proceed to consider that we would eventually want to calculate a Nusselt number. From the definitions one concludes that the instantaneous Nusselt number at some time t is $-\Theta'(0)L/\sqrt{(\alpha t)}$, i.e., to within a numerical constant (which has to be of order unity) it is $L/\sqrt{(\alpha t)}$, where L is the characteristic length used in the definition of the Nusselt number. We are, however, interested in the average value of **Nu** over some interval of time t^*, which measures how long an element of the viscous phase stays at the interface. (Some sort of average is of course involved, but that only affects the numerical coefficient, and we do not need to bother with it). Since the average of the inverse square root is again the inverse square root (yes, Ralph, to within a numerical factor — we have decided we don't want to bother with those), one finally concludes that $\mathbf{Nu} \approx L/\sqrt{(\alpha t^*)}$.

Nice and clean. However, we are accustomed to Nusselt numbers which depend on the Reynolds and Prandtl numbers, aren't we? Well, the viscosity of the phase of interest can't play any role, since we have essentially assumed it to be infinitely large. Thus only the product of the Reynolds and Prandtl numbers can appear — that is the only combination which eliminates the viscosity. On the other side, we know (from the result just obtained) that **Nu** is proportional to the square root of the specific heat c, since $\alpha = k/\Phi c$, and hence we conclude that $Nu \approx \sqrt{(\mathbf{RePr})}$. All we need is the proportionality factor, and that can be determined by one single experiment. It is important to notice that the result has been obtained without ever trying to integrate any differential equation, and that the similarity variable ϕ has been deduced by straightforward application of dimensional analysis. From that point on, it has simply been a matter of applying the definitions. We have purposefully discussed this first case in terms of heat transfer theory, so that its being nothing else that the classical penetration model of mass transfer wasn't immediately evident. The results so far are very well known, and we haven't done anything which hadn't in essence been done already in Chapter 1.

But now let's focus our attention to the other phase, where things are not quite as obvious as they have been so far. Here $\partial v/\partial x$ is most certainly not zero; let's say it has some value τ at the interface. Now in the viscous phase we were willing to conclude that it moved as a rigid body because $\partial v/\partial x$ was zero — we were willing to use for v a Taylor series expansion truncated at the second term. By the same token, we should now be willing to approximate the velocity tangential to the interface with τx (x now pointing into the less viscous phase). We have normalized the velocity at the interface itself to be zero. Now let y be the direction of the velocity v. The differential equation is:

$$x\partial\Theta/\partial y = (\alpha/\tau)\partial^2\Theta/\partial x^2 \tag{3.2.3.4}$$

where conduction in the y direction has been neglected since convection is presumably predominant. The BCs are $\Theta(x,0) = 0$, $\Theta(0,y) = 1$, and again the lack of an external length scale leads us to think that there must be a similarity variable. However, this time we seem to be in trouble. We have again two independent variables (x and y) and one parameter (α/τ), but this time there are infinitely many different ways of constructing a dimensionless combination of those. Well, let's try to understand where the problem lies. In the previous case, the two independent variables (x and t) had different dimensions; in this case x and y have the same dimensions, that's the source of our troubles. If only x and y had different dimensions, it would work beautifully, wouldn't it? But what's wrong with thinking as if they did have different dimensions? After all, nobody forbids us to measure distances orthogonal to the interface in miles, and distances in the flow direction in inches. Let's see what happens; let's say that y has dimensions L, and x has dimensions L'. It follows that α/τ has dimensions L'^3/L. Now there is only one possible dimensionless combination, $\phi = x(\alpha y/\tau)^{-1/3}$, and this is thus our similarity variable – $\Theta = \Theta(\phi)$. Of course Ralph doesn't believe it, and the fact that, upon substitution of $\Theta(\phi)$ into Eq. (3.2.3.4) it indeed reduces to an ordinary differential equation leaves him a bit at a loss.

Now the local value of **Nu** corresponding to some value of y is again obtained from the definitions by straightforward algebra; to within a numerical factor (which has to be of order unity) it is $L(\tau/\alpha y)^{1/3}$, and the same is true of the average **Nu** over some finite distance y^* which is $L(\tau/\alpha y^*)^{1/3}$ to within a numerical factor. This means that **Nu** is proportional to the cubic root of **Pr** — as we know to be the case. We cannot deduce how **Nu** depends on **Re**, though. In fact, τ equals the characteristic velocity divided by the thickness of the momentum boundary layer δ, and hence we get a D/δ ratio in our final expression which leaves us at a loss, since we don't know how this ratio may depend on **Re**.

Lessons to be learned

1. We have seen an even more extreme example of the freedom of choice of the fundamental dimensions. In this case, we have used two different dimensions for length. Of course, the freedom of choice has to be used with quite a bit of care.
2. The existence of a similarity variable, or lack thereof, can be found out by judicious dimensional analysis. Of course a similarity solution may exist even when there is no similarity variable.

3.2.4 A problem in oceanography

Here we consider a somewhat unusual problem for a chemical engineer. Suppose that wind blows on the surface of the ocean in a steady way for some time T. Correspondingly, the water is set in motion. We ask ourselves the following questions: do the Coriolis forces due to rotation of the earth have any influence on the flow pattern? Does the answer depend on the value of T? Can we say something about the flow pattern which may result? The nice thing about this problem is that Bob doesn't know anything about its solution; he is an engineer, not an oceanographer, although he suspects that T must be one day, since the Earth makes one revolution a day.

Well, let's begin by considering what the problem looks like in the absence of Coriolis forces. The wind blows steadily, and this can perhaps be modeled by assuming that a constant tangential stress is applied to the free surface of the water. There is no reason (in the absence of Coriolis forces) to suppose that the water is set in motion other than in the direction of the blowing wind, and therefore we are interested in only one component of velocity, say U, and this depends on depth X and time t. Since a tangential stress is imposed at the free surface, we recognize the problem to be the dual one discussed at the end of Section 1.4. However, even this recognition is not needed here. There is no external length scale, so there must be a similarity solution (for the stress, not the velocity) in terms of the variable $X/\sqrt{(\mu t / \Phi)}$. There might be some doubt about μ, since perhaps turbulence is of importance in the ocean; but let's proceed anyhow, with the reservation that perhaps we will need to use some sort of estimate of the turbulent viscosity to obtain any significant result. Viscous (with possibly the turbulent viscosity being relevant) forces and inertia forces have to balance each other, so they are of the same order of magnitude. We therefore only need to estimate one of them, say the viscous one, which is of order $\mu V/L^2$, where V is the appropriate scaling factor for velocity, and L the appropriate length scale.

Well, we don't know V, but we know L to be $\sqrt{(\mu t / \Phi)}$, so the viscous force per unit volume is $\approx \Phi V(t)/t$. Now this is nice because viscosity has cancelled out, and we don't need to bother whether it is the turbulent or the ordinary one. We could have reached the same result by estimating the inertia forces. Since the velocity is horizontal and it changes only along the vertical direction, there is no convective acceleration ($\mathbf{v} \cdot \text{grad}\mathbf{v} = 0$), and hence the inertia force term reduces to $\Phi\partial\mathbf{v}/\partial t$. Now velocity is 0 at time zero, and hence inertia forces are of order $\Phi V(t)/t$: indeed, this result guarantees that the relevant length scale is $\sqrt{(\mu t / \Phi)}$. The order of magnitude of both viscous and inertia forces will depend on time t (unless $V(t)$ happens to be linear in t, which we don't expect to be the case). So in the absence of Coriolis forces we have an estimate of the order of magnitude of the relevant forces as a function of time elapsed since the wind started blowing; notice that the order of magnitude decreases with time, since $V(t)$ is less then linear in time.

Now we need an estimate of the order of magnitude of Coriolis forces. We go to our book on mechanics, and we find that the Coriolis force per unit volume due to an angular velocity Ω is $-2\Phi\Omega x\mathbf{v}$, where $\mathbf{v}$ is the local velocity vector. So the order of magnitude of the Coriolis forces is $\Phi\Omega V(t)$, and this becomes comparable with the other forces at a t value of $1/\Omega$. Since Ω is obviously 1 day^{-1}, we conclude that if the wind blows steadily for at least one day rotation of the earth will start having an influence on the flow pattern in the Ocean. Bob is not impressed by this; he reminds us that he had guessed it right away, since 1 day is the only time scale imposed on the problem, and he is, as usual, right. Ralph does not believe any of it, and we suggest that he should study the appropriate section in the book by Batchelor. Sue is entirely happy, but in expressing her happiness she tries rather unsuccessfully to hide the fact that she doesn't have the slightest idea about what the vector product of two vectors might be, except for remembering that one has to look downwards and then decide to move counterclockwise — or perhaps it was clockwise — to do something she is rather hazy about.

Well, later on we'll need to know *something* about vector products, but right now dimensional analysis is all we need. The Coriolis force per unit volume can only depend on density (it is an inertia force, isn't it?), angular velocity (there would be no Coriolis force without rotation), and local velocity (a bicycle does fall if one is riding it at zero velocity and has forgotten to put a foot on the ground). The only combination of Φ, Ω and V having dimensions of a force per unit volume is $\Phi\Omega V$, so why should we care about whether one has to do whatever one may wish to do clockwise or counterclockwise?

We now ask ourselves a rather more sophisticated question. If Coriolis forces don't play any role, we know that the surface layer where the velocity is significant has a thickness which grows steadily in time — in fact proportionally to $\sqrt{t}$. If Coriolis forces do play a role, could we have some steady state thickness of this boundary layer? Well, let's suppose that's possible, and let L_0 be the thickness of this layer. Now viscous forces are of order $\mu V/L_0^2$, and Coriolis forces of order $\Phi\Omega V$, so they will balance each other at $L_0^2 \approx \mu/\Phi\Omega$. This time we do need to decide whether μ is the ordinary or the turbulent viscosity. With the ordinary viscosity, $\mu/\Phi = 0.01\ \text{cm}^2/\text{sec}$, which yields $L_0 \approx 3$ meters. The turbulent viscosity may be perhaps four orders of magnitude larger, so we get 300 meters instead — that's a significant difference even for Bob. Ralph has some problem this time: what about inertia forces? Couldn't the situation be such that Coriolis forces balance inertia forces, in which case $\Phi V^2/L_0 \approx \Phi\Omega V$ would yield $L_0 \approx V/\Omega$, a quite different result? Ralph needs to be congratulated for having finally acquired the spirit of making order of magnitude estimates. However, there is a flaw in the argument: there is no convective acceleration, and as steady state is reached there is no change of velocity in time, so there are no inertia forces — Coriolis forces must necessarily be balanced by viscous forces. "Ma'm? Didn't you just say that Coriolis forces are inertia forces?". Ralph, it's your good day. They are. But in spite of their being inertia forces, they may exist even without any acceleration do you see?

We finally proceed to ask ourselves what the flow field may look like once this steady state situation has been reached. The velocity is still horizontal everywhere, and we may thus ask ourselves simply what is the angle ϕ it may form with the wind direction at any given depth x — leaving aside the question of how large it might be. The effect of earth's rotation clearly depends on the latitude Θ (how? more about this later). Thus a blind list of variables would be:

$$\phi = \phi(\mu, \Phi, \Omega, x, \Theta) \tag{3.2.4.1}$$

However, we know that the only inertia forces are the Coriolis ones, and hence density and angular velocity only appear through their product $\Phi\Omega$. Therefore one concludes that:

$$\phi = \phi(\Phi\Omega x^2/\mu, \Theta) \tag{3.2.4.2}$$

The angle ϕ thus depends on latitude and on a Reynolds number based on depth and angular velocity. Now at $x = 0$ one concludes that $\phi = \phi(\Theta)$, i.e., *the angle between surface velocity and wind direction depends only on latitude and is independent of angular velocity* — it is the same angle for all planets in the Universe.

Bob is a bit worried about this result, since he sees something funny happening when Ω approaches zero: at $\Omega = 0$ one would expect $\phi = 0$ no matter what, he rightly points out. Well, this is the steady state result, and we have seen before that Coriolis forces become comparable to viscous forces when time approaches $1/\Omega$ — on a planet with a zero rate of rotation, one would simply never reach the steady state (a day is infinitely long on such a planet), and on one with a very low rate of rotation presumably the wind would never blow steadily for long enough to reach steady state, so we have no paradox on our hands.

We now come to our final question, which is specifically addressed to Sue — nobody else in the class is allowed to advance a hypothesis. Does ϕ vanish at the poles, at the equator, or perhaps nowhere at all? Sue struggles a bit and then comes to the conclusion that it must vanish at the poles by reasons of symmetry — if one walks away from the North pole, one walks south no matter in which direction one does move. Wrong, Sue — the direction of the wind provides the needed asymmetry. We do need to know enough about vector products to know that the vector product is a vector orthogonal to the plane formed by the two vectors (though admittedly we still may ignore whether to do that mysterious thing clockwise or not). The wind velocity is always tangential to the surface, and the angular velocity always goes pole to pole; hence at the equator the plane containing them is tangential to the sea surface, and thus the Coriolis force is vertical, and hence it has no effect for our problem. Its effect is maximum at the poles. Sue, you can't quite be *that* cavalier about the algebraic details of the problems you are considering.

Lessons to be learned

1. One can do quite a bit with a problem in an entirely new field by judicious application of the techniques discussed in this Chapter.
2. Sometimes it is useful to analyze the unsteady state form of a problem before analyzing the steady state form of it.

3.2.5 The vibrating sphere

Our next example is one where the ideas from dimensional analysis, proper scaling, and order of magnitude estimates come to bear fruitfully and synergetically — together with some fundamental understanding of the physics involved, and of the thermodynamic implications. This seems a rather grandiose program, and the formulation of our example appears to be exceedingly simple to fulfil our promises. We consider a sphere of diameter D which is vibrating with frequency Ω and amplitude δ, with $\delta << D$. The sphere is immersed in an incompressible fluid having density Φ and viscosity μ, we ask ourselves questions such as what the force F with which the fluid acts on the sphere may be, and how much mechanical power is needed to keep the sphere vibrating.

We begin with straightforward dimensional analysis and we write the following list of parameters:

$$F = (\mu, \Phi, \Omega, D, \delta, t) \qquad (3.2.5.1)$$

since of course the force will vibrate itself and therefore it will depend on time t. Well, we have been rather clever in selecting the right list of parameters, and we are going to be clever in what follows as well. The force F is presumably to appear in something like a drag coefficient, say something like $F/\Phi V^2L^2$. But we don't have a velocity — that doesn't scare us at all, we have a frequency, and $D\Omega$ has the dimensions of a velocity, doesn't it? So we decide that the dimensionless group containing F (we have already decided there will be only one group containing the unknown F) is going to be $F/D^4\Phi\Omega^2$. Our little counting at the beginning has told us there are going to be four dimensionless groups (seven variables and three dimensions), and two of those are easily identified, say Ωt and δ/D — the first one a dimensionless time along the cycle of vibration, the second one a geometrical form factor. From the statement of the problem, we are looking for the case where $\delta/D \approx 0$, so perhaps we may forget about that one. Perhaps. Anyhow, we only need one more group, and it has got to be something like a Reynolds number, and again the lack of a velocity scale doesn't scare us, so we write:

$$F/D^4\Omega^2 = f(\Omega t, \delta/D, D^2\Omega\Phi/\mu) \qquad (3.2.5.2)$$

This has required some cleverness, and we are rather proud of it, but unfortunately it isn't very illuminating, is it? So we try to be even more clever, and we decide that the force F is going to vibrate itself, and that therefore we are really interested in two separate things: the amplitude of the force vibration, say F^*, and possibly its phase angle with the sphere vibration, ϕ. Let's be a bit careful about how we define the phase angle ϕ. If the force F which the sphere exerts on the fluid is always in phase with its velocity, the instantaneous rate of work done by the sphere on the fluid will be always positive, and let's say we call this situation $\phi = 0$. All the work done by the sphere is dissipative when $\phi = 0$. Conversely, if F is out of phase with the velocity (say it is in phase with the acceleration), for half the cycle the sphere will do work on the fluid, and for the other half it will be the other way around — the total work done by the sphere over a cycle thus being zero (no dissipation). We'll call this situation $\phi = \pi/2$. Both F^* and ϕ do not depend on time, so we have got rid of one of the arguments of function $f()$. Furthermore, we are seeking a solution for δ/D

$$F^*/D^4\Phi\Omega^2 = f(0, D^2\Omega\Phi/\mu) \tag{3.2.5.3}$$

with ϕ given by a similar equation. Does anybody in the class have any other clever idea about what to do with the dimensional analysis of this problem? Our three students all have a suggestion. Ralph suggests we write down the Navier–Stokes equations, with the appropriate boundary conditions, develop a numerical code to solve the problem, and be done with it. Well, that's a swell idea, and Ralph is assigned as a homework, due next week, to actually do that. Bob's suggestion is as follows: at very small frequencies, Stokes' law is going to apply, and therefore ϕ is going to be zero, or $\pi/2$, he doesn't remember how we have defined it. At high frequencies, it's like a swing — we'll need very little power to keep it going, just a tiny little push at each swing, so the force will be almost out of phase with what it was before, and ϕ will be $\pi/2$ or 0 as the case may be. Very nice indeed, Bob, that can't be too wrong — but how small a frequency is small enough and all that, what do you think? Bob suggests that any frequency will be small enough in tar pitch, and any one will be large enough in air — he's probably right, but what in, say, water? Sue agrees with Bob entirely —she always does that lately, much to our worry that the agreement may be due to reasons which aren't entirely technical — and she thinks Ralph's suggestion is ridiculous, the computer code will never run in just one week. Well, if she had to develop it, it wouldn't run in one year, but that's not the point: Sue's suggestion is to make an order of magnitude analysis of the Navier–Stokes equations, and she feels confident we will pull the rabbit out of the hat as we have always done lately.

Since Ralph's suggestion is to do something which we have no desire of doing, and Bob's suggestion is to do nothing, we'll follow Sue's advice. There are four terms to deal with: the Eulerian inertia $\Phi\partial\mathbf{v}/\partial t$, the convected inertia $\Phi\mathbf{v}\cdot\text{grad}\mathbf{v}$, the pressure gradient, and the viscous force term. Now our velocity is best scaled with $\delta\Omega$, isn't it? That must be the right order of magnitude of the velocity. The pressure gradient term can't be scaled — it will adjust itself to whatever the other terms demand. But what we do know is that the force will be of the same order of magnitude as the form drag we would calculate, and hence that $\text{grad}p \approx F/D^3$. Time is obviously to be scaled as Ωt, even Ralph readily agrees to that, but we have some doubt as to how gradients, laplacians and all those paraphernalia have to be scaled — what is the appropriate length scale, δ or D? Well, we don't really know so far, so let's call it L, remembering that L is at most D and at least δ. So we estimate the following orders of magnitude:

Eulerian inertia, $\Phi\delta\Omega^2$
Convected inertia, $\Phi\delta^2\Omega^2/L$
Viscous forces, $\mu\delta\Omega/L^2$
Pressure gradient, F^*/D^3

Now let's suppose that $\Omega << \mu/\Phi D^2$ hence our estimate is $F^* \approx \mu D\Omega\delta$. Even if we hadn't remembered the expression for Stokes law (and Sue in fact didn't quite remember it), we would have argued that grad$p \approx \mu V/D^2$ (D is known to be the appropriate length scale if Stokes law applies), and this would have produced the same result. So we obtain our first conclusion:

1. If $\Omega << \mu/\Phi D^2$, then $F^* \approx \mu D\Omega\delta$, and $\phi = 0$. The force is almost entirely dissipative. This estimate holds up to $\Omega \approx \Omega_1 = \mu/\Phi D^2$, where $F^* \approx \mu^2\delta/\Phi D = F^*_1$.

Now let's jump to the other extreme, let's say that $\mu/\Phi\delta^2 << \Omega$. Now, no matter what L may be, the viscous forces are negligible. and the convected inertia is either of the same order of the Eulerian one or negligible with respect to it. Hence the Navier–Stokes equations reduce approximately to $\Phi\partial\mathbf{v}/\partial t = -$grad$p$, and since $F^* \approx D^3|$grad$p|$ we get an estimate of $F^* \approx \Phi\delta\Omega^2D^3$. Notice that this time grad$p$ has the opposite sign — for half a cycle the fluid pushes the sphere, and for the other half the sphere pushes the fluid. This is not surprising, because if viscous forces are negligible there is very little dissipation, and hence F is almost entirely conservative — not quite entirely; as Bob said, we need just a tiny little push at each swing. So we get our second conclusion:

2. If $\Omega >> \mu/\Phi\delta^2$, then $F^* \approx \Phi\delta\Omega D^3$, and $\phi = \pi/2$. The force is almost exactly conservative. This estimate holds down to $\Omega \approx \Omega_2 = \mu/\Phi\delta^2$, where $F^* \approx \mu D^3/\Phi^2\delta^3 = F^*_2$.

Now $F^*_2/F^*_1 \approx (D/\delta)^4$, while $\Omega_2/\Omega_1 \approx (D/\delta)^2$. It follows that the situation, as far as F^* is concerned, is as sketched in Fig. 3.2.5.1. There is a very smooth transition between the slope 1 line holding for $\Omega < \Omega_1$ and the slope 2 line holding at $\Omega > \Omega_2$. The two lines cross each other at $\Omega = \Omega_1$, which suggests that the true critical frequency is Ω_1, and hence that the appropriate length scale for the viscous forces is indeed D. The transition has got to look something like the dashed line in the figure. As far as ϕ is concerned, it is ≈ 0 at $\Omega < \Omega_1$, and it is $\approx \pi/2$ at $\Omega > \Omega_1$.

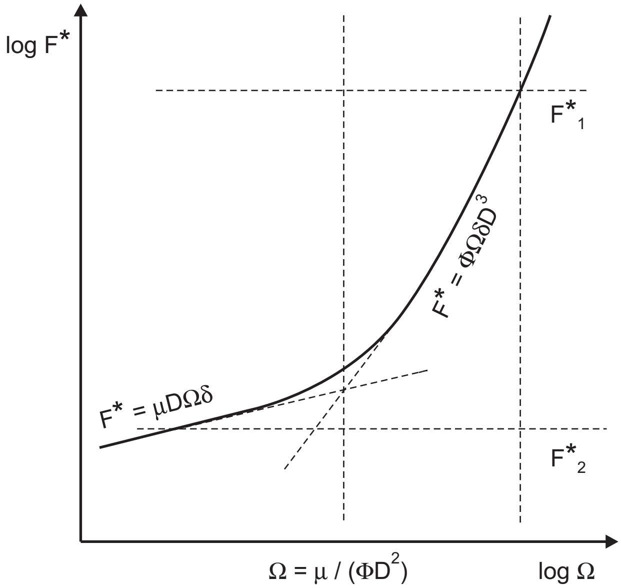

Fig. 3.2.5.1. Plot of F^* vs. frequency.

Today is one of Sue's best days. She comes up with the perfect question, which runs something like this. In Section 1.10, we discussed an oscillating flow field, and we concluded that the behaviour was reminiscent of that of a nondissipative solid at high frequencies, while it was reminiscent of viscous fluids at low frequencies. We then proceeded to infer that the effect was due to the viscoelasticity of the fluid — we sought the explanation in the form of the constitutive equation. This time we have discussed a similar problem, with dissipation becoming negligibly small as the frequency increases, and yet we don't ascribe the effect to anything funny in the constitutive equation, it comes out directly from the Navier–Stokes equations. Can't we make up our minds, that's the main thrust of Sue's point. When dissipation becomes negligible at high frequencies, is it to be attributed to viscous forces becoming negligible as compared to inertia forces, or to some viscoelasticity?

There are two ways of becoming confused about something; one is to know very little about the subject, and the other one is to know too much about it. Now this particular confusion is one which generally arises in the second mode, and Sue must be complimented for getting confused on this particular point in the first mode — there is no chance for her ever getting into the second mode. The confusion is related to the fact that the effects of elasticity and of inertia are often of a similar nature in fluid flow. In Section 1.10, we considered the case where we had satisfied ourselves that, even at the highest frequency of the experiment, inertia forces were still negligible; hence the need to find the explanation in the constitutive equation, i.e., to attribute the effect to the elasticity of the fluid. The phenomenon considered in this subsection is one where inertia determines the behaviour at high frequency.

Lessons to be learned

1. Dimensional analysis, scaling, and order of magnitude analysis are intertwined with each other.
2. Some thermodynamic thinking about dissipation is very often useful.
3. The same type of macroscopic behaviour may be due to quite different inherent mechanisms in different situations.

3.2.6 Two examples of scaling

In this subsection, we will discuss the scaling of free and forced convection, presenting a technique which is on occasion extremely useful, particularly when there are several equations to be considered simultaneously. We begin with a problem of free convection, where some solid body immersed in a fluid at temperature T_0 is kept at temperature $T_0 + \delta T$. We will assume that the boundary layer which develops on the body does not go turbulent — that assumption obviously sets an upper bound on the value of the Grashof number. What we wish to establish is, possibly, the form of the relationship **Nu** = *f*(**Gr**, **Pr**), as well as some more fundamental understanding than was acquired in Section (3.2.2).

The scaling of temperature does not present any problem, since quite obviously $\Theta = (T - T_0)/\delta T$ is going to be of order unity. Let starred quantities be dimensional; x^* is a coordinate moving upwards from the lowest point of the body considered, and u^* is the tangential velocity in the x^* direction. The coordinate orthogonal to the body surface is y^*, and the corresponding velocity component is v^*. D is the diffusivity of momentum, $D = \mu/\Phi$, and α is the diffusivity of heat, $\alpha = \mathrm{k}/\Phi c$, so that **Pr** $= D/\alpha$. The dimensional equations are as follows.

Momentum in the x^* direction:

$$u^*\partial u^*/\partial x^* + v^*\partial u^*/\partial y^* = g\beta\delta T\Theta\phi + D\partial^2 u^*/\partial y^{*2} \tag{3.2.6.1}$$

where ϕ is a geometrical parameter describing the local orientation of the body surface; ϕ is expected to be of order unity.

Mass:

$$\partial u^*/\partial x^* + \partial v^*/\partial y^* = 0 \tag{3.2.6.2}$$

Energy:

$$u^*\partial\Theta/\partial x^* + v^*\partial\Theta/\partial y^* = \alpha\partial^2\Theta/\partial y^{*2} \tag{3.2.6.3}$$

Now there is no doubt that x^* should be scaled with respect to the vertical dimension of the body, L, say $x = x^*/L$. We therefore need to establish three scaling factors: U, V and Y, appropriate to the scaling of $u^* = Uu, v^* = Vv, y^* = Yy$. The scaling factors have to be such that the dimensionless variables are of order unity.

The continuity equation requires $\partial u^*/\partial x^*$ and $\partial v^*/\partial y^*$ to be of the same order of magnitude, and that fact sets a first constraint, say:

$$UY = VL \tag{3.2.6.4}$$

which guarantees that the dimensionless continuity equation exhibits no parameters:

$$\partial u/\partial x + \partial v/\partial y = 0 \tag{3.2.6.5}$$

Now let's turn attention to the energy equation. If (3.2.6.4) is satisfied, it reduces to:

$$(U/L)(u\partial\Theta/\partial x + v\partial\Theta/\partial y) = (\alpha/Y^2)\partial^2\Theta/\partial y^2 \tag{3.2.6.6}$$

Now we have to be careful. Equation (3.2.6.6) could conceivably be satisfied with the two terms on the LHS of equal order of magnitude and opposite sign, and then the RHS would have a much smaller order of magnitude. This would however imply that conduction (the RHS) is negligible, and this does not make sense. Hence we conclude that the RHS and at least one of the terms on the LHS must have the same order of magnitude, and hence we obtain one more constraint:

$$Y^2 = \alpha L/U \tag{3.2.6.7}$$

which reduces the dimensionless energy equation to the following parameter free form:

$$u\partial\Theta/\partial x + v\partial\Theta/\partial y = \partial^2\Theta/\partial y^2 \tag{3.2.6.8}$$

Having established two constraints on the scaling factors U, V and Y, we have only one degree of freedom left, and we now move to the balance of momentum, which, if (3.2.6.4) and (3.2.6.7) are satisfied, reduces to:

$$(U^2/L)(u\partial u/\partial x + v\partial u/\partial y) = \mathrm{g}\beta\delta T\phi\Theta + (U^2\mathbf{Pr}/L)\partial^2 u/\partial y^2 \tag{3.2.6.9}$$

Now there is no doubt that viscous forces must play a role near the solid surface, and we therefore multiply the whole equation by $L/U^2\mathbf{Pr}$ so as to make unity the coefficient of the viscous forces:

$$(1/\mathbf{Pr})(u\partial u/\partial x + v\partial u/\partial y) = (Lg\beta\delta T/U^2\mathbf{Pr})\phi\Theta + \partial^2 u/\partial y^2 \tag{3.2.6.10}$$

We finally observe that buoyant forces have to be significant, and in fact near the solid surface they have to be of the same order of magnitude as the viscous forces; it follows that the coefficient of $\phi\Theta$ must be unity, which sets the value of the scaling factor U:

$$U^2 = Lg\beta\delta T/\mathbf{Pr} \tag{3.2.6.11}$$

Now all our scaling factors have been determined, say:

$$V^2 = Lg\beta\delta T\,\mathbf{Pr}^{-3/2}\mathbf{Gr}^{-1/2} \tag{3.2.6.12}$$

$$Y = L\mathbf{Pr}^{-1/4}\mathbf{Gr}^{-1/4} \tag{3.2.6.13}$$

Furthermore, no parameters appear in any of the equations or the boundary conditions, except for the $1/\mathbf{Pr}$ coefficient of the inertia forces in the equation of motion. We can therefore envisage two cases of interest. First, consider gases: the Prandtl number is of order unity, and so no term drops out of our equations. What that means, physically, is that the thickness of the "thermal boundary layer", Y (the distance from the solid surface over which Θ is appreciably different from zero), is also the distance over which inertia terms become significant — it is equal to the thickness of the momentum boundary layer. Notice that in this case (3.2.6.13) yields $Y/L = \mathbf{Gr}^{-1/4}$, and recalling that $\mathbf{Gr} = \mathbf{Re}^2$ we recover the classical result that $Y/L = 1/\sqrt{\mathrm{Re}}$. The other case of interest is that of liquids, for which $\mathbf{Pr} >> 1$. In this case the inertia terms drop out — the thermal boundary layer is much thinner than the momentum boundary layer, and its thickness Y is small enough that inertia forces are negligible when Θ is significantly different from unity.

For both cases considered above, **Nu** would be calculated as $-L\partial\Theta/\partial y^*_0$. In the case where $\mathbf{Pr} \approx 1$ (gases), this yields $\mathbf{Nu} = K\mathbf{Gr}^{1/4}f(\mathbf{Pr})$. We don't need to worry too much what the function $f()$ may be, since it is unity when $\mathbf{Pr} = 1$ and it can't be much different from unity in any gas; furthermore, the constant K is expected to be of order unity. In the case where $\mathbf{Pr} >> 1$, the inertia terms drop out, and one obtains $\mathbf{Nu} = K(\mathbf{GrPr})^{1/4}$, with again K expected to be of order unity. Notice that the powers 1/4 come out simply from our scaling of Y, see Eq. (3.2.6.13).

We now move to our second problem, forced convection heat transfer at a submerged body on which the boundary layer does not go turbulent. The only difference in the equations is that, instead of having the forcing term of the buoyant forces $\Phi g\beta\delta T\phi\Theta$, we have a term which represents the pressure gradient along the x^* direction which is forced by the external velocity field. Say $U_E(x^*)$ is the slip velocity at the wall in the x^* direction (which is now the external flow direction) as calculated from the Euler equations of motion. Given an external velocity scale U, U_E can be written as $U\phi(x)$, with $\phi(x)$ of order unity, and its functional form being determined by the shape of the object. Thus in this case (3.2.6.4) and (3.2.6.7) hold again, with the same meaning as before, and the external velocity scale is fixed at U. Thus the dimensionless equation of motion becomes:

$$u\partial u/\partial x + v\partial u/\partial y = \phi \mathrm{d}\phi/\mathrm{d}x + \mathbf{Pr}\partial^2 u/\partial y^2 \tag{3.2.6.14}$$

The scaling factors are $V = \sqrt{(\alpha U / L)}$, $Y = \sqrt{(\alpha L / U)}$. For the case of gases, one obtains $\mathbf{Nu} = K\sqrt{(RePr)f(Pr)}$, and again we don't need to worry too much about the Prandtl number — the Nusselt number is proportional to the square root of the Reynolds number. For the case of liquids, (3.2.6.14) again tells us that the velocity profile is essentially linear within the thermal boundary layer (viscous forces predominate over inertia forces), but the Prandtl number does not drop out of the equations this time, because of the $\phi d\phi/dx$ term. One again gets the result that the Nusselt number is proportional to the square root of the Reynolds number, and, based on the argument in Section 3.2.3, presumably to the cubic root of the Prandtl number.

3.2.7 Too much of a good thing

We have always tried to reduce the parameters to their minimum possible number, to have the differential equations and boundary conditions all in terms of zeros and 1's, and so on; sometimes we have been successful, sometimes a dimensionless group or maybe two were left over — but all our cunning was in eliminating as many parameters as we possibly could. Well, sometimes there is too much of a good thing, and in this section we give an example of that.

We have got a steady point source of some obnoxious chemical, of strength S (moles/sec). There is no wind, it's all steady diffusion in an infinite body of fluid — what is the steady concentration distribution? It doesn't take much to find out that, with r distance from the source, it is:

$$c = S/4\pi Dr \tag{3.2.7.1}$$

Fine. Now we go to the other extreme: we have a very strong steady wind, with velocity u in the z direction. This time we use cylindrical coordinates, with $\Theta = 0$ the downwind direction, and y the distance from the downwind axis, and we decide, as usual, that since the wind is strong we may as well neglect diffusion in the downwind direction z. After some more cumbersome algebra than in the first case, we find that the solution is:

$$c = (S/4\pi Dz)\exp(-uy^2/4Dz) \tag{3.2.7.2}$$

Now we consider the case of a wind which is not quite so strong. Bob, it's not academic. It's the confined leak problem: we have the leak in a big lab, the wind is certainly not strong, but air conditioning or whatever may create a gentle breeze. So what do we do? Has anybody a suggestion?

Sue has one. "We should find a dimensionless group, say Q, containing the wind speed u, say in the numerator. When Q is large, we expect (3.2.7.2) to hold, when it is small, we expect (3.2.7.1) to hold". Very well indeed. Let's proceed along these lines. What coordinate system are we going to use? "It doesn't really matter, sir — spherical or cylindrical". OK, Ralph, let's choose the spherical one, with Θ the longitude from the downwind direction, and the latitude being the symmetry coordinate. We have only one length scale available, D/u, and so $R = ru/D$. We don't have any immediate concentration scale available, right? Concentration is going to be ∞ right at the source and 0 far away from it, so ...? Bob has a surge of mathematical awareness. "Ma'm, let's make the dimensionless concentration $\tau = 4\pi D^2 c/Su$, so that Eq. (3.2.7.1) reduces to $\tau = 1/R$". Very good indeed. Extremely good — (3.2.7.1) does not contain u, but both τ and R do, and so u cancels out in $\tau = 1/R$.

We now need to write the convective term in the downwind direction in spherical coordinates, and Sue knows how to do that: look it up in Bird, Stewart and Lightfoot. It turns out to include nasty trigonometric terms, but what can we do? The equation turns out to be:

$$\mathrm{Lapl}\tau = \cos\Theta\,\partial\tau/\partial R - \sin\Theta\,\partial\tau/\partial\Theta \qquad (3.2.7.3)$$

Never mind the nastiness of the equation — the problem is, we haven't got a dimensionless group Q, have we? This is too much of a good thing: we wanted a dimensionless group to come out of the equations, and it hasn't. What are we going to do?

Ralph has a suggestion. "Solve the equation numerically". Good. What are the boundary conditions? τ goes to zero as R goes to ∞; $\partial\tau/\partial\Theta$ is zero at $\Theta = 0$ and at $\Theta = \pi/2$; and the fourth one? It takes some time, but Sue comes up with the answer: "Given any closed surface encompassing the source, whatever comes out of it per unit time must be S". Very good Sue — Ralph, do you like that as the fourth boundary condition to be satisfied by your numerical code? Much to his annoyance, Ralph has to admit that (a), Sue is right, and (b), the condition is very nasty indeed.

Bob has kept quiet for some time, but now he comes to life. "We only need to worry about the EPA recommendations — there will be some concentration value c_0 which cannot be exceeded. So we do have a concentration scale, and we'll get a dimensionless group of the Q type as Sue has suggested". Ralph hates this. "The solution of the problem cannot depend on what value of c_0 the EPA decides to choose. Bob, you are being silly". Ralph, it's always dangerous to regard good engineering sense as silly. Let's just give it a try, shall we? Let's define Γ as c/c_0 and let's see what happens.

We decide to switch to cylindrical coordinates this time, because Sue is scared of sines and cosines, so $Y = yu/D$, and we end up with:

$$\text{Lapl}\Gamma = \text{Q}\partial\Gamma/\partial Y \tag{3.2.7.4}$$

where $Q = Su/4\pi D^2c_0$ — what Bob predicted would happen has indeed happened. "Sheer luck, Ma'm. Q has no physical meaning". Well, it may not have meaning, but it does tell us that, when Q approaches zero, the equation is $\text{Lapl}\tau = 0$, and that has (3.2.7.1) as the solution, right? So that if c_0 is large (3.2.7.1) holds — or, in other words, it must hold near the origin, right? It takes quite some time to convince the class that it does indeed mean that. We'll leave Q to its ill fate — we will simply use the fact that we have discovered: the nasty boundary condition is now substituted by (in spherical coordinates):

$$R \to 0,\ \tau \to 1/R \tag{3.2.7.5}$$

Isn't that better for your computer code, Ralph? This problem admits a surprisingly simple exact solution:

$$\tau = (1/R)\exp[R(\cos\Theta - 1)/2] \tag{3.2.7.6}$$

Exactly downwind ($\cos\Theta = 1$), the wind has no effect whatsoever on the concentration distribution!

3.3 SOME CONCEPTUAL ISSUES

3.3.1 Peculiarity of the power law

The problem we consider here is the following one. Suppose one has decided to write down a constitutive equation for some variable y, and one has decided that its value is presumably uniquely determined by the value of some other variable x, say:

$$y = f(x) \tag{3.3.1.1}$$

The case of interest is the one where both y and x are dimensional, and their dimensions are different from each other. Should one wish to consider a concrete physical example, one may focus attention on a chemical reaction, for which y would be the rate of reaction (with dimensions of moles per unit volume and time), and x would be the concentration of one particular species — the concentration of any other one being related to it by stoichiometrical constraints and by an initial composition. Another possible example would be a purely viscous non-Newtonian fluid, where y would be the stress tensor, and x the rate of deformation tensor. We ask ourselves the following question: if (3.3.1.1) is one of the equations which enter any problem we may wish to consider, how many dimensional parameters does it contribute to our starting list of classical dimensional analysis?

We begin by noticing that we can identify the dimensionality of an operator such as $f()$. What do we mean by saying that some physical quantity y is dimensional? We mean that the number which measures its value depends on the set of units we have chosen. Well, we can extend this concept: an operator $f()$ is dimensional if its mathematical form depends on the units chosen to measure both its argument x and its value y. Now quite obviously if at least one of either x or y is dimensional, $f()$ is dimensional. Can we transform Eq. (3.3.1.1) in such a way that a dimensionless operator $F()$ — dimensionless in the sense that its mathematical form is independent of the set of units used for either x or y — appears in it? Well, the answer is affirmative: all that we need to do is to introduce two dimensional parameters Y and X, having the same dimensions of y and x, respectively, and then write:

$$y/Y = F(x/X) \tag{3.3.1.2}$$

The operator $F()$ is now guaranteed to be dimensionless. We have of course unlimited latitude of choice for the parameters X and Y, but the result to be stressed at this point is the following one: in order to have a dimensionless operator, we need two dimensional parameters. Thus in general (3.3.1.1). will introduce two dimensional parameters. (Should there be more than one independent variable, say $x_1, x_2, \dots x_N$, we would in general need $N+1$ dimensional parameters).

The question which arises now is the following one: are there exceptions? In other words, are there special mathematical forms of $f()$ which will result in less than two — say only one — dimensional parameters? Well, the answer is obviously affirmative, since we have the example of a power law:

$$y = Kxn \tag{3.3.1.3}$$

Equation (3.3.1.3) produces, when it is cast in the form of (3.3.1.2):

$$y = (KY/Xn)xn \tag{3.3.1.4}$$

which contains only one dimensional parameter, the group KY/Xn. Notice that the linear case ($n = 1$) is a special case of the power law form.

Now it takes a bit of algebra — not in any significant way different from the one leading to the classical formulation of Buckingham's theorem —to convince oneself that the power law form of constitutive equation is the only one for which a single dimensional parameter is all one needs. The result is related to the fact that the only algebraic operations which are legitimate with dimensions are multiplication and raising to a power, and therefore only the power law transforms to a legitimate operation in the algebra of dimensions. Be that as it may, the fact remains that the power law is dimensionally peculiar. In order to convince oneself of how peculiar it is, consider the deceivingly simple case where (3.3.1.1) can be expressed as:

$$y = a + bx \tag{3.3.1.5}$$

We have two dimensional parameters: a (having units equal to those of y, $a[=]y$), and b ($b[=]y/x$). No amount of juggling of algebra can collapse those two parameters into one.

Well, one might here reason that perhaps the linear case has some special status — in the sense of a Taylor series expansion truncated at the first order in x, after appropriate normalization to guarantee that $f(0) = 0$ — but the power law has not, and that therefore we really have only two cases: either (3.3.1.1) is linear, and then we have only one parameter; or it is not, and then we have two. That is Sue's argument, and there's no doubt it has considerable merit. It is hand-waving as all Sue's arguments are, but then that's why they have merit. Ralph, however, doesn't buy it, recalling that we have been using chemical kinetics as an example, and damn it, chemical reactions can in general be described as endowed with n-order kinetics, don't they? Bob utterly dislikes having to agree with Ralph, but this time he can't avoid it — nth order kinetics work most of the time, so the power law has got to have some special status.

Let's discuss this question of nth order kinetics. We have established that they are dimensionally peculiar, since any other nonlinear constitutive equation would give rise to two dimensional parameters rather than one. Let's examine in some detail, for a specific problem, what the consequences of this dimensional peculiarity may be. In particular, let's consider the isothermal effectiveness factor in a porous catalyst having a flat plate geometry with half thickness L. For n-order kinetics, the differential equation is:

$$D\mathrm{d}^2c/\mathrm{d}x^2 = bc^n \tag{3.3.1.6}$$

where, for reasons which will soon become clear, we use the somewhat unusual symbol b for the "n-order kinetic constant". Equation (3.3.1.6) is subject to $c = c_S$ at $x = 0$, $\mathrm{d}c/\mathrm{d}x = 0$ at $x = L$, and hence it is natural to scale distances and concentrations as $y = x/L$, $\tau = c/cS$. This reduces the differential equation to:

$$\mathrm{d}^2\tau/\mathrm{d}y^2 = \phi_S{}^2\tau^n \tag{3.3.1.7}$$

subject to $\tau = 1$ at $y = 0$, $\mathrm{d}\tau/\mathrm{d}y = 0$ at $y = 1$. The Thiele modulus ϕ_S is defined by $\phi_S{}^2 = bL^2/Dc_S{}^{1-n}$, i.e., it is recognized as the usual Thiele modulus for linear kinetics, with the pseudo-first-order kinetic constant being evaluated at the surface conditions, $k_{EQ} = bc_S{}^{1-n}$. Now the effectiveness factor E is given by:

$$E = -D(\mathrm{d}c/\mathrm{d}x)0/bc_S{}^nL = (\mathrm{d}\tau/\mathrm{d}y)_0/\phi_S{}^2 \tag{3.3.1.8}$$

Since the solution of (3.3.1.7) is obviously of the form $\tau = \tau(y,\phi_S)$, (3.3.1.8) implies that the effectiveness factor is a unique function of the Thiele modulus calculated at the surface conditions, $E = E(\phi_S)$.

Now suppose that the actual kinetics are of the Langmuir isotherm type, say:

$$D\mathrm{d}^2c/\mathrm{d}x^2 = kc/(1 + Kc) \tag{3.3.1.9}$$

This time it is natural to define $\tau = Kc$. The equivalent kinetic constant at the surface conditions is now based on $k_{EQ} = kc_S/(1 + Kc_S)$, and the second part of (3.3.1.8) still holds, $E = (\mathrm{d}\tau/\mathrm{d}y)_0/\phi_S{}^2$. The differential equation is:

$$\mathrm{d}^2\tau/\mathrm{d}y^2 = \phi_S{}^2\tau(1 + \tau_S)/\tau_S(1 + \tau) \tag{3.3.1.10}$$

The solution of (3.3.1.10) is of the type $\tau = \tau(y,\phi_S,\tau_S)$, and so one cannot conclude that E is a unique function of the Thiele modulus calculated at the surface conditions, say in this case $E = E(\phi_S,\tau_S)$. The result is not surprising: we have one more dimensional parameter, and hence we have one more dimensionless group than we had in the former case.

The discussion above shows that the dimensional peculiarity of the power law form is not a trivial one, since it leads to conclusions which would not hold for any other nonlinear form. There are other examples in the literature. For a power law non-Newtonian fluid, one can define a modified Reynolds number (analogous to ϕ_S in the example above), and conclude, for instance, that the ratio of the entry length in a pipe to its diameter is a unique function of this modified Reynolds number; the same conclusion is not reached for any other form of the constitutive equation for the stress. Also, for any fluid for which the stress is a unique, though not necessarily linear, function of the rate of deformation, a variational principle for creeping flow of the type of the Helmholtz Korteweg one can be constructed. In the special case of power law fluids (including Newtonian ones) the quantity to be minimized is proportional to the rate of entropy production, so that the variational principle can be interpreted as a principle of minimum dissipation; for any other nonlinear form of the constitutive equation, the quantity to be minimized is not the total rate of energy dissipation. This peculiarity of the power law equation is again related to its dimensional peculiarity.

Now we come to the question of why power law equations have this habit of propping up here, there, and everywhere. One looks at where they come from, and one reaches a somewhat unexpected conclusion: data are plotted on logarithmic paper, as engineers are wont to do for a variety of reasons, and straight lines are fit to them — again, engineers like straight lines. Of course the equation of a straight line in logarithmic paper is a power law, and that's where the origin lies, in most of the cases. Now those same data could have been plotted on some other type of paper (semilogarithmic, hyperbolic, whatever) and perhaps a straight line would have fit them equally well. But now one would have obtained a nonlinear constitutive equation of type other than power law to represent those very same data. One thus concludes that one should be very careful about conclusions drawn from dimensional analysis when power law constitutive equations are used (unless $n = 1$): such conclusions are peculiar, and different conclusions would be reached with any other type of data-fitting constitutive equation.

3.3.2 Time and length scales

In many problems in chemical engineering, a very powerful tool is the comparison of time scales of different phenomena. For instance, all approximations of the "quasi steady state" type are obviously based on the consideration that the time scale of unsteadiness is in some sense very large as compared to the time scale of the phenomenon one is considering. In this section, we discuss some conceptual issues associated with time scales, and we also show that often a comparison of time scales is equivalent to a comparison of length scales.

First consider classical transport phenomena. One always has a diffusivity D, and we have seen repeatedly that the appropriate time scale of a transport phenomenon is L^2/D, where L is some external length scale. Now consider any problem where a transport phenomenon takes place simultaneously with some other phenomenon which is characterized by an intrinsic time scale Θ. A first order reaction is a good example: the kinetic constant k is the inverse of the intrinsic time scale of the reaction, $\Theta = 1/k$. The ratio of the two time scales is $L^2/D\Theta$, or, in the case of a first order reaction, it is the square of the Thiele modulus ϕ. Now of course ϕ is usually regarded as a dimensionless measure of the catalyst's linear dimension L, and in this sense one concludes that the intrinsic length scale one is comparing L with is $\sqrt{(D\Theta)}$. Given any phenomenon where a diffusivity D and an intrinsic time scale Θ play simultaneously a role, the corresponding intrinsic length scale is $\sqrt{(D\Theta)}$.

Where do intrinsic time scales come from? They may arise from two different types of equations, both being constitutive equations. The first type is that of kinetic phenomena, i.e., phenomena the rate of which is governed by their own intrinsic kinetics. Chemical reactions are of course the example which immediately comes to mind, but they do not exhaust the field. Whenever the thermodynamic state includes some internal state variable, the rate of change of the latter is given by a kinetic equation, and hence an intrinsic time scale can — with some attention —be identified. The second type of intrinsic time scale is the one arising when relaxation phenomena are considered; any relaxation phenomenon is or is not relevant to the analysis of any given problem according to whether the intrinsic time scale, i.e., the relaxation time, is or is not comparable with other time scales of the process.

Intrinsic time scales arise from constitutive equations, but as we have seen in the previous section some care must be exercised in extracting the time scale when the constitutive equation is not linear. For instance, the value of $\phi_S{}^2$ in the previous section can be interpreted as the ratio of the diffusion time L^2/D to the intrinsic time scale evaluated at the surface conditions, say $\Theta = 1/k_{EQ,S} = \Theta_S$. We have also seen, however, that one cannot simply carry over results from the linear case to the nonlinear one by using such time scales evaluated at some particular point. For instance, in the case of Langmuir isotherm kinetics, one could well argue that the intrinsic time scale is $1/k$, not $(1 + KC_S)/k$. There is, in fact, a degree of arbitrariness: one may choose any definition one may wish for the intrinsic time scale, but one should then be careful about order of magnitude analysis.

Let's consider a rather general case of kinetic equation, say one where the rate of change of some internal state variable w depends on the value of w itself, say:

$$\partial w/\partial t = f(w) \tag{3.3.2.1}$$

Since w is an internal state variable, it will have some equilibrium value w^*, such that at $w = w^*$ $\partial w/\partial t = 0$. It is therefore useful to rewrite (3.3.2.1) in the following form:

$$\partial w/\partial t = f(w^* - w) \tag{3.3.2.2}$$

so that the function $f()$ is now constrained by $f(0) = 0$, $\text{sgn} f = \text{sgn}(w^* - w)$. The operator $f()$ is still dimensional, even if w has been normalized to be dimensionless (as can always be easily done). One can easily rewrite (3.3.2.2) in a form which contains a dimensionless operator $F()$ subject to the same constraints:

$$\partial w/\partial t = F(w^* - w)/\Theta \tag{3.3.2.3}$$

where Θ is an intrinsic time scale. However, the constraints $F(0) = 0$, $\text{sgn} F = \text{sgn}(w^* - w)$ do not uniquely determine the value of Θ, and there is therefore still freedom of choice.

A possibility which is always present is as follows. If we are considering this kinetic problem, it must be because we know that w is either initially, or at some external boundary, different from w^* — otherwise there would be no kinetics to consider. In other words, we have some physically significant driving force $\delta w = w^* - w$ which is different from zero. In the example discussed in the previous section, c_S played the role of the driving force. We can now require $F(\delta w)$ to be unity, which of course uniquely determines the value of Θ. Furthermore, we have now a physical interpretation of Θ: it is the intrinsic time scale corresponding to the imposed driving force δw. This is essentially the choice Θ_S done in the previous section.

Another possibility is as follows. Since we know that $F(0) = 0$, $\mathrm{sgn}F = \mathrm{sgn}(w^* - w)$, we can (perhaps) expand $F()$ in a Taylor series near equilibrium and write:

$$F(w^* - w) = (w^* - w)F'(0) + O[(w^* - w)^2] \tag{3.3.2.4}$$

If we now require $F'(0) = 1$, the value of Θ is uniquely determined. In this case, Θ is recognized as the intrinsic time scale near equilibrium — in the Langmuir isotherm case it would be the constant $1/k$ (since equilibrium corresponds to $c = 0$). Now quite obviously the two choices are both legitimate, but one should be very careful about the implications of dimensional analysis based on either one of them.

While the first choice, $F(\delta w) = 1$, is always possible —after all, $f(\delta w)$ must have some finite value — the second choice is possible only if the kinetic constitutive equation can be expanded in a Taylor series at the equilibrium point, i.e., if $F'(0)$ is finite. Now this may seem a trivial requirement (Bob is losing interest fast, since his philosophy of life is that everything of even remotely pragmatical interest is always differentiable and smooth), but it deserves some discussion.

Let's first consider the case of an irreversible dimerization reaction, for which $\partial c/\partial t = -bc^2$. In this case the equilibrium point is $c = 0$, and the kinetic function is not expandable in a Taylor series at the equilibrium point, since $F'(0) = 0$ and we cannot require it to be unity. This does seem a rather convincing counterexample, doesn't it, Bob? However, as usual, Bob was right — that's the mark of the great practical engineer, to be right without knowing exactly why. We know that no reaction is truly irreversible, and that therefore the correct kinetic equation for a dimerization reaction is:

$$-\partial c/\partial t = b(c^2 - c_2/K) \tag{3.3.2.5}$$

where c_2 is the dimer concentration and K is the concentration based equilibrium constant. Now, given a dimer concentration c_2, one may define an equilibrium concentration of monomer as $c^* = \sqrt{(c_2\ /\ K)}$: c^* is the monomer concentration which would equilibrate the local instantaneous dimer concentration. The RHS of 3.3.2.5 would thus be written as $b(c^2 - c^{*2})$. Now suppose that $c = c^*(1 + \varepsilon)$, with $\varepsilon << 1$, i.e., the actual concentration is close to the equilibrium one. One obtains:

$$-\partial c/\partial \mathrm{t} = 2b\varepsilon c^{*2} + O(\varepsilon^2) \tag{3.3.2.6}$$

i.e., the rate of reaction would be linear in the displacement from equilibrium εc^*. It follows that $F'(0)$ would be nonzero, and Θ (near equilibrium) would be identified with $1/(2bc^*)$ — which of course becomes zero in the case of an irreversible reaction. But if we believe that there is no such thing in nature as a truly irreversible reaction, one would always get a finite Θ according to the second procedure, if one which is exceedingly large for reactions which are almost irreversible.

The other point to be made is that the assumption that $F()$ is Taylor series expandable near equilibrium is equivalent to the assumption that it is invertible near equilibrium — an assumption which was shown in the Chapter on thermodynamics to be needed in order to get the usual result that the affinity is zero at equilibrium. Since the latter result is invariably regarded as the correct one in the theory of equilibrium, the assumption that $F'(0)$ is finite is nothing else but the assumption one makes, albeit in most cases implicitly, to obtain the classical theory of equilibrium. So Bob was right after all: in all pragmatically relevant cases, $F()$ is indeed expandable in a Taylor series near equilibrium.

Now go back to the case where we wish to identify Θ from the requirement that $F(\delta w) = 1$, as we have done in the previous section. While this sets the value of Θ, the value one gets does depend on the imposed driving force δw. This may seem to be a consequence of the nonlinearity of the constitutive equation one is considering, since for a first order reaction $\Theta = 1/k$ whatever the driving force may be. However, this conclusion is not true in general, and the following example shows that one may have a perfectly linear problem where the ratio of two relevant time scales depends on the imposed driving force.

The example is that of the Stefan problem, which was discussed in Section 1.7. Indeed, Sue has been pointing out that we are redoing the analysis in Sections 1.7–9 over and over again, and if one accepts the rather sloppy mathematics of all her statements we are compelled to say she is right. Still, we'll revisit the Stefan problem from the viewpoint of dimensional analysis, and we hope we can get some better understanding of it. We have seen that the time scale of a straightforward transport problem is L^2/D. If we have a sample with half-thickness L initially at some temperature T_0, the faces of which are brought to temperature $T_0 + \delta T$ at time zero and kept there afterwards, the amount of heat to be removed per unit exposed surface is $cL\delta T$, and under a driving force δT it will take a time of order L^2/D to do that: the amount to be removed and the rate of removal are both proportional to the imposed driving force, and hence the time needed for removal is independent of δT. However, consider the case where we have to remove an amount of heat governed by the latent heat Γ, say an amount per unit surface which is ΓL — this is not proportional to δT. The driving force is the same as before, and thus also the rate of removal is of the same order of magnitude, so that now the time required is $(L^2/\mathrm{D})(\Gamma/c\delta T)$. So the Stefan number $St = \Gamma/c\delta T$ is seen to be the ratio of the time needed to remove the latent heat to that needed to remove the sensible heat under the same driving force — and that ratio does depend on the imposed driving force δT in spite of the fact that all equations governing the phenomenon are linear.

3.3.3 Scaling, Nusselt and Biot type numbers

In this section, we address the following deceivingly simple question: what is the physical meaning of dimensionless groups of the type of the Nusselt number? Let's first see what we mean by the classical Nusselt number arising in the theory of heat transfer, $\mathbf{Nu} = hL/K$, where h is the heat transfer coefficient, L is some characteristic length, and k is the thermal conductivity. The heat transfer coefficient is in turn defined from $q = h\delta T$, where q is the heat flux orthogonal to the walls which are the boundary of our system, and δT is the driving force. The obvious analog in the theory of mass transfer is the Sherwood number kL/D, where k is the mass transfer coefficient, and D is diffusivity; k is again defined from $N = k\delta c$, with N the mass flux and δc the driving force. Now consider the case of momentum transfer, where things aren't quite so trivial. For the sake of simplicity, consider first turbulent flow in a tube. The equivalent of q and N would be the momentum flux at the wall, say the wall shear stress τ. The imposed driving force would be the average velocity V. Thus the equivalent of h and k would be τ/V, and since $\tau = LP/4$, where L is the tube diameter and P is the pressure drop per unit length, one would have a transfer coefficient given by $LP/4V$. Now P is expressible in terms of the friction factor f as $P = f\Phi V^2/L$, and therefore the transfer coefficient is (to within numerical factors) $f\Phi V$. The equivalent of the Nusselt number would thus be $f\Phi VL/\mu = f\mathbf{Re}$. The result can be extended to the case of flow around a submerged object, if τ is now interpreted as the average flux of the momentum component in the direction of relative motion, and one would have the drag force divided by the cube of the linear dimension as the equivalent of P. The Nusselt number equivalent would again come out to be the product of the drag coefficient and the Reynolds number.

Now let's go back to the simple heat transfer problem — we know by now how to carry over results to the cases of mass and momentum transfer. Let x be a coordinate orthogonal to the wall, with $x = 0$ identifying the bounding wall itself. Should we know the space distribution of temperature T, we would calculate q as $k(\partial T/\partial x)_0$, and hence $\mathbf{Nu} = (\partial T/\partial x)_0 L/\delta T$. Now there is no doubt that variations of T are of order δT, so that we estimate $(\partial T/\partial x)_0$ to be of order $\delta T/l$, where l is the distance over which the significant variation of T takes place (l is often called, somewhat improperly, the boundary layer thickness). We therefore identify $\mathbf{Nu}$ with L/l. Had we chosen the appropriate scaling factor properly, we would have chosen l instead of L, and $\mathbf{Nu}$ would have been of order unity by definition. We are unlikely to choose the scaling factor for length smaller than the actual one — for instance, in a tube how could the proper scaling factor for distance be significantly more than the diameter, which we always choose as the scaling factor? — and thus the Nusselt number type of dimensionless group is expected to be either of order unity or much larger than that. (It will be much larger than unity if we overestimated the scaling factor for length by taking, say, the tube diameter while in actual fact the whole variation of T takes place in a thin layer of thickness l near the tube wall).

The situation is entirely different in the case of the Biot number. Bi is defined as hL/k, where L is the characteristic dimension of the phase where heat conduction is being considered, k is the conductivity of that same phase, but h is the heat transfer coefficient in another phase. It follows that the Biot number may have any value whatsoever, ranging from zero (the exposed surface is essentially adiabatic) to infinity (there is no external heat transfer resistance). It is instructive to consider the deceivingly simple problem of a slab of half thickness L, initially at temperature T_0, which at time zero is brought in contact with a fluid phase at temperature T_1; h is the heat transfer coefficient in the fluid phase. In dimensional form, the differential equation and boundary conditions are:

$$\Phi c \partial T/\partial t = k\partial^2 T/\partial x^2 \tag{3.3.3.1}$$

$$t = 0, T = T_0 \tag{3.3.3.2}$$

$$x = 0, -k\partial T/\partial x = h(T_1 - T) \tag{3.3.3.3}$$

$$x = L, \partial T/\partial x = 0 \tag{3.3.3.4}$$

The dimensionless temperature Θ needs to be defined as $(T_1 - T)/(T_1 - T_0)$, so that (3.3.3.3) becomes homogeneous. As far as scaling distances is concerned, there is no doubt that the appropriate scale is L, and thus one defines $y = x/L$. The scaling of time is more delicate. One possible scale is $\Phi cL^2/k$, which is the heat diffusion time inside the slab; this would lead to defining $\tau = kt/\Phi cL^2$, which is known to eliminate all parameters from the differential equation. However, $\Phi cL/h$ also has dimensions of time, and it could be used as the time scale. The question which arises is of course which one is the realistic time scale; in other words, how long does it take for the slab to approach the final condition $T = T_1$? The answer may be $\Phi cL^2/k$ (the internal diffusion time), or $\Phi cL/h$ (the external resistance time), and at this stage the best one can do is to suspect that it will be the one which happens to be larger. Now notice that the ratio of the internal to the external time scale is the Biot number. We choose (somewhat arbitrarily) the internal time scale, and hence the dimensionless time τ. The dimensionless formulation becomes:

$$\partial\Theta/\partial\tau = \partial^2\Theta/\partial y^2 \tag{3.3.3.5}$$

$$\tau = 0, \Theta = 1 \tag{3.3.3.6}$$

$$y = 0, \partial\Theta/\partial y = Bi\Theta \tag{3.3.3.7}$$

$$y = 1, \partial\Theta/\partial y = 0 \tag{3.3.3.8}$$

Now let's first consider the case of a negligible external resistance to heat transfer, say $1/Bi$ approaching zero. Boundary condition (3.3.3.7) becomes $\Theta(0,\tau) = 0$, and thus the problem reduces to the classical one which was discussed in section 1.6, and the time scale of interest is indeed $\tau \approx 1$. So far, nothing is very exciting. However, let's now consider the case of a very high external resistance to heat transfer, say Bi approaching zero. Does anybody in class have an idea of how to deal with this asymptotic case?

Ralph says that the solution for any value of *Bi* is in Carslaw and Jaeger, and in fact there is the solution for an arbitrary initial distribution of temperatures, provided it can be expanded in a Fourier series, so why are we bothering with asymptotes and all that? He is perfectly right, of course, the complete solution is in Carslaw and Jaeger, and perhaps we shouldn't bother, but let's see if there are any other suggestions. Bob has one, based on a different idea. He says that, if external heat transfer is the predominant resistance, temperature within the slab will be uniform at all times, $T = T(t)$, and hence that one simply needs to integrate $\Phi c L \mathrm{d}T/\mathrm{d}t = h(T_1 - T)$ subject to (3.3.3.2). Very nice indeed, since now the alternate time scale $\Phi cL/h$ emerges quite naturally. We need the integral of Bob's equation, and Sue is asked to perform the necessary algebra and to express the result in terms of Θ, τ and *Bi*. Once Sue's horrible mess is cleared up there is an agreement that the result is $\Theta = \exp(-Bi\tau)$. So now we conclude that the time scale for final equilibration is $\tau \approx 1/Bi$, which is very significantly larger than unity. Well, this is very nice, but could we have obtained the same result from (3.3.3.5–8), without reformulating the problem? After all, $\Theta = \exp(-Bi\tau)$ must be the zero order approximation to the solution of (3.3.3.5–8) if we believe it to have any validity at all.

Let's try a trick which has worked other times (see again Section 1.6). Let's take the integral of (3.3.3.5) over the whole thickness to obtain:

$$\int (\partial\Theta/\partial\tau)\mathrm{d}y = -Bi\Theta(0,\tau) \tag{3.3.3.9}$$

Equation (3.3.3.9) simply states that the instantaneous rate of accumulation of heat in the body equals the rate of supply of heat at the surface. But as *Bi* approaches zero, (3.3.3.7) suggests that $\partial\Theta/\partial y = 0$, and hence, by a somewhat hand-waving argument, (3.3.3.9) reduces to $\partial\Theta/\partial\tau = -Bi\Theta$, which of course integrates to the result we expected. Sue likes this rabbit out of the hat enormously; Bob finds it irrelevant; but Ralph points out that our solution $\Theta = \exp(-Bi\tau)$ is not a solution at all, since the differential equation is not satisfied: $\partial^2\Theta/\partial y^2 = 0$, but $\partial\Theta/\partial\tau = -Bi \exp(-\mathrm{Bi}\tau)$ which is not zero. It is not zero indeed, but it is at most of order *Bi*, and hence the differential equation is satisfied to within $O(Bi)$, isn't it?

Ralph isn't convinced by our argument, and even Sue this time finds our technique slightly too cavalier. So let's be a little more sophisticated. We suspect that $\exp(-Bi\tau)$ is the zero order solution, and so we define ϕ as the difference between the actual and the (presumed) zero order solution, $\phi = \Theta - \exp(-Bi\tau)$. We wish to find a first order solution, so we'll express $\exp(-Bi\tau)$ as $1-Bi\tau$, and we'll drop all terms which are quadratic in *Bi*. The result is:

$$\partial\phi/\partial\tau = \partial^2\phi/\partial y^2 + Bi \tag{3.3.3.10}$$

$$y = 1, \partial\phi/\partial y = 0 \tag{3.3.3.11}$$

$$y = 0, \partial\phi/\partial y = Bi(\phi + 1) \tag{3.3.3.12}$$

$$\tau = 0, \phi = 0 \tag{3.3.3.13}$$

So ϕ starts from a value of zero, and at time zero $\partial\phi/\partial\tau = Bi$, so that at small times ϕ is of order $Bi\tau$. At large τ, ϕ will approach some constant distribution $\phi(y)$, which is trivially calculated from (3.3.3.10–12) by setting $\partial\phi/\partial\tau = 0$: one obtains $\phi = Bi(y - y^2/2)$. This shows that ϕ is always at most of order Bi, i.e., it proves that $\exp(-Bi\tau)$ is indeed the zero order solution. For $\tau >> 1$, $\Theta = \exp(-Bi\tau) + Bi(y - y^2/2)$ is the first order solution. (Just to make Ralph happy, we point out that Bi approaching zero corresponds, in Carlslaw and Jaeger's analysis on pp. 114–115, to hl approaching zero. With this, the first root $\alpha_1 = \sqrt{(hl)} << 1$, and since the second root must be $> p$ one has $\alpha_2 >> \alpha_1$, and hence the second term in the sum in Eq. (8) decays very rapidly indeed as compared to the first one. The first term is of course our zero order solution).

3.4 MODELLING AND APPROXIMATIONS

In this section, we are going to discuss two examples of approximations in the modelling of phenomena. A very common, and very powerful approximation which chemical engineers do all the time —sometimes rather unconsciously— is the quasi steady state approximation (QSSA). The QSSA is based on the following general considerations: often the solution of a problem is available for steady state conditions, but it is hard to obtain for unsteady state ones. In such cases, one wishes to apply the steady state solution to conditions which are known not to be quite steady, but perhaps in some sense they may be regarded as being almost steady. Let's consider a specific problem.

In the Ziegler–Natta process of polymerization, a heterogeneous catalyst is used, which is introduced into the system in the form of very small particles, which we are willing to regard as approximately spherical, with radius R_0. The catalyst is, let's say, sparged into a liquid phase of monomer, and polymerization occurs until rather large particles of polymer are produced. In fact the particles are large enough that the volume ratio of polymer to catalyst is very, very large indeed, so that one does not need to worry about recovering the catalyst. Let's say one is willing to consume one part of catalyst per 1,000,000 parts of polymer; the final volume ratio needs to be 10^6. Since volume is proportional to the cube of radius, this means that the final polymer particle must have a radius 100 times larger than R_0. Well, that does not sound impossible. Let's proceed to model the process.

First the physics. A growing spherical shell of polymer forms on the catalyst; let $R(t)$ be its radius. The monomer surrounding the growing polymer particle must:

(a) dissolve into the polymer.
(b) diffuse towards the catalyst surface.
(c) react on the catalyst surface.

Hence we need to write a diffusion equation in the polymer shell. Step (a) is described by the requirement that, at the outer limit of that layer, the concentration of monomer in the polymer layer equals its solubility c_0:

$$r = R(t), c = c_0 \tag{3.4.1}$$

Step (b) is modelled by writing the diffusion equation in the polymer shell:

$$D[\partial^2 c/\partial r^2 + (2/r)\partial c/\partial r] = \partial c/\partial t \tag{3.4.2}$$

Step (c) may perhaps be modelled by assuming that the reaction rate (per unit catalyst surface) is first order with respect to the monomer concentration at the catalyst surface. That's reasonably simple, and it may do for the time being. One thus writes, with k a kinetic constant per unit surface area having units of a velocity:

$$r = R_0, D\,\partial c/\partial r = kc \tag{3.4.3}$$

Anybody have anything to say? Do we have a well posed mathematical problem? Ralph points out that we don't know what $R(t)$ may be, and hence (3.4.1) is not a legitimate boundary condition; in fact his hand was raised since the time we wrote it down. Good point, Ralph; so how are we going to get $R(t)$? Bob suggests a mass balance should do: whatever reacts on the catalyst surface contributes to the growth of $R(t)$. Good point again. Sue, how would you write an equation describing mathematically what Bob just said? Sue makes a horrible mess of it, and when finally the smoke is cleared up there is agreement that, if c_1 is the density of the polymer expressed as moles of monomer units per unit volume, the equation should be:

$$kc(R_0,t)R_0^{\,2} = R^2c_1\mathrm{d}R/\mathrm{d}t \tag{3.4.4}$$

This is a differential equation, and a boundary condition is needed, and everybody agrees that it should be:

$$R(0) = R_0 \tag{3.4.5}$$

Bob raises his hand here. "Obviously a quasi steady state approximation is legitimate provided the reaction rate is low, say k is small enough. From the list of parameters there is only one combination having units of a velocity, and that is D/R_0. Hence k is the only combination having units of a velocity, but what about c_1/c_0? That is an independent dimensionless group, and maybe the condition is k.

Sue has a suggestion: c_1 appears only in Eq. (3.4.4), and there what matters is really the ratio k/c_1, so her suggestion is that we'll find out that the condition is $kc_0/c_1 << D/R_0$. This boils down to whether Bob's presumption that the QSSA holds provided the reaction is slow enough is or is not reasonable. Well, the QSSA consists in regarding the diffusion phenomenon as steady, and hence the other phenomenon — the reaction — has to be slow: the time scale of the reaction, which is the time scale of unsteadiness, must be large as compared to the diffusion time scale. The trouble here is that we have a velocity scale and not a time scale for the reaction, perhaps. Bob grants that there is some room for further analysis, though he insists that Sue's inequality must be right.

Let's first make everything dimensionless. We are experts by now at that, but in order to avoid another mess of algebra, this time Ralph is asked to suggest the appropriate dimensionless variables and parameters, and his list is impeccable as we expected it to be:

$$b = c/c_0;\ z = r/R_0;\ \phi = R/R_0;$$

$$\tau = tD/R_0^{\,2};\ A = kc_0R_0/Dc_1 \tag{3.4.6}$$

Somewhat annoyed at the fact that the parameter A seems to hint at the fact that Sue was perhaps right, Ralph also writes down the form to which the equations reduce to, with some half-spoken comments about people who just make lucky guesses:

$$\partial^2 b/\partial z^2 + (2/z)\partial b/\partial z = \partial b/\partial \tau \tag{3.4.7}$$

$$z = \phi(\tau),\ b = 1 \tag{3.4.8}$$

$$z = 1,\ \partial b/\partial z = Ab \tag{3.4.9}$$

$$Ab(1,\tau) = \phi^2 \mathrm{d}\phi/\mathrm{d}\tau \tag{3.4.10}$$

$$\phi(0) = 1 \tag{3.4.11}$$

"I'll get a computer code to solve those in no time at all". That's not the point, Ralph. We want to ask ourselves about the QSSA, right? In other words, suppose we formulate the problem as follows: in the diffusion equation, we forget about time dependence, say we set the RHS of (3.4.7) to zero. We now solve the problem — what does the solution mean? "It does not mean anything: (3.4.8) guarantees that *b* will be in fact a function of time, and so you can't set the RHS of (3.4.7) to zero". Well, let's give it a try anyhow. Just go ahead and solve the QSS problem, and let's look at the solution, shall we? Ralph, somewhat annoyed at this futile exercise, does it correctly anyhow, and he obtains:

$$b = [\phi + A\phi - A\phi/z]/[\phi + A\phi - A] \tag{3.4.12}$$

$$\phi \mathrm{d}\phi/\mathrm{d}t = A/(\phi + A\phi - A);\ \phi(0) = 0 \tag{3.4.13}$$

We have to exert all our authority to refrain Ralph from integrating (3.4.13), something he wants to do in spite of his strong belief that (3.4.12–13) have nothing to do with our original problem. But that's exactly the point we wish to investigate, and so we proceed to consider that the QSS hypothesis holds provided $\partial b/\partial \tau$ is negligibly small as compared to any one of the terms on the LHS of (3.4.7), say:

$$\partial b/\partial \tau \tag{3.4.14}$$

Ralph, please, if (3.4.14) is satisfied $\partial b/\partial\tau$ will be much less than $\partial^2 b/\partial z^2$ as well, don't you see? Now let's first consider what happens at the outer layer of the polymer shell, say at $z = \phi(\tau)$. Let $\mathrm{D}b/\mathrm{D}\tau$ be the "substantial" derivative of b along $\phi(t)$, which is zero because of (3.4.8):

$$0 = \mathrm{D}b/\mathrm{D}\tau = \partial b/\partial\tau + (\partial b/\partial z)\,\mathrm{d}\phi/\mathrm{d}\tau \tag{3.4.15}$$

$$\partial b/\partial\tau = -(\mathrm{d}\phi/\mathrm{d}\tau)\partial b/\partial z \tag{3.4.16}$$

Hence (3.4.14) is satisfied provided that:

$$|\mathrm{d}\phi/\mathrm{d}\tau| \tag{3.4.17}$$

Now (3.4.13) tells us that $\mathrm{d}\phi/\mathrm{d}\tau$ decreases more than linearly with increasing ϕ, and hence that (3.4.17) is hardest to satisfy the smaller ϕ is. Since the smallest value of ϕ is unity, we require (3.4.17) to be satisfied at $\phi = 1$ to obtain:

$$A << 2 \tag{3.4.18}$$

Is everybody happy? Bob and Sue are quite happy, of course, since their wild guess has been validated, but Ralph has a serious objection. He wants to check that (3.4.14) is satisfied at $z = 1$ as well, and he proceeds to show that it is not, since he calculates the following from (3.4.12) at $z = 1$, $\tau = 0$:

$$\partial b/\partial\tau = -A \tag{3.4.19}$$

$$2\partial b/\partial z = 2A \tag{3.4.20}$$

Very good point indeed, Ralph. There is indeed a neighbourhood of time zero where the QSSA is not valid near the surface of the catalyst. At times other than zero, we obtain at $z = 1$:

$$(\partial b/\partial\tau)/(2\partial b/\partial z) = -1/2\phi(\phi + A\phi - A) \tag{3.4.21}$$

Since, as discussed at the beginning, the final value of ϕ of interest is 100 or more, (3.4.21) in fact guarantees that, at later times, the QSSA is justified near the catalyst surface whatever the value of A may be.

Having convinced ourselves that if (3.4.18) is satisfied we can indeed use the QSSA, we now proceed to integrate (3.4.13):

$$A\tau = (A + 1)(\phi^3 - 1)/3 - A(\phi^2 - 1)/2 \tag{3.4.22}$$

However, since (3.4.13) holds only if (3.4.18) is satisfied, we have to be coherent with ourselves and therefore we can only trust (3.4.22) to within $O(A)$, i.e., we need to write:

$$\phi^3 = 1 + 3A\tau \tag{3.4.23}$$

This is an important point of a general nature. If a simplified model is used, one needs to establish the order of approximation; in the case at hand, the approximation is valid to within terms of order A. It follows that the solution of the equations describing the simplified model is valid only to within the same order, and hence that terms of order A^2 have to be neglected: it does not make sense to keep terms of an order which has been considered as negligibly small when setting up the model. This may seem as a trivial conclusion, but in fact there is an abundance of results in the literature where terms are retained in the solution which are of the same order as terms which have implicitly been neglected.

We now move to another example which provides a somewhat unusual type of QSSA. Consider a CVD process where a gaseous stream containing an organometallic compound C is produced by exposing a metal surface M to a low pressure gas containing a finite concentration of free organic radicals R (for example, an ethane gas at low pressure and high temperature will have a finite concentration of $CH_3^{\cdot}$ radicals). The following chemical reaction may take place:

$$R + M = C \tag{3.4.24}$$

The intrinsic rate of this reaction is believed to be extremely high, so that it may be regarded as instantaneous in a diffusion-reaction model. Let p_R be the partial pressure of R in the gas phase; if one assumes that the solid surface is exposed metal M, the partial pressure of radicals at the surface will be zero, and thus the flux of radicals to the metal surface is given by:

$$N_R = k_R p_R \tag{3.4.25}$$

where k_R is the pressure-based mass transfer coefficient of the radical in the gas phase.

The flux of radicals to the surface equals the number of moles of C formed per unit surface, and if the metal has to stay exposed an equal flux of C from the surface to the gas phase must establish itself. Now what is the transport equation which determines the flux of C away from the interface? Ralph says the question is trivial: it is $k_C(p_{CI}-p_{C0})$, with p_{CI} the partial pressure of C at the interface, and p_{C0} its value in the bulk of the gas. Very good; presumably we know what p_{C0} might be, but how do we determine what p_{CI} might be? Ralph has the solution: since the flux of C must balance that of R, p_{CI} will adjust itself to the appropriate value. Again very well indeed; but isn't there an upper bound to the value of p_C? After all, the partial pressure of the organometallic cannot exceed its vapour pressure p^0. So what if the value of p_{CI} needed to balance the radical flux exceeds p^0? Bob has seen the light: if that is the case, a layer of liquid organometallic will form, and the radical flux will be slowed down by the need to diffuse through the liquid layer in order to reach the metal surface.

Having reached agreement on the basic physics of the problem, we may proceed. The highest possible value of p_{CI} is p^0, and hence the largest possible flux of C is:

$$N_{CMAX} = k_C p^0 \tag{3.4.26}$$

Now if N_R as given by Eq. (3.4.25) is larger than N_{CMAX}, a liquid layer will indeed form, and the partial pressure of radicals at the surface will be larger than zero. In the limiting case where diffusion through the liquid layer is the rate-determining phenomenon, the interface partial pressure will in fact be equal to the bulk gas partial pressure p_R. Let's begin by analyzing the latter situation.

Let $c(x,t)$ be the concentration of radicals in the liquid organometallic layer, with x being distance from the original position of the exposed surface, and t being time elapsed since the surface was exposed to the gas. The distribution $c(x,t)$ is determined by the one-dimensional unsteady diffusion equation:

$$D\partial^2 c/\partial x^2 = \partial c/\partial t \tag{3.4.27}$$

The liquid–metal surface is located at some position $x = s(t)$. At that surface, the chemical reaction takes place at an essentially infinite rate, and therefore:

$$x = s(t),\ c = 0 \tag{3.4.28}$$

Now let Φ_M and Φ_C be the molar densities of the metal and of the organometallic, respectively, and define $V = N_C/\Phi_C$ and $\beta = \Phi_M/\Phi_C$ (notice that β is likely to be close to unity). A simple mass balance thus yields the result that the thickness $h(t)$ of the liquid layer increases at a rate given by:

$$\mathrm{d}h/\mathrm{d}t = \beta \mathrm{d}s/\mathrm{d}t - V \tag{3.4.29}$$

It follows that the position of the exposed surface $q(t)$ moves downwards with a velocity $V + (1 - \beta)\mathrm{d}s/\mathrm{d}t$:

$$\mathrm{d}q/\mathrm{d}t = V + (1 - \beta)\mathrm{d}s/\mathrm{d}t;\ q(0) = 0 \tag{3.4.30}$$

The boundary condition imposed on Eq. (3.4.27) are, if c_0 is the solubility of the radical corresponding to p_R, and recalling that the reaction may be regarded as instantaneous:

$$x = q(t),\ c = c_0 \tag{3.4.31}$$

$$x = s(t),\ c = 0 \tag{3.4.32}$$

However, in order to formulate a well posed free boundary problem, we need an additional boundary condition, since the position of the liquid–metal interface $s(t)$ is not known in advance (there are two moving boundaries in the problem, $q(t)$ and $s(t)$, but once $s(t)$ is determined, $q(t)$ is given by Eq. (3.4.30)). This is obtained from a mass balance at that surface. The flux of radicals reaching the surface is $-D\partial c/\partial x$, and this represents the number of metal moles transformed to organometallic per unit surface area and time. Thus one can write:

$$x = s(t), -D\partial c/\partial x = \Phi_M\, \mathrm{d}s/\mathrm{d}t;\ s(0) = 0 \tag{3.4.33}$$

The equations above constitute a well posed if somewhat unusual Stefan-like problem. The problem as formulated contains four dimensional parameters: D, V, Φ_M, and c_0. An intrinsic length scale D/V and an intrinsic time scale D/V^2 suggest themselves naturally. Thus the dimensionless distance y, time τ, exposed surface position Θ, liquid–metal surface position μ, concentration ϕ, and the Stefan number *St*, are defined as:

$$y = xV/D \tag{3.4.34}$$

$$\tau = tV^2/D \tag{3.4.35}$$

$$\Theta(\tau) = q(t)V/D \tag{3.4.36}$$

$$\mu(\tau) = s(t)V/D \tag{3.4.37}$$

$$\phi = c/c_0 \tag{3.4.38}$$

$$\mathbf{St} = \Phi_M/c_0 \tag{3.4.39}$$

The dimensionless formulation of the problem thus becomes:

$$\partial^2\phi/\partial y^2 = \partial\phi/\partial\tau \tag{3.4.40}$$

$$\mathrm{d}\Theta/\mathrm{d}\tau = 1 + (1-\beta)\mathrm{d}\mu/\mathrm{d}\tau \tag{3.4.41}$$

$$y = \Theta(\tau),\ \phi = 1 \tag{3.4.42}$$

$$y = \mu(\tau),\ \phi = 0 \tag{3.4.43}$$

$$y = \mu(\tau),\ -\partial\phi/\partial y = \mathbf{St}\,\mathrm{d}\mu/\mathrm{d}\tau;\ \mu(0) = 0 \tag{3.4.44}$$

The problem represented by Eqs. (3.4.40–44) cannot of course have a completely steady solution, i.e., an asymptotic solution where $\partial\phi/\partial\tau = 0$. However, a quasi-steady state asymptote may exist where the thickness of the liquid layer is constant, so that the exposed surface and the metal–liquid surface travel at equal and constant speed. Indeed, suppose that this is the case, with *K* the (dimensionless) constant thickness of the liquid layer:

$$\mu(\tau) - \Theta(\tau) = K \tag{3.4.45}$$

It is easy to find that Eqs. (3.4.41) and (3.4.45) are satisfied if $\Theta = \tau/\beta$ and $\mu = K + \tau/\beta$. Now we have got a liquid layer of constant thickness K, and can we call this a steady state situation? We cannot, because the concentration in the liquid layer is not constant in time at any value of y, since y is distance from the *original* position of the gas–metal surface. Bob says this is trivial: ϕ is not constant in time at any value of y, but it is constant in time at any value of $z = y - \tau/\beta$, since z measures distance from the instantaneous position of the liquid–gas surface. Good, and based on this would anybody care to suggest what the appropriate similarity variable might be? Sue beats everybody else on this, stating that she feels that it must be z itself. Ralph has a somewhat disgusted expression on his face, but when he substitutes $\phi = f(z)$ in the equations he finds:

$$f'' + f'/\beta = 0 \tag{3.4.46}$$

$$f(0) = 1, f(K) = 0 \tag{3.4.47}$$

$$-f'(K) = \mathbf{St}/\beta \tag{3.4.48}$$

Ralph's residual objection that $\mu(0) = 0$ is not satisfied is brushed aside by the consideration that the QSSA only holds at sufficiently long times.

The solution is trivially calculated as:

$$f = [\exp(-z/\beta) - \exp(-K/\beta)]\,/\,[1 - \exp(-K/\beta)] \tag{3.4.49}$$

$$K = \beta \ln(1 + 1/\mathbf{St}) \tag{3.4.50}$$

It is now of interest to estimate the time required in order to approach the quasi-steady state asymptote. One would be tempted to estimate the initial rate of growth of the liquid layer thickness (i.e., $d\mu/d\tau - d\Theta/d\tau$ at time zero) and thus estimate the time needed to reach the value K with the initial rate. This however cannot be done, since the initial rate of growth of the liquid layer thickness is infinitely large. However, the estimate can be obtained by another argument. At small times, $|d\Theta/d\tau|$ is significantly less than $d\mu/d\tau$. In fact, at time zero the conditions are identical with those of the classical Stefan problem, and thus $d\mu/d\tau$ is proportional to $1/\sqrt{\tau}$. Equation (3.4.41) now shows that the dominant term for $d\Theta/d\tau$ at small times is the second one, and since β

has a value close to unity, one concludes that $|d\Theta/d\tau|$ is much less than $d\mu/d\tau$. (Notice that if $\beta > 1$, $d\Theta/d\tau$ will be negative at small values of τ). Thus initially the process is dominated by the classical Stefan problem equations:

$$\mu(\tau) = 2A\sqrt{\tau} \tag{3.4.51}$$

$$\exp(-A^2) = \mathbf{St}A\, erfA\sqrt{\tau} \tag{3.4.52}$$

We now observe that **St** is, for the problem at hand, necessarily a large number, since Φ_M is guaranteed to be significantly larger than c_0. It follows that Eqs. (3.4.50) and (3.4.52) can be expanded in powers of 1/**St**, and with the expansions truncated at the first term one obtains:

$$K \approx \beta/\mathbf{St} \tag{3.4.53}$$

$$A^2\,\mathbf{St}\sqrt{\tau} \approx 1 \tag{3.4.54}$$

Requiring $\mu(\tau)$ as given by Eq. (3.4.51) to be equal to K provides the required estimate of the time needed to approach the quasi-steady state solution, τ^*:

$$\tau^* \approx \beta^2\sqrt{\pi}/4\mathbf{St} \tag{3.4.55}$$

This is guaranteed to be a very small number, and hence the time needed to reach the quasi-steady state solution is significantly less than D/V^2.

PROBLEMS

3.1 For the problem considered in Section 3.2.4, is it possible that Coriolis forces due to revolution of the earth around the sun may play a role? If yes, for how long should the wind blow steadily before they become of importance?

3.2 Consider settling of a sphere in a liquid. Let δ be the momentum boundary layer thickness, and D the diameter of the sphere. How does δ/D change with changing **Re**? Is the answer the same at **Re** values larger and smaller than unity?

3.3 A cylindrical wire of diameter D is withdrawn continuously from an infinite pool of liquid; the purpose is to coat the wire with some finite thickness of the liquid, which is achieved by freezing the liquid layer at some distance well above the pool surface.

First consider the case where D is large (as compared to what?), so that we are willing to approximate the problem with that of withdrawal of an infinitely wide flat plate strip. This eliminates D from consideration. What are the relevant dimensional parameters of the problem?

There are four forces to be considered: inertial, viscous, gravitational, and capillary. There is no external length scale, so we need to establish an internal one. How many different internal length scales can we establish? What do these mean physically? Is it possible to find a scale such that viscous and gravity forces are of the same order of magnitude? Or such that the capillary and viscous forces are of the same order of magnitude?

A preliminary analysis should yield the result that near the pool's surface capillary and gravity forces predominate. What is the vertical length over which this is true?

At sufficiently large distances from the pool surface, we conclude that only viscous and gravity forces are important (how do we reach this conclusion?) Choose the appropriate length scale for this region, and calculate the Reynolds and Froude numbers — they should come out equal. Why?

Finally, we wish to investigate the possible importance of inertia forces, and thus to establish a dimensionless group which measures this. What is the appropriate group?

Now perhaps our initial assumption that D is large may be wrong. How do we deal with this?

3.4 Consider flow of a liquid out of a cylindrical tube into the atmosphere. The tube is vertical, and flow is downwards. We are interested in the diameter of the free jet as a function of distance from the tube outlet. Make as complete a dimensional and order of magnitude analysis of the problem as you can think of.

3.5 Consider a mixer consisting of an impeller rotating inside a baffled vessel. All geometrical factors are kept constant, so that the geometry is fixed once the impeller diameter D is assigned. Determine the dimensionless groups relevant to the problem, taking care that the unknown (the required mechanical power) appears in only one of them. For at least some limiting conditions, determine the presumable relationship between the required power and the rate of rotation of the impeller. Do the analysis without consulting a book on mixing.

3.6 Consider settling of a sphere through a less dense viscous liquid. The sphere starts from a state of rest, and conditions are such that the terminal velocity will be low enough for Stokes' law to apply. Can we apply Stokes' law during the unsteady part of the process? Doing that would be a QSSA; establish the conditions for its validity. Estimate the time needed to reach the terminal velocity in terms of the appropriate parameters of the problem.

3.7 For the polymerization problem considered in Section 3.4, consider the case where the parameter A is much larger than unity. What kind of approximations are possible, if any? At large τ, to which power of τ will ϕ be proportional?

3.8 Consider the motion of waves on the surface of a liquid; on the average, the surface is horizontal. Suppose the body of the liquid is infinitely deep (how deep needs it to be for this to be justified?), so the length L of the wave is the only length scale available. There are four forces to deal with: inertia, viscous, gravitational, and capillary. Presumably the relationship between the wave velocity U and L is determined by balancing two such forces.

(a) Estimate the value of U for all possible pairs of forces. Is there some pair which seems to lead nowhere? Why?

(b) There may be critical lengths L^* at which control may possible pass from one pair to another. What are the characteristic length scales? Discuss the patterns emerging from your estimates.

(c) Make a sketch in the $\log U$–$\log L$ plot of your estimates. Choose the properties of water for concreteness.

(d) Decide what the actual $U(L)$ curve looks like. Explain why some of the $U(L)$ relationships may never apply.

SUGGESTED READING

3.1 The book by Bridgman (1922) is the best presentation of dimensional analysis. The fundamental theorem of dimensional analysis is known as Buckingham's (1914) theorem, although in fact Rayleigh (1879), Vashy (1892) and Riabouchinsky (1911) gave earlier versions of it. The history of dimensional analysis is discussed by Macagno (1971). Scaling is discussed beautifully in a chapter of the book by M.M. Denn, "Process Modelling", Longman 1986.

3.2 Boundary layer problems are discussed in detail in the monograph by Schlichting (1955). The free convection problems follow the line of thought presented by Acrivos et al. (1960a,b), see also Ng and Hartnett (1988). The problem in oceanography is known as the Ekman (1905) layer, and is discussed by Defant (1961) and by Batchelor (1967). The latter book also discusses the vibrating sphere problem. Using different length scales in different directions was considered as early as 1892 by Williams; many applications of the method are discussed by Huntley (1953). It should be used cautiously, as discussed by Panton (1984).

3.3 The peculiarity of the power law is discussed by Astarita (1985) and by Astarita and Ocone (1988); its relationship with the variational theorem for creeping flow is discussed by Astarita (1977, 1983). Intrinsic time scales near equilibrium, and their relationship to the usual theory of equilibrium, are discussed in the context of a continuous description of reacting mixtures by Astarita and Ocone (1988). The fundamental textbook for heat conduction is Carlslaw and Jaeger (1959).

REFERENCES

Acrivos, A., AIChEJ, 6 (1960a) 584.
Acrivos, A., M.J. Shah, and E.E. Pete, AIChEJ, 6 (1960b) 312.
Astarita, G., J. Non-Newt. Fluid Mech., 2 (1977) 343.
Astarita, G., *ibid*., 13 (1983) 223.
Astarita, G., in: Scaleup, A. Bisio and E. Kabel (Eds.). J. Wiley, New York, 1985.
Astarita, G. and R. Ocone, AIChEJ, 34 (1988) 1299.
Batchelor, G.K., An Introduction to Fluid Dynamics. Cambridge University Press, 1967.
Buckingham, E., Phys. Rev. Ser. 2, 4 (1914) 345.
Carlslaw, H.S. and J.C. Jaeger, Conduction of Heat in Solids, 2nd edn. Oxford University Press, Oxford, 1959.

Defant, A., Physical Oceanography. Pergamon Press, Oxford, 1961.
Ekman, W.V., Ark. Math. Astr. och Fys, 2 (1905) 11.
Huntley, H.E., Dimensional Analysis. McDonald, London, 1953.
Macagno, E.O., J. Franklin Inst., 292 (1971) 391.
Ng, M.L. and J.P. Hartnett, Intl. J. Heat Mass Tr., 31 (1988) 441.
Panton, R.L., Incompressible Flow. J. Wiley, New York, 1984.
Rayleigh, Lord, Proc. Roy. Soc., 29 (1879) 71; Nature, 95 (1915) 66.
Riabouchinsky, D., L'Aerophile, p. 407.
Schlichting, H., Boundary Layer Theory. McGraw Hill, New York, 1955.
Vashy, A., Annales Telegraphiques, 19 (1892) 25.
Williams, W., Phil. Mag., 34–35 (1892) 234.

Chapter 4

Fluid Mechanics

4.1 BALANCE AND CONSTITUTIVE EQUATIONS

There are excellent books on Fluid Mechanics, and this chapter does not in any way try to be a substitute for them. We limit ourselves to discussion of some conceptual points which we believe are of particular importance; the reader is referred to the existing textbooks for more detail. In particular, the book by Batchelor is an excellent exposition of classical Newtonian fluid mechanics; the book by Lamb is the bible as far as hydrodynamics of ideal fluids are concerned, and the book by Schlichting is the best available presentation of boundary layer theory. Other useful books are quoted in the literature section.

In this section, we concentrate attention on the question of how to write down the equations governing fluid mechanics, without any discussion of the techniques needed to solve them. The equations to be written down are balance equations and constitutive equations; of course the former have general validity, while the latter hold only for restricted classes of materials and flow fields.

The principle of conservation of mass can be applied to the neighbourhood of a material point to obtain the continuity equation in its most general form:

$$\mathrm{D}\Phi/\mathrm{D}t = -\Phi \operatorname{div}\mathbf{v} \tag{4.1.1}$$

where $\mathbf{v}$ is the velocity vector, Φ is density, and $\mathrm{D}/\mathrm{D}t$ is the substantial derivative operator, $\mathrm{D}/\mathrm{D}t = \partial/\partial t + \mathbf{v}\cdot\mathrm{grad}$. Equation (4.1.1) applies to any material whatsoever, since it contains no constitutive assumptions.

Now in classical fluid mechanics one usually restricts attention to isothermal systems, and we will do so in this Chapter. [One could, incidentally, ask oneself whether such an assumption makes sense. Even if the boundaries of the flowing fluid are kept at constant temperature, frictional heating could cause temperature to be nonuniform. However, it is conceivable that on the one side frictional heating is of minor importance, and on the other that heat due to friction can be removed with negligibly small temperature gradients if the thermal conductivity of the fluid is sufficiently large (see Appendix 4.1). Under such assumptions, of course, the balance of energy does not play any role]. Another very common assumption is that of incompressibility, and this should be discussed in some detail.

In thermodynamics, one usually assumes that density is a unique function of temperature and pressure. [Again there is a subtlety involved. For instance, what is the density of H_2O at 100°C and 1 atm? It could be anything between the density of saturated steam and that of boiling water. However, this problem does not arise with one-phase systems]. If temperature is uniform, density would thus depend only on pressure, and the assumption of incompressibility would mean that we regard density as being the same at all pressures — that does seem to make sense, does it? The class does not like this reference to thermodynamics; we are dealing with a respectable subject such as fluid mechanics, and we should not muddy the water with thermodynamics. However, let's ask ourselves the following question: at equilibrium, density is indeed a unique function of temperature and pressure, or at least we are willing to grant this point. In a fluid, at equilibrium the internal stress is isotropic, and hence pressure completely determines the state of stress. However, under nonequilibrium conditions the internal stress in a fluid needs not be an isotropic tensor, and so why should density still depend, at any given temperature, only on pressure? Recall that pressure was defined in Section 2.1 as 1/3 of the trace of the stress tensor $\mathbf{T}$. An inquisitive mind could suspect that (a) maybe in a flowing fluid density depends on $\mathbf{T}$, and not only on p; and (b) that even if it depends only on p, it may do so in a way which needs not be the same as observed at equilibrium.

With Bob recording his protest for our losing time with such trivial questions, we proceed by first considering point (a). Consider for instance a solid which has a Poisson ratio different from 0.5. In such a solid the density at any given pressure and temperature depends on the state of stress — so why couldn't something of the same kind happen in a fluid as well? We are, however, fortunate in that no such complication does in fact arise. However, the additional step of assuming that density is independent of pressure should be taken with care.

First consider the case of liquids. These are known to be very nearly incompressible at equilibrium: in order to change the density of, say, water at room temperature by 1% one would need to increase pressure to several hundred atmospheres. So maybe we can go ahead with incompressibility. However, one should be careful: there are problems for which the assumption of incompressibility would lead to paradoxes. For instance, in a truly incompressible liquid sound would propagate at an infinitely large speed, which is nonsense. In the classical water hammer problem of interest to hydraulic engineers, pipes would inevitably burst on closing a valve if water were truly incompressible. The philosophical point to be made is that one should not think in terms of the following statement: "Water is incompressible". One should think in terms of the following much more qualified statement: "Most flow problems for water can be adequately described by the theory of incompressible fluids".

The case of gases is even more illuminating. When we calculate something like the pressure drop of a gas flowing through a circular tube, we do not take compressibility into account, and yet gases are known to be compressible. What's going on here? Well, the density of ideal gases is proportional to pressure, so that if the pressure drop is negligible as compared to the average pressure, density variations will be negligible as compared to the average density. But what if the condition is not satisfied? Well, if pressure drops are comparable to the average pressure one is approaching the speed of sound, and one should take into account the Mach number. That is normally done in Gas Dynamics, but then one usually forgets about viscosity in Gas Dynamics. [One hates to refer to thermodynamics again, but look in any standard textbook on thermodynamics the section on compressible flow, and where is viscosity? See also Section 1.5]. The point is that taking into account simultaneously both viscosity and compressibility makes the resulting equations hopelessly complex, and so in classical fluid mechanics one usually lets go of compressibility, while in Gas Dynamics one lets go of viscosity.

For incompressible fluids, the equation of continuity degenerates to:

$$\mathrm{div}\mathbf{v} = 0 \tag{4.1.2}$$

i.e., the velocity distribution is *solenoidal*.

There is however another consequence of the assumption of incompressibility. This has to do with what is called the "principle of internal constraints" in continuum mechanics. The principle states that the constitutive equation for the stress delivers it to within an arbitrary additive tensor which does no work in all deformations allowed by the internal constraints. Stated this way, the principle appears mysterious, and an example seems in order. Consider a rigid body. Here the internal constraints are that no deformation whatsoever is possible. It follows that it is impossible for internal stresses to do any work, no matter what. [Recall that work is as defined in Section 2.1]. Since any stress does no work in processes allowed by the internal constraints, the stress tensor is determined to within an arbitrary tensor — i.e., it is entirely arbitrary. And in fact internal stresses are not defined in the theory of rigid bodies (or in rigid-body motions, see Section 4.9).

Now incompressibility is an internal constraint — density cannot change. Since the isotropic part of the stress can do work only as a consequence of changes in density, it does no work in the case of an incompressible fluid. It follows that the constitutive equation for an incompressible fluid determines the stress to within an arbitrary additive isotropic tensor. While this point does not cause any major difficulty in the mechanics of Newtonian fluids, it is a very important one in the case of non-Newtonian ones

Let **T** be the stress tensor, and **g** the field acceleration vector (usually gravity). The balance of linear momentum as applied to the neighbourhood of a material point is:

$$\Phi \mathrm{D}\mathbf{v}/\mathrm{D}t = \Phi\mathbf{g} - \mathrm{div}\mathbf{T} \tag{4.1.3}$$

Equation (4.1.3) is *general*, in the sense that it applies to any material whatsoever. The discipline of fluid mechanics diverges from, say, that of elasticity in that a different constitutive equation is written for the stress tensor **T**; but at the level of generality of Eq. (4.1.3) the two disciplines have not diverged yet.

In addition to the balances of mass and of linear momentum, another balance equation is needed in isothermal continuum mechanics, namely, the balance of angular momentum. One needs at this point to make a distinction between polar and nonpolar materials. A nonpolar material is one which is not susceptible to body couples, i.e., couples acting directly on interior parts of the material without any contact forces. An example of body couple is the magnetic couple exerted by an external magnetic field (this is the couple which moves a compass needle to point north on Earth). If we are willing to exclude such effects from consideration (which of course implies we are excluding magnetofluid dynamics), the balance of angular momentum on the neighbourhood of a material point reduces to the requirement that the stress tensor is symmetric (see Appendix 4.1):

$$\mathrm{T} = \mathbf{T}^{\mathrm{T}} \tag{4.1.4}$$

Now we may come to the question of how to write the constitutive equation for stress in a fluid. First of all, we recognize that, by definition of a fluid, at equilibrium the stress tensor is isotropic, $\mathbf{T} = p\mathbf{I}$, where p is pressure. (We'll need to be a little careful about what we mean by equilibrium, but our intuitive notion of "standing still" will do for the time being). The simplest possible constitutive equation would of course be the assumption that also in a flowing fluid the stress is always isotropic. This reduces Eq. (4.1.3) to:

$$\Phi \mathrm{D}\mathbf{v}/\mathrm{D}t = \Phi\mathbf{g} - \mathrm{grad}p \tag{4.1.5}$$

Equation (4.1.5) is the momentum balance used in the classical hydrodynamic theory of ideal fluids, which is discussed in Section 4.3. Hydrodynamics of ideal fluids is too elementary a theory, unfortunately, since it essentially neglects all effects of viscosity and hence is incapable of describing dissipation of energy, as was discussed in Section 2.3. We need to allow for the possibility that stress in a flowing fluid is not necessarily isotropic. How do we distinguish mathematically the difference between a flowing fluid and one at rest? It would seem at first sight that the rest condition is simply $\mathbf{v} = 0$, but this is too simplistic: if the whole fluid moves with the same velocity $\mathbf{v}$, it is really at rest, only that we have chosen to describe its kinematics in a reference which is moving with steady translational motion with respect to the fluid. So perhaps the condition for being at rest should be $\mathrm{grad}\mathbf{v} = 0$. Thus one could be tempted to assume that the difference between the actual stress and an isotropic one is proportional to $\mathrm{grad}\mathbf{v}$, but this is not legitimate because of the requirement in Eq. (4.1.4) ($\mathrm{grad}\mathbf{v}$ is in general not symmetric). In fact, the antisymmetric part of gradv cannot contribute anything to the stress, due to principles of invariance for constitutive relations (see Appendix 4.1). Thus one concludes that a linear constitutive equation for stress should have the form in Eq. (2.3.18). [It is useful to recall that such an equation may well yield a pressure which is not the equilibrium pressure, unless the Stokes hypothesis in Eq. (2.3.20) holds true]. For incompressible fluids, the last term on the RHS of

(2.3.18) is identically zero, and the scalar p^* is not determined by the constitutive equation, which is thus written as:

$$\mathbf{T} = p\mathbf{I} - 2\mu\mathbf{D} \tag{4.1.6}$$

with the proviso that the value of the "pressure" p at any given point is not determined by the kinematics at that point (say the value of **D**), but it is to be found from the solution of the equations of motion.

Substitution of Eq. (4.1.7) into (4.1.3) yields the Navier–Stokes equations for incompressible fluids:

$$\Phi \mathrm{D}\mathbf{v}/\mathrm{D}t = \Phi\mathbf{g} - \mathrm{grad}\mathbf{p} + \mu\mathrm{lapl}\mathbf{v} \tag{4.1.7}$$

where lapl is the laplacian operator, lapl = div·grad.

Equations (4.1.2) and (4.1.7) constitute the basis of the classical theory of fluid mechanics of incompressible fluids. It should however be noticed that the constitutive equation (4.1.6) is not guaranteed to be valid; two examples have already been encountered. In Section 2.3 the Reiner Rivlin constitutive equation was discussed (Eq. 2.3.22); that is the most general form for the case where the stress tensor is a unique (though not necessarily linear) function of **D**. Many materials in nature are known to be "shear thinning", i.e., their apparent viscosity decreases with increasing rate of shear, and this can be described by Eq. (2.3.22) by taking appropriate functional relationships between the viscosity coefficients and the invariants of **D**. [Often the second viscosity coefficient is taken to be zero, and the first one is assumed to depend only on the magnitude of **D**]. However, the experimental evidence discussed in Section 1.10 cannot be described by any equation where **T** is a unique function of **D**, and indeed Eq. (1.10.7) shows that at any given value of $\partial U/\partial X$ one may have different values of F (for the problem considered in Section 1.10, $\partial U/\partial X$ is the only nonzero component of **D**, and F is a component of **T**).

4.1.1 Expansion of a liquid containing gas bubbles

Consideration of the uniform expansion of a liquid containing gas bubbles offers an interesting example not only of concepts of compressibility, but also on the possible existence of a finite value of the expansion viscosity k (corresponding to the quantity $2\mu + 3\mu'$ discussed in Section 2.3).

First consider a single gas bubble which is expanding, so that its radius R depends on time t. The flow field induced in the liquid as a whole is purely radial, and the radial velocity component is immediately obtained from the equation of continuity:

$$r^2u(r) = R^2\mathrm{d}R/\mathrm{d}t \tag{4.1.1.1}$$

The corresponding local rate of dissipation ϕ (defined exactly in the next section) is:

$$3\mu(\mathrm{d}u/\mathrm{d}r)^2 = 12\mu R^4(\mathrm{d}R/\mathrm{d}t)^2/r^6 \tag{4.1.1.2}$$

This is seen to decay very rapidly with increasing r, and hence one concludes that the whole dissipation takes place in a thin shell of fluid surrounding the bubble. The total rate of energy dissipation due to a single bubble is obtained by integrating (4.1.1.2):

$$\phi_{\mathrm{TOT}} = 16\Pi\mu R(\mathrm{d}R/\mathrm{d}t)^2 \tag{4.1.1.3}$$

Since for any given bubble the dissipation takes place essentially in a thin shell surrounding it, the rate of dissipation for a suspension of many bubbles may simply be obtained by summing up over all bubbles, provided the latter are not so closely spaced as to result in an overlap of the dissipation shells. Thus if there are N bubbles per unit volume of suspension, the dissipation rate per unit volume of suspension is, on the average:

$$\phi = 16N\Pi\mu R(\mathrm{d}R/\mathrm{d}t)^2 \tag{4.1.1.4}$$

Now suppose one regards the suspension as a homogeneous medium with an expansion viscosity k, and let Q be the rate of expansion of the suspension. Of course, in actual fact the liquid does not expand, and Q simply measures the overall effect of the bubbles expansion, say:

$$Q = N\mathrm{d}(4\Pi R^3/3)/\mathrm{d}t = 4\Pi NR^2\mathrm{d}R/\mathrm{d}t \tag{4.1.1.5}$$

If the dissipation rate is attributed to an "effective" expansion viscosity of the suspension, k, one would calculate $\phi = kQ^2$. Equating this with (4.1.1.4) one obtains:

$$k = \mu/N\Pi R^3 = 4\mu/3C \tag{4.1.1.6}$$

where C is the instantaneous volume fraction of bubbles

Equation (4.1.1.6) should be discussed in some detail. First of all, it provides a concrete example of a system endowed with a finite expansion viscosity. The fact that we identify the system as a two-phase one does not make any difference: materials can easily be envisaged which, if viewed at a sufficiently small scale, are two-phase systems. The other interesting point about Eq. (4.1.1.6) is that k is predicted to go to infinity when C goes to zero. This could be interpreted in the sense that, if there are no bubbles, the system is incompressible, and hence an infinitely large tensile stress would be needed to expand it at a finite rate. In this interpretation, one would conclude that incompressible fluids have an infinitely large expansion viscosity, which is certainly very much in contrast with the Stokes hypothesis that the expansion viscosity is zero. More interestingly, the result that k approaches ∞ when C approaches zero is removed if one includes in the model the small, but finite compressibility of the liquid (as very large tensile stresses tend to develop, the liquid will in fact expand somewhat on its own, and this brings k back to some finite value). This can therefore be seen as another example where the small compressibility of liquids needs to be taken into account.

Appendix 4.1

Consider the question of the conditions under which the isothermal assumption is a reasonable one (when the boundaries of the flow field are kept isothermal). The viscous forces are responsible for frictional heating, and their order of magnitude is seen to be $\mu V/L^2$, as will be discussed in more detail in the next section. Hence the rate at which viscous forces do work is of order $\mu V^2/L^2$, since the velocity scale V is the appropriate yardstick for the displacement per unit time. It follows that V^2/L^2 is the rate at which frictional heat is being produced per unit volume. In order to remove frictional heat, a temperature difference δT needs to develop, and the rate of heat removal per unit bounding surface is $h\delta T$, with h the heat transfer coefficient; correspondingly, the rate of removal per unit volume is $h\delta T/L = \mathbf{Nu}k\delta T/L^2$, where **Nu** is the Nusselt number and k the thermal conductivity. One can now estimate the required δT by equating the rate of frictional heating with the rate of heat removal, to obtain $\delta T \approx \mu V^2/k\mathbf{Nu}$. It is of interest to observe that δT is independent of the characteristic dimension L, and the reader would do well in trying to understand why this is so. Recalling the discussion in Section 3.3.3, the Nusselt number is at least of order unity, and thus one concludes that δT is at most $\mu V^2/k$. For water, $\mu/k = 2\times10^{-7}\ C\ \mathrm{s}^2/\mathrm{cm}^2$, and thus even with a velocity of 1 m/s one calculates $\delta T = 10^{-3}\ C$, which is certainly small enough to justify the isothermal approximation. If there is an externally imposed δT (as typical of heat transfer problems), $\mu V^2/k\delta T$ is the Brinkman number, which measures the upper bound on the relative importance of frictional heating and heat transfer; it turns out very often to be a very small number, which justifies the classical simplification of neglecting frictional heating in the energy balance, which thus becomes a linear equation for temperature, $\mathrm{D}T/\mathrm{D}t = (k/\Phi \mathrm{c})\mathrm{lapl}T$.

Next consider the balance of angular momentum. It states that the sum of all torques exerted on a body equals the rate of change of its angular momentum; at any given rate of rotation, the angular momentum is proportional to the moment of inertia, and hence to the fourth power of l, the linear dimension of the body. Body forces do not produce any torque, and contact forces (which are proportional to the surface of the body and hence to l^2) exert torques which are proportional to l^3. Body couples (such as exerted by a magnetic field), if present, are proportional to l^3. Now consider the limit where l approaches zero, which is the one we are interested in, since we want to obtain results concerning the stress tensor at a point. If there are no body couples, the balance of angular momentum reduces to the requirement that the couples exerted by contact forces sum up to zero, since a finite value proportional to l^3 could not be balanced by the rate of change of angular momentum which is of order l^4. The couples are indeed balanced if the stress tensor is symmetric, and hence one obtains Eq. (4.1.4). Should there be body couples, $\mathbf{T}$ would not need to be symmetric; its antisymmetric part $\mathbf{T}^-$ would then balance the effect of the body couples.

The fact that the antisymmetric part of grad$\mathbf{v}$ cannot contribute anything to the stress is a subtle point. If we consider two flow fields which differ only by a superimposed rigid body rotation, one could regard them as the same flow field described in two frames of reference which rotate with respect to each other. Now for instance the balance of momentum is not invariant under such a change of frame, but only for a Galileian change (the two frames only translate at constant speed with respect to each other in a Galileian change). However, the constitutive equation has to be invariant under an arbitrary change of frame, and hence the antisymmetric part of grad$\mathbf{v}$ cannot enter it. It is difficult to derive this principle of invariance rigorously, and a detailed discussion is postponed to Section 4.9.

As far as the question of incompressibility is concerned, there are in fact five conditions which need to be met in order to justify the assumption of incompressibility. These conditions are discussed in detail by Batchelor [1967]. Let c be the speed of sound, β the coefficient of cubic expansion, θ the largest temperature difference in the fluid, c_P its specific heat, and k what Batchelor calls the "expansion damping", corresponding to the quantity $2\mu + 3\mu'$ discussed in Section 2.3. The five conditions are expressible in terms of a characteristic velocity for the flow V, a characteristic length scale L, and, for unsteady flows, a characteristic frequency n. The five conditions are as follows:

$$V^2/c^2 << 1 \qquad (4.1.a.1)$$

i.e., the Mach number must be significantly less than unity. This is violated in Gas Dynamics, where the assumption of incompressibility is not made

$$n^2L^2/c^2 << 1 \qquad (4.1.a.2)$$

This is obviously violated in acoustics, where the wave length L is c/n.

$$gL/c^2 << 1 \qquad (4.1.a.3)$$

which implies that the length scale is less than the scale of density variations due to gravity. This is violated in dynamic meteorology, since c^2/g is of the order of 8 km in air.

$$\beta\mu V/c_P\Phi L << 1 \qquad (4.1.a.4)$$

implying that the expansion due to frictional heating is negligible; this condition is practically always satisfied.

$$\beta\theta k/LV << 1 \qquad (4.1.a.5)$$

which could be violated in some extreme case in gases.

4.2 DIMENSIONAL ANALYSIS AND CLASSIFICATION OF FLOWS

The Navier–Stokes equations contain three-dimensional parameters: viscosity μ, density Φ, and gravity **g**. Additional parameters are of course present for any given problem because of the boundary conditions, and one can easily envisage three possible ones: a velocity scale V, a length scale L, and, for boundary conditions which are not steady, a time scale T. (Of course, one could have an external force or pressure difference scale, but in such cases there would be no external velocity scale). This would result in an inconveniently large number of dimensionless groups (3), and some possible simplification is in order.

First, one may wish to restrict attention to the case where the boundary conditions are steady ones. This guarantees that there is no externally imposed time scale, though it does not guarantee that the flow field will be a steady one (turbulent flows are of course the classical example of an unsteady flow field developing with steady boundary conditions). Another very important simplification arises because in a large variety of problems gravity can be eliminated from the governing equations. This is accomplished as follows. In hydrostatics, Eq. (4.1.8) reduces to:

$$\mathrm{grad}p_0 = \Phi\mathbf{g} \tag{4.2.1}$$

where the subscript 0 reminds us that the hydrostatic distribution is being considered. Now one can define a reduced pressure P as $p - p_0$, i.e., the difference between the pressure at a point and the pressure one would have at that same point in hydrostatics. With this definition, Eq. (4.2.8) reduces to:

$$\Phi\mathrm{D}\mathbf{v}/\mathrm{D}t = -\mathrm{grad}P + \mu\mathrm{lapl}\mathbf{v} \tag{4.2.2}$$

which contains no gravity term, and the dimensional parameter g seems to have been neatly disposed of.

The class is not convinced, and rightly so. “This is a dirty trick — you can’t make gravity irrelevant by introducing a definition” is the gist of the several observations that are made. Well, but where has gravity disappeared to? Sue suggests that it has, perhaps, ended up in the boundary conditions. Very good point. Will it inevitably do so? No; it will only do so if the original boundary conditions were somehow imposed on p, and when we transform to BCs for P gravity will appear. Two examples may clarify the issue.

Let's first consider motion of a wave in the sea. Let the wave surface be located at a height $y = A\cos(x/L)$ over the average sea level, A being the height of the wave, L its length, and x the horizontal coordinate across the wave. Now one of the boundary conditions will be that at $y = A\cos(x/L)$, $p = 1$ atm, which translates to $P = 1$ atm $+ \Phi gy$, and g has indeed appeared in the boundary conditions. Now consider flow in a vertical tube. Suppose there are two pressure gauges located at a vertical distance of 10 meters from each other, and consider the following four possible couples of readings of p:

	1	2	3	4
high	1.2 atm	2.2 atm	1.0 atm	0.8 atm
low	2.0 atm	3.0 atm	2.0 atm	2.0 atm

Does anybody care to guess whether water is flowing up, down, or not at all? Agreement is immediately reached that water is not flowing in case 3 (the two readings give the hydrostatic pressure difference). It is also readily agreed that water flows downhill in cases 1 and 2 (pressure on the high side exceeds the hydrostatic value), and uphill in case 4. It takes some little time to convince the class that the flow-rate will be the same in cases 1, 2 and 4, because the difference of P values is the same — 0.2 atm. So this is a problem where the boundary conditions are in fact assigned on P, not p, and gravity plays no role. Ralph disagrees: he says that, given the same readings on the moon, he would come to different conclusions. Given the same readings of p, yes; but let's say that on the moon 10 meters correspond to 0.2 atm. Let's now take a high reading of 1.2, and a low reading of 1.2 as well. The δp is now zero, but δP is 0.2 and pushing down — wouldn't we expect the same flow-rate as in cases 1 and 2 on the Earth? So indeed gravity does not play any role in this problem.

In fact it is easy to recognize that gravity will enter the boundary conditions when there are *free surfaces*, where the liquid considered is in contact with a constant pressure gas. For *confined* flows, i.e., flows without free surfaces, gravity does not play any role, and we are reduced to only four parameters and one dimensionless group, the Reynolds number $LV\Phi/\mu$. If gravity does play a role, the additional dimensionless group is the Froude number, $\mathbf{Fr} = V^2/Lg$, and for unsteady boundary conditions one has also the Strouhal number $\mathbf{Sr} = VT/L$.

Let's try to understand what these dimensionless groups stand for. Let's begin with the case where there is no g and no $\mathbf{T}$ to worry about. If the velocity scale has been chosen appropriately, the dimensionless velocity $\mathbf{v}^* = \mathbf{v}/V$ will be of order unity. Let's suppose the flow is steady — the boundary conditions are, after all. Therefore $D/Dt = \mathrm{v}\cdot\mathrm{grad}$, and hence has an order of magnitude of V/L if we have chosen the length scale L appropriately. (One should be careful here: if **v** and the grad operator are orthogonal to each other, $D/Dt = 0$ — this happens in laminar flows). So $D^*/Dt = (L/V)D/Dt$ is of order unity. By the same token, $\mathrm{grad}^* = L\mathrm{grad}$ and $\mathrm{lapl}^* = L^2\mathrm{lapl}$ are also operators of order unity. (The order of magnitude of an operator is the ratio of the orders of magnitude of the value and of the argument of the operator). The equations of motion reduce to:

$$(\Phi V^2/L)D^*\mathbf{v}^*/Dt = -\mathrm{grad}^*(P/L) + (\mu V/L^2)\mathrm{lapl}^*\mathbf{v}^* \tag{4.2.3}$$

The analysis above guarantees that the order of magnitude of the LHS (the inertia force) is $\Phi V^2/L$, and the order of magnitude of the second term on the RHS (the viscous force) is $\mu V/L^2$. The order of magnitude of the pressure gradient will adjust itself so as to balance the combined effect of viscous and inertia forces. We now recognize the Reynolds number as the ratio of inertia to viscous forces. In fact, if (4.2.3) is divided throughout by $\Phi V^2/L$ one obtains:

$$D^*\mathbf{v}^*/Dt = -\mathrm{grad}^*(P/\Phi V^2) + (1/\mathbf{Re})\mathrm{lapl}^*\mathbf{v}^* \tag{4.2.4}$$

Should there be an external time scale T, the order of magnitude of $\Phi\partial\mathbf{v}/\partial t$ would be estimated as $\Phi V/T$, and hence the Strouhal number is interpreted as the ratio of the unsteady to the convective inertia force. Should g play any role, the $\Phi\mathbf{g}$ term would be estimated at an order of Φg, and hence the Froude number is interpreted as the ratio of inertia to gravity forces. Finally, should surface tension σ play a role, the surface tension force per unit volume is estimated at σ/L^2. The ratios of the latter to other forces of importance produce different capillary numbers: the ratio of surface to viscous forces is $Ca_1 = \sigma/\mu V$; surface to inertia forces is $Ca_2 = \sigma/\Phi L V^2$; etc. Some examples are discussed in Appendix 4.2.

We are now in a position to discuss the classification of flow fields of Newtonian fluids. First, we restrict attention to the case where the boundary conditions imposed are steady ones (a few examples of unsteady boundary conditions have been discussed in Sections 1.4 and 3.2.5). Now with steady boundary conditions one may look for a steady solution of the Navier–Stokes equations. Since there is no theorem of uniqueness of solutions for the equations, having found a steady one does not guarantee that it will be observed in nature, since it may well be unstable. This leads to the question of instability, which is very concisely discussed below.

Suppose one writes the Navier–Stokes equations with $D\mathbf{v}/DT$ set at $\mathbf{v}\cdot\text{grad}\mathbf{v}$ at the start; these will be called the steady NS equations. For a problem with steady BCs, it is possible to find a solution to the steady NS equations, and such a solution is guaranteed to be steady, since time t never appears anywhere in the equations; hence the solution one has found is also a solution of the complete NS equations. However, it could be an unstable solution, in the sense that, should the velocity field undergo an even infinitesimally small disturbance from the solution one has found, it would not tend to evolve back to it. This means that a stability analysis would be in order, and a stability analysis is necessarily based on the complete NS equations, since one is asking oneself what happens to a disturbance as time progresses, a question which is of course meaningless if $\partial\mathbf{v}/\partial t$ is set to zero. The classical example of an unstable solution is of course that of laminar flow in a tube, which is known to go unstable at sufficiently high Reynolds numbers. Instability and turbulence will be discussed in Section 4.7.

Once attention is restricted to steady solutions, a first class of flow fields is that of *laminar* flows. We will use the word “laminar” in a restricted sense: a laminar flow is an *exact* solution of the NS equations which satisfies all BCs and yields an inertia force which is identically zero. It is easy to convince oneself that laminar flows are possible only if all streamlines are parallel straight lines. Should a streamline not be straight, there would be a centrifugal acceleration, and hence the flow would not be laminar. Should straight streamlines not be parallel, constancy of flow-rate along a stream tube would imply that the velocity is changing, and thus there would be a finite convective inertia. Laminar flow solutions, as will be seen in Section 4.4, at most require solution of the two-dimensional Laplace equation.

There are very few problems, in addition to laminar flows, for which an exact solution to the NS equations can be found, and some will also be discussed in Section 4.4. For the great majority of problems, one can only look for an approximate solution, and this is done by trying to find the *exact* solution of *approximate* equations — a technique which is not guaranteed to yield what one is really looking for, i.e., an *approximate* solution to the *exact* equations. The presence of the Reynolds number in the NS equations allows a fundamental classification in two classes of flows. If $\mathbf{Re} << 1$, inertia forces are (presumably) negligible, and one thus looks for a solution to the *creeping flow* equations:

$$\mathrm{grad}P = \mu\mathrm{lapl}\mathbf{v} \tag{4.2.5}$$

Of course, a creeping flow solution will *not* have a zero inertia force, since there is nothing in Eq. (4.2.5) to guarantee that its solutions are such that $\mathrm{D}\mathbf{v}/\mathrm{D}t = 0$. However, the order of magnitude of the inertia force calculated from a creeping flow solution should be negligible as compared to that of the viscous and pressure forces (if $\mathbf{Re} << 1$). As will be seen in Section 4.5, this is not generally true, but in most cases there are no serious problems arising. Since Eq. (4.2.5) is linear, solutions can in general be found, at least by numerical techniques.

The converse case is of course the one where $\mathbf{Re} >> 1$, when viscous forces are negligible and thus one can write the Euler equations of motion:

$$\Phi\mathrm{D}\mathbf{v}/\mathrm{D}t = -\mathrm{grad}P \tag{4.2.6}$$

However, very serious problems arise in this case. Since the viscous force term has been dropped from the equations, their order has been lowered — there are no second derivatives in Eq. (4.2.6). This shows that the $1/\mathbf{Re} = 0$ limit is a singular one. In particular, not all the original BCs can be imposed on the Euler equations, and in finding solutions to them one lets go of the no-slip condition (one cannot let go of the condition that the velocity component normal to the bounding surfaces is zero, because otherwise one would lose the problem altogether). However, this means that the solution of the Euler equations is at best an "outer" solution to the NS ones, and a boundary layer theory needs to be developed for the region near the bounding surfaces where the no-slip condition is violated by the solution of the Euler equations. The Euler equations (which are nonlinear) will be discussed in Section 4.3, while boundary layers are discussed in Section 4.6.

There is a final conclusion of a general character to be discussed here. Equation (2.3.17) shows that for isothermal flow of an incompressible fluid $\mathbf{T}:\mathbf{D} \leq 0$. The term $p\mathbf{I}$ in Eq. (4.1.6) does not contribute anything to $\mathbf{T}:\mathbf{D}$, and hence one concludes that $2\mu\mathbf{D}:\mathbf{D} = \phi = 0$. The quantity ϕ is the rate of energy dissipation, and on occasion results can be obtained by calculating the overall rate of dissipation in the whole body of fluid. That ϕ is indeed the rate of energy dissipation can be seen also from a different argument. Since we have seen in Section 2.3 that free energy cannot depend on **D** (if Eq. (2.3.13) applies, as it certainly does in the case at hand), we also conclude that internal energy U and entropy S cannot depend on **D** (U and S are here expressed per unit mass). Hence the classical Maxwell relation $T\mathrm{d}S = \mathrm{d}U + p\mathrm{d}V$ applies, and this reduces to $T\mathrm{d}S = \mathrm{d}U$ for an incompressible fluid. Now the balance of energy can be written as:

$$\Phi \mathrm{D}U/\mathrm{D}t = -\,\mathrm{div}\mathbf{q} + \varnothing \tag{4.2.7}$$

where $\mathbf{q}$ is the conductive heat flux vector, and $-\mathrm{div}\mathbf{q}$ represents the local influx of heat per unit volume, q. Notice that Eq. (4.2.7) is general, i.e., no constitutive assumptions are made in deriving it. But if the relation $T\mathrm{d}S = \mathrm{d}U$ is used (and this is now based on the constitutive assumption in Eq. (2.3.13)), one obtains:

$$\mathrm{D}S/\mathrm{D}t = q/T + \phi/T \tag{4.2.8}$$

Since the second law requires, for a locally isothermal system ($\mathrm{grad}T = 0$), that $\mathrm{D}S/\mathrm{D}t = q/T$, the quantity ϕ/T is recognized as the local rate of entropy generation, and hence ϕ as the local rate of energy dissipation.

Appendix 4.2

Consider the motion of a wave on the sea. Let U be the celerity of the wave — the speed at which its trough moves (it is called a celerity because it is not the velocity of any real mass: the water simply moves up and down as the wave passes through, but its horizontal velocity is essentially zero). As long as the water depth is large compared with the wave height, what the depth may be becomes irrelevant, and the Reynolds number is large enough to make viscous forces irrelevant. It follows that the only forces of importance are gravity and inertia forces, so that the Froude number U^2/Lg must have some fixed value of the order of unity. That in turns implies that the celerity is $U \approx \sqrt{(Lg)}$. Larger waves do indeed move faster than small ones — they appear to slide under the latter. Now consider a "floating" boat (a floating boat is one whose weight is sustained by gravity also when it moves). Such a boat sheds waves behind it, and if one looks at them from the boat they appear to be stationary, so that their celerity is proportional to the speed of the boat, and their length to the length of the boat. One thus concludes that the boat speed V is given by something like $\beta\sqrt{L_B g}$, with β a coefficient which may perhaps depend slightly on the shape of the boat, but is essentially restricted to be significantly less than unity, say to a value of 0.1. With this, one calculates a speed of 7 knots for a 24-ft boat — not a bad estimate at all. For a 600-ft ship one estimates 35 knots, again not a bad estimate. Those are the maximum speeds that a floating boat can reach, no matter how large the power of the engine. If the engine is oversized, the boat will just get out of its natural horizontal alignment, its effective length will decrease, and its speed will decrease as well. Have you ever noticed small boats with an oversized outboard engine moving at an awkward angle and actually slower than similar boats with a smaller engine? The restrictions discussed do not apply to flying boats, which have a keel so designed as to give them lift at sufficiently high speed, so that their weight is partly sustained by hydrodynamic lift. Such boats shed waves which may have lengths comparable to that of the boat itself, and this gives a factor of about 3 or even more, so that speeds of 20 knots or more are possible with a 24-ft flying boat. However, the engine needs to have considerably more power — 30 HP is quite reasonable for a floating 24-ft boat, but 250 HP are just barely enough for a flying 24-ft boat.

In the vibrating sphere problem discussed in Section 3.2.5, $1/\Omega$ was of course the external time scale. Since the velocity scale was $\Omega\delta$, the Strouhal number was L/δ — its magnitude was not clear at the start of the analysis. The final result suggested that the Strouhal number was of order D/δ, i.e, very large for the conditions of the problem: the unsteady inertia force dominated over the convective one.

Consider the shape vibration of a 0.1 cm diameter water droplet in air. Surface, inertia and viscous forces may play a role. First assume that the vibration is dominated by the interplay of surface and viscous forces. One would conclude that $\mathbf{Ca}_1 = \sigma/\mu V \approx 1$, which yields an estimate for V of the order of 10^4 cm/s, and a vibration period of the order 10^{-5} s. However, one now calculates a Reynolds number of 10^5, which is in contrast with our assumption that viscous forces are more important that inertia ones. So we conclude that capillary forces are balanced by inertia forces, and thus that $\mathbf{Ca}_2 = \sigma/L\Phi V^2 \approx 1$, which yields an estimate for the velocity of the order of 100 cm/s, and a period of oscillation of 10^{-3} s. The Reynolds number is of the order of 1,000, consistent with our assumption.

4.3 HYDRODYNAMICS OF IDEAL FLUIDS

In this section, we discuss the hydrodynamics of incompressible ideal fluids. Of course, today we know that the Euler equations of motion are obtained from the Navier–Stokes equations if one is willing to neglect the viscous forces term $\mu \mathrm{lapl}\mathbf{v}$, i.e., for sufficiently large Reynolds numbers. We also know that the $1/\mathbf{Re} = 0$ limit is a singular one, and that therefore a boundary layer analysis is required. However, the theory of hydrodynamics was developed much before the concept of internal stresses, let alone that of viscous stresses, was available; the body of the theory rests on the work of scientists working in the second half of the eighteenth century, notably Euler and Bernoulli. At those times, the theory was regarded as *the* theory of fluid dynamics, not as an asymptotic limit of a more general theory, be it a regular or, as we know today, a singular one. The theory was based on the simplest possible constitutive assumption. Since it was known that in hydrostatics (which was developed in the second century BC by Archimedes) a fluid could only exhibit an isotropic internal stress (i.e., it only exhibits a pressure, but can sustain no tangential stresses and no anisotropic normal stresses), the assumption was simply that such was the case also for a fluid in flow. It is useful to consider hydrodynamics of ideal fluids from such a historical viewpoint: in the second half of the eighteenth century, there was no thermodynamics whatsoever available (the very word ‘energy’ had not been invented), the only kind of internal stress which was understood was a state of isotropic pressure, and the only mechanical theory available was that of rigid bodies and mass points. It was on such a restricted basis that classical hydrodynamics was developed.

There is also another reason for dealing with classical hydrodynamics as an independent theory standing on its own feet. While it is true that the Euler equations are obtained from the Navier–Stokes ones by setting $1/\mathbf{Re}$ to zero, they can also be viewed in a more general context. Given a constitutive equation other than Newton’s linear law of friction, one could always in principle write down the equations of motion, and identify in these one or more terms which represent the contribution of the fact that internal stresses are different from an isotropic pressure. It is also possible to envisage situations where one convinces oneself that such terms are negligibly small as compared to those arising from inertia and pressure, and thus one is willing to neglect them — the result being, of course, again that inertia and pressure forces balance each other, i.e., one would again obtain classical hydrodynamics.

The Euler equations of motion are of course:

$$\Phi D\mathbf{v}/Dt = -\mathrm{grad}P \tag{4.3.1}$$

where P is the reduced pressure, i.e., the difference between the actual pressure at a point and the pressure one would have at that same point should the pressure distribution be hydrostatic:

$$P = p - p_o + \Phi gh \tag{4.3.2}$$

where h is height over a reference horizontal plane where the actual pressure is p_o. Since only the gradient of P appears in Eq. (4.3.1), one can always set $\mathrm{p_o}$ to zero, and we will do so in the following.

The two most fundamental results in classical hydrodynamics were obtained by applying to fluids techniques which had worked well in the mechanics of rigid bodies. The first one is as follows. For a rigid body, the first law of mechanics can be written as:

$$\mathbf{F} = m\mathbf{a} = m\mathrm{d}\mathbf{v}/\mathrm{d}t \tag{4.3.3}$$

where $\mathbf{F}$ is the (vectorial) sum of all forces acting on the body of mass m, $\mathbf{a}$ is its acceleration, and $\mathbf{v}$ is the velocity of its centre of mass. Now the concepts of work and of kinetic energy were available at the times (the latter was called living force, since the word energy did not exist). Work is force times displacement, and if the body is in translational motion (no rotation), all its points move with the same velocity (because of its rigidity), and hence the total work done per unit time is simply the scalar product $\mathbf{F}\cdot\mathbf{v}$. Thus from Eq. (4.3.3) one calculates:

$$\mathbf{F}\cdot\mathbf{v} = m\mathbf{v}\cdot\mathrm{d}\mathbf{v}/\mathrm{d}t = \mathrm{d}(\tfrac{1}{2}mv^2)/\mathrm{d}t \tag{4.3.4}$$

i.e., the total work done by all forces per unit time is equal to the rate of change of kinetic energy. Of course Eq. (4.3.4), being a scalar equation, contains less information that Eq. (4.3.3) from which it is derived; however, the result was a very useful one in the mechanical theory of rigid bodies. Incidentally, Eq. (4.3.4) shows that rigid bodies are thermodynamically degenerate, since work cannot be converted to internal energy.

If the same algorithm is applied to the Euler equations, one obtains:

$$\Phi v \cdot \mathrm{D}\mathbf{v}/\mathrm{D}t = -\,\mathbf{v} \cdot \mathrm{grad}P \tag{4.3.5}$$

The kinetic energy per unit volume, K, is obviously $\frac{1}{2}\Phi\mathbf{v}^2$, and little algebra is needed to show that the LHS of (4.3.5) is the substantial derivative of K, say:

$$\mathrm{D}K/\mathrm{D}t = -\mathbf{v} \cdot \mathrm{grad}P \tag{4.3.6}$$

so that the same result is obtained again — the rate of work done equals the rate of accumulation of kinetic energy. Of course, today we may interpret this result from the vantage point of thermodynamics: since we have left out viscous forces, there is no mechanism of dissipation, and any work done on the system is accumulated in the form of kinetic energy, i.e., in a form where it can be converted back to useful work without any loss.

If one now restricts attention to steady flows ($\partial\mathbf{v}/\partial t = 0$), $\mathrm{D}K/\mathrm{D}t = \mathbf{v} \cdot \mathrm{grad}K$, and Eq. 4.3.6 becomes:

$$v \cdot \mathrm{grad}(P + \tfrac{1}{2}\Phi v^2) = 0 \tag{4.3.7}$$

which is the celebrated Bernoulli equation. The important point to be made here is that it is a straightforward consequence of Euler's equations of motion, nothing else but a consequence of the principle of conservation of linear momentum.

Equation (4.3.7) implies that the gradient of $P + \frac{1}{2}\Phi v^2$, if nonzero, is orthogonal to the velocity vector, and hence is everywhere orthogonal to the trajectories (or streamlines). It follows that $P + \frac{1}{2}\Phi v^2$ is constant along every streamline. This is a very useful result, and many problems can be solved easily and quickly based on this result. It does, however, and unfortunately, have a misleading scent of a thermodynamic result. $p + \Phi gh + \frac{1}{2}\Phi v^2$ seems, in some sense, to be an energy content: kinetic, potential, and pressure energy (whatever the latter may mean). There is no thermodynamic content whatsoever in the Bernoulli equation. $p + \Phi gh + \frac{1}{2}\Phi v^2$ may, if anything, be interpreted as *mechanical* energy (it would be better to say that it may be *defined* as mechanical energy) per unit volume. There is, of course, no principle of conservation of mechanical energy, and hence the fact that it happens to be constant along streamlines is simply a mechanical result, which only a posteriori can be justified and interpreted by considering that, since the Euler equations don't incorporate any mechanism of dissipation, they predict the conservation of mechanical energy.

The Bernoulli equation has less information than the Euler equations, since it is a scalar equation. There is another transformation of the Euler equation which does not result in a loss of information — taking the curl of both sides of Eq. (4.3.1). Again, this was a natural for eighteenth century scientists. In rigid body mechanics, rotations are handled by using the curl operator. Furthermore, it was known already that the curl of the gradient is zero, and hence that taking the curl of both sides of Eq. (4.3.1) would result in the elimination of pressure, say that:

$$\mathrm{curl}(\mathrm{D}\mathbf{v}/\mathrm{D}t) = 0 \tag{4.3.8}$$

which constitutes a vectorial equation for velocity which, together with the continuity equation, is in principle sufficient to determine the kinematics of motion. Once those are obtained, the reduced pressure distribution can be obtained from Eq. (4.3.1), with the LHS a known quantity and the RHS being linear.

It is now a matter of commuting the D/Dt and the curl operators to have an equation for the rotation vector $\mathbf{w} = \mathrm{curl}\mathbf{v}$. The result is:

$$\mathrm{D}\mathbf{w}/\mathrm{D}t = \mathbf{w}\cdot\mathrm{grad}\mathbf{v} \tag{4.3.9}$$

Equation (4.3.9) has a very important consequence: if the rotation is zero at one point of a streamline, it stays at zero along the whole streamline. (Rigorous proof that this result follows from Eq. (4.3.9) is somewhat more involved than discussed here). It follows that, if a flow field is known to be such that presumably all trajectories originate at $-\infty$ where the rotation is zero, the rotation will be zero everywhere. This situation is encountered, for instance, for flow around a submerged object.

There exists therefore a large class of pragmatically significant solutions of the Euler equations which are *irrotational*. For an irrotational flow field, the curl of the velocity vector is zero, and hence the velocity can be expressed as the gradient of some scalar $\varnothing$: the velocity vector admits a potential. (The fact that $\mathrm{curlgrad}\varnothing = 0$ for every scalar field $\varnothing$ is obvious: its proof lies on the fact that the order of differentiation is irrelevant. The converse proof, i.e., that a vector field which has a curl of zero everywhere is always expressible as the gradient of a scalar is not so trivial).

Now consider the special case of plane flows, where there are only two nonzero components of velocity (say the x and y ones), and all derivatives with respect to the third direction (say z) are zero, and suppose that the flow considered is irrotational. Therefore, the velocity vector admits a potential ϕ, and the continuity equation implies the existence of a "stream function" $Q(x,y)$ defined below. (Notice that all two-dimensional flows, be they irrotational or not, allow the definition of a stream function. Also notice that the analysis given below in essence applies also to flows which are bidimensional in some coordinate system which needs not be Cartesian). Thus one has, with u and v the x and y components of velocity:

$$u = \partial\phi/\partial x = \partial Q/\partial y \tag{4.3.10}$$

$$v = \partial\phi/\partial y = -\partial Q/\partial x \tag{4.3.11}$$

The conditions on $\varnothing$ and Q in Eqs. (4.3.10–11) are the Cauchy Riemann conditions for a function of complex variable to be analytic. Specifically, let w and z be two complex variables, $z = x + iy$, $w = \phi + iQ$, and let $w = W(z)$ — this of course means that ϕ and Q are functions of x and y. The function $W()$ is analytic (i.e., it is differentiable) if ϕ and Q satisfy Eqs. (4.3.10–11). What this means in practice is that it is extremely easy to construct two-dimensional irrotational solutions of the Euler equations. All one needs to do is to write an analytic function $W(z)$, which is very easy indeed: anything like $w = z^n$, $w = tgz$, whatever, will do (it is much harder to think of a nonanalytic function). Once one has written such a function, little algebra will produce ϕ and Q, and hence a velocity field which is guaranteed to be irrotational and to satisfy the continuity equation. It is a very strange method of doing things: it is easy to construct solutions, but one does not know what is the problem one for which one is constructing the solution. However, given 200 years of mathematicians playing this game, it is no surprise that a very large inventory of solutions of significant problems has been constructed this way.

There is however another important point to consider. For irrotational flows, the governing equations are.

$$\mathrm{div}\mathbf{v} = 0 \tag{4.3.12}$$

$$\mathrm{curl}\mathbf{v} = 0 \tag{4.3.13}$$

These are both linear, and therefore a variety of mathematical techniques are available for their solution. Once a solution has been obtained, the pressure distribution is calculated from the Euler equations, where now the nonlinear term is known. Since Eqs. (4.3.12–13) are linear, solutions can be superimposed, and therefore a whole new variety of solutions can be found. Notice that, while the velocity fields can be superimposed, the pressure fields cannot, but this poses no great difficulty, since the sum of the two velocity fields can be inserted into the Euler equation to calculate the pressure distribution.

Now let us consider the Euler equations from the vantage point of knowing what viscosity is. By dropping the term expressing the action of viscous forces, we have dropped the mechanism for dissipation of energy — the Euler equations describe reversible (in the thermodynamic sense) flows. Indeed, we have seen that the work done is always accumulated as kinetic energy. It follows that the Euler equations cannot describe any dissipative phenomenon. The most important example is that of flow past a submerged body. If the object is not symmetrical with respect to the direction of approach of the fluid, we expect the force which the fluid exerts on the body not to be in the direction of flow, and hence we can think of decomposing that force into a component in the direction of flow, which we'll call the drag force, and an orthogonal component, which is called the lift (since it is the component responsible for vertical support of an airfoil). Now the lift does not do any work, since it is orthogonal to the flow direction, but the drag does work, and it is dissipative work. It follows that the Euler equations will always predict a drag of zero, no matter what the shape of the submerged body may be. This is known as the D'Alembert paradox, and it puzzled no end scientists in the early 19th century, since of course the existence of a finite drag force was well known from experiments. The Euler equations do quite well in predicting the lift, though. An aeronautical engineer can design the shape of the wings based on the Euler equations, but as far as the engine is concerned, the Euler equations would tell him that, in a steady horizontal course, the pilot might as well turn off the engine — in fact he should do that, if he doesn't want his plane to accelerate all the time (work done results in the accumulation of kinetic energy). Of course our readers — or at least some of them — may at this point ask themselves why we are making this point again — after all, it was already discussed in Section 2.3 (see Fig. 2.3.1). Well, we have to confess that we are afraid some of our readers may have just decided to skip the chapter on Thermodynamics entirely, and we dare not be critical of such readers, should there be any.

Now suppose one considers a problem of irrotational flow around a submerged object, for which the solution to the Euler equations is available. However, we now wish to solve D'Alembert's paradox by pointing out that we need a solution to the Navier–Stokes equations, not the Euler equations. We are having our discussion in the late nineteenth century with a respected hydrodynamicist who doesn't believe this idea of viscosity anyhow; he is (sexist grammar comes in here, but let's face it, there were no female hydrodynamicists in the late nineteenth century), however, an open-minded person and is willing to follow our argument. He is also a good mathematician, and therefore he is quick to point out a problem. The difference between the Euler and the Navier Stokes equations is the existence of a term containing the Laplacian of the velocity. But, he points out, the Laplacian of the velocity lapl**v** is also zero (the divergence of the transpose of the velocity gradient is always zero; hence if the velocity gradient is antisymmetrical, as it must be in an irrotational flow, the Laplacian is zero). It follows that *an irrotational solution of the Euler equations is also a solution of the Navier Stokes equations*. Now it's an 1890 vintage Ralph on our side, and it takes our hydrodynamicist painstaking efforts to convince Ralph that he is right. Indeed the available solution to the Euler equations is also a solution of the Navier–Stokes equations, and what does Ralph want with his viscosity? There might be such a monstrous thing, but it isn't going to solve D'Alembert's paradox.

Our 1890 Ralph ponders on it a bit, and finally he comes up with an answer. The flow field is right, OK; but the calculation of the drag is not. The net effect of the pressure distribution on the submerged object is indeed zero, but with viscosity one needs to also include the net effect of tangential forces, and you hydrodynamicists didn't do that. (Poor Ralph. He is painting himself into a corner — he is giving in to the idea that there is no such thing as form drag, only skin drag. Scores of engineers working in the automobile industry to design the right hydrodynamic shape of automobiles so the Madison Avenue guys can use a low value of C in their TV spots are risking their jobs, should Ralph be right). The hydrodynamicist hadn't thought of that, but he proceeds to a quick calculation to show that, at least in the case of a sphere, the net effect of the tangential stresses as calculated from his flow field is also zero. Furthermore, he produces an order of magnitude estimate to show that even in cases where it could be nonzero — and he suspects there are no such cases — it would be way too small to account for experimental evidence. After all, the tangential velocity gradient at the solid boundary coming out from his solution is always quite small. Ralph is annoyed at this argument; the hydrodynamicist is supposed to be a good mathematician, not somebody who comes up with order of magnitude arguments which are hardly convincing anyhow.

Fortunately enough, at the beginning of this century there was a Sue who got interested in this sort of things; it was a male Sue with the last name of Prandtl. He came up with a brilliant idea: the solution we have is a solution of the Navier– Stokes equations, but it doesn't satisfy the no-slip condition at the solid boundary — one can't impose that condition on the Euler equations, since there are no second derivatives — the Euler equations are of lower order than the Navier– Stokes equations. So, he reasoned, we need a boundary layer analysis — it's a singular limit, isn't it? Being a Sue type, he made a horrible mess of the actual analysis (Ralph was absolutely flabbergasted at his hand-waving arguments concerning orders of magnitude, though Bob followed them with sympathy if not with understanding), but he did hit on the right point. The essence of the physical argument is that the viscous forces predominate necessarily at a solid bounding surface, since on that surface the no-slip condition implies that the inertia forces are zero. Viscous forces generate rotation, and hence the motion ceases to be irrotational. When the rotation diffuses away from the solid surface, viscous forces may become negligible again, but the flow is not any more an irrotational one, even if it is still governed by the Euler equations. These points will be discussed in detail in the section on boundary layers.

Appendix 4.3

Hydrodynamic theory is heavily based on an accurate and rigorous mathematical description of the kinematics of motion, and we have been a bit cavalier about that. In this appendix, we try to offer a glimpse of what is involved.

Let a material point (to have a concrete idea in mind: one particular molecule) be identified by some label **X**, which may conveniently be chosen as the position it occupies at some time $t = 0$. At any other time t, that same material point will in general occupy a different position **Y**, which depends on time itself, as well as on the particular material point considered:

$$\mathbf{Y} = \mathbf{f}(\mathbf{X},t);\ \mathbf{f}(\mathbf{X},0) = \mathbf{X} \qquad (4.3.a.1)$$

Now **f**() describes the trajectory of the material point **X**, and its velocity is:

$$\mathbf{v}(\mathbf{X}) = \partial \mathbf{f}/\partial \qquad (4.3.a.2)$$

while the substantial derivative of velocity is:

$$\mathrm{D}\mathbf{v}/\mathrm{D}t = \partial^2\mathbf{f}/\partial t^2 \tag{4.3.a.3}$$

The velocity gradient grad**v** is $\partial\mathbf{v}/\partial\mathbf{Y}$ (remember that **Y** is the position measured in the frame of reference). If **F** is defined as $\partial\mathbf{Y}/\partial\mathbf{X}$ (this is called the deformation gradient), we have:

$$\mathrm{grad}\mathbf{v} = \mathbf{F}^{\mathrm{T}} \cdot \partial\mathbf{f}/\partial\mathbf{X} \tag{4.3.a.4}$$

Now it is possible to decompose grad**v** in its symmetric part $\mathbf{D} = \frac{1}{2}(\mathrm{grad}\mathbf{v} + \mathrm{grad}\mathbf{v}^{\mathrm{T}})$ (which is called the rate of deformation tensor), and $\mathbf{W} = \frac{1}{2}(\mathrm{grad}\mathbf{v} - \mathrm{grad}\mathbf{v}^{\mathrm{T}})$, the latter being called the vorticity tensor. In a rigid body motion, $\mathbf{D} = 0$, which justifies its name. Now notice that **W** is antisymmetrical, and is therefore completely identified by three components (say the ones above the main diagonal). Those components are the same as those of the rotation vector **w**.

So far, it has all been very formal. Now let's try to understand what the rotation vector **w** really means, and let's consider plane flows to make things easier. Is it possible to travel a curved trajectory with a rotation of zero? Well, let's ask a more drastic question: is it possible to travel a closed trajectory — one that eventually leads back to its own starting point — with zero rotation? Well, think of a very devout Moslem who always walks looking towards the Mecca. Can he circle a fountain doing that? The answer is obviously yes: he'll find himself in the awkward condition of having to walk backwards at times, but he can do it. Now if it is possible to travel a closed path without rotating, it certainly is also possible to travel a curved but open one. But now let's ask ourselves what happens if there are several devout Moslems trying to do the same thing, keeping their arms locked at all times — let's say just two of them. Well, they can still do it, but the one who started closer to the fountain must at some time move on an outside path, and then go back inside. Infinitely many material points which are constrained by continuity and impenetrability of matter can't quite do that, and therefore the appearance of closed streamlines is a guarantee that rotation will not be zero. All this can be (and has been) formalized by introducing the concept of *circulation*, which in some sense is the integral of rotation along a closed path. Fortunately enough, we don't need to worry with such complications here.

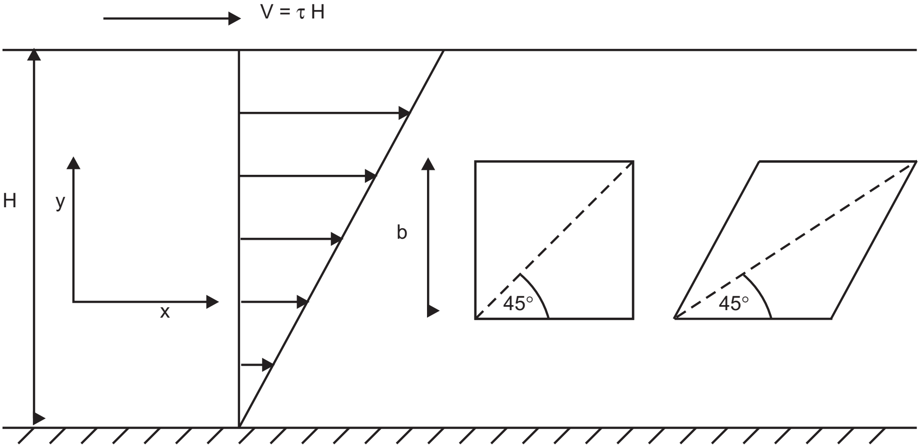

Fig. 4.3.1. Deformation and rotation of a material element in shear flow.

It is, however, of interest to understand in a bit more detail the decomposition of grad**v** into **D** and **W**, and in order to do so it is useful to consider steady flow between two parallel plates (see Fig. 4.3.1). The only nonzero component of velocity is u, and it depends only on y, $u = \tau y + \text{const}$. Consider a material element which at time $t = 0$ has a square shape with sides oriented along the x and y axes (say its bounding surfaces are $y = 0, y = b, x = 0, x = b$). At some later time, the $y = 0$ segment has moved to some new position along x without changing length — all of its points move with the same velocity. The same is true of the $y = b$ segment, only that it has been displaced farther along the x axis. Thus what was a square has become a diamond — there is deformation, and indeed **D** has two nonzero components, $D_{XY} = D_{YX} = \tau/2$. But notice that the material line which was at 45 degrees with the coordinate axes at time zero forms a different angle later — the element has also rotated, and indeed **W** has two nonzero components, $W_{XY} = -W_{YX} = \tau/2$. So the flow between two parallel plates is not irrotational. Now consider a flow described by $u = \tau y$, $v = \tau x$. A square will now deform to a diamond, but will not rotate (the reader is asked to convince herself of this by drawing a little sketch). And indeed, **D** has again two nonzero components, $D_{XY} = D_{YX} = \tau$, i.e., it is exactly the same as before; but this time $\mathbf{W} = 0$. The value of **D** being the same as before, the same viscous stresses will be exhibited by a Newtonian fluid; but there is no rotation.

The example of flow between parallel plates leads to an important conclusion: *shearing a fluid induces rotation*. The second flow field considered is one where the fluid *is not* sheared, but simply elongated, and no rotation is thus induced. In flow past submerged bodies, the no-slip condition at the body surface causes the fluid to be sheared locally, and thus rotation is induced in it: *the no-slip condition is a source of rotation*.

There are some purely kinematic constraints which are imposed on flow fields which are worth mentioning here. First, the divergence of the rotation vector is necessarily zero, $\text{div}\mathbf{w} = 0$. While the velocity field is solenoidal only in incompressible fluids, the rotation field is invariably solenoidal. Second, the velocity field, as any vector field, can be expressed as the sum of the gradient of a scalar and the curl of another vector field, say $\mathbf{v} = \text{grad}\varnothing + \text{curl}\mathbf{b}$, with $\text{grad}\varnothing$ having zero curl (and possibly finite divergence), and $\text{curl}\mathbf{b}$ being necessarily solenoidal.

The Bernoulli equation implies that $P + \frac{1}{2}\Phi v^2 = H$ is constant along any streamline. It can be proved that $\text{gradH} = \mathbf{vxw}$, which implies that H is also constant along every vortex line (a vortex line is everywhere tangential to the rotation vector). It also follows that in irrotational flow fields $\text{grad}H = 0$, i.e., $P + \frac{1}{2}\Phi v^2$ has the same value everywhere in the flow field. If Coriolis forces due to rotation of the frame of reference are of importance, one gets the result that $\text{gradH} = \mathbf{vx}(\mathbf{w} + 2\mathbf{W})$, where $\mathbf{W}$ is the angular velocity of the rotating frame (typically, the Earth rotation, but often also some intrinsic rotation when flow fields in rotating machinery are described in a frame rotating with the machine).

When fluid viscosity is taken into account, Eq. (4.3.9) becomes:

$$\text{D}\mathbf{w}/\text{D}t = \mathbf{w}\cdot\text{grad}\mathbf{v} + (\mu/\Phi)\text{lapl}\mathbf{w} \tag{4.3.a.5}$$

which again shows that viscous forces are a source of rotation. The Bernoulli equation becomes:

$$\text{grad}(P + \tfrac{1}{2}\Phi v^2) = \mathbf{vxw} + (\mu/\Phi)\text{lapl}\mathbf{v} \tag{4.3.a.6}$$

which of course implies that H needs not be constant along either streamlines or vortex lines.

4.4 LAMINAR FLOWS AND THE LUBRICATION APPROXIMATION

Laminar flows constitute the largest class of known exact solutions of the Navier–Stokes equations, and, as was discussed in Section 4.2, laminar flows are possible only for steady flows which can be realized with parallel rectilinear streamlines. One can choose a Cartesian coordinate system with the z axis along the flow direction, and assume that the following kinematics hold:

$$w = f(x,y); u = v = 0 \tag{4.4.1}$$

Ralph has a protest to lodge here. He says this is not an assumption, the solution will in fact be of the type given. He remembers the solution for flow down a circular tube and an infinite slit by heart, and could work out that for a triangle, a square, an ellipse, a rectangle — he did that in his undergraduate days. Well, if indeed the solution is of the type of Eq. (4.4.1), that means we have made a good guess — at this stage it's still a guess. It isn't, however, a wild guess, or one which is made in hindsight by knowing the solution. There certainly must be a z component of velocity to have any flow-rate; there is no reason to believe that there is secondary flow within the plane of the cross section; and we are looking for a laminar flow solution, so $\mathbf{v}\cdot\mathrm{grad}\mathbf{v}$ has got to be zero. Equation (4.4.1) satisfies that: $\mathbf{v}$ points in the z direction, and grad$f$ certainly lies in the plane of the cross section and is thus orthogonal to the velocity, so *if* we can satisfy the NS equations exactly with Eq. (4.4.1) we'll have a laminar flow solution. Notice that this time our estimate of the order of $\mathbf{v}\cdot\mathrm{grad}\mathbf{v}$ as V^2/L is entirely wrong: the scalar product of two vectors is *usually* of the order of the product of their magnitudes, but not when they are orthogonal to each other (or almost orthogonal, as will be seen later). We have already established that (4.4.1) yields $\mathrm{D}\mathbf{v}/\mathrm{D}t = 0$, so the NS equations along x and y reduce to $\partial P/\partial x = 0$, $\partial P/\partial y = 0$, and hence (4.4.1) can work only if $P = P(z)$. That makes sense — we don't expect P to change in the plane of the cross section. The z component of the NS equations now reduces to μlapl$f$ = dP/dz. Since lapl$f$ depends only on x and y, and dP/dz only on z, this can be satisfied only if they are separately equal to the same constant $-K$, K being the pressure drop per unit length. Notice two things: first, K could be zero, like in the flow between two parallel plates one of which is in steady motion and thus drags the fluid along. Second, in the gravity driven flow down an inclined plane, the actual pressure on the free surface is constant along z, but P is not, since z is not horizontal. (Going back to the case of flow down a tube, we note that we are happy that $dP/dz = -K$: we expect the pressure drop per unit length of tube to be the same everywhere along the tube axis). Thus the problem is reduced to solving the two-dimensional Laplace equation laplf $= -K/\mu$ subject to whatever boundary conditions are relevant to the problem at hand. Solving the latter problem is trivial, and we are done with laminar flows — or are we? Well, there is the little matter that for the case of flow down a tube (where the boundary conditions impose that f is zero on the wall of the tube) μf$/K$ is independent of K, which means that the flow-rate is proportional to the pressure gradient and inversely proportional to viscosity. It also means that at any given flow-rate the velocity profile is independent of viscosity. Now we are really almost done with laminar flows, except to observe that it has all worked because of the simple constitutive equation for stress involved. For viscoelastic fluids it doesn't work, except for the circular tube and the infinite slit; in all other cases the equations of motion will not be satisfied by anything of the form (4.4.1), and a secondary flow in the plane of the cross section will develop. Finally, it is interesting to observe that an exact solution could be obtained because inertia drops out and the problem becomes linear.

Is the dropping out of inertia the only situation where the equations become linear? Can anybody think of another possibility? Sue suggests that perhaps sometimes inertia is linear. Good point: $\partial\mathbf{v}/\partial t$ is linear, and if $\mathbf{v}\cdot\text{grad}\mathbf{v} = 0$ one can have a linear problem with nonzero inertia. The problem discussed in Section 1.4 is an example of this, and so is the one discussed in Section 3.2.4 (for the latter one, the Coriolis force is linear). For such problems an exact solution is possible, as well as for problems which are variations thereof (for example, the problem in Section 1.4 with a fixed plate at some distance from the moving one). There is another problem (and variations thereof) where inertia is nonlinear, and yet an exact solution can be found. That is the problem of an infinite circular rod rotating around its own axis in an infinite sea of fluid. Here the solution is found because the streamlines are circular, and the inertia force is thus radial and is balanced by a pressure distribution in the radial direction. In the tangential direction inertia is zero, and thus the kinematics of motion are determined by a linear equation. The flow in a Couette viscometer is a variation of the latter flow, with a twist on it. In a Couette viscometer with the inner cylinder rotating, the tangential velocity may decrease with distance r from the axis of rotation rapidly enough for the centrifugal force decreasing with increasing r. This is clearly a possibly unstable situation, since the material near the axis is pushed outwards by the centrifugal force, and indeed it is the classical problem of the Taylor instability. Now we all know that all laminar flows eventually become unstable (at sufficiently high values of the Reynolds number), and the flow becomes turbulent. This is however a different kind of instability, and its arising at some particular value of **Re** should not be constructed to imply that it's a question of relative importance of inertia and viscous forces; as long as the flow stays laminar, inertia forces are exactly zero, and **Re** cannot be interpreted as the ratio of their value to that of the viscous forces. The Taylor instability, on the contrary, does arise from an existing inertia force (the centrifugal force), which at some rotational speed cannot any more be stably balanced by a pressure distribution.

There are a few additional problems where an exact solution of the Navier–Stokes equations can be obtained, but we will not discuss these. We focus our attention instead on a class of flows which might be called "quasi-laminar", which constitute the body of the so called lubrication approximation. The essence is as follows. Suppose that we have a flow field where the streamlines are *almost* rectilinear and *almost* parallel to each other. Then, in some sense to be made precise, the flow will be *almost* laminar. Bob finds this statement trivial, Ralph thinks it is meaningless, and Sue finds it of some moderate interest. Ralph is the easiest one to satisfy with an example where an exact solution can be found, that of flow between two plates converging at some small angle α to each other, with the flow taking place under a pressure difference in the direction of decreasing cross section. For that particular case, the exact solution can be found, and indeed it degenerates properly to the laminar flow situation when α approaches zero. It is all worked out in Batchelor's book, and Ralph is urged to study that section in detail. But in most cases one cannot obtain an exact solution, and one must rely on a scaling argument of the type discussed in Chapter 3.

The scaling argument, however, is very delicate, and the problem of flow between converging planes forming an angle α is of help. We do not need to actually find the solution for that problem, but simply observe that, in a cylindrical coordinate system with the z axis orthogonal to the plane of the flow and r measured from the singular point, a similarity solution can be obtained where the only nonzero component of the velocity is the radial one, say w, with $w = f(\theta)g(r)$. Continuity now forces $g(r)$ to be of the form $-A/r$, and the constant A can be so chosen that f is of order unity. Now suppose we concentrate attention on a region far away from the singular point, with length L. Let R be the (average) value of r in the region of our interest, and let $H = R\alpha$

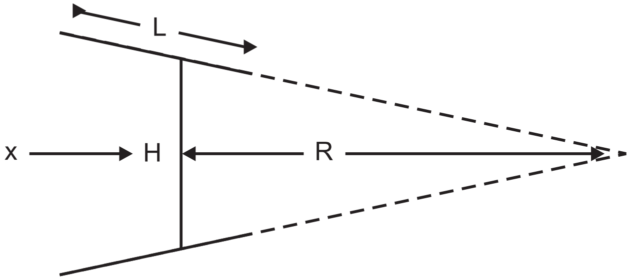

Fig. 4.4.1. Converging flow geometry for the estimate of the scaling parameter.

be the (average) spacing of the plates in the same region. Then the velocity scale is $U = A/R$. We now choose Cartesian coordinates as shown in Fig. 4.4.1, with u and v the x and y components of velocity, and observe that $\partial u/\partial x \approx \partial w/\partial r \approx A/R^2$. But $A/R^2 = \alpha U/H$, and since u is known is to be of order U we conclude that the significant distance in the x direction is H/α. This is all that we need: without the considerations in this paragraph, we might have guessed that the significant distance in the x direction would be L, which needs not be of order H/α (though it is in the classical inclined sliding block example of lubrication theory).

The equation of continuity now furnishes the appropriate scale for v as αU. The equation of motion in the y direction is:

$$\Phi(u\partial v/\partial x + v\partial v/\partial y) = -\partial P/\partial y + \mu(\partial^2 v/\partial x^2 + \partial^2 v/\partial y^2) \tag{4.4.2}$$

The two terms on the LHS are both of order $\Phi U^2\alpha^2/H$.In the viscous force term, the two derivatives are of order $\mu\alpha^3 U/H^2$ and $\mu\alpha U/H^2$, respectively, and hence only the $\mu\partial^2 v/\partial y^2$ term needs to be retained, which is of order $\mu\alpha U/H^2$. We now take the ratio of the inertia force to the viscous force and find it to be of order $\alpha\mathbf{Re}$, where the Reynolds number is based on the separation H, $\mathbf{Re} = \Phi UH/\mu$. Thus inertia can be neglected if:

$$\alpha\mathbf{Re} << 1 \tag{4.4.3}$$

which is a much less stringent condition than that of requiring the Reynolds number itself to be much less than unity. What happens in this case is that the velocity vector and the direction of its gradient operator form an angle which is of the order of $\Pi/2 - \alpha$, i.e., they are *almost* orthogonal, so that the Reynolds number overestimates the ratio of inertia to viscous forces by a factor $1/\alpha$.

The y-momentum equation has reduced to:

$$\partial P/\partial y = \mu \partial^2 v/\partial y^2 \tag{4.4.4}$$

Let δP_Y be the variation of P across the channel. The LHS of (4.4.4) is of order $\delta P_Y/H$, and thus δP_Y is estimated to be of order $\alpha\mu U/H$. Notice that in the laminar flow case δP is exactly zero; in the present case it is of order α and, as will be seen, it can be taken to be zero when considering the x-momentum equation.

Now consider the x-momentum equation. Condition (4.4.3) is again what is needed to neglect the inertia term, and the derivative with respect to y is the only one retained in the viscous force term, so that it reduces to:

$$\partial P/\partial x = \mu \partial^2 u/\partial y^2 = O(\mu U/H^2) \tag{4.4.5}$$

Since the order of magnitude is independent of α, at sufficiently small α pressure P can be taken to be only a function of x, and hence the differential equation of the laminar flow case has been recovered: *the flow at every value of x is the same as one would have in laminar flow with the local spacing between the plates and the local pressure gradient*. The conclusion is a general one, valid for all flows where α (even if variable) is always small, and condition (4.4.3) is satisfied.

The lubrication approximation is extremely successful. It applies, in addition to the classical problems in lubrication, for all flows where the geometrical conditions are met, such as squeezing flows, superimposed drag and pressure driven flows, and the like. Possibly the most striking example of successful application of the approximation is in the calculation of the flow field for a liquid layer falling under the influence of gravity over the surface of a sphere. Here one assumes that the flow at any angular position θ (measured from the top of the sphere) is the same as one would have on an inclined plane forming the same angle with the vertical, and with the same flow-rate per unit width. (The latter flow is a laminar flow driven by gravity, i.e., under the influence of a finite grad$P$, with one no-slip surface and a parallel no stress surface). If Q is the volumetric flow-rate, one obtains from the lubrication approximation the following value of the film thickness δ:

$$\delta(\theta) = (3\mu Q/2\Pi\Phi g)^{1/3} \sin\theta^{-2/3} \tag{4.4.6}$$

This of course becomes degenerate at $\theta = 0$ and $\theta = \Pi$, as one would expect since at the poles the tangent plane is horizontal. (The fluid is actually observed to remain attached to the sphere all the way to the lower pole at $\theta = \Pi$). If x is distance into the liquid layer measured from the free surface, the velocity distribution is:

$$v = V[1-(x/\delta)^2] \tag{4.4.7}$$

where V is the velocity of the free surface:

$$V = \sin\theta^{1/3}\,(3Q/4\Pi R\delta_o) \tag{4.4.8}$$

and δ_o is the layer thickness at the equator $\theta = \Pi/2$:

$$\delta_o = (3\mu Q/2\Pi\Phi g)^{1/3} \tag{4.4.9}$$

Of course, v and V also become degenerate at the two poles. One however proceeds without worrying about that, and one calculates the residence time of a liquid element travelling at the exposed surface:

$$t_D = \int_0^{\Pi} R\mathrm{d}\theta \,/\, V \tag{4.4.10}$$

The value of t_D so calculated is finite although V is zero at both extremes of integration. Ralph doesn't believe any of this, and he points out that we may have obtained a value for t_D with this sort of spurious way of doing things, but how do we know it is the correct value? Bob is worried about how to actually do an experiment which could prove or disprove our result. Sue doesn't see anything even slightly awkward in what we have done, except that she wouldn't know how to calculate the integral of the cubic root of $\sin\theta$ — and she is not interested in learning how to do it. Well, the experiment can be set up, and the value of t_D can be used to calculate the rate of mass transfer of an absorbing component into the liquid, and the result agrees remarkably well with experiments.

4.4.1 The Hele Shaw cell

The lubrication approximation yields equations of motion which are exactly those of a laminar flow. The other characteristic of the lubrication approximation is that, in order to neglect inertia forces, the required condition is not that the Reynolds number should be small as compared to unity, but the less restrictive condition that the product of the Reynolds number times a geometrical shape factor (which by itself is small) should be small. In this subsection, we discuss a special flow field which has the second one of these properties but not the first one: the Hele Shaw cell. The Hele Shaw flow field is interesting in another respect which will be discussed below, and it represents a smooth transition to the flow fields which will be discussed in the next section.

Consider two fixed parallel plates located at a distance d apart from each other. Between the two plates there are obstacles which for simplicity will be regarded as circular disks of length d and diameter $L >> d$; the average spacing between the disks is also L. The x and y coordinates are in the plane of the lower plate, and the z one is orthogonal to them. Let u, v and w be the corresponding velocity components. The length scale in both the x and y direction is obviously L, while the one in the z direction is d. Let V be the velocity scale in the planes parallel to the fixed plates. The equation of continuity now yields the result that the velocity scale in the z direction is $V = U\mathrm{d}/L << U$. Let K be the order of magnitude of the pressure gradient in the planes orthogonal to z, and consider the equation of motion in the x or y direction. The inertia force is estimated at $\Phi U^2/L$, and the viscous force as $\mu U/\mathrm{d}^2$, with only the second derivative with respect to z being retained. Hence inertia forces can be neglected provided $\mathbf{Re}d/L << 1$, where the Reynolds number is based on the spacing d. Hence again one concludes that a (small) geometrical parameter is involved in the condition for neglecting inertia forces, in perfect analogy with the lubrication approximation. In other words, again **Re** overestimates the ratio of inertia to viscous forces because the velocity vector and the direction of its gradient are *almost* orthogonal to each other.

The analogy with the lubrication approximation stops here, however, since there are two velocity components which are significant, u and v, and the streamlines are expected to be anything but straight lines. Indeed, the governing equations become:

$$\partial P/\partial x = \mu\partial^2 u/\partial z^2; \partial P/\partial y = \mu\partial^2 v/\partial z^2 \tag{4.4.11}$$

which, when subject to the no-slip condition at $z = 0$ and $z = d$, have the solution:

$$2\mu u = -(\partial P/\partial x)z(d-z); 2\mu v = -(\partial P/\partial y)z(d-z) \tag{4.4.12}$$

The interesting point about Eq. (4.4.12) is as follows. At any given value of z (as well as for the average value of u and v over z), the flow is irrotational, and the velocity admits a potential which is proportional to pressure P. The flow is not irrotational if the z direction is considered as well, since $\partial u/\partial z$ and $\partial v/\partial z$ are nonzero, and cannot possibly be balanced by $\partial w/\partial x$ and $\partial w/\partial y$ terms, since the latter are known to be of a much smaller order of magnitude; but it is irrotational in the planes orthogonal to z. In fact the Hele Shaw cell can be used as a demonstration of the kinematics of an irrotational flow field.

4.5 CREEPING FLOWS AND THE OSEEN APPROXIMATION

In the preceding section, we have analyzed flow fields where inertia forces can be neglected provided $\varepsilon\mathbf{Re} << 1$, where ε is a (small) geometrical parameter. In this section, we examine flow fields where the inertia force can (perhaps) be neglected because **Re** itself is small as compared to unity. If indeed **Re** represents a proper estimate of the ratio of inertia to viscous forces *over the whole flow field*, then it is (perhaps) appropriate to drop the inertia term from the Navier Stokes equation (NSE), and obtain the *creeping flow* equation (CFE) below:

$$\text{grad}P = \mu\text{lapl}\mathbf{v} = -\mu\text{curl}\mathbf{w} \tag{4.5.1}$$

Equation (4.5.1) implies that:

$$\text{lapl}P = 0 \tag{4.5.2}$$

$$\text{lapl}\mathbf{w} = 0 \tag{4.5.3}$$

Creeping flow problems are analyzed by seeking an (exact) solution of the CFE. This is hoped to be an approximate solution of the NSE, but, as will be seen below, this is not guaranteed to be the case when the flow field extends to infinity.

There is a theorem of uniqueness of the solutions of the CFE, and hence when one such solution is found which satisfies the relevant boundary conditions, it is guaranteed to be *the* solution of the problem. There is also a variational theorem which will be discussed later, and which offers the possibility of finding (approximate) solution by variational techniques.

The CFE is linear, and its solution can often be found, though the mathematical and numerical techniques involved may be quite sophisticated. We begin by considering a very simple problem, i.e., the classical problem of a sphere moving at constant speed in an infinite body of fluid which is at rest far from the sphere. This is well known to lead to Stokes' law for the drag force F exerted by a fluid moving with velocity U on a sphere of radius R:

$$F = 6\Pi\mu RU \tag{4.5.4}$$

Equation (4.5.4) is obtained as follows. First, the velocity and pressure distributions in the fluid are calculated by solving the CFE with the appropriate boundary conditions. Next, the form drag is calculated by integrating the effect of P on the sphere. (Notice that the effect of the actual pressure p can be thought of as arising from two contributions: the hydrostatic pressure p_o, and the reduced pressure $P = p - p_o$. Since the effect of the hydrostatic pressure is the Archimedean buoyant force, only P contributes to the drag force). Finally, the skin drag is calculated by integrating the effect of the tangential stresses on the sphere, and Eq. (4.5.4) is obtained by summing the form and skin drag so calculated.

Ralph is a bit bored by the statements in the previous paragraph, because he knows this by heart and sees no point in dwelling on it. We wish however to make a point here. Stokes' law is very well confirmed by experimental results; this however only guarantees that the solution of the CFE is correct near the surface of the sphere, since only the pressure and kinematics at the surface itself are used in calculating the drag. The solution found could be seriously in error at large distances from the sphere, without invalidating Eq. (4.5.4). Bob finds this point irrelevant, since he couldn't care less about what happens far away from the sphere — in actual fact the flow never really extends to infinity anyhow. Sue thinks that the velocity has got to be very close to the free stream velocity far away from the sphere, and P has got to be almost zero, both in actual fact and as predicted from the CFE solution, so she rather agrees with

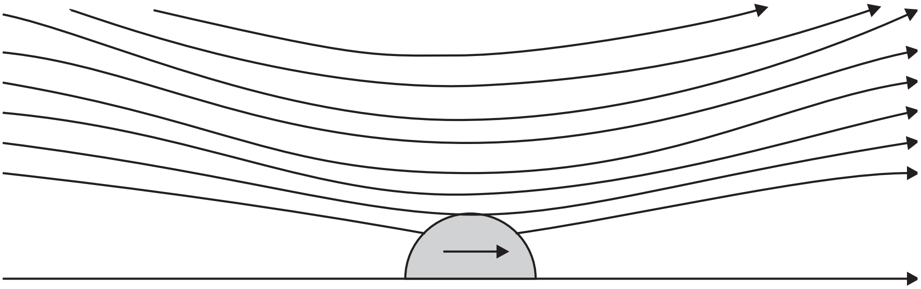

Fig. 4.5.1. Streamlines calculated from the derivation of Stokes' law.

Bob. Well, let's see what happens anyhow, maybe we can learn something of interest. Let's look at the streamlines predicted by the CFE solution, which are given in Fig. 4.5.1. They are perfectly symmetrical with respect to the equatorial plane, which means that, should the sphere move in the opposite direction, all that would happen is simply that **v** would change sign everywhere. Does that make sense? Does really the sphere affect the motion upstream exactly the same way as it does downstream?

Enough doubt has been raised by now for the class to show some interest. Let's try to look at this result from the viewpoint of Eq. (4.5.3). The rotation is zero far away from the sphere, but the shearing caused by the sphere in its proximity is a source of rotation, and Eq. (4.5.3) asserts that this rotation diffuses out but is not convected by the moving fluid — it is, after all, the steady diffusion equation. One would however expect *some* convection of the rotation, and this of course would imply that the flow downstream is not the mirror image of the flow upstream. The question is not, of course, that the solution found is not an exact solution of the CFE; the problem is that it might not be an approximate solution of the NSE. Perhaps what happens is as follows: **Re** is an appropriate yardstick for the ratio of inertia to viscous forces near the sphere, but not far away from it. Indeed, if we consider a point at distance r from the centre of the sphere, the appropriate length scale is r itself, and hence the ratio of inertia to viscous forces is $\Phi Ur/\mu = \mathbf{Re}r/R$. (This is confirmed by calculating the ratio from the solution of the CFE for flow around a sphere). So we conclude that, no matter how small **Re** might be, sufficiently far away from the sphere inertia forces predominate over viscous ones (though both are quite small indeed). *The CFE solution is not a uniformly valid approximation of the NSE solution for flows which extend to infinity*. This should be contrasted with the results in the previous section: the lubrication approximation, and the Hele Shaw one, are uniformly valid, since the flow is restricted to the region within which the approximations are valid ones, and does not extend to infinity.

A uniformly valid approximation can be obtained from the Oseen approximation. The reasoning is as follows. Define **u** as **v** – **U**, where **U** is the free stream velocity. Since grad**U** = 0, the neglected inertia force for steady flow can be written as:

$$\Phi \mathrm{D}\mathbf{v}/\mathrm{D}t = \Phi \mathbf{u}\cdot\mathrm{grad}\mathbf{v} + \Phi \mathbf{U}\cdot\mathrm{grad}\mathbf{v} \qquad (4.5.5)$$

Near the sphere, the CFE is appropriate, but it is not so far away. But far away from the sphere u is much less than U (as Sue had shrewdly pointed out), and hence the first term on the RHS of (4.5.5) can be neglected. One thus obtains the Oseen equation (OE):

$$\Phi \mathbf{U}\cdot\mathrm{grad}\mathbf{v} + \mathrm{grad}P = \mu\,\mathrm{lapl}\mathbf{v} \qquad (4.5.6)$$

The OE is again linear, and hence solutions can (with some difficulty) be found. Equation (4.5.3) does not follow from the OE, and one has instead:

$$\Phi \mathbf{U}\cdot\mathrm{grad}\mathbf{w} = \mu\,\mathrm{lapl}\mathbf{w} \qquad (4.5.7)$$

It follows that convection of rotation can be accounted for, although the effective convecting velocity decreases with decreasing density. The solution of the OE for flow around a sphere can be found, and it can be shown to be a uniformly valid approximation to the solution of the NSE, i.e., the ratio of inertia to viscous forces calculated from the solution of the OE is everywhere of order **Re**.

The streamlines obtained from the OE for flow around a sphere are shown in Fig. 4.5.2, and they are seen to be significantly different from those in Fig. 4.5.1: the velocity far from the sphere may be quite close to the free stream velocity, but still the streamlines are quite different from what one would calculate from the CFE. And, Bob, there might be some pragmatical problems where the shape of the streamlines far away from the sphere is relevant; after all, things do get convected along streamlines.

Ralph is quick to catch something wrong here. The solution of the OE, he points out, must depend on the value of **Re**, and so what are the streamlines in

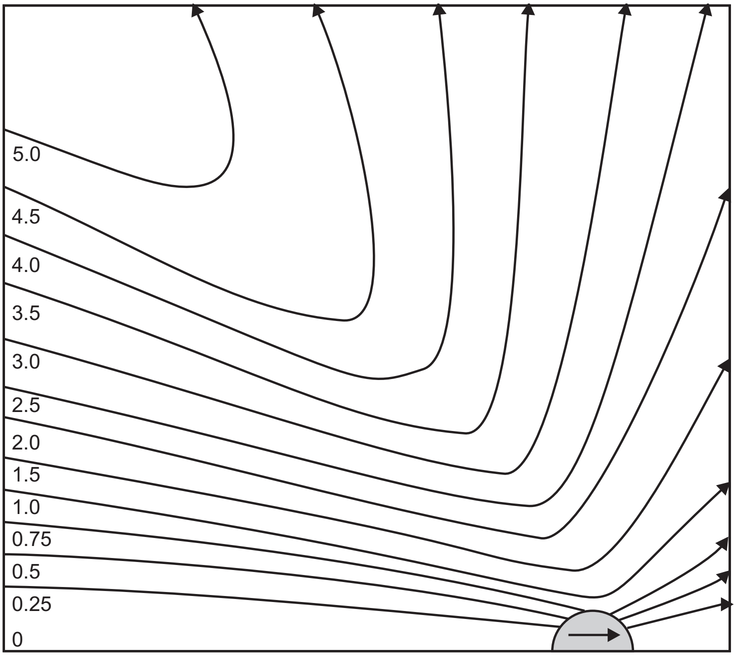

Fig. 4.5.2. Streamlines for flow around a sphere calculated from the OE.

Fig. 4.5.2? To which value of **Re** do they correspond? Good point indeed. It so happens that the stream function calculated from the OE is the sum of two terms, one of which becomes negligible far from the sphere, and has not been included in the streamlines plotted in Fig. 4.5.2. (The neglected term is the stream function obtained from the CFE solution). The dominant term far from the sphere becomes independent of **Re** when **Re** approaches zero, and hence the streamlines in Fig. 4.5.2 are correct in this limit and independent of **Re**.

The solution of the OE also furnishes a first order correction to Stokes law, i.e., the RHS of (4.5.4) should be multiplied by $1 + 3\mathbf{Re}/16$. This however is not very significant, since the OE only holds to within $O(\mathbf{Re})$ anyhow, and hence the correction term is negligible when it is appropriate. There is however another problem where the OE furnishes a very fundamental result, and this is discussed in the following.

Consider the two-dimensional version of the Stokes problem, i.e., flow orthogonal to a cylinder which extends to infinity in the axial direction. It takes some heavy algebra to prove a negative, but it can be shown that there is no steady solution of the CFE for this flow, and this seems to create a paradox (it is in fact referred to in the literature as the Stokes paradox). Let's first look at the problem from a dimensional viewpoint. In the Stokes problem F must depend only on U, R and μ if we believe the CFE — density cannot play any role. Since the product $UR\mu$ is the only possible combination having the dimensions of a force, the whole Stokes analysis is really needed only to find the numerical value 6Π. Now in the cylinder case we are interested in the force per unit length of cylinder, F' (the total force on an infinitely long cylinder being of course infinity). Now the only combination of U, R and μ having the dimensions of F' is μU, but the equation $F' = K\mu U$ does not make any sense: can we really believe that the drag force on the cylinder is independent of its diameter?

This time the OE comes to the rescue: a solution for flow across a cylinder can be found. and, in contrast with the sphere case, there is no term contributing to the drag which is independent of **Re**. The drag force is given by:

$$F = 4\Pi\mu U/\log(7.4\mathbf{Re}^{-1}) \qquad (4.5.8)$$

which of course becomes degenerate when **Re** = 0.

It is worthwhile to consider an analogous problem in heat transfer which makes the difference between the cylinder and sphere cases more clear. If one considers a sphere maintained at some temperature T_o in a quiescent fluid kept at some other temperature far away from the sphere, the steady diffusion equation has a solution, and correspondingly a Nusselt number can be calculated (the value is 2). If one formulates the analogous problem for a cylinder, there is no steady state solution other than temperature being T_o everywhere. One may now ask oneself the following question: if there is a finite Nusselt number, there is a steady flow of heat into the fluid — how does this agree with a steady temperature distribution? In the case of a sphere, the flow of heat is nonzero but finite, and the amount of fluid receiving it is infinitely large, so that there is no contradiction. In the case of the cylinder, a finite heat transfer coefficient implies an infinitely large flow of heat into (or out of) the fluid, and no steady solution can be found. In the fluid mechanics analog, what diffuses out according to the CFE is rotation, and for the cylinder case there is an infinitely large flow of rotation into the fluid.

Let's now move to a similar problem. Say we want to calculate the steady rising velocity of an air bubble through water, and we are willing to limit attention to the case where the bubble is spherical and conditions are such that the Reynolds number in the liquid is small. It also seems reasonable to assume that the gas has zero viscosity, and zero density —they both are certainly much smaller than the corresponding ones for water. If the gas viscosity is zero, one certainly cannot use the no-slip condition at $r = R$, and what should one use instead? We are experts by now, we need a condition of zero tangential stress at the interface, since a gas with zero viscosity cannot sustain a tangential stress. This means that the skin drag is zero before we even begin doing any calculation, so the drag must be all form drag. Now pressure within the gas bubble is presumably constant, and so the pressure in the liquid at the interface is constant too — lower by an amount $2\sigma/R$ than that within the bubble, where σ is the surface tension. Any objections so far? Ralph is quick to point out that the drag force is $4\Pi\mu RU$ and doesn't want to be bothered. Sue says she senses something fishy in our argument but is unable to pinpoint it. But let's proceed. If p is constant, then P has (minus) the hydrostatic distribution, and hence the form drag is equal and opposite to the buoyant force. But that we knew from the beginning, since we are looking at the *steady* rising of the gas bubble. This time trying for a shortcut has not been very helpful. One has to solve the CFE with the appropriate boundary conditions, and the final result is indeed the one Ralph remembered; the drag is of course all form drag. One can also solve a more complex problem, that of a spherical liquid drop with viscosity μ' moving steadily through an infinite body of liquid. In this case one has to write down the CFE for both fluids (inside and outside), and write the appropriate boundary conditions. At infinity the (outside) fluid is at rest and $P = 0$. Velocity and pressure are everywhere finite in the inside fluid. At the interface itself, the no-slip condition implies that the inside and outside velocity are equal, and the first law of mechanics implies that the tangential stress must also be the same. The result is:

$$F = \Pi\mu RU(4\mu + 6\mu')/(\mu + \mu') \tag{4.5.9}$$

which degenerates to Stokes' law for $\mu' = \infty$ (a solid sphere), and to $F = 4\Pi\mu UR$ for $\mu' = 0$ (a gas bubble). Of course, the solutions of the CFE discussed above all suffer from the same limitations as discussed before in the regions far away from the sphere, but there are no such limitations for the flow inside the sphere, where r never exceeds R.

Notice that P is defined to within an arbitrary additive constant (since only its gradient enters the CFE), but this arbitrary constant needs not be the same in the inside and outside fluid. Notice also that no boundary condition for P

has been written at the interface for that very reason. Now the solution of the problem (which we know to be accurate near the interface and in the inside fluid) allows us to calculate the normal stress difference at the interface, and it turns out to be a constant over the whole interface: thus the stresses which are generated by the flow do not tend to deform the sphere from its spherical shape, and surface tension enters the problem only in determining the constant normal stress difference at the interface to be $2\sigma/R$. This is an important result, because in its absence one would estimate $\sigma >> \mu U$ as the condition required for surface tension to keep the sphere from being distorted by the flow, and in actual fact one observes that liquid drops retain their spherical shape also when that condition is grossly violated.

There is turmoil in the class here - where did we get $\sigma >> \mu U$ from? Well, σ/R is the force per unit surface resulting from surface tension (and this force tends to keep the shape spherical, since at an assigned volume the spherical shape has the smallest possible surface area), and $\mu U/R$ is the correct estimate of the stress levels at the interface (which could conceivably tend to distort the shape from being spherical). Hence if the stresses had a tendency to deform the sphere one would need σ/R to predominate over $\mu U/R$ in order to neglect shape distortions due to flow.

We now come to the question of the Helmholtz–Korteweg variational theorem. The theorem is usually derived (see e.g. Batchelor's book) by considering directly steady flows of incompressible Newtonian fluids, and the procedure is essentially analogous to that needed to obtain the uniqueness theorem for the CFE. We here choose an alternate procedure, which has the advantage of being applicable also to a (restricted) class of Non-Newtonian fluids. Let's first state the theorem: *of all incompressible flows corresponding to assigned steady boundary conditions, the flow corresponding to the solution of the CFE has the minimum value of the integral over the volume of fluid of the quantity* Γ *defined below*.

Our procedure is to consider a situation where the flow is at some instant in time different from the solution of the CFE, and to ask ourselves what happens as time proceeds. Since initially the flow is not the solution of the CFE, the kinematics will change in time, and we need to retain inertia terms if we want to consider the time evolution. However, at sufficiently small values of **Re** the convective inertia is negligible, and hence we consider the following equation for the evolution in time of the flow field:

$$\Phi \partial v/\partial t = -\mathrm{div}\mathbf{T} \tag{4.5.10}$$

Equation (4.5.10) implies that:

$$0 = (\partial \mathbf{v}/\partial t) \cdot \mathrm{div}\mathbf{T} = \mathrm{div}(\mathbf{T} \cdot \partial \mathbf{v}/\partial t) - \mathbf{T}\text{:}\mathrm{grad}(\partial \mathbf{v}/\partial t) \tag{4.5.11}$$

If (4.5.11) is integrated over the whole flow field volume V, the Gauss theorem can be applied to the first term, which is thus seen to be zero since the boundary conditions are steady ones. In the second term the grad and $\partial/\partial t$ operators commute, and thus one obtains (indicating with $<>$ volume integrals):

$$<\mathbf{T}\text{:}\partial \mathrm{grad}\mathbf{v}/\partial t> = 0 \tag{4.5.12}$$

Now $\mathbf{T}$ is symmetric, and thus only the symmetric part of $\partial \mathrm{grad}\mathbf{v}/\partial t$, $\partial \mathbf{D}/\partial t$, contributes to the integrand in (4.5.12); thus one has:

$$<\mathbf{T}\text{:}\partial \mathbf{D}/\partial t> = 0 \tag{4.5.13}$$

Now suppose one could find some quantity Γ such that $-\partial\Gamma/\partial t$ equals the integrand in (4.5.13); one would conclude that $<\Gamma>$ steadily decreases, until the CFE solution is obtained, and thus a variational theorem has been established. Now suppose that the stress can be expressed as:

$$\mathbf{T} = p\mathbf{I} - f(\alpha)\mathbf{D} \tag{4.5.14}$$

$$\alpha = \mathrm{tr}(\mathbf{D}\cdot\mathbf{D}) \tag{4.5.15}$$

Equation (4.5.14) is more generally applicable than (4.1.7), since $f(\alpha)$ is not required to be constant. Notice that (4.5.13) reduces now to the requirement that $<\tfrac{1}{2}f(\alpha)\partial\alpha/\partial t>$ is nonpositive. Since the fluid is incompressible, $p\mathbf{I}$ contributes nothing to the integrand in (4.5.13). The quantity Γ is easily seen to be:

$$\Gamma = \tfrac{1}{2}\int_0^{\alpha} f(\alpha)\mathrm{d}\alpha \tag{4.5.16}$$

and thus a variational principle has been established. In the Newtonian case, $f(\alpha) = 2\mu$, and hence $\Gamma = \mu\alpha$. Now for any fluid satisfying (4.5.14) the rate of energy dissipation is given by $\varnothing = f(\alpha)\alpha$, and hence for the Newtonian case Γ is proportional to the rate of energy dissipation (of course, if Γ is minimized at the solution of the CFE, so is any quantity proportional to it). In the case of a power law fluid, $f(\alpha) = K\alpha^{n-1}$, and thus $\Gamma = K\alpha^n/2n$, while $\phi = K\alpha^n$: Γ turns out again to be proportional to the rate of energy dissipation. However, this is not true in the case of any other nonlinear form of $f(\alpha)$. For instance, suppose that $f(\alpha) = K + K'\alpha$: one finds $\Gamma = K\alpha/2 + K'\alpha^2/4$, but $\phi = K\alpha + K'\alpha^2$, which is not proportional to Γ (see Section 3.3.1).

The fact that, for Newtonian fluids, the quantity to be minimized is the rate of energy dissipation gives a misleading scent of a thermodynamic result. This is not so, as the counterexample above shows. The Helmholtz Korteweg theorem (and variations thereof which can be obtained with non-constant $f(\alpha)$) is a purely mechanical result which is derived directly from the CFE without any appeal to thermodynamics.

4.5.1 D'Arcy's Law

Consider flow of a fluid through a porous medium. Let d be the (average) dimension of the particles constituting the bed, or equivalently the average pore size, and ε the void fraction of the bed. It is in this case obvious that d represents the appropriate length scale — the actual fluid velocity will change (in both magnitude and direction) over a length scale d. Now let V be the ratio of the volumetric flow-rate of fluid divided by the total cross section. Only a fraction ε of the total cross section is available for flow, and hence the appropriate velocity scale is $U = V/\varepsilon$. The Reynolds number is thus given by:

$$\mathbf{Re} = \mathrm{d}U\Phi/\mu \tag{4.5.1.1}$$

If the Reynolds number given by (4.5.1.1) is small as compared to unity, we have a creeping flow problem, but one where the approximation of neglecting inertia forces is uniformly valid, since the flow is a confined one and does not extend to infinity. There is no Oseen correction to worry about for flow through a porous medium. The CFE implies that the pressure gradient is equal to the laplacian of the velocity, and hence that its order of magnitude is $\mu U/\mathrm{d}^2$. Over a length L of bed, the pressure drop δP is thus estimated as:

$$\delta P = KL\mu U/\mathrm{d}^2 = KL\mu V/\varepsilon \mathrm{d}^2 \tag{4.5.1.2}$$

Equation (4.5.1.2) is known as D'Arcy's law. Notice that the argument given in the preceding paragraph implies that the numerical constant K is of order unity.

D'Arcy's law is often presented as a three-dimensional generalization of the Hele Shaw cell, but this is misleading. D'Arcy's law follows from the CFE, which are assumed to hold because $\mathbf{Re} << 1$, without any (small) geometrical parameter multiplying the Reynolds number, in contrast with the case of the Hele Shaw cell where the condition is $\mathbf{Re}d/L << 1$. In the Hele Shaw cell one expects velocity to change significantly over a length scale L, and yet the condition for neglecting inertia forces is not that the Reynolds number based on L should be much less than unity. In the case of D'Arcy's law, the condition is indeed the more stringent one for the CFE validity.

4.6 HIGH REYNOLDS NUMBER FLOWS AND BOUNDARY LAYERS

When the Reynolds number is large, an approximation similar to the one leading to the CFE in the converse case leads to the formulation of the Euler equation of motion (EE), which have been discussed in Section 4.3. However, the solution of the EE is not simply a nonuniform approximation to the solution of the NSE (as was the case for the CFE solutions for flows which extend to infinity), since the $1/\mathbf{Re} = 0$ limit is a singular one: the highest order derivative in the NSE is dropped, and hence not all the boundary conditions can be imposed on the Euler equations.

We begin by considering flow around a submerged object, say a sphere, for the case where the Reynolds number based on the diameter of the sphere is much larger than unity. Of the boundary conditions which need to be imposed, those at infinity cannot be dropped since one would be describing another problem altogether. On the surface of the sphere, one cannot let go of the requirement that the radial velocity is zero, since if one did the solution would yield the ridiculous result that fluid either enters the sphere or comes out from it. One is thus left with the need to give up the no-slip condition, and the solution of the EE will in fact yield a finite slip velocity. Now consider for instance flow down a constant section tube. Without the no-slip condition, the solution is simply that one has plug flow, and no pressure drop. This is obvious, since the pressure drop is a measure of dissipation, and the EE invariably predicts a zero dissipation.

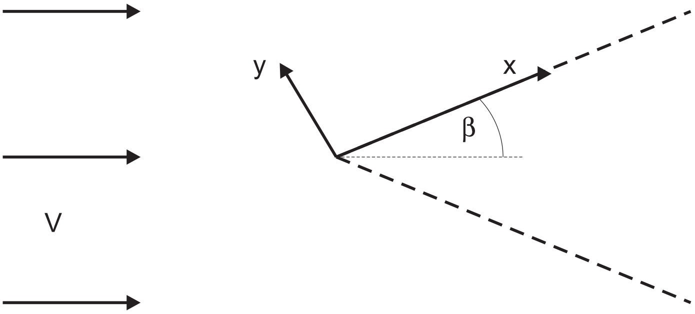

Fig. 4.6.1. Geometry of flow past a wedge.

Of course, one needs to develop an appropriate boundary layer (BL) analysis, as is the case in all singular limits; the solution to the EE can (hopefully) be regarded as the "outer" solution to which the BL solution will need to be matched. It is important to realize that, in order to develop a BL analysis, the solution of the EE for the case at hand needs to be available. The classical example of the flat plate at zero incidence is degenerate in this regard, since the solution of the EE is, trivially, that the velocity is the same everywhere and the pressure gradient is zero. We will therefore begin by considering a slightly more general case, that of a symmetric wedge of semi-angle β invested by a fluid stream moving with velocity V towards the wedge at infinity (see Fig. 4.6.1). Of course, the wedge has no characteristic length, and so it is a bit awkward to understand what the Reynolds number may mean. In reality, an infinite wedge cannot exist, and we are therefore considering some approximation to a finite wedge case; let's say that the Reynolds number based on distance along the wedge from the leading point is large except in a negligibly small region near the wedge tip.

Since the problem has symmetry, we consider only the upper half of the wedge, and take the x axis along it with the origin at the tip; the y axis is away from the wedge surface, and the z axis is the indifferent one. The solution of the EE will yield a finite slip velocity $u_o(x,0) = U(x)$, and this is all we need to know. We also need, of course, the pressure distribution at $y = 0$ as obtained from the EE, but this is simply related to $U(x)$ by $\mathrm{d}P_o/\mathrm{d}x = -\Phi U\mathrm{d}U/\mathrm{d}x$ (we are using a 0 subscript for quantities as obtained from the EE). The boundary layer is located near the surface $y = 0$ where the no-slip condition is violated by the solution to the EE, and its thickness has an order of magnitude $\delta(x)$ which needs to be determined, but is expected to be small as compared to x. The matching conditions are that $u(x,y)$ should merge smoothly into $U(x)$ as the stretched coordinate y/δ approaches infinity.

Our first task is to estimate δ. To this end we reason as follows, to the chagrin of Ralph. At the outer edge of the boundary layer, the EE hold, and hence inertia forces predominate over viscous forces. At $y = 0$, inertia forces are zero because of the no-slip condition, but viscous forces are finite. Hence the F_I/F_v ratio ranges from zero to infinity within the boundary layer, and we conclude that on the average it is of order unity. Yes, Ralph, the average between zero and infinity could be anything, but let's try to be good sports, shall we? The fluid develops a boundary layer thick enough to make inertia and viscous forces comparable; as the viscosity decreases, the boundary layer thickness decreases as well, and thus the ratio of inertia and viscous forces remains constant. At least tentatively it makes sense. Now viscous forces are clearly of order $\mu U/\delta^2$ (if indeed $\delta << x$, only the $\partial^2/\partial y^2$ derivatives need to be retained), while inertia forces are of order $\Phi U^2/x$ (at any given y, U changes by an amount comparable to its magnitude over a distance x). Hence we conclude that:

$$\delta/\mathrm{x} \approx 1/\sqrt{\mathbf{Re}} \tag{4.6.1}$$

where **Re** is based on $U(x)$ and x.

Equation (4.6.1) tells us that indeed, except in a small neighbourhood of the tip, $\delta << x$ as we expected. The equation of continuity now tells us that $v \approx U\delta/x$, i.e., it is much smaller than $u \approx U$. We now first investigate the equation of motion along y, which (neglecting the $\partial/\partial x^2$ term in the viscous force) is:

$$\Phi(u\partial v/\partial x + v\partial u/\partial y) = -\partial P/\partial y + \mu\partial^2 v/\partial y^2 \tag{4.6.2}$$

The two terms on the left, and the second term on the right, have the same order of magnitude, $\sqrt{(\Phi U^3\mu / x^3)}$. Hence $\partial P/\partial y$ has at most the same order, so that the pressure variation across the BL, δP, is of order $\mu U/x$. We now turn attention to the equation of motion in the x direction:

$$\Phi(u\partial u/\partial x + v\partial u/\partial y) = -\partial P/\partial x + \mu\partial^2 u/\partial y^2 \tag{4.6.3}$$

Again the two terms on the left, and the viscous term, all have the same order of magnitude, $\Phi U^2/x$. Now at the outer edge of the BL $\partial P/\partial x = -\Phi U\mathrm{d}U/\mathrm{d}x$, i.e., it is of order $\Phi U^2/x$ as well (except in the degenerate case of the flat plate, $\beta =$

0, where $\mathrm{d}U/\mathrm{d}x = 0$). At the solid surface, the pressure can differ from its value at the outer edge by at most $\mu U/x$, and hence one calculates:

$$\partial P/\partial x \approx \Phi U \mathrm{d}U/\mathrm{d}x + \mu U/x^2 = \Phi U \mathrm{d}U/\mathrm{d}x(1 + 1/\mathbf{Re}) \tag{4.6.4}$$

Equation (4.6.4) implies that, to within order $1/\mathbf{Re}$, the pressure distribution within the boundary layer is the same as calculated from the EE. This has two important consequences. First, since P is known, one may discard the y component of the momentum equation and work only with the x component and the continuity equation to calculate the two unknowns u and v. This of course considerably simplifies the analysis. The second important consequence seems to be that the form drag will again turn out to be zero, (since it is zero according to the EE), but this is an artifact, as will be seen later, since for any realistic shape of the submerged body (and an infinite wedge is not realistic) the boundary layer will eventually detach and a wake will form.

The analysis given above is clearly reminiscent of the lubrication approximation one, with the y-component of the momentum balance used only to infer that the pressure gradient in the transverse direction is negligibly small. The analogy ends there, though, since in the lubrication approximation the inertia force is neglected, while it obviously is not in the BL equations. For what it is worth, the analogy is a physical one. In the case of lubrication, one would choose a lubricant with high enough a viscosity to insure that inertia forces are negligible. Here the fluid has no such luxury available, it must use itself as a lubricant: inertia forces tend to make it slip along the wall, and the fluid creates its own lubricant layer by developing a boundary layer thin enough to generate viscous forces comparable to the inertia ones.

The infinite wedge geometry is degenerate in that a *similarity* solution can be found. The reason for this is essentially a dimensional one. The infinite wedge does not furnish any length scale (any other two-dimensional geometry would: as soon as there is some curvature of the body considered, the radius of curvature would be a length scale). Now let V be the velocity at infinity. Should there be a length scale, the $U = f(x)$ function could be made dimensionless by defining $U/V = F(x/L)$, and no information could be extracted from this result concerning the form of $f()$. However, since there is no length scale, $f()$ must be of the only form which induces only one-dimensional parameter, see Section 3.3.1, and hence one may conclude that $U = Ax^m$. Indeed this is the case when the solution to the EE is calculated, with the exponent m being related to the angle β by:

$$\beta = \Pi m/(m + 1) \tag{4.6.5}$$

Notice that the flat plate at zero incidence corresponds to $m = 0$, and the flat plate transverse to the flow direction corresponds to $m = 1$. One may also wish to consider cases where $m < 0$, which would correspond to external flow towards the tip of the wedge, as well as cases where $m > 1$, corresponding to inverse wedges where the external flow moves backwards towards its source. The case $m < 0$ corresponds to a decelerating external flow field, and this is of interest for what will be discussed later concerning the separation of the boundary layer. The case $m = 1$ corresponds to a constant thickness of the boundary layer, and cases where $m > 1$ correspond to boundary layer thicknesses which decrease along the flow direction.

The similarity solution is sought as follows. The similarity variable has to be $z = y/\delta$, with δ given by Eq. (4.6.1). One thus assumes that some $f(z)$ exists such that:

$$u/U = \mathrm{d}f/\mathrm{d}z \tag{4.6.6}$$

(Writing the RHS as $\mathrm{d}f/\mathrm{d}z$ rather than f makes it easier to work with the continuity equation).

When this is substituted into the equations of motion, one indeed obtains the following ordinary differential equation for f:

$$mf'^2 - \tfrac{1}{2}(m + 1)ff'' = m + f''' \tag{4.6.7}$$

subject to $f(0) = f'(0) = 0, f'(\infty) = 1$. This is nonlinear, but can often be solved numerically. The solution is plotted in Fig. 4.6.2 for m values between –0.0904 and 4. The reader would do well doing the algebra leading to Eq. (4.6.7), to convince himself that a similarity solution is obtained because $U\mathrm{d}U/\mathrm{d}x$ is proportional to U^2/x for the power law case $U = Ax^m$, and that no similarity solution could be obtained otherwise.

There are several important observations to be made concerning the results given above. First consider the shear stress at the wall, $\tau_w = \mu(\partial u/\partial y)_o$. One obtains from the definitions that $\tau_w = \Phi f''(0)\sqrt{(\mu u^3 / \Phi x)}$, i.e., that the wall shear stress is proportional to the $(3m - 1)/2$ power of x. For a flat plate ($m = 0$), the wall shear stress decreases along the flow direction: this is due to the fact that rotation diffuses out, and the boundary layer thickness thus increases. When $m > 0$, the external flow is accelerated, and this by itself would cause the

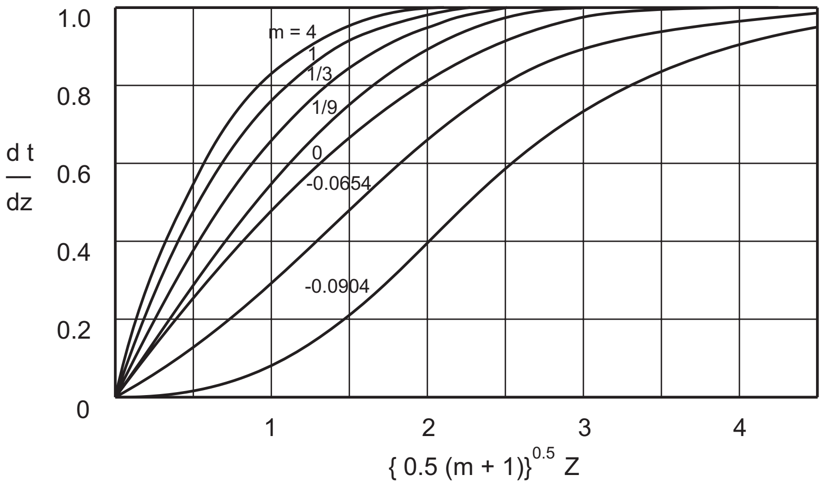

Fig. 4.6.2. Numerical solutions of Eq. (4.6.7).

shear stress at the wall to increase; however, the opposite effect of diffusion of rotation is still present, and the wall shear stress still decreases, though less rapidly, until *m* reaches the value 1/3 (corresponding to β = 22.5 degrees). At this particular value of the angle the two effects balance each other, and the wall shear stress stays constant. As *m* is further increased, the effect of the acceleration of the external flow predominates, and the wall shear stress decreases.

If one now considers negative values of *m*, the external flow is decelerated. This means that the flow takes place against a pressure gradient ($\partial P/\partial x$ is positive), and hence there is a tendency for the wall shear stress to become negative. Indeed, for $m = -0.0904$ (a value corresponding to a very small negative angle of the wedge, such as might occur in the rear part of a streamlined airfoil), the wall shear stress is zero, as the figure shows (the value of $f''(0)$ becomes zero). At values of m less than –0.0904 numerical integration becomes unstable, and this is related to the fact that the wall shear stress has become negative. In fact, a negative wall shear stress means that there is a region of backwards flow near the wall, which would result in the shedding of vortices and the formation of a wake. Indeed, the velocity profiles for negative *m* all show a point of inflection, i.e., a hint of a tendency to form a region of backwards flow.

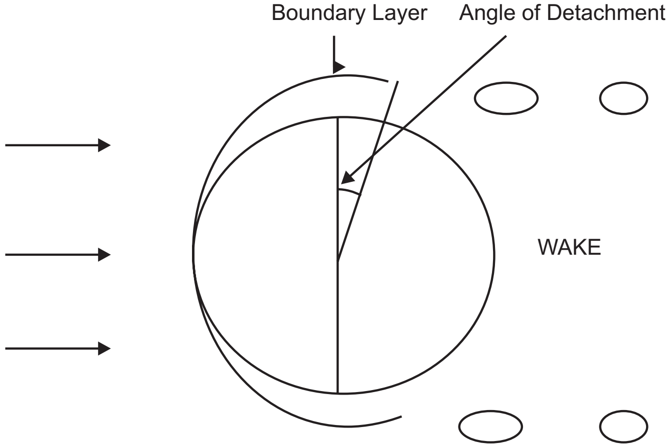

Fig. 4.6.3. Sketch of the flow field around a sphere.

It is important to realize that, in any realistic geometry, there will always be a region where the external flow is decelerated. Consider for instance the case of a sphere. From the forward stagnation point (one of the poles) up to the equator, the external flow is accelerated, since the available cross section steadily decreases. However, from the equator downstream, the external flow tends to be decelerated, and a behaviour similar to the one calculated for $m < 0$ is to be expected. In very rough terms, one would expect that formation of vortices and the subsequent *detachment* of the boundary layer would occur at an angle θ corresponding to a local value of $m = -0.0904$. This estimate yields an angle of detachment of 104.9 degrees, which is very close to the one which is observed experimentally.

Once vortices are formed and shed into the wake (see Fig. 4.6.3) the external flow, though still governed by the EE, is not irrotational any more, and hence severe distortions from the EE solution are found in the wake. Notice that the vortices shed have kinetic energy associated with their rotation, and that the EE would predict that the kinetic energy cannot be dissipated — the wake would be predicted to be infinitely long. Of course, the viscosity is not exactly zero, and hence eventually the wake will die out. However, wakes are indeed much larger than the dimensions of the object which causes them, and a rough estimate of the ratio of the volume of the wake to that of the object is the Reynolds number. The physics are as follow. The moving sphere does work on the fluid, and the rate at which it does work is of

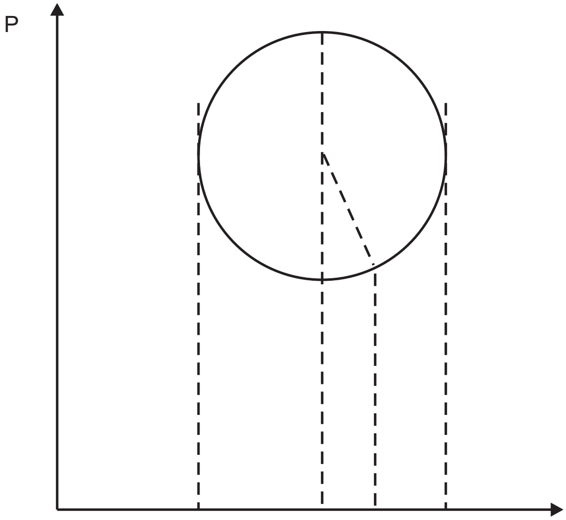

Fig. 4.6.4. Pressure distribution on the sphere.

course the product of the drag force and the velocity. This work is in part dissipated by the effect of viscosity in the boundary layer, but a large fraction of it results in the kinetic energy of the vortices in the wake. Once kinetic energy has been provided to the wake, there is no way of getting it back — it may not have dissipated yet, but it will eventually far downstream as the wake dies out. Examples of the importance of this result are discussed in Appendix 4.6.

Now what can one do concerning the calculation of the drag on a sphere? One could begin by integrating the wall shear stress effect up to the point of detachment of the boundary layer, and if one does that one comes up with a drag which is very significantly lower than the observed one. Any other suggestion? Bob suggests that we have at best estimated a fraction of the skin drag, and we have entirely left out the form drag. Good point, Bob; any suggestion about how to calculate the form drag? We know that up to the point of detachment the pressure has the EE distribution — what about after that? Let's look at Fig. 4.6.4. The EE pressure distribution is symmetrical with respect to the equator, so that its net effect is zero. However, detachment of the boundary layer now leaves us at a loss concerning what the pressure distribution at θ exceeding 104 degrees might be — any suggestions? Sue comes up with a nice idea: perhaps in the wake the pressure is simply what it was at the point of detachment. This would give the pressure distribution plotted as a dashed line, and now we have plenty of form drag — the high pressure on the fore side is not balanced by the low pressure on the hind side. If one does the calculation according to Sue's suggestion, the result is not too bad — within 20% of the experimentally observed drag. Ralph regards this as sheer luck, and doesn't believe any of it.

For the case of a sphere (or of any *bluff* body) the form drag is much more than the skin drag. This is also confirmed by the fact that the Colburn analogy fails for flow around bluff objects: the Colburn factors for heat or mass transfer are significantly less that the drag coefficients at the same Reynolds number. Of course, there is no analog of the form drag in heat or mass transfer, and hence the Colburn factors, if anything, are expected to be of the order of the skin drag only, which is much less than the total drag. If one now moves to streamlined bodies, the Colburn factors tend to become closer and closer to the drag coefficients, until for a perfectly streamlined airfoil they are approximately equal to the drag coefficient.

Ralph points out here that we are talking of streamlined bodies without having defined them. Good point. Let's suppose that we wish to design a streamlined body (see Fig. 4.6.5). We are given a minimum transverse direction H. On the fore side, we do not need to be sophisticated, since the boundary layer has no tendency to detach there, and so we make it something like a circular cylinder of radius $H/2$ (shaded area in Fig. 4.6.5). This way we have reached our required dimension H, and now we have to be careful. The hind region has to correspond to an external flow deceleration, and perhaps we can avoid detachment of the boundary layer by keeping the angle small enough, say less that 14.9 degrees. So we end up with the shape in the Figure, which is roughly the cross section of an aeroplane wing. Skin drag will be rather large, because by keeping the angle small we have had to make the body

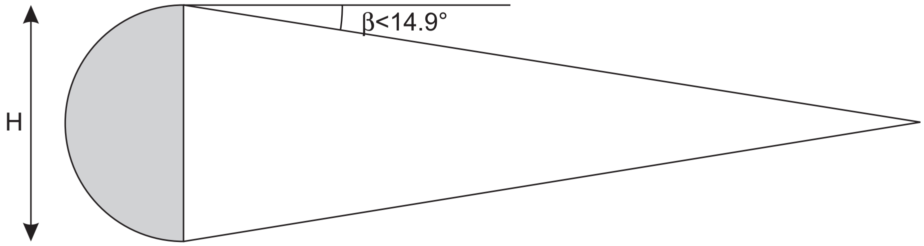

Fig. 4.6.5. A streamlined body.

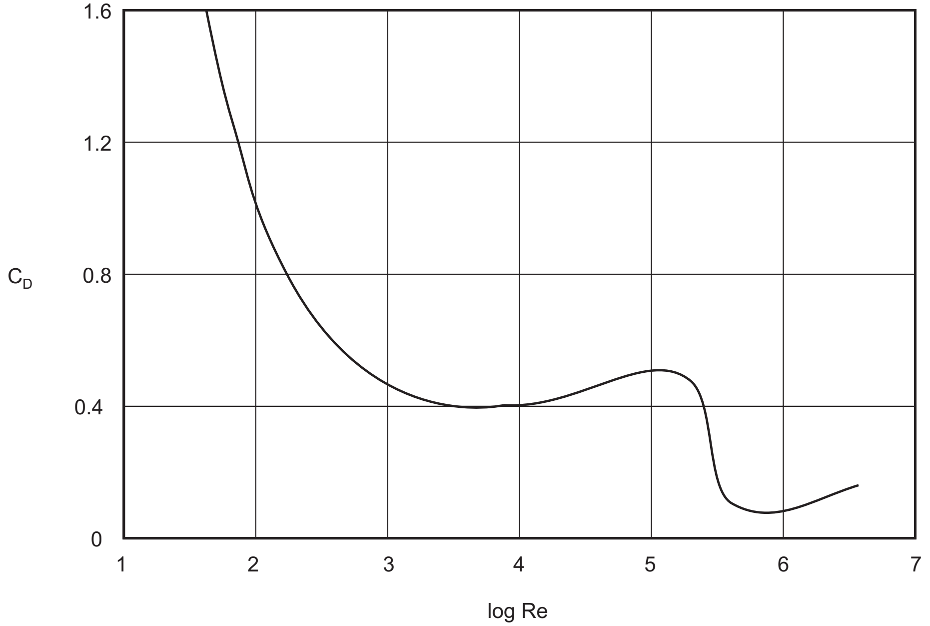

Fig. 4.6.6. Drag coefficients for flow past a sphere.

quite a bit longer than required by its minimum dimension H, but we don't care, as long as we can avoid detachment of the boundary layer, and hence save on form drag much more than we lose on skin drag.

Drag coefficients for flow around a sphere are plotted in Fig. 4.6.6 vs. the Reynolds number. The leftmost part of the curve represents Stokes' law, and it holds reasonably well up to **Re** values close to unity. A transition region follows, and after that the drag coefficient is approximately constant, as one would expect for a form drag dominated situation (the form drag is obviously proportional to U^2, as Sue's suggestion indicates). At a value of **Re** somewhat larger than 10^5, a second transition is observed, with the drag coefficient becoming again constant at a lower value than it was before. What is happening here?

A plausible explanation is as follows. As the Reynolds number is increased, the flow *within* the boundary layer may become turbulent. Does anybody in the class have a suggestion concerning how we might estimate the value of **Re** where this occurs? Ralph suggests that it should be about 1,000 — that's when

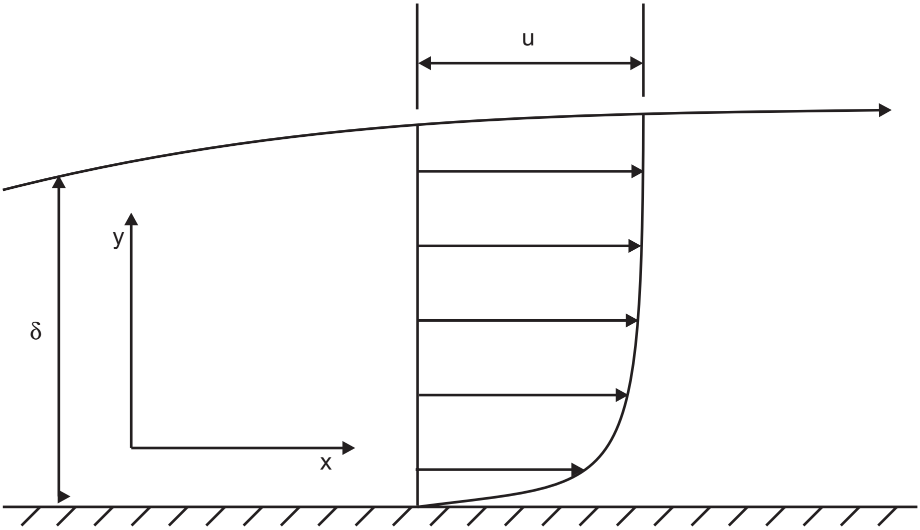

Fig. 4.6.7. Average velocity profile within a turbulent boundary layer.

laminar flows get turbulent. Good point, but there is a twist. Laminar flows go turbulent at **Re** ≈ 1,000, with **Re** based on the characteristic dimension of the tube or channel. In our case that characteristic dimension would be the boundary layer thickness δ, so let's say that the flow goes turbulent when $\delta U\Phi/\mu$ approaches 1,000. But Eq. (4.6.1) now tells us that the corresponding $xU\Phi/\mu$ (which can't be very different from **Re**) is 10^6. Well, that's not so different from the value at which the kink occurs in Fig. 4.6.6, so perhaps the decrease of the drag coefficient is related to the flow within the boundary layer becoming turbulent.

Ralph entirely disagrees — when the flow becomes turbulent pressure drops become larger, not smaller. Much to their distaste, Sue and Bob this time have to agree with Ralph. Well, let's look at the situation. Suppose the flow within the boundary layer is turbulent; the velocity profile would certainly be much flatter (see Fig. 4.6.7), and hence a larger wall shear stress is expected. Hence indeed the skin drag will increase significantly. For flow in a tube, that's the only thing which happens, since all the drag is skin drag (the pressure force at the wall being always orthogonal to the flow direction). However, here we have form drag to take into account. Since the velocity profile is flatter, and the wall velocity gradient larger, it will take more time for the deceleration of the external flow to invert the sign of the wall shear stress, and hence detachment of the boundary layer will take place at an angle larger than 104 degrees, won't it? This makes for a smaller region exposed to the low pressure of the wake, and hence for a smaller form drag. Since the form drag dominates, its decrease will result in a decrease of the total drag, in spite of the fact that the skin drag increases when the flow within the boundary layer goes turbulent. Indeed, the angle of detachment has been observed experimentally to become larger after the kink in Fig. 4.6.6. Moreover, the kink can be displaced to lower values of **Re** by purposefully inducing turbulence in the boundary layer (this can be accomplished by having a small wire fastened to the sphere at some value of θ less that 90 degrees).

4.6.1 The integral approximation

For shapes other than the wedge, a similarity solution cannot be found, and approximate techniques of solution are in order. The usual approximation scheme is the integral approximation, which is based on the following considerations. The momentum equation in the x direction can be written as:

$$\Phi(u\partial u/\partial x + v\partial u/\partial y - U\mathrm{d}U/\mathrm{d}x) = \mu\partial^2 u/\partial y^2 \qquad (4.6.1.1)$$

Now instead of trying to find a velocity distribution which satisfies (4.6.1.1) at every point within the boundary layer, one looks for one which satisfies it *on the average*. This is accomplished by taking the integral of (4.6.1.1) over the whole thickness of the boundary layer, and then looking for a solution of the integral equation. First consider the integral of the viscous term on the RHS, which is simply the difference between the values of $\mu\partial u/\partial y$ in the upper limit of integration and at zero. As long as the boundary layer thickness is small, the external flow field at its upper edge can still be approximated with the solution of the EE for $y = 0$, and hence $\partial u/\partial y \approx 0$ in the upper limit. It follows that the integral of the RHS of (4.6.1.1) is simply minus the tangential stress at the wall τ. The integral of (4.6.1.1) is thus written as:

$$\tau / \Phi = \int_0^\infty (-u\partial u / \partial x - v\partial u / \partial y + U\mathrm{d}U / \mathrm{d}x)\mathrm{d}y \qquad (4.6.1.2)$$

The integrand in (4.6.1.2) is best expressed as follows:

$$-u\partial u/\partial x - v\partial u/\partial y + U\mathrm{d}U/\mathrm{d}x = (U-u)\mathrm{d}U/\mathrm{d}x + u\partial(U1u)/\partial x - v\partial u/\partial y \tag{4.6.1.3}$$

This has introduced the quantity $U-u$, which approaches zero as y approaches the external flow field, and hence the upper limit in (4.6.1.3) may be conveniently taken as large as one wishes. (Ralph dislikes this vagueness, but Sue just loves it). The last term on the RHS of (4.6.1.3) can also be written as $v\partial(U-u)/\partial y$, since (to within the stated restriction), $\partial U/\partial y = 0$. The last term can now be integrated by parts to obtain, by considering the continuity equation:

$$\int_0^\infty v\mathrm{d}(U-u) = v(U-u)\int_0^\infty + \int_0^\infty (U-u)\frac{\partial u}{\partial x}\mathrm{d}y = \int_0^\infty (U-u)\frac{\partial u}{\partial x}\mathrm{d}y \tag{4.6.1.4}$$

It is useful now to define a *displacement thickness* δ', and a *momentum thickness* θ as follows (it is easy to convince oneself that both are of the order of the boundary layer thickness):

$$\delta' = \int_0^\infty (1-u/U)\mathrm{d}y; \quad \theta = \int_0^\infty (u/U)(1-u/U)\mathrm{d}y \tag{4.6.1.5}$$

Substitution of the definitions and algebraic manipulation yields the final result:

$$\tau/\Phi = (\delta' + 2\theta)U\mathrm{d}U/\mathrm{d}x + U^2\mathrm{d}\theta/\mathrm{d}x \tag{4.6.1.6}$$

For the special case of the flat plate, $\mathrm{d}U/\mathrm{d}x = 0$, and since $\tau \approx 1/\theta$ one recovers the result $x^2/\delta^2 \approx \mathbf{Re}$. In general, (4.6.1.6) shows that an accelerating external flow will tend to result in a positive wall shear stress, but a decelerating flow could result in a negative one, and thus in a tendency to form a backflow region near the wall.

So far, no approximation has been made except for the classical thin boundary layer approximations. Now, in order to find a solution, one has to assume a shape of the velocity distribution in the boundary layer. Of course we know that in general no similarity solution will exist, but one is looking for an *approximate* solution, and we will assume similarity anyhow. The rationale is as follows: As long as the boundary layer remains thin, the equations of motion cannot be satisfied everywhere with a similarity solution, but they may be satisfied on the average. In fact, there is freedom of choice in the shape of the velocity distribution, since different shapes can be assumed which lead to "solutions" of the equations in their integral form.

Now the assumed shape has to be a reasonable one, and the following conditions will need to be fulfilled. First, the derivative $\partial u/\partial y$ at the wall needs to be finite. It is useful to define a *stress thickness* δ as follows:

$$\tau = \mu U/\delta \tag{4.6.1.7}$$

which sets the derivative at the wall at the value U/δ. The velocity u needs to merge smoothly into U at some y value L of order δ, say $u/U = 1$ at $y > L$; this sets the upper limits of integration in (4.6.1.5) at L rather than ∞, and imposes that $\partial u/\partial y = 0$ at $y = L$. One shape satisfying these constraints is half a sinusoidal wave of wavelength $\Pi\delta$, say:

$$u/U = \sin(y/\delta); L = \Pi\delta/2 \tag{4.6.1.8}$$

For such a shape, the quantities δ' and θ are calculated as:

$$\delta' = (\Pi/2 - 1)\delta; \theta = (1 - \Pi/4)\delta \tag{4.6.1.9}$$

and are thus seen to indeed be of order δ.

When all this is substituted back into (4.6.1.6), one obtains the following differential equation for δ:

$$U^2(1 - \Pi/4)\mathrm{d}\delta/\mathrm{d}x = \mu U/\delta - \delta U \mathrm{d}U/\mathrm{d}x \tag{4.6.1.10}$$

Equation (4.6.1.10) should be considered carefully. The two terms on the RHS represent the effect of the wall shear stress and of the acceleration of the external flow on the tendency of the boundary layer to thicken in the flow direction. The wall shear stress tends always to thicken the boundary layer, since it induces rotation which is diffused into the flow field. This tendency is counteracted by the second term when $\mathrm{d}U/\mathrm{d}x > 0$, i.e., an accelerated external flow tends to compress the boundary layer thickness (for the wedge case, the two effects balance each other at $m = 1/3$, as discussed earlier). However, a decelerated external flow contributes to the thickening of the boundary layer, and hence the assumption that it is thin rapidly breaks down.

The integral approximation method is believed to be rather accurate, and is used extensively. Its accurateness can be gauged by applying it to the flat plate geometry for which an exact solution is available, and for that case the integral method yields a wall shear stress which is within less than 2% of the exact solution value.

4.6.2 The boundary layer at a free surface

Consider the case of a high Reynolds number flow field bounded by a *free* surface, say the interface with a gas phase. In order to have a concrete example, we consider the steady rise of a spherical gas bubble under conditions where the Reynolds number based on bubble velocity U and diameter D for the liquid flow field is large. Since the motion is steady, one may as well regard the bubble as stationary and the fluid moving downwards.

Contrary to what happens in the case of a solid sphere, the velocity at the interface needs not be zero, and one could superficially conclude that there is no reason why a boundary layer should develop at all. However, if one regards the viscosity of the gas as being essentially zero, the tangential stress at the interface needs to be zero, and this condition is not satisfied by the solution of the EE. It is however *almost* satisfied, since the radial gradient of the tangential velocity at the sphere surface is in fact rather small (of order U/D). It follows that a very thin boundary layer will indeed form, within which the velocity deviates slightly from the EE solution so as to establish a zero radial gradient at the interface itself. The sketch in Fig. 4.6.2.1 shows a plausible profile of tangential velocity at the equator of the sphere; the boundary layer thickness δ is expected to be of the order of $R/\sqrt{\mathbf{Re}}$. This in turns implies that the deviation of the velocity from the EE value is of order $U/\sqrt{\mathbf{Re}}$, i.e., it is very small indeed as compared to what happens on a solid surface (where the deviation is of order U).

Since the shear stress at the interface is zero everywhere, it cannot become negative in the downstream region, and hence there is very little tendency of the boundary layer to detach; a small wake will form on the rear pole, but its influence on the flow field is minor and localized. It follows that the flow is almost exactly the one predicted by the EE, i.e., it is irrotational and the velocity admits a potential ϕ, $\mathbf{v} = \text{grad}\phi$. For flow around a sphere, ϕ is given by:

$$\phi = -UR^3\cos\theta/2r^2 \tag{4.6.2.1}$$

Now if one wants to calculate the drag force F, one may reason as follows. The product FU represents the total rate of energy dissipation, and hence it must

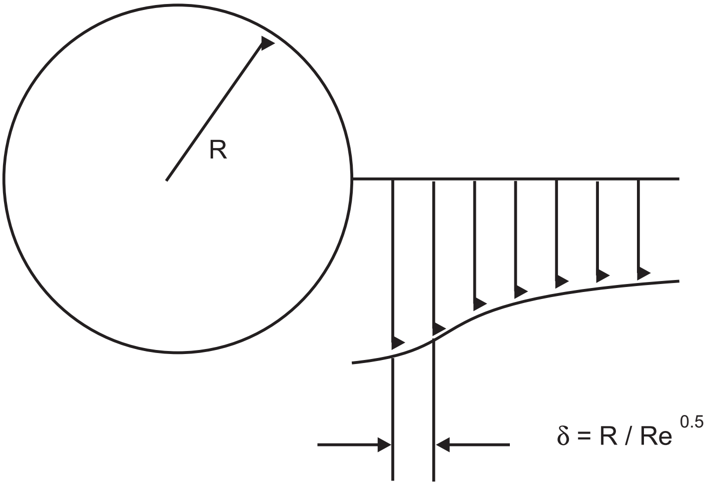

Fig. 4.6.2.1. Sketch of the tangential velocity profile near the equator.

be equal to the integral of the rate of dissipation per unit volume ϕ discussed in Section 4.2. While this is true also for flows past solid objects, it is of little utility, since most of the dissipation takes place in the boundary layer and in the wake, where the velocity distribution is not known. However, for the present problem the rate of dissipation in the boundary layer and the wake contribute little to the volume integral of ϕ, both because the volume of the boundary layer and wake is small, and because the velocity gradients within that volume are small (they are zero at the interface itself). Hence most of the dissipation takes place in the external flow field, and F can be calculated by integrating ∅ over the whole flow field given by Eq. (4.6.2.1). The result is:

$$F = 12\Pi\mu RU \tag{4.6.2.2}$$

i.e., the drag is exactly twice the Stokes law drag.

Equation (4.6.2.2) is interesting in that it is the only case where a drag force for a large Reynolds number flow is proportional to viscosity and to velocity, rather than being approximately proportional to U^2 and independent of μ. The effect of the boundary layer and wake can be calculated, and it is, as expected, in the direction of decreasing the drag: the RHS of (4.6.2.2) should be multiplied by $(1 - 2.2/\sqrt{\mathbf{Re}})$, and this is well confirmed by experimental data. The correction is of some importance, since very large Reynolds numbers are impossible because gas bubbles become nonspherical at large velocities: for gas bubbles in water, significant deviations from sphericity occur at $\mathbf{Re} > 300$. For $\mathbf{Re} = 100$, the correction term is 0.78, which is not significantly less than unity. Of course, the bubble is kept spherical by surface tension, provided $\sigma >> \Phi R U^2$.

Appendix 4.6

Addition of minute amounts of polymers to, say, water result in a significant reduction of the pressure drop for turbulent flow in a tube. The phenomenon is known as "drag reduction", and has been observed repeatedly. The amount of polymer added is small enough for the viscosity of the solution to be still essentially that of water. It is believed that the effect is due to the *viscoelasticity* of the polymer, which somehow reduces the possibility of energy dissipation in a turbulent flow field. The underlying mechanism needs not be discussed here.

Now suppose one considers flow around a submerged object of such a polymer solution: will the drag be reduced significantly? There was much hope at some time that the answer would be yes, for reasons which are not too ethical: submarine launched nuclear missiles would get out of the water at a higher speed, it was hoped, if they shed some polymer from the nose. Well, that turned out not to be the case. Vortices are formed and shed on the rear end of the submerged object, and thus kinetic energy is acquired by the fluid. Now the fluid has difficulty in dissipating this kinetic energy, and hence a longer wake will form — but the length of the wake doesn't influence the drag on the body.

There is harder experimental evidence than the one discussed above. Consider flow through a sudden enlargement from a small diameter pipe into a large diameter one. The frictional pressure drop due to such a flow configuration can be calculated from overall momentum and mass balances, and is thus independent of the details of the flow structure. If a drag reducing polymer solution is forced to flow through a sudden enlargement, the frictional pressure drop is the same as in an ordinary liquid. However, since dissipation of energy is more difficult to achieve, the region in the downstream tube where dissipation takes place becomes longer: the linear pressure profile develops farther downstream. Drag reduction only occurs in flow down a tube where all the energy dissipation has to take place within the tube itself, i.e., essentially within the boundary layer (the whole tube cross section is a boundary layer, since nowhere within it is the EE solution adequate).

4.7 INSTABILITY AND TURBULENCE

Flow fields have a nasty tendency to become unstable. The problem of course has to do with the fact that solutions to the Navier–Stokes equations are not unique, and that therefore once one has found a solution to them there is no guarantee that it will be the one which is going to be observed in actual fact. We will not discuss flow instabilities in any detail in this section, and we limit ourselves to the essential features and available results in the area.

Suppose one has some flow problem for which the boundary conditions are steady. One then looks for a solution of the steady state Navier–Stokes equations SNSE:

$$\Phi\mathbf{v}\cdot\mathrm{grad}\mathbf{v} = -\mathrm{grad}P + \mu\mathrm{lapl}\mathbf{v}$$

Suppose one finds such a solution, say $\mathbf{v} = \mathbf{v}(x,y,z)$. Since time t has not entered our analysis, $\partial\mathbf{v}/\partial t$ is guaranteed to be zero, and hence $\mathbf{v}$ is also a solution of the complete Navier–Stokes equations (CNSE). One can now ask oneself the following question: if some disturbance is superimposed on the steady flow field, will it eventually die out or not? Quite obviously, the answer to this question can only emerge from consideration of the CNSE, since one is asking a question about evolution in time which is meaningless in the context of the SNSE.

Now the question as formulated above is quite hard to answer, and one may settle for a less ambitious one. We ask whether the basic flow field is stable to infinitesimally small disturbances (this presumably will lead to some kind of linearization). Now suppose the answer to this question is no. Since infinitesimally small disturbances are impossible to avoid, we conclude that the flow field is unstable. This means that the disturbances will grow, and hence they will not stay infinitesimal — so our analysis won't tell us what is going to happen, it will only tell us that something rather dramatic will happen. Conversely, suppose the answer to our question is yes — infinitesimal disturbances will tend to die out. That does not guarantee that large disturbances will die out — the flow may still be unstable to some large enough disturbances. Stability analysis only answers the questions which one asks (when it does answer them), and one should not expect it to give results it is not meant to give.

Now suppose we limit ourselves to linear stability, i.e., stability to infinitesimal disturbances. We still have a major problem: what kind of disturbances are we going to analyze? After all, an infinite variety of disturbances could be considered, and the flow may be stable to some of them but not all of them. The standard procedure is to consider disturbances which are decomposable into normal modes: that kind of decomposition is known to work for a variety of linear problems, though it is not guaranteed to be appropriate for all flow fields. In order to be more specific, consider a very simple basic flow where **v** has only one nonzero component, say the x-one V, and that V depends only on y (this is a shear flow). One would now consider a disturbance expressible as:

$$v = v^*(y)\exp[i(\alpha x + \beta y - \alpha c^* t)] \tag{4.7.1}$$

where starred quantities are complex. The wave number k and complex wave speed are:

$$k = \sqrt{(\alpha^2 + \beta^2)} \tag{4.7.2}$$

$$c^* = c_R + ic_I \tag{4.7.3}$$

The plan of operation is as follows. One substitutes (4.7.1) into the CNSE (with **v** being the basic flow plus the disturbance), drops out all terms which are quadratic in the disturbance (that's when we use the infinitesimal disturbance assumption), and one obtains a linear equation for the disturbance. One now looks for the eigenfunctions and eigenvalues of the solution, and if everything works out smoothly one can determine the conditions under which c_I is negative, which are the stable ones (the disturbance decays as $\exp(-c_I t)$. A nice program of work, but a very hard one to implement.

Now let's look at some limiting cases. As was discussed in section 4.5, there is an uniqueness theorem for the creeping flow equations, and therefore every flow field becomes stable at **Re** $= 0$. However, flows with vanishingly small **Re** may well be unstable, and one should not conclude that viscosity has always the effect of stabilizing the flow. Indeed, we know that viscosity results in a diffusion of both rotation and shear stress, the diffusivity being μ/Φ. There may well be conditions where this diffusion results in shear stress getting in phase with the pressure fluctuation corresponding to a travelling wave, thus destabilizing an otherwise stable flow field.

At the other extreme, there is no uniqueness theorem for the Euler equations. There is a theorem of uniqueness for irrotational solutions of the Euler equations (the solution is the one which minimizes the total kinetic energy of the flow field, a result due to Kelvin which is vaguely reminiscent of the Helmholtz–Korteweg variational principle), but this does not apply to shear flows since such flows are rotational. Indeed, for the shear flow considered before, there are two general theorems due to Rayleigh which apply to the inviscid Euler equations:

1. A necessary condition for flow instability is that $V(y)$ has a point of inflection, say $d^2V/dy^2 = 0$ at some position $y = y_o$. (A slightly stronger form of this first theorem was obtained by Fjortoft: if there is a point of inflection, and U_o is the velocity at y_o, a necessary condition for instability is that $(U - U_o)d^2U/dy^2 < 0$ at some y).
2. The wave speed c is intermediate between the minimum and maximum values of U. It follows that $U = c$ somewhere in the flow field. That position is called the critical shear layer, for reasons which will become clear later.

The simplest case of a shear flow which produces this type of inviscid instability is the one where, for $x > 0$, $V = V_1$ at $y > 0$, and $V = V_2$ at $y < 0$. This is the kind of inviscid flow which could result from two streams with different velocities on the two sides of a flat plate coming in contact with each other at the plate edge $x = 0$. This flow is irrotational everywhere except at $y = 0$, where a vortex sheet exists (the local shear is infinitely large). This flow is unstable to all wavelengths, i.e., under no conditions will it be observed in actual fact. This is not surprising, since decelerating flows are rather generally unstable, and the stream with a higher initial velocity will tend to decelerate as soon as there is any transverse disturbance.

Things become remarkably more complex if both viscosity and inertia are taken into account, as unfortunately one has to do in most case of pragmatical relevance. One can however limit oneself to flows which are *almost* parallel, say flows where the basic flow field is expressible as:

$$\mathbf{v} = U(x,y), V(x,y), 0 \tag{4.7.4}$$

$$V << U; \partial/\partial x << \partial/\partial y \tag{4.7.5}$$

Lubrication type flows, and two-dimensional boundary layers, are of this type. Here the very powerful Squire theorem comes to our help: if a three-dimensional disturbance results in an instability, a two-dimensional disturbance exists which is more unstable. This means that we can confine attention to two-dimensional disturbances. Two-dimensional flows admit a stream function, and this is true for the disturbance as well because the continuity equation holds for both the basic flow and the disturbance. Hence one can work with the disturbance stream function q:

$$u = \partial q/\partial y; v = -\partial q/\partial x \tag{4.7.6}$$

$$q = p^*(y)\exp[i(\alpha x + \beta y - c^* t)] \tag{4.7.7}$$

When this is substituted into the CNSE, terms are discarded according to (4.7.5), and quadratic terms in the disturbance are also discarded, one obtains the Orr– Sommerfeld equation:

$$(U-c)[\mathrm{d}^2 p^*/\mathrm{d}y^2 - \alpha p^*] - p^*\mathrm{d}^2 U/\mathrm{d}y^2 = [\mathrm{d}^2/\mathrm{d}y^2 - \alpha^2]^2 p^*/i\alpha\mathbf{Re} \tag{4.7.8}$$

The Orr–Sommerfeld equation is linear, but is nonetheless extremely hard to solve. When the solution can be found, it produces both the eigenfunction p^* and the complex wave speed c, and hence one obtains:

$$c_\mathrm{I} = c_\mathrm{I}(\alpha, \mathbf{Re}) \tag{4.7.9}$$

The flow is stable for **Re** values such $c_\mathrm{I} < 0$ for all values of α; it becomes unstable at the smallest **Re** value for which $c_\mathrm{I} > 0$ for at least one wavelength.

The Orr–Sommerfeld equation simplifies for inviscid flows, since the RHS becomes zero. When $U = c$ the LHS becomes degenerate, and this is the reason why the position at which $U = c$ (which is guaranteed to exist by Rayleigh's second theorem) is the critical shear layer.

Sue has lost all interest in what we are saying, since she is convinced that stability analysis is way too complicated for her to learn to do it; furthermore, she sees no point in even trying. Ralph entirely disagrees, and he would like to embark on a program of massive algebra and numerical computation. Books on stability theory are available, and he is urged to go ahead and study them in detail. Bob has a different idea: since all this stability analysis does not tell us what is going to happen, only that something is going to happen, couldn't we just leave it at that and move to consideration of what actually does happen? His proposal gets a 2 to 1 majority vote, and thus we move to turbulence.

Let's first describe somewhat qualitatively what a turbulent flow might be. In particular, we consider a very simple geometry, say flow between two

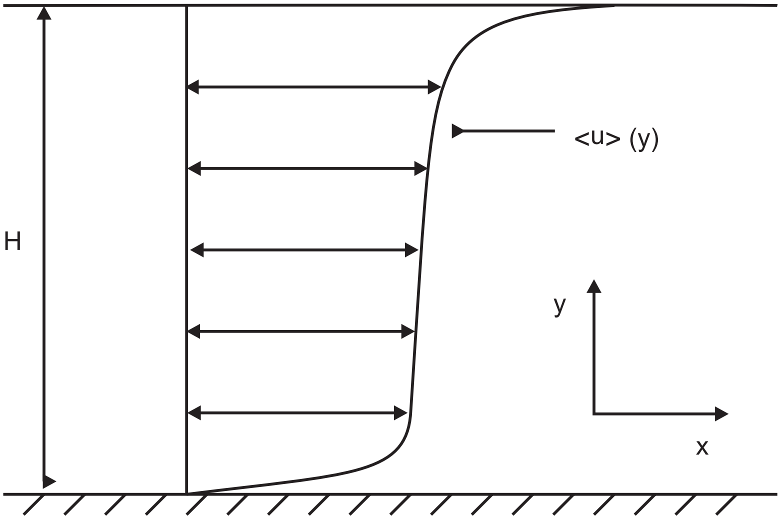

Fig. 4.7.1. Turbulent flow between two parallel plates.

parallel plates, one stationary and one moving with steady velocity U (see Fig. 4.7.1). The steady flow solution is trivial:

$$\mathbf{v} = Uy/H, 0, 0 \tag{4.7.10}$$

$$\mathrm{grad}P = 0 \tag{4.7.11}$$

Now if the flow goes turbulent, one would observe that pressure and all three components of velocity oscillate in what may appear as a random fashion at first sight; a representative response of a probe might look something like sketched in Fig. 4.7.2. In principle, we believe that the CNSE should be able to describe the turbulence. In practice, this is a hopeless task, and even if we could do it we wouldn't really want to: after all, we are only interested in the average values of all quantities, not a detailed description of their random fluctuations. So we plan on the following program. First, we need to define *average* and *fluctuating* quantities unequivocally. We then substitute these in the CNSE, and we'll try to get rid of the fluctuations by some sort of averaging technique. If this program is successful, we'll have obtained equations for the average values, and hopefully we can try to solve those.

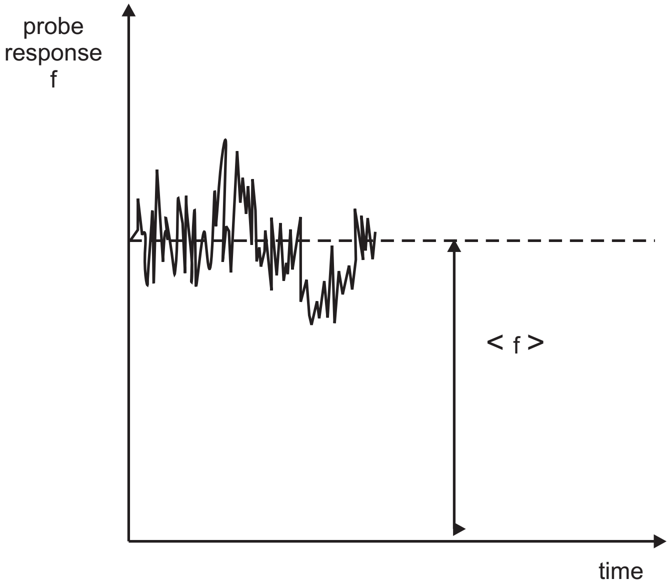

Fig. 4.7.2. Typical response of a probe in turbulent flow.

Let's first define the average $<f>$ of any quantity f. The average is over some appropriate time interval T, say:

$$<f>(t) = (1/T) \int_{t}^{t+T} \mathrm{d}t \tag{4.7.12}$$

Now how are we going to choose the value of T? Well, if we choose it too small, the value of $<f>$ may well depend on what T we choose: we might, for instance, have averaged over a peak in Fig. 4.7.2. But if we take T large enough to contain a statistically significant sample of the fluctuations, we should be OK. For basic flows which are steady, we expect $<f>$ to be independent of time t.

Having defined the average, we can now define the fluctuation f' as:

$$f = <f> + f' \tag{4.7.13}$$

Now notice that, by definition, $<f'> = 0$, i.e., the average of the fluctuation is zero. This is what gives some hope to our program of getting rid of the fluctuations by averaging. The other important point is that the average is a linear operator, and that therefore it commutes with all linear operators. Let L be any linear operator. We can write:

$$<Lf> = L<f>;\ <Lf'> = 0 \tag{4.7.14}$$

This means that when we substitute (4.7.13) into the CNSE, and then take the average of all terms, all the linear terms will reproduce themselves with the average instead of the instantaneous value. Our program would be completed if all the terms were linear, but this is unfortunately not the case, and we have to deal with the nonlinear term $\mathbf{v}\cdot\mathrm{grad}\mathbf{v}$. Substitution of (4.7.13) yields four terms. The first one is $<\mathbf{v}>\cdot\mathrm{grad}<\mathbf{v}>$, and this simply reproduces the mean convected inertia. The two terms which contain both $<\mathbf{v}>$ and $\mathbf{v}'$ are linear in $\mathbf{v}'$, and thus their average is zero. The only term containing $\mathbf{v}'$ which survives is:

$$<\mathbf{v}'\cdot\mathrm{grad}\mathbf{v}'> = \mathrm{div}<\mathbf{v}'\mathbf{v}'> \tag{4.7.15}$$

For reasons which will soon become clear, $\Phi<\mathbf{v}'\mathbf{v}'>$ is called the Reynolds stress tensor $\mathbf{T}'$. The averaged turbulence equations are thus written as:

$$\Phi\mathrm{D}<\mathbf{v}>/\mathrm{D}t = -\mathrm{grad}<P> + \mu\mathrm{lapl}<\mathbf{v}> - \mathrm{div}\mathbf{T}' \tag{4.7.16}$$

Equation (4.7.16) shows that the turbulence equations are just the same as the Navier–Stokes ones, except for the additional term $\mathrm{div}\mathbf{T}'$. That term appears in a form which suggests regarding $\mathbf{T}'$ as an additional stress tensor, hence the name. Let's try to understand the physics involved by looking at our simple geometrical example. We expect to have only one nonzero average velocity $<U>$, but fluctuating velocities may exist in all directions. The upper plate exerts an average tangential stress Φ on the fluid, and since on the average the fluid is not accelerated, on the average that same stress is exerted by the fluid above any plane parallel to the bounding planes on the fluid below it. Now viscous forces are linear, and hence the average stress exerted by viscous forces is simply $\mu\mathrm{d}<U>/\mathrm{d}y$. However, there is now an additional mechanism for transferring x momentum in the y direction. Due to velocity fluctuations, instantaneously there is some flow in the y direction, though it averages to zero. However, when some material crosses the plane in the y direction, it comes from a region with a higher value of x-velocity, and *vice versa* when it crosses the plane in the other direction. It follows that, although the average flux of mass is zero, the average flux of momentum is not.

Indeed, let's do the calculation. The instantaneous mass flow-rate per unit area in the y direction is $\Phi v'$. It follows that the instantaneous flux of x momentum is $\Phi v'u = \Phi v'(<U> + u')$. If we now take the average of this expression, the term containing $<U>$ averages to zero, and we obtain $\Phi<v'u'>$, i.e., the x–y component of the Reynolds stress tensor. Notice that the Reynolds stress for this flow can be measured without actually measuring any fluctuation, i.e., even with slow probes. First, measurement on either the upper or the lower plate yields τ. If the profile of average velocity is measured (which is easy: one needs a velocity probe with a slow enough response to smooth out the fluctuations), one can calculate the Reynolds stress from $\tau = \mu \mathrm{d}<U>/\mathrm{d}y + \Phi<u'v'>$.

The profile of $<U>$ which is observed experimentally is of the type shown in Fig. 4.7.1: it is almost flat everywhere, except in thin regions near the two plates where the velocity gradient is very steep. Now at the wall itself there can be no v' fluctuation, and hence Φ is transmitted to the fluid entirely by viscosity, which sets the value of the gradient of average velocity at the two walls. As one moves towards the centre, the viscous contribution rapidly decreases in importance, and momentum is transported essentially by the turbulent fluctuations. These observations apply to all turbulent flow fields, where the average viscous stress is always negligible except in thin layers adjacent to solid boundaries.

Bob has a comment at this stage. He is dissatisfied with Eq. (4.7.16), since it contains too many unknowns ($<\mathbf{v}>$ and $\mathbf{T}'$), and thus cannot be solved for either one. Very good point, we were not able to get rid of the fluctuations after all. Does anybody have a suggestion about what to do next? This time not even Ralph thinks that one could solve the problem by going back to the CNSE and instantaneous values. Sue has an idea: we wrote the Navier–Stokes equations by substituting a constitutive equation for stress in the div**T** term; couldn't we do something of the same kind here? After all, the mechanism of momentum transfer we described reminded her of something she had studied about ordinary viscosity being due to molecules which fluctuate — from then on her statements become rather vague, but we have to admit she has a very good point here. Indeed, viscous transfer of momentum is due to molecular-scale fluctuations, and the Reynolds stresses are due to fluctuations on a scale which is not very clear, but is certainly much larger than molecular. Well, shall we try to pursue this idea? Does anybody have any suggestion?

Ralph thinks that our analogy with molecular transport of momentum is nonsense. His argument is as follows: if the analogy is correct, then the constitutive equation for $\mathbf{T}'$ should look something like $\mathbf{T}' = 2\mu'<\mathbf{D}>$, with μ' a "turbulent" viscosity. So the net effect of turbulence would simply be that of increasing the effective viscosity, but this does not make sense, since if that were the case the velocity profile of $<\mathrm{U}>$ would be the same as in laminar flow, and it isn't. Very good argument indeed; we run into trouble if we think of a turbulent viscosity as such. Still, our analogy with molecular transport is tempting, and let's try to see where the difficulty lies. In the molecular theory, the diffusivity of momentum μ/Φ comes out to be a characteristic molecular velocity times a characteristic molecular length scale. So if we want to pursue this analogy, we need to consider a turbulent diffusivity of momentum $\mu/\Phi = E$, and we expect E to be the product of a characteristic velocity W and a characteristic length L, $E = WL$. W has got to be of the order of magnitude of the fluctuation velocity itself, and L has to be of the order of a mixing length of the fluctuation. However, we now have a problem: the value of W is zero at any solid boundary, and hence we expect E to depend on distance from the solid wall.

Now Sue has a problem. The same difficulty applies also to the molecular model, and yet we consistently regard the molecular viscosity as independent of the distance from the wall. Well, that's quite true, but the scale of distance from the wall for which the problem arises is the molecular scale, and that we are willing to regard as vanishingly small. Here the scale is of the order of the thickness of the high velocity gradient layer, and that is by no means negligibly small. Indeed, looking back at Fig. 4.7.1, the mean velocity profile in the central region is not very different from what it would be in the laminar case, $\mathrm{d}<U>/\mathrm{d}y$ is approximately constant. Only that this time we cannot simply forget about the layer at the boundary, since most of the action is taking place there.

Bob suggest that we go ahead in a more constructive way. He says that E depends on distance from the wall, so let's get down to business and let's write a constitutive equation for E. Good idea. Let's sketch out the program of what we want to do. Limiting ourselves to the simple geometry in Fig. 4.7.1, we know the average xy component of the momentum flux, Φ, and we'll write this as:

$$\Phi = (\mu + \Phi E)\mathrm{d}<U>/\mathrm{d}y \tag{4.7.17}$$

Next we need to write E as a function of y (which is distance from the wall if we consider the lower plate). How are we going to do that? Bob suggests that we begin by considering that, since $E = 0$ at the wall itself, there must be a region within which $E = 0$ is still a good approximation. Following his advice, we calculate $<U>$ from (4.7.17) setting E to zero, and we obtain:

$$<U> = \Phi y/\mu \tag{4.7.18}$$

This is not very exciting, and we add a little twist to it. Let's make $<U>$ dimensionless — what velocity scale do we have? Not the mean flow velocity, because we are only looking at a region near the wall; so we need something determined by conditions at the wall. τ/Φ has dimensions of a velocity squared, so we'll call u^* the quantity $\sqrt{\tau / \Phi})$, and u^+ the corresponding dimensionless mean velocity, $u^+ = <U>/u^*$. We substitute into (4.7.18) and obtain $u^+ = \Phi u^*y/\mu$. $\Phi u^*y/\mu$ is properly dimensionless, and we call it y^+, so that (4.7.18) takes the very simple form:

$$u^+ = y^+ \tag{4.7.19}$$

What do we do next? Bob says that since E is zero at the wall, assuming it to be a linear function of distance from the wall should be reasonable. Well, it's a good idea, but we need to be a bit careful about it. We concluded that $E = 0$ at the wall because there can be no fluctuations at the wall itself, i.e., that the velocity scale W went to zero there. Now as we move away from the wall, what may W be? Turbulence decays in the absence of average shear, and hence W must somehow be related to $\mathrm{d}<U>/\mathrm{d}y$. Given some intrinsic length scale L, the best we can do is to assume that W is proportional to $L|\mathrm{d}<U>/\mathrm{d}y|$. With that much background, we write the Prandtl *mixing length* constitutive equation for E:

$$E = L^2|\mathrm{d}<U>/\mathrm{d}y| \tag{4.7.20}$$

Now we are in much better shape, since Bob's suggestion is more easily implemented. $\mathrm{d}<U>/\mathrm{d}y$ is an external quantity, and thus we need to write a constitutive equation for L, which has units of a length, and hence is quite naturally written as $L = ky$ (k is known as von Karman's constant. A somewhat philosophical problem about Eq. (4.7.20) is discussed in Appendix 4.7). Since we are trying to write an equation which holds at distances from the wall where E is not close to zero any more, we might as well jump to the other extreme and neglect μ in comparison to ΦE in Eq. (4.7.17), which thus yields:

$$\tau = \Phi k^2 y^2 (\mathrm{d}<U>/\mathrm{d}y)^2 \tag{4.7.21}$$

We find out with hindsight that u^* was indeed very well chosen, since Eq. (4.7.21) can be rearranged to:

$$1 = ky + \mathrm{d}u^+/\mathrm{d}y^+ \tag{4.7.22}$$

Equation (4.7.22) integrates to:

$$u^+ = (1/k)\ln y^+ + \mathrm{C} \tag{4.7.23}$$

Now the two constants k and C are not known at this stage — any idea about how we could determine them? There are two suggestions - Ralph is silent because he doesn't believe any of this, our arguments are way too hand-waving for him to accept them. Sue suggests that (4.7.23) must merge with (4.7.19), and that therefore we can at least extract a relationship between k and C, though she doesn't feel confident doing that herself. Bob suggests that if (4.7.23) has any value, it must be confirmed by experimental data, and hence k and C can be extracted from those data. Let's follow these ideas. Sue's suggestion means that, if (4.7.19) holds (at least approximately) up to a y^+ value of, say, m, (4.7.23) should give the same value of u^+ at $y^+ = m$. This translates to the following equation:

$$m = (1/k)\ln m + C \tag{4.7.24}$$

It was a good try, Sue, but it didn't lead us anywhere, since we don't know what m may be. If we follow Bob's advice, we get the result that $k = 0.4$, but C depends on how smooth the surface is, with values between 3 for a smooth wall and 5 for a very rough one. Equation (4.7.24) yields $m \approx 8$ for a smooth wall, and this is confirmed by experimental data.

What have we accomplished with all this? Have we done anything more than elaborate curve fitting? Yes, indeed we have: the velocity profile near a wall in turbulent flow, expressed as $u^+(y^+)$, is a *universal* velocity profile, which does not depend on the geometry of the main flow field. This conclusion is well confirmed by experimental data, and hence we have acquired a procedure for calculating the mean tangential velocity profile near the wall for any turbulent flow past a solid wall.

From the viewpoint of chemical engineers, the most important result is, however, another one. Suppose we consider a heat transfer problem in turbulent flow. We could go through the same type of analysis, and come to the conclusion that, in addition to the average conductive flux orthogonal to a solid boundary, $k\mathrm{d}<T>/\mathrm{d}y$, there is a turbulent flux which is given by $\Phi c<v'T'>$, so that the analog of Eq. (4.7.17) would be:

$$q = (k + \Phi c E')\mathrm{d}<T>/\mathrm{d}y \tag{4.7.25}$$

E' is now the turbulent diffusivity of heat. At this point, one can make a reasonable assumption: since the same turbulent fluctuations are responsible for diffusion of both momentum and heat, it makes sense to suppose that $E' = E$. With that assumption, we may not be able to calculate either the friction factor or the Nusselt number, but we can obtain a relationship between the two quantities. If this avenue of thought is pursued, one does indeed get a relationship which is very close to the classical Colburn analogy for turbulent flow in a tube.

The point discussed in the preceding paragraph should be analyzed in somewhat more detail. First of all, consider the fact that turbulent transport predominates, in the case of momentum, if $E >> \mu/\Phi$, and in the case of heat if $E' >> k/\Phi c$. Now if the flow is turbulent, we know that $E >> \mu/\Phi$ except in a thin layer adjacent to the wall. If $E = E'$, we also know that $E' >> \mu/\Phi$, but that does not guarantee that $E' >> k/\Phi c$, unless the Prandtl number **Pr** $= \mu/kc$ is of the order of unity or larger. Now **Pr** ≈ 1 in gases, and **Pr** > 1 in ordinary liquids, so no problem arises. But with liquid metals **Pr** $<< 1$, and hence even in turbulent flow of liquid metals conductive heat transport may be significant, and indeed liquid metal heat transfer is not adequately correlated by the Colburn analogy.

Appendix 4.7

The philosophical problems arising in connection with writing a constitutive equation for the turbulent diffusivity are as follows. One needs necessarily to allow for the variation of E with distance from the wall, y. This however means that any given local mean kinematics (say at an assigned value of $\mathrm{d}<U>/\mathrm{d}y$), the Reynolds stress is not uniquely determined — it depends also on how far from the wall that kinematics is achieved. It is easy to convince oneself that this is indeed the case, since one can realize any value of $\mathrm{d}<U>/\mathrm{d}y$ at the wall itself, and hence one could have, under appropriate conditions, zero Reynolds stress at whatever average velocity gradient one chooses to consider. This may seem a minor matter, but it is not. Practically all the constitutive equations used in engineering science conform to the so called principle of local action, i.e., that the observed behaviour at a material point depends only on what has happened to a neighbourhood of that same material point. The principle of local action excludes long range interactions, and it allows to write the equations governing any phenomenon as differential equations. An E value changing with y clearly does not conform to the principle of local action, and indeed physically the effect of the wall is a long distance interaction. One comes away with relatively simple equations because one considers very simple geometries (essentially, tangential flow past an approximately plane solid boundary), but in general a long distance interaction would give rise to integro differential equations.

Another important point to be stressed is that constitutive equations for the Reynolds stress do not need to conform to the strict requirement of invariance of other constitutive equations; specifically, vorticity can enter the constitutive equation for the Reynolds stress. The reason is here that a constitutive equation for the Reynolds stresses is one for an averaged quantity, and hence inertia at the microscale subsumed by the averaging procedure is relevant. Inertia is a mechanical concept, and the balance of momentum is not invariant under a non-Galileian change of frame, such as would occur if one chooses a frame with respect to which the local vorticity is zero.

4.8 PURELY VISCOUS NON-NEWTONIAN FLUIDS

In this and the next section, we discuss some fundamental points about the mechanics of non-Newtonian fluids. The subject is a very wide one, and we do not in any way try to cover it in any detail, but only to present its essential features.

The constitutive assumption which is the definition of an incompressible purely viscous fluid is that the stress tensor **T** is a unique, though possibly non-linear, function of the rate of deformation tensor **D**. The most general constitutive equation of this type is of the form of Eq. (2.3.22), which is rewritten below with slightly different symbols:

$$\mathbf{T} = p\mathbf{1} - 2\mu\mathbf{D} - \mu'\mathbf{D}\cdot\mathbf{D} \qquad (4.8.1)$$

where μ and μ' are the viscosity coefficients, which may depend only on the invariants of **D**, and p is pressure, which is not delivered by the constitutive equation because the fluid is incompressible (see Section 4.1). The first invariant of **D** is zero by definition in an incompressible fluid, and hence we only have to worry about the second and third one.

Consider steady flow between two parallel plates, the lower one fixed and the upper one moving. Since the flow is steady, and there is no externally imposed pressure gradient, the tangential stress Φ must be the same at all values of y, as can be seen by writing a force balance between any two values of y. Since at the upper plate the tangential stress is the tangential force per unit area needed to keep the plane in steady motion, the value of Φ is measurable. Presumably there is only one nonzero velocity component, u, and it depends only on y. Let Γ be the value of du/dy. The matrix of the components of **D** is given by:

$$\mathbf{D} = \tfrac{1}{2}\Gamma \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}; \quad \mathbf{D}\cdot\mathbf{D} = (\Gamma^2 / 4) \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix} \tag{4.8.2}$$

If (4.8.1) holds, the matrix of the components of stress is:

$$\mathbf{T} = p1 = \mu\Gamma \begin{vmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix} + (\mu''\Gamma^2 / 4) \begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{vmatrix} \tag{4.8.3}$$

Let α be defined as in Eq. (4.5.15). For the flow field considered, both μ and μ'' can depend only on α, since the third invariant of **D** (the determinant of its matrix, β) is zero. Notice that we have established that τ is the same at all values of y, and (4.8.3) tells us that $\tau = \mu\Gamma$; but α is uniquely determined by Γ (it is in fact $<<b^2/4$), and hence constancy of τ implies constancy of Γ: the velocity profile is triangular. This in turn implies that Γ equals the velocity of the upper plate divided by the distance between the two plates, and is thus measurable. It follows that in principle $2\mu = f(\alpha)$ is measurable. Of course there are no infinitely long plates viscometers available in real life, but a variation of the experiment considered can in fact be performed.

Of course, such an experiment doesn't tell us what $\mu(\alpha,\beta)$ might be when β is not zero. However, let's keep with the parallel plate experiment a bit longer, and let's ask ourselves the following question: what is the orthogonal force needed to keep the upper plate from moving towards (or away from) the lower one? It's a legitimate question, and (4.8.3) implies that it should be zero, since one can easily convince oneself that the orthogonal force is proportional to the "first normal stress difference", i.e., the difference between the *xx* and *yy* components of stress. Now in actual fact, whenever $f(\alpha)$ is observed not to be constant, one also measures a finite normal force, and this is something which Eq. (4.8.3) cannot predict, no matter what the $\mu''(\alpha,\beta)$ function might be. Inclusion of the second viscosity coefficient μ'' does not serve any purpose, since it yields the wrong prediction in the simplest possible flow field, and hence we may apply Occam's razor and simply assume the $\mu'' = 0$. Of course, with this one gives up the hope of predicting what the first normal stress coefficient might be, but one has a simple way of dealing with the fact that the apparent viscosity in a pure shear flow, i.e., the ratio of τ to Γ, is not a constant but does depend on Γ itself. This is the reason why the constitutive equation for purely viscous fluids is usually written in the form of Eq. (4.5.14), which is reported below:

$$\mathbf{T} = p\mathbf{1} - f(\alpha)\mathbf{D} \tag{4.8.4}$$

Notice that (4.8.4) always yields the result that tr$\mathbf{T} = 3p$ for an incompressible fluid. That was not the case for Eq. (4.8.1).

It was already shown in Section 4.5 that a variational principle of the Helmholtz–Korteweg type can be written for any assigned form of $f(\alpha)$; the quantity whose integral has to be minimized in the solution of creeping flow problems is proportional to the entropy production (or energy dissipation) rate only when $f(\alpha)$ is of the power law form (see also Section 3.3.1). The existence of a variational principle also ensures that solutions of the creeping flow equations are unique for any form of $f(\alpha)$; this is a powerful result, because fluids which exhibit shear dependent viscosity generally have such large values of the apparent viscosity as to make the creeping flow approximation almost universally valid.

The point in the preceding paragraph is somewhat hand-waving, as Ralph is quick to point out. What is the Reynolds number of a purely viscous fluid with nonconstant viscosity? Good point indeed. The order of magnitude of the inertia forces per unit volume is still $\Phi V^2/L$; what about the viscous forces? Bob suggests that it is $f(\alpha)V/L^2$. That's reasonable, but what is the order of magnitude of the value of $f(\alpha)$? We need to estimate the value of α itself, and that is clearly V^2/L^2; now if we know the function $f(\alpha)$ we can estimate the value of it, and we can calculate the Reynolds number as $\Phi VL/f(V^2/L^2)$. If we do that, in most flow fields of interest we will conclude that the creeping flow equations are indeed appropriate.

A common form taken for $f(\alpha)$ is the power law form:

$$f(\alpha) = 2K\,\alpha^{(n-1)/2} \tag{4.8.5}$$

where K is called the "consistency" and n the "power law index", with K becoming the viscosity in the Newtonian case $n = 1$. Equation (4.8.5) usually fits data rather well over a comparatively wide range of α values. However, it fails to fit the data at sufficiently small α values, where all fluids approach Newtonian behaviour.

In the case of a power law fluid, the Reynolds number is given by:

$$\mathbf{Re} = \Phi V^{1-n} L^{1+n}/K \tag{4.8.6}$$

and it is the only dimensionless group of importance, since (4.8.5) introduces only one-dimensional parameter, the consistency K (see Section 3.3.1). Of course the dimensionless power law index will appear in the equations as well, but that is a constant for any given fluid if (4.8.5) is assumed to hold. One should however be very careful with conclusions based on the argument that there is only one dimensionless group. One such argument might be as follows: the ratio of the entry length to the diameter for flow down a pipe is a unique function of the Reynolds number; hence if we change V and L in such a way as to keep **Re** the same, we should observe the same entry length to diameter ratio. Well, it doesn't work that way, because (4.8.5) is only an approximation. Since the behaviour approaches that of a Newtonian fluid at sufficiently low values of Γ, there must be some characteristic value Γ^* at which the transition from power law to Newtonian behaviour occurs. This is a now dimensional quantity, and now we have a second dimensionless group $\Gamma^* L/V$.

Let's now consider the solution of actual flow problems for purely viscous fluids. An important consideration is the following one. In the creeping flow approximation, inertia forces are neglected. This, however, does not make the resulting equations linear, unless $f(\alpha)$ is a constant; hence the solution is generally far from being easy. In particular, the powerful technique of superimposing solutions of simple problems to get that of a more complex one is not available. An example of this is discussed in the next subsection. The fact that the techniques of solution are not easy is the reason why, in spite of the shortcomings discussed before, the power law constitutive equation is in fact often used: it offers the great advantage of comparative simplicity.

The considerations above show that non-Newtonian fluid mechanics is a very different game than its Newtonian counterpart. Unless a fluid is Newtonian, its behaviour in response to deformation is generally so complex that finding a constitutive equation capable of describing it in all conceivable flow fields is a hopeless task. Hence one chooses to write a tractable constitutive equation, even though one knows that it is not an accurate description of the fluid considered; at best it describes some aspect of its behaviour in some types of flow fields. For instance, (4.8.4) is hoped to be appropriate if the fact that the viscosity is variable in a shear flow is supposedly the only thing of importance. This means, among other things, that one needs to have an idea of what the kinematics of the flow field of interest might be before even choosing the appropriate constitutive equation, and yet the kinematics of the flow field are just what one is trying to calculate by solving the flow problem.

Another problem is as follows. Suppose one considers the possibility of solving flow problems for no matter how complex a constitutive equation might be. Perhaps advances in computer hardware could make such an assumption reasonably realistic. One then could embark in the following program: establish by experiment what the appropriate form of the constitutive equation might be. The program would go something like sketched in the logical diagram in Fig. 4.8.1. One first writes down the simplest possible

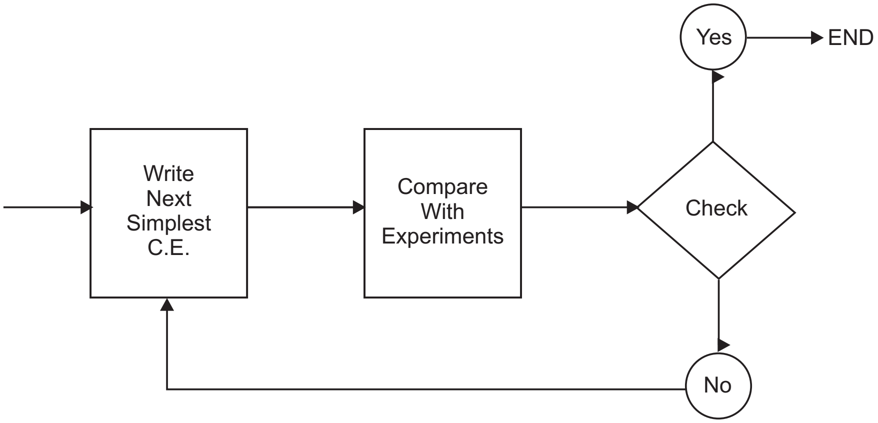

Fig. 4.8.1. Logical flowsheet of program for developing constitutive equations.

constitutive equation, say the Newtonian one. One then compares with experiments, and say experiments don't confirm our equation — viscosity is not constant, for instance. We now write down the next more complex constitutive equation (we'll call this the NMC step), and try again. For instance, we have written (4.8.1) to accommodate a variable viscosity. We now go to the experiments again, as in the example of the orthogonal force, and we are again unlucky. So we go back to the NMC step, and try again. This is the program, and we hope eventually to come out of the loop in the "check? Yes" step. There are three problems with this program. First, there is no guarantee that we'll ever come out of the loop. Second, the NMC step is hard to perform: linearity is of course the simplest assumption, but what about the subsequent choices? And, finally, even if we do the NMC steps cleverly, and eventually come out of the loop, the resulting constitutive equation may well be so horrendously complex that our initial hope to use it in the solution of flow problems turns out to be overly optimistic. That is why one decides to come out of the loop at some intermediate stage, where the fact that the "check? yes" step has been only moderately to our liking is balanced by the relative ease of use of the resulting constitutive equation.

4.8.1 Extruder flow

Consider flow between two parallel plates, see Fig. 4.8.1.1. The upper plate moves steadily rightwards with velocity V, and the lower plate is fixed. A positive pressure gradient G is imposed in the x direction, and this by itself would cause the flow to be in the negative x direction. However, superimposed on this pressure driven flow there is the drag driven flow due to the motion of the upper plate, and this is in the opposite direction. We wish to calculate the net flow rate, which might be in either direction. This idealized flow pattern is not very different from what happens in an extruder, and therefore we consider positive the flow-rate if it is rightwards. The flow is laminar, and hence the creeping flow equations apply.

If the fluid is Newtonian, the problem is very simple. In the absence of a pressure gradient ($G = 0$), the drag driven flow-rate is $VH/2$. Should the upper plate be fixed ($V = 0$), the pressure driven flow-rate would be $-GH^3/12\mu$. We simply need to superimpose these two to obtain the net flow-rate as $VH/2-GH^3/12\mu$. The situation is considerably more complex for a non-Newtonian fluid, even if we choose a power law equation.

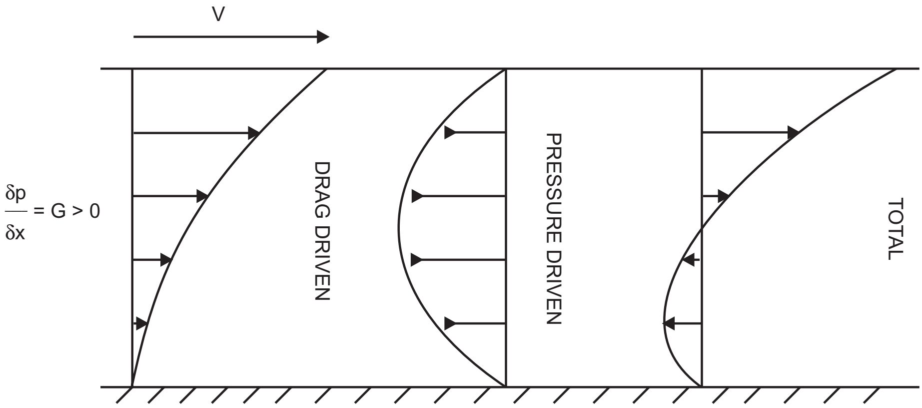

Fig. 4.8.1.1. Flow between parallel plates: superposition of drag driven and pressure driven flows.

First consider the two subproblems separately. The drag driven flow-rate is still $VH/2$. The pressure driven flow is slightly more complex. The shear stress τ is now variable along y, and a simple force balance yields:

$$G = \mathrm{d}\tau/\mathrm{d}y \tag{4.8.1.1}$$

We now have a problem. The constitutive equation can be written as $\tau = K|\Gamma|^{n-1}\Gamma$, and as long as Γ is nonnegative this can be simplified to $\tau = K\Gamma^n$. However, one has to be careful when Γ changes sign, since τ is required to always have the same sign as Γ. For the case at hand, however, there is a shortcut. The problem clearly has symmetry, and we can limit attention to the lower halfwidth, $0 \leq y \leq H$. By symmetry, the shear stress is zero at the midplane, and it is therefore $-GH/2$ at $y = 0$; it follows that $\tau = G(y - H/2)$. This is always nonpositive in the lower half, and hence the constitutive equation can be written as (with $s = 1/n$):

$$\mathrm{d}u/\mathrm{d}y = -[G(H/2 - y)/K]^S \tag{4.8.1.2}$$

This can now be integrated with the boundary condition $u = 0$ at $y = 0$, and the flow-rate can be calculated in the lower half; the total flow-rate is, by symmetry, twice that in the lower half, and one obtains:

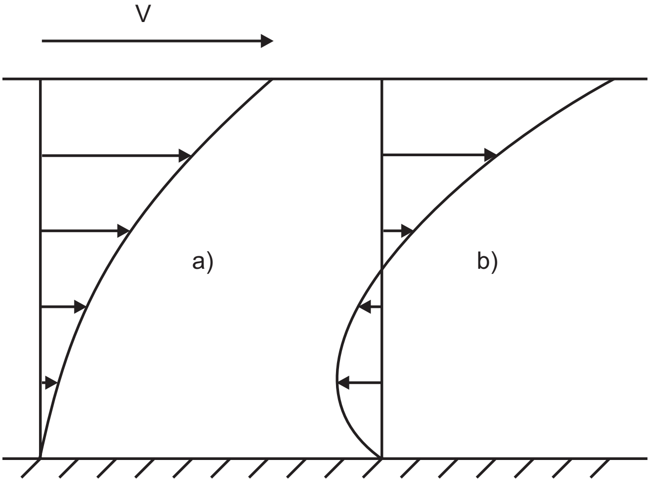

Fig. 4.8.1.2. Possible velocity profiles in extruder flow: (a) velocity is positive everywhere; (b) velocity is negative over part of the flow field.

$$Q_P = -(G/K)^S H^{S+2}/(s+2)2^{S+1} \tag{4.8.1.3}$$

Equation (4.8.1.3) of course reduces to the Newtonian result when $n = s = 1$.

The difficulty that arises is that one cannot simply add Q_P to $VH/2$ to obtain the result for the combined case: the equations are nonlinear, and therefore superposition does not work. Physically, the question is as follows. The backwards pressure driven flow induces a shear rate Γ which depends on y; consequently, the apparent viscosity depends on y. This has an effect on the drag driven flow, because while shear stress due to drag is constant, a nonconstant viscosity results in a nonconstant shear rate. One therefore needs to solve the coupled problem anew.

In trying to do this, one meets another difficulty: there is no symmetry argument to tell us whether the shear stress is positive or negative. Indeed, consider Fig. 4.8.1.2. The velocity profile may be as in (a), in which case the shear stress is everywhere positive; or it might be as in (b), in which case there is some position $y = h$ where the shear stress changes sign. The problem considered is still simple enough to allow us to obtain an analytic solution, which will not be discussed here; but it shows how even extremely simple flow problems may become quite complex when the fluid is non-Newtonian.

4.9 ELEMENTS OF VISCOELASTICITY

There are a number of phenomena observed with non-Newtonian fluids which are manifestations of viscoelasticity, i.e., of a behaviour somewhat intermediate between that of fluids and that of elastic solids. The most important ones are discussed below.

Response to oscillations. The phenomenon has been discussed in its simplest geometrical form in Section 1.10; the thermodynamic implications have been discussed in Section 2.3. At high frequencies, the behaviour of elastic solids is approached, which shows in what sense one may think of the fluid possessing a degree of elasticity. Since the stress is out of phase with the rate of deformation, this behaviour can never be described by the constitutive equation of a purely viscous fluid. When actual experiments of forced oscillation are performed, the amplitude of the imposed strain is kept small, and when that is the case the amplitude of the stress oscillation is proportional to the amplitude of the strain. Such a situation is referred to as linear viscoelasticity.

Finite first normal stress difference. This has been discussed in the preceding section, and it has been shown that it cannot be predicted by a purely viscous constitutive equation. The effect is reminiscent of elasticity, because elastic solids do develop a first normal stress difference when shear deformation is imposed.

Stress relaxation. If a non-Newtonian fluid is suddenly brought to rest after some deformation, the stress tensor does not become an isotropic pressure instantaneously, but it decays to an isotropic pressure over some time interval. The suggestion of elasticity here lies in the fact that in a strained solid at rest the stress does not decay at all. Again a purely viscous constitutive equation cannot describe stress relaxation, since $\mathbf{D} = 0$ during stress relaxation and yet $\mathbf{T}$ is changing in time.

Response to elongational flow. Consider two flow fields described by the following kinematics:

$$u = \Gamma y, v = w = 0 \tag{4.9.1}$$

$$u = \Gamma y/2, v = \Gamma x/2, w = 0 \tag{4.9.2}$$

The rate of deformation tensor $\mathbf{D}$ is the same for these two kinematics, and hence any purely viscous fluid will exhibit the same stresses. This is not the case by far with many non-Newtonian fluids, which exhibit much larger stresses in the second type of kinematics. Notice that the second flow is elongational, while the first one is a pure shear flow. (This is a subtle point, and will be discussed in detail later).

When viscoelasticity is supposed to be relevant for the problem at hand, one needs a constitutive equation capable of describing the phenomena considered above. Now notice that in all four cases one may identify an intrinsic time scale θ. In oscillatory experiments, the phase angle depends on frequency, and hence a characteristic frequency $1/\theta$ can be identified. The first normal stress difference is, at low enough shear rates Γ, proportional to Γ^2, while the tangential stress is proportional to Γ. It follows that the ratio of normal to tangential stress is proportional to Γ, and the proportionality constant is an intrinsic time scale. The phenomenon of stress relaxation obviously introduces a time scale, which is a yardstick of the time required for stresses to decay to some small fraction of their initial value. The response to elongational flow is the hardest one to describe without heavy mathematics, but since the difference between (4.9.1) and (4.9.2) is that the first flow is rotational while the second one is not, some measure of the rate of rotation is of importance, and hence again one identifies an intrinsic time scale.

The time scale θ is best regarded as the length of the memory span of the material (see Section 2.7). The question to be addressed is the formulation of a constitutive equation containing such a time scale, in addition to a viscosity. A simple one-dimensional formulation was given in Section 1.10, and this is reported below:

$$T_{XY} + \theta \partial T_{XY}/\partial t = -\mu \partial u/\partial x \qquad (4.9.3)$$

If one tries to write a properly invariant three-dimensional form of (4.9.3), one meets with several problems. The RHS is easily generalized to $-2\mu\mathbf{D}$, and the stress itself with $\mathbf{T}$; the problems arise with the term containing θ. The difficulty lies in the fact that even the substantial time derivative of stress is not invariant with respect to a non-Galileian change of frame. It is possible to define time derivatives which are properly invariant, and in fact one may define an infinite number of such derivatives. We only discuss one of them, which is the most widely used. Let $\mathbf{L}$ be the velocity gradient; the "upper convected" derivative of stress (or of any other tensor) is defined as follows:

$$\delta\mathbf{T}/\delta t = \mathrm{D}\mathbf{T}/\mathrm{D}t - [\mathbf{T}\cdot \mathbf{L}_{\mathrm{T}} + \mathbf{L}\cdot\mathbf{T}] \qquad (4.9.4)$$

Correspondingly, one can generalize (4.9.3) by writing the so-called Maxwell equation (with **T** the stress in excess of an arbitrary isotropic pressure):

$$\mathbf{T} + \theta \delta \mathbf{T}/\delta t = -2\mu \mathbf{D} \tag{4.9.5}$$

Now let's consider how well the Maxwell equation does with the phenomena discussed above. It does degenerate to (4.9.3) for the oscillating problem considered in Section 1.10, and hence it describes the essential features of the response to oscillations. For pure shear flow (described by Eq. (4.9.2)), it produces (after some algebra) a shear stress $\mu\Gamma$, and a first normal stress difference $\mu\theta\Gamma^2$, which is not too bad: the variable viscosity is not obtained, but a finite first normal stress difference is obtained. For stress relaxation, it predicts that the extra stress decays to zero like $\exp(-t/\theta)$, which again is not too bad. Finally, for the elongational flow described by Eq. (4.9.3) it predicts stresses which grow more than linearly with Γ, and in fact become infinitely large at a stretch rate $2\theta\Gamma = 1$; hence the Maxwell equation is capable of distinguishing between the two flows in (4.9.1) and (4.9.2), and it does predict that (4.9.2) will result in larger stresses. Thus one concludes that the Maxwell equation does a reasonable job in describing viscoelasticity, though not a perfect one. A variable viscosity could be accommodated by allowing either μ or both μ and θ to depend on the invariants of **D**, but this is rather arbitrary.

We now come to the most subtle, and perhaps most significant problem in viscoelasticity. Suppose one has a problem like flow of a polymer melt in an extruder, where one convinces oneself that the flow is very nearly a shear flow, and normal stress differences don't play any significant role. For such a problem, the viscoelasticity of the fluid is not of importance, and a purely viscous constitutive equation is probably adequate. On the other side, one may have a problem like fibre spinning, where the flow field is very nearly an elongational one, and this time we want to use a viscoelastic constitutive equation, since any viscous theory would yield gross underestimates of the stresses. The problem which arises is deceivingly simple: how do we distinguish flow fields as to their being almost pure shear, almost pure elongation, or something in between? Or maybe something more extreme than either one? One feels that there is a spectrum between pure shear and pure elongation, and that such a spectrum may in fact not be bounded by those two types of kinematics; and one would wish to quantify the position of any given kinematics in such a spectrum.

A very simple two-dimensional formulation may help in visualizing the problem; this is an extension of the example given in Eqs. (4.9.1–2). Let the flow field be described by:

$$u = \alpha\Gamma y;\, v = (1-\alpha)\Gamma x;\, w = 0 \tag{4.9.6}$$

The examples in (4.9.1) and (4.9.2) are immediately recognized as being the cases where $\alpha = 1$ (shear flow) and $\alpha = 0.5$ (elongational flow. So this is a simple problem where, as α spans the range 0.5 to 1, one moves from purely elongational flow to pure shear flow. Now values of α less than 0.5 do not need to be considered, since by switching the x and y axes one could always make α not smaller than 0.5. However, values of α larger than unity are clearly possible, and indeed as α approaches infinity one would be describing a rigid body rotation with constant rotation rate $\alpha\Gamma$. The rate of strain tensor **D** is given by:

$$\mathbf{D} = \begin{bmatrix} 0 & \Gamma/2 \\ \Gamma/2 & 0 \end{bmatrix} \tag{4.9.7}$$

and is thus independent of α: any purely viscous theory will predict the same stresses no matter what the value of α.

Now let's first understand well the difference between a shear and an elongational flow. Consider a square material element located at time zero with a corner at the coordinate origin (see Fig. 4.9.1). At some later time, in shear flow its base line will still lie on the x axis, and in fact no material point has moved in the y direction, since $v = 0$. The square has been deformed, of course, but it has also *rotated*: the material line which was at 45 degrees to the x axis at time zero forms a smaller angle with it at any later time. Since the kinematics are uniform in space, the direction of maximum stretching is everywhere at 45 degrees, and so the material line which was most severely stretched at time zero is not the most severely stretched one at later times. The rotation has the effect of relieving the stress, because it causes different material lines to be subjected to the highest stretch rate at different times.

Conversely, in the case of elongational flow the square will have deformed, but it has not rotated: the line which was at 45 degrees at time zero is still at 45 degrees at any later time. Indeed, if one rotates the axes by 45 degrees, the description of the kinematics in the new (primed) coordinate system becomes:

$$u' = \Gamma x'/2;\, v' = -\Gamma y'/2;\, w' = 0 \tag{4.9.8}$$

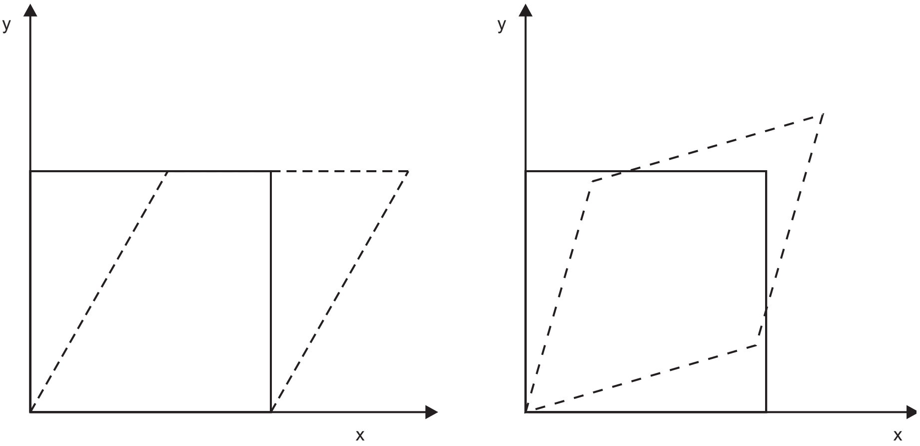

Fig. 4.9.1. Deformation of material elements subjected to the flow fields (4.9.1) and (4.9.2).

and this is immediately recognized as an elongational flow, the material being continuously stretched in the *x*′ direction, and compressed in the *y*′ direction.

Consideration of the streamlines is also of interest. In shear flow ($\alpha = 1$), all the streamlines are straight lines, at 45 degrees to the direction of highest stretch. In elongational flow ($\alpha = 0.5$), the streamlines are hyperbolas coming from the direction of highest squeezing (–45 degrees to the *x* axis) and approaching the direction of highest stretching, see Fig. (4.9.2). Now what about the case $\alpha > 1$? This is given as a homework problem due next time, and as usual Sue has made a horrible mess of it. Ralph has got it right — the trajectories are ellipses, which tend to become circles as α approaches infinity. In fact if one considers, for the case $\alpha > 1$, a material element having the shape of an ellipse similar to the streamlines, it gets squeezed until the major axes are reversed, and it then gets squished back to its initial shape; for very large values of α, the total deformation is a small one, and hence this is a flow within the realm of linear viscoelasticity — small periodic oscillation.

The other homework was the calculation of the stresses from the Maxwell constitutive equation, and after some debate the Ralph result is accepted as correct:

$$T_{XY} = \mu\Gamma/(1-b) \tag{4.9.9}$$

$$T_{XX} - T_{YY} = (4\alpha-2)\theta\mu\tau^2/(1-b) \tag{4.9.10}$$

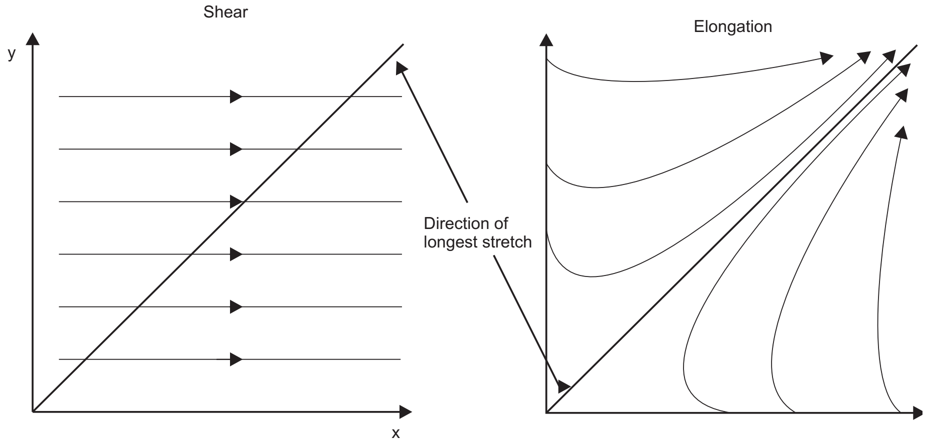

Fig. 4.9.2. Streamlines for elongational and shear flow.

$$b = 2\theta^2\tau^2\alpha(1-\alpha) \tag{4.9.11}$$

Now this is rather interesting. In shear flow, $b = 0$ and the tangential stress is proportional to Γ — no shear thinning, as we knew was the case for the Maxwell constitutive equation. In elongational flow, T_{XY} approaches infinity as $\theta\tau$ approaches $\sqrt{2}$. And the interesting thing is that, when $\alpha > 1$, T_{XY} is less than proportional to Γ: a Maxwell fluid exhibits shear thinning in elliptical flows. In the limit of α approaching infinity (rigid body motions) all stresses become zero, as one would expect to be the case.

So we are now in a position to understand the physical basis for the different response of viscoelastic fluids as α changes. A viscoelastic fluid has memory — it exhibits relaxation phenomena, which means it remembers what deformations it has been subjected to in the past. And the material resents being stretched always along the same material line. It seems to be saying "give me a break", and in elliptical flows it gets quite a break — a little bit of stretching, and then some compression to compensate for it. The break the material wishes is, in some sense to be elucidated, some rotation. So how are we going to describe this physical reality in a mathematical sense?

Bob stumbles here. He suggests that we should look at the rate of rotation, say the vorticity tensor **W**. He has quickly calculated it:

$$\mathbf{W} = \begin{bmatrix} 0 & (2\alpha - 1)\Gamma / 2 \\ -(2\alpha - 1)\Gamma / 2 & 0 \end{bmatrix} \tag{4.9.12}$$

and this now does depend on α — so what is all the fuss about?

The fuss is about the little problem that vorticity cannot enter the constitutive equation for stress, which has to be frame invariant. Bob does not believe this: who says it has to be frame invariant? The momentum balance is not invariant under superimposed rotations, so why should the constitutive equation be invariant? Well, let's make a thought experiment. We have the two flows described by (4.9.1) and by (4.9.2) set up in our lab, complete with gauges measuring the stresses. We now decide to place the graduate student who does the experiment on a platform rotating around the z axis with angular speed Ω. We have a second graduate student who is used as a blank and just sits in the lab, so he describes the kinematics by (4.9.1) and (4.9.2). But the student in the merry-go-round sees a superimposed rigid body rotation, and so his kinematic description is:

$$u = (\Gamma + \Omega)y; v = -\Omega x; w = 0 \tag{4.9.1$'$}$$

$$u = (\Gamma/2 + \Omega)y; v = (\Gamma/2 - \Omega)x; w = 0 \tag{4.9.2$'$}$$

We have in fact been nasty, and we have set Ω at the value Γ/2, and so our student on the merry-go-round concludes that the shear flow is irrotational, and the elongational flow is rotational, which leads to a terrible argument with the other student. They, however, agree on what the stresses are — the transducers have transmitted the same signals to both. Since they can't agree on what rotates and what does not, they conclude that the stresses must be independent of vorticity.

Now Ralph has a difficulty — actually his wording suggests that he thinks we are talking nonsense. The student on the merry-go-round will experience centrifugal forces, and hence he knows he is on a merry-go-round — he should subtract his own rotation from the kinematic description, and then the two students would agree. Well, there are some difficulties here, because the student sitting in the lab experiences gravity, and how can he tell it is not artificial gravity obtained by rotating the whole lab? It is true that the momentum balance is not invariant under a non-Galileian change of frame (such as going from the lab to the merry-go-round), but the constitutive equation needs to be invariant under such a change. The lab is (presumably) an inertial frame, and if we had moved the equipment to the rotating platform centrifugal forces would perhaps have changed the kinematics of motion. But that does *not* mean that the centrifugal forces change the *relationship* between the kinematics of motion and the stresses. Perhaps another thought experiment where one rotates the equipment rather than the student is in order.

Suppose we have a spring which has been measured to elongate by one cm when loaded with a weight of 1 kg. Now suppose we mount the string horizontally on the rotating platform, with one end hooked at the axle, and the free end hooked to a mass of 1 kg. We start rotating the platform, and at some angular speed we observe it has elongated by 1 cm under the influence of the centrifugal pull of the 1 kg mass. Are we willing to calculate the angular speed from the assumption that the centrifugal acceleration is exactly equal to gravity? Of course yes; but that means that we think the spring's elongation under a certain load has nothing to do with whether it is rotating or not.

The point is, somewhat reluctantly, granted. (Incidentally, the nasty term in brackets in (4.9.4) comes out from exactly this requirement of invariance). So Bob's original suggestion has to be abandoned, and does anybody have a bright idea?

Sue raises to the occasion. "Rotation is somehow involved - but not rotation with respect to our frame of observation. It's the rotation with respect to the direction of highest stretching that matters, and I bet that rotation is invariant". You would win the bet, Sue, though there is some merit in convincing oneself through careful algebra rather than by rather wild guesses. The stretching tensor **D** at a material point has three principal axes, and these may rotate (with respect to whatever frame we choose) at some rate Ω' which is not frame invariant. With respect to the same frame, the material element rotates at some other rate Ω which is also not frame-invariant; but the difference $\Omega^* = \Omega - \Omega'$ is frame invariant. The proof for the general three-dimensional case involves some subtleties, and we won't discuss those. In the case of (4.9.6), the flow has been described in a frame with respect to which $\Omega' = 0$, and hence the actual rotation is the significant one.

In general, we may define a *relative vorticity* $\mathbf{W}^*$, and then define the flow parameter R as follows:

$$R = |\mathbf{W}^*|/|\mathbf{D}| \tag{4.9.13}$$

For the flows described by (4.9.6), a little algebra leads to $R = 2(2\alpha -)$. So the flow parameter is zero in elongational flow (no stress relieving rotation at all), it is unity in shear flow (stress relieving rotation as effective as the stress inducing deformation rate), and is larger than unity in elliptical flows, approaching infinity as the flow approaches a rigid body rotation. One can now play the game of devising constitutive equations which change character as the value of R changes.

4.10 ELEMENTS OF GAS DYNAMICS

In this section, we discuss some elementary concepts which form the basis of classical Gas Dynamics, say of compressible flow. Compressibility is a difficult property to deal with, and one therefore begins the analysis by leaving out other "complications", such as viscosity and thermal conductivity (more about this later). We further restrict ourselves to one-dimensional flows, as was done in Section 1.5, so that there is only one component of velocity to worry about, u. The variables of interest are u, pressure p, temperature T, cross section of the duct considered A, density Φ, internal energy per unit mass E, enthalpy per unit mass H, and entropy per unit mass S; all these are regarded as possibly depending on position x (though not on time, since we also restrict attention to phenomena which, in some appropriate frame of reference, are steady ones). Furthermore, we assume that the gas is an ideal gas endowed with a specific heat per unit mass at constant volume c_v which is independent of temperature; the constant pressure specific heat is of course $c_P = c_v + R$, where R is the mass based gas constant, i.e., the actual gas constant R divided by the molecular weight of the gas. Ralph thinks these restrictions are going to make the subject too easy, but that is far from being true.

Everybody agrees what the balance of mass looks like:

$$\Phi u A = \text{constant} \tag{4.10.1}$$

As for the momentum balance, it takes some more time to reach agreement on the following (viscosity has been neglected):

$$\mathrm{d}(\Phi u^2 A)/\mathrm{d}x = -\mathrm{d}p/\mathrm{d}x \tag{4.10.2}$$

One can now make use of Eq. (4.10.1) to simplify this to:

$$\Phi u \mathrm{d}u/\mathrm{d}x = -\mathrm{d}p/\mathrm{d}x \tag{4.10.3}$$

Now this could have been obtained directly from the Euler equations of motion, and since we are neglecting viscosity we expect of course the latter to hold. So far so good, but now we enter the muddy waters of thermodynamics in writing down the balance of energy. There is no shaft work done on the system (the bounding walls are stationary), and so there is only flow work to worry about. The heating rate is taken to be zero, and so the rate of shaft work done on the system equals the rate of accumulation of internal and kinetic energy. After the dust has settled on this agreement is reached that, if gravity is neglected (after all, the gas has a small density, and the duct may well be horizontal), and if the heat flux is taken to be zero, one has:

$$\mathrm{d}(E + u^2/2 + p/\Phi)/\mathrm{d}x = \mathrm{d}(H + u^2/2)/\mathrm{d}x = 0 \tag{4.10.4}$$

Now we make use of our assumption that $H = c_\mathrm{P}T$ + const. to obtain the following equation, where T^0 is the "stagnation temperature", i.e., the temperature which is attained when the fluid is brought to rest:

$$c_\mathrm{P}T + u^2/2 = c_\mathrm{P}T^0 \tag{4.10.5}$$

So far everything has flowed out smoothly from our constitutive assumptions and balance equations. Now we are going to perform a rather devious trick: we *define* the quantity c as the square root of $RT\tau$, where τ is the ratio of the two specific heats. Of course we know that c is the speed of sound, Ralph; or at least we believe it to be the speed of sound if we believe that sound propagates isoentropically; but entropy is a strange animal, and we don't want to have it cluttering our results for as long as possible. So we are going to simply regard c as a distorted measure of temperature itself. We also define the Mach number M as u/c, and this is now seen to depend on both velocity and temperature. We substitute the definitions into (4.10.5), do a bit of rearrangement, and obtain:

$$T = T^0\left[1 + \frac{\tau - 1}{2}M^2\right]^{-1} \tag{4.10.6}$$

Bob has a protest to lodge here: he says that we are simply juggling around the definitions and getting nowhere; we haven't even got as far as we did in Section 1.5. Be a little patient with us, will you, Bob? Ralph has a different protest: he knows that we are going to end up in converging-diverging nozzles, and he does not know why we refrain from using entropy, since he knows we'll never get anywhere without assuming isoentropic flow. Well, let's see how long we can keep up the game without using entropy, shall we? First of all, we make (temporarily) one more assumption, namely, that the duct has a constant cross section, $\mathrm{d}A/\mathrm{d}x = 0$. That's certainly quite different from converging diverging nozzles, but it is certainly appropriate to the consideration of shocks, since over the thickness of the shock layer the cross section is unlikely to change by any significant amount. Let ′ and ″ identify cross sections located at opposite sides of the shock (or, for that matter, two sections in any constant section conduit). Equation (4.10.6) now tells us that:

$$\frac{T'}{T''} = \frac{1 + (\tau - 1)M''^2 / 2}{1 + (\tau - 1)M'^2 / 2} \tag{4.10.7}$$

The ideal gas law gives:

$$p'/p'' = (M''/M')\,(T'/T'')^{1/2} \tag{4.10.8}$$

Notice that so far we have *not* used the constant cross section assumption. We do so in using the mass balance, to obtain:

$$\Phi'/\Phi'' = u''/u' = (M''/M')\,(T''/T')^{1/2} \tag{4.10.9}$$

So now we have expressed all the ratios in terms of the two Mach numbers, and we immediately recognize that the case where $M'' = M'$ is trivial: everything is the same at the two cross sections. Even pressure — does that make sense? If gas is flowing isothermally through a constant section tube, wouldn't we expect an axial pressure gradient to develop? Not if we neglect viscosity, and we have done so. The Euler equations apply, and there is no pressure drop in the motion through a constant section tube, as far as the governing equations are concerned. We are dealing with compressibility, and we are letting go of everything else.

If the two Mach numbers are different, one sees immediately that the section where the Mach number is smaller has a larger temperature, a larger pressure, and (with some algebra with Eqs. (4.10.7) and (4.10.9)) a larger density. Which way the flow is going is left unspecified so far —we have not used the second law, and everything is perfectly reversible. Going back to what was discussed in Section 1.5, we see that if the flow is towards the low Mach number section, we have a detonation, and that is known to propagate supersonically, so the gas approaching the shock must be moving supersonically. In the converse case where the gas is moving towards the high Mach number section, we recognize a deflagration which propagates subsonically.

Ralph just cannot stand all this any more. He proposes that we go on to calculate S'/S'', and he maintains that we will be sorry for what we will conclude from that — he remembers that the Mach number at the exit of the nozzle can only be *smaller* than the one upstream. Well, let's do that. From the appropriate Maxwell equation we get, after substituting our constitutive assumptions:

$$S'' - S' = c_P \ln(T''/T') - R\ln(p''/p') \tag{4.10.10}$$

We now indulge in a bit of algebra so as to express everything in terms of the two Mach numbers:

$$S'' - S'' = (C_P - R/2)\ln\frac{1+(\tau-1)M'^2/2}{1+(\tau-1)M''^2/2} - R\ln\frac{M'}{M''} \tag{4.10.11}$$

So now what do you conclude from this equation, Ralph? "Simple. Let the flow be towards the double prime section. Then S'' has got to be larger than S', and hence M' has got to be smaller than M''. The shock is stationary in the diverging portion of the nozzle with the upstream flow being supersonic, and hence it is a detonation". Good, Ralph, as expected you remember correctly your undergraduate stuff. However, why should S'' be larger than S'? "The second law, of course". Well, may be. Does the second law imply that flow is always towards a higher entropy condition? Water flowing into a very cool tube may freeze, and ice has less entropy than liquid water, so perhaps things are not quite as simple as that. And, Ralph, are you willing to go all the way and state that you believe deflagrations to be impossible? The problem we are facing is, of course, that the second law is a tricky business, a statement with which Sue entirely agrees.

Having reached the stage where entropy has made everybody very confused indeed, we proceed to reason as follows. There is no viscosity, and hence no viscous dissipation. If there is no heat flux either (and we have already made this assumption in writing the energy balance), we are left with no cause for dissipation, and hence everything is reversible, perhaps — though Ralph has opened Pandora's box by now and we have some difficulty, since we have just concluded that, at least under some qualifying conditions, the flow direction is not quite so reversible. But anyhow, let's see where the assumption of reversibility brings us. Since we have also assumed the heating rate to be zero, the process is adiabatic, and, well, if it's adiabatic and reversible it must be isentropic. Of course that can't quite be true, since at least for the constant section conduit we would lose everything of even mild interest, shocks would be impossible, and life would just be made miserable (or beautiful, according to Sue's tastes). But let's see what happens if we assume that the flow is indeed isentropic, and let's go back to a varying section conduit, since we want to have something to play with. So we are finally into converging–diverging nozzles, as Ralph had known we would be at some time.

Recalling the pressure-temperature relationship for isoentropic transformations of an ideal gas, Eq. (4.10.6) can be transformed to an equation relating the actual pressure and the stagnation pressure p_o:

$$p = p_o[1 + (\tau - 1)M^2/2]^{\tau/(1-\tau)} \tag{4.10.12}$$

This is not very illuminating, but let's try to use it for a very simple case. Consider the case where M is very small; to within order zero in M we obtain:

$$p_o = p + \Phi u^2/2 \tag{4.10.13}$$

This is the equation we believe is right when we use Pitot tubes to measure velocities, isn't it? So the result is rather satisfying. If one keeps the lowest order terms in M, one obtains an equation which, at M values less than 0.5, agrees to within one percent with (4.10.13). We thus conclude that, perhaps, one does not need to worry about compressibility unless the Mach number is larger than 0.5.

Now we do something else: we calculate the speed of propagation c of a wave (the wave has a constant speed of propagation, and a constant shape; it needs not be infinitesimal). How the wave has been generated does not concern us. We also assume the cross section area to be constant. Now we only need to use the mass and the momentum balance to obtain the following result:

$$(u - c)^2(\mathrm{d}\Phi/\mathrm{d}x) = \mathrm{d}p/\mathrm{d}x \tag{4.10.14}$$

If $u << c$ (or if $u = 0$, i.e., the wave is propagating through a stationary medium), one obtains $c^2 = \mathrm{d}p/\mathrm{d}\Phi$, and if the flow is isoentropic one of course identifies c with the usually accepted speed of sound. So far so good. We now come to a rather interesting point. The energy balance (4.10.4) can be written as:

$$\mathrm{d}H/\mathrm{d}x + u\mathrm{d}u/\mathrm{d}x = 0 \qquad (4.10.15)$$

If we substitute the Maxwell relationship $\mathrm{d}H = T\mathrm{d}S + \mathrm{d}p/\Phi$, which for isoentropic flows reduces to $\mathrm{d}H = \mathrm{d}p/\Phi$, we obtain:

$$\Phi u\mathrm{d}u/\mathrm{d}x = -\mathrm{d}p/\mathrm{d}x \qquad (4.10.16)$$

This is somewhat surprising: we have only used the energy balance and the assumption of isoentropicity, and, lo and behold, we have obtained the momentum balance. What's going on here? There certainly is some room for thought even without nozzles. The point is that, by *assuming* isentropicity, we have really negated any possibility of energy dissipation, and so mechanical energy is conserved. For a one-dimensional flow, this is as much information as the momentum balance can conceivably yield. *For isentropic one-dimensional flow the energy and momentum balance contain the same information*.

There is still more to come before we actually deal with nozzles. Differentiation of the mass balance and division by ΦuA gives:

$$\frac{1}{\Phi}\frac{\mathrm{d}\Phi}{\mathrm{d}x} + \frac{1}{u}\frac{\mathrm{d}u}{\mathrm{d}x} + \frac{1}{A}\frac{\mathrm{d}A}{\mathrm{d}x} = 0 \qquad (4.10.17)$$

Since the flow is isentropic, $\mathrm{d}p/\mathrm{d}x = c^2\mathrm{d}\Phi/\mathrm{d}x$, and if we substitute this and the energy balance into (4.10.17), and we eliminate $\mathrm{d}x$, we obtain:

$$\mathrm{d}u/\mathrm{d}A = [A(M^2 - 1)/u]^{-1} \qquad (4.10.18)$$

As long as the flow is subsonic ($M < 1$), this tells us that velocity decreases in a diverging duct, and it increases in a converging one, as we are accustomed to expect. But in supersonic flow ($M > 1$), the opposite is true, much to our surprise (and the mind refuses to even consider what happens at exactly the speed of sound).

There are other interesting points to be considered. Let W be the mass flow-rate, and $W/A = \Phi u$ the mass flux (which is constant in a steady flow). By substituting the definitions and Eqs. (4.10.6) and (4.10.12) one obtains:

$$\frac{W}{A} = p^{0} M \left[\frac{\tau}{RT^{0}} \right]^{1/2} \left[1 + \frac{(\tau - 1)M^{2}}{2} \right]^{(1+\tau/2(\tau-1)} \tag{4.10.19}$$

Now Eq. (4.10.12) tells us that, as p/p_{o} approaches zero, the Mach number approaches ∞ — that seems to make at least superficial sense: if we keep decreasing the back pressure in the downstream reservoir into which a converging nozzle discharges, the gas velocity at the exit increases beyond limits. But Eq. (4.10.19) tells us that, as the Mach number approaches ∞, W/A approaches zero — that *does not* seem to make much sense, does it? If it were true, space shuttles could have as many holes as one wishes, since no gas would escape into the vacuum of space.

What happens in reality is that, as long as Eq. (4.10.12) yields a Mach number not in excess of unity (no supersonic flow), the observed mass flow-rate is the one given by (4.10.19). For an ideal gas with $\tau = 5/3$, the value of p/p_{o} corresponding to $M = 1$ is 0.487. If the back pressure is lowered below that value, the actual Mach number at the exit stays at unity, the mass flow-rate stays at the value given by (4.10.19) for $M = 1$, and a shock develops at the exit of the nozzle; this is known as choke flow. The important point is that flow through the shock is not isoentropic — and that a supersonic flow condition *has not* been achieved in a convergent nozzle.

Now it is possible to ask oneself whether one *can* achieve supersonic flow, and here we finally come to convergent–divergent nozzles. As the flow becomes sonic at the nozzle throat, the subsequent part of the duct *must* be divergent because in supersonic flow W/A decreases as the Mach number increases. It is therefore possible to design a converging-diverging nozzle (i.e., to choose $A(x)$ appropriately) which will give supersonic flow at the exit and such that Eqs. (4.10.12) and (4.10.16) are satisfied everywhere along the nozzle.

However, suppose we have an assigned converging–diverging nozzle, and we consider the situation where the back pressure is continuously lowered. At sufficiently high back pressures, the flow is everywhere subsonic, and the nozzle acts as a Venturi meter, with the lowest pressure (and the highest Mach number) at the throat. However, when the Mach number at the throat is unity, the $p(x)$ curve is double valued downstream of the throat, and there is a range of back pressures which cannot be achieved by the equations. In practice, if a back pressure in the “forbidden” region is imposed, a shock will develop somewhere in the diverging region, through which the flow goes from supersonic to subsonic. Again, flow through the shock is not isoentropic.

PROBLEMS

4.1 In Section 1.5, we discussed how shocks may be both detonations and deflagrations. In Section 4.10, our second law analysis seems to allow only detonations, which correctly propagate supersonically into the gas. There seems to be a paradox here, since we might ask ourselves whether deflagrations are or are not possible. Explain the solution of this apparent paradox.

4.2 Consider a large spherical gas bubble formed by cavitation; its initial radius is R and the pressure inside the bubble is essentially zero. Calculate the time it takes for the bubble to collapse.

4.3 Consider a Maxwell fluid between two parallel plates. The lower one is fixed, and the upper one is forced to undergo oscillations so that a mark on it is at a position z given by $Z\sin(\Omega t)$. Neglect inertia completely, and express the tangential stress as $T\sin(\Omega t + \varnothing)$. Find the relationship between T and $\varnothing$ on the one side, and viscosity and relaxation time on the other one.

4.4 Consider a "start up of shear" experiment, where the fluid is standing still at times <0, and is flowing in pure shear flow ($u = \Gamma y, v = w = 0$) at $t > 0$. Let $T(t)$ and $s(t)$ be the tangential stress and the first normal stress difference, and let $T(\infty)$ and $s(\infty)$ be their steady state values. The "stress overshoot" phenomenon is the one where either $T(t)$ or $s(t)$ or both attain, at some time, values larger than their steady state values. Is such a phenomenon describable by the Maxwell fluid constitutive equation?

4.5 A forced and a driven shaft are connected by a cylindrical joint: the forcing shaft part of the joint is a cylinder of length 20 cm and radius 10 cm; the driven shaft part is a hollow cylinder of length 20 cm and radius 10.05 cm. The forcing shaft turns at 6,000 RPM, and transmits to the joint a power of 10 HP. The fluid in the joint has a viscosity of 10 poise. Calculate the RPM of the driven shaft and the power dissipated in the joint.

4.6 Suppose you need to determine the viscosity of a liquid, which you know beforehand to be of the order of 100 poise. You need to design both a concentric cylinder and a capillary viscometer for the job. Decide on the appropriate geometrical dimensions for both.

4.7 Compute the steady rising velocity of a ping pong ball through water, and the initial acceleration of the ball when starting from a state of rest.

4.8 A fireman's hose and nozzle are designed to eject 10 litres of water per second; the water jet must be capable of reaching a height of 10 meters. The fireman can hold the nozzle with a force not exceeding 100 Newtons. The hose leading to the nozzle is 10 meters long and has a diameter twice that of the nozzle. Calculate the pressure required at the inlet to the hose, the nozzle diameter and the minimum pumping power required for feeding hose and nozzle.

4.9 Estimate a lower and an upper bound for the drag on a Corvette driving at 70 mph. Calculate the corresponding power requirements. Compare with the actual power of a Corvette engine, and comment.

4.10 Two fluids, 1 and 2, flow between fixed parallel plates spaced a distance H apart. The flow is perfectly stratified, and the interface is located at a height h over the lower plate. The flow is laminar, and the two flow-rates are equal. Express h/H in terms of the ratio of the viscosities and densities of the two fluids.

4.11 Consider flow of a fluid in the wake downstream of a flat plate at zero incidence of length L; the free stream velocity is U. Write the differential equations and boundary conditions which describe flow in the wake. Describe which approximations, if any, you would make before trying to construct a solution to this problem. Estimate the thickness of the wake as a function of distance from the trailing edge of the plate.

4.12 Consider the classical problem of a free jet of liquid impinging on a flat plate which is located at an angle θ with respect to the flow direction of the jet. Look up the problem in any standard book on engineering fluid mechanics, and discuss in detail on which assumptions the solution (i.e., the calculation of the force exerted on the plate) is based. Using the same assumptions, find a shortcut to the solution where there is no need to write down the Bernoulli equation.

SUGGESTED READING

The literature on fluid mechanics is so ample that it is impossible to give a succinct guide to it. The following books are strongly recommended: Aris (1962) for mathematical preliminaries; Astarita and Marrucci (1974) for non-

Newtonian Fluid Mechanics; Batchelor (1967) for a general introduction, viscous flow, and kinematics; Brodkey (1967) for two-phase flow; Denn (1980) for an engineering approach to the subject; Finlayson (1972) for variational methods; Happel and Brenner (1965) for low Reynolds number flows; Hinze (1975) for turbulence; Lamb (1932) for the hydrodynamics of ideal fluids; Landau and Lifshitz (1959) as a general reference book and for compressible flow; Panton (1984) for incompressible flow theory; Schlichting (1979) for boundary layer theory; Serrin (1959) for mathematical background; Whitaker (1968) for a lively introduction to all areas of fluid mechanics. The few other references quoted below refer to material in the chapter in an obvious way.

REFERENCES

Aris, R., Vectors, Tensors, and the Basic Equations of Fluid Mechanics. Prentice-Hall, Englewood Cliffs, NJ, 1962.

Astarita, G. and G. Marrucci, Principles of Non-Newtonian Fluid Mechanics. McGraw-Hill, New York, 1974.

Batchelor, G.K., An Introduction to Fluid Dynamics. Cambridge University Press, 1967.

Brodkey, R.S., The Phenomena of Fluid Motions. Addison Wesley, Reading, MA, 1967.

Denn, M.M., Process Fluid Mechanics. Prentice-Hall, Englewood Cliffs, NJ, 1980.

Finlayson, B.A., The Method of Weighted Residuals and Variational Principles. Academic Press, New York, 1972.

Happel, J. and H. Brenner, Low Reynolds Number Hydrodynamics. Prentice Hall, Englewood Cliffs, NJ, 1965.

Hele Shaw, H.J.S., Nature, 58 (1898) 34.

Hinze, J.O., Turbulence. McGraw-Hill, New York, 1975.

Lamb, H., Hydrodynamics. Cambridge University Press, 1932.

Landau, L.D. and E.M. Lifshitz, Fluid Mechanics. Addison-Wesley, Reading, MA, 1959.

Orr, W.M.F., Proc. R. Irish Acad., 27 (1907) 9; Sommerfeld, A., Atti 4o Congr. Intern. Matem., Rome, 3 (1908) 116.

Panton, R.L., Incompressible Flow. Wiley, New York, 1984.

Schlichting, H., Boundary Layer Theory. McGraw-Hill, New York, 1950 (7th edn., 1979).

Serrin, J., Mathematical Principles of Classical Fluid Mechanics. Enc. of Physics, 8/1, Springer-Verlag, Berlin, 1959.

Squire, H.B., Proc. Roy. Soc. London, A, 142 (1933) 621.

Whitaker, S., Introduction to Fluid Mechanics. Prentice-Hall, Englewood Cliffs, NJ, 1968.

Chapter 5

Distributions

5.1 INTRODUCTION

There are some well known problems in chemical engineering where distribution functions come into play; the classical example is that of the residence time distribution (RTD) in a reactor. In this chapter, we discuss some basic features of the classical problems, and we present some extensions to nontraditional problems. In particular, we focus attention on problems where in actual fact the distribution is a histogram, but things are made easier by considering continuous distribution functions rather than histograms. The most important example is that of mixtures with very many components.

Suppose one asks oneself what is the distribution of the age of the population of, say, New York City. That might be of interest to somebody in charge of social services there, on the improbable hypothesis that there is somebody like that. Now the population of New York is large but finite, and hence in principle the age distribution is represented by a histogram — some finite number of people being 35 years old, some other finite number being 36 years old, and so on. A histogram of this type would have slightly more than 100 vertical bars, the length of each one being of the order of 1% of the total population. Suppose, however, we wish to be more detailed than that, and to distinguish between people who are 35 years and one month old and those who are 35 years and two months old. Now we have got a histogram with something like 1,500 bars, which would look almost like a continuous curve if plotted on anything but very large paper indeed. It's at this point that we decide life is made easier by introducing a continuous distribution function.

But let's keep the histogram in mind a bit longer. The height of the bar between 35 and 36 years represents the fraction of the population with age between 35 and 36 years: the vertical scale is dimensionless. The width of the bar is one year, so the area of the bar has dimensions of time. The total area of the histogram has dimensions of time, and what does it really represent? Little thinking is needed to convince oneself that the total area is in fact one year — it coincides with the arbitrarily chosen interval of ages we have decided to lump into one bar. Indeed, when we go to the one-month histogram, the length of the bars will on the average decrease by a factor of 12, but the same length is spanned on the ages axis — from 0 to something like 108 years. Hence the new histogram has an area which is 12 times smaller than the preceding one. This doesn't make much sense, does it? It seems like we are going to end up with a histogram of zero area if we increase our level of fussiness down to days, hours, minutes, seconds, and so on. The length of the bars would go to zero, since if we specify an age with excessive accuracy, the probability of finding a citizen of exactly that age becomes negligibly small. We can avoid this unpleasant situation by requiring the *area* of a bar to be proportional to the fraction of the population having an age in the specified interval: say if δt is the specified interval, and f is the height of the bar, the area is $f\delta t$. Now we can go the limit $\delta t = 0$ without any problem, and the continuously varying height of the bars, $f(t)$, will be interpreted as follows: $f(t)\mathrm{d}t$ is the fraction of the population having age between t and $t + \mathrm{d}t$. Notice that the value of the distribution function $f(t)$ has dimensions of an inverse time; more generally, it has dimensions inverse to those of its argument.

Sometimes we may wish to have another type of information. For instance, we want to know simply the fraction of the total population which is younger than 65 years. This could be of interest since the remaining fraction cannot be expected to do any useful work, and neither can it be expected to be taken care of by the parents, so somehow society as a whole has to take care of them. Well, that's easy; the required fraction is simply the integral of $f(t)$ between zero and 65 years. Had we wished to know the fraction of the total population younger than any specified age t, $F(t)$, this would simply be the integral of $f(t)$ from zero to t. We have now obtained the cumulative distribution function $F(t)$.

Now let's look at the RTD case. First consider a CSTR. If the mixing is really as efficient as the definition of a CSTR requires it to be, the probability of an element of fluid getting out of the reactor is independent of how long it has resided in the reactor. It follows that the number of fluid elements with residence time t getting out per unit time is simply proportional to how many such elements there are in the reactor. Little thinking is needed to convince oneself that this implies that $f(t)$ is proportional to minus its own derivative, i.e., that $f(t)$ is of the form $A\exp(-t/\theta)$. The integral of $f(t)$ between 0 and ∞ is thus equal to $A\theta$. Since we want to normalize the distribution function in such a way that its integral is unity, we choose $A\theta = 1$, and hence we have obtained the RTD as being $f(t) = \exp(-t/\theta)/\theta$. What is the physical significance of the parameter θ? It is, in some sense, the *average* residence time in the reactor. Now define $x = t/\theta$, i.e., x is the residence time of a particular fluid element measured as a fraction of the average residence time. $f(t)\mathrm{d}t$ becomes simply $\exp(-x)\mathrm{d}x$.

Let's jump to the other extreme, the plug flow reactor (PFR). Here it is easier to work with $F(t)$: if θ is the residence time (which is the same for all elements in the reactor), $F(t)$ is zero for $t < \theta$, and it is unity for $t > \theta$, i.e., $F(t) = H(t-\theta)$, where $H()$ is the Heaviside step function. Since $f(t)$ is the derivative of $F(t)$, one concludes that $f(t) = \delta(t-\theta)$, where $\delta()$ is the delta function. The delta function has a value of zero for all x except $x = 0$, and its integral between any negative number and any positive number is unity. $\delta(x)$ can be written in a variety of ways, but the most useful one is as follows:

$$\delta(x) = \lim_{\varepsilon=0} \exp[-(x/\varepsilon)^2]/\varepsilon\sqrt{\Pi} \tag{5.1.1}$$

For any finite value of the parameter ε, the RHS is a Gaussian distribution, and ε measures the *spread* of the distribution. Suppose for instance that a shooter is aiming at a target; the space distribution of his hits is Gaussian; the better the shooter is, the smaller the value of ε. In the limit of a perfect shooter, who hits invariably on the target, the distribution becomes a delta function and $\varepsilon = 0$.

The RTDs of the CSTR and the PFR are, intuitively, the two extreme cases. This often happens to be the case, though not invariably, since it is possible to have a distribution which is even more heavily weighted in favour of the low values of t than the exponential function. However, let's consider the case where we are willing to presume that the exponential and the delta distributions are the two extreme cases. We want to consider intermediate cases, since in actual life we don't expect things to be as extreme as we have assumed them to be in the analysis of the CSTR and the PFR. There is a very powerful form of distribution function, which contains only one parameter, α, and yet is capable of describing every intermediate case between the exponential and the delta distributions; it is called the gamma distribution $\phi(x, \alpha)$. The definition of ϕ is as follows:

$$\phi(x,\alpha) = \alpha^{\alpha} x^{\alpha-1} \exp(-\alpha x)/\Gamma(\alpha); \quad \alpha \geq 1 \tag{5.1.2}$$

where Γ(α) is the gamma function or generalized factorial:

$$\Gamma(\alpha) = \int_0^\infty \exp(-y) y^{\alpha-1} \mathrm{d}y \tag{5.1.3}$$

Notice that, if α is integer, Γ(α) is the factorial of α – 1, with the factorial of zero being unity. Also, Γ(α) = (α –1)Γ(α – 1), so that Γ(α) is completely specified once its values between α = 0 and α = 1 are given. The gamma distribution ∅(x, α) degenerates to the exponential distribution when α = 1, and to the delta distribution as α approaches ∞. Notice that ∅(0,α) = 0 unless α = 1.

Given a distribution function $f(x)$, one can define the Nth moments of it, M_N, as follows:

$$M_N = \int_0^\infty f(x) x^N \mathrm{d}x \tag{5.1.4}$$

For a normalized distribution, the zeroth moment is unity as a consequence of the normalization. The first moment is simply the average value of x, and if x has been normalized with respect to its average the first moment is unity as well. Higher order moments give information about the shape of the distribution function; in particular, the second moment measures the skewness of the distribution. The moments of the gamma distribution are as follows:

$$M_N = (\alpha + N - 1)(\alpha + N - 2)\ldots(\alpha + 1) / \alpha^{N-1} \tag{5.1.5}$$

In particular, for α = 1 (the exponential distribution), M_N is simply the factorial of N, while for α = ∞ (the delta distribution), all the moments are unity.

Another classical problem where distribution functions arise is that of the liquid side mass transfer coefficient k_C in gas–liquid systems. By considering the analog of the heat transfer problem discussed in Section 3.2.3, we conclude that the Sherwood number $k_C L/D$ is proportional to $L/\sqrt{(Dt)}$, where L is the characteristic length and D is the liquid side diffusivity. The equation for k_C is thus:

$$k_C = K\sqrt{(D/t)} \tag{5.1.6}$$

with K a number of order unity. The interesting question is what the time t appearing in Eq. (5.1.6) might be. Equation (5.1.6) can be obtained by considering the unsteady diffusion in a surface element of liquid, in which case k_C is the instantaneous mass transfer coefficient at time t; so t is identified with the time elapsed since the liquid element considered reached the interface. In this sense, t is really seen to be the "age" of the element considered. Efficient mixing of the liquid phase will decrease the average life of elements at the interface, and hence will increase the value of k_C.

If we consider cases such as laminar flow down a wetted wall column or a sphere (the latter was discussed in Section 4.4), or the steady uprising motion of a gas bubble (see Appendix 5.1), all elements of fluid stay at the interface for the same amount of time θ, and hence the average mass transfer coefficient is simply $2K/\sqrt{(D\theta)}$. However,the situation is not always so simple, and hence one needs to consider the distribution of t values for all the interface, and one calculates an average mass transfer coefficient given by (the constant K is in fact $1/\sqrt{\Pi}$):

$$k_c = \int_0^\infty f(t)\sqrt{(D / \Pi t)}\,\mathrm{d}t \tag{5.1.7}$$

$f(t)$ is really a distribution of ages, and our example of the population of New York is going to be of help. At any given time the distribution is presumably the same as that at any other time: our equivalent New York has a fixed total population, with births balancing exactly deaths at every time (when a surface element is sucked into the liquid by mixing, another one must appear if the interface area is constant). Now what kind of $f(t)$ can we imagine? The stated restrictions imply that $\mathrm{d}f/\mathrm{d}t$ cannot be positive: there cannot be more 50-year-olds than 40-year-olds, since the 50-year-olds were 40-year-olds ten years before. If the probability of being sucked in from the surface is independent of the age of the element (the probability of death is independent of the age of the individual) we again conclude that $f(t)$ is of exponential form, with $1/\theta$ being the fraction of the interface area sucked in per unit time. What for the case where all elements stay at the surface for the same time θ? Everybody dies at age 50, on the spot. So we won't find anybody older than 50, say $f(t) = 0$ when $t > \theta$. Since $F(\infty) = 1$ and $\mathrm{d}F/\mathrm{d}t = f$, we conclude that $F(t) = 1$ when $t > \theta$. Since nobody dies any younger than 50, $\mathrm{d}f/\mathrm{d}t = 0$ at $t < \theta$: all those who now are 30 will be 31 one year hence, and so f is a constant up to age 50. The constant value must be $1/\theta$ for the zeroth moment to be unity, and hence we conclude that:

$$f(t) = H(\theta -)/\theta \tag{5.1.8}$$

If we now want to normalize so that the first moment is unity as well, we define $x = t/\theta$ and we get $f(x) = H(1 - x)$, as opposed to $\exp(-x)$ in the random case. It again looks like the two cases we have considered are the extreme ones, and that reality will lie somewhere in between. We would like to use the gamma distribution, but this time one of the extremes is not the delta function. However, this is easily taken care of. Define $g(x)$ as $-\mathrm{d}f/\mathrm{d}x$. The random case produces $g = \exp(-x)$, and the other extreme produces $\delta(x - 1)$, so one can use a gamma distribution for $g(x)$.

Appendix 5.1

In the steady upwards motion of a gas bubble through a liquid at large Reynolds number, the flow pattern in the liquid phase is that of an ideal fluid, see Section 4.6.2. Consider a fluid element which at time zero is right at the upper pole of the bubble. This element will stay at the interface until it reaches the lower pole. Since the liquid is not in vertical motion, the element considered reaches the lower pole when the bubble has moved upwards a distance equal to its diameter; hence the time of exposure to the gas phase is equal to the ratio of the bubble diameter to its velocity. This time is the same for all fluid elements which come into contact with the gas phase, and hence the distribution is a delta one.

5.2 CONTINUOUS THERMODYNAMICS OF PHASE EQUILIBRIA

When dealing with mixtures which have very many components, it is sometimes useful to describe the mixture as a continuous one. There are two advantages in such a description. First, one often is interested only in some average property of the mixture: for example, in desulphurization of an oil cut one is only interested in the residual amount of chemically bound sulphur, and not in the detailed concentrations of all the sulphur containing compounds. A continuous description offers simple direct methods of estimating such averages. The second advantage lies in the fact that, when the number of components becomes very large, keeping track of all of them becomes very cumbersome, and one wants to substitute for the real mixture a model one consisting of a few pseudocomponents. Since a continuous description contains integrals of quantities weighted by an appropriate distribution function, one can look for the best quadrature points and thus identify the pseudocomponents rationally.

In the discrete description of mixtures, individual components are identified by a subscript $I = 1, 2, \ldots N$, where N is the total number of components. The assignment of the values of I to individual components is entirely arbitrary, although often one chooses some assignment which somehow relates I to some significant property such as carbon number. The subscript I is really simply a label assigned arbitrarily to the components in order to identify them. In a continuous description, the label is chosen as some continuous variable u which is allowed to take values over some range, typically 0 to ∞. Bob rather dislikes this abstraction, and we point out that this is less strange than it may appear at first sight. For instance, oil cuts are often identified by their boiling points, but in actual fact the boiling temperature is a continuous variable — there will be some cut on a distillation tray at whatever temperature that tray may operate.

The fundamental distribution in a continuous description is the mass distribution $m(u)$: $m(u)\mathrm{d}u$ represents the mass of components with label between u and $u + \mathrm{d}u$. Notice that, if one chooses a dimensional label such as a boiling point, the value of $m(u)$ does not have units of mass. Since the label u identifies components, the molecular weight distribution $M(u)$ is in principle known. The distribution of number of moles n(u) is then simply $n(u) = m(u)/M(u)$. Whenever an equation applying to the discrete description contains terms involving the summation over the index I, the corresponding continuous equation contains integrals over the variable u; for simplicity, such integrals will be indicated with triangular brackets. Thus for instance the total mass and the total number of moles of the mixture are given by:

$$m = <m(u)>; \; n = <n(u)> \qquad (5.2.1)$$

The only real difficulty with the thermodynamic theory of mixtures is the definition of partial molar properties. Let L stand for any extensive thermodynamic property such as free enthalpy. At fixed pressure and temperature, the value of L depends on the composition of the mixture, i.e., on the number of moles distribution. This means that a functional exists whose argument function is $n(u)$ and whose value is L:

$$L = \mathbf{L}[n(u)] \tag{5.2.2}$$

Since L is an extensive property, the functional must satisfy the following condition of homogeneity:

$$\alpha\mathbf{L}[n(u)] = \mathbf{L}[\alpha n(u)] \tag{5.2.3}$$

Now we define ε from $\alpha = 1 + \varepsilon$, and we subtract L from both sides of (5.2.3) to obtain:

$$\varepsilon\mathbf{L}[n(u)] = \mathbf{L}[(1+\varepsilon)n(u)] - \mathbf{L}[n(u)] \tag{5.2.4}$$

We next proceed to express the first term on the right via a Frechet expansion truncated at the first term (see Appendix 5.2):

$$\mathbf{L}[(1+\varepsilon)n(u)] = \mathbf{L}[n(u)] + \delta L[n(u);\varepsilon n(u)] + O(\varepsilon^2) \tag{5.2.5}$$

The Frechet differential is by definition linear in its second argument, and hence substituting (5.2.5) into (5.2.4), dividing through by ε, and taking the limit for ε approaching zero yields:

$$\mathbf{L}[n(u)] = \delta\mathbf{L}[n(u);n(u)] \tag{5.2.6}$$

Since the Frechet differential is linear in its second argument, it must be expressible as the integral of $n(u)$ times some weighing function $l(u)$:

$$\delta\mathbf{L}[n(u);n(u)] = <l(u)n(u)> \tag{5.2.7}$$

Notice however that the kernel $l(u)$ (which is called the functional derivative of **L**) is itself a functional of the mole number distribution $n(u)$. Equations (5.2.6) and (5.2.7) together give the final result:

$$L = <l(u)n(u)> \tag{5.2.8}$$

Equation (5.2.8) identifies $l(u)$ as the distribution of the partial molar values of L: indeed, in the discrete description (5.2.8) would hold with < > interpreted as the sum over I.

Now consider a small change of the number of moles $\mathrm{d}n(u)$. Using again the functional derivative, one calculates that the corresponding $\mathrm{d}L$ is:

$$\mathrm{d}L = \langle l(u)\mathrm{d}n(u)\rangle \tag{5.2.9}$$

However, differentiation of (5.2.8) yields:

$$\mathrm{d}L = \langle l(u)\mathrm{d}n(u) + \langle n(u)\mathrm{d}l(u)\rangle \tag{5.2.10}$$

Comparison of (5.2.9) and (5.2.10) yields the continuous form of the Gibbs–Duhem equation:

$$\langle n(u)\mathrm{d}l(u)\rangle = 0 \tag{5.2.11}$$

The most important quantity to consider at conditions of constant temperature and pressure is the free enthalpy G. The corresponding partial molar free enthalpy is the distribution of chemical potentials $\mu(u)$, say:

$$G = \langle n(u)\mu(u)\rangle \tag{5.2.12}$$

This is all that is needed, in principle, to analyze phase equilibria, the condition of equilibrium being that $\mu(u)$ is the same in all phases at equilibrium. Chemical equilibria pose some additional difficulty, which will be discussed in the next section.

Let's now move on to some concrete problems. We begin by considering the simplest possible case: vapour liquid equilibrium in a system where the liquid phase is an ideal solution, so that Raoult's law applies. In this case, the only parameter entering the vapour liquid equilibrium equation is the vapour pressure b of each individual component: every component is entirely identified by its vapour pressure. Of course, more than one component may have the same vapour pressure, but if that is the case all such components really behave as if they were a single one. It follows that one may use the vapour pressure b itself as the label identifying the components. We have learned a lesson here: the value of some specific physical property can be used as the label only if the behaviour of that component is entirely determined by the value of that one property. If that is not the case, the value cannot be used as a label, since two components with the same value of the chosen property may well have different values of some other property which is of importance.

For the case at hand, we are going to use b as the label. Let $x(b)$ and $y(b)$ be the mole fraction distributions in the liquid and in the vapour, respectively (the mole fraction distribution is simply $n(b)/n$). For the time being, we assume that conditions are such that both a liquid and a vapour phase will in fact be present at equilibrium. Raoult's law can be written as:

$$py(b) = bx(b) \tag{5.2.13}$$

If the integral of (5.2.13) is taken, the LHS simply yields the pressure p, since $<y(b)> = 1$. Hence one concludes that pressure equals the first moment of the liquid phase mole fraction distribution:

$$p = <bx(b)> \tag{5.2.14}$$

Ralph has something important to say at this stage. The way we have defined $x(b)$ and $y(b)$, we have a very funny value of the mole fraction distribution, since it has units of an inverse pressure. This is in fact true, since we have not yet normalized the label b with respect to its average value. Let's do that, so that things go back to having their usual units; let B be the "average" vapour pressure (to be defined more exactly in a minute), and let β be the normalized vapour pressure, $\beta = b/B$. This of course implies that both the zeroth and the first moment of $x(\beta)$ are unity. This identifies the value of B as the one which sets to unity the first moment of the $x(\beta)$ distribution. Equation (5.2.14) now becomes:

$$p = B < \beta x(\beta)> = B \tag{5.2.15}$$

We first observe that we have correctly recovered the asymptotic case where there is only one component. Indeed, if there is only one component, its vapour pressure must be B, its mole fraction in the liquid phase is unity, and hence the pressure is going to be equal to B. Formally, this is described by setting $x(\beta)$ to be $\delta(\beta - 1)$, and Eq. (5.2.15) now yields the required result $p = B$. In the remainder of this chapter, we will often make use of this single component identity (SCI) requirement.

The second observation about (5.2.15) is that, since $x(\beta)$ has been normalized, its first moment is unity, and hence the "average" vapour pressure of the mixture is simply the equilibrium total pressure. Equation (5.2.13) becomes, in normalized form:

$$y(\beta) = \beta x(\beta) \tag{5.2.16}$$

Should the liquid phase mole fraction distribution be a gamma one, one obtains immediately:

$$y(\beta) = \phi(\beta, \alpha + 1) \exp\beta \tag{5.2.17}$$

The mole fraction distribution in the gas phase is thus *not* a gamma one, and it tends to be sharper than the one in the liquid phase. The physical significance of this is best understood by considering that Eq. (5.2.17) always yields $y(0) = 0$, since $\alpha > 0$. Thus even if the liquid phase contains a finite amount of components with a vanishingly small vapour pressure ($\alpha \leq 1$), none of these are present in the vapour phase, as one would expect. Flash calculations for this case are discussed in Appendix 5.2.

Liquid mixtures, however, have a rather nasty tendency to be nonideal. This we deal with by defining activity coefficients, as we engineers always do when we don't know what to do on a problem: we bury our ignorance into a correction coefficient. This time we cannot use b as a label, because two components with the same vapour pressure may well have different activity coefficients; so we need an unequivocal label u. If $\mu^*(u)$ is the chemical potential of component u in its pure form, for an ideal solution the chemical potential distribution would be given by:

$$\mu(u) = \mu^*(u) + RT\ln x(u) \tag{5.2.18}$$

We are now in the position of defining the distribution of activity coefficients $\tau(u)$ as follows:

$$\mu(u) = \mu^*(u) + RT\ln[\tau(u)x(u)] \tag{5.2.19}$$

The vapour liquid equilibrium conditions are now:

$$py(u) = b(u)\tau(u)x(u) \tag{5.2.20}$$

$$p = <b(u)\tau(u)x(u)> \tag{5.2.21}$$

One cannot of course proceed any further without writing down constitutive equations for the activity coefficients. However, generalizations of equations usually written down for mixtures with a finite number of components is comparatively trivial. We discuss only one example in the following, namely, the one-constant Margules equation. For a three-component mixture, the equations for the activity coefficients are as follows (those for the other components are obtained by permutation of indices):

$$\ln\tau_1 = A_{12}x_2{}^2 + A_{13}x_3{}^2 + (A_{12} + A_{13} - A_{23})x_2x_3 \tag{5.2.22}$$

In this equation, the A parameters are thought of as unlike species binary interaction parameters. The corresponding distribution of interaction parameters is of course some function $A(u,u')$. Thus the equation for the activity coefficient distribution becomes:

$$\begin{aligned}\ln\tau(u) = {} & <A(u,u')x(u')^2> + <[A(u,u') \\ & + A(u,u'') - A(u',u'')]\, x(u')x(u'')>\end{aligned} \tag{5.2.23}$$

We now turn our attention to liquid–liquid equilibria, and we first have to address the question of the possibility of phase splitting. In a two-component mixture, phase splitting in two liquid phases is known to be possible. In a three-component mixture, splitting into three phases is a possibility. If all three-component pairs exhibit a large miscibility gap, there will be one phase regions near the corners of a classical triangular plot, two phase regions near the sides and away from the corners, and a three phase region in the central part. In fact we know that even a one-component system may exhibit three phases (at the triple point). So the number of possible phases can be larger than the number of components, and it keeps growing as the number of components increases. So what about a continuous mixture, where the number of components has essentially gone to infinity? Yes, Bob, it does run against physical intuition, but to be consistent we have to allow for an infinite number of phases. So we introduce a phase label z, and we indicate with curled brackets integrals over z.

Let $q(z)$ be the fraction of the total mass in phase z. The fractional mass distribution $q(z)$ must of course satisfy the requirement $\{q(z)\} = 1$. Given any quantity L the value of which would usually be calculated by the lever rule for two-phase systems, it is given by $\{L(z)q(z)\}$.

When there is only one phase the free enthalpy per mole of mixture is given by:

$$G = <\mu^*(u)x(u)> + \mathbf{F}[x(u)] \tag{5.2.24}$$

where the value of **F**[] is the free enthalpy of mixing. For ideal solutions, the free enthalpy of mixing is given by:

$$\mathbf{F}[x(u)] = RT<x(u)\ln x(u)> \tag{5.2.25}$$

Now notice that, since $\mu^*(u)$ is arbitrary (the free enthalpy of a pure component is known to within an arbitrary additive constant, which is different for different components), all statements about free enthalpy must be invariant under addition of an arbitrary linear transform of $x(u)$. However, this does not rule out statements about the convexity of the free enthalpy hypersurface, since a linear transform is neither convex nor concave.

Now suppose the system has split into a phase distribution $q(z)$. Let $x(u,z)$ be the mole fraction distribution in phase z. The average mole fraction of the system as a whole, $x(u)$, is given by:

$$x(u) = \{q(z)x(u,z)\} \tag{5.2.26}$$

Let $\varepsilon X(u,z)$ be defined as:

$$x(u,z) = x(u) + \varepsilon X(u,z) \tag{5.2.27}$$

Whatever the value of the parameter ε, the two functions $q(z)$ and $X(u,z)$ must of course satisfy the following conditions:

$$\{q(z)\} = 1;\ <X(u,z)> = 0;\ \{q(z)X(u,z)\} = 0; \tag{5.2.28}$$

The free enthalpy per mole of any of the phases is given by:

$$G(z) = \mathbf{G}[x(u,z)] = <\mu^*(u)x(u,z)> + \mathbf{F}[x(u,z)] \tag{5.2.29}$$

Thus the total free enthalpy per mole of the system is given by:

$$G = \{q(z)G(z)\} \tag{5.2.30}$$

Now notice that the first term on the RHS of (5.2.29), when substituted into (5.2.30), yields simply the arbitrary term of the one phase system:

$$\begin{aligned}\{q(z)<\mu^*(u)x(u,z)>\} &= <\mu^*(u)\{q(z)x(u,z)\}> \\ &= <\mu^*(u)x(u)>\end{aligned} \tag{5.2.31}$$

This is important, because it gives meaning to the following statement: "the free enthalpy corresponding to a given composition $x(u)$ as given by (5.2.24) is no more than the free enthalpy as given by (5.2.30) for *any* choice of $q(z)$ and $\varepsilon X(u,z)$ which satisfies (5.2.28)". A composition $x(u)$ for which this statement holds is called a globally convex point in composition space, and little thought shows that no phase splitting will occur at a globally convex point.

We have not made any use of the parameter ε so far. We now proceed to define a locally convex point as a composition $x(u)$ such that the statement above holds true for vanishingly small ε, but not necessarily in general. Of course, a globally convex point is also locally convex, but the converse is not true. Now consider a composition which is not locally convex. That means that there is some phase distribution $q(z)$, with all phases having compositions differentially close to the average one, which has a lower free enthalpy than the one phase system. This is recognized as a spinodal point: it is unstable to infinitesimal phase splitting, which can always occur simply by Brownian motion. If there are compositions which are spinodal points, phase splitting is guaranteed to be possible for some average compositions.

Finally, consider a composition which is locally convex but not globally convex. That is a metastable composition: it is stable to infinitesimal phase splitting, but some distribution of phases exist which will result in a lower free enthalpy than that of the one phase state.

Notice that the free enthalpy of mixing in an ideal solution is globally convex everywhere, and its convexity approaches infinity in the neighbourhood of the pure components compositions, i.e., in the neighbourhoods of $x(u) = \delta(u - u^*)$, where u^* identifies the pure component considered. This has two consequences. First, the nonideal contribution to the free enthalpy of mixing needs to be concave for phase splitting to be possible. Second, since the concaveness of the nonideal contribution never approaches infinity (as the success of the theory of dilute solutions proves), any system is guaranteed to have globally convex compositions in the neighbourhoods of $\delta(u - u^*)$. If a system has a spinodal region of compositions, metastable regions are thus guaranteed to exist between the spinodal and the globally convex regions.

Appendix 5.2

Functionals and their derivatives have been discussed in the Appendix to Section 2.6. Here we only need to add that, in the case of multicomponent mixtures, a simple Euclidean norm presumably induces the appropriate topology with respect to which the constitutive functionals are Frechet differentiable.

Now consider again the case where Raoult's law applies; we wish to consider the solution of a classical flash problem. Let $z(b)$ be the mole fraction distribution of the feed to the flash chamber. In this case it is *not* useful to normalize the label b, because there is no guarantee that the pressure in the flash chamber is such that two phases will in fact exist. Therefore we proceed as follows. *If* there will be two phases in the flash chamber where the pressure has some preassigned value p, Eq. (5.2.13) will hold. Now let τ be the mass fraction of the vapour phase (which is restricted to $0 \le \tau \le 1$). The mass balance equation reads:

$$\tau y(b) + (1-\tau)x(b) = z(b) \tag{5.2.a.1}$$

Equations (5.2.13) and (5.2.a.1) can be solved for either $y(b)$ or $x(b)$, yielding:

$$x(b) = \frac{pz(b)}{p + \tau(b - p)} \tag{5.2.a.2}$$

$$y(b) = \frac{bz(b)}{p + \tau(b - p)} \tag{5.2.a.3}$$

The bubble point is $\tau = 0$, and one gets $x(b) = z(b)$ and $y(b) = bz(b)/p$, as it should be; the condition $<y(b)> = 1$ establishes the value of p corresponding to the bubble point; any pressure higher than that will result in only a liquid phase with composition $z(b)$. Conversely, the dew point is $\tau = 1$, and one gets as expected $y(b) = z(b)$ and $x(b) = pz(b)/b$. Again, the condition $<x(b)> = 1$ establishes the dew point value of p, and any pressure lower than that will result in only a vapour phase with composition $z(b)$. If p is between the dew and the bubble point values, Eqs. (5.2.a.2–3) have nontrivial solutions, and the conditions $<x(b)> = <y(b)> = 1$ establish the phase split τ. With anything less trivial than Raoult's law, the algebra becomes cumbersome, but it is obviously possible to satisfy both the mass balance and the equilibrium equations, though not with the same type of distribution in the vapour and liquid phases.

5.3 CONTINUOUS HOMOGENEOUS CHEMICAL EQUILIBRIA

We begin by considering homogeneous (i.e., one-phase) chemical equilibria. A chemical equilibrium composition is guaranteed to be homogeneous if the mixture of all components present (reactants, products and inerts) is miscible in all proportions. When a mixture contains many components, it is easier to determine the composition corresponding to chemical equilibrium by looking directly for the minimum of free enthalpy over all compositions which are stoichiometrically possible (i.e., accessible from the initial one), rather than solving the coupled nonlinear equilibrium conditions for every reaction. Searching directly for the minimum is obviously the only possibility when the number of components and of reactions is going to infinity, as indeed it does in a continuous description. This time Bob has no protest to lodge: having swallowed infinitely many phases, infinitely many reactions don't bother him at all.

However, one meets immediately with a difficulty which is best explained by considering a very simple case: a dimerization reaction $2A = B$. To make things even simpler, we'll assume the A–B mixture to be ideal, and the standard free enthalpy of the reaction to be zero. This means that $\mu^*_{\mathrm{B}} = 2\mu^*_{\mathrm{A}}$: this time knowledge of the standard free enthalpy of reaction (which can be measured) makes the two pure components free enthalpies determined to within a single additive constant, rather than two constants. Since the solution is ideal, the equilibrium condition reduces of course to $y_{\mathrm{B}} = y_{\mathrm{A}}^2$, and setting $y = y_{\mathrm{A}}$ one obtains the equilibrium value $y = 0.618$. Not very exciting, is it? But now let's search the minimum of $g = G/RT$, and let $Q = \mu^*_{\mathrm{A}}/RT$. One is tempted to proceed as follows. Since the solution is ideal, we write:

$$g = Qy + Q(1-y)/2 + y\ln y + (1-y)\ln(1-y) \tag{5.3.1}$$

We now want to find the minimum, and hence we set the derivative to zero and ... yes, Sue, you've spotted the problem. We get:

$$0 = Q/2 + \ln[y/(1-y)] \tag{5.3.2}$$

but this does not yield the value of y because Q is known only to within an arbitrary additive constant. If for instance we set arbitrarily Q to zero, we would get the wrong answer $y = 0.5$, Well, Sue, you've spotted the problem — any idea on how we may resolve it? Why hasn't Q cancelled out of the equations? Try it at home with the isomerization reaction A = B, and you'll see that Q does cancel out. Try it with the reaction $A+B = C+D$, and you'll get *three* Q's floating around, and yet they all cancel out. Why is $2A = B$ so nasty? If you only think of it a little bit, the answer becomes obvious: Q cancels out when the reaction is equimolar, it does not cancel out otherwise. Now since most reactions are not equimolar, we don't have a freak with dimerization; we are in serious trouble.

Sue comes up with another brilliant idea, it's her good day. Chemical reactions, she says, may result in a change of the total number of moles, but they don't cause any change in the total mass. So why don't we work with masses instead? That's exactly what one needs to do, much to the dislike of Ralph, who remembers our mass based formalism in Section 2.4 and hasn't quite swallowed it yet. Anyhow, how are we going to write our dimerization reaction in the mass based formalism? We'll have to write $A = B$: when one gram of monomer dimerizes, the result is one gram of dimer. In fact, let σ stand for a stoichiometric coefficient in the mole based formalism; the corresponding mass based formalism coefficient τ is simply σM, with M the molecular weight of the component considered. The dimer has twice the molecular weight of the monomer, and half the stoichiometric coefficient, so in the mass based formalism the two stoichiometric coefficients are equal (and can be normalized to be unity).

We'll write a generic reaction as $\sum\tau_I B_I = 0$, where B_I stands for the chemical symbol of the Ith component, and τ_I is its mass based stoichiometric coefficient, positive for products and negative for reactants. Since if all stoichiometric coefficients are multiplied by the same nonzero scalar one is describing the same reaction, consideration of the case where the scalar is –1 shows that the distinction between reactants and products is arbitrary. Conservation of mass is guaranteed by the requirement that $\sum\tau_I = 0$. Of course, atoms have to be preserved as well, but we will not consider this for the time being. The composition of the mixture (in the continuous description) will be described by either the mass distribution $m(u)$, or the mass fraction distribution $w(u) = m(u)/m$. Chemical potentials are of course to be understood as mass based ones, and so on.

However, before discussing continuous mixtures, we need a few more things with discrete ones first. Suppose there is only one reaction, and let w_{I0} be the initial mass fraction of component I. At any later time, the mass fraction will be:

$$w_I(t) = w_{I0} + \tau_I X(t) \tag{5.3.3}$$

where $X(t)$ is the *extent of reaction* at time t. This shows that the reaction restricts the composition to a one-dimensional trajectory in composition space. In our dimerization example, there were only two components, and so the composition space was also one-dimensional (the composition was identified by the value of y). This, however, is the exception rather than the rule, as consideration of the simple reaction $A + \mathrm{B} = 2C$ shows: only one trajectory, but the composition space is two-dimensional.

What about the case when there are more than one reaction? This time we have a matrix of stoichiometric coefficients, τ_{IK} being the stoichiometric coefficient of component I in reaction K, and Eq. (5.3.3) becomes:

$$w_I(t) = w_{I0} + \Sigma\tau_{IK}X_K(t) \tag{5.3.4}$$

The trajectory is now an M-dimensional one, if $K = 1, 2, \ldots M$, and the M reactions are *independent*. Now there are many ways of defining independence of reactions, which are all equivalent to each other, but the most useful one for generalization to continuous mixtures is as follows: the reactions are independent if the following equation for the multipliers p_K has only the trivial solution $p_K = 0$ for all I:

$$\Sigma\tau_{IK}p_K = 0 \tag{5.3.5}$$

This guarantees that the set of $X_K(t)$'s which satisfies Eq. (5.3.4) is unique: should there be some nonzero p_K's, $X_K(t) + p_K$ would also satisfy it. We are now ready for business, and we move to continuous mixtures.

As we have already hinted before, we are going to allow for infinitely many reactions. A simple example may clarify the issue: cracking of an oil cut. There are very many components, and we want to use a continuous description; but most of the components can crack, and so there are about as many reactions as there are components. We will use a label v for the reactions, and we indicate with curled brackets integrals over v. The stoichiometric coefficients distribution function is $\tau(u,v)$, which represents the stoichiometric coefficient of component u in reaction v. $\tau(u,v)$ can also be thought of as a continuum of reactions: a reaction is entirely identified once the stoichiometric coefficients of it are given. Of course, $\tau(u,v)$ is expected to be a sum of a few delta functions for every v — after all, there aren't going to be more than a few components participating in any given reaction. $\tau(u,v)$ must satisfy the condition $<\tau(u,v) = 0$ for mass to be preserved; hence $\tau(u,v)$ is, for all values of v, orthogonal to all constant functions $f(u)$ = constant. The continuum of reactions $\tau(u,v)$ is independent if the following equation for the multipliers $p(v)$ has only the trivial solution $p(v) = 0$:

$$\{\tau(u,v)p(v)\} = 0 \tag{5.3.6}$$

This of course means that $\tau(u,v)$ is orthogonal only to the zero function $f(v) = 0$ for all values of u. Now let $w(u,t)$ be the mass fraction distribution at time t. Given the initial mass fraction distribution $w(u,0)$, the extent of reaction distribution at time t, $X(v,t)$, determines $w(u,t)$ through the following equation:

$$w(u,t) = w(u,0) + \{\tau(u,v)X(v,t)\} \tag{5.3.7}$$

and the independence condition guarantees that the distribution $X(v,t)$ which satisfies (5.3.7) is unique.

The free enthalpy per unit mass of mixture, G, is of course given by a functional of the mass fraction distribution:

$$G(t) = \mathbf{G}[w(u,t)] = <\mu(u,t)w(u,t)> \tag{5.3.8}$$

where the second equality is obtained because the functional derivative of **G** is the chemical potential distribution $\mu(u,t)$.

The reaction rate distribution $r(v,t)$ and the affinity distribution $\theta(v,t)$ are defined as follows:

$$r(v,t) = \partial X(v,t)/\partial t \tag{5.3.9}$$

$$\theta(v,t) = \mathrm{d}G/\mathrm{d}X(v,t) = <\tau(u,v)\mu(u,t)> \tag{5.3.10}$$

where the derivative called for in (5.3.10) is legitimate because substitution of (5.3.7) into (5.3.8) shows that G is a unique functional of $X(v,t)$. Combination of (5.3.9) and (5.3.10) yields the following fundamental result:

$$\mathrm{d}G/\mathrm{d}t = \{\theta(v,t)r(v,t)\} \tag{5.3.11}$$

The importance of (5.3.11) lies in the fact that, at constant pressure and temperature, the second law of thermodynamics requires dG/dt to be non-positive, and hence (5.3.11) gives a constraint on the possible kinetics of the reactions, say on the relationship between $r(v,t)$ and $X(v,t)$. However, we won't go into kinetics yet, and we limit attention to equilibrium.

At equilibrium, $r(v,t) = 0$, and hence the RHS of (5.3.11) has a maximum at the equilibrium composition $X(v,t) = X^*(v)$. We may drop the explicit dependency of everything of interest upon time, and observe that (5.3.7) gives the composition at any one time in terms of the extent of reaction distribution $X(v)$. Since both the affinity $\theta(v)$ and the reaction rate $r(v)$ obviously depend only on composition (at constant temperature and pressure), it follows that functionals exist which map $X(v)$ into $r(v)$ and $\theta(v)$; such functionals may of course still be specific to the reaction considered, say they may depend parametrically on v itself. Thus we write:

$$\theta(v) = \theta[X(z);v] \tag{5.3.12}$$

$$r(v) = \mathbf{R}[X(z);v] \tag{5.3.13}$$

Every hand in the class is raised. "What is z?" is the essence of what is being asked. Well, z is a dummy reaction label: the affinity of reaction v depends on the whole distribution of extents of reactions, in a way which is specific to reaction v itself; so we need a parametric dependence on v, and a dummy reaction label to identify the extent of reaction distribution. The turmoil does not die out easily, but when Sue suggests that z is something like the integration variable, which we may call anything we wish since it doesn't matter anyhow, Bob declares he is seeing some light.

Now the RHS of (5.3.11) has a maximum at equilibrium, and hence its Frechet differential at the equilibrium composition is zero. But at equilibrium $r(v)$ is zero by definition, and hence the requirement reduces to the following equation, in which an asterisk identifies conditions at equilibrium, z is again a dummy reaction label, and $s(z)$ is some arbitrary displacement of the extent of reaction distribution from the equilibrium one:

$$\{\{\theta[X^*(z);v]u^*(z;v)s(z)\}\} = 0 \tag{5.3.14}$$

The function $u^*(z;v)$ is the functional derivative of $\mathbf{R}$ at the equilibrium extent of reaction distribution $X^*(w)$; this functional derivative still depends parametrically on v, since it may well be different for different reactions.

Sue, we know you are completely lost — forget about the algebra, only think of the basic implications. We are trying to do the analog of what was done in Section 2.5, i.e., we are trying to show that some sort of assumption of invertibility of the kinetic equation near equilibrium is needed in order to obtain an equilibrium condition. In (5.3.13), v and z span the same range, and hence the question of whether **R** is invertible is a legitimate one. Only this time we can't go by with something as simple as the Jacobian of the transformation. Sue didn't find the Jacobian a simple concept, and hence anything more complicated isn't going to bother her in the least, it's just as mysterious as the Jacobian was. Be that as it may, let's proceed.

Now let **D** be the set of functions $f(z)$ to which the functional derivative $u^*(z;v)$ is orthogonal, say functions $f(z)$ such that:

$$\{u^*(z;v)f(z)\} = 0 \tag{5.3.15}$$

and let $\theta^*(v)$ be the affinity distribution at equilibrium:

$$\theta^*(v) = \theta[X^*(z);v] \tag{5.3.16}$$

Equation (5.3.14) is now satisfied with $\theta^*(v)$ being *any one* of the functions $f()$ in **D**, and we would have an indeterminate equilibrium condition.

Now suppose we have small displacements from the equilibrium extent of reaction distribution, say $X(z) = X^*(z) + \varepsilon s(z)$, and correspondingly we have a small reaction rate distribution $\varepsilon r(v)$. Onme would calculate:

$$r(v) = \{u^*(z;v)s(z)\} \tag{5.3.17}$$

It follows that displacements $\varepsilon[s(z) + f(z)]$ would result in the same reaction rates, and hence **R** would not be invertible. We can therefore conclude that *if* **R** is invertible at the equilibrium composition (an assumption which Bob regards as a self-evident truth, of course), the set **D** must contain only the zero function $f(z) = 0$. Be it self-evident or not, if we do make this assumption we get the result:

$$\theta^*(v) = 0 \tag{5.3.18}$$

If this is substituted into (5.3.10), we finally get the equilibrium condition in terms of the chemical potential distribution:

$$<\tau(u,v)\mu^*(u)> = 0 \tag{5.3.19}$$

This determines, in principle, the equilibrium composition $m^*(u)$. Notice that, as long as we limit attention to mixtures which are miscible in all proportions, all compositions are globally convex, and hence only one minimum of free enthalpy can exist, so that the $m^*(u)$ which satisfies (5.3.19) is guaranteed to be unique.

5.4 KINETICS IN CONTINUOUS MIXTURES

In this section, we discuss the continuous description of chemical kinetics in multicomponent mixtures. In order to simplify the problems, in most of the following we consider the case where all chemical reactions involved are irreversible, so that we may avoid the issue of chemical equilibrium (of course, real reactions are never truly irreversible; however, the limit of irreversibility is, for the problems at hand, a regular one, and thus the theory below applies to *almost* irreversible reactions). The two essential points which the discussion below hopefully brings out are:

1. The kinetic behaviour of a multicomponent mixture is quite different from that of any one of the components which constitute it.
2. When a lumped (or overall) kinetic equation can be found which describes the behaviour of the mixture in a batch or plug flow reactor, it cannot be applied to other reactor configurations as such, but the intrinsic kinetics of individual components need to be taken into account.

5.4.1 Linear kinetics

We begin by considering the simplest possible case, i.e., a mixture of many reactants all of which are consumed by first order reactions. Furthermore, we begin by considering the simplest possible type of reactor, i.e., a batch reactor — where everything depends only on time t^* elapsed since the reactor was started (we use starred symbols to identify dimensional quantities). Let $g^*(k^*,t^*)$ be the concentration distribution at time t^*, in the sense that $g^*(k^*,t^*)\mathrm{d}k^*$ is the concentration of reactants with kinetic constants between k^* and $k^* + \mathrm{d}k^*$. We have already made a rather crucial choice, i.e., we have decided to use the kinetic constant k^* itself as the label. Is that legitimate? Yes, because if all the reactions are first order and irreversible, the behaviour of every individual reactant depends only on the value of k^*.

Now the feed to the reactor has a concentration distribution $g^*(k^*,0) = G^*(k^*)$, and we may proceed to a bit of normalization. First, we normalize G^* to G, and g^* to g, by requiring the zeroth moment of G to be unity, and by defining $g = g^*/<G^*(k^*)>$:

$$<G(k^*)> = <g(k^*,0)> = 1 \qquad (5.4.1.1)$$

Next, we require the first moment of G to be unity as well, by defining the initial average value of k^* as $\tilde{k} = <k^*G(k^*)>$ and the dimensionless kinetic constant as $u = k/\tilde{k}$, so that:

$$<uG(u)> = <ug(u,0)> = 1 \qquad (5.4.1.2)$$

Finally, we normalize time so that $t = \tilde{k}t^*$, and we define the *lumped concentration* at time t, $U(t)$, as:

$$U(t) = <g(u,t);\ U(0) = 1 \qquad (5.4.1.3)$$

So far, as Ralph is quick to point out, we have done nothing more than laying down some definitions; the fact that the reactions involved are first order ones has not been made use of. Well, that's not exactly true, because we have used k^* itself as the label, which is legitimate only for first order reactions. Bob has another objection: why did we introduce the quantity U? Let's consider a specific example, say a hydrodesulphurization process (we are *not* claiming that hydrodesulphurization reactions are necessarily governed by first order kinetics). One would only be interested in the total residual sulphur content, not in the details of the concentration distribution of all sulphur containing species; hence we are trying to obtain equations for the lumped concentration U. (In actual fact, for the desulphurization example one would be interested in a *weighted* lumped concentration $<h(u)g(u,t)>$, where the weighting function $h(u)$ measures the content of sulphur in species u.)

We now make use of the assumption of first order kinetics by writing:

$$-\partial g(u,t)/\partial t = ug(u,t) \qquad (5.4.1.4)$$

This has so obvious a solution that even Sue gets it without any trouble (well, almost without any):

$$g(u,t) = G(u)\exp[-ut] \qquad (5.4.1.5)$$

We now substitute this into (5.4.1.3) to obtain:

$$U(t) = \langle G(u)\exp[-ut]\rangle \tag{5.4.1.6}$$

i.e., $U(t)$ is the Laplace transform of $G(u)$, with however t playing the rather unconventional role of the transform parameter. Now we need to perform an important check: the equations must be such that, should there be only one reactant, one recovers the classical result; this will be called the single component identity (SCI) requirement. For the case at hand, the check is trivially satisfied: the single component case is described by $G(u) = \delta(u - 1)$, and for this case we get $U(t) = \exp[-t]$, which we know to be the correct result.

In order to proceed, we need to assign the initial distribution G(u), and we'll consider the case where $G(u) = \phi(u,\alpha)$. One obtains:

$$U(t) = \langle \phi(u,\alpha)\exp[-ut]\rangle = [\alpha/(t + \alpha]^{\alpha} \tag{5.4.1.7}$$

Differentiation of (5.4.1.7) with respect to time yields:

$$-\mathrm{d}U/\mathrm{d}t = -\langle uG(u)\exp[-ut]\rangle = [\alpha/(t + \alpha)]^{\alpha+1} \tag{5.4.1.8}$$

Now the crucial point is that t can be eliminated between Eqs. (5.4.1.7) and (5.4.1.8) to yield directly a *lumped kinetic equation* for $U(t)$:

$$-\mathrm{d}U/\mathrm{d}t = U^{(\alpha+1)/\alpha};\ U = [1 + t/\alpha]^{-\alpha} \tag{5.4.1.9}$$

Equation (5.4.1.9) should be discussed in some detail. It tells us that the *apparent lumped order of reaction* is $(\alpha + 1)/\alpha$, i.e., that the mixture as a whole (insofar as only U is observable) behaves kinetically in a way which is quite different from that of any one of its components: the overall reaction appears to be of an order different from unity (except in the limit $\alpha = \infty$ where, as expected, the order is one). Now what is the *physical* reason for this behaviour? Let's look at the problem from a somewhat different viewpoint. The average kinetic constant of the mixture at time zero is $\widetilde{k} = \langle k^*G(k^*)\rangle$, and normalization has made it equal to unity. However, as time progresses, components with large kinetic constants (large u) disappear more quickly, and hence the average dimensionless kinetic constant at some later time, $\langle ug(u,t)\rangle$, is less than unity: as the lumped concentration U decreases, so does the average kinetic constant. But this is the kind of behaviour associated with reactions which are of order larger than unity, and hence the lumped concentration U appears to decrease with an order larger than one.

Now suppose we consider the behaviour of the same mixture in a CSTR. Let T^* be the (dimensional) residence time in the CSTR, and $T = \tilde{k}T^*$ the corresponding dimensionless residence time. Let $g(u)$ be the concentration distribution in the exit stream and in the reactor itself. The mass balance on component u yields:

$$g(u) = G(u)/(1 + uT) \tag{5.4.1.10}$$

The dimensionless concentration at the exit, U_E, is $<g(u)>$, and one would like to have an explicit equation for it in terms of T. First of all, however, let's perform the SCI check: if $G(u) = \delta(u - 1)$, we obtain $U_E = 1/(1 + T)$ which is the correct result. Now suppose $G(u) = \varnothing(u,\alpha)$: we find, with some mild surprise, that $<g(u)>$ is not so easily calculated: the CSTR is a more difficult problem than the batch or PFR ones. Ralph does not agree: he says that we can expand $1/(1 + uT)$ in a power series and there is really no difficulty. Well, let's do that, and after a bit of algebra we obtain:

$$U_E = \sum_0^\infty (-T)^N \frac{\Gamma(\alpha + N)}{\alpha^N \Gamma(\alpha)} \tag{5.4.1.11}$$

Suppose we had been stupid enough to assume that (5.4.1.9) was a *true* kinetic equation — say we had measured it in a batch reactor by measuring U and had forgotten U is a lumped concentration; and suppose for the conditions of our experiment $\alpha = 1$, so that we would have determined $-\mathrm{d}U/\mathrm{d}t = U^2$. Now for a CSTR we would have written the overall mass balance as $1 - U_E = TU_E^2$, and we would have obtained a result which is *different* from what (5.4.1.11) yields for $\alpha = 2$. *An experimentally determined lumped kinetic equation can not be applied blindly to a reactor configuration other than the one of the reactor used to determine it.*

There is also another problem with (5.4.1.11). If T is a small number, the series converges rather quickly, but if T is in excess of unity it converges very slowly indeed. Now the interesting case is the one where T is larger than unity: after all, T is the actual residence time times the average kinetic constant of the mixture, and we would be very poor engineers indeed if we didn't set T to be larger than unity. So what we need is to have an equation which works easily at large values of T, and how are we going to obtain that? Ralph suggests to do the same kind of trick by expanding in a power series around $1/uT = 0$, rather than around $uT = 0$, but when we try to do that we rapidly run into troubles — terms with $N > \alpha$ yield diverging integrals. Rather disappointing, isn't it? Any suggestions?

Bob says that, if we are interested in large T values, we might as well go all the way and only calculate the leading term in the expansion — he suggests neglecting 1 with respect to uT in the denominator of (5.4.1.10). Good, let's give it a try. We obtain after some algebra:

$$U_E = [\alpha/(\alpha - 1)]/T \tag{5.4.1.12}$$

We perform the usual SCI check which yields $U_E = 1/T$ for $\alpha = \infty$, and this is OK. But we most certainly seem to have some serious problem with (5.4.1.12) if $0 < \alpha \leq 1$. Well, in this case one can never neglect 1 with respect to uT in the denominator of (5.4.1.10), because the corresponding integral does not converge. There is no need for the solution at large T's to have a leading term which is linear in $1/T$, and for sufficiently small α it does not. We therefore conclude that, while for single components, and for mixtures with $\alpha 1$, U_E at large T becomes inversely proportional to T, for $0 < \alpha \leq 1$ that is not the case, and a log–log plot of U_E vs. T would never reach a slope of –1.

Having dealt with batch reactors (and equivalently with plug flow ones) as well as with a CSTR, we now move to consideration of reactors with an arbitrary residence time distribution (RTD), $f(t)$, and we define $F(t)$ as the complementary cumulative RTD, say the integral of $f(t)$ from t to ∞, so that $f(t) = -\mathrm{d}F/\mathrm{d}t$, and $F(0) = 1$. The first moment of $f(t)$ is the average dimensionless residence time T. Any suggestions on how to proceed? Do we need to distinguish between maximum mixedness (MM) and maximum segregation (MS)? Ralph suggest that we do not need to make the distinction, since for first order kinetics the MS and MM case yield the same result. Good point; however, while the intrinsic kinetics are first order, the lumped one is not, and there is some room for doubt, isn't there? Let's first consider again the CSTR, which is known to be the MM case with $Tf(t) = \exp[-t/T]$. If we consider a MS reactor with the same RTD, we would calculate U_E as follows (we indicate integrals over t ranging from 0 to ∞ with curly brackets):

$$U_E = \{[1 + t/\alpha]^{-\alpha}\exp(-t/T)/T\} \tag{5.4.1.13}$$

This does not look quite the same as $<\phi(u,\alpha)/(1+uT)>$ does it? However, Ralph turns out to be right, because a representation in powers of T of (5.4.1.13) yields in fact again (5.4.1.11), and hence indeed, at least for the case where $Tf(t) = \exp(-t/T)$ the MS and MM calculations yield the same result. What about the general case? And, even more crucially, we need to ascertain that the MS case is indeed always represented by:

$$U_E = \{[1 + t/\alpha]^{-\alpha} f(t)\} \tag{5.4.1.14}$$

Let's consider the classical representation of a MM reactor as a plug flow one with a distribution of exit streams resulting in the appropriate RTD, see Fig. 5.4.1.1. We use t itself as the axial coordinate, noting that, if $V(t)$ is the local velocity, the geometrical coordinate z is given by $\mathrm{d}z = V\mathrm{d}t$. An overall local mass balance yields $v(t) = -\mathrm{d}V/\mathrm{d}t$ for the side stream distribution of flow-rates per unit cross sectional area. The local balance for component u reads:

$$Vg = vg\mathrm{d}t + (V + \mathrm{d}V)(g + \mathrm{d}g) + ugV\mathrm{d}t \tag{5.4.1.15}$$

This reduces to:

$$\partial g/\partial t = -ug \tag{5.4.1.16}$$

and hence $g(u,t)$ is again given by (5.4.1.5). Now if the feed velocity has been normalized to be unity, $v(t)$ is identified with $f(t)$, and hence we obtain:

$$\begin{aligned} U_E &= <\{\phi(u,\alpha)\exp(-ut)f(t)\}> \\ &= \{<\phi(u,\alpha)\exp(-ut)>f(t)\} \end{aligned} \tag{5.4.1.17}$$

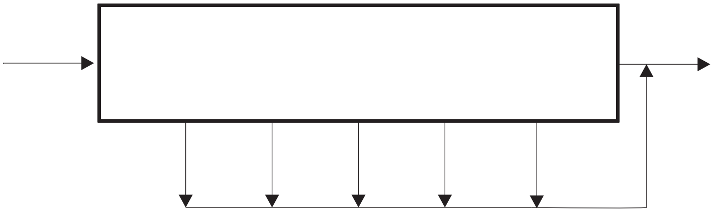

Fig. 5.4.1.1. Flow pattern equivalent to a maximum segregation reactor, and the mass balance on a differential element of it.

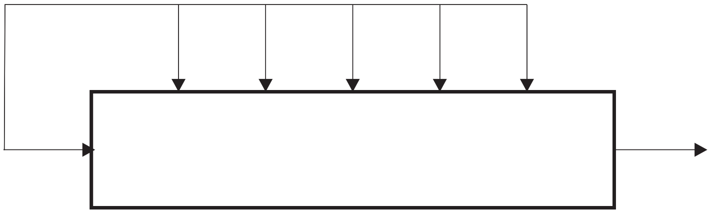

Fig. 5.4.1.2. Flow pattern equivalent to a maximum mixedness reactor, and the mass balance on a differential element of it.

We have now satisfied ourselves that (5.4.1.14) is the correct equation for MS reactors, and we move to the consideration of MM ones (see Fig. 5.4.1.2). This time t has to be understood as the life expectancy, recalling that the distribution of life expectancies in the feed is the same as the distribution of residence times in the product stream [in particular, there is no value of t which reduces $g(u,t)$ to $G(u)$]. This time the overall mass balance yields $v = \mathrm{d}V/\mathrm{d}t$, and the geometrical distance is given by $\mathrm{d}z = -V\mathrm{d}t$ (since life expectancy decreases as the fluid moves forward). The local balance for component u reads:

$$Vg + vG\mathrm{d}t = (V + \mathrm{d}V)(g + \mathrm{d}g) - ugV\mathrm{d}t \tag{5.4.1.18}$$

which reduces to:

$$\partial g/\partial t = ug + v(G - g)/V \tag{5.4.1.19}$$

Now a little thinking shows that $v/V = -f/F$, and hence we obtain:

$$\partial g/\partial t = ug + f(g - G)/F \tag{5.4.1.20}$$

This can be integrated by first defining $\tilde{g} = g\exp(-ut)$, which yields after some algebra:

$$F\partial\tilde{g}/\partial t = [\tilde{g} - G\exp(-ut)]f \tag{5.4.1.21}$$

Recalling that $f = -\mathrm{d}F/\mathrm{d}t$ this transforms to the following directly integrable form:

$$\partial(F\tilde{g})/\partial t = -G\exp(-ut)f \tag{5.4.1.22}$$

When t approaches ∞, $\tilde{g}$ approaches zero (provided g is bounded, which it is guaranteed to be), and hence one obtains, after some algebra:

$$g = G\int_t^\infty e^{-uz} f(z)\mathrm{d}z \tag{5.4.1.23}$$

The concentration distribution in the exit stream is of course $g(u,0) = G(u)\{f(t)\exp(-ut)\}$, and hence one concludes that (5.4.1.14) indeed holds also for the MM case. Ralph is entirely vindicated: since the intrinsic kinetics are linear, there is in fact no difference between the lumped exit concentrations in the MM and MS cases, in spite of the fact that the lumped kinetics are not first order ones.

5.4.2 Nonlinear kinetics

So far we have beaten to death the case where the intrinsic kinetics are linear - but what if they are not? Suppose for instance that, in the case of a single reactant, the kinetic equation may be written as follows (c^*, t^* and k^* are dimensional concentration, time and "kinetic constant"):

$$-\mathrm{d}c^*/\mathrm{d}t^* = k^*r(c^*);\ t^* = 0,\ c^* = c^*_0 \tag{5.4.2.1}$$

How are we going to deal with the case of a continuous mixture? Bob has an idea. He says that (5.4.2.1) can certainly be integrated to yield something like:

$$c^* = B[c^*_0, k^*t] \tag{5.4.2.2}$$

That sounds reasonable; now what about a continuous mixture? Well, in the first order case we had (5.4.1.4), and for a single component we would have $c^* = c^*_0 e^{-k^*t^*}$; hence perhaps this time all we need is to write, after appropriate normalizations along lines which have been cleared up in the previous subsection:

$$g(u,t) = B[G(u), ut] \tag{5.4.2.3}$$

$$U(t) = \langle B[G(u), ut] \rangle \qquad (5.4.2.4)$$

Nice try, Bob. Should we try the SCI check? Bob is convinced it's going to work anyhow, but we proceed to try it nonetheless, and we have some trouble right away. When $G(u) = \delta(u - 1)$, (5.4.2.4) becomes an integral which does not converge unless $B[]$ is linear in its first argument, and that is true only if $r^*()$ is a linear operator, i.e., in the case of first order kinetics. Bob, taking shortcuts didn't quite work this time.

Sue has seen the light. She points out that the concentration is $g\mathrm{d}u$, not g, and that therefore the equivalent of (5.4.2.1) in the mixture should be:

$$-\partial(g\mathrm{d}u)/\partial t = ur^*(g\mathrm{d}u) \qquad (5.4.2.5)$$

This time there is turmoil in the class. Ralph maintains that (5.4.2.5) doesn't mean anything, and he is, as usual, right — it only means something if $r^*()$ is a linear operator, but we want to consider the case where it is not. Bob is entirely flabbergasted, and somewhat cross because he has to agree with Ralph about the lack of significance of (5.4.2.5). Sue says that, if (5.4.2.5) is so patently wrong, then (5.4.2.3) which she suspects is derived from it must be wrong as well, and that is the point she was trying to make. Agreement is finally reached that things should be reformulated from the very beginning, without any attempts at shortcuts, and we choose the following line of approach.

For the single component case, we write the kinetic equation in the following form:

$$-\mathrm{d}c^*/\mathrm{d}t^* = k^*c^*R[K^*c^*] \qquad (5.4.2.6)$$

Several hands are raised, asking what is K^*? Well, some parameter having units of an inverse concentration, so that k^* is an honest-to-God kinetic constant having units of a frequency, and $R[]$ is a dimensionless operator (transforming a dimensionless argument, K^*c^*. into a dimensionless value). We should rather be asking ourselves what the terms k^*c^* and $R[]$ represent physically, keeping in mind that, for first order reactions, $R[]$ would be identically unity.

If the probability of a molecule to undergo the reaction is a constant, then the number of molecules which do undergo it per unit time is simply proportional to how many such molecules there are, and that's the first order case where $-\mathrm{d}c^*/\mathrm{d}t^* = k^*c^*$. So in the case where $R[]$ is not identically unity, its value represents the effect of other molecules being around on the probability of the given molecule to undergo the reaction. Now when we consider a mixture, is it reasonable to assume that R depends *only* on the concentration of the particular component being considered? Isn't it more likely that *all* components influence the probability of one particular reactant to undergo the reaction? Let's give it a try, and let's be as general as possible at this stage. We'll write:

$$-\partial g/\partial t = ug\mathbf{R}[K(v)g(v,t);u] \qquad (5.4.2.7)$$

Nobody knows what we are trying to do; what are $\mathbf{R}[]$, $K()$, and v — apart from the funny looking semicolon in (5.4.2.7)? $\mathbf{R}[]$ is a *functional* of the whole distribution of concentrations, and hence we are using v as a dummy component label; $K()$ is the distribution of dimensionless values of K^*; and we are allowing for $\mathbf{R}[]$ to be different for different components, hence writing, after the semicolon, an explicit dependence on u. After some discussion the points are granted, and this time even Bob asks for a SCI check. Good; if $G(u) = \delta(u - 1)$, (5.4.2.7) guarantees that $g(u,t) = \delta(u - 1)h(t)$, since its solution for u different from unity is simply $g = 0$. So we get:

$$-\mathrm{d}h/\mathrm{d}t = h\mathbf{R}[K(1)\delta(v-1)h(t);1] \qquad (5.4.2.8)$$

The SCI requirement is therefore satisfied provided that:

$$\mathbf{R}[K(1)\delta(v-1)h(t);1] = R[K(1)h(t)] \qquad (5.4.2.9)$$

and in fact this is very easy to satisfy — there are infinitely many functionals $\mathbf{R}[]$ which may satisfy it. This is all very nice, but the hope of getting anywhere with (5.4.2.7) without making some drastic simplifying assumptions is practically zero, and Bob, having realized this, is losing interest quickly.

Well, it turns out that there is an assumption about $\mathbf{R}[]$ which simplifies life very considerably, i.e., that its value *does not* depend on u — no funny looking semicolon. This will be discussed in the next subsection. However, let's keep in mind that it is not a necessarily sound assumption, and that we therefore need to keep in mind the case where it does not hold.

5.4.3 Cooperative kinetics

In the case of first order intrinsic kinetics, it is easy to visualize realistic systems for which the kinetic assumption is reasonable: the case of thermal cracking of a mixture of many components presents itself naturally. It is instructive to consider some specific cases where we expect nonlinear kinetics, and we analyze two such cases in the following.

First consider reactions which take place on a heterogeneous catalyst and are dominated by Langmuir isotherm adsorption. In the case of a single component, we would write the kinetic equation as follows:

$$-\mathrm{d}c^*/\mathrm{d}t^* = k^*c^*/(1 + K^*c^*) \qquad (5.4.3.1)$$

This has the form of (5.4.2.6), with $R[K^*c^*] = 1/(1 + K^*c^*)$. The term K^*c^* in the denominator accounts for the fact that part of the catalyst surface is covered by adsorbed molecules and is not available for reaction. Equation (5.4.3.1) gives an order of reaction which varies smoothly from zero at very high values of c^* to unity at very low values. Now suppose one has a mixture with many components, and before moving to a continuous description let's write down the kinetic equation for component I in a discrete description. Now all reactants compete for adsorption on the catalyst, and we would therefore write:

$$-\mathrm{d}c^*_I/\mathrm{d}t^* = k^*_I c^*_I/(1 + \Sigma K^*_J c^*_J) \qquad (5.4.3.2)$$

Equation (5.4.3.2) shows that the rate of disappearance of an individual component is the product of a linear term ($k^*_I c^*_I$) and a "nonlinearity factor" $1/(1 + \Sigma K^*_J c^*_J)$ *which depends on the whole concentration distribution and is the same for all components* (its value does not depend on I). In the continuous description, this would correspond to the assumption of *uniformity*, i.e., that **R**[] does not depend on u.

Now let's consider the case of oligomerization of olefins. This is clearly, in the case of a single reactant, a case of second order kinetics, at least in its simplest possible description. In a mixture of very many olefins, every one of them may oligomerize with any other one, and hence we would write, in a discrete description (and neglecting the fact that the product of an oligomerization reaction is itself an olefin which may react):

$$-\mathrm{d}c^*_I/\mathrm{d}t^* = c^*_I \Sigma k^*_{IJ} c^*_J \qquad (5.4.3.3)$$

If there are very many olefins, knowing the values of all bimolecular kinetic constants k^*_{IJ} would be out of the question, and perhaps the best one could do would be to assume that k^*_{IJ} is somehow expressible as the product of the intrinsic reactivities of the two olefins, say something like $(k^*_{II}k^*_{JJ})^{1/2}$. With this assumption, (5.4.3.3) becomes again the product of a term which is linear in c^*_I and a nonlinearity factor which depends on the whole concentration distribution and is the same for all components. The two examples discussed above give some physical support to the assumption of uniformity for the functional $\mathbf{R}[]$. So we write down the most general form of cooperative uniform kinetics as follows:

$$-\partial g(u,t)/\partial t = \theta(u)g(u,t)\mathbf{R}[K(v)g(v,t)] \tag{5.4.3.4}$$

$\theta(u)$ in (5.4.3.4) is a distribution of dimensionless kinetic constants, which has been normalized so that its initial average value is unity, $<\theta(u)G(u)> = 1$.

The important point is that the solution of (5.4.3.4) is easily found by defining a *warped time* $w(t)$, with $w(0) = 0$. One has:

$$g(u,t) = G(u)\exp[-\theta(u)w(t)] \tag{5.4.3.5}$$

$$\mathrm{d}w/\mathrm{d}t = \mathbf{R}[\exp[-\theta(u)w(t)]K(v)G(v)] \tag{5.4.3.6}$$

Equation (5.4.3.6) is an ordinary differential equation for the warped time, and its solution can always be found for any given problem.

We may however obtain much more powerful results without great loss of generality. First of all, under mild assumptions of well-behavedness u can be rescaled so that $\theta(u) = u$, i.e., components are labelled in order of increasing kinetic constant. Furthermore, for any practical case where the single component kinetic equation has been written in the form (5.4.2.6), the functional $\mathbf{R}[]$ can be written as:

$$\begin{aligned}\mathbf{R}[] &= R[<K(v)g(v,t)>]\\ &= R[<K(v)G(v)\exp[-uw(t)]>] = R[Q(w)]\end{aligned} \tag{5.4.3.7}$$

Equation (5.4.3.7) is guaranteed to satisfy the SCI requirement; furthermore, for any given concrete problem, the form of the $R[]$ operator would be known from our understanding of the underlying chemistry. In order to proceed any further, one needs to assign the distributions $G(u)$ and $K(u)$, and we consider the following ones:

$$G(u) = \phi(u, \alpha), \alpha > 0 \tag{5.4.3.8}$$

$$K(u) = \widetilde{K}(\alpha u)^{\beta}\Gamma(\alpha)/\Gamma(\alpha + \beta), \quad \alpha + \beta > 0 \tag{5.4.3.9}$$

Notice that the condition $\alpha + \beta > 0$ does not forbid β to be negative: hence (5.4.3.9) may describe both the case where K decreases with u (components with a high kinetic constant have a low normalization factor), and *vice versa*. $\widetilde{K}$ is simply the initial average value of $K(u)$, $\widetilde{K} = <K(u)G(u)>$, which cannot be normalized to be unity except in the case where $R[]$ is of a power law form.

Now one can calculate $U(t)$ by integrating (5.4.3.5) to obtain:

$$U = [\alpha/(\alpha + w)]^{\alpha} \tag{5.4.3.10}$$

The argument of $R[]$ in (5.4.3.7), $Q(w)$, which is the average value of $K()$ at w, is obtained as:

$$Q(w) = \widetilde{K}[\alpha/(\alpha + w)]^{\alpha+\beta} = U^{(\alpha+\beta)/\alpha} \tag{5.4.3.11}$$

Finally, differentiating (5.4.3.5) one obtains:

$$-\mathrm{d}U/\mathrm{d}t = [\alpha/(\alpha + w)]^{\alpha+1}\mathrm{d}w/\mathrm{d}t \tag{5.4.3.12}$$

The interesting point is that w may be eliminated from these equations to obtain a lumped kinetic function:

$$-\mathrm{d}U/\mathrm{d}t = U^{(\alpha+1)/\alpha}R[\widetilde{K}U^{(\alpha+\beta)/\alpha}];\ U(0) = 1 \tag{5.4.3.13}$$

which will yield in general an explicit solution for the lumped concentration:

$$U = Z(t) \tag{5.4.3.14}$$

Now let's look at our two specific examples. For olefin oligomerization, one would have $R[Q] = Q$, and hence $\widetilde{K}$ can be normalized to be unity; furthermore, since $k_{IJ} = k_{JI}$, one would have $\beta = 1$. Thus one obtains:

$$-\mathrm{d}U/\mathrm{d}t = U^{2(\alpha+1)/\alpha} \tag{5.4.3.15}$$

i.e., the apparent overall order would be $(\alpha + 1)/\alpha$ times the underlying intrinsic order of two. For the Langmuir isotherm case, $R[Q] = 1/(1 + Q)$, and one obtains:

$$-\mathrm{d}U/\mathrm{d}t = U^{(\alpha+1)/\alpha}/(1 + \widetilde{K}U^{(\alpha+\beta)/\alpha}) \tag{5.4.3.16}$$

This is particularly interesting, since, if $\beta > 1$ (which it might very well be) the apparent overall order of reaction may become negative at large U's, in spite of the fact that the intrinsic kinetics are always of positive order. Indeed, for intrinsic n-order kinetics with $n > 0$ (a possibly not very realistic case), $R[Q] = Q^{n-1}$, $\widetilde{K} = 1$, and one obtains:

$$-\mathrm{d}U/\mathrm{d}t = U^{[1+n\alpha+(n-1)\beta)]/\alpha} \tag{5.4.3.17}$$

and therefore the apparent lumped order is only restricted to be $>1 - \beta$ and might thus very well be negative.

Now let's move to consideration of a CSTR reactor, with $G(u)$ and $g(u)$ defined as in Section 5.4.1. The mass balance on component u is:

$$G(u) - g(u) = Tug(u)R[<K(u)g(u)>] \tag{5.4.3.18}$$

This suggests very strongly that the actual dimensionless residence time T has been warped to a value W given by:

$$W = TR[<K(u)g(u)>] \tag{5.4.3.19}$$

which yields trivially the solution of (5.4.3.18):

$$g(u) = G(u)/(1 + uW) \tag{5.4.3.20}$$

The value of W is found by substituting (5.4.3.20) into (5.4.3.19): the solution can be found for any given special case, but not in general form: again the CSTR turns out to be harder than the PFR. An interesting problem arises if one considers the Langmuir isotherm case with $\beta > 1$ and a sufficiently large value of $\widetilde{K}$. Should (5.4.3.13) be a true kinetic equation, one could have the possibility of multiple steady states, since the lumped mass balance would be $1 - U = TH(U)$, and hence a straight line of slope $1/T$ could have three intercepts with $H(U)$. However, in actual fact multiple steady states are impossible — because the intrinsic kinetics are never of negative order. Yes, Ralph, that's not proof enough, the proof is rather involved and we'll skip it.

Now if one considers an MS reactor, one would again calculate, with $Z(t)$ the solution of Eq. (5.4.3.13):

$$U_E = \{Z(t)f(t)\} \tag{5.4.3.21}$$

This time, however, with $Tf(t) = \exp[-t/T]$ the result is most definitely *not* $U_E = <g(u)>$ with $g(u)$ given by (5.4.3.20), except in the special case where $R[]$ is identically unity.

This seems to suggest that we might be in a bit of a trouble when dealing with MM reactors. Well, let's give it a try anyhow. We look back at Fig. 5.4.1.2 and we quickly come to the conclusion that the label u only plays the role of a parameter in the component balance, which therefore becomes:

$$\frac{\partial g}{\partial t} = ugR[Q(t)] + \frac{f(t)}{F(t)}(g - G) \tag{5.4.3.22}$$

$$Q(t) = <K(u)g(u,t)> \tag{5.4.3.23}$$

The idea of warping time has served us well so far, and we'll give it another chance. Let again $w(t)$, with $w(0) = 0$, be a warped life expectancy variable, delivered by the solution of:

$$\mathrm{d}w/\mathrm{d}t = R[Q(t)] = 0 \tag{5.4.3.24}$$

with $\mathrm{d}w/\mathrm{d}t$ strictly positive if we accept the usual theory of equilibrium, i.e., that the kinetics are linear in a neighbourhood of the equilibrium value.

Furthermore, we define $\tilde{g}(u,t) = g(u,t)\exp[-uw(t)]$, which is guaranteed to approach zero as t approaches ∞, since g is bounded. One only needs to duplicate the algebra done in Section 5.4.1 to obtain:

$$\frac{g(u,t)}{G(u)} = \frac{\exp[uw(t)]}{F(t)} \int_t^\infty f(z)e^{-uw(z)}\,\mathrm{d}z \tag{5.4.3.25}$$

One can now easily calculate the concentration distribution and the lumped concentration at the exit:

$$g(u,0) = \{f(t)G(u)\exp[-uw(t)]\} \tag{5.4.3.26}$$

$$U_E = \langle g(u,0)\rangle = \{f(z)\langle G(u)\exp[-uw(t)]\rangle\} \tag{5.4.3.27}$$

This looks deceptively simple, and in fact it looks just like the result for a MS reactor. However, this is not so, because the warped time function for a MM reactor is different from the one for a MS reactor. The former is delivered by the solution of (5.4.3.24), which, after substitution of (5.4.3.23) and (5.4.3.25) yields:

$$\frac{\mathrm{d}w}{\mathrm{d}t} = R\left[\left\langle K(u)G(u)\int_t^\infty f(z)e^{-uw/z}\,\mathrm{d}z\right\rangle\right] \tag{5.4.3.28}$$

The difficulty with (5.4.3.28) is that the integral within the <> one is a functional of the unknown function $w(t)$, and therefore the solution of (5.4.3.28) is very hard to find.

An important point is to establish what $g(u,t)$ does as t approaches ∞ (what is the concentration distribution in those material elements which have an infinitely large life expectancy?). We have obtained (5.4.3.25) by imposing $\tilde{g}(u,\infty) = 0$, and this is correct provided $g(u,\infty)$ is bounded, which it certainly is since $g(u,\infty) \le G(u)$. As t approaches ∞, $g(u,t)$ may approach some asymptotic value $g(u,\infty)$, in which case $\partial g/\partial u$ approaches zero; but it may also oscillate. If one takes the limit of (5.4.3.25) for t approaching ∞, one obtains after some algebra:

$$\lim_{t=\infty} g(u,t)/G(u) = 1\Big/\left[1 + uT^*\lim_{t=\infty} w'(t)\right] \tag{5.4.3.29}$$

$$T^* = \lim_{t=\infty} F(t)/f(t) \tag{5.4.3.30}$$

The derivative $w'(t)$ is guaranteed to have some finite value as t approaches ∞, and hence the limit in (5.4.3.29) exists provided the one in (5.4.3.30) does, i.e. provided T^* exists. The existence of T^* is a property only of the RTD, and it is easy to convince oneself that there are RTDs which do not have that limit: for instance, a PFR with recycle has a RTD given by a sum of infinitely many delta functions, and hence T^* does not exist. The existence, or lack thereof, of the relevant limits has to do with the question of the possible multiplicity of steady states, but we do not discuss this here.

PROBLEMS

5.1 Consider the two classical problems of sequential $A \to B \to C$ and parallel $C \leftarrow A \leftrightarrow B$ reactions. Now suppose one has a mixture of very many components, initially containing only $A(u)$-type components, and let both steps in both cases be governed by first order kinetics. Develop the equations which describe the continuous mixture, and discuss your result in detail.

5.2 Consider the case of a continuous ideal mixture, for which the free enthalpy of mixing is given by (5.2.26). We are interested in the derivatives of the functional $<x(u)\ln x(u)>$.

(a) Calculate the first and second functional derivative.

(b) Show that, as $x(u)$ approaches $\delta(u - u^*)$, where u^* is the solvent label, the second functional derivative approaches infinity, and draw conclusions from this result.

SUGGESTED READING

The idea of a continuous description of multicomponent mixtures goes back to De Donder (1931). The paper by Gavalas and Aris (1966) is a classical reference on the application of linear vector algebra to multicomponent mixtures. The papers by Amundson and Acrivos (1955) Briano and Glandt (1984) and Shibata et al. (1987) are useful for the analysis of phase equilibria in multicomponent mixtures. The Shibata et al. paper discusses the problem of finding the appropriate quadrature points and hence the best choice of pseudocomponents. The continuum approach to homogeneous chemical equilibria is best illustrated by Alberty and Oppenheimer (1984), and the case of simultaneous chemical and phase equilibria (i.e., heterogeneous chemical equilibria) is discussed by Astarita and Ocone (1989). Linear kinetics in continuous mixtures have been solved in 1968 by Aris; the paper by Krambeck

(1984) is a very useful one. Nonlinear kinetics are discussed by Ho and Aris (1987) and by Aris (1989). Cooperative kinetics have been introduced by Astarita and Ocone (1988), and extended by Astarita (1989) and by Astarita and Nigam (1989). A different approach to nonlinear kinetics has been presented by Chou and Ho (1988). A review on continuous kinetics is available (Aris and Astarita 1989). The original conception of a maximum mixedness reactor is due to Zwietering (1959); its application to continuous mixtures is discussed by Astarita and Ocone (1990), and the problem of multiplicity of steady states by Astarita (1990). A book by Astarita and Sandler (1991) is a review of the whole area of thermodynamics and kinetics in continuous mixtures.

REFERENCES

Alberty, R.A. and L.A. Oppenheimer, J. Chem. Phys., 81 (1984) 4603.
Amundson, N.R. and A. Acrivos, Chem. Eng. Sci., 4 (1955) 29, 68, 141, 159, 206, 249.
Aris, R., Arch. Ratl. Mech. Anal., 27 (1968)356.
Aris, R., AIChEJ, 35 (1989) 539.
Aris, R. and G. Astarita, Chem. Eng. Process., 26 (1989) 63.
Astarita, G., AIChEJ, 35 (1989) 529.
Astarita, G., Chem. Eng. Comm., 93 (1990) 111.
Astarita, G. and A. Nigam, AIChEJ, 35 (1989) 1927.
Astarita, G. and R. Ocone, AIChEJ, 34 (1988) 1299.
Astarita, G. and R. Ocone, Chem. Eng. Sci., 44 (1989) 2323.
Astarita, G. and R. Ocone, Chem. Eng. Sci., 45 (1990) 3399.
Astarita, G. and S.I. Sandler (Eds.), Kinetic and Thermodynamic Lumping in Multicomponent Mixtures. Elsevier, Amsterdam, 1991.
Briano, J.G. and E.C. Glandt, J. Chem. Phys., 80 (1984) 3336.
Chou, M.Y. and T.C. Ho, AIChEJ, 34 (1988) 1519.
De Donder, Th., L'Affinité. Seconde Partie. Gauthiers-Villars, Paris, 1931.
Gavalas, G.R. and R. Aris, Proc. Roy. Soc. London, Ser. A, 260 (1966) 351.
Krambeck, F.J., Inst. Chem. Engrs. Symp Ser., Proc. ISCRE, 8 (1984) 152.
Ho, T.C. and R. Aris, AIChEJ, 33 (1987) 1050.
Shibata, S.K., S.I. Sandler and R.A. Behrens, Chem. Eng. Sci., 42 (1987) 1977.
Zwietering, Th.N., Chem. Eng. Sci., 11 (1959) 1.

Chapter 6

Granular Materials

6.1 INTRODUCTION

Granular systems consist of a large number of solid particles usually dispersed in a fluid. Granular materials are two-phase systems; however, in many cases, the interstitial fluid does not play a major role, and most existing theories apply to "dry" granular systems. In this chapter, compressible flow of granular materials is analyzed only for dry systems. The philosophical concepts behind the derivation of the large-scale statistical thermodynamics are generally applicable to both dry and wet systems. In writing balances and constitutive equations, the implicit assumption is that the interstitial fluid (gas) does not play a major role. (Indeed, we assume that there is no influx of fluctuating energy due to the action of the fluid, and we assume that constitutive equations do not depend upon the fluid either.) In Section 6.3.2 the influence of the fluid is considered, and consequently a thermodynamic theory is presented, which is however still at its exordium.

One of the approaches to the analysis of the flow of granular systems is based on continuum modelling, i.e. the assembly of particles is regarded as an equivalent continuum phase. Such a model is of necessity based on an evaluation of local average values in the equivalent continuum: to understand this, one simply needs to realize that even the concept of a local solid volume fraction ν is defined only in an average sense. Let d be the average particle diameter. Local averages (indicated with triangular brackets in the following) are intended as averaged over a volume of linear dimension l, with $l >> d$. Since ν is regarded as a point value into the appropriate equations for the equivalent continuum, the scale of description of the entire system, L, needs to be well in excess of l, $L >> l$ [for instance, in packed bed flow, $L/d >> 30$ (Beavers et al., 1973; Zou and Yu, 1995)].

The question of how to correctly define averages is well discussed in the literature (Bowen, 1971; Drew, 1983; Joseph et al., 1990; Lundgren, 1972); the requirement for any appropriate averaging procedure is that the exact solution of the averaged equations should coincide with the average solution of the exact equations.

In the following we will regard the granular phase as a compressible one, and we will focus our attention on the analysis of the flow of granular systems under conditions where the compressibility of the granular phase is of relevance. Indeed, the local average density of the granular phase is $\rho = \pi_p \nu$, ρ_p being the particle density, which is a constant. ν may very well not be constant, therefore it follows that the granular phase is compressible.

6.2 LARGE-SCALE STATISTICAL THERMODYNAMICS

Compressible flow is a thoroughly studied subject in gas dynamics. Classical compressible flow theory is rooted in a firm background of thermodynamics, and several analogies can be used to develop the theory of compressible flow of granular materials. In classical statistical thermodynamics, averages are taken over a "local" scale which is required to be large with respect to the molecular mean free path. For granular materials, a large-scale statistical theory of thermodynamics is needed, based on averages at the l-scale, where solid particles play the role that molecules play in the Maxwellian theory. In particular, large-scale or granular analogues of classical thermodynamic quantities such as temperature, entropy, etc. need to be available.

The concept of granular temperature, which was first explicitly introduced by Ogawa (1978), but can in fact be traced back to the work of Bagnold (1954), has been widely used in the literature on granular systems. Let $\mathbf{v}$ be the instantaneous velocity of a particle. The local average velocity of the particulate phase, $\mathbf{u}$, is:

$$\mathbf{u} = <\mathbf{v}> \tag{6.2.1}$$

Let $\mathbf{c} = \mathbf{v} - \mathbf{u}$ be the instantaneous fluctuating velocity of the particle with respect to the local average velocity. Thus, $\mathbf{c}$ has by definition zero average, $<\mathbf{c}> = 0$. Since $<\mathbf{c} \cdot \mathbf{u} = 0$, the average total kinetic energy per unit mass of the granular phase, K_{TOT}, is:

$$K_{\mathrm{TOT}} = <v^2/2> = <u^2/2> + <c^2/2> \tag{6.2.2}$$

$<u^2/2>$ is the kinetic energy associated with the main motion of the granular phase, denoted by K. In analogy with the classical Maxwell (1967) [but the credit should really go to Waterston (1845)] statistical theory of rarefied gases, $<c^2/2>$ has been identified (to within a numerical factor) with the "granular temperature", T, so that:

$$K_{\mathrm{TOT}} = K + T \tag{6.2.3}$$

Temperature is of course a thermodynamic concept, and therefore, the introduction of the concept of granular temperature can be considered as the first step towards the development of a large-scale thermodynamic theory of granular systems.

Indeed, Jenkins and Savage (1983) and Haff (1983), almost simultaneously, introduced the idea of a "balance of granular temperature". A subtle conceptual point arises in this regard. In the Maxwellian theory of rarefied gases, the internal energy per unit mass, U, is proportional to temperature, the proportionality constant (the constant volume specific heat) being a dimensional universal constant. In classical gas dynamics, U and T do not have the same dimensions; if they had been assigned the same dimensions, the proportionality constant could easily have been set to unity by normalisation, and the distinction between U and T would have become blurred. This is what has happened in the literature on granular systems, where T has not been assigned its own dimensions, and hence U and T have essentially been identified with one another as if they were the same concept. This is strictly true when the stress tensor (and all other quantities such as entropy, internal energy, etc.) of the granular material is not influenced by the interstitial fluid (the former situation will be identified in the next section as the "inertial" limit). It is therefore important to keep the conceptual distinction between U and T in mind, even if the two quantities are assigned the same dimensions and, in the inertial limit, the same value. The "balance of granular temperature" is then more clearly recognized as a balance of "granular energy", which clearly identifies it with the first law of large-scale thermodynamics.

Let σ be the particulate phase stress tensor. Regarding σ as the momentum flux tensor, as is usual in fluid mechanics, forces the convention that compressive stresses are positive. Let $\mathbf{D}$ be the rate of strain tensor of the granular phase:

$$\mathbf{D} = [\mathrm{grad}\mathbf{u} + \mathrm{grad}\mathbf{u}^{\mathrm{T}}]/2 \tag{6.2.4}$$

The net rate of work per unit volume done by internal stresses in the particulate phase (i.e., that part of the work which does not result in an increase of the kinetic energy of the main motion), w, is:

$$w = -\sigma : \mathbf{D} \tag{6.2.5}$$

Let q be the local average net influx of granular energy per unit volume. A non-zero q may exist for two different reasons. First, there might be a direct influx Q due, e.g., to vibration of the container; since one is trying to develop a theory of granular thermodynamics, it is useful to recognize Q as the analog of the radiant heat influx in classical thermodynamics. Second, there may well be a conductive flux of granular energy, $\mathbf{q}$, related to the possibility that granular energy may tend to move in the direction of decreasing granular temperature (energy). Hence one has:

$$q = Q - \mathrm{div}\mathbf{q} \tag{6.2.6}$$

Finally, let I be the rate of dissipation of granular energy, per unit volume, due to inelasticity of particle–particle collisions. In the Maxwellian theory, collisions between molecules are assumed to be perfectly elastic; however, such an assumption cannot be extended to granular systems.

Remark 6.2.1

> In Eq. (6.2.6) only two contributions for the net influx of granular energy are taken into account. Of course, one can think of many other contributions; for instance, if an interstitial (viscous) fluid is considered, its action can result in damping or enhancing granular energy (Ocone, 1992).

The first law of large-scale thermodynamics (or the balance of granular energy) is written as (D/Dt is the substantial derivative operator; in the classical inertial theories, $U = T$):

$$\rho \mathrm{D}U/\mathrm{D}t + I = w + q \tag{6.2.7}$$

There is no analog of the term I in classical thermodynamics and this immediately shows that granular thermodynamics is of necessity a non-equilibrium theory. In the absence of a net influx of granular energy ($w + q = 0$), the (positive) term I implies that DU/Dt is negative, i.e., that granular energy steadily decreases. A granular system cannot stay thermalized without a continuous influx of energy, and hence its only "equilibrium" state is one where $T = 0$. Notice also that, since the only "equilibrium" state is $T = 0$, no argument of the "quasi-static process" kind is available —any non-trivial state of a granular system is not in any sense close to an equilibrium state. Another consequence of the presence of the term I is that in general granular energy is not equipartitioned (Maddox, 1995; Ocone and Astarita, 1995a).

In order to close the system of relevant equations, constitutive equations need to be written down for the particulate phase constitutive variables (or functions of state) U, σ, $\mathbf{q}$, I, and for the interaction force between the interstitial fluid and particulate phases (Jackson, 1986). These have been discussed in the literature, and for the aim of the present discussion their detailed form is not needed here. In order to develop a large-scale thermodynamic theory, one only needs to establish the constitutive class, i.e., the list of independent state variables which determine the values of the constitutive properties (Truesdell, 1984). The list, as deduced from the relevant literature, includes the solid volume fraction ν, the granular temperature T, its gradient $\mathbf{n} = \text{grad}T$, the rate of strain tensor $\mathbf{D}$. Some of the published theories include gradν in the list, but the thermodynamic analysis (Ocone and Astarita, 1993) has shown that internal consistency excludes gradν from the list.

Having identified the constitutive class, one may proceed to write down the cornerstone of large-scale statistical thermodynamics, i.e. the second law. Physically, the second law of granular thermodynamics aims at describing the "irreversibility" of the transfer of mechanical energy to the level of particle fluctuations: when it is transferred, although still present in the form of kinetic energy, it is no longer entirely retrievable. In developing the formulation of the second law of granular thermodynamics, Ocone and Astarita (1993) have followed the philosophical approach of Truesdell (1984) and of Coleman and Noll (1963). This approach is consistent with the fact that granular thermodynamics is a non-equilibrium theory where no analog of the vague idea of "quasi-static process" can be used. Hence granular entropy S (a concept without which no compressible flow theory can be developed) is regarded as a primitive concept which (a) cannot be defined in terms of more fundamental concepts, and (b) is only determined by the axioms laid down for it. The first axiom is that granular entropy is a function of state, and hence, by the principle of equipresence (Truesdell; 1984), a constitutive function $s()$ exists which delivers the value of S:

$$S = s(\nu, T, \mathbf{n}, \mathbf{D}) \tag{6.2.8}$$

Remark 6.2.2

> If the interstitial fluid is a viscous one, entropy depends upon the fluid viscosity as well, as will be discussed in Section 6.3. The complete theory of fluid viscosity dominated granular flow rheology can be found in Ocone and Astarita (1995b).

The second axiom is of course the second law of large-scale thermodynamics, and here a subtle point arises in connection with the inelasticity of collisions. First, consider the case where there is no net influx of entropy in the neighbourhood of the point considered. In classical thermodynamics, one would for this case write the second law by requiring $\mathrm{D}S/\mathrm{D}t$ to be non-negative. Here however the term I provides a new source of irreversibility, and hence the approach of Ocone and Astarita (1993) is equivalent to writing down the second law for this case as:

$$\rho\,\mathrm{D}S/\mathrm{D}T + I/T \geq 0 \tag{6.2.9}$$

i.e., for the case at hand granular entropy may increase, but not faster than $I/\rho T$.

For the general case, the net rate of influx of entropy per unit volume, J, is non- zero, and thus the second law takes the form:

$$\rho\mathrm{D}S/\mathrm{D}t + I/T \geq J \tag{6.2.10}$$

By analogy with the classical theory, J is assumed to be made up of two parts. First, there may be a direct entropy influx per unit volume Q/T; second, there might be an imbalance of the conductive entropy flux q/T:

$$J = Q/T - \mathrm{div}(\mathbf{q}/T) = q/T + \mathbf{q}\cdot\mathbf{n}/T^2 \tag{6.2.11}$$

Let $Z = -\,\mathbf{q}\cdot\mathbf{n}/T$ be the “thermal dissipation rate”, i.e., the rate of dissipation of granular energy due to its flowing towards regions of lower granular temperature. By substitution and rearrangement one obtains the final expression of the second law in its general form involving granular entropy:

$$\rho T\mathrm{D}S/\mathrm{D}t + I + Z \geq \mathrm{q} \tag{6.2.12}$$

Equation (6.2.12) can be considered as the fundamental axiom of our formulation of “pseudo-thermodynamics”: it might be called the “pseudo-Clausius–Duhem” inequality.

Since the concepts of granular (pseudo-) energy *U*, the granular temperature *T*, and the granular entropy *S* are available, we do not need to introduce any new concept in defining the (Helmholtz) pseudo-free energy *A* as:

$$A = U - TS \tag{6.2.13}$$

One may now eliminate *q* between Eqs. (6.2.7) and (6.2.13) (the first and the second laws) and re-arrange to obtain a general formulation of (the local form of) the second law of pseudo-thermodynamics:

$$\rho[\mathrm{D}A/\mathrm{D}t + S\mathrm{D}T/\mathrm{D}t] \leq w + Z \tag{6.2.14}$$

It is noteworthy that *I* does not appear in Eq. (6.2.14) because it has been included in Eq. (6.2.12) (and has therefore been eliminated between the first and the second law). The fact that the term *I* does not appear in Eq. (6.2.14) shows that the latter is a description of the irreversibility as observed at the level of pseudo-energy, i.e. independent of the fact that pseudo-energy may be dissipated all the way down to true thermal energy.

6.3 CONSTITUTIVE PROPERTIES

Theories of granular flow are essentially constitutive models of the mechanical behaviour of two-phase systems consisting of a fluid phase and a collection of solid particles. Balance equations alone cannot be used to solve problems of practical interest, and constitutive equations need to be written for the fluid and solid phase as well as for the interaction between the two phases.

In most of the available classical theories, the interstitial fluid is regarded as being “almost” inviscid, and the stress in the particulate phase is not influenced by any property of the interstitial fluid. This limit has been referred to as “inertial” by Jenkins and McTigue (1990, 1992) and Ocone and Astarita (1995b). Conversely, the stress tensor of the particulate phase may be dominated by stresses of the type described in lubrication theory (Ocone and Astarita, 1995b). Such kind of stresses arises from the squeeze flow of the interstitial fluid between two close particles in relative motion toward each other, and the corresponding flow is referred to as “viscous” (“viscous” limit).

6.3.1 Inertial limit

In this section we consider some of the results obtained for the inertial limit, and we therefore regard the interstitial fluid (gas) as being inviscid. It is noteworthy that most of the literature on granular flow deals (implicitly) with the inertial limit; indeed (as pointed out also in Remarks 6.2.1 and 6.2.2), the results in Section 6.2 are obtained assuming that the interstitial fluid does not play a significant role.

Before presenting part of the results for the inertial limit obtained by Ocone and Astarita (1993), it is useful to decompose the stress tensor, σ, into the sum of its isotropic and deviatoric part (**1** is the identity tensor):

$$\sigma = p^* \mathbf{1} + \sigma^{\mathrm{D}} \tag{6.3.1.1}$$

The isotropic part of the stress tensor σ is $p^*\mathbf{1}$, where p^* is the particulate phase "pressure". p^* can be expressed as the sum of a term p which does not depend on the rate of compression of the particulate phase, and another term which may depend on it. The part p of the particulate pressure only does reversible work, and it is therefore not surprising that analogues of classical reversible thermodynamics results can be obtained for p (Ocone and Astarita, 1993).

To obtain the constitutive relations needed, we will follow here "scaling" arguments — or simply we will apply dimensional analysis in an intelligent way. The independent variables, which may influence the problem, are the particle density ρ_{p}, the granular temperature T, the fluid viscosity μ_{f}, the solid volume fraction ν, and the particle diameter d. We then assume that all the parameters appearing in the constitutive equations for the particulate phase depend on the above list. Application of the Buckingham π theorem leads to two dimensionless groups: a possible choice is the volume fraction ν, and the group $R = \rho_{\mathrm{p}} \mathrm{d}\sqrt{T} / \mu_{\mathrm{f}}$. We will comment on the latter later on.

Again making use of a simple-minded dimensional argument, an expression for the conservative pressure p is easily obtained (Ocone and Astarita, 1995b):

$$p = f(\nu)\rho_{\mathrm{p}} T \tag{6.3.1.2}$$

where $f(\nu)$ is a dimensionless function which does not need to be specified for the present purposes. Notice that both Lun et al. (1983) and Haff (1983) derive for p expressions in accordance with Eq. (6.3.1.2).

Next, a constitutive expression for the fluctuation energy U is needed. In the inertial limit, the fluctuation energy U can be identified with the granular temperature T. Apart from the actual equality, the important point is that U is taken to be a unique function of T, and hence the particulate phase is identified as one endowed with entropic elasticity (Ocone and Astarita 1993). Subsequently, p is required to be proportional to T, as indeed it is in the inertial theories. The constitutive assumption made for U makes it possible to derive an appropriate pseudo-thermodynamic theory (all the results in Section 6.2 are derived using the constitutive assumption $U = T$, and therefore they are valid in the "inertial" limit).

6.3.1.1 Inertial limit: how to calculate entropy

As pointed out in Section 6.2, the introduction of the concept of temperature is the first step towards formulating a thermodynamic theory. First and second laws can be deducted by simply writing balances (or inequalities), after constitutive equations for U and the equilibrium collisional pressure p are written down. Let these have the following general form, which is typical of granular flow models:

$$U = m(\nu)\, T \qquad (6.3.1.1.1)$$

$$p = p(\nu, T) \qquad (6.3.1.1.2)$$

[Notice that m is a dimensionless function of the volume fraction and Eq. (6.3.1.1.2) is a general form of Eq. (6.3.1.2).] We assume (but this can in fact be proven — see Ocone and Astarita, 1993) that perfect analogy with classical thermodynamics exists, and then the pseudo-entropy equals minus the partial derivatives of pseudo-free energy with respect to pseudo-temperature:

$$S = -\partial A/\partial T \qquad (6.3.1.1.3)$$

Substitution of the definition of A [Eq. (6.2.13)] into the above equation gives:

$$\partial A/\partial T = \partial U/\partial T - S - T\partial S/\partial T \qquad (6.3.1.1.4)$$

Given Eq. (6.3.1.1.3), Eq. (6.3.1.1.4) can also be written as:

$$\partial U/\partial T - S - T\partial S/\partial T = -S \qquad (6.3.1.1.5)$$

And finally:

$$\partial S/\partial T = \partial U/\partial T/T = m(\nu)/T \tag{6.3.1.1.6}$$

Let S_0 and A_0 be the pseudo-entropy and pseudo-free energy at some reference state ν_0, T_0. One can integrate Eq (6.3.1.1.6) to obtain:

$$S(\nu_0,T) = S_0 + m(\nu_0)\ln(T/T_0) \tag{6.3.1.1.7}$$

so that pseudo-entropy is known at any pseudo-temperature once a reference state can be identified.

6.3.2 Viscous limit

The analogues of Eq. (6.3.1.2) can be obtained again using a scaling argument (Jenkins and McTigue, 1990; Ocone and Astarita, 1995b):

$$p = g(\nu)\mu_\mathrm{f}\sqrt{T/d} \tag{6.3.2.1}$$

where μ_f is the viscosity of the interstitial fluid and $g(\nu)$ is a dimensionless (unspecified) function of the solid volume fraction.

Remark 6.3.2.1

> It is natural to ask oneself what kind of equation for p (and for other functions of state) should be used in general, i.e., when both μ_f and ρ appear in the list of independent variables. Such a general equation for p should of course reduce to either Eq. (6.3.1.2) or Eq. (6.3.2.1) in the appropriate limits. When both μ_f and ρ are included in the list of independent variables, one can construct, as Jenkins and McTigue (1990) have shown, a dimensionless group R which is very vaguely reminiscent of a Reynolds number:
>
> $$R = \rho_\mathrm{p} \mathrm{d}\sqrt{T/\mu_\mathrm{f}} \tag{6.3.2.2}$$
>
> and a simple dimensional argument would yield $p = \rho_\mathrm{p} T F(\nu, R)$.
>
> The function $F(\nu, R)$ is totally arbitrary from the viewpoint of dimensional analysis. From a physical viewpoint, however, is subject to the constraint

that it should degenerate into the appropriate limits. The viscous limit corresponds to $R << 1$, and hence $F(\nu, R)$ should in this limit approach $g(\nu)/R$:

$$p = \rho_p g(\nu)T/R \tag{6.3.2.3}$$

Remark 6.3.2.2

The inertial limit obviously includes the case where $\mu_f = 0$, and hence it is expected to hold when R approaches ∞: hence we require that $F(\nu,\infty) = f(\nu)$.

As seen in the previous paragraph, in classical inertial theory, U is identified (to within a numerical factor) with pseudo-temperature itself, $U = T$. Apart from the actual equality, the important point is that in the inertial theory U is taken to be a unique function of T. Hence, the particulate phase is identified as one endowed with entropic elasticity (Ocone and Astarita 1993), so that p is required to be proportional to T. Now in Eq. (6.3.2.1) p *is not* proportional to T, and hence by an inverse implication, U is not a unique function of T (Astarita and Sarti, 1975). It follows that one has to distinguish between the average square velocity of particle oscillation, T, and the energy per unit mass associated with particle fluctuation, U. It is this (necessary) distinction between T and U which gives to the pseudo-thermodynamic theory in the general case much more conceptual body than one has in the inertial limit.

Now that one has to exclude the case in which U is a unique function of T, one must decide what to do — and it is here that a conceptual problem arises. Given that T is the average square velocity of particle oscillation about their mean motion, what is U? In what sense can the fluctuation energy per unit mass be anything but the average square velocity of oscillation? We refrain from trying to answer these conceptually bothering questions, and we simply choose to write:

$$U = G(\nu)T/R \tag{6.3.2.4}$$

where G is a dimensionless function.

We now want to proceed to establishing an appropriate relation for the entropy in the viscous limit. We consider the conceptually fundamental part of Ocone and Astarita (1993) analysis, which is based on writing the appropriate analogues of the Clausius–Duhem inequality (Truesdell, 1985). One can reproduce the same kind of analysis presented in Section 6.3.1:

$$\partial S/\partial T = \partial U/\partial T/T = G(\nu)\kappa/(2T\sqrt{T}) \qquad (6.3.2.5)$$

and hence:

$$\partial/\partial T(\partial S\,/\partial\nu) = G'(\nu)\kappa/(2T\sqrt{T}) \qquad (6.3.2.6)$$

with $G'(\nu) = \partial G/\partial\nu$ and $\kappa = \mu_f/(d\rho_p)$.
Notice that, in the inertial limit, the condition that $\partial/\partial T(\partial S/\,\partial\nu)$ [or $\partial/\partial\nu(\partial S/\partial T)$] calculated from the two approaches should be the same is trivially satisfied in the form $0 = 0$: this result was presented in a less compact integral form by Ocone and Astarita (1993). In the general case, the condition $\partial/\partial T(\partial S/\partial\nu) = \partial/\partial\nu(\partial S/\,\partial T)$ is not trivially guaranteed, and it can only be satisfied if the functions $g(\nu)$ and $G(\nu)$ are related to each other by the following relationship:

$$g(\nu)/\nu^2 = 2G'(\nu) \qquad (6.3.2.7)$$

which shows that the constitutive functions $g(\nu)$ and $G(\nu)$ cannot be chosen arbitrarily and independently of each other. Indeed, one only needs to assign $g(\nu)$, which is a strictly mechanical assignment.

We (arbitrarily) choose to regard $G(\nu)$, rather than $g(\nu)$, as the fundamental function (from which $g(\nu)$ is obtained from Eq. (6.3.2.7)). This yields:

$$p = 2\rho_p\nu^2 G'(\nu)\kappa\sqrt{T} \qquad (6.3.2.8)$$

$$-\partial S/\partial\nu = \kappa G'(\nu)/\sqrt{T} \qquad (6.3.2.9)$$

$$\partial S/\partial T = \kappa\, G(\nu)/2T\sqrt{T} \qquad (6.3.2.10)$$

which represent the basic equations of the pseudo-thermodynamic theory. In particular, Eqs (6.3.2.9)–(6.3.2.10) allow us to calculate the pseudo-entropy S at some state ν, T given its value S^* at some reference state ν^*, T^*:

$$S - S^* = \left[\kappa G(\nu)\,/\,2\right]\left[1\sqrt{T}\right] - \int_{\nu^*}^{\nu} \frac{f(\nu)}{\nu^2} - (k\sqrt{T})\left[G(\nu) - G(\nu^*)\right] \qquad (6.3.2.11)$$

Once one has an explicit equation for both U and S, a pseudo-thermodynamic theory is clearly available, and one only needs to develop the details of it — whatever details may be needed for any given problem one wants to consider.

6.4 COMPRESSIBLE FLOW

As gases, granular materials are intrinsically compressible. In chemical engineering there are not many cases where the material is intrinsically compressible, and none of them bears any resemblance to granular flow. However, this subject is in strict analogy with the classical Gas Dynamics (GD) already treated in Sections 1.5 and 4.10, and therefore every analogy will be emphasized. Nevertheless, the subject will be covered following a slightly different approach. The basic idea is to show how it is possible to treat new problems on the basis of the knowledge of a related scientific field. In particular, in the inertial limit case, which will be considered in the following, the analogies with GD are both in the balance and constitutive equations. In this case, for dilute systems, both the pressure and the internal energy are linear functions of the temperature. No effort will be made to consider complex constitutive relations (in most of the analysis we will assume a direct proportionality between the pressure and the solid volume fraction) or difficult geometry (bi-dimensional or tri-dimensional flows).

6.4.1 Wave stability

We wish to discuss the existence and the stability of waves, and in doing that we need to consider unsteady balance equations. Under the hypothesis of one-dimensional flow, the mass balance can be written as:

$$\frac{\partial \nu\rho_{\mathrm{p}}}{\partial t}+\frac{\partial \nu\rho_{\mathrm{p}}u}{\partial x}=0 \tag{6.4.1.1}$$

To simplify the notation (and to have a stronger resemblance to the GD equations) we can substitute the product between the solid volume fraction and the particle density with the particulate averaged density ρ:

$$\frac{\partial \rho}{\partial t}+\frac{\partial \rho u}{\partial x}=0 \tag{6.4.1.2}$$

The momentum balance in conservative form is:

$$\frac{\partial \rho u}{\partial t}+\frac{\partial(\rho u^2+p)}{\partial x}=\frac{\partial}{\partial x}\left(\mu\frac{\partial u}{\partial x}\right) \tag{6.4.1.3}$$

The granular bulk viscosity, μ, is a function of the temperature and, therefore, may well vary in space. For a steady and non-viscous flow this equation simplifies to (4.10.2).

The energy balance is conceptually more difficult because of the dissipation term due to inelastic collisions between the particles. Since granular materials have non-zero thermal conductivity, k, the conservative form of the energy equation is written as:

$$\frac{\partial \rho(U+u^2/2)}{\partial t}+\frac{\partial \rho u(H+u^2/2)}{\partial x}=\frac{\partial}{\partial x}\left(u\mu\frac{\partial u}{\partial x}+k\frac{\partial T}{\partial x}\right)+Q-1 \qquad (6.4.1.4)$$

The term Q may be regarded as an external imposed energy flux and at this stage we do not need to specify its form. Neglecting the imposed energy flux and all the other dissipative terms the balance of energy at steady state reduces to the condition of constant total enthalpy, see Eq. (4.10.4). In GD it is common to indicate the quantity $(H + u^2/2)$ as the total enthalpy, H_0, and the total energy as $E_t = (U + u^2/2)$, so that the energy equation reduces in the more compact form:

$$\frac{\partial \rho(E_t)}{\partial t}+\frac{\partial \rho u H_0}{\partial x}=\frac{\partial}{\partial x}\left(u\mu\frac{\partial u}{\partial x}+k\frac{\partial T}{\partial x}\right)+Q-1 \qquad (6.4.1.5)$$

Before attacking more difficult problems let us see what results (and probably which inconsistencies) it is possible to reach neglecting the dissipative terms, the external imposed heat flux and the time dependence. Clearly these are strong hypotheses, and immediately Sue has some perplexity "but are we sure that we will not lose completely the problem? It seems to me that we are going too far and we will probably reduce our problem to a probably meaningless case".

Under such hypotheses the problem reduces to something very similar to GD; nevertheless, some useful results can be found. Using the same arguments as in classical GD, the speed of sound in granular systems is found to be:

$$a^2=\left.\frac{\partial p}{\partial p}\right|_s \qquad (6.4.1.6)$$

Notice that the derivative is taken at constant entropy; we expect the speed of sound to be a non-dissipative phenomenon (so that the hypotheses made are correct and the entropy is constant).

Bob now has a problem: he thinks that in GD the speed of sound has an immediate physical meaning but it is not so in granular materials. Good point! Indeed, the speed of sound does not measure the celerity of acoustic sound waves in granular materials but it is a measure of the velocity of small disturbances (for instance of density). Moreover, even if compressible materials always undergo compressible flow, in solving practical problems, in some situations one may assume incompressible flow. Bob can easily understand that: if the flow is only slightly compressible, the assumption of compressible flow may produce numerical errors in the solution procedure. Ralph points out that the definition "slightly compressible" is too simplistic, and even if one needs to do "only engineering calculations" there should be a way to measure the importance of compressibility. Well, in GD the difference between an incompressible and a compressible flow may be easily defined and quantitatively verified introducing the concept of steady and isentropic flow. In this way, it is possible to reduce the independent thermodynamic variables to only one (there is a unique relation between density and temperature, density and pressure and between pressure and temperature). Furthermore, after having defined the Mach number, M, as the ratio of the local velocity and the speed of sound in the same point, it is possible to consider the ratio of the static and stagnation values of each thermodynamic variable, which is a function of the Mach number only. In particular, from the value of the density ratio, it is possible to have an insight of "how" compressible the flow is. Since the density ratio is a monotonically decreasing function of the Mach number, by fixing the accuracy, it is possible to find an M such that the flow can be assumed incompressible.

Remark 6.4.1.1

There is a big conceptual difference between the static and stagnation condition of a thermodynamic variable. To explain such a difference we take as an example the temperature (static and stagnation), the generalization to all other thermodynamic variables being trivial. The static temperature, in a point, is that a thermometer, moving with the same speed of the fluid in that point, would measure. The stagnation temperature is the one measured by a stagnant thermometer after the fluid has been decelerated completely with an isentropic and adiabatic transformation. From the definition is immediately evident that the static temperature does not depend on the reference coordinate system, while the stagnation term does. From a very rough balance of energy we can immediately obtain that the stagnation temperature should be greater than (or at least equal to, for zero velocity) the static one. This means that the kinetic energy will be converted in thermal energy.

In granular flow the situation is far more complex; due to the inelasticity of particle–particle and/or particle–wall collisions, there is always a dissipation of internal pseudo energy and hence a change of the entropy level (even for a situation of rest). A way to overcome this problem is to assume that there is a positive influx of energy from the walls, considering, for instance, vibrating walls. This is an example for Q in Eq. (6.4.1.4).

Using the constitutive relation defined by eq. 6.3.1.2 the speed of sound, a, turns out to be:

$$a^2 = T[f'(\nu) + f^2(\nu)/\nu^2] \tag{6.4.1.7}$$

If in the above equation one forgets about the second term within the brackets it is tantamount to falling into the same error as Newton did (i.e., taking the isothermal — rather than the isentropic — compressibility as the relevant one).

Since $f'(\nu) + f^2(\nu)/\nu^2$ is of order unity (it equals 2 in the dilute limit where ν approaches zero and $f(\nu) \cong \nu$, and is somewhat larger than that at moderate values of ν), Eq. (6.4.1.7) shows that the speed of granular sound is of the order $\sqrt{T}$. For typical values of T, as extracted from the literature, the speed of sound is of the order 1 m/s. Since the external velocity scale for the particulate system is typically of the same order of magnitude, one concludes that compressible flow of granular systems is likely to take place at Mach number, M, of order unity.

6.4.2 Propagation of finite discontinuities

The next problem that may be approached with this simplified form of the governing equations is that of the propagation of finite discontinuities. Again, using the same arguments of classical GD it is possible to rewrite the balance equation in the form:

$$[\rho u] = 0 \qquad \Rightarrow \qquad \rho u = W \tag{6.4.2.1}$$

$$[p + \rho u^2] = 0 \qquad \Rightarrow \qquad p + \rho u^2 = N \tag{6.4.2.2}$$

$$[H + u^2/2] = 0 \quad \Rightarrow \quad H + u^2/2 = H_0 \tag{6.4.2.3}$$

where the square brackets indicate a jump through the discontinuity, and equations on the right hand side are equivalent to the ones on the left side as discussed in Chapter 1. In particular, the first states that the mass flow W should be constant through a discontinuity and this is an obvious consequence of the steady state hypothesis. The second one simply shows that, in absence of dissipative effects the momentum flux is conserved, i.e. the impulse N is constant. Equation (6.4.2.3) is a little trickier, and in fact the constancy of the total enthalpy is not a completely correct statement, as will be seen later.

To follow the way paved by GD, and to avoid all possible difficulties, in the following we will restrict our attention to dilute systems so that the function $f(\nu)$ can be taken equal to ν. Therefore, the constitutive relation reduces to the one valid for a perfect gas with a specific heat ratio equal to 2. In this case both the constant volume specific heat and the gas constant are equal to 1. Squaring Eq. (6.4.2.1) and expressing the velocity, for the three balance equations, in terms of the Mach number ($u^2 = 2T M^2$) we obtain:

$$2\rho p M^2 = W^2 \tag{6.4.2.4}$$

$$p(1 + 2M^2) = N \tag{6.4.2.5}$$

$$2T(1 + M^2) = H_0 \tag{6.4.2.6}$$

Combining these equations, the dependence from ρ, T and p can be eliminated, and it is possible to find a relation of the Mach number only:

$$\frac{M^2\left(1 + \dfrac{M^2}{2}\right)}{(1 + 2M^2)^2} = \cos t \tag{6.4.2.7}$$

If we assume the existence of a stationary discontinuity, it is possible to regard the left hand side as a relation between the Mach numbers on the left (L) and right (R) side of the discontinuity:

$$(M_L^2 - M_R^2)\left(1 - 2M_L^2 M_R^2 + \frac{M_L^2 + M_R^2}{2}\right) = 0 \qquad (6.4.2.8)$$

A possible solution of this equation is the trivial one $M_L = M_R$: in the limiting case where the discontinuity degenerates into a continuous flow (no discontinuity is present) nothing should change across it. If the second bracket is taken equal to zero, a unique relationship between the Mach numbers on the two sides of the discontinuity is obtained. As expected, the formula is symmetric (it is possible to switch the left with the right side without changing anything) and admits the solution $M_L = M_R = 1$.

So far we have just approached GD from a different point of view and it looks as everything works for granular materials as well. However, two points remain unclear: is the one obtained an admissible steady solution? How about the stability of the discontinuity?

In the above discussion we have implicitly assumed that on both sides of the discontinuity nothing changes (this is true in GD). If we put a gas in a closed container, in the absence of externally imposed heat flux and/or shaft work, the gas remains in the same thermodynamic state. The situation is completely different in granular flow because of the dissipation of internal energy: to have a thermalized state, it is necessary to impose an external heat flux which balances the term I (vibrated conduit, for instance). The initial assumption to neglect in the energy balance equation the term $Q - I$ has been proved to be wrong, and therefore an account on the implications of such a term is needed. For dilute systems the term I may be written as:

$$I = \rho T^{3/2}/D \qquad (6.4.2.9)$$

Remark 6.4.2.1

> D is an intrinsic length scale and is equal to $d/(1 - e)$, where e is the coefficient of restitution of the inelastic particle–particle collisions. Notice that, should the collisions be perfectly elastic ($e = 1$), as they are in Maxwellian theory of rarefied gases, the intrinsic length scale would degenerate into $D = \infty$. This explains why in the Maxwellian theory, there is no intrinsic scale.

Equation (6.4.2.9) imposes a relationship between possible couples of values of ρ and T which can be maintained at steady state. Condition (6.4.2.9) can be satisfied on one side of the discontinuity, say on the left hand side, by T_L

and ρ_L. On the right side of the discontinuity the values of T_R and ρ_R, which can be calculated from Eqs. (6.4.2.1)–(6.4.2.3), will not satisfy Eq. (6.4.2.9). Indeed, consider for instance a compression shock, for which, according to Eqs. (6.4.2.1)–(6.4.2.3), $T_R \geq T_L$ and $\rho_R \geq \rho_L$. Equation (6.4.2.9) is certain not to be satisfied (except in the trivial case), since it gives T as a decreasing function of ρ. Hence the shock, if indeed it can be maintained at all, is of necessity followed by a wake where thermal relaxation takes place, until (presumably) far downstream Eq. (6.4.2.9) is satisfied again, and a new steady state is reached. There is no equivalent of such a relaxation phenomenon in classical GD theory, where any couple of values of temperature and density can be maintained indefinitely.

6.4.3 Why classical (GD) methods do not work — a first example

The difficulties involved in treating granular systems are best appreciated by trying to follow classical GD in the formulation of the general one-dimensional granular compressible flow problem. Consider a smooth variation of, say, ρ initially imposed on a one-dimensional flow field of the particulate phase such as a plug down a vibrated conduit. Reinforcement of compression waves is simply due to the fact that the speed of sound increases with increasing temperature, so that successive infinitesimal compression disturbances propagate at an increasing speed. In a particulate phase, the speed of pseudo-sound also increases with increasing pseudo-temperature (Ocone and Astarita, 1995c), but the latter undergoes the relaxation phenomenon discussed before, and this, as will be seen, makes a lot of difference.

Let $\wp$ be the difference $Q - I$. Based on the usual approximations which are standard in gas dynamics (essentially, that viscous and thermal conductivity are set to zero), the balance equations can be written, in the non-conservative form, as:

$$\frac{\partial \rho}{\partial t} + u\frac{\partial \rho}{\partial x} + \rho\frac{\partial u}{\partial x} = 0 \tag{6.4.3.1}$$

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{1}{\rho}\frac{\partial p}{\partial x} = 0 \tag{6.4.3.2}$$

$$\frac{\partial U}{\partial t} + \frac{p}{\rho}\frac{\partial u}{\partial x} + u\frac{\partial U}{\partial x} = \frac{\wp}{\rho} \tag{6.4.3.3}$$

The culprit is the energy balance, because of the appearance of the term $\wp$ which makes the system of Eqs. (6.4.3.1)–(6.4.3.3) inhomogeneous.

We now try to develop the same compact method of solution which is standard in GD. The pressure p is a unique function of ρ and U, and the following two quantities can be defined:

$$\upsilon = \left.\frac{\partial p}{\partial \rho}\right|_{U} \tag{6.4.3.4}$$

$$X = \left.\frac{\partial p}{\partial \rho}\right|_{\rho} \tag{6.4.3.5}$$

We also define the following two vectors and the following matrix:

$$\mathbf{b} = (\rho, u, U) \tag{6.4.3.6}$$

$$\mathbf{p} = (0, 0, \wp/\rho) \tag{6.4.3.7}$$

$$\mathbf{A} = \begin{vmatrix} u & \rho & 0 \\ \upsilon/\rho & u & X/\rho \\ 0 & p/\rho & u \end{vmatrix} \tag{6.4.3.8}$$

so that the system of Eqs. (6.4.3.1)–(6.4.3.3) reduces to:

$$\frac{\partial \mathbf{b}}{\partial t} + \mathbf{A}\frac{\partial \mathbf{b}}{\partial x} = \mathbf{d} \tag{6.4.3.9}$$

Contrary to the case of gas dynamics, where $\mathbf{d} = 0$, the quasi-linear system (6.4.3.9) is inhomogeneous. However, just as in the case of GD, the three eigenvalues of **A**, which define the characteristic curves of the hyperbolic system (6.4.3.9), are u, $u - a$, and $u + a$.

The system (6.4.3.9) represents a typical hyperbolic problem and the classical approach is to diagonalize it. Suppose one can find the three Riemann invariants of (6.4.3.9) which would constitute a 3-dimensional vector **y**. In that case, the system (6.4.3.9) reduced to:

$$\frac{\partial \mathbf{y}}{\partial t} + \mathbf{B}\frac{\partial \mathbf{y}}{\partial x} = \mathbf{Cd} \tag{6.4.3.10}$$

Here matrix **B** is a diagonal matrix, so that the three components of Eq. (6.4.3.10) are uncoupled. The elements of **B** are the eigenvalues of **A**:

$$\mathbf{B} = \mathbf{CAC}^{-1} \tag{6.4.3.11}$$

and $\mathbf{C} = \partial\mathbf{y}/\partial\mathbf{b}$ is a matrix whose lines are the left eigenvectors of **A**:

$$\mathbf{C} = \begin{vmatrix} p/\rho & 0 & -1 \\ \upsilon/X & -a\rho/X & 1 \\ \upsilon/X & a\rho/X & 1 \end{vmatrix} \tag{6.4.3.12}$$

The Riemann invariants can in fact be found if each row of **C** derives from a pseudo-potential, say if **y** is expressible as $r(\mathbf{b})\mathrm{grad}z(\mathbf{b})$, with $r()$ and $z()$ being scalar-valued functions. This in turn implies that the following equation holds for $i = 1, 2, 3$:

$$curl(\partial y_i / \partial \mathbf{b}) \cdot (\partial y_i / \partial \mathbf{b}) = 0 \tag{6.4.3.13}$$

Since the relation between p and T is linear (which is true for ideal gases as well as for granular materials), $(\partial y_i/\partial\mathbf{b})$ admits the integrating factor $1/T$ and hence it can be directly integrated. Equation (6.4.3.13) is satisfied since $(\partial y_i/\partial\mathbf{b})$ is solenoidal. It is important to realize that y_1 is in fact the entropy, and that the equation for it is:

$$\frac{DS}{Dt} = \frac{\partial S}{\partial t} + u\frac{\partial S}{\partial X} = \frac{\wp}{\rho T} \tag{6.4.14}$$

In GD, $\wp = 0$, and hence Eq. (6.4.3.14) implies that in the absence of shocks, if the initial conditions are isentropic, they will stay isentropic at all times. This in turn implies that (6.4.3.3) can be discarded and substituted by the condition of isentropicity. Thus p reduces to a unique function of ρ, $\mathrm{d}p/\mathrm{d}\rho = a^2$, and the problem formulation reduces to only two equations. Eq. (6.4.3.13) is satisfied by the reduced set of equations, and hence the Riemann invariants are established. With this, the system reduces to two ordinary differential equations along the characteristic curves, and the solution can be obtained along the classical lines. However, Eq. (6.4.3.14) shows that, in the case of granular flow, even if the initial conditions are isentropic, they will not stay so at later times — and one cannot follow the shortcut of classical GD. This is related to the fact that the thermal relaxation phenomenon discussed before is not isentropic, and relaxation takes place only when $\wp$ is nonzero. In the system of Eqs. (6.4.3.1)–(6.4.3.3), $\wp$ is a lower order term, and yet its presence gives an entirely new structure to the problem.

At this stage, we have concluded that the classical approach for the problem of waves propagation cannot be followed in the case of granular flow. Formally, there is also the difficulty, that since Riemann invariants cannot be found, there is no guarantee of existence and uniqueness of the solution; this however we believe is not a problem. Of course, Eqs. (6.4.3.1)–(6.4.3.3) could be solved numerically for any assigned initial condition (in the following this will be done for a particular case). However, we are interested first in establishing the structure of the analysis of compressible flow of granular materials, and we analyze another problem in the following.

6.4.4 Why classical (GD) methods do not work — a second example

Another problem which can be analyzed is the determination of the "thickness" of a compression wave. The thickness will be finite if the effects of a nonzero solid bulk viscosity, μ, and thermal conductivity, k are included. Inclusion of bulk viscosity and thermal conductivity makes the problem parabolic (unsteady state formulation), or elliptic (steady state), so that the question of possible existence of Riemann invariants disappears entirely.

Consider plug flow of a granular material, kept thermalized by vibration, down a constant section conduit. Consider a wave which is moving in the negative x direction at some speed V; the wave can be regarded as stationary if one subtracts V from all velocities. The resulting standing wave problem is elliptic, and it bears of course a strong analogy with the standing wave problem in classical GD.

The mass, momentum and energy balances reduce to:

$$\frac{\partial \rho u}{\partial x} = 0 \tag{6.4.4.1}$$

$$\frac{\partial(\rho u^2 + p)}{\partial x} = \frac{\partial}{\partial x}\left(\mu \frac{\partial u}{\partial x}\right) \tag{6.4.4.2}$$

$$\frac{\partial \rho u(H + u^2 / 2)}{\partial x} = \frac{\partial}{\partial x}\left(u\mu \frac{\partial u}{\partial x} + k\frac{\partial T}{\partial x}\right) + \wp \tag{6.4.4.3}$$

In principle one would need to determine the distributions of u, p, ρ, and $U = T$ within the wave. Of these, only three are independent, since p is a unique function of ρ and U. The three distributions can be obtained, in principle, by integrating Eqs. (6.4.4.1)–(6.4.4.3). If Eq. (6.4.4.3) is integrated over the whole thickness of the wave, the terms including μ and k integrate to zero.

The balance of energy, however, poses again a problem. Let $\Re$ be defined as the total unbalance between energy supply and dissipation $(Q - I)$ within the wave:

$$\Re = \int_{\mathrm{L}}^{\mathrm{R}} \wp \, \mathrm{d}x \tag{6.4.4.4}$$

Integrating Eq. (6.4.4.4) between L and R yields:

$$T_{\mathrm{L}} + p_{\mathrm{L}}/\rho_{\mathrm{L}} + u^2_{\mathrm{L}}/2 + \Re = T_{\mathrm{R}} + p_{\mathrm{R}}/\rho_{\mathrm{R}} + u^2_{\mathrm{R}}/2 \tag{6.4.4.5}$$

which implies that the total enthalpy on the right is different from what it is on the left (the sign of $\Re$ cannot be established *a priori*).

In classical GD, $W = 0$, and hence the total enthalpy is the same on the two sides; from this one obtains the classical Hugoniot conditions relating temperature, velocity and density on the two sides of a shock. The Hugoniot conditions have originally been derived for an existing shock (not a finite-thickness standing wave), and in their derivation bulk viscosity and thermal conductivity are set at zero. However, the equations apply in spite of the fact that one is considering a wave of finite thickness and not a shock. The bulk viscosity and the conductivity are non-zero, and their values determine the thickness of the wave, as well as the distributions of T, u and p within it, but they do not influence the relationship between the values of these quantities on the right and on the left.

In the case of granular materials, Eq. (6.4.4.5) shows that the standing wave equations do not result in the same left and right conditions as those of a shock; this of course implies that the very possibility of the existence of finite shocks cannot be taken for granted. The mathematical status of the problem is as follows. The analysis of an existing shock is in fact the "outer" solution of a classical singular perturbation problem, and in this sense the standing wave is an inner boundary layer. In GD, the inner boundary layer equations integrate to the shock ones, indicating that the outer solution is a proper asymptote. This is not the case for granular systems, for which the asymptotic status of the shock equations is unclear.

In both the previous examples it has been evidenced the importance of the lower order term $Q-I$, and it is now clear that it is not possible to have a steady solution unless the dissipation of internal energy is balanced by the imposed energy flux. At the same time, we can neglect both k and μ without losing the problem. Under these hypothesis Eq. (6.4.1.4) does not state that the total enthalpy should be constant on the two sides of the discontinuity, the term $Q-I$ behaves as a source term.

6.4.5 The Rayleigh curve — a stability analysis

Let us put aside for a moment the energy equation and analyze the implication of Eqs. (6.4.2.1)–(6.4.2.2). The mass and momentum balances do not, of course, determine the values of p, ρ, u and T, but they do establish a unique relationship between any two of them once W and N have been assigned. It is, therefore, possible to draw a curve in the (T–S) plane, every point of which satisfies Eqs. (6.4.2.1)–(6.4.2.2); such a plot, that in GD is called the Rayleigh curve, is presented in Fig. 6.4.5.1.

The Rayleigh curve in Fig. 6.4.5.1 may be thought of as made up of two different branches, separated by the point of maximum entropy where $M = 1$. (The ordinate axis is the temperature T divided by the temperature corresponding to $M = 1$, T^*. The shape of the curve is the same, no matter what the values of Wand N might be). The upper branch of the curve corresponds to subsonic conditions ($M < 1$), and the lower branch to supersonic ones ($M > 1$), with the point $T = T^*$, $M = 1$ corresponding to the maximum of entropy (a square in Fig. 6.4.5.1). Note that in the upper branch M increases when entropy is increased, and *vice versa* in the lower branch.

As can be seen in Fig. 6.4.5.1, temperature has a maximum (a circle in Fig. 6.4.5.1) when $M = 1/\sqrt{2}$. If one accepts the idea that an extra supply of energy can only result in an increase of entropy, one reaches the somewhat counterintuitive conclusion that, for subsonic M values larger than $1/\sqrt{2}$ (the branch between the circle and the square), addition of energy results in a decrease of

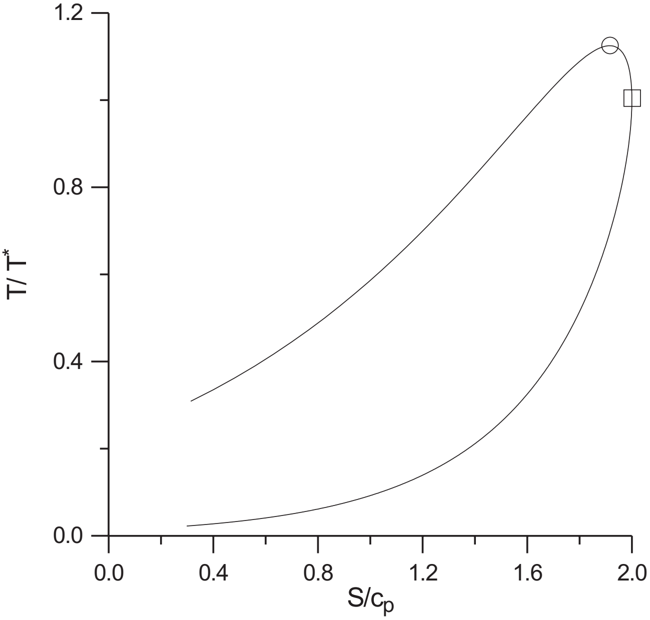

Fig. 6.4.5.1. Rayleigh curve in the temperature entropy plane. The lower branch is supersonic, the upper branch is subsonic. The circle (maximum temperature) corresponds to $M = 1/\sqrt{2}$. The critical point $M = 1$ (square) corresponds to the maximum of both entropy and total enthalpy.

temperature. Notice also that it is impossible to add energy to the system, once $M = 1$, without changing the mass flow rate; this point is discussed in more detail later.

At this point we need to identify the couples of steady state values on the two sides of the discontinuity and we need to use the condition that $Q = I$ on both sides. With the same procedure already used to find Eq. (6.4.2.8) it is possible to find:

$$M_2(1+2M_1^2) = M_1(1+2M_2^2)^2 \tag{6.4.5.1}$$

or, which is the same:

$$(M_1 - M_2)(4M_1M_1 - 1 + 4M_1M_2(M_1^2 + M_1M_2 + M_2^2)) = 0 \tag{6.4.5.2}$$

This relation admits the trivial solution $M_1 = M_2$ but, and this is a big difference with GD, the second term of the left hand side does not admit the solution $M_1 = M_2 = 1$; it admits the solution $M_1 = M_2 = M^\wedge = 1/\sqrt{6}$ instead. [$M^\wedge$ can be considered as "analogous" of $M = 1$ in GD, i.e. it represents the switch for subsonic–supersonic flow. Also, $M^\wedge$ is the only Mach number where a wave cannot occur]. If M_1 is less than $M^\wedge$, then $M_2 > M^\wedge$, and *vice versa*. Hence, unless the Mach number in front of the wave is exactly $M^\wedge$, a non-trivial wave may well exist, with M_1 and M_2 the two Mach numbers on the two sides of the wave. This is a crucial difference from classical GD where it is not possible to have a completely subsonic wave.

Before continuing with the analysis it is useful to make the equations dimensionless. However, new difficulties arise: which are the correct reference quantities? In our problem we have four externally imposed quantities: W, D, N and Q and we need only three of them to make the equations dimensionless. If we use the first we obtain:

$$\begin{gathered} p^* = p/N;\ \rho^* = \rho N/W^2;\ T^* = T\,W^2/N^2;\ u^* = uW/N; \\ Q^* = QWD/N^2 \end{gathered} \tag{6.4.5.3}$$

Clearly U, H and H_0 can be made dimensionless in the same way as the temperature. If we suppose that the relevant imposed quantities are W, D and Q, we obtain:

$$\begin{gathered} \beta = p/\sqrt{(QWD)};\ r = \rho\sqrt{(QD/W^3)};\ \theta = TW/QD;\ \phi = u\sqrt{(W/Qd)}; \\ N^* = N/\sqrt{(QWD)}\ i = I/Q \end{gathered} \tag{6.4.5.4}$$

Let $\Im$ be defined as QWD/N^2, each term of the first non-dimensionalisations differs from the corresponding term in the second one for a power of $\Im$. Expressing $\Im$ as a function of the Mach number one obtains:

$$\Im = \frac{\sqrt{2}M}{(1+2M^2)^2} \tag{6.4.5.5}$$

Now this seems a little strange because we would have expected that $\Im$ should be constant. The problem arises because Q must be equal to the dissipative term I on the two sides of the wave, since it is a function of the condition ahead of the wave; this is why Eq. (6.4.4.5) is practically equivalent to Eq. (6.4.5.1). Using Eq. (6.4.5.3) the relevant equations become:

$$p^* = \rho^* T^* \tag{6.4.5.6}$$

$$\rho^* V^* = 1 \tag{6.4.5.7}$$

$$p^* + \rho^* V^{*2} = p^* + V^* = 1 \tag{6.4.5.8}$$

$$p^* \sqrt{T^*} = Q^* \tag{6.4.5.9}$$

The values of p^*, ρ^*, T^*, $p^*\sqrt{T^*}$, and H_0^* are unique functions of the Mach number:

$$p^* = 1/(1 + 2M^2) \tag{6.4.5.10}$$

$$\rho^* = (1 + 2M^2)/2M^2 \tag{6.4.5.11}$$

$$T^* = 2M^2/(1 + 2M^2)^2 \tag{6.4.5.12}$$

$$p^*\sqrt{T^*} = \sqrt{2M}/(1 + 2M^2)^2 \tag{6.4.5.13}$$

$$H_0^* = 2(2M^2 + M^4)/(1 + 2M^2)^2 \tag{6.4.5.14}$$

The quantity $p^*\sqrt{T^*}$ is plotted in Fig. 6.4.5.2 versus M. It starts from the origin, grows until $M = \hat{M} = 1/\sqrt{6}$, where it has its maximum value of $9/16\sqrt{3} = 0.325$, and then decays continuously to zero. *It follows that no solution exists if Q^* is larger than 0.325.* If Q^* is less than 0.325, there are always two values of M that satisfy all the equations. For all values of M intermediate between those two values, $p^*\sqrt{T^*}$ is larger than Q^* (fluctuation energy is dissipated at a rate exceeding Q), and *vice versa* for all values outside the interval.

Equation (6.4.5.14) shows that H_0^* increases with increasing M in the subsonic region ($M < 1$), and *vice versa*; the function $H_0^*(M)$ is plotted in Fig. 6.4.5.3. It starts from zero, grows to a maximum of 2/3 at $M = 1$, and then decreases steadily to the asymptotic value ½. With reference to the classical Rayleigh curves in Fig. 6.4.5.1, for both branches of the curve, H_0^* increases as one moves right towards the critical point $M = 1$, i.e., H_0^* always increases with increasing granular entropy S.

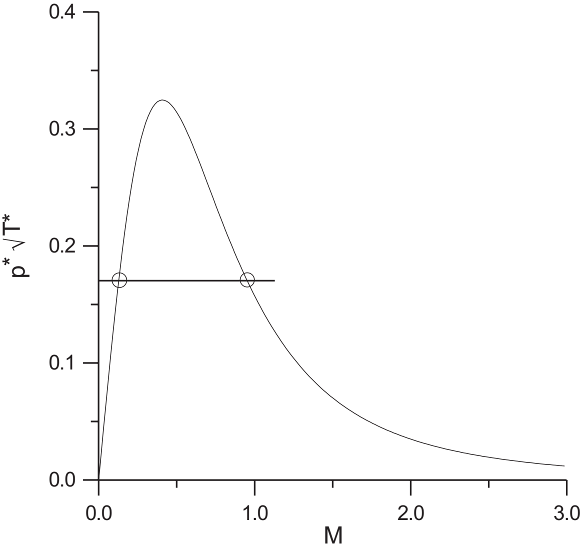

Fig. 6.4.5.2. $p^*\sqrt{T^*}$ versus M. Given the value of Q (horizontal line the two steady-state solutions are the intersection with the curve (circles).

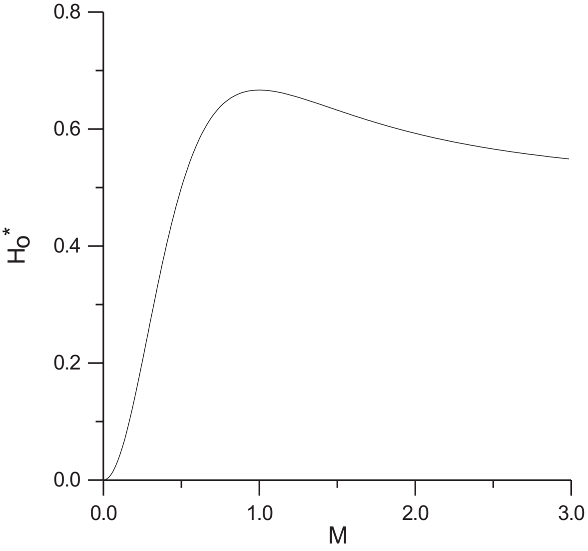

Fig 6.4.5.3. H_0* versus M.

This is as far as one can possibly go without using the balance of energy. We now can write this in a dimensionless form as:

$$\mathrm{d}H_0{}^*/\mathrm{d}x = (Q - P^*\sqrt{T^*})/D \tag{6.4.5.15}$$

or, defining a dimensionless distance $y = x/D$ (i.e. scaling distances with respect to the intrinsic length scale D):

$$\mathrm{d}H_0{}^*/\mathrm{d}y = (Q - P^*\sqrt{T^*}) \tag{6.4.5.16}$$

Remark 6.4.5.1

We have implicitly assumed that we are not dealing with discontinuity any longer; indeed, Eq. (6.4.5.16) is valid only if the mass flux W is constant across the wave and H_0 itself should be a continuous function of the spatial coordinate.

Remark 6.4.5.2

To avoid the problem that H_0 is not an invertible function of the Mach number Eq. (6.4.5.16) may be transformed into a form where $\mathrm{d}M/\mathrm{d}y$ is given by a known function of M itself.

In the Rayleigh approach to classical gas dynamics, the right-hand side of Eq. (6.4.5.15) is imposed from outside, and hence, Eq. (6.4.5.16) can simply be integrated between stations 1 and 2 to yield $H_0{}^*{}_2 - H_0{}^*{}_1$ in terms of the total energy supplied between the two stations. *This cannot be done in the granular flow case, where Q is imposed from outside but I depends on the local conditions (local values of p and T).* Let M' and M'' ($>M'$) be the two steady-state solutions, as determined from the horizontal line $p^*\sqrt{T^*} = Q^*$ in Fig. 6.4.5.2. M' is always subsonic, while M'' may be both subsonic or supersonic, depending on the value of Q^*.

It is easy to convince oneself that any value of M that is a *steady-state solution is locally stable if the sign of* $\mathrm{d}(p^*\sqrt{T^*})/\mathrm{d}M$ *is the same as that of* $\mathrm{d}H_0^*/\mathrm{d}M$. This identifies three regions of interest and two singular points:

1. $0 \le M < \hat{M}$, where both $\mathrm{d}(p^*\sqrt{T^*})/\mathrm{d}M$ and $\mathrm{d}H_0{}^*/\mathrm{d}M$ are positive: M' is always locally stable.
2. $\hat{M} < M < 1$, where $\mathrm{d}(p^*\sqrt{T^*})/\mathrm{d}M$ is negative but $\mathrm{d}H_0{}^*/\mathrm{d}M$ is positive: a subsonic M'' is always unstable.

3. $1 < M$, where both $\mathrm{d}(p^*\sqrt{T^*})/\mathrm{d}M$ and $\mathrm{d}H_0^*/\mathrm{d}M$ are negative: a supersonic M'' is always stable.
4. $M = \hat{M}$. Here, $\mathrm{d}(p^*\sqrt{T^*})/\mathrm{d}M = 0$, $\mathrm{d}^2(p^*\sqrt{T^*})/\mathrm{d}M^2$ is negative, and $\mathrm{d}H_0^*/\mathrm{d}M$ is positive. This point is stable to left perturbations but it is unstable to right ones. However, this point is of interest only in the case where the imposed value of Q is exactly $9/16\sqrt{3}$, and hence, it does not need special analysis. *It is, however, of interest that there is no* $\hat{M}$ *in classical gas dynamics*.
5. $M = 1$. Here, $\mathrm{d}(p^*\sqrt{T^*})/\mathrm{d}M$ is negative, $\mathrm{d}H_0^*/\mathrm{d}M = 0$, and $\mathrm{d}^2H_0^*/\mathrm{d}M^2$ is negative. This point is stable to right perturbations but it is unstable to left ones. The special interest of this point is that it is (perhaps) reached for any right perturbation of a subsonic M'', but the condition $Q^* = p^*\sqrt{T^*}$ will not be satisfied.

Figure 6.4.5.4 gives the axial distributions of velocity, pressure, temperature, density, Mach number, and $Q^* - p^*\sqrt{T^*}$ calculated for a left perturbation of a subsonic M''. Velocity and temperature decrease as one moves downstream, while pressure and density increase. Note that, as expected, $Q^* - p^*\sqrt{T^*}$ is negative throughout the wave, and has a minimum at the same axial position where M takes the value $\hat{M}$. Figure 6.4.5.4 shows that a non-trivial standing wave can exist in a granular system, although *it has quite a different logical status from that enjoyed by the classical standing-wave analyses* (Zucrow and Hoffman, 1976).

Now consider the case where the initial perturbation is positive, $\delta M > 0$. The stability analysis predicts that M will keep growing (and, hence, diverging from M''), until M reaches the value of unity. At that point, Q^* still exceeds $p^*\sqrt{T^*}$, and the situation is analogous to what happens in classical gas dynamics if one keeps adding energy to a system when $M = 1$. In other words it is impossible to keep adding energy without changing the flow rate (i.e., the value of W). The physics underlying this situation are discussed below.

In the classical case, one considers a conduit (which might be, e.g., a converging nozzle followed by a straight tube) emerging from a very large reservoir kept at some constant upstream pressure (this implies that N is fixed, since it is simply the pressure in the reservoir where $u = 0$). As one adds energy through the walls, one may approach the situation where the sonic velocity is reached. At this point, further addition of energy will simply decrease the flow rate W (the flow has become choked), thus displacing to the right and

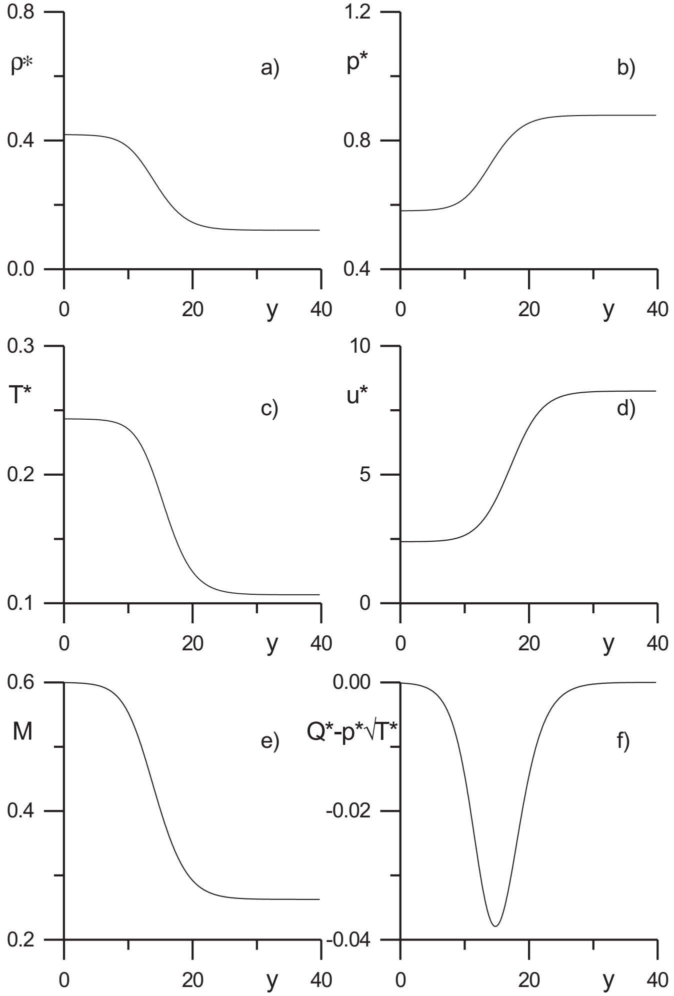

Fig 6.4.5.4. Stationary waves resulting from a left perturbation about a subsonic M''.

enlarging the Rayleigh curve in Fig. 6.4.5.1. The velocity will stay at the critical value corresponding to $M = 1$ (but to progressively higher values of entropy).

In the granular flow case, *one cannot reach the same conclusions*, and a complete analysis of what happens for a positive perturbation around a subsonic M'' would require two things that are well beyond the scope of the present chapter: first, one would need a model of the feeding mechanism (the

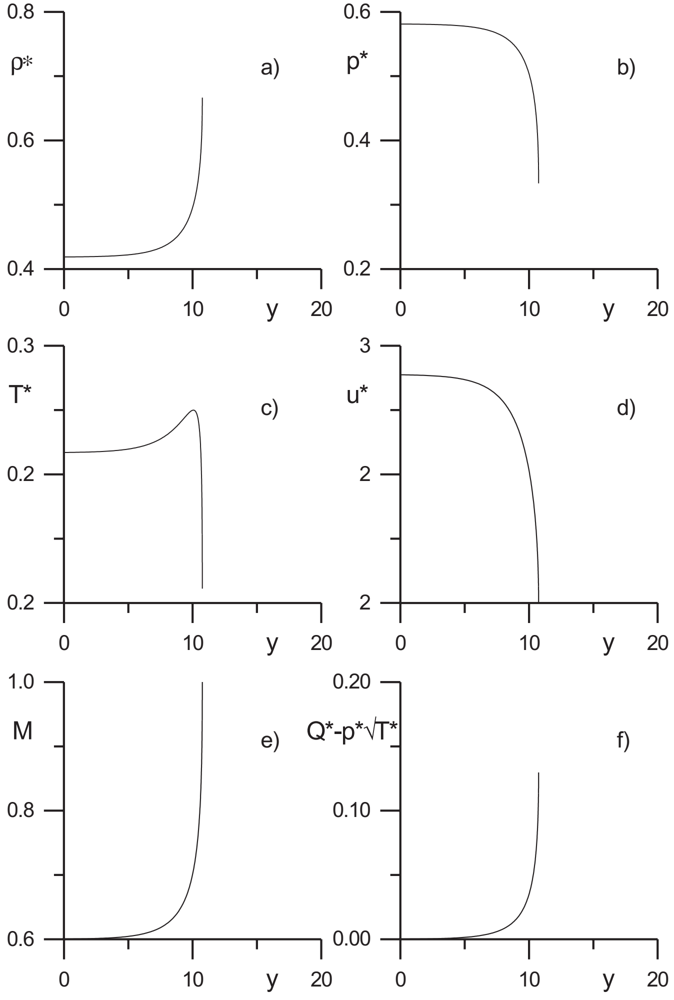

Fig 6.4.5.5. Results from a right perturbation about a subsonic *M*″.

equivalent of the constant pressure large upstream reservoir and the nozzle connecting it to the conduit under consideration); second, one would need to write down the unsteady-state governing equations coupled with the feeding mechanism.

Figure 6.4.5.5 gives p^*, ρ^*, T^*, u^*, M, and $Q^* - p^*\sqrt{T^*}$ as a function of y as calculated for a positive infinitesimal perturbation around a subsonic M''. The

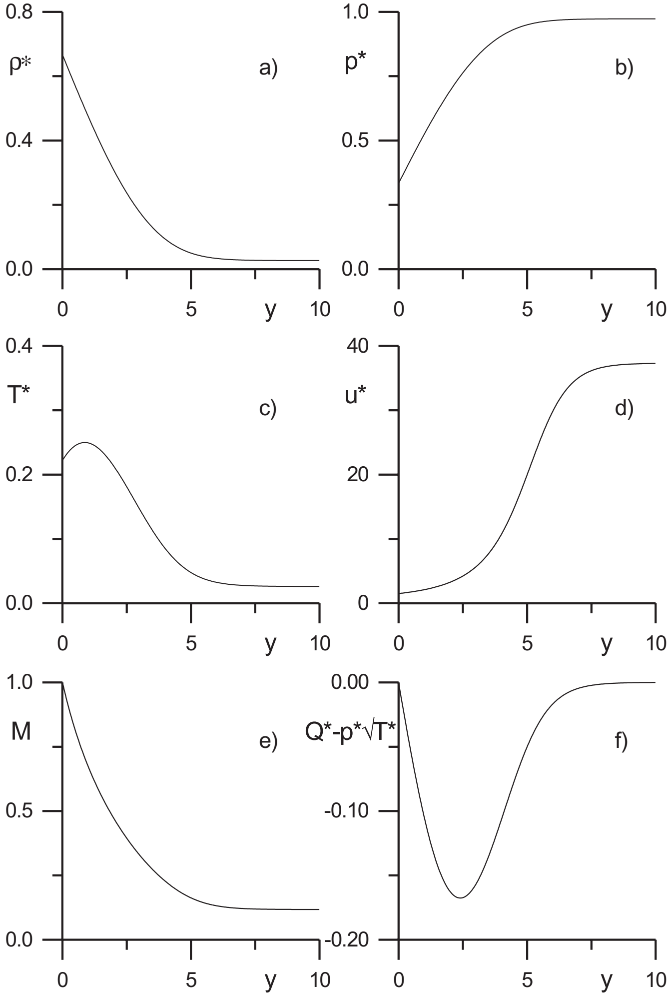

Fig. 6.4.5.6. Waves resulting from a lift perturbation about $M = 1$.

numerical scheme diverges as M approaches unity, because the model equations do not make sense at $M = 1$, as discussed above. If one assumes (by way of a very loose analogy with the classical gas dynamics case) that, indeed, the condition $M = 1$ is reached, and that then somehow W and N change until $Q^* = \sqrt{2/9}$, one becomes interested in the stability of the $M = 1$ point. Figure 6.4.5.6 gives the same quantities as in Fig. 6.4.5.5 as functions of y once the initial M value has been set to unity, and the Q^* value to $\sqrt{2/9}$. The final asymptotic value of M is indeed 0.117, i.e., the M' value corresponding to $M'' = 1$.

The stability analysis discussed above has some interesting features. In chemical engineering, there are scores of problems that admit more than one steady-state solution; however, the number of solutions is usually odd, and in the simplest case where there are three solutions, the two outer ones are locally stable, and the middle one is locally unstable; the corresponding stability analysis is comparatively trivial. In the problem considered here, there are only two steady-state solutions, and the stability analysis has more conceptual content.

6.4.6 Shock waves

The analysis presented above is as far as one can possibly go without considering the unsteady balance equations. We now want to examine the case of what in GD is called a shock wave i.e. a discontinuity. From the previous analysis we know that a supersonic point is stable, and at first sight we might expect that there should be no equivalent of a shock wave in compressible granular flow. In fact, this is not correct at all: in GD each point is a stable one, and nevertheless a shock wave may well exist and the same may occur for granular flows.

In the following we will use the dimensionless form of the balance equations derived by the use of Eq. (6.4.5.4) (in particular $t^* = t\sqrt{(Q / WD)}$) and we will consider only the case of a compression wave. Taking into account the difficulties found before it is not possible to exactly integrate the governing equation, so in order to find some results it is necessary to use a numerical method. Also in this case we will neglect the viscosity and thermal conductivity. Numerical methods for hyperbolic problems are known to introduce some diffusive effects and one should be very careful in the interpretation of the results.

Figure 6.4.6.1 represents what, in GD, may be called a shock wave, showing the solution reached at steady state. A step from the left to the right fixed (in time) values is taken for all variables as initial condition. Moreover, the two values are chosen such that they satisfy Eq. (6.4.5.2). Given the above dimensionless form of the equations, both the mass flow rate and the dissipation rate at the two boundary points are equal to 1. All the profiles start with a zero derivative, then, the curves present a first region (A) where a very sharp

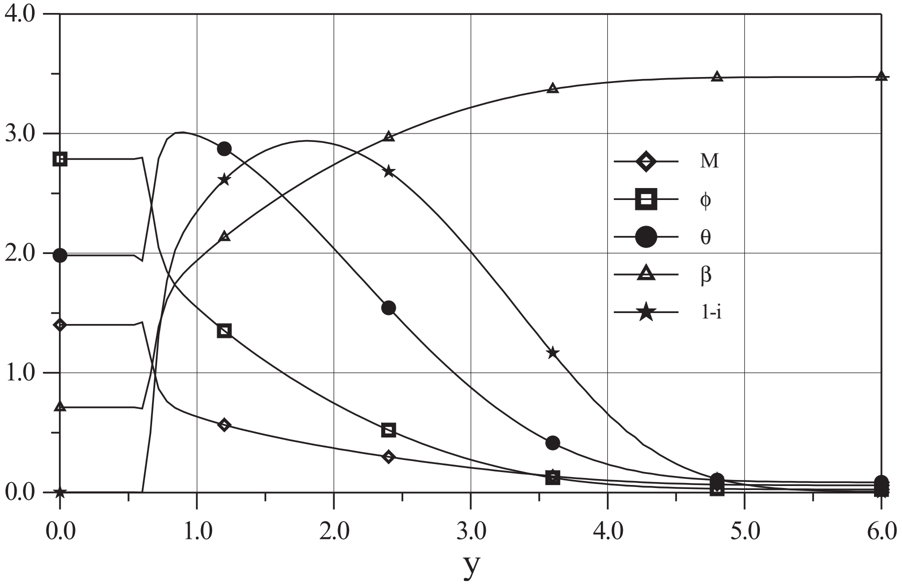

Fig 6.4.6.1. Spatial variation of Mach Number (*M*), velocity (ϕ), temperature (θ), pressure (β), energy unbalance (1 – *i*) for a weak shock wave.

increase steep slope is observed. Following this, a second region (B) can be observed where the variations become smoother, and in the limit the derivative goes to zero. This behaviour is more clear in the temperature profile, where the first derivative changes sign and the temperature, after having reached a maximum, decreases to an absolute minimum. The presence of the two regions is related to two different length scales: the first, very small, is due to artificial viscosity effects introduced by the numerical integration; the second is related to relaxation phenomena occurring after the compression wave. It is interesting to note that the maximum of temperature is practically equal to the value that one would calculate from the analysis of Rankine–Hugoniot shock wave (Zucrow and Hoffman, 1976), and the width of the relaxation region is of the same order of magnitude of completely subsonic waves. This last result is of practical importance; indeed, one can get all the information needed on the wave, solving algebraic equations from a Rankine–Hugoniot analysis, and then separately consider the relaxation zone for completely subsonic waves, starting from a non-equilibrium state (for non-equilibrium state we mean a situation where the condition $Q = I$ is not

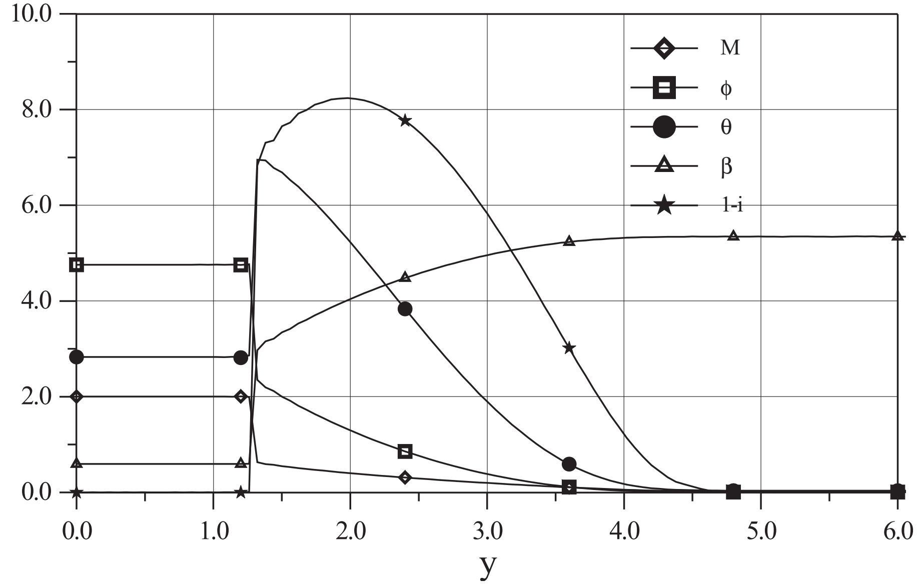

Fig 6.4.6.2. Spatial variation of Mach Number (*M*), velocity (ϕ), temperature (θ), pressure (β), energy unbalance ($1 - i$) for a strong shock wave.

satisfied). In other words, the global problem can be solved by solving two simpler problems and superimposing the solution.

The difference between the two regions is even more evident in Fig. 6.4.6.2 where a "shock wave" with a higher Mach number on the left side is reported. A well known result in GD is that for a diffusive fluid the width of a shock wave decreases as M_1 increases; analogously, here the length scale of region A drastically reduces with respect to the one reported in Fig. 6.4.6.1, whilst the relaxation scale decreases only slightly.

REFERENCES

Astarita, G. and G.C. Sarti, in: Theoretical Rheology, J.F. Hutton, J.R. Pearson and K. Walters (Eds.). Elsevier, Barking, 1975.

Bagnold, R.A., Proc. Roy. Soc. London, A225 (1954) 49.

Beavers, G.S., E.M. Sparrow and D.E. Rodenz, Trans. ASME/J. Appl. Mechanics, Sept. 1973.

Bowen, R.M., In: Continuum Physics, A.C. Eringen (Ed.), Vol. II. Academic Press, New York, 1971.

Coleman, B.D. and W. Noll, Arch. Ratl. Mech. Anal., 13 (1963) 167.
Drew, D., Ann. Rev. Fluid Mech., 15 (1983) 261.
Haff, P.K., J. Fluid Mech., 134 (1983) 401.
Jackson, R., J. Rheol., 30 (1986) 907.
Jenkins, J.T. and S.B. Savage, J. Fluid Mech., 130 (1983) 187.
Jenkins, J.T. and D.F. McTigue, Two Phase Flow and Waves, D.D. Joseph and D.G. Schaeffer (Eds.), Springer Verlag, New York, 1990.
Joseph, D.D., T.S. Lundgren (with an Appendix by R. Jackson and D.A. Saville), Int. J. Multiphase Flow, 16 (1990) 35.
Lun, C.K.K., S.B. Savage, D.J. Jeffrey and N. Chepurnity, J. Fluid Mech., 140 (1984) 223.
Lundgren, T.S., J. Fluid Mech., 51 (1972) 273.
Maddox, J., Nature, 374 (6517) (1995) 11.
Maxwell, J.C., Phil. Trans. Roy. Soc. London, 157 (1867) 49.
McTigue, D.F. and J.T. Jenkins, Advances in Micromechanics of Granular Materials, H.H. Shen, M. Satake, M. Mehrabadi, C.S. Chang, and C.S. Campbell (Eds.). Elsevier, Amsterdam, 1992.
Ocone, R., PhD Thesis, Princeton, 1992.
Ocone, R. and G. Astarita, J. Rheol., 38 (1993) 727.
Ocone, R. and G. Astarita, Nature, 375 (6530) (1995a) 254.
Ocone, R. and G. Astarita, Rheol. Acta, 34 (4) (1995b) 323.
Ocone, R. and G. Astarita, Powder Tech., 82 (1995c) 231.
Ogawa, S., Proc. of the US-Japan Symp. on Cont. Mech. and Statist. Appr. in Mech. of Granular Materials, S.C. Cowin and M. Satake (Eds.). Gukujatsu Bunken Fukyakai, 1978.
Truesdell, C.A., Rational Thermodynamics, 2nd edn. Springer-Verlag, Berlin, 1984.
Waterston, J.J., Phil. Trans. Roy. Soc. London, A183, (1893) 1 [Published posthumously, 48 years after its submission in 1845].
Zou, R.P. and A.B. Yu, Chem. Eng. Sci., 50 (9) (1995) 1504.
Zucrow, M.J. and D.J. Hoffman, Gas Dynamics, Vol 1. Whiley, New York, 1976.

Index