

# Mathematical Biofluidynamics

SIR JAMES Lighthill

University of Cambridge  
Cambridge, England

CBMS-NSF  
REGIONAL CONFERENCE SERIES  
IN APPLIED MATHEMATICS

---

SPONSORED BY  
CONFERENCE BOARD OF  
THE MATHEMATICAL SCIENCES

---

SUPPORTED BY  
NATIONAL SCIENCE  
FOUNDATION

---

# **Mathematical Biofluidynamics**

*This page intentionally left blank*

---

Sir James Lighthill  
University of Cambridge  
Cambridge, England

---

# Mathematical Biofluidynamics

A treatment in book form of the material in the lecture course delivered to the "Mathematical Biofluidynamics" Research Conference of the National Science Foundation held from July 16–20, 1973 at Rensselaer Polytechnic Institute, Troy, New York.

**siam.**

SOCIETY FOR INDUSTRIAL AND APPLIED MATHEMATICS

PHILADELPHIA

Copyright ©1975 by the Society for Industrial and Applied Mathematics.

10 9 8 7 6 5 4

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 University City Science Center, Philadelphia, PA 19104-2688.

Portions of this volume are reprinted with the permission of Annual Reviews Inc. and the International Centre for Mechanical Sciences.

ISBN 0-89871-014-6

**siam** is a registered trademark.

# Contents

Acknowledgements .....	ix
<b>Chapter 1</b>	
INTRODUCTION TO BIOFLUIDDYNAMICS	
1. General description .....	1
2. External biofluidynamics .....	3
3. Internal biofluidynamics .....	5
PART I. EXTERNAL BIOFLUIDDYNAMICS	
<b>Chapter 2</b>	
HYDROMECHANICS OF AQUATIC ANIMAL PROPULSION: A SURVEY	
1. Scope of the survey .....	11
2. Aquatic propulsion in twelve classes of invertebrates .....	12
3. Introduction to fish propulsion .....	15
4. Anguilliform propulsion .....	19
5. Introduction to the carangiform mode .....	22
6. Measurements of the carangiform mode .....	25
7. The carangiform mode with lunate tail .....	26
8. Two-dimensional analysis of the lunate tail .....	29
9. Need for more detailed study of fast percomorph fishes .....	32
10. Fishes with weight unbalanced by buoyancy .....	34
11. Propulsion in 'armoured' fishes .....	37
12. Limbs in swimming: amphibians and arthropods .....	38
13. Aquatic propulsion in reptiles .....	39
14. Aquatic propulsion in mammals .....	41
15. Conclusion .....	42
<b>Chapter 3</b>	
MATHEMATICS OF AQUATIC ANIMAL LOCOMOTION AT LOW REYNOLDS NUMBER	
1. Introduction .....	45
2. The fluid equations and their fundamental singular solutions .....	45
3. Line distributions of singular solutions .....	49
4. Resistive-force theory of flagellar propulsion .....	53
5. Swimming motions with minimum rate of working .....	59
6. A more accurate slender-body theory .....	62

**Chapter 4****AQUATIC ANIMAL PROPULSION OF HIGH HYDROMECHANICAL EFFICIENCY**

1. Introduction .....	67
2. The pure anguilliform mode of propulsion .....	72
3. Vortex sheets shed by fins .....	79
4. Mechanics of the carangiform mode .....	85
5. Two-dimensional theory of the lunate tail .....	92

**Chapter 5****LARGE-AMPLITUDE ELONGATED-BODY THEORY OF FISH LOCOMOTION**

1. Introduction .....	103
2. Calculation of the reactive force .....	106
3. Discussion of the thrust-drag balance in relation to observations ..	111

**Chapter 6****AQUATIC ANIMAL LOCOMOTION: A SURVEY OF RECENT THEORETICAL DEVELOPMENTS**

1. Introduction .....	117
2. The subdivisions of hydrodynamic theory relevant to aquatic animal locomotion .....	119
3. Hydrodynamics of ciliary propulsion .....	126
4. Large-amplitude elongated-body theory .....	132
5. Vortex wakes .....	137

**Chapter 7****SOME CURRENT INVESTIGATIONS OF AQUATIC ANIMAL MOTIONS**

1. Developments at low Reynolds numbers .....	141
2. Developments at high Reynolds numbers .....	144

**Chapter 8****ANIMAL FLIGHT**

1. The conquest of the air .....	151
2. Sustained forward flight of an insect .....	155
3. Bird forward flight .....	159
4. Hovering flight .....	169

**Chapter 9****ON THE WEIS-FOGH MECHANISM OF LIFT GENERATION**

1. Introduction .....	179
2. Two-dimensional inviscid-flow theory .....	181
3. Modifications due to viscous effects .....	186
4. Conclusion .....	190

## PART II. INTERNAL BIOFLUIDDYNAMICS

### **Chapter 10**

#### **PHYSIOLOGICAL FLUID DYNAMICS: A GENERAL SURVEY**

1. Introduction .....	199
2. Steady secondary flows .....	201
3. Entry regions .....	202
4. Incipient atheroma .....	203
5. Distribution of shear in branched systems .....	203
6. Distribution of resistance in branched systems .....	204
7. Bronchial resistance .....	205
8. Velocity distributions in pulsatile flow .....	206
9. Pulse propagation .....	206
10. Turbulence in the blood stream .....	207
11. Urinary tract .....	208

### **Chapter 11**

#### **RESPIRATORY FLOW PATTERNS**

1. Introduction .....	211
2. Flow patterns in human bronchi .....	214
3. Lung flow patterns in birds .....	217

### **Chapter 12**

#### **PULSE PROPAGATION THEORY**

1. Introduction .....	227
2. Wall stress systems associated with pulse propagation .....	229
3. Propagation through a junction .....	232
4. Interaction between junctions in a branching system .....	237
5. Amplitude gradations along a single tube .....	241
6. Some comparisons with experiment .....	246

### **Chapter 13**

#### **BLOOD FLOW AND ARTERIAL DISEASE**

1. Observations of turbulence in blood flow .....	253
2. Poststenotic dilatation .....	255
3. Aneurysms in the cranial circulation .....	256
4. Atherogenesis by intercellular lipid deposition .....	259
5. Intracellular lipid accumulation .....	262

### **Chapter 14**

#### **THE MICRO CIRCULATION**

1. Introduction .....	269
2. Vasomotor control of peripheral perfusion .....	270
3. Pulmonary perfusion and ventilation .....	272
4. Axial concentration .....	274
5. Lubrication problems in very narrow capillaries .....	276

*This page intentionally left blank*

## Acknowledgements

It is a pleasure to thank many friends, and in particular Richard Bainbridge, John Blake, Quentin Bone, Colin Caro, Jim Fitz-Gerald, James Gray, Ken Machin, Norman Marshall, Tim Pedley, Margot Roach, Bob Schroter, Tony Seed, Daniel Weihs and Torkel Weis-Fogh for their much valued help and cooperation generously given at various times during the five years of preparation of the material in this book. My sincere thanks are due also to Dick DiPrima and Les Rubenfeld for arranging so admirably the 'Mathematical Biofluidynamics' Research Conference at Rensselaer Polytechnic Institute, as well as to the National Science Foundation for sponsoring it.

JAMES LIGHTHILL

*This page intentionally left blank*

## CHAPTER 1

# Introduction to Biofluidynamics

### 1. General description

It can be argued that two important practical aims have underlain most researches in fluid dynamics:

- (i) improved locomotion through fluid media,
- (ii) enhanced understanding of transport by fluids in circulating systems.

Researches with these aims over many years led to (i) the design of efficient vehicles for locomotion in water or through the air, and (ii) better analysis of how great circulating systems (the atmosphere, the hydrologic cycle, or fluid-flow systems in industrial plant) transport heat or various substances in solution or suspension.

Biofluidynamics has a similar duality of purpose, reflected in the design of this book. *External* biofluidynamics is concerned with animal locomotion through surrounding fluid media (water and air), while *internal* biofluidynamics is concerned with heat and mass transport by fluid flow systems within an animal. Nevertheless, these two halves of biofluidynamics have in common one general characteristic that distinguishes them from the rest of fluid dynamics: the fluid movements are energized by the working of an animal's motile external or internal surfaces, parts of which are highly flexible. Our clear limitation of the field to animal systems (which perhaps might make more suitable the name 'zoofluidynamics') with their characteristic *motility* (ability to make actively controlled movements in response to stimuli) gives it this unifying feature.

The purposes of the research are mainly biological, including pure and applied aspects. The fundamental biological goals are to understand not only the functioning, but also the evolution, of the various systems for locomotion of animals within external fluids or for mass transport and heat transport by internal fluid flows. Features of such systems which, in the animal's environment, slightly increase its chance of surviving and producing offspring can of course appear in slightly greater numbers in the next generation, and greater still in the next, with long-term effects of progressive incorporation into the characteristics of a species. The pathways by which this evolution may have proceeded to meet gradually changing environments or fill vacant ecological niches can gradually be traced through a combination of evidence from general biology and palaeontology and from biofluidodynamic research.

Such work, though of primarily biological interest, has nevertheless a secondary interest to engineers because progressive improvement by animals of solutions to particular fluid-dynamical problems posed by environment or ecology is sometimes found on analysis to have led to remarkable results that suggest new engineering possibilities. The applied arts founded on biology benefit also from our studies:

medicine needs information from the internal biofluidynamics of man (and of those mammals that are used as experimental animals for medical research), and the wise use of natural resources demands inputs from many parts of biofluid-dynamics into fisheries research, pest-control research and veterinary research.

Progress towards biofluidodynamic goals can occur only when workers from the biological sciences and from the mathematical and engineering sciences work closely together. Complex movements of fluids energized by motile surfaces, parts of which are highly flexible, demand for their analysis professional fluid-dynamic skills founded on the mathematical and engineering sciences but the analyses required must be carefully defined and interpreted in a full and intricate biological context. Effective biofluiddynamic research demands therefore collaboration between mathematicians/engineers and physiologists/zoologists, and a survey of such researches is necessarily a story of particular collaborations along these general lines. This kind of survey, from a particular mathematician's standpoint, of various such collaborations in which he personally has been involved (either closely or more remotely) is offered in the present book.

The value of seeing any biofluidodynamic problem, whether concerning locomotion in external fluids or transport of internal fluids, against the background of a systematic comparative survey of the biological function in question in many different groups of animals, can hardly be overestimated: such a systematic approach is essential for answering fundamental zoological questions of how biofluidodynamic mechanisms evolved, and also is an important source of the understanding needed to tackle practical problems in medical, veterinary and conservation science. In preparing an account of mathematical biofluidynamics, however, there may be difficulties in finding the right balance between zoological-survey material and analytical studies of fluid motions.

That balance has been struck in the present book by preceding the account of the *first* of the four main subjects treated by an extensive comparative survey (Chapter 2). That systematic survey of aquatic animal locomotion, put forward to indicate the broad background against which the selection and treatment of analytical problems has to be made, is followed by two chapters (3 and 4) which can be viewed as mathematical supplements, giving the analytical reasoning (for motions at low and high Reynolds numbers respectively) behind much of what is stated in the comparative survey. Chapters 5, 6 and 7 then describe recent developments in the analysis of aquatic locomotion.

With later subjects, by no means such a full comparative survey has been attempted. In the hope that Chapter 2 has indicated the general spirit of how biofluiddynamic studies need to be approached, the comparative background to those later subjects is set out more briefly so that analysis can begin earlier. Thus, Chapter 8 combines a shorter comparative survey of animal flight, primarily among the insects and the birds, with preliminary analysis of its biofluidodynamic aspects. For large areas of the subject, classical aerodynamic principles are found to be adequate to explain the effectiveness of the observed motions. Some exceptional areas, however, require the development of quite new analysis, and one example of this is given in detail in Chapter 9.

Part II of the book, concerned with the internal biofluidynamics of respiration and circulation, includes even less comparative physiology and, furthermore, confines its attention exclusively to the vertebrates. The preliminary survey (Chapter 10) is devoted largely to comparing the patterns of internal fluid flow occurring in different parts of the human respiratory and circulatory systems with each other and with internal-flow patterns known to engineering science. Chapter 11, on the other hand, seeks to compare respiratory flow patterns in the different classes of vertebrates, showing for example the quite different types of analysis needed for birds' lungs and mammals' lungs.

Finally, circulatory biofluidynamics is developed in Chapters 12, 13 and 14 exclusively for mammalian cardiovascular systems. Here the practical applications of the analysis are stressed, particularly in Chapter 13 on the interactions between blood flow patterns and arterial diseases, for which it has been found necessary to describe the biofluidynamic problems in a broad medical and biochemical context. Again, circulatory flow patterns in physiologically normal conditions are emphasized in Chapters 12 and 14, for large and small blood vessels respectively.

The preparation of this book has taken five years. Certain parts of it (Chapters 2, 4, 5, 10 and 14) had already been published as separate articles before the present Research Conference was planned. The material in question appeared in a suitable form, however, for later use as constituent parts of a broad survey of biofluidynamics. The author expresses his particular gratitude to Annual Reviews, Inc. for permission to reprint Chapter 2, and also to the International Centre for Mechanical Sciences at Udine, Italy for permission to reprint Chapters 10 and 14 which were originally given as lectures at the International Centre.

Two other chapters were written with a dual purpose, partly for purposes of the Research Conference and the resulting book and partly for other purposes: Chapter 6 as a General Lecture at the 13th International Congress of Theoretical and Applied Mechanics and Chapter 9 as a paper in the Journal of Fluid Mechanics. The remaining seven chapters (1, 3, 7, 8, 11, 12 and 13) have been written expressly for the Research Conference and the resulting book to complete as full as possible an account of present-day mathematical biofluidynamics as viewed by the present author.

Careful consideration was given in finalizing the book to the decision whether or not to make minor modifications to those articles that had been or were being published elsewhere. It was finally concluded that this was undesirable and unnecessary: undesirable, because it would be confusing to readers generally if the same article existed in two forms that differed to a minor degree, and unnecessary because all the points supplementing that material which it was thought desirable to make could be satisfactorily inserted into the new chapters.

## 2. External biofluidynamics

The necessary mathematical background for external biofluidynamics is set out in the two companion volumes edited by Thwaites (1960) and Rosenhead (1963) on the fundamentals of incompressible flow. The necessary biological background on animal locomotion is authoritatively expounded by Gray (1968).

Study of locomotion involves investigation of propulsion and control. For a general survey of aquatic animal propulsion and the associated fluid dynamics, see the general review in Chapter 2, compiled by the author in 1969 in collaboration with scientists in the Cambridge University Department of Zoology, the British Museum of Natural History (London) and the Marine Biological Association (Plymouth).

A general mathematical method useful in this field is the elongated-body theory, analogous to slender-body theory in aerodynamics. It depends on the idea that each cross-section of an elongated body interacts with the water in a manner dependent on its velocity *relative* to the undisturbed fluid. In the low-Reynolds-number limit (Chapters 3 and 7) the force between the cross-section and the fluid, or in the high-Reynolds-number limit (Chapters 4 and 5) the fluid momentum attached to the cross-section, depends on that relative velocity, leading respectively to theories of 'resistive' or 'reactive' type.

Further problems at high Reynolds number are posed by the crescent moon tail propulsion system known as the 'lunate tail' (Marshall 1965, and Chapter 2) used by most of the fastest aquatic animals. This high-aspect-ratio propulsion system is suitable for analysis by airfoil theory (Chapter 4). Numerous ideas from mathematical aerodynamics prove valuable, also, in analysing the motion of many groups of aquatic animals *heavier than water* (Chapter 2).

Further problems at low Reynolds number include those of ciliary motion: the coordinated waving of large numbers of microscopic elongated bodies known as cilia attached to the surface of an organism. Studies by a Cambridge mathematician (Blake 1971, 1972), inspired by experimental work of protozoologists at London and Bristol Universities, produced a valuable statistical picture of their combined fluid dynamic effects in the presence of the organism's surface (Chapter 6). This work on propulsion of ciliated organisms is now beginning to throw a bridge over into Internal Biofluidynamics by extension to the study of how tubes lined with waving cilia bring about transport of particular fluids within various organisms (Chapter 7).

Not only propulsion problems but also control problems of great interest arise at high Reynolds number in studies of fish manoeuvrability (Marshall 1965, Weihs 1972, and Chapters 6 and 7). By contrast, little is known concerning control of a single-celled organism's movements, and it will be important in the future to direct biofluidynamic research in protozoan locomotion towards studies that will throw light on control mechanisms.

Fast forward flight by an *insect* (whose impermeable wings lend themselves to aerodynamic analysis) has been fruitfully studied in collaborative work between a zoologist and an aerodynamicist (Weis-Fogh & Jensen 1956). Several interesting additional problems arise in the forward flight of birds, both due to the complex properties of wings constructed with feathers and due to birds' utilisation of wind shears (see Brown 1948 & 1953, Gray 1968). Chapter 8 describes those now classical investigations and reports also some very new work, which goes beyond studies of fast forward flight, and raises matters of rather general fluiddynamic interest.

This includes a broad study of hovering animals (including birds, bats and many orders of insect) by Weis-Fogh (1972, 1973) with some mathematical collaboration by Lighthill. The vast majority of hovering animals support their weight by motions that lend themselves to analysis on classical aerodynamic principles. There are important exceptions, however, where substantially better performance is achieved by use of mechanisms of lift generation never discovered by engineers: one of these mechanisms (used by the small wasp *Encarsia formosa*) is described in full mathematical detail in Chapter 9.

### 3. Internal biofluidynamics

The internal biofluidynamics of the human body is usefully surveyed in the book edited by Wolstenholme & Knight (1969) and also in part of the book edited by Fung, Perrone & Anliker (1972), which goes also into 'internal biosolid-mechanics'. Both are reports of valuable and interesting interdisciplinary symposia with medical, physiological, engineering and mathematical participants. Again, Part II of this book describes many parts of the field into which the author has been drawn through collaboration with the interdisciplinary Physiological Flow Studies Unit of London University under Dr. Caro's direction.

Caro (1966) edited a book with many interesting articles describing aspects of the physiology and mechanics of mammalian respiration. More recent work in this field by the Physiological Flow Studies Unit is described in articles by Pedley et al. (1970, 1972). Several further aspects of air flow and blood flow in the lungs and the gas exchange between them are reported and, where possible, analysed mathematically, in Chapters 11 and 14.

There are interesting biofluidynamic problems of respiration in all classes of vertebrates: see Piiper (1972) for a general survey of the field. Many such problems are picked out for discussion in Chapter 11. Flows in and out of fish gills are discussed: these again form a bridge to external fluid dynamics since they influence the boundary layer on the fish's external surface (Chapter 2). Parts of Piiper's survey concerned with bird lungs are particularly interesting fluidodynamically: there is a *rectification*, bringing about one-way flow through more than half of the lung. This phenomenon is discussed in detail following a 1973 interdisciplinary symposium on the subject.

Important advances in the analysis of the complex motions in the cardiovascular system of mammals were described by Womersley (1955) and McDonald (1960), a mathematician and a medical physiologist who worked in close collaboration on these problems. There are many difficulties involved in analysing details of how a complex fluid, blood, is induced by the heart's pulsations to circulate through an intricate branching system of elastic tubes. Womersley and McDonald worked principally on the 'local' biofluidynamics within one or two tubes and branches, while their colleague Taylor (1966) extended the work to an overall view of the system as a whole (see especially Chapter 12).

Several attempts to improve on those now classical researches are described in Part II. Important geometrical aspects there neglected are indicated in Chapter 10: for example, effects of curvature (as in the arch of the aorta) which may lend

themselves to treatment in terms of secondary-flow theory (McConalogue 1970), or various entry-length effects (now being studied).

There are many interactions, some of which are discussed in Chapter 13, between these fluydynamic problems and problems of cardiovascular disease. Atherosclerosis, for example, has been linked in more than one way with details of the distribution of blood flow: some preliminary idea of these connections is given by Caro et al. (1971). A rather detailed account of these different connections based on a recent interdisciplinary symposium chaired by the author is given in Chapter 13.

Finally, we return (Chapter 14) to low-Reynolds-number fluydynamics in the problems of the microcirculation: both the pulmonary microcirculation (which interacts with the respiratory process) and the systemic microcirculation. These are where the particulate character of blood, as a suspension of red-cells and other bodies, is of very special significance. The collections edited by Copley (1968) and by Wolstenholme & Knight (1969) contain many articles concerned with how the particles redistribute themselves in blood flow through arterioles and capillaries. They include some studies of the exchange of gas and nutrients in the peripheral circulation or of gas in the lung, much of which occurs in capillaries so narrow that red-cells pass through them in Indian file, often in a considerably deformed condition. On this type of low-Reynolds-number internal biofluydynamics, which has been the subject of many mathematical studies (Fitz-Gerald 1969, Aroesty & Gross 1970), a concluding account of various recent developments is given.

The present author, as a lifelong devotee of fluydynamics, has attempted in this book to demonstrate that during the past two decades there has come of age a new major division of the subject: biofluydynamics, to join all the other divisions such as geofluydynamics, technofluydynamics and so on. Within the limits of a short book he has only been able to scratch the surface of this important new field, but hopes nevertheless to have indicated how much may be lying there beneath the surface for those who are prepared to dig down deep.

## REFERENCES

- AROESTY, J. & GROSS, J. 1970 *Convection and Diffusion in the Microcirculation*. Rand Corp. Rep. RM-6214-NIH.
- BLAKE, J. R. 1971 *A note on the image system for a Stokeslet in a no-slip boundary*, Proc. Camb. Phil. Soc. **70**, 303–310.
- , 1972 *A model for the micro-structure in ciliated organisms*, J. Fluid Mech. **55**, 1–23.
- BROWN, R. H. J. 1948 & 1953 *The flight of birds*, J. Exptl. Biol. **25**, 322–333 & **30**, 90–103.
- CARO, C. G. (ed.) 1966 *Advances in Respiratory Physiology*, Edward Arnold, London.
- CARO, C. G., FITZGERALD, J. M. & SCHROTER, R. C. 1971 *Atheroma and arterial wall shear*, Proc. Roy. Soc. B, **177**, 109–159.
- COPLEY, A. L. (ed.) 1968 *Hemorheology*, Pergamon, London.
- FITZ-GERALD, J. M. 1969 *Mechanics of red-cell motion through very narrow capillaries*, Proc. Roy. Soc. B, **174**, 193–227.
- FUNG, Y. C., PERRONE, N. & ANLICKER, M. (ed.) 1972 *Biomechanics: its Foundations and Objectives*, Prentice-Hall, Englewood Cliffs, N.J.
- GRAY, J. 1968 *Animal Locomotion*, Weidenfeld & Nicolson, London.
- MARSHALL, N. B. 1965 *The Life of Fishes*, Weidenfeld & Nicolson, London.

- McCONALOGUE, D. J. 1970 *The effects of secondary flow on the laminar dispersion of an injected substance in a curved tube*, Proc. Roy. Soc. A, **315**, 99–113.
- MCDONALD, D. A. 1960 *Blood Flow in Arteries*, Edward Arnold, London.
- PEDLEY, T. J., SCHROTER, R. C. & SUDLOW, M. F. 1970 *Energy losses and pressure drop in models of human airways*, Respiration Physiol. **9**, 371–386.
- PEDLEY, T. J., SUDLOW, M. F. & MILIC-EMILI, J. 1972 *A non-linear theory of the distribution of pulmonary ventilation*, Ibid. **15**, 1–39.
- PIPER, J. (ed.) 1972 *Proceedings of the Symposium “Comparative Physiology of Respiration in Vertebrates.”* Ibid. **14**, 1–236.
- ROSENHEAD, L. (ed.) 1963 *Laminar Boundary Layers*, Oxford University Press.
- TAYLOR, M. G. 1966 *Wave transmission through an assembly of randomly branching elastic tubes*, Biophys. J. **6**, 697–716.
- THWAITES, B. (ed.) 1960 *Incompressible Aerodynamics*, Oxford University Press.
- WEIGHS, D. 1972 *A hydrodynamical analysis of fish turning manoeuvres*, Proc. Roy. Soc. B, **182**, 59–72.
- WEIS-FOGH, T. & JENSEN, M. 1956 *Biology and physics of locust flight*, Phil. Trans. Roy. Soc. B, **239**, 415–584.
- WEIS-FOGH, T. 1972 *Energetics of hovering flight in hummingbirds and drosophila*, J. Exptl. Biol. **56**, 79–104.
- , 1973 *Quick estimates of flight fitness in hovering animals, including novel mechanisms for lift production*, Ibid. **59**, 169–230.
- WOLSTENHOLME, G. E. W. & KNIGHT, J. (ed.) 1969 *Circulatory and Respiratory Mass Transport*, J. & A. Churchill, London.
- WOMERSLEY, J. R. 1955 *Method for the calculation of velocity, rate of flow and viscous drag in arteries when the pressure gradient is known*. J. Physiol. **127**, 553–563.

*Note added in proof.* Edward Arnold published a revised 2nd edition of *Blood Flow in Arteries* by D. A. McDonald during 1974, when also an article *Aerodynamic aspects of animal flight* by M. J. Lighthill, giving a more detailed comparative survey of that subject, was scheduled to appear in the Tenth Anniversary Number of the Bulletin of the Institute of Mathematics and its Applications.

*This page intentionally left blank*

PART I  
EXTERNAL BIOFLUIDDYNAMICS

*This page intentionally left blank*

## CHAPTER 2

# Hydromechanics of Aquatic Animal Propulsion—A Survey\*

### 1. SCOPE OF THE SURVEY

This is a survey of the hydromechanics of aquatic animal propulsion. It includes no discussion of the aquatic movements of animals that are mainly nonaquatic (e.g., the feeble swimming attempts of *Homo sapiens*). Conversely, there is no reference to how normally aquatic animals move in environments that are wholly or partly nonaquatic; for example, to the creeping and burrowing movements of bottom-living creatures, or to locomotion upon the surface, or to any of the special problems of normally aquatic animals that temporarily leave the weight-supporting aqueous environment to walk, leap, glide, or fly.

The mechanical discussion concentrates similarly on the aqueous medium, and on its interactions with the movements of the animal's external (and, where relevant, internal) surfaces. By contrast, little is said either about the internal mechanics of the animal, or about means of coordination and stimulation of those movements (which include the sense organs and neural channels of communication as well as the muscles and other motile organs). Within the hydromechanical field, furthermore, the survey concentrates on the problems of propulsion at the expense of related problems of stability and control.

The subject so delimited is, however, of particular hydromechanical interest, because about  $10^9$  years of animal evolution in an aqueous environment, by preferential retention of specific variations that increase ability to survive and produce fertile offspring, have inevitably produced rather refined means of generating fast movement at low energy cost, which merit study. This is mainly because, at every point in the food chain, or network of predator-prey relationships, capability of movements that are fast relative to those of predator or prey, or to the distances that must be traversed before prey is found, is one out of the many important factors influencing such ability to survive.

Some of the other factors, to be sure, may in certain ecological situations be of overriding importance, so that evolution proceeds in a direction (for example, development of protective armour) that may reduce mobility although net chance of survival is improved. This survey notes several such cases (see especially Section 11), but concentrates principally on those extensive portions of the animal kingdom where capability of motion in

\* Reprinted with permission, from the Annual Review of Fluid Mechanics, Volume 1, pages 413–446. Copyright © 1969 by Annual Reviews Inc. All rights reserved.

water at speeds comparable with those of other animals of similar mass has successfully been achieved and retained.

It is likely that most readers may possess specialised knowledge in either zoology or the mechanics of fluids, but not both. The survey is written as far as possible to be equally intelligible to either kind of reader.

## 2. AQUATIC PROPULSION IN TWELVE CLASSES OF INVERTEBRATES

Out of the phyla, or major divisions of the animal kingdom (in number, on most systems of classification, around twenty), we discuss representatives of ten (see Table I), including first of all the Protozoa. Among these single-celled creatures we draw special attention to the class Mastigophora which propel themselves by undulatory motions of a whip-like flagellum, and whose structure is reflected in that of the spermatozoa of higher animals. A continuing theme of the survey is the success of this basic undulatory mode of propulsion, and of various modifications of it that are found in practically all the aquatic animals that are most successfully mobile.

Relatively less success, in the evolutionary sense (partly for mechanical reasons set out below), has been achieved by the two other main means of aquatic animal propulsion to which we refer. One of these (which would repay further hydromechanical study) is propulsion by movements of large numbers of attached cilia [see especially Gray (1)], as found already within the Protozoa in the class Ciliophora. Also, within the phylum Coelenterata (which includes the jelly-fish and sea-anemones) members of one class, the Ctenophora or comb-bearers, possess a well-developed ciliary means of propulsion in the form of eight combs of waving cilia. Again, most members of the phylum Rotifera are propelled by means of a usually double ring of cilia around the oral cavity.

The second main alternative to undulatory propulsion in water, besides ciliary movement, is propulsion by jet reaction. The medusoid forms of two classes of Coelenterata, the Hydrozoa and Scyphozoa, have the characteristic 'parachute' jellyfish shape, and by contraction of the circular subumbrial muscles can expel water from the subumbrial cavity and thereby slowly propel themselves in the opposite direction. Similar propulsion by jet reaction is found, within the phylum Mollusca comprising animals at a rather higher level of organisation, in the class Cephalopoda, the squids, which in cases where the jet is concentrated in a narrow funnel and the body is well streamlined may travel at quite high speeds (2 to 4 m/s according to size).

Most cephalopods (particularly the decapod squids) possess also fins, which are given an undulatory mode of motion for manoeuvring at relatively low speed, whereas a sudden retreat is made possible by a jet of water forcibly expelled from the forward-facing funnel. In certain squids, however, the habitual mode of swimming is by periodic rapid expulsion, through the funnel, of fluid that has been slowly drawn into the mantle cavity. Siekmann (2) has investigated this from the hydromechanical point of view, but classi-

TABLE I

Phylum	Classes referred to	Section in which discussed
Protozoa	Mastigophora	2
	Ciliophora	2
Coelenterata	Hydrozoa	2
	Scyphozoa	2
	Ctenophora	2
Rotifera	Rotifera	2
Nematoda	Nematoda	2
Nemertini	Nemertini	2
Annelida	Polychaeta	2
	Hirudinea	2
Arthropoda	Insecta	12
	Crustacea	12
Mollusca	Cephalopoda	2
Chaetognatha	Chaetognatha	3
Chordata	Urochordata	2
	Cyclostomata	3
	Pisces	3 to 11
	Amphibia	12
	Reptilia	13
	Mammalia	14

Table I. To assist readers grounded in fluid mechanics rather than biology, the ten phyla from which animals discussed in the text are drawn are listed on the left. Within each phylum, it is only animals in the classes listed against it on the right that are discussed. For amplification of the meaning of the terms used, see text. The main phyla excluded, as largely irrelevant to the subject of the survey, are Porifera (sponges), Platyhelminthes (tape-worms and related animals), Polyzoa (animals forming sea-mats and other sessile colonies), Brachiopoda (which are all sessile solitary animals with bivalve shells) and Echinodermata (the large phylum of bottom-living animals that includes the starfish and sea-urchins).

cal propulsion hydromechanics already tells us that only rather a small proportion of the energy devoted to making this very narrow jet can be yielding useful propulsive work.

In addition, within the phylum Chordata which contains all the verte-

brates and other animals based on an internal skeletal rod, members of the class Urochordata (the sea-squirts and salps, headless and accordingly invertebrate and indeed mostly losing the skeletal rod in adult life) characteristically draw in water through a branchial cavity and expel it through an atrial cavity, the latter expulsion being, in free-swimming solitary members of the order Thaliacea, of sufficient force to produce substantial movement. Propulsion by jet reaction is, however, rare among the vertebrates, if the view expressed in Section 9 be accepted that such dynamical role as is played by expulsion of water from the gills of fish is mainly reduction of drag by boundary-layer control.

As already stated, the undulatory mode must, along with its modifications described below, be regarded as the principal means of aquatic animal propulsion. One reason for this is its relative insensitivity to changes of scale; in the language of fluid mechanics it operates well at all values of the Reynolds number

$$R = Ul/\nu \quad 1.$$

where  $U$  and  $l$  are the speed and length of the animal in mm/s and mm, while  $\nu$  is the 'kinematic viscosity' (viscosity divided by density) which is between 1 and 2 mm<sup>2</sup>/s in most naturally occurring aqueous media.

Thus, although as a means of propulsion jet reaction is appropriate only when  $R$  is rather large (so that inertia dominates over viscosity) and ciliary movement only when  $R$  is rather small, the undulatory mode gives satisfactory results for all values of  $R$ . This has permitted its use to spread through animals in all ranges of size. A typical Reynolds number for a flagellate single cell is about 10<sup>-3</sup>. The motion takes the form of a transverse wave passing down the flagellum from head to tail at velocity  $V$  say, with amplitude increasing towards the tail. The consequential forward velocity  $U$  of the animal [Gray & Hancock (3)] is of the order 0.2  $V$ .

Similar results with  $R$  of order 1 are found within the phylum Nematoda (the thread-worms). Although the majority of Nematoda are parasitic, and a few of these attain great lengths (even exceeding 1 m), it is the quite large minority of free-living nematodes with which we are here concerned; their lengths are around 1 mm. They swim [Gray & Lissmann (4)] by passing waves down their body at a speed  $V$  of order 1 mm/s giving a Reynolds number of order 1. The resulting forward velocity can be as much as 0.4  $V$ , as in Figure 1; one reason for the greater value than in single-celled animals is that there is no relatively large head to propel.

For similar behaviour at Reynolds numbers of order 10<sup>3</sup>, we may turn to the aquatic members of the phylum Annelida (worms regularly segmented) with characteristic lengths of order 100 mm and swimming speeds of order 10 mm/s. For example, the largely smooth-surfaced members of the class Hirudinea (leeches), when swimming, pass down their body transverse waves with amplitude increasing towards the rear, and travel forward [Gray (5)]

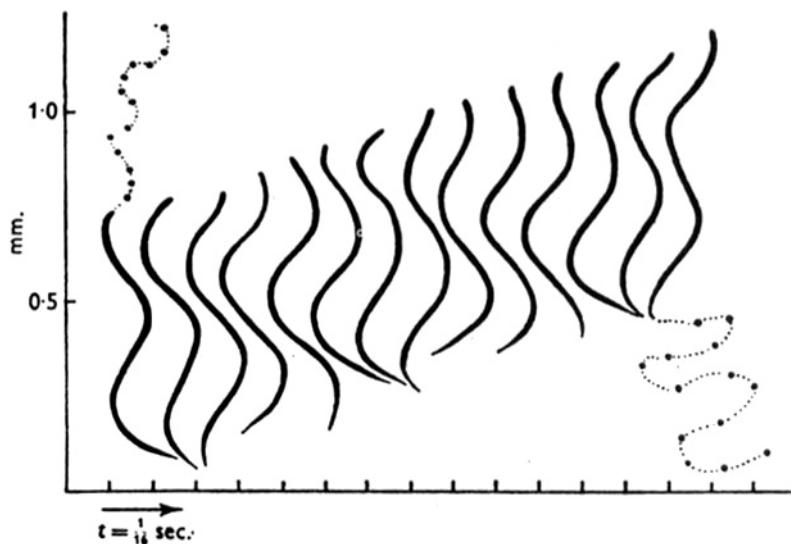


FIG. 1. Successive positions, at intervals of 0.0625s, of a nematode (*Turbatrix aceti*) swimming in water. The amplitude of transverse movements of the tail was almost four times that of the head. Note backward movement of waves relative to the water [from Gray & Lissmann (4)].

at a speed around  $0.3 V$ , and Gray noted similar behaviour in free-swimming members of the phylum Nemertini (the unsegmented worms with lengthy retractile proboscis). It is, however, when we inspect aquatic members of another annelid class, the Polychaeta, that we find, in those animals with lengthy parapodia attached to each segment, our first major modification of the undulatory mode of propulsion.

These polychaete worms have, in fact, put the undulatory mode into reverse [Gray (6)], so that the transverse wave is propagated along the body from tail to head, again with increasing amplitude, but the animal as a consequence moves in the direction of propagation of the wave; we may say, in the notation used above, that the induced velocity  $U$  of the animal is about  $-0.2 V$ . The same reversal is found both in genera like *Nereis* (Figure 2) where the parapodia appear to participate in the swimming action (those on any one side being more extended at crests than at troughs of the wave) and in genera like *Nephthys* where the parapodia seem to be passive. Taylor (7) first explained mechanically how this reversal of the wave is necessitated (see Section 13) by the presence of the parapodia.

### 3. INTRODUCTION TO FISH PROPULSION

On the other hand, hydrodynamic analysis of undulatory propulsion in relatively smooth and elongated animals at still higher Reynolds number (by the 'elongated-body' theory; see Section 4) shows the great advantage of

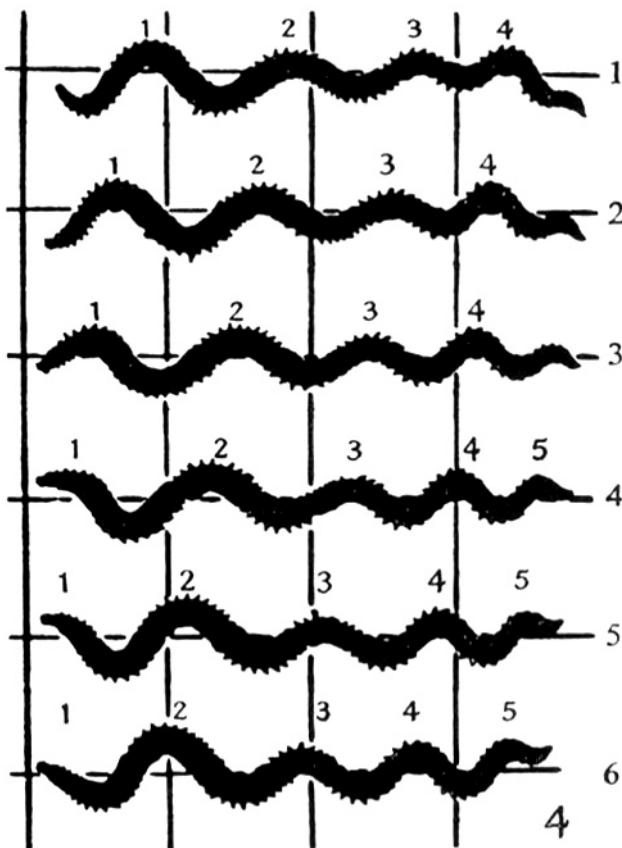


FIG. 2. Successive positions, at intervals of 0.05s, of a polychaete worm (*Nereis diversicolor*) swimming to the left. Numbers attached to moving 'crests' show that the wave travels in the same direction as the worm, although considerably faster [Gray (6)].

transverse 'compression' (that is, flattening) at the posterior end of the animal, so that the end takes the form of a broad edge at right angles both to the length of the animal and to the transverse motion which passes down it in the undulatory mode. We speak here of 'transverse' compression rather than 'lateral' compression (the form it takes in fish, amphibia and reptiles) to emphasize that the direction of flattening is necessarily that of the back-and-forth motion that is being propagated. Where that motion is dorso-ventral, that is, vertical (as in cetacean mammals and some crustacea), the posterior flattening is dorso-ventral.

The very early vertebrate forms were already showing this advantageous transverse compression at the posterior end (rare among the invertebrates, although present within the small isolated phylum Chaetognatha, the marine

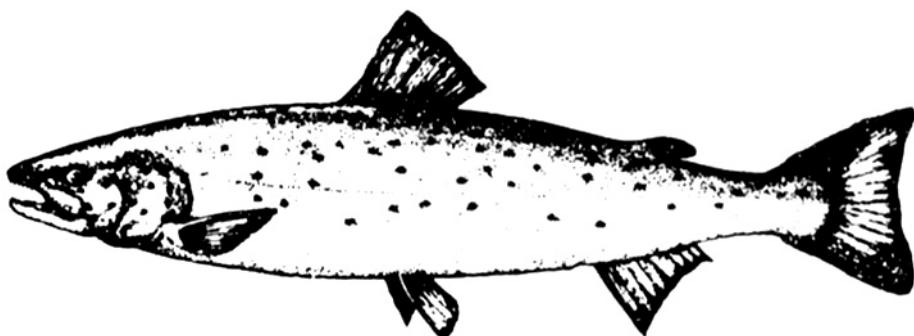


FIG. 3. A specimen of the salmon *Salmo salar*, to illustrate the description of fin arrangements in fishes [from Marshall (8)].

arrow-worms). In particular, the jawless vertebrates, represented today by the class Cyclostomata (lampreys and hag-fishes), possess a continuous median fin, stretching both dorsally and ventrally around the posterior end; there, while the body cross-section is reduced in area almost to zero, the combined depth of fin and body is maintained (as in the eels; see Figure 4 page 422.) The advantage of this is very clear from the hydrodynamical 'elongated-body' theory. Members of the higher vertebrate class Pisces, with their usually more complicated arrangement of fins, maintain and sometimes improve on this advantage owing to the detailed characteristics of their caudal fins.

We think it necessary to give a more detailed account of methods of aquatic propulsion in fishes, the pre-eminent class within the aquatic environment, than we have given in relation to the other classes so far mentioned. Out of about thirty orders into which the class is subdivided [Marshall (8)], we discuss representatives of 17 (see Table II). Using them as illustrations, we first discuss fishes in which the undulatory mode of propulsion is little modified from that found in flagellates, nematodes, leeches and lampreys, and then describe modifications: (a) increasing concentration of the movement into the tail region, (b) development of tails of high aspect-ratio, (c) modifications special to fishes that are heavier than water or have lost flexibility owing to a covering of 'armour.'

At the outset, we may mention the typical fin arrangements in the class Pisces, and define the terms used to describe them (see Figure 3). The fins in the fish's plane of symmetry (typically vertical) are the caudal fin at the posterior end, which is normally the most important for propulsion, one or more dorsal fins which effectively extend the fish cross-section at its uppermost point, and similarly at the lowest point one or more fins that are usually called anal if they are near the posterior end of the fish or otherwise ventral. Evidently dorsal and ventral fins act in combination with the caudal fin to

TABLE II

Order	Examples	Section in which discussed
Selachii	sharks and dogfishes	10
Batoidei	rays, skates, and sawfishes	10
Dipnoi	lung-fishes	4
Palaeoniscoidea	sturgeons, etc.	10
Isospondyli	herring, pilchard, salmon, trout, etc.	6
Haplomi	pike, etc.	9
Ostariophysi	carp, minnow, dace, bream, tench, etc.	6
Apodes	eels	4
Heteromi	spiny-eels, etc.	4
Lyomeri	gulper-eels	4
Allotriognathi	ribbon-fishes, unicorn-fishes, etc.	4
Anacanthini	cod, whiting, haddock, hake, etc.	4
Percomorphi	(see Table III)	7, 8, 9
Zeomorphi	John Dories, boar-fishes, etc.	11
Plectognathi	e.g., file-, puffer-, porcupine-, and sun-fishes	11
Solenichthyes	tube-mouths (e.g., sea-horses)	11
Heterosomata	flatfishes	10

Table II. Orders of the class Pisces discussed in the text. The first two belong to the sub-class Chondrichthyes (having cartilaginous skeletons). Within the sub-class Osteichthyes (having bony skeletons), the Dipnoi are members of the group (Crossopterygii) from which the Amphibia evolved, and the Palaeoniscoidea represent another archaic group, whereas the remaining orders belong to the most advanced fish group Teleostei, with the swim-bladder as an effective hydrostatic organ (except where it has become vestigial as in some bottom-living or fast-moving fishes). Certain orders are excluded from discussion as not specially relevant to the subject of the survey, e.g., the bottom-living Holocephali (chimaeras, etc.), Xenopterygii (cling-fishes) and Scleroparei (gurnards, sea-snails, etc.), and the highly specialized Discophthali (shark-suckers) and Pediculati (angler-fishes, etc.). Other are excluded for relative lack of prominence or interest from the point of view of propulsion: the Holosteii (another archaic group), the Synentognathi (of which the most interesting members are flying fishes), the Thoracostei (sticklebacks, interesting in many ways unconnected with propulsion), the Berycomorphi (lacking prominence only relatively to the huge order Percomorphi, which they resemble), and the Microcyprini, Giganuroidea, Cetunculi, and Miripinnati.

give stability against yawing disturbances, but we shall see (Section 5) that they can also contribute to improving propulsion efficiency.

Off the plane of symmetry fins naturally occur in pairs. The anterior pair, which is almost always important, consists of the pectoral fins; the posterior pair, which is not quite so regularly present and important, consists of the pelvic fins. Stability and control in pitch and roll can, perhaps,

be described as the commonest function of the paired fins [Harris (9)], but we shall note a wide range of other roles that the pectorals have acquired in different groups of fish (a range which would be even wider if propulsion in nonaquatic environments had been included in the survey).

As an immediate illustration of this, we may mention the special problems of stopping and starting (although later discussion is concerned mostly with propulsion at a steady speed). These special problems do not exist at low Reynolds number (about 1 or less), when viscosity is so dominant that an animal stops as soon as it ceases to propel itself, and begins to move at full speed with negligible time-lag as soon as the propulsive movements begin. By contrast, at high Reynolds number inertia is important; extra energy is required to accelerate a fish from rest; conversely, a moving fish of good 'streamlined' shape may glide for a considerable distance after its whole body has assumed a rigid form.

For starting at high speed from rest [Hertel (10), p. 160], and indeed also in relation to the problem [Gray (11), p. 76] of making a sudden change of direction, fish benefit greatly from a well-known property of muscle, that its power output for a short period of time (a few seconds) can exceed by a large factor (of the order of 4) the power output produced over an extended period (which is limited by oxygen supply). These two problems can be solved, therefore, by making normal propulsive or yawing movements with (for a short time) greatly increased force. Stopping, however, is not so easily accomplished, which is why fish that have developed rather flexible pectoral fins benefit greatly from being able to use them as an aircraft pilot uses his air brakes. Again, some fishes have problems of accurate 'station-keeping' in slowly and sometimes irregularly moving water containing stationary obstacles of complex shape, and employ the pectoral and other fins for the fine adjustments continuously required in doing this.

#### 4. ANGUILLIFORM PROPULSION

Ichthyologists, beginning with Breder (12), describe as 'anguilliform' (after the common eel, of genus *Anguilla*) the pure undulatory mode of propulsion in which the whole body participates (even though amplitude may increase substantially towards the tail). In the typical eels, of the large order Apodes, a continuous dorsal and anal fin over the posterior portion of the fish is present (as in the class Cyclostomata mentioned above) so that, while the body cross-section is reduced almost to zero at the tail, the 'span' (combined depth of fin and body) is maintained approximately constant (Fig. 4). A similar configuration is found in the order Dipnoi (the lung-fishes, specially interesting from an evolutionary standpoint).

One way in which the propulsive effectiveness of waves of lateral displacement propagated at a given speed  $V$  down such a shape (flattening into a vertical trailing edge at the posterior end) can be estimated [Lighthill (13)] is as follows. When the lateral velocity of the trailing edge is  $W$ , the water

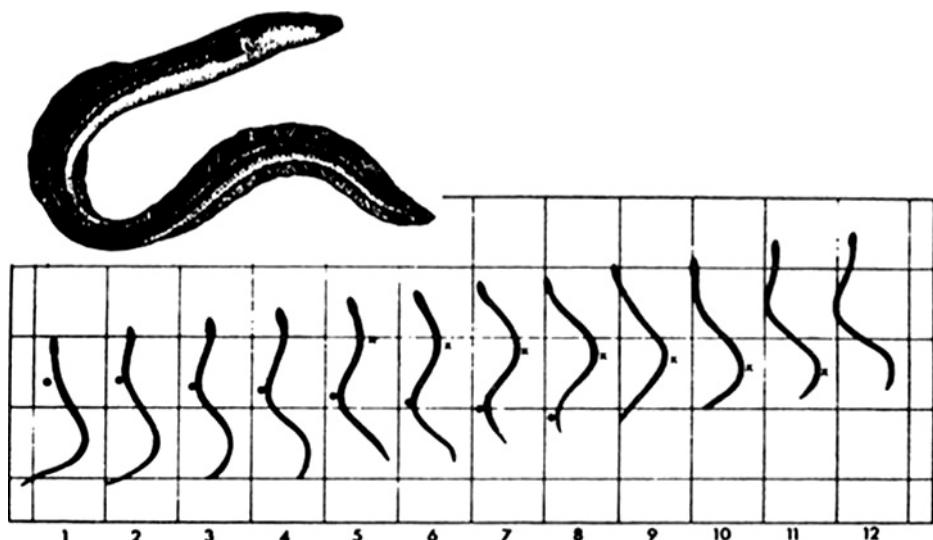


FIG. 4. The upper picture shows the common eel *Anguilla vulgaris*, while the lower picture shows successive positions, at intervals of 0.09s, of a young specimen swimming in water [Gray (11)]. The side of each square is 25.4 mm. Positions of maximum curvature are marked with dots and crosses to show backward movement of waves.

near the trailing edge does not feel itself being pushed to the side at speed  $W$  but at the lower speed

$$w = W(V - U)/V < W \quad 2.$$

where  $U$  is the net velocity of the fish (typically about 0.6  $V$  for anguilliform propulsion). This formula is explained in Figure 5, but we may note that  $w$  must obviously get smaller and smaller as  $U$  tends towards  $V$ , because in the limiting case when  $U = V$  the fish would be slipping through the water without giving it any lateral displacement at all.

The undulating fish as it moves through the water has the effect on any particular vertical slice of water (at right angles to the direction of motion) that it gets a part of the slice near the fish into lateral motion, which by the time the fish's trailing edge reaches the slice is motion with velocity  $w$ . If the span of the trailing edge is substantial, this motion has a substantial lateral momentum, proportional to the velocity  $w$ . When this lateral momentum gets left behind (due to the fish's motion at speed  $U$ ) in the form of a vortex wake shed behind the trailing edge, the force of reaction on the fish is a rate of change of water momentum proportional to  $Uw$ .

Against this force, the trailing edge moving at velocity  $W$  does work at a rate proportional to  $UwW$ . However, part of this work is producing the 'wasted' energy of the vortex wake, which is the energy of motion at velocity

$w$  shed at velocity  $U$  and so is proportional<sup>†</sup> to  $U(\frac{1}{2}w^2)$ . The remainder of the fish's energy output (that is, the difference between these quantities) is available for propulsion, that is, for producing movement through the water at velocity  $U$  against the viscous resistance opposing that movement.

To produce thrust efficiently, then, the undular mode with  $U$  a substantial fraction of  $V$  is appropriate, because then Equation 2 makes  $w$  considerably smaller than  $W$  and therefore the energy wasted in the vortex wake, proportional to  $U(\frac{1}{2}w^2)$ , is small compared with the total energy expended, proportional to  $UwW$ . On the other hand,  $w$  must not be too small (that is  $U$  cannot be too close to  $V$ ) for given trailing-edge span, or the energy input will not be sufficient to overcome the viscous resistance. Obviously these are only crudely approximate arguments (depending on the elongated shape of the fish to justify considering the dynamics of different water slices independently of each other; for this reason they may be called elongated-body theory); but they do give some insight into the hydromechanics of anguilliform propulsion.

The eel-like fishes of the order Heteromi, together with the ribbon-fishes and unicorn-fishes of the order Allotriognathi, are similar to the Apodes except that the 'span' (total depth of body and fin) is not maintained constant quite as far as the posterior end but begins to taper sharply at a point a small fraction of the length away from that end. The loss of efficiency that results, because the tapering portion of the body contributes to resistance but effectively not to thrust, is fairly small because that portion is a small fraction of the whole. In some 'gulper-eels' of the order Lyomeri, and in two families of percomorph fishes (see Table III), it is a much greater fraction, and the presumed resulting loss of efficiency is accordingly much greater.

Actually, eels and similar fishes use strict anguilliform propulsion only in their higher ranges of speed; for relatively slow motion, undulation is often restricted mainly to the dorsal and anal fins themselves. Indeed, slow backward motion is also achieved, by putting those fin undulations into reverse. On the other hand, something close to anguilliform propulsion is achieved by fish without continuous dorsal and anal fins; in particular,

<sup>†</sup>The word 'proportional' has just been used four times, and the reader may be wondering what is the coefficient of proportionality. In each case it is, in hydro-mechanical jargon, the 'virtual mass per unit length,' which to a good approximation is the density of water times the area of a circle with the trailing-edge span as diameter [Lighthill (13)]. It may be crudely explained as follows: if any body in water suddenly starts moving, the water near the body is instantaneously set in motion (strictly, the influence spreads out from the body at the speed of sound, but this is so fast that the effect can be regarded as instantaneous), and so a momentum has to be given to the body equal to its velocity times the mass of the body augmented by an additional 'virtual mass' of water. It is worth noticing that the corresponding momentum changes in oscillatory motion differ from any effects of hydrodynamic resistance, and in particular occur at a phase earlier by 90°.

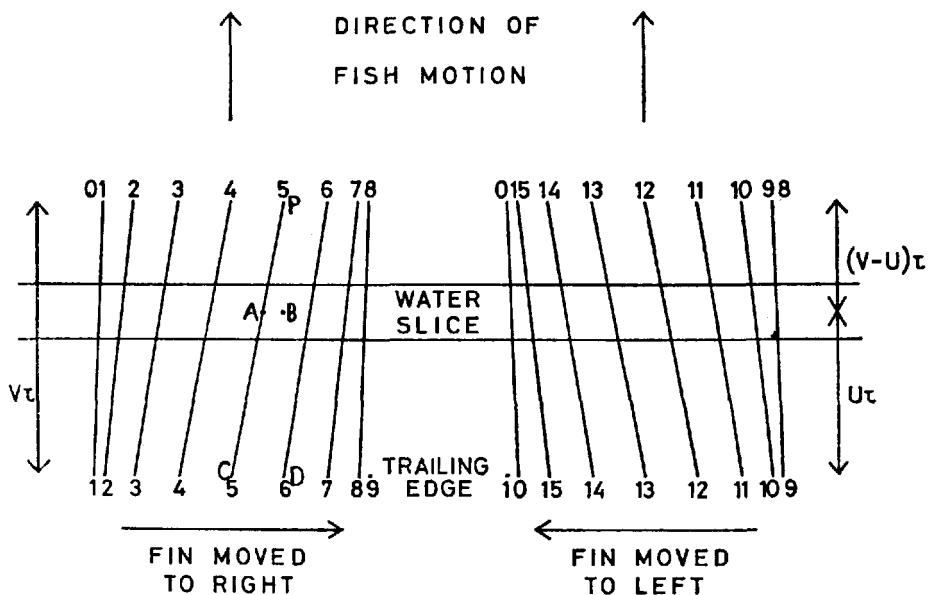


FIG. 5. Motion of caudal fin to the right (left-hand diagram) and then to the left (right-hand diagram) as a transverse wave reaches it at velocity  $V$  while fish moves forward at velocity  $U$ . The period of the undulation (time after which it repeats itself) is  $16\tau$ , and the numbers  $1, 2, 3, \dots$  on the diagram indicate positions of the trailing edge and of an anterior section of the fin at times  $\tau, 2\tau, 3\tau, \dots$  after a certain instant (when they are at position number 0). The wave takes a time  $\tau$  (that is, one-sixteenth of a period) to travel from the anterior section to the trailing edge, because that distance is taken as  $V\tau$ , and so the numbers on the trailing-edge positions are the same as those on the anterior-section positions but increased by 1 in each case. A water slice which at time  $5\tau$  (say) is anterior to the trailing edge by a distance  $U\tau$ , as shown, has been reached by the trailing edge (owing to the fish's forward motion with velocity  $U$ ) at time  $6\tau$ . Therefore the lateral distance moved by the fish's surface through the water slice during this time-interval  $\tau$  is  $AB$ . But similar triangles show us that  $AB/CD = PB/PD = (V - U)/V$ . Also,  $CD = W\tau$  when the lateral velocity of the trailing edge is  $W$ . Hence the velocity  $w$  with which the fish surface pushes the water slice is  $AB/\tau = (CD/\tau)(V - U)/V = W(V - U)/V$ .

members of the order Anacanthini like cod, whiting, haddock, and hake, with numerous separate dorsal fins and a substantial anal and caudal fin, use something close to it with much the same result in spite of the gaps between fins. Theory explains this by predicting that the 'vortex sheet,' filling in the area between fins, behaves sufficiently like a solid surface for the arguments set out above to be still operative.

### 5. INTRODUCTION TO THE CARANGIFORM MODE

We must now discuss the most important modification of the basic undulatory or anguilliform mode of propulsion; namely, the modification to what

shed into the wake will have the same values as before, and the inferred propulsive thrust and efficiency must be unchanged. Efficiency may even be improved, because lateral acceleration of water slices will proceed for a shorter time and is less likely to be accompanied by extra parasitic losses due to vortex formation at edges in the cross flow.

In fishes adopting the carangiform mode of propulsion, only a rather small posterior fraction of body length is capable of a high degree of flexure. The undulations are confined to that region and no complete wavelength is at any time apparent. It is most important, however, that within that region the propagated character of the undulation is retained. This means, as Figure 5 shows, that the caudal fin changes its angle of inclination as it moves from side to side in such a way that its movement always has a backward-facing component. (One may think of the resistance to this backward-facing component as providing the thrust, but this approach is almost impossible to make quantitative, or indicative of how much propulsive energy is wasted in making a vortex wake.)

One important hydromechanical difficulty prevents the carangiform mode of propulsion being successful unless certain conditions on the shape of the fish are satisfied. Because at no time is any one complete wavelength apparent, side forces produced by the lateral movements show no tendency to cancel out in their net effect, as they do for anguilliform propulsion. By contrast, those side forces can have a large resultant, fluctuating in direction from side to side as the fish's tail moves. These side forces, resulting from the fish's flexural movements, can force the fish to make additional body movements (namely, nonflexural movements of sideslip and yaw) that can be thought of as a sort of recoil process [Lighthill (13)].

Large recoil movements of this kind would be undesirable in many ways, and in particular would increase the proportion of energy wasted in making a vortex wake since they are in no way subject to the rules (Section 4) that make propagating undulatory movements avoid this. Most fish that have adopted the carangiform mode of propulsion have, however, succeeded in minimising recoil (for example, their heads do not make large periodic sideslipping or yawing movements). To achieve this, in the process of evolution, they adopted two main methods.

The more obvious is by minimising the side force itself. Wave amplitude in the carangiform mode has to increase from practically zero over the front portion of the fish to a rather large value over a short rear portion. The intermediate portion, where this increase occurs, is one that would produce large, sudden changes of momentum in slices of water passed through by it, and therefore also large local side forces, unless it happened to be a region of small local fish span, able to move only a small amount of water. This argues the value of a narrow necking (local reduction of depth of fish cross-section) immediately anterior to the caudal fin, with lateral oscillation growing mainly in that narrow region to a large, approximately uniform value over the fin itself.

There must still be a net fluctuating side force, of course, but under these circumstances it is kept down in magnitude to the rate of loss of momentum to the wake, noted above as proportional to  $Uw$ ; the recoil resulting from it makes changes in both  $W$  and  $w$ , but these are only by fractions of  $w$  and do not produce very large reductions in efficiency provided that the new ratio  $w/W$  remains relatively small. It can be helped to change little by a second feature of fish form; if the overall depth of body and fins for the fish as a whole is large, any sideslip or yaw involves the acceleration of considerable quantities of water (see previous footnote), and so a given side force produces only a relatively small recoil.

## 6. MEASUREMENTS OF THE CARANGIFORM MODE

When members of the order Isospondyli (not in many ways one of the most advanced of fish orders) attained their present form, very effective carangiform propulsion was already being achieved, and many of these fish are well known as strong, active swimmers. Necking of body shape anterior to the caudal fin, together with extension of depth in other regions through dorsal and anal fins (Figure 3), have helped to minimise recoil, while the body resistance that thrust must balance is minimised by their good streamlined shape.

Members of the large order Ostariophysi (that contains most of the successful freshwater fishes, and has the advantage, in general, over the Isospondyli of a specialised aid to acoustic sensitivity, as well as motile benefits from typically more flexible pectoral fins) have mainly adopted the carangiform mode of propulsion. The most notable exceptions are members of the suborder Siluroidea (catfishes) which use a practically anguilliform movement, and the highly specialised electric eel *Gymnotus*, which commonly uses anguilliform undulations confined to its long broad anal fin, and uncorrelated (in contrast to the Apodes) with any body movements.

The details of carangiform propulsion in several members of the Isospondyli and Ostariophysi have been investigated by Bainbridge (14–16) in his 'fish wheel.' More accurate measurements of speed are possible with this device than from observations under natural conditions (where, for example, velocities of water movement are often unknown), and accurate cinematographic measurements of flexural movements can be made. By observing species of different size and shape (the goldfish *Carassius auratus*, dace *Leuciscus leuciscus*, trout *Salmo trutta*, and far less slender bream *Abramis brama*) and individuals of very varying lengths within each species, he was able to show that fish length,  $l$ , is the principal parameter governing carangiform propulsion.

The maximum amplitude of caudal-fin movement (that is, lateral displacement from unstretched position) for these fish is found by Bainbridge to be scattered around  $0.2l$ , with no significant dependence on any other parameters. The hydromechanically important parameter  $\omega l/U$  (where the radian frequency  $\omega$  is  $2\pi$  times the frequency in Hz) takes values clustering

around ten. Scattered evidence from other sources is regarded by Bainbridge as suggesting that these values are typical of carangiform propulsion in general.

As an individual fish grows, the maximum frequency with which it can move its tail gradually decreases, so that its maximum cruising speed in lengths per second  $U/l$  (with typical values 3 to  $6\text{ s}^{-1}$ ) decreases, although  $U$  itself increases. Similarly, the smallest species have generally the highest frequency, and among the Ostariophysi it is particularly entertaining to observe the high-frequency carangiform propulsion used by small members of the family Characidae such as tetras. The propagated character of the posterior undulations, although not visually clear, can be established cinematographically; for example, Bainbridge found that  $U/V$  is about 0.7 for goldfish.

The caudal fin in many of the Ostariophysi is remarkably flexible; as one consequence of this, the side force resisting its lateral movement causes it to become somewhat concave towards its direction of motion. This passive mechanical effect of the carangiform mode of motion seems unlikely to have much back-reaction on the hydromechanics (for example, by altering the amount of momentum that the fin's motion can give to the water). On the other hand, Bainbridge (16) has shown that actively produced changes in fin area by stretching and unstretching occur, which must give the fish useful additional control over its propulsive mechanism; furthermore, vertically propagated undulations in the caudal fin are often used to generate slow vertical displacements.

The hydrodynamical resistance of a fish, which its carangiform tail movements must overcome, has been measured [Gray (11), p. 52] by several methods (rate of deceleration in glide, terminal velocity of a dead fish made rigid, etc.) and there is general agreement that resistance is close to that of any fairly smooth body of the same streamlined shape. Typical drag coefficients  $C_D$  based on maximum frontal area are about 0.25, showing some reduction with increasing Reynolds number. Gray pointed out that, if as a fish grows the power that its muscles can exert increases roughly in proportion to volume, and therefore to  $l^3$ , whereas the power needed to overcome drag is  $\frac{1}{2}\rho U^3 C_D S$  in terms of a representative area  $S$  proportional to  $l^2$ , then the speed achieved should vary as  $(l/C_D)^{1/3}$ ; this can explain various velocity dependences on  $l$  from a 0.33 power to a 0.60 power according to the dependence of  $C_D$  on  $R = Ul/\nu$ , and these are fairly close to the range of dependences observed.

## 7. THE CARANGIFORM MODE WITH LUNATE TAIL

The spiny-finned members of the very large order Percomorphi (Table III) have in general adopted a standard carangiform mode of propulsion, but there are three important areas of deviation that must be mentioned; the first, on which no further comment is necessary, is that a few families (see Table III) have retained anguilliform propulsion. By contrast, the

second area of deviation is of paramount interest because it has produced the highest velocities found in any kind of fish: speeds of 10 to as much as 20 m/s (for limited periods) were carefully measured by Fierstine & Walters (17). Seven important families are involved, including above all the Scombridae (named after the mackerel, *Scomber scombrus*, but in this context represented more impressively by the tunnyfishes, albacores, wahoo, skipjacks, and bonitos).

Actually, members of the family Carangidae itself already, to a somewhat greater extent than other families listed in the left-hand column of Table III, and like the herrings and other members of the suborder Clupeoidea among the Isospondyli, have caudal fins so scooped out internally as to make a V shape, the two arms of the V being 'sweptback' (in the language of aerodynamics) by an angle between 50° and 60°. This sweepback is, however, just large enough to make the estimation of propulsive efficiency by elongated-body theory (as above) a reasonable rough approximation, and that theory confirms that, when such a fin moves as a whole, the scooped-out area works (through the properties of the vortex sheet that fills it) just as effectively as the rest of the tail.

For further increases in thrust, that may permit still faster speeds, it is evident that caudal-fin span must be further increased, but that structural difficulties may prevent this being achieved by further extension of the arms of the V if the same degree of sweepback is retained. This is where the evolutionary advantages of speed, particularly in the highly competitive struggle for survival in the surface waters of the ocean, have brought a solution through the development of an important group of fish with caudal fins of high 'aspect-ratio' (another aerodynamicist's phrase, meaning span squared divided by planform area). A large change of shape is involved, typically to the crescent-moon fin shape (Figure 6) often described as a 'lunate tail' [Marshall (8), p. 10], although in its motion, photographed for example by Fierstine & Walters (17), essentially the same carangiform mode is retained.

Typical members of the seven families (headed by the Scombridae) noted in Table III have acquired both such a lunate tail and powerful musculature to move it. They have effectively separated the thrust-producing and drag-generating parts of their structures, especially because they retain in a notably intensified form the body necking anterior to the caudal fin, that is, a drastic reduction in span where the amplitude of lateral oscillation is rapidly increasing (Figure 6). In consequence, the thrust obtained is practically the same as if that 'high-aspect-ratio wing' that is the fin were performing its oscillations while moving through otherwise undisturbed water. The analogy to the method of thrust production in the commonest mode of bird flight, implied here by the use of the word 'wing,' can be seen to be rather complete if we turn Figure 5 on its side (Figure 7).

Admittedly, a bird needs high-aspect-ratio wings in any case, for the efficient weight support that only their high lift-drag ratio (and consequent low angle of glide) can give. However, to make the carangiform mode of

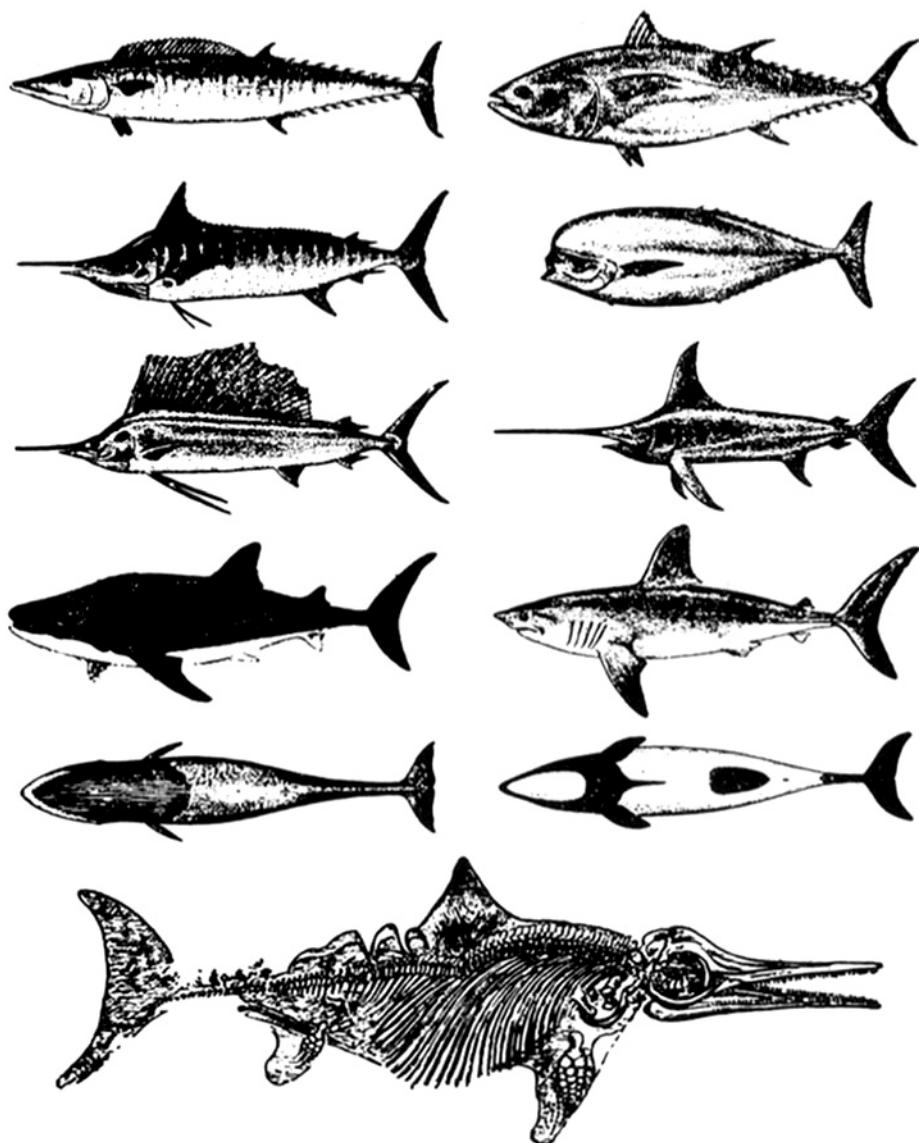


FIG. 6. Variants of the lunate tail in 6 percomorph fish, 2 sharks, 2 cetacean mammals and 1 reptile. Reading from left to right down the page, there are lateral views of specimens of wahoo, tunnyfish, striped marlin, louvar, sailfish, and swordfish (of genera *Acanthocybium*, *Thunnus*, *Tetrapturus*, *Luvarus*, *Istiophorus*, and *Xiphias*) from the Percomorpha, and of the whale shark and porbeagle (*Rhineodon* and *Lamna*) from the Selachii, and ventral views of the sei whale *Balaenoptera* and of a dolphin of genus *Cephalorhynchus* [all from Norman & Fraser (18)]. Lastly, there is an unusual impression of *Ichthyosaurus*, showing clearly the fin structure, found in the Lias of Württemberg [Woodward (19)].

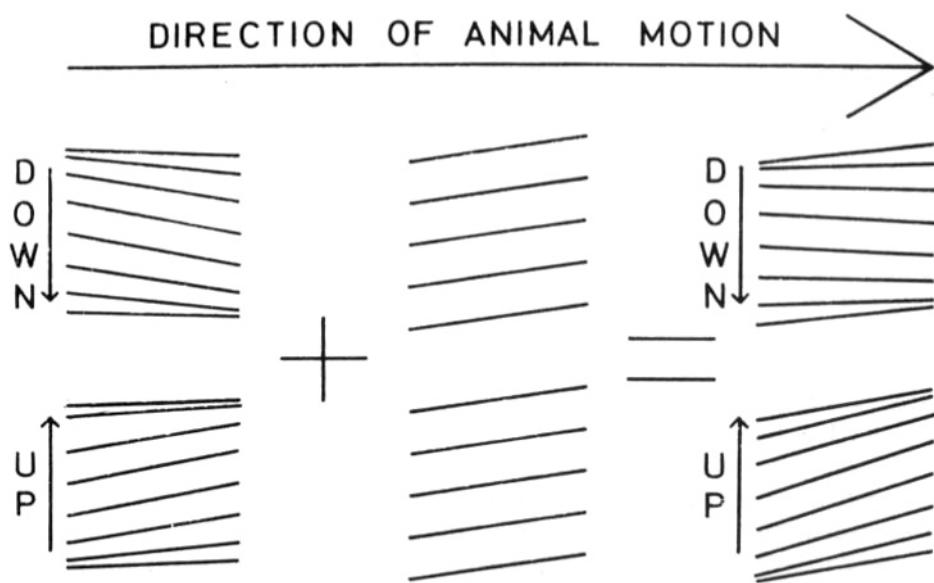


FIG. 7. The left-hand diagram shows vertical oscillations of carangiform type, obtained by rotating Figure 5 clockwise through  $90^\circ$ . The middle diagram shows a weight-supporting régime with constant angle of attack, and the addition of the two in the right-hand diagram portrays a typical bird-flight mode.

oscillation equivalent to the normal mode of bird propulsion, one only need add on to it (Figure 7) the constant angle of attack required for weight support. The result is seen to be the characteristic combination [Gray (11), p. 222] of a relatively straight downward stroke, where the wings are exerting a force considerably greater than the bird's weight, and the approximately 'feathered' upward stroke where little force is exerted; to make this possible, birds possess of course a far larger volume of muscle associated with downward than with upward wing movement.

#### 8. TWO-DIMENSIONAL ANALYSIS OF THE LUNATE TAIL

It is evident from this analogy that the action of propulsive caudal fins of high aspect ratio can be analysed by the same methods that have been used to analyse either the flight of birds and insects or the 'flapping' and 'torsional' modes in which aircraft wings may (undesirably) 'flutter.' Among such methods of analysis [Garrick (20)] the best developed is the 'two-dimensional' theory of oscillating aerofoils, of which an excellent review with emphasis on how to estimate the efficiency of thrust production was given by Wu (21). It is, however, important to recognise that the proper area of application of this two-dimensional theory of fish propulsion is to the study of hydrodynamic surfaces of high aspect-ratio (that is, principally the lunate tail, and in certain cases, as we shall see, the pectoral fins), and not to misapply it to the movements of whole fish, as some have done with the unfor-



FIG. 8. Depicts vortices cast off by the caudal-fin trailing edge as fish moves to the left, and also the jet-like streamline pattern induced by those vortices. This pattern was, actually, traced from streamlines experimentally obtained for such a configuration of vortices [Goldstein (23), plate 1].

tunate consequence that the aerofoil frequency parameter  $\omega c/U$  based on aerofoil 'chord' was taken of order 10 (like  $\omega l/U$ ; see above) rather than of order 1 or less.

In this two-dimensional approach (that is, one that considers horizontal, not vertical, slices of water as influenced independently by the caudal fin's lateral oscillations) careful study is once more given to the balance between (a) need to avoid too high a proportion of the fish's energy being wasted in making the vortex wake, and (b) need for some vortex wake to be present if any thrust is to be produced at all. Karman & Burgers (22) explained (b) clearly in a diagram (their figure 105) here reproduced in a modified form as Figure 8.

This shows a horizontal slice of water being traversed by the caudal fin as it propels the fish towards the left. Clockwise vorticity shed when the fin is at one of the lowest points in the diagram combines with anticlockwise vorticity shed when it is at one of the highest to generate in between something like a propulsive jet. On this simplified two-dimensional view, then, one can say that the caudal fin propels on the jet-reaction principle, but with good efficiency because the jet cross-section is large. (Note that the 'Karman vortex street' associated [Goldstein (23)] with the drag on a wire in a wind is here present in reverse, and so is associated with a thrust. On the other hand, the resistance of the fish's body is associated with a wake of more complex, essentially three-dimensional shape, with the water moving the other way; it and the wake of the fin do not, of course, cancel each other out, except in respect of the net horizontal force associated with them.)

Just as in the elongated-body theory, the proportional energy predicted as lost in the vortex wake can be minimised by partially feathering the fin movements in accordance with the principles of the carangiform mode (Figure 5), because the thrust increases as the first power of vortex strength (Figure 8) and energy loss as its square. Calculation of either has to take into account how water movements induced by wake vorticity modify the fin's effective angle of attack. No details can be given here, but it is worth noting that the mathematical device most effective for facilitating the calculation [Wu (21)] has a physical meaning that, again, reminds us of elongated-body theory; it calculates the pressure distribution from the fact that its

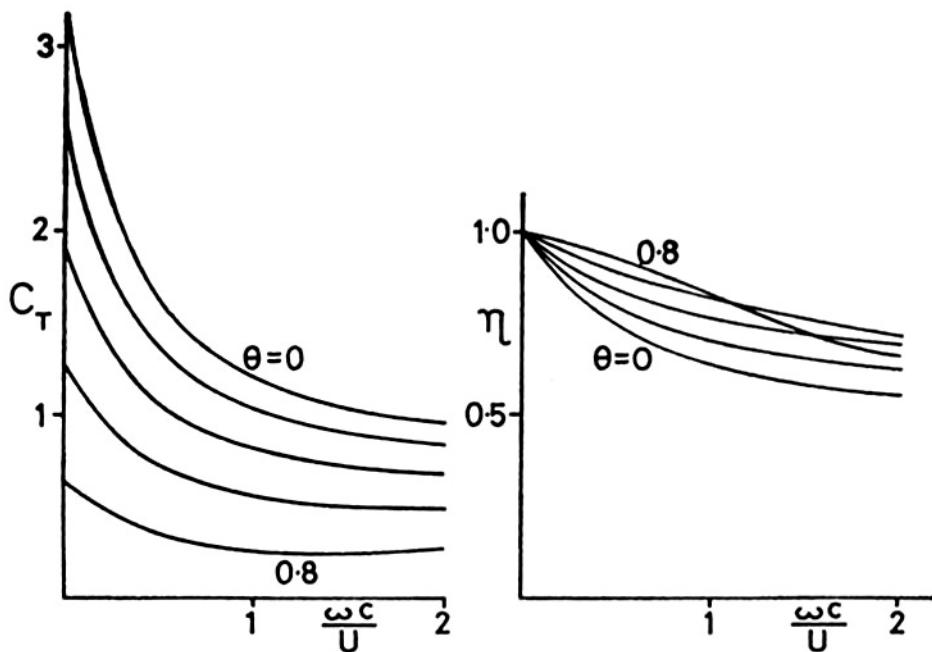


FIG. 9. Results of calculations, using two-dimensional theory, of thrust coefficient  $C_T$  and propulsive efficiency  $\eta$ , as functions of frequency parameter  $\omega c/U$  and of proportional feathering  $\theta$ , for lunate-tail motion in the carangiform mode.

gradient at right angles to the fin equals the rate at which the fin's movement is causing the local water momentum in that direction to change.

Figure 9 shows results of some calculations of this kind. The tail is supposed to oscillate laterally with maximum velocity  $W$  and maximum angle of attack (reached at the same phase)  $\theta W/U$ . Thus,  $\theta$  is a number usually less than 1 and can be described as the proportional feathering, since  $\theta=1$  would correspond to geometrically accurate feathering of the fin. The thrust coefficient  $C_T$  is thrust per unit fin area divided by  $\frac{1}{2}\rho W^2$ , and the efficiency  $\eta$  is the proportion of power exerted that goes into providing thrust. The figure shows clearly the competing tendencies of increasing efficiency but decreasing thrust coefficient as proportional feathering increases, between which the fish has somewhere to strike a balance.

To apply Figure 9 to estimate thrust and efficiency for a truly lunate fin, whose chord  $c$  (dimension in the direction of fish motion) tapers towards the tip, it would be necessary to use the variation in  $C_T$  and  $\eta$  with  $\omega c/U$  appropriate to the actual proportional feathering  $\theta$  to infer the appropriate averages (weighted with respect to fin area). Evidently, this is quite feasible. However, readers familiar with aerodynamic wing theory will recognize that such a purely two-dimensional method of calculation must overestimate effi-

ciency, because it takes into account only the energy of cross-stream wake vorticity (at right angles to the direction of motion) although in reality trailing vorticity (parallel to that direction) must also be present.

### 9. NEED FOR MORE DETAILED STUDY OF FAST PERCOMORPH FISHES

Thus, a more fully three-dimensional theory, using the ideas of 'lifting-line' theory, is needed, both to estimate efficiency more precisely, taking trailing vorticity into account, and to give more insight into the reasons for the evolution of the precise crescent-moon external fin shape involved in the lunate tail. We shall see that animals from completely different lines of evolution have converged (in the biological sense) on to the lunate tail as a means of fast propulsion through the water; not only the seven families of percomorph fish noted in Table III, but also a large group of sharks from a completely different sub-class of the class Pisces, and (from even another class) most of the cetacean mammals (see Figure 6). All these tails, furthermore, have acquired an aerofoil-shaped cross-section; its smoothly rounded leading edge is of particular importance, since both the two- and three-dimensional wing theories state that much of the thrust that they predict has to be derived from suction by water moving around such a leading edge. (Elongated-body theory, by contrast, indicates that a highly sweptback caudal fin would gain no comparable advantage from a blunt leading edge.)

It might, of course, be supposed that the advantages of the crescent shape depend on internal mechanics, but the varied structure of the tail (with a bony skeleton in the percomorphs, a cartilaginous one in the sharks, and no skeleton at all in the tail flukes of Cetacea) suggests that the advantages may well be hydromechanical. A possible reason for such hydromechanical advantage is indicated, bearing in mind the principle [Goldstein (23)] that vortex lines must be closed and cannot end in the fluid, when we consider the three-dimensional form of the vortex wake, a two-dimensional slice of which was depicted in Figure 8.

Because the cross-stream vorticity shed from the central part of the fin span is much greater than can be shed from the tapered tips, the vortex lines shed from the central part must bend downstream. Later, they must turn inwards, and finally close up when they reach the vorticity of opposite sign shed by the fin half a period earlier. This gives the idea that the wake consists of a series of vertical vortex rings travelling alternately to the right and to the left (of which Figure 8 represents a horizontal cross-section). They would be approximately circular vortex rings if the frequency parameter  $\omega l/U$  was about equal to  $\pi l/s$  where  $s$  is the span of the caudal fin, and this does typically take values of the order of 10 such as follow from the meagre measurements in the literature (see especially Fierstine & Walters (17), although a greater number of reliable frequency measurements are urgently needed for all the families of animals with lunate tails).

It is just possible, then, that the lunate tail is a favourable form because, at convenient frequencies of oscillation for fast movement, it can especially

readily shed vortex rings of approximately circular shape. These characteristically carry a large amount of momentum in relation to their energy, and so rate of shedding of wake energy might be minimised as a proportion of effective power exerted, which is related to  $W$  times the maximum rate of shedding of cross-stream momentum into the wake. The hydromechanical validity or otherwise of this conjecture is at present being probed quantitatively by the author, in a version of lifting-line theory specially devised for wing shapes with curved centre-lines.

The fast percomorph fishes in the Scombridae and associated families have, of course, achieved speed not only by improving thrust but also by reducing drag. First, the combined depth of dorsal fin, body, and ventral fin is in all cases large, so that lateral recoil of body in response to fluctuating side force on the caudal fin is minimised, because of the high 'virtual mass' of water that would also have to move (Section 4). This is especially notable in the sailfishes of the family Istiophoridae. Secondly, the body has in all cases a good streamlined shape, with a notably smooth surface, and in many species the pectoral and pelvic fins are folded into special slots during fast motion to make for reduced hydrodynamic resistance.

The question whether any further reductions in drag are achieved by fast fish through methods of boundary-layer control is still a difficult one to answer positively. However, it seems likely that the considerable quantities of water expelled through the gill slits may in many cases be important for keeping boundary-layer fluid charged with enough kinetic energy to avoid separation. An interesting case quoted by Breder (12) in his great study of fish locomotion (which he regarded as significantly influenced by gill efflux, though he does not frame his arguments in hydrodynamical terms) is that of a turning fish that closes the gill slits on the inside bend and concentrates all the efflux in those on the outside bend, which is just where the danger of boundary-layer separation would be greatest.

A traditional view of ichthyologists was that fish help to reduce their resistance by the lubricating effect of the mucus that they secrete, and indeed such mucus is secreted in particularly large quantities in the skin of many fast fish. Admittedly the mucus has other functions, including protection against bacterial attack, and its hydrodynamical relevance seems distinctly uncertain. However, the possibility of reduction in drag associated with boundary-layer turbulence, as a result of damping by very small concentrations of high-polymer substances in the water, or by viscous fluids that can move within cavities in an elastic boundary wall, cannot be ruled out as such effects have been observed under specially favourable hydrodynamics-laboratory conditions [Lumley (24), Kramer (25)].

We must not leave the order Percomorpha without briefly noting the third modification of the basic undulatory mode of propulsion that some of them exhibit (Table III). This rather radical modification is one where the undulation is entirely confined to the two pectoral fins, which make vertical undulations very similar to the horizontal undulations of the fish's tail in the caran-

giform mode (Figure 5). The rest of the body is relatively inflexible. Families in which propulsion is largely pectoral (Table III) have certain other features in common [Marshall (8), p. 210] that favour a somewhat cautious and tortuous type of motion; often they are frequenters of coral reefs; again, they are often 'cleaners' of other fishes. Pectoral propulsion, furthermore, is used in combination with carangiform propulsion in certain fishes, including members of the order Haplomi such as the pike.

#### 10. FISHES WITH WEIGHT UNBALANCED BY BUOYANCY

This survey has concentrated so far on the Teleostei (see legend to Table II), bony fishes that mostly have an effective hydrostatic organ (the swim-bladder) through which they can ensure that at all times their weight is balanced by their buoyancy. The organ is vestigial or absent only (a) in certain bottom-living teleosts, which gain evolutionary advantage from a tendency to sink back, when inactive, on to the bottom (where, often, protective colouring makes them relatively secure), and (b) in certain families of fast-moving fish such as the Scombridae, which might be at a disadvantage from restrictions on rapid movement between different levels imposed by a swimbladder, and in any case are easily able to maintain their vertical level through the propulsive force available to them and the fact that they are in constant motion (as, indeed, their particular respiratory mechanism requires).

Where fish do not live specially close to the bottom and are not propulsively active to a special degree, the absence of a swimbladder poses special problems. For example, members of the sub-class Chondrichthyes (see Table II), with cartilaginous skeletons and no swimbladders, have to support an excess of weight over buoyancy of as much as 5 per cent [Marshall (8), p. 69], except where this is reduced owing to their liver oil and muscle fat content. This need is reflected in their form. A similar excess of weight over buoyancy is found in a prominent family Acipenseridae (the sturgeons, whose swim-bladders are not fully effective hydrostatically) within that relatively archaic order of bony fishes, the Palaeoniscoidea.

Both the Acipenseridae and most members of the order Selachii (sharks and dogfishes) possess caudal fins whose external shapes are markedly heterocercal; such a fin has no symmetry about a horizontal plane (Figure 10). Its upper half is like a long, highly sweptback wing, pointing diagonally upwards and backwards. Its lower half is a far more modest wing-type surface whose span is typically between 0.1 and 0.5 times as great. These fishes with markedly heterocercal fins use a mode of propulsion that is essentially anguilliform, though there is substantial posterior increase of amplitude. The upper half of the caudal fin, however, produces an additional force that may appropriately be analysed by the two-dimensional theory described in Section 8.

Evidently, such analysis must consider the motion of the sweptback half-fin through the water as resolved into two parts, a resultant perpendicular

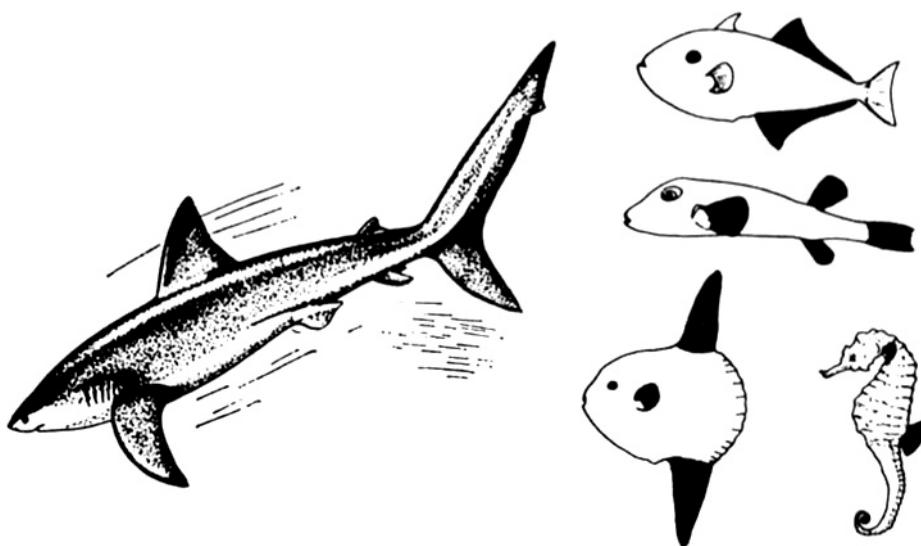


FIG. 10. Left: the blue shark *Prionace glauca*, of the family Carcharinidae, showing the markedly heterocercal tail. Right: the trigger-fish *Xanthichthys*, puffer-fish *Sphaeroides*, ocean sunfish *Mola*, and sea-horse *Hippocampus*, with fins active in propulsion indicated in black [Marshall (8)].

to its span with which is associated a thrust in that direction, and a resultant parallel to its span with which negligible thrust is associated. The vertical component of the thrust perpendicular to the half-fin's span is a 'lift.' This acts in combination with the hydromechanical lift on the pectoral fins, which in these animals can readily be held at a positive angle of attack to the axis of the body, to support the excess of weight over buoyancy. Admittedly the pectoral fins, nearer (and anterior) to the mass centre, support most of the excess, but this support could not be provided stably if it were not supplemented by the posterior caudal-fin support. The two sources of lift give the animal dynamical properties like an aeroplane, and it can even 'loop the loop' (as frequently observed in the genus *Acipenser* in particular).

The caudal fin is markedly heterocercal throughout many families in the order Selachii; for example, the Odontaspidae, Carcharinidae, Squalidae, and Heterodontidae. In others, there are exceptions where the caudal fin is not prominent at all, like bottom-living dogfishes of the family Scyliorhinidae. The most interesting exceptions, however, are the very fast moving sharks, including such members of the family Lamnidae as the porbeagles, mackerel- and mako-sharks of genera *Lamna* and *Isurus*, and the great white shark *Carcharodon* (but not the fox shark *Alopias vulpes*), and such members of the family Orectolobidae as the whale shark *Rhineodon* (but not the long-tailed zebra shark *Stegostoma*).

It is these fast sharks that are noted above (see Figure 6) as having developed lunate tails of practically the same external form as in the fastest

teleost fishes. (Essentially, they have achieved this by acquiring lower, skeletonless halves to their caudal fins, similar to the upper halves and with span at least 0.75 times as great.) They move these tails, furthermore, in practically the same carangiform mode. They need no special morphological feature for balancing their excess of weight over buoyancy for the same reason as in the Scombridae, that is, because they are such fast, active swimmers.

Within the other great order, Batoidei, of cartilaginous fishes, members of one family, the Pristidae (sawfishes), are rather similar to the faster sharks in aspects of their external form relevant to propulsion and weight support. Other families show clearer evidence of a very long process of adaptation of form to bottom-living conditions. The development of much larger pectoral fins, able to balance most of the excess of weight over buoyancy, is pronounced already in the Rhinobatidae (guitar-fishes) and Torpedinidae (electric-rays). However, effective caudal fins are still present, combining with pectoral undulation to provide the propulsive effort.

In the other four families, the caudal fin is negligibly small or absent. The Rajidae (skates) have adopted a particularly graceful form of propulsion. Their enormous pectoral fins together make approximately a square, with the body stretched along its diagonal and only a very thin tail spilling over beyond the diagonal. Passing backwards over the pectoral fin is an undulation very close in character to the basic undulatory mode from which our survey started (but vertical, rather than horizontal as in anguilliform propulsion), and comprising considerably more than one whole wavelength. The Dasyatidae (sting-rays) use a similar propulsive mode.

Hydromechanically, this would be hard to analyse, because for such a shape of aspect-ratio 2 neither elongated-body theory, nor a wing theory based on two-dimensional studies, can be used. However, by a process of interpolation between the results of both theories we may guess that reasonable propulsive efficiency is achieved when the wave speed only slightly exceeds the forward speed. The wing-like shape of these animals, which are heavier than water (except for certain deep-ocean forms), is additionally valuable in giving them a low angle of glide when not actively swimming.

Two families, the Myliobatidae (eagle-rays) and Mobulidae (devil-rays), by a later adaptation to a more free-swimming existence, have become still closer in their swimming movements to the flight of birds, as the name 'eagle-rays' implies. In fact, their pectoral fins have a 'diamond' planform of aspect-ratio about 4 (with the body along the shorter diagonal), and less than a wavelength of undulation is present at any one time, so that a motion not unlike that of Figure 7 is achieved (although with the added constant angle of attack less because lift needs to balance only the excess of weight over buoyancy, and with rather more curvature retained in the undulatory mode).

A remarkable case of convergent evolution has brought one order of teleost fishes, the Heterosomata (see Table II; familiar examples are plaice,

sole, halibut, turbot) into some crude similarity of external form and motion with their fellow bottom-livers the *Raiidae* (skates). A flatfish has, in fact, no swimbladder and at a certain stage in its life turns over on to one side while the eye on that side moves round to the other. It retains the anguilliform mode of motion (which as in *Apodes* uses the whole body for faster motion and only the dorso-ventral fins for slower motion) but the fish's oscillatory movements, after it has turned over, are of course vertical. The whole fish, now, is like a wing whose 'diamond' planform has aspect-ratio about 1 (with the head on the longer diagonal).

### 11. PROPULSION IN 'ARMOURED' FISHES

The remaining three out of the fish orders to be discussed (see Table II) have developed defensive armour at the expense of speed. Not only the body in general but in most cases even the tail has lost flexibility, to such an extent that the main means of propulsion is by fin undulations. The deep-bodied *Zeomorphi* use undulations in highly flexible posterior dorsal and anal fins, and basically the same mode appears in the families *Monacanthidae* (file fishes) and *Balistidae* (trigger-fishes; see Figure 10) among the *Plectognathi*, and in the slender-bodied *Centriscidae* (shrimp-fishes) among the *Solenichthyes*. In this latter order, the also slender-bodied *Fistularidae* (cornet-fishes) and *Aulostomatidae* (trumpet-fishes) can supplement this mode with caudal-fin movement when hard pressed.

External armour is even more rigid in the family *Syngnathidae* (sea-horses and pipe-fishes) among the *Solenichthyes*, and these dwellers in 'jungles' of marine plant life are adapted for precision manoeuvrability rather than speed. They use dorsal and pectoral fins of very small total area for this purpose (Figure 10), but to produce various speeds can undulate them at a great variety of frequencies (70 Hz has been recorded). The downward-pointing flexible tails of sea-horses have acquired mainly prehensile function, but are sometimes undulated to counteract the fish's excess of weight over buoyancy or provide upward movement; horizontal displacements, on the other hand, are made through fin vibration.

Among the *Plectognathi*, there are some dwellers in coral reefs bearing exceptionally heavy armour, such as the *Diodontidae* (porcupine-fishes), *Tetraodontidae* (puffer-fishes; see Figure 10) and *Ostraciontidae* (trunk-fishes), which combine the methods characteristic of the *Zeomorphi* and *Syngnathidae*, using pectoral dorsal and anal undulations. On the other hand, the *Molidae* (ocean sunfishes) show fascinating traces of readaptation to pelagic life. With almost complete absence of effective caudal fin, they have had to reacquire propulsive efficiency by a huge increase in the aspect-ratio of their dorso-ventral fins (Figure 10). These operate in the bird-flight mode of the *Myliobatidae* (eagle-rays) but turned through 90°; apparently a unique case within the animal kingdom of a disjoint pair of high-aspect-ratio wings deployed in a vertical plane.

## 12. LIMBS IN SWIMMING: AMPHIBIANS AND ARTHROPODS

This relatively lengthy survey of the class Pisces leaves space for only a brief treatment of aquatic propulsion in those classes of vertebrates whose tetrapod structure is an adaptation to operation on land. In most of the Amphibia, of course, there is a larval state, which has inherited anguilliform motion, and a laterally compressed tail to go with it, from fish ancestors. The corresponding tetrapod adult forms, even in the tail-retaining sub-class Urodeles, are sometimes aquatic to only a negligible extent (as in the genus *Salamandra*, whose tail has lost its suitability for aquatic propulsion), but the majority of Urodeles, from the newts to the giant salamander *Megabatrachus*, retain a laterally compressed tail for a life that is aquatic to a considerable degree.

Of the tailless amphibians (the frogs and toads of the sub-class Anura) the majority are aquatic to an insignificant extent in adult life, but in that considerable minority that are mainly aquatic we encounter for the first time the problem of adapting the tailless-tetrapod structure for aquatic propulsion. The mode that they use is unrelated to any used by fish (and its symmetry about the animal's plane of symmetry occurs among the fish only in those that rely on pectoral-fin propulsion). With their flexible back legs and webbed feet they execute something close to the leg movements of a human 'breast stroke,' rather more efficiently because the legs are better streamlined.

We may now look back momentarily to the invertebrates, among which we earlier failed to mention modes of aquatic propulsion in the great phylum Arthropoda (see Table I). Most arthropods, to be sure, are either terrestrial or bottom-living, but some are considerably adapted for swimming. In the class Insecta, the problem of so adapting the basic hexapod form is similar to that just discussed, and it has been solved by many families in the order Coleoptera (beetles) through remoulding of the cross-sections of the hind and intermediate legs as in *Hydrophilus*, or addition of fringes to them as in *Dytiscus*, in each case converting them to a good aerofoil shape, able to exert force with a minimum of wasted energy. The mode of motion is, however, antisymmetrical, with legs on each side sculling alternately.

In the more characteristically aquatic class Crustacea, we consider only three orders: Stomatopoda, Euphausiacea, and Decapoda, containing all the larger animals [for an excellent survey of the class as a whole, see Lochhead (26)]; in these, the body is composed of 19 somites (13 in head and thorax and 6 in the abdomen), each bearing two limbs which may however be greatly modified or suppressed. A rather more fish-like adaptation of legs is therefore possible. In the first place, the telson, or extensible tail-fan formed from the appendages of the last abdominal somite, allows propulsion by abdominal motions in a vertical plane, which are presumably effective owing mainly to its substantial horizontal span.

The other ten abdominal limbs (or pleopods), when significantly retained

in members of these orders, allow them to swim, additionally, by passing a wave of propulsive movement along them. This wave in shrimps and similar animals may be thought reminiscent of pectoral propulsion in fishes. It is used alone for slow movement and in combination with the telson's motion in a vertical plane for faster movement.

By contrast, the abdominal limbs are suppressed in the crabs (suborders Anomura and Brachyura). These are mainly crawlers, but a crab family that excels at swimming is the Portunidae, with their thoracic limbs (especially the last pair) modified to an aerofoil cross-section to give by a wing-like beating thrust enough for speeds as high as 1 m/s. However, after this brief reference to just a few Arthropoda we must return to the vertebrates.

### 13. AQUATIC PROPULSION IN REPTILES

From the point of view of aquatic propulsion, the descent of modern members of the class Reptilia from archaic tailed Amphibia is most obvious in the case of the order Crocodilia (crocodiles and alligators), which use their deep, laterally compressed tails in an essentially carangiform mode of propulsion. The other reptile orders significantly represented to-day are the Squamata (snakes, lizards, and chameleons) and the Chelonia (tortoises and turtles), but before discussing these we may, perhaps, note the striking fact that the extinct marine reptile *Ichthyosaurus*, though descended from terrestrial ancestors, possessed a lunate tail, arranged much as in the Scombridae (Figure 6).

The 'serpentine' mode [Gray (27)] of terrestrial propulsion in the majority of snakes (sub-order Ophidia) is, to be sure, reminiscent of anguilliform propulsion, but is a very atypical special case of it with  $U = V$ ; that is, the snake glides along a fixed path as undulations pass backward down its body at the same speed as the whole animal moves forward. This motion is possible only on a surface with many protuberances, and depends on the fact that those protuberances give a 'normal' force (i.e., force at right angles to the body axis) already when the normal component of velocity is zero, whereas tangential force is not strongly dependent on the tangential component of velocity. In other words, there is an incommensurability between normal and tangential resistance which is absent for motion in a fluid.

Terrestrial snakes, in spite of their 'bad' hydromechanical shape with roughly circular cross-section and gradual posterior taper, swim fairly readily when required, by a similar undulatory movement, which however produces of course a forward velocity  $U$  less than the propagation velocity  $V$  [Gray (5) observed that  $U = 0.7V$  in the grass-snake *Tropidonotus natrix*]. The following very crude theory of undulatory propulsion in 'bad' hydromechanical shapes, simplified from Taylor (7) and Gray & Hancock (3), is constructed by supposing that Figure 5 now describes the motion of any short section of the animal, and by adopting a crude linearised law for resistance to both its normal and tangential velocities of motion.

Figure 5 shows that the slope of such a short section, that is, the tangent of the angle it makes with the direction of motion, is  $CD/PD = W/V$  at any one time. We make the common approximation of supposing this angle small enough for its cosine to be replaced by 1, so that the section's normal velocity relative to the fluid is  $w$ . Then the normal resistance to the section's motion is taken as  $K_N$  times this. The tangential velocity is  $W$  times the slope, giving  $W^2/V$ , and the tangential resistance to the section's motion is similarly taken as  $K_T$  times this tangential velocity. Here  $K_N$  and  $K_T$  are resistance coefficients for normal and tangential motion respectively.

We can now approximate the propulsive thrust  $P$  on the section as normal resistance times slope, minus tangential resistance, that is, as

$$P = (K_N w)(W/V) - K_T(W^2/V) \quad 3.$$

At the same time, the animal expends energy at a rate  $E = (K_N w)W$  by its movements at velocity  $W$  against the resistance  $K_N w$ . The efficiency  $\eta$ , or useful work done,  $UP$ , divided by  $E$ , is

$$\eta = UP/E = (U/V)[1 - (K_T/K_N)(W/w)] \quad 4.$$

It is easy to see from Equation 3 that positive thrust is obtained only if  $w/W$  lies between 1 and the resistance-coefficient ratio  $K_T/K_N = r$ , say. Furthermore, calculations using Equations 2 and 4 show that the maximum efficiency is achieved under the conditions

$$w/W = r^{1/2}, \quad \text{giving} \quad U/V = 1 - r^{1/2} \quad \text{and} \quad \eta = (1 - r^{1/2})^2 \quad 5.$$

For a snake on the ground, bracing itself against protuberances, the ratio  $r$  between the coefficients of tangential and normal resistance is effectively zero, since the coefficient is infinite for normal resistance, which is obtainable with zero normal velocity; Equation 5 then implies that  $U = V$  and that the purely mechanical efficiency of thrust production is 1. On the other hand, for aquatic motion of *Tropidonotus natrix* at a Reynolds number of order  $10^6$  based on length, a value of  $r$  around 0.1 is not unreasonable, giving  $U/V = 0.7$  to one significant figure, as observed, and  $\eta$  a little less than 0.5. Furthermore, as Reynolds number  $R$  becomes less,  $r$  should slowly increase, giving a slow decrease in  $U/V$ , until for very small  $R$  we have, to a useful rough approximation discovered by Hancock (28),  $r \doteq 0.5$ , giving  $U/V \doteq 0.3$  and  $\eta \doteq 0.1$ .

This trend is fairly well in agreement with a number of observations quoted earlier (Section 2) on invertebrates, and it also explains the anomalous behaviour of the polychaete annelids, whose lengthy parapodia, impeding tangential but not normal motion, could well cause  $r$  to exceed 1, so that  $U/V$  would have the negative value observed. The crude approximation is improved on and made to apply to a wider range of types of undulation at low  $R$  by Gray & Hancock (3) and at high  $R$  by Taylor (7). For hydrome-

chanically 'good' shapes, however, like those of fish, it would be quite unsatisfactory, since it ignores the large 'virtual mass' effect (Section 4), and the resulting large additional thrust appearing whenever there is a deep, laterally compressed tail. For shapes of intermediate character, it would be desirable if possible to combine the two approaches.

The life of one sub-family of Ophidia, namely the Hydrophinae, is purely marine, and in spite of terrestrial ancestry they have acquired for propulsive purposes just such a deep, laterally compressed tail ending in a vertical trailing edge, with presumably considerable advantage in respect of speed. To conclude discussion of the Reptilia, we may note that, in the order Chelonia, a hydromechanically even more sophisticated propulsive mode has been adopted by the Chelonidae (true turtles). They have found a beautiful solution to the problem of tetrapod swimming through the development of their forelimbs into flippers of high-aspect-ratio wing shape, with which they imitate a bird's propulsive mode (Figure 7) more accurately than do the eagle-rays themselves.

#### 14. AQUATIC PROPULSION IN MAMMALS

Among the Mammalia (the last class in Table I), a mainly aquatic mode of life is found in only three orders: Carnivora, Cetacea, and Sirenia. Within the first of these, furthermore, it is only the seals (families Phocidae and Otariidae) and walruses (Odobenidae) to which the comment applies. These are true quadrupeds, and have evolved interesting compromises, in their use of limbs for swimming, with the demands of the not negligible terrestrial portion of their life; for example, the forelimbs of Phocidae (the earless seals, including all the European species) are specialised for locomotion on land, and in swimming are retracted into the axillary hollows (except for purposes of changing direction), while alternately each hind flipper, using the capacity of its pentadactyl skeleton for stretching, sweeps backwards and inwards with toes extended before returning forwards with knee flexed and toes closed [Backhouse (29)]. The result is a rather crude approximation to carangiform propulsion.

In the Cetacea (whales and dolphins), by contrast, there is uncompromising adaptation to a totally marine life, and the hind limbs have completely disappeared. At the tapered posterior end of the vertebral column, a new organ in the form of a horizontal lunate tail, consisting of two 'flukes' (lateral expansions of skin supported by fibrous tissue), provides through its vertical movements all necessary propulsive force, while the flippers derived from the fore-limbs are responsible only for stability and control. The effectiveness of this solution is indicated by the fact that the quite different evolutionary line of development represented by the order Sirenia, with only two existing genera *Halicore* (dugongs) and *Trichechus* (manatees), has led to almost an identical method of propulsion, involving loss of hind limbs and laterally expanded tail, although only in the former genus has the expansion attained the truly lunate form.

Some excellent cinematographic records by Parry (30) and Lang (31) prove that the vertical movements of cetacean lunate tails are exactly of the same type (depicted in Figure 5) as the horizontal movements in the carangiform mode of fish propulsion. Recoil (which here would be vertical, resulting from fluctuating lift forces on the tail) is again minimised, especially because the body has a very small horizontal span in the anterior tail region where amplitude is rapidly increasing.

If a cetacean has often to swim in conditions when its buoyancy exceeds its weight, its flukes and flippers must provide negative lift. The flukes, being farther from the mass centre, need provide only a small fraction of the excess, which however for a large cetacean can be a quite substantial force in relation to tail area. The carangiform mode must then be altered by addition of a constant negative angle of attack, the opposite of the alteration portrayed in Figure 7 in connection with bird flight. It follows that the greatest effort would be made on the upstroke, and it is interesting that in both *Phocaena* [Gray (11), p. 289] and *Balaenoptera* [Kermack (32)] the volume of musculature that is used in the upstroke is substantially greater than that involved in the downstroke, again in opposition to the arrangement found in birds (Section 7).

Most cetaceans have very smooth, streamlined bodies whose relatively low drag is the more easily maintained because the propulsive mechanism is confined to the tail, and then the combination of powerful musculature, efficient propelling mechanism, and low drag produces a high speed (for example,  $U=8$  m/s with  $l=2$ m) when required. As with the fast fish, however, there is no clear evidence [see Bainbridge (33), Slijper (34) and Lang (31)] that skin friction is reduced below that appropriate to the actual surface smoothness and Reynolds number (of order  $10^7$  or more). It should be noted that many reports of extraordinary speeds in dolphins have to be discounted because of the likelihood that the animals were riding currents, wakes, or waves.

### 15. CONCLUSION

All the fastest marine animals, whether among the mammals, the cartilaginous fishes, or the bony fishes, have adopted the carangiform mode of propulsion with a large lunate tail. Further hydromechanical study of this propulsive mechanism, by theory and experiment, coupled with considerable increase in observational knowledge on details of its operation under natural conditions, is extremely desirable.

Next, a more complete elongated-body theory, taking into account simultaneously the virtual-mass effect and viscous resistance to the cross-flow, would be particularly useful. To supplement these two main lines of hydrodynamic research that appear necessary, special studies would be appropriate on (a) propulsion by undulations in a fin alone, (b) propulsion in skates, other rays, and flatfishes, (c) the starting and turning mechanisms in fishes,

(d) the manner in which shrimp-like crustacea use their ten abdominal limbs for propulsion, and (e) application of the methods of Gray & Hancock (3) to ciliary propulsion. No doubt many other suitable lines of research will occur to readers of this survey.

#### ACKNOWLEDGEMENTS

I have pleasure in expressing my gratitude to Dr. Quentin Bone, Dr. Victor Howarth, and Dr. Richard Bainbridge for extremely valuable discussions, as well as to the Zoological Society of London and the British Museum (Natural History) for benefits received from intensive study of their outstanding collections. My debt to Dr. N. B. Marshall of the Museum will be particularly evident from the text. Finally, warmest thanks are due from me, as from all other workers in the field of animal locomotion, to Sir James Gray for the wonderful stimulus that he has been giving to the subject throughout the past fifty years.

## LITERATURE CITED

1. Gray, J., *Ciliary Movement* (Cambridge Univ. Press, 162 pp., 1928)
2. Siekmann, J., *J. Fluid Mech.*, **15**, 399–418 (1963)
3. Gray, J., Hancock, G. J., *J. Exptl. Biol.*, **32**, 802–14 (1955)
4. Gray, J., Lissmann, H. W., *J. Exptl. Biol.*, **41**, 135–54 (1964)
5. Gray, J., *Proc. Roy. Soc. (London) Ser. B*, **128**, 28–62 (1939)
6. Gray, J., *J. Exptl. Biol.*, **16**, 9–17 (1939)
7. Taylor, G. I., *Proc. Roy. Soc. (London) Ser. A*, **214**, 158–83 (1952)
8. Marshall, N. B., *The Life of Fishes* (Weidenfeld & Nicolson, London, 402 pp., 1965)
9. Harris, J. E., *J. Exptl. Biol.*, **15**, 32–47 (1938)
10. Hertel, H., *Structure, Form and Movement* (Reinhold, New York, 251 pp., 1966)
11. Gray, J., *Animal Locomotion* (Weidenfeld & Nicolson, London, 479 pp., 1968)
12. Breder, C. M., *Zoologica*, **4**, 159–297 (1926)
13. Lighthill, M. J., *J. Fluid Mech.*, **9**, 305–17 (1960)
14. Bainbridge, R., *J. Exptl. Biol.*, **35**, 109–33 (1958)
15. Bainbridge, R., *J. Exptl. Biol.*, **37**, 129–53 (1960)
16. Bainbridge, R., *J. Exptl. Biol.*, **40**, 23–56 (1963)
17. Fierstine, H. L., Walters, V., *Mem. S. Calif. Acad. Sci.*, **6**, 1–31 (1968)
18. Norman, J. R., Fraser, F. C., *Giant Fishes, Whales and Dolphins* (Putnam, London, 375 pp., 1937)
19. Woodward, A. S., *Encycl. Britann.* (11th Ed.), **23**, 141–47 (1910)
20. Garrick, I. E., Nonsteady Wing Characteristics. In *Aerodynamic Components of Aircraft at High Speeds*, 658–793 (Donovan, A. F., Lawrence, H. R., Eds., Princeton Univ. Press, 845 pp., 1957)
21. Wu, T. Y., *J. Fluid Mech.*, **10**, 321–44 (1961)
22. von Karman, T., Burgers, J. M., *General Aerodynamic Theory: Perfect Fluids. Vol. II* (367 pp.) of *Aerodynamic Theory* (Durand, W. F., Ed., Springer Verlag, Leipzig, 6 vols., 1934)
23. Goldstein, S., Ed., *Modern Developments in Fluid Dynamics* (Oxford Univ. Press, 702 pp., 1938)
24. Lumley, J. L., *Appl. Mech. Rev.*, **20**, 1139–49 (1967)
25. Kramer, M. O., *J. Am. Soc. Naval Engr.*, **72**, 25–34 (1960)
26. Lochhead, J. H., Locomotion. In *Physiology of the Crustacea*, vol. 2, 313–64 (Waterman, T. H., Ed., Academic Press, New York, 1961)
27. Gray, J., *J. Exptl. Biol.*, **23**, 101–20 (1946)
28. Hancock, G. J., *Proc. Roy. Soc. (London) Ser. A*, **217**, 96–121 (1953)
29. Backhouse, K. M., *Symp. Zool. Soc. London*, **5**, 59–75 (1961)
30. Parry, D. A., *J. Exptl. Biol.*, **26**, 24–34 (1949)
31. Lang, T. G., Hydrodynamic analysis of cetacean performance. In *Whales, Dolphins and Porpoises: Proceedings of the First International Symposium on Cetacean Research*, 410–32 (Norris, K. S., Ed., Univ. of California Press, 789 pp., 1966)
32. Kermack, K. A., *J. Exptl. Biol.*, **25**, 237–40 (1948)
33. Bainbridge, R., *Symp. Zool. Soc. London*, **5**, 13–32 (1961)
34. Slijper, E. J., *Symp. Zool. Soc. London*, **5**, 77–94 (1961)

## CHAPTER 3

# Mathematics of Aquatic Animal Locomotion at Low Reynolds Number

### 1. Introduction

We saw in Chapter 2 that it is the animals around 1 mm in size or less whose motions, with Reynolds numbers around 1 or less, are essentially uninfluenced by inertia. Some mathematical aspects of their locomotion are investigated in Chapter 3.

In such motions the forces acting on each element of fluid must be in balance, just as in statics, since those inertial terms (mass times acceleration) which might account for any imbalance are negligible. Similarly the fluid acts on the swimming animal with a system of forces that must effectively be in static equilibrium at all times, since the animal's rate of change of either momentum or angular momentum is negligible compared with the magnitude of individual viscous forces.

These statical principles are supplemented by thermodynamic and kinematic constraints: the fluid velocity field must satisfy an equation of continuity allowing the density to remain constant, and the viscous forces must take the form that follows from Newton's viscosity law for the fluid medium. Similarly kinematic theory tells us that at any instant the total movement of the animal's external surface must be a sum of

- (i) motions *relative to* the mass centre associated with the distribution of body curvature and other local body strainings (including extensions or relaxations) generated by the animal's motile activity; and
- (ii) rigid-body motions, in general composed of a uniform translation at the velocity  $\mathbf{U}$  of the animal's mass centre and a rotation at angular velocity  $\boldsymbol{\Omega}$  about that mass centre.

Given the part (i) generated by the animal's motile activity, this total instantaneous motion is specified by two unknown vector quantities  $\mathbf{U}$  and  $\boldsymbol{\Omega}$ . Evidently the condition that the forces acting between the fluid and the animal are a system with zero resultant and zero moment provides the right number of equations (two vector equations) to determine  $\mathbf{U}$  and  $\boldsymbol{\Omega}$ . Note that their instantaneous values are determined solely by the animal's instantaneous motile activity, since lag-generating inertial effects are negligible. Dynamics degenerates here into a blend of kinematics and statics: an interesting and, really, a simplifying feature although it takes some getting used to!

### 2. The fluid equations and their fundamental singular solutions

The forces acting on any element of a Newtonian fluid, per unit volume, are a pressure gradient force  $-\nabla p$  and a viscous force  $\mu \nabla^2 \mathbf{u}$ , where  $p$  is pressure,  $\mu$  is

viscosity and  $\mathbf{u}$  the fluid velocity vector. The condition

$$-\nabla p + \mu\nabla^2\mathbf{u} = 0 \quad (1)$$

that they are in equilibrium, taken with the equation of continuity,

$$\nabla \cdot \mathbf{u} = 0, \quad (2)$$

leads to the important conclusion that the pressure  $p$  satisfies Laplace's equation

$$\nabla^2 p = 0. \quad (3)$$

The pressure, then, is a *harmonic function* in any inertialess flow, even though all we can say about the velocity vector  $\mathbf{u}$  from (1) and (3) is that it is a *biharmonic function* (solution of  $\nabla^4\mathbf{u} = 0$ ; actually, a *vector* solution regarding which (2) gives some further information).

To represent the fluid motions induced by a flagellum, different parts of which exercise different forces on the fluid, we may wish to use the equations of an inertialess fluid under some general distribution of forces. This may be built up from point forces, each corresponding to a distribution

$$\mathbf{F}\delta(\mathbf{r}) \quad (4)$$

of force per unit volume, where  $\delta$  is the delta function and  $\mathbf{r}$  is the vector displacement from the point of application of the force  $\mathbf{F}$ . Equilibrium of a fluid element under the distribution (4) of external forces implies the modified form of (1):

$$-\nabla p + \mu\nabla^2\mathbf{u} + \mathbf{F}\delta(\mathbf{r}) = 0, \quad (5)$$

which with (2) gives

$$\nabla^2 p = \nabla \cdot [\mathbf{F}\delta(\mathbf{r})]. \quad (6)$$

The solution of (6) in unbounded fluid is the classical *dipole field*

$$p = -\nabla \cdot [\mathbf{F}/4\pi r] \quad (7)$$

of *strength* equal to the external force  $\mathbf{F}$  applied to the fluid at  $r = 0$ . This conclusion is an interesting parallel with the theory of sound, where however the corresponding *velocity* field is quite different, being found by equating  $-\rho\partial\mathbf{u}/\partial t$  (rather than  $\mu\nabla^2\mathbf{u}$ ) to  $\nabla p$ .

We determine the velocity field when

$$\mathbf{F} = (F, 0, 0) \quad (8)$$

is a force in the  $x$ -direction. Then (7) becomes

$$p = Fx/4\pi r^3, \quad (9)$$

and (5) requires that, except at the singularity  $r = 0$ ,

$$\nabla^2\mathbf{u} = \frac{1}{\mu}\nabla p = \frac{F}{4\pi\mu}\left(\frac{1}{r^3} - \frac{3x^2}{r^5}, -\frac{3xy}{r^5}, -\frac{3xz}{r^5}\right). \quad (10)$$

There is a well-known technique for determining a biharmonic function whose Laplacian is equal to a given harmonic function that is homogeneous of degree  $n$ . It depends on the identity

$$\begin{aligned}\nabla^2(r^2\phi) &= r^2\nabla^2\phi + \phi\nabla^2(r^2) + 2\nabla\phi \cdot \nabla(r^2) \\ &= r^2\nabla^2\phi + 6\phi + 4\mathbf{r} \cdot \nabla\phi,\end{aligned}$$

which if  $\nabla^2\phi = 0$  and  $\phi$  is homogeneous of degree  $n$  becomes  $(6 + 4n)\phi$ . Hence  $(6 + 4n)^{-1}r^2\phi$  has Laplacian equal to  $\phi$ .

The right-hand side of (10) is homogeneous of degree  $-3$ , so a particular solution of (10) for  $u$  is

$$\mathbf{u} = -\frac{1}{6}r^2\left(\frac{F}{4\pi\mu}\nabla\frac{x}{r^3}\right) = \frac{F}{8\pi\mu}\left(\frac{x^2 - \frac{1}{3}r^2}{r^3}, \frac{xy}{r^3}, \frac{xz}{r^3}\right). \quad (11)$$

This as it stands does not satisfy the equation of continuity (2) however; indeed,

$$\begin{aligned}\nabla \cdot [r^2\nabla(x/r^3)] &= r^2\nabla^2(x/r^3) + \nabla(r^2) \cdot \nabla(x/r^3) \\ &= 0 + 2r \cdot \nabla(x/r^3),\end{aligned} \quad (12)$$

which, since  $x/r^3$  is homogeneous of degree  $-2$ , is

$$-4x/r^3 = \nabla \cdot (4/r, 0, 0). \quad (13)$$

It follows that by adding the vector  $(F/6\pi\mu r, 0, 0)$  to (11) we obtain a sum that does satisfy the equation of continuity (2), while not of course changing  $\nabla^2\mathbf{u}$  from its value (10) since  $1/r$  is a harmonic function.

The velocity and pressure fields associated with the concentrated force (8) at  $r = 0$  are therefore

$$\mathbf{u} = \frac{F}{8\pi\mu}\left(\frac{x^2 + r^2}{r^3}, \frac{xy}{r^3}, \frac{xz}{r^3}\right), \quad p = \frac{Fx}{4\pi r^3}. \quad (14)$$

At each point the *radial* component of the velocity field is

$$u_r = (\mathbf{r} \cdot \mathbf{u})/r = (F/8\pi\mu)(2 \cos \theta/r), \quad (15)$$

where  $\theta$  is a spherical polar coordinate defined so that  $\cos \theta = x/r$ , and the transverse component  $u_\theta$  is

$$u_\theta = (-y^2 - z^2, xy, xz) \cdot \mathbf{u}/r^2 \sin \theta = (F/8\pi\mu)(-\sin \theta/r). \quad (16)$$

Two noteworthy features of these equations are illustrated in Figure 1. One is that *only* the factor 2 in (15) prevents them from representing a unidirectional velocity field in the  $x$ -direction, with magnitude equal at all points on a spherical surface of radius  $r$ . We note also the similarity to the flow field  $\nabla\phi$  derived from a dipole *velocity potential*

$$\phi = Gx/4\pi r^3 \quad (17)$$

with dipole strength  $(G, 0, 0)$  which generates velocity and pressure fields

$$\mathbf{u} = \frac{G}{4\pi} \left( \frac{1 - 3x^2}{r^3}, -\frac{3xy}{r^5}, -\frac{3xz}{r^5} \right), \quad p = 0, \quad (18)$$

since  $\mathbf{u} = \nabla\phi$  satisfies  $\nabla^2\mathbf{u} = 0$ . This velocity field has radial and transverse components

$$u_r = (-G/4\pi)(2 \cos \theta/r^3), \quad u_\theta = (-G/4\pi)(\sin \theta/r^3), \quad (19)$$

which decay like  $1/r^3$  instead of like  $1/r$ , but otherwise differs from the velocity field (15) and (16) of a concentrated force only through a difference of sign of one of the components.

With any linear partial differential equation, the building-up of solutions from fundamental singular solutions is to be recommended as an effective route towards satisfying boundary conditions. For the Stokes equations (1) and (2) of inertialess fluid motions, the two singular solutions that have proved especially fruitful in this respect have been (14) and (18). The solution (14) representing the effect of a concentrated external force  $(F, 0, 0)$  acting at the origin has been named a *stokeslet* of strength  $(F, 0, 0)$ . The solution (18) involves no external force on the fluid; it is called a dipole velocity field of strength  $(G, 0, 0)$  or sometimes a *source-doublet*.

An immediate illustration of their use is obtained if to the stokeslet field (Figure 1a) we add just that multiple of the dipole-field (Figure 1b) which on the spherical

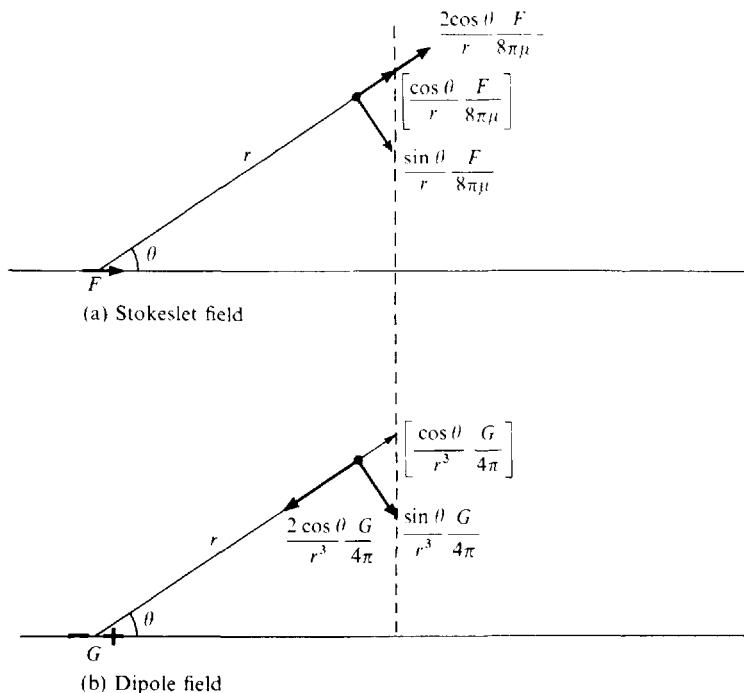


FIG. 1. Illustrating the velocity components in spherical polar coordinates for (a) a stokeslet field of strength  $F$ ; (b) a dipole field of strength  $G$ .

surface  $r = a$  will make the velocity field unidirectional and of uniform magnitude. Figure 1 shows us that the choice of dipole strength

$$G = Fa^2/6\mu \quad (20)$$

has the effect on  $r = a$  of changing the stokeslet velocity field into a *uniform* velocity  $F/6\pi\mu a$  parallel to the direction of the applied force, since it reduces the radial component to this multiple of  $\cos\theta$  and increases the transverse component to the same multiple of  $(-\sin\theta)$ . Those preferring an analytical to a geometrical derivation will equally see the result as evident from (14) and (18).

Thus we have rediscovered Stokes's law that the external force  $F$  required to move a sphere of radius  $a$  through a fluid of viscosity  $\mu$  at velocity  $U$  is given by the equation

$$F = 6\pi\mu a U. \quad (21)$$

For we can exactly satisfy the boundary condition that the velocity of the fluid equals that of the solid surface, say  $(U, 0, 0)$ , on  $r = a$  by means of a stokeslet of strength  $(F, 0, 0)$  at  $r = 0$ , representing an external force  $(F, 0, 0)$  on the fluid, plus a dipole of strength (20) involving no application of an external force.

### 3. Line distributions of singular solutions

We consider now methods of using the singular solutions (14) and (18) to study motions of microscopic organisms propelled (see Chapter 2) by a flagellum. Stokes's sphere law (21) is already of relevance to those organisms where the drag of a head of nearly spherical shape is balanced by the thrust from an undulatory movement of the flagellum. It must however be supplemented by information about the force between the fluid and an organelle with the long slender shape of a flagellum: nothing like a sphere but much more like a long cylinder.

In this context it is worth remembering that no law such as (21) exists for the drag of an *infinitely* long cylinder. The famous Stokes paradox states that the equations (1) and (2) of inertialess flow possess no solution satisfying uniform boundary conditions on an infinite circular cylinder. We can understand this, and also learn much that is relevant to flagellar motions, if we study the flow field due to a distribution of stokeslets along a finite line segment.

We investigate, then, the integrated effect of a distribution of stokeslets (14) spread between, say,  $Z = -b$  and  $Z = c$  with the total strength  $F$  in an interval  $(Z, Z + dZ)$  equal to  $(fdZ, 0, 0)$ . Thus the stokeslet strength, still in the  $x$ -direction, has uniform magnitude  $f$  per unit length. The flow field is obtained by integrating (14) from  $Z = -b$  to  $Z = c$ , with  $z$  replaced by  $z - Z$  and  $F$  replaced by  $fdZ$ .

We evaluate that flow field on the cross-section  $z = 0$  of a circular cylinder

$$x^2 + y^2 = a^2 \quad (22)$$

whose axis is the line of stokeslets and whose radius  $a$  is small compared with both  $b$  and  $c$ . (Thus, we have chosen the origin of  $z$  on that cross-section where we are evaluating the flow field, and we are denoting by  $b$  and  $c$  the distances of that cross-section from the two ends of the distribution.) In (14), when (22) is satisfied and  $z$  is

replaced by  $z - Z$ , the value of  $r^2$  on  $z = 0$  is  $Z^2 + a^2$ . This makes

$$\int_{-b}^c \frac{dZ}{r^3} \doteq \frac{2}{a^2}, \quad \int_{-b}^c \frac{Z dZ}{r^3} \doteq 0, \quad (23)$$

since with  $b$  and  $c$  large compared with  $a$  both these integrals take values close to the convergent integrals from  $-\infty$  to  $\infty$ . The corresponding integral of  $1/r$ , however, does not converge at infinity. Its exact and approximate values are

$$\int_{-b}^c \frac{dZ}{r} = \sinh^{-1}\left(\frac{c}{a}\right) + \sinh^{-1}\left(\frac{b}{a}\right) \doteq \log\left(\frac{4cb}{a^2}\right), \quad (24)$$

where the approximation  $\log(2x)$  to the inverse hyperbolic sine function  $\sinh^{-1} x$  has an error of order  $x^{-2}$  as  $x \rightarrow \infty$ . Using all this information in (14), we obtain an approximate velocity field on the circle  $x^2 + y^2 = a^2$ ,  $z = 0$  as

$$\mathbf{u} = \frac{f}{8\pi\mu} \left( \frac{2x^2}{a^2} + \log \frac{4cb}{a^2}, \frac{2xy}{a^2}, 0 \right). \quad (25)$$

Although this velocity field due to a line distribution of stokeslets varies around the circle, it is possible just as with the sphere to remove that variation by combining it with a similar arrangement of dipole velocity fields, of strength  $(g dZ, 0, 0)$  in each interval  $(Z, Z + dZ)$ . The corresponding flow field can be inferred from (18) exactly as above, using the additional integral relationships

$$\int_{-b}^c \frac{dZ}{r^5} \doteq \frac{4}{3a^4}, \quad \int_{-b}^c \frac{Z dZ}{r^5} \doteq 0, \quad (26)$$

and the approximate velocity field takes the form

$$\mathbf{u} = \frac{g}{4\pi} \left( \frac{2}{a^2} - \frac{4x^2}{a^4}, -\frac{4xy}{a^4}, 0 \right) \quad (27)$$

independent of  $b$  and  $c$ . If we choose

$$g = fa^2/4\mu \quad (28)$$

(not, in fact, the same choice (20) that was needed for the sphere problem) we obtain as the sum of the velocity fields (25) and (27)

$$\mathbf{u} = \frac{f}{8\pi\mu} \left( 1 + \log \frac{4cb}{a^2}, 0, 0 \right); \quad (29)$$

a unidirectional velocity, uniform around the cylinder.

Here the stokeslets of strength  $(f, 0, 0)$  per unit length correspond to forces  $(f, 0, 0)$  per unit length applied to the fluid. Thus, a circular cylinder of radius  $a$  and length  $b + c \gg a$  exerts a uniform force  $(f, 0, 0)$  on the fluid per unit length of cylinder when a typical section of it (the section  $z = 0$  in the above analysis)

moves with the velocity (29), where  $b$  and  $c$  are the distances of the section from the ends of the cylinder. Of course the Stokes paradox is connected with the lack of convergence in (24) and the failure of (29) to tend to a limit as  $b$  and  $c$  tend to infinity.

Indeed, there is a slow (logarithmic) dependence of the force  $f$  per unit length, required to give a certain motion to a circular cylinder, on the actual value of the ratio of length to radius. Furthermore, a uniform force  $f$  per unit length produces a slow (logarithmic) variation in the velocity (29) generated at different distances  $b, c$  from the ends of the cylinder. This rises to a maximum at the midpoint where  $bc/a^2$  is greatest, possibly due to the pushing effect of an extensive distribution of forces on both sides, and becomes rather smaller near the ends, but the percentage variation is only slight over (say) 80% of the length of the cylinder (we omit the calculation of what happens when either  $b$  or  $c$  is no longer large compared with  $a$ ).

It is very interesting to compare the above conclusions for cylinders moving normal to their axes with the corresponding ones for tangential motion. To calculate these from the flow field (14) for a stokeslet with vector strength parallel to the  $x$ -axis, we take that strength to be  $(fdX, 0, 0)$  in each element of the segment  $-b < X < c$  of the  $X$ -axis, replace  $x$  by  $x - X$  in (14) and integrate on the circle  $y^2 + z^2 = a^2$ ,  $x = 0$  using the results (23) and (24) with the  $x$ - and  $z$ -coordinates interchanged to give

$$\mathbf{u} = \frac{f}{8\pi\mu} \left( -2 + 2 \log \frac{4cb}{a^2}, 0, 0 \right). \quad (30)$$

Note that the highly significant factors of 2 in (30) arise from the fact that in (14) an  $(x - X)^2 + r^2$  numerator can be replaced by  $-a^2 + 2r^2$  on  $y^2 + z^2 = a^2$ .

With a line distribution of *tangential* stokeslets we see that the achievement of a unidirectional velocity on the cylinder  $y^2 + z^2 = a^2$ , uniform over the section at  $x = 0$ , requires no addition of any dipole distribution. Other aspects of the conclusions are qualitatively as before: the velocity generated is in the direction of the stokeslets (that is, of the force being applied) and shows a slow logarithmic dependence on the ratio of cylinder length  $b + c$  to cylinder radius  $a$ ; furthermore, it exhibits a slow (logarithmic) variation with the distances  $b, c$  from the ends of the cylinder, taking its maximum value at the midpoint as before.

Quantitatively, however, there is a significant difference between the magnitudes of the velocity fields (29) and (30) produced by forces  $f$  per unit length acting respectively in directions normal and tangential to the cylinder. If  $\log(4cb/a^2)$  is large, their ratio is almost a factor of 2: given forces allow the cylinder to move almost twice as fast when applied tangentially as when they are applied normally.

This result is often stated in terms of *resistances* to motion. In motion of a cylinder normal to itself the ratio between the force  $f$  per unit length that must be applied and the local velocity  $u$  of cylinder that results can be described as a coefficient  $K_N$  of resistance to normal motions of the cylinder satisfying

$$K_N = 8\pi\mu / \left( 1 + \log \frac{4cb}{a^2} \right). \quad (31)$$

It is a crude concept whose approximate character is emphasized by the fact that the velocity generated by a uniform distribution of force varies along the cylinder (although only slightly along most of its length). It may nevertheless have real value for analysing microscopic motions that can only be measured rather crudely.

From (30) we see that the corresponding tangential resistance coefficient  $K_T$  (ratio of the tangential force  $f$  applied per unit length of cylinder to the resulting local tangential velocity  $u$  of the cylinder surface) is

$$K_T = 4\pi\mu \left/ \left( -1 + \log \frac{4cb}{a^2} \right) \right.. \quad (32)$$

The suggestion from (31) and (32) that the resistance to normal motions of a cylinder significantly exceeds the resistance to tangential motions helps in understanding the propulsive effect of undulatory movements of a flagellum. This is because transverse undulations in a flagellum, in which normal motions have components in the direction of wave motion, exert a greater force on the fluid in that direction than do the tangential motions with components in the opposite direction. The net reaction on the organisms is then available for overcoming its resistance to bulk motion through the fluid.

On this line of argument it is principally the ratio

$$r_K = K_T/K_N = \frac{1}{2} \left/ \left( 1 + \log \frac{4cb}{a^2} \right) \right/ \left( -1 + \log \frac{4cb}{a^2} \right) \quad (33)$$

of tangential resistance to normal resistance (a ratio used already but without any subscript in Chapter 2, § 13) that determines the swimming speed generated by given undulatory movements. Fortunately, the value of this ratio does not depend very sensitively on  $b/a$  and on  $c/a$ : it varies only between 0.6 and 0.7 as  $(cb)^{1/2}/a$  varies between 120 and 10.

Nevertheless, it is desirable to consider what value of  $(cb)^{1/2}$  (and hence what value of  $r_K$ ) may be appropriate to the description of an undulation of wavelength  $\lambda$ . Since the force per unit length, far from being uniform, undulates with period  $\lambda$ , this measure  $(cb)^{1/2}$  of the distance from a given point of the flagellum within which it is effectively constant should presumably be a relatively small fraction of  $\lambda$ . Actually a simple calculation shows that fraction to be  $0.09\lambda$  for undulations of small amplitude. This depends on the idea that when in the calculations of this section  $f$  is replaced by  $fe^{ikZ}$ , where  $k = 2\pi/\lambda$  and  $\lambda \gg a$ , the additional  $e^{ikZ}$  factor changes none of the integrals in (23) and (26), but changes (24) to an integral that converges, even with the limits replaced by  $(-\infty, \infty)$ , to the value

$$2K_0(ka) \doteq 2[\log(2/ka) - \gamma], \quad (34)$$

where  $K_0$  is the modified Bessel function and  $\gamma = 0.577$  is Euler's constant. This agrees with (24) if

$$(cb)^{1/2} = e^{-\gamma}/k = \lambda/(2\pi e^\gamma) = 0.09 \lambda. \quad (35)$$

The above calculation leading to a lower value (35) than has commonly been adopted depends on the idea that in steady swimming by an undulation of small

amplitude the stokeslet strength has very small components in the direction of swimming but substantial undulatory components in the direction at right angles. It follows that the *local effect* of the stokeslet components in the *local normal* direction (or the local tangential direction) is given to good approximation by (31) (or (32) respectively) with  $(cb)^{1/2}$  determined by (35). Even for undulations of large amplitude it may be reasonable to use this result for a straight cylinder acting on the fluid with an undulatory distribution of force, since within a distance  $0.09\lambda$  of a given point the departure from straightness is only moderate.

#### 4. Resistive-force theory of flagellar propulsion

The results of § 3 can be applied to the study of flagellar propulsion in two alternative ways:

- (i) approximate use for each section of the flagellum of the effective resistance coefficients (31) and (32) for normal and tangential motion of the section, with a suitable choice such as (35) made for  $(cb)^{1/2}$ ; this is the method pioneered by Gray & Hancock (1955) whose results are given in an extended form in the present section;
- (ii) a more accurate but considerably more complicated method due to Hancock (1953), which is sketched in § 6.

We consider a flagellum, whose length is  $L$  when stretched straight, and which may or may not be propelling a cell body. We suppose that the waveform passing backwards along the flagellum has the general form

$$(x, y, z) = (X(s), Y(s), Z(s)), \quad (36)$$

where  $s$  is a parameter representing distance along the flagellum from the head, so that

$$X'^2(s) + Y'^2(s) + Z'^2(s) = 1. \quad (37)$$

The functions (36) are taken to have period  $\Lambda$ , in such a way that

$$Y(s + \Lambda) = Y(s), \quad Z(s + \Lambda) = Z(s), \quad \text{but} \quad X(s + \Lambda) = X(s) + \lambda. \quad (38)$$

Thus  $\Lambda$  is a wavelength measured along the curved body of the organism, but  $\lambda = \alpha\Lambda$  is the wavelength measured in the direction of propulsion: here  $\alpha$  is a factor  $< 1$  which specifies the contraction in the organism's apparent length due to waviness.

In a frame of reference *travelling with the propulsive wave*, an inextensible flagellum moves purely *tangentially* along the waveform (36), at a uniform velocity  $c$ , so that at time  $t$ ,

$$(x, y, z) = (X(s - ct), Y(s - ct), Z(s - ct)). \quad (39)$$

Note that  $c$  is the speed with which the wave passes along the curved body of the

organism. This must be distinguished from  $V$ , defined as the apparent wave velocity in a frame of reference in which the organism has no net forward velocity. Actually,

$$V = \alpha c, \quad (40)$$

since the period of undulation can be written as either  $\Lambda/c$  or  $\lambda/V = \alpha\Lambda/V$ .

If the swimming speed of the animal is  $U$  (in the opposite direction to the wave velocity (40)) then the propulsive wave moves backwards at a velocity  $V - U$  relative to the fluid. The velocity relative to the fluid of any section of the flagellum is then the vector sum of the velocity  $c$  directed along the forward tangent (this was noted already as the motion in a frame of reference travelling with the wave) and a velocity  $(V - U, 0, 0)$ . Its component along the backward tangent, then, is

$$(V - U)X'(s - ct) - c \quad (41)$$

while its component along the backward normal is

$$(V - U)[1 - X'^2(s - ct)]^{1/2}. \quad (42)$$

Here  $X'$  and  $(1 - X'^2)^{1/2}$  are the direction cosines of the tangent and the normal respectively.

The total thrust  $P$  on the flagellum can now be written as the  $x$ -component of the sum of the tangential force resisting (41) with resistance coefficient  $K_T$  and the normal force resisting (42) with resistance coefficient  $K_N$ . Taking the  $x$ -component multiplies the tangential force by a further  $X'$  factor and the normal force by a further  $(1 - X'^2)^{1/2}$  factor. Hence,

$$\begin{aligned} P = & \int_0^L \{ K_T[(V - U)X'(s - ct) - c]X'(s - ct) \\ & + K_N(V - U)[1 - X'^2(s - ct)] \} ds. \end{aligned} \quad (43)$$

In (43) we use the fact that

$$\int_0^L X'(s - ct) ds = l = \alpha L = VL/c, \quad (44)$$

where  $l$  is the apparent length of the undulatory organism measured in the direction of propulsion, so that  $l/L = \alpha$ . We also define

$$\int_0^L X'^2(s - ct) ds = \beta L, \quad (45)$$

where  $\beta$  is a mean value of the square of the tangential direction cosine. With this notation,

$$P = K_T L[(V - U)\beta - V] + K_N L(V - U)(1 - \beta). \quad (46)$$

When the flagellum propels an inert head, this thrust must balance the force resisting forward motion of the head at the swimming velocity  $U$ . We write this

force as  $K_N LU\delta$ , where  $\delta$  is a coefficient defined as

$$\delta = \frac{\text{resistance to forward motion of the head}}{\text{resistance to uniform normal motion of whole flagellum}}. \quad (47)$$

The coefficient  $\delta$  is usually a rather small fraction for flagella of typical length, as is suggested by comparing the formula (21) for the drag on a sphere of radius  $a$  to the formula (31) for  $K_N$ .

Equating (46) to  $K_N LU\delta$  we obtain the expression

$$\frac{U}{V} = \frac{(1 - \beta)(1 - r_K)}{1 - \beta + r_K\beta + \delta} \quad (48)$$

for the ratio of swimming speed to wave speed. As  $\beta$  (the mean square tangential direction cosine) decreases from 1 (its value for no swimming motions) this ratio  $U/V$  increases from 0 but cannot exceed

$$\left(\frac{U}{V}\right)_{\max} = \frac{1 - r_K}{1 + \delta}, \quad (49)$$

its value in the limit  $\beta \rightarrow 0$ .

Gray & Hancock (1955) obtained these results in the case  $r_K = \frac{1}{2}$ : a limiting case for an exceedingly thin flagellum suggested by (33). We prefer to write them for general  $r_K$  in order to be able to gauge the sensitivity of the conclusions to the value of that parameter; especially since even for  $(cb)^{1/2}/a = 120$  the predicted value of  $r_K$  still exceeds  $\frac{1}{2}$  by 20%. Gray & Hancock (1955) gave data (Figure 2) suggesting that for the spermatozoa of a sea-urchin (*Psammechinus miliaris*) the wave was approximately a sine-wave of length  $\lambda = 24 \mu\text{m}$  and amplitude  $b = 4 \mu\text{m}$ , and proposed an approximate value  $0.2 \mu\text{m}$  for the radius of the tail.

With our improved value for  $K_N$  suggested by equations (31) and (35) these data make

$$\delta = \frac{6\pi\mu A}{8\pi\mu L} \left[ 1 + 2 \log \frac{(2)(0.09)(24 \mu\text{m})}{(0.2 \mu\text{m})} \right] \div \frac{5A}{L}, \quad (50)$$

where  $A$  is the radius (or effective radius) of the head. Figure 2 suggests  $A = 1 \mu\text{m}$ ,  $L = 45 \mu\text{m}$  giving  $\delta = 0.11$ . At the same time the value of  $\beta$  is close to 0.65, as calculated by Gray & Hancock (1955) through an approximation  $\beta \div (1 + 2\pi^2 b^2 / \lambda^2)^{-1}$ .

By contrast, the value  $r_K = 0.5$  which they used is far away from the value  $r_K = 0.7$  suggested by equations (33) and (35). The value (48) of the swimming velocity ratio depends critically on  $r_K$ : for  $r_K = 0.5$  it takes a value 0.23 in the midst of the range of ratios of swimming velocity to wave velocity observed for these spermatozoa, whereas the estimate (48) with  $r_K = 0.7$ , though more soundly based theoretically, takes a value only half as great, far below the observed values. Gray & Hancock were aware of such a paradox although they used the alternative method of § 6 below to obtain an improved calculation instead of merely changing the value of  $r_K$ . While noting that the observations might be influenced in an

Table I

Sperm no.	Length of envelope in $\mu$	Wave-length ( $\lambda$ ) in $\mu$	Amplitude ( $b$ ) in $\mu$	Total length of tail in $\mu$	Ratio $b/\lambda$
1	35	23	4	45	0.174
2	33	24	4	38	0.167
3	34	27	3.5	42	0.130
4	35	27	4	42	0.148
5	35	23	4	40	0.174
6	34	21	4	42	0.190
7	32	24	3.5	42	0.146
8	34	24	4	42	0.167
9	32	22	3.8	40	0.170
10	33	24	3.8	40	0.156
11	32	23	4	38	0.174
12	35	29	5	40	0.172
13	29	24	5	40	0.209
14	30	21	3.2	40	0.160
Average	33	24	4	41	0.167

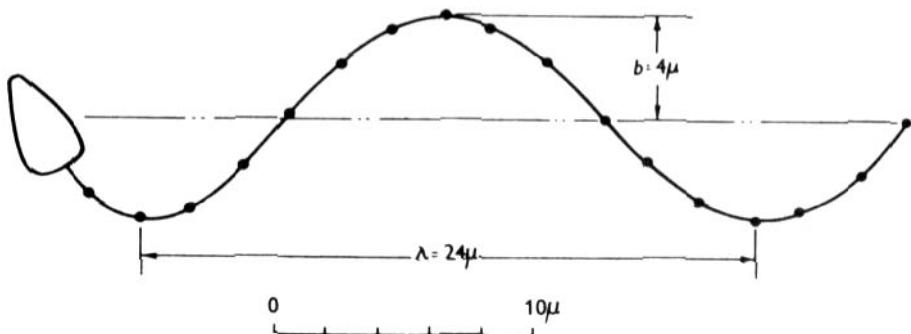


FIG. 2. This Table I and Text-figure 4 from Gray & Hancock (1955) illustrate their method of determining a typical sinusoidal representation for a flagellar wave.

unknown way by the fact that the spermatozoon was swimming in very close proximity to the glass of the microscope slide, they concluded that the balance of evidence favoured the view that the theory with  $r_K = 0.5$  was closely correct for the swimming of these spermatozoa.

Hancock (1953, p. 114) had come to the same paradoxical conclusion for the swimming of a nematode, *Turbatrix aceti*, the flow around which is made visible in Figure 3 (the sequence of swimming movements having been given in Figure 1 of Chapter 2). Two out of three observations on *Turbatrix* were found to be consistent with values of  $V/U$  corresponding to  $r_K = 0.5$  (the third, actually, would correspond to a still smaller value of  $r_K$ ). Hancock suggested that this 'zero-radius' value  $r_K = 0.5$  might be closely correct for flagella, because they

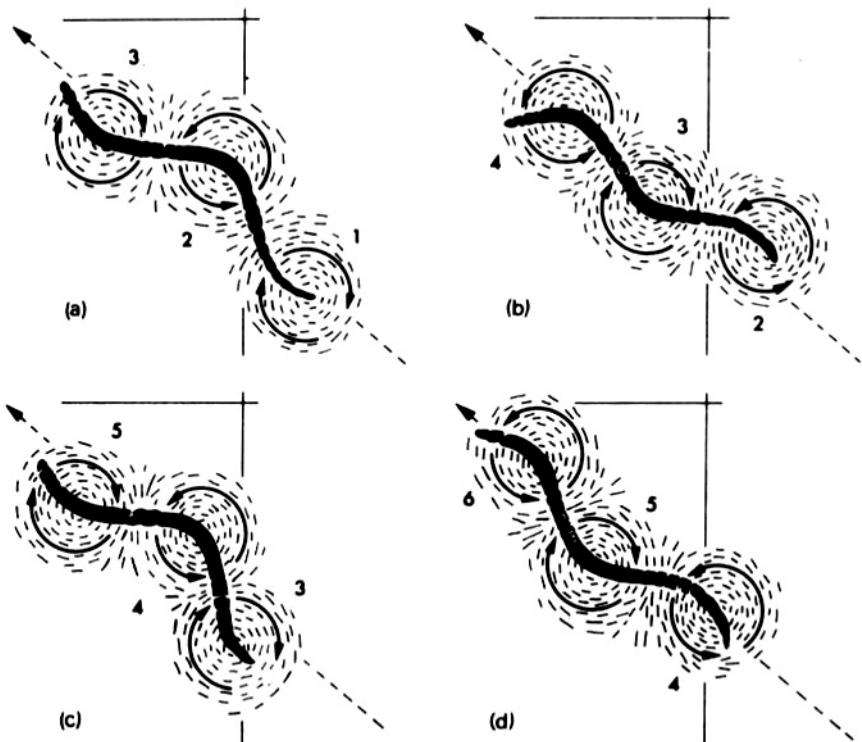


FIG. 3. This Figure 16.12 from Gray (1968) gives a determination of flow directions (suggesting the streamline pattern) in the fluid flow around a swimming nematode.

might be so flexible (responding almost like the liquid itself in their outer portions) that their effective radius might be far less than their geometric radius.

This resolution of the paradox, always uncertain because (see § 3) the value (33) of  $r_K$  shows only a slow (logarithmic) response to the radius  $a$ , has since become unacceptable through increased knowledge of the microstructure of flagella. The true resolution will be given in Chapter 7, where the effect of swimming in close proximity to a microscope slide is indeed found to be a substantial reduction in the value of  $r_K$ .

By contrast, an accurately observed experiment by Taylor, using a self-propelled mechanical device (of a size convenient for accurate observation) in a large volume of fluid, sufficiently viscous to ensure a very low Reynolds number, gave results for  $U/V$  substantially below the values for  $r_K = 0.5$ : indeed (see Figure 4) with  $\lambda/a = 100$  they are consistent with values for  $r_K = 0.7$  as suggested by (33) and (35). In these experiments the 'head' contained a motor which steadily rotated a wire 'tail' of spiral shape. Such a spiral makes a uniform angle  $\psi$  with the direction of motion, so that the tangential direction cosine  $X'$  takes a uniform value  $\cos \psi$  and (44) and (45) become

$$\alpha = \cos \psi, \quad \beta = \cos^2 \psi. \quad (51)$$

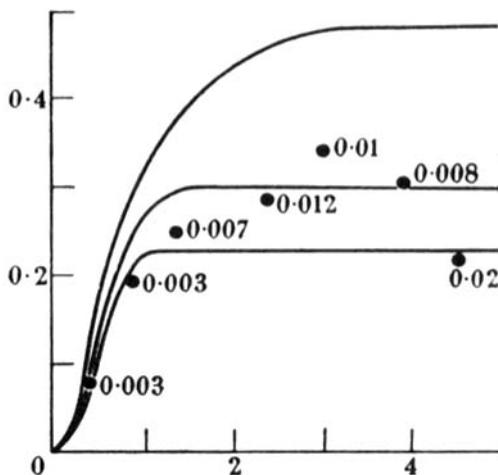


FIG. 4. Plot of  $U/V$  against  $2\pi b/\lambda$ . (where  $b$  is amplitude of spiral motion) for Taylor's propulsive spiral wire, compared with theory by Hancock (1953). The upper, middle and lower curves give theoretical data for  $a/\lambda = 0$ , 0.01 and 0.02, respectively; the value of  $a/\lambda$  appropriate to each experimental point is marked against it.

This motion, to be sure, is one where the system of forces on the spiral flagellum ( $K_T$  times (41) plus  $K_N$  times (42)) produces not only a resultant force (the thrust) but also a resultant twisting couple about the axis of the spiral. Taylor used an eccentric weight distribution in the 'head' of his device to cancel out this twisting couple.

It has since become increasingly clear that flagellar propulsion of many organisms takes place by means of a spiral wave. It is, however, most unlikely that features in the cell body of those organisms are able to reduce significantly the twisting couple due to the flagellar movement and, indeed, some such organisms such as spirochetes possess no 'head' at all. Recently Chwang & Wu (1971) pointed out that in all these cases the effect of the twisting couple can only be to generate what they call 'body rotation': a twisting of the whole flagellum (and of the head if any) about its own (curved) centreline. The rate of twist must be such that the couples resisting it have a resultant exactly cancelling those due to the normal and tangential forces. The micro-structure of flagella, involving nine independently extensible and contractible fibrils in a circular arrangement around the two central fibrils, would appear to allow such body rotation.

When a circular cylinder of radius  $a$  rotates with angular velocity  $\omega$  in a viscous fluid, the resulting velocity field is that due to a line vortex of strength  $2\pi a^2 \omega$  stretched along the axis. Both the viscous force on any fluid element and the pressure gradient are zero, but the rotation is resisted by a viscous stress  $2\mu\omega$  at the cylinder surface, producing no net force between the cylinder and the fluid but a resisting couple  $4\pi a^2 \mu\omega$  per unit length.

For a flagellum propagating a spiral wave we can similarly assume that any self-rotation will leave the *forces* between the flagellum and the fluid unaffected, while being resisted by a couple  $4\pi a^2 \mu \omega$  per unit length about the centreline. This generates a resultant couple

$$4\pi a^2 L \mu \omega \cos \psi \quad (52)$$

about the axis of the spiral. The value of  $\omega$  must be determined so that this quantity (52) together with the couple resisting twisting of the head (if any†) cancels the moment of the tangential and normal forces on the flagellum. Those forces themselves, however, together with the thrust that they generate, are uninfluenced by the body rotation and may be calculated by the methods described in this section.

## 5. Swimming motions with minimum rate of working

Having described a rather general theory of swimming by undulatory motions at low Reynolds number, we now consider the possibility of determining swimming motions for maximum 'efficiency' just as was done on the basis of a much cruder theory in Chapter 2 (see especially equation (5)). The idea is to find motions that achieve a given swimming speed  $U$  with minimum rate of working.

In the motion described in § 4, the rate of working by the animal against the different resistances (to the tangential motion (41), the normal motion (42), and the motion of the head) is

$$\begin{aligned} E = & \int_0^L \{ K_T [(V - U) X'(s - ct) - c]^2 \\ & + K_N (V - U)^2 [1 - X'^2(s - ct)] \} ds + K_N L U^2 \delta, \end{aligned} \quad (53)$$

since for example the tangential force per unit length  $K_T$  times (41) must be multiplied again by the tangential velocity to give rate of working, and *squares* of velocities appear for just the same reasons in the other terms. The rate of working (53) has been calculated in the frame of reference in which the fluid is at rest, but since the resultant force exerted by the animal is zero the value of  $E$  must be the same in any frame of reference.

The dependence of  $E$  on squares of velocities implies that we could study the problem of minimising  $E$  for given  $U$  by seeking a minimum of  $E/U^2$ . Actually we shall work with a nondimensional multiple of this, defined as

$$\eta^{-1} = E / (K_T + K_N \delta) L U^2; \quad (54)$$

---

† We may note here that rotation of a sphere of radius  $A$  is resisted by a couple  $8\pi A^3 \mu \omega$ , which a comparison with (52) suggests may be dominant in determining  $\omega$  for many cases of flagellar propulsion of a large cell body.

namely, the ratio of the animal's rate of working at the swimming speed  $U$  to the rate of working required by an external force to drag the animal through the water at that speed with its flagellum stretched straight behind it. The quantity  $\eta$  is of the nature of a Froude efficiency.

The integrals in (53) can be simply expressed in terms of the definitions (44) and (45) of  $\alpha$  and  $\beta$ , giving

$$\begin{aligned} E = & K_T L[(V - U)^2 \beta - 2(V - U)c\alpha + c^2] \\ & + K_N L(V - U)^2(1 - \beta) + K_N L U^2 \delta. \end{aligned} \quad (55)$$

Hence, expressing  $c$  in terms of  $V$  by (40) and  $V$  in terms of  $U$  by (48), we obtain

$$\begin{aligned} \eta^{-1} = & (1 - r_K)^{-2}(1 - \beta)^{-2}[(r_K + \delta)(1 - \beta + \beta r_K) - 2r_K(1 - \beta + r_K\beta + \delta)] \\ & + \delta(r_K + \delta)^{-1} + \alpha^{-2}r_K(1 - r_K)^{-2}(1 - \beta)^2(r_K + \delta)^{-1}(1 - \beta + r_K\beta + \delta)^2. \end{aligned} \quad (56)$$

The two waveform parameters that appear in (56) are the mean values  $\alpha$  and  $\beta$  of the tangential direction cosine and of its square. We now proceed to calculate the minimum of (56), and begin by obtaining the minimum for given  $\beta$ .

We use the theorem that the mean square of any function is not less than the square of its mean, with equality only when the function is constant (more strictly, constant except in a set of measure zero). This 'Schwarz inequality' applied to the tangential direction cosine tells us that

$$\alpha^2 \leq \beta, \quad (57)$$

with equality for constant direction cosine. It follows that for given  $\beta$  the minimum of  $\alpha^{-2}$  is  $\beta^{-1}$  so that the minimum of (56) is obtained by substituting  $\beta^{-1}$  for  $\alpha^{-2}$  in the last term. This minimum, obtained whenever the tangential direction cosine is constant except in a set of measure zero, takes the simplified form

$$\min_{\alpha} (\eta^{-1}) = \frac{1}{(1 - r_K)^2} \left[ \frac{r_K(1 + \delta)^2}{(r_K + \delta)\beta} + \frac{r_K + \delta}{1 - \beta} \right] + \frac{\delta}{r_K + \delta}. \quad (58)$$

A final minimisation of (58) with respect to  $\beta$  is achieved when

$$\frac{\beta}{1 - \beta} = \frac{r_K^{1/2}(1 + \delta)}{r_K + \delta}. \quad (59)$$

This determination of  $\beta$  gives the maximum efficiency  $\eta_m$  as

$$\eta_m = (1 - r_K^{1/2})^2 / (1 + \delta) \quad (60)$$

and the ratio of swimming speed to wave velocity as

$$U/V = (1 - r_K^{1/2}) / (1 + \delta). \quad (61)$$

It is interesting that the maximum efficiency (60), and associated swimming speed ratio (61), take in the case  $\delta = 0$  of a flagellum without a head exactly the values obtained for that case in equation (5) of Chapter 2 by a much cruder procedure. Probably this is because the assumption there made, that the undulation

involves only transverse motion, is actually correct in the optimum swimming motion with tangential direction cosine constant. These conclusions confirm all the remarks made in Chapter 2 about the critical dependence of both quantities on the value of the resistance ratio  $r_K$  (there written  $r$ ). For a flagellum propelling a cell body with small positive  $\delta$ , both quantities are slightly reduced by the same factor  $1 + \delta$ .

Since maximum efficiency is achieved in those swimming modes with the tangential direction cosine constant (except perhaps in a set of measure zero), the angle  $\psi$  which the tangent makes with the direction of swimming must also be constant, with  $\alpha = \cos \psi$  and  $\beta = \cos^2 \psi$  as in (51). The optimal value of  $\psi$  is

$$\psi_0 = \tan^{-1}[r_K^{1/4}(1 + \delta r_K^{-1})^{1/2}(1 + \delta)^{-1/2}] \quad (62)$$

by (59). This optimal angle  $\psi_0$  is exceedingly insensitive to the values of  $r_K$  and of  $\delta$ : as  $r_K$  increases from 0.5 to 0.7 it increases from  $40^\circ$  to  $42^\circ$  when  $\delta = 0$ , from  $41^\circ$  to  $43^\circ$  when  $\delta = 0.1$ , and from  $42^\circ$  to  $43^\circ$  when  $\delta = 0.2$ . A typical value may be taken as  $\psi_0 = 42^\circ$ .

Undulatory motions in a single plane, for which  $\psi = \psi_0$  except in a set of measure zero, can only be of 'saw-tooth' shape: in other words the optimum two-dimensional undulation passes backwards a sequence of alternating diagonal straight segments each making an angle of about  $42^\circ$  with the direction of swimming. Probably real flagella possess some minimum achievable radius of curvature of bending, which would prevent them from achieving this saw-tooth waveform: subject to this side condition the optimum undulation is probably a saw-tooth with the corners rounded to circular arcs of that radius, a form observed for certain spermatozoa (Brokaw 1965) and close to that illustrated for a nematode in Figure 1 of Chapter 2.

On the other hand, a three-dimensional undulation which has already been noted as making  $\psi$  constant and which raises no difficulties of locally small radii of curvature is the spiral. It is interesting that a backward-moving spiral wave with the tangent to the spiral everywhere at about  $42^\circ$  to the direction of swimming satisfies the optimality condition; particularly since the discovery of Chwang & Wu (1971) that, as noted at the end of § 4, the twisting couple formerly thought to raise difficulties in this kind of mode could be cancelled by body rotation.

An example of this type of spiral flagellar propulsion is the locomotion of *Euglena viridis*, with a flagellum of length  $L = 100 \mu\text{m}$  along which a spiral wave is propagated with a wave-length  $\lambda = 35 \mu\text{m}$  and an amplitude  $b = 6 \mu\text{m}$  at a beat frequency of 12 Hz (Holwill 1966): these figures correspond to  $\psi = 47^\circ$  but, with  $b$  quoted to one significant figure, are not accurate enough to be distinguished from the theoretical optimum  $\psi_0 = 42^\circ$ . This is an animal with a relatively large cell body about  $60 \mu\text{m}$  long and  $10 \mu\text{m}$  in diameter, on which the viscous couple resisting body rotation would be large enough to keep that down to a modest value. In fact, the body is observed to gyrate at about 1 Hz. Holwill (1966) analysed former speculations regarding the possible propulsive effects of the observed gyrations of this large body but concluded that the propulsive effect of the flagellar motions of *Euglena* are at least an order of magnitude greater.

## 6. A more accurate slender-body theory

Although in biological researches the approximate theoretical approach of § 4 and § 5 may be adequate in relation to the potential accuracy of observation and the general variability of biological material, it is at least mathematically interesting to look for a theory lacking the rather arbitrary features involved in the selection of a value for  $(cb)^{1/2}/a$  and hence for the resistance ratio  $r_K$ . Such a theory was initiated by Hancock (1953) although its implications have not as yet been adequately followed through.

The Hancock method applies the general set of ideas collectively known as 'slender-body theory' (see also Batchelor 1970, Cox 1970). With a body (such as a flagellum) whose radius is extremely small compared with other relevant lengths, including its bending radius of curvature, we approximate the local distribution of flow by a two-dimensional distribution in the plane of a cross-section, whose motion, however, is recognised to take place not through undisturbed fluid but through a fluid flow generated by the far fields of the motion of other cross-sections. In this connection we recognise that the far fields of stokeslets cannot be ignored, since (14) shows them to decay only like  $r^{-1}$  as  $r \rightarrow \infty$ , the law which led to the convergence problems of (24). Dipoles, on the other hand, have negligible far fields, decaying like  $r^{-3}$ .

As described in § 1, we must suppose the flagellar motion to be a sum of the true swimming movements designated there as (i) and the rigid-body movements designated as (ii) with the swimming velocity  $\mathbf{U}$  of the mass centre and with an angular velocity  $\boldsymbol{\Omega}$  about it. Here  $\mathbf{U}$  and  $\boldsymbol{\Omega}$  are to be determined from the absence of any resultant force or couple between the organism and the fluid.

To solve this problem Hancock selects a length which he calls  $\gamma$  which is large compared with the radius  $a$  of the flagellum but small compared with its bending radius of curvature. He approximates the flow field associated with the flagellar movements (i) and (ii) by means of a distribution of fundamental singular solutions along the curved centreline of the flagellum. They include *tangential* stokeslets, with their strength per unit length specified as an unknown function of the distance  $s$  along the flagellum, and also *normal* stokeslets with a separately specified distribution of strength, accompanied by normal *dipoles* with strength related to that by equation (28) which facilitates satisfaction of the boundary condition on the cylinder.

The flow around any one cross-section is then written as a sum of:

(a) the flow due to stokeslets and dipoles within a distance  $\gamma$  of that cross-section; this, since  $\gamma$  is small compared with the radius of curvature of the centre-line, can be approximated as a distribution of singularities along a straight segment, generating (by the cases  $b = c = \gamma$  of equations (29) and (30)) a velocity

$$\frac{\mathbf{f}}{4\pi\mu} \left( \frac{1}{2} + \log \frac{\gamma}{a} \right) \quad \text{or} \quad \frac{\mathbf{f}}{2\pi\mu} \left( -\frac{1}{2} + \log \frac{\gamma}{a} \right) \quad (63)$$

on the cross-section boundary according as the stokeslet strength  $\mathbf{f}$  is normal or tangential to the centreline; and

(b) the flow due to stokeslets at distances greater than  $\gamma$  from that cross-section; since  $\gamma \gg a$  we can neglect any variation of this flow over the cross-section and approximate the velocity all over the cross-section boundary by the velocity generated by those stokeslets at the centre of the cross-section; for the same reason the field of the associated doublets can be neglected. The portion (b) takes the form of an integral, from 0 to  $s - \gamma$  and from  $s + \gamma$  to  $L$ , of the fields of the stokeslets in those intervals calculated at a point on the centreline whose distance measured along it from the head is  $s$ .

Note that the sum of (a) and (b) is independent of the exact value of  $\gamma$  chosen: for example, a slight reduction in  $\gamma$  would reduce both logarithms in (63) but make an equal and opposite change in (b) by adding to it the far field of the distribution of stokeslets along two additional short straight segments. Thus it is mathematically a well-posed problem to determine both the normal and tangential components of  $f$ . Indeed by equating the sum of (a) and (b) to the actual velocity of the cross-section (which in turn is the sum of (i) and (ii) above) we obtain for  $f$  a Fredholm integral equation of the second kind. When that has been solved to give the distribution of the force per unit length  $f$  between the flagellum and the fluid, the requirement that that distribution of force is in static equilibrium determines both the swimming velocity  $U$  and any angular velocity  $\Omega$ .

There is still a serious lack of carefully checked solutions of Hancock's integral equation. With modern computational aids it should be possible to carry out the solution for interesting waveforms like the saw-tooth form described in § 5 and compare the results for different values of the angle  $\psi$  with those given by resistive-force theory.

## REFERENCES

- BATCHELOR, G. K. 1970 J. Fluid Mech. **44**, 419–440.  
BROKAW, C. J. 1965 J. Exptl. Biol. **43**, 155–169.  
CHWANG, A. T. & WU, T. Y. 1971 Proc. Roy. Soc. B, **178**, 327–346.  
COX, R. G. 1970 J. Fluid Mech. **44**, 791–810.  
GRAY, J. 1968 *Animal Locomotion*, Weidenfeld & Nicolson, London.  
GRAY, J. & HANCOCK, G. J. 1955 J. Exptl. Biol. **32**, 802–814.  
HANCOCK, G. J. 1953 Proc. Roy. Soc. A, **217**, 96–121.  
HOLWILL, M. E. J. 1966 J. Exptl. Biol. **44**, 579–588.

*This page intentionally left blank*

## CHAPTER 4

# Aquatic Animal Propulsion of High Hydro-mechanical Efficiency\*

This paper attempts to emulate the great study by Goldstein (1929) ‘On the vortex wake of a screw propeller’, by looking for a dynamical theory of how another type of propulsion system has evolved towards ever higher performance. An ‘undulatory’ mode of animal propulsion in water is rather common among invertebrates, and this paper offers a preliminary quantitative analysis of how a series of modifications of that basic undulatory mode, found in the vertebrates (and especially in the fishes), tends to improve speed and hydromechanical efficiency.

Posterior lateral compression is the most important of these. It is studied first in ‘pure anguilliform’ (eel-like) motion of fishes whose posterior cross-sections are laterally compressed, although maintaining their depth (while the body tapers) by means of long continuous dorsal and ventral fins all the way to a vertical ‘trailing edge’. Lateral motion of such a cross-section produces a large and immediate exchange of momentum with a considerable ‘virtual mass’ of water near it.

In § 2, ‘elongated-body theory’ (an extended version of inviscid slender-body theory) is developed in detail for pure anguilliform motion and subjected to several careful checks and critical studies. Provided that longitudinal variation of cross-sectional properties is slow on a scale of the cross-sectional depth  $s$  (say, if the wavelength of significant harmonic components of that variation exceeds  $5s$ ), the basic approach is applicable and lateral water momentum per unit length is closely proportional to the square of the local cross-section depth.

The vertical trailing edge can be thought of as acting with a lateral force on the wake through lateral water momentum shed as the fish moves on. The fish’s mean rate of working is the mean product of this lateral force with the lateral component of trailing-edge movement, and is enhanced by the virtual-mass effect, which makes for good correlation between lateral movement and local water momentum. The mean rate of shedding of energy of lateral water motions into the vortex wake represents the wasted element in this mean rate of working, and it is from the difference of these two rates that thrust and efficiency can best be calculated.

Section 3, still from the standpoint of inviscid theory, studies the effect of any development of discrete dorsal and ventral fins, through calculations on vortex sheets shed by fins. A multiplicity of discrete dorsal (or ventral) fins might be

---

\* Reprinted from J. Fluid Mech., 44 (1970), pp. 265–301 in the Journal’s ‘Goldstein number’.

thought to destroy the slow variation of cross-sectional properties on which elongated-body theory depends, but the vortex sheets filling the gaps between them are shown to maintain continuity rather effectively, avoiding thrust reduction and permitting a slight decrease in drag.

Further advantage may accrue from a modification of such a system in which (while essentially anguilliform movement is retained) the anterior dorsal and ventral fins become the only prominent ones. Vortex sheets in the gaps between them and the caudal fin may largely be reabsorbed into the caudal-fin boundary layer, without any significant increase in wasted wake energy. The mean rate of working can be improved, however, because the trailing edges of the dorsal and ventral fins do work that is not cancelled at the caudal fin's leading edge, as phase shifts destroy the correlation of that edge's lateral movement with the vortex-sheet momentum reabsorbed there.

Tentative improvements to elongated-body theory through taking into account lateral forces of viscous origin are made in § 4. These add to both the momentum and energy of the water's lateral motions, but may reduce the efficiency of anguilliform motion because the extra momentum at the trailing edge, resulting from forces exerted by anterior sections, is badly correlated with that edge's lateral movements. Adoption of the 'carangiform' mode, in which the amplitude of the basic undulation grows steeply from almost zero over the first half or even two-thirds of a fish's length to a large value at the caudal fin, avoids this difficulty.

Any movement which a fish attempts to make, however, is liable to be accompanied by 'recoil', that is, by extra movements of pure translation and rotation required for overall conservation of momentum and angular momentum. These recoil movements, a potentially serious source of thrust and efficiency loss in carangiform motion, are calculated in § 4, which shows how they are minimized with the right distribution of total inertia (the sum of fish mass and the water's virtual mass). It seems to be no coincidence that carangiform motion goes always with a long anterior region of high depth (possessing a substantial moment of total inertia) and a region of greatly reduced depth just before the caudal fin.

The theory suggests (§ 5) that reduction of caudal-fin area in relation to depth by development of a caudal fin into a herring-like 'pair of highly sweptback wings' should reduce drag without significant loss of thrust. The same effect can be expected (although elongated-body theory ceases to be applicable) from widening of the wing pair (sweepback reduction). That line of development of the carangiform mode in many of the Percomorphi leads towards the lunate tail, a culminating point in the enhancement of speed and propulsive efficiency which has been reached also along some quite different lines of evolution.

A beginning in the analysis of its advantages is made here using a 'two-dimensional' linearized theory. Movements of any horizontal section of caudal fin, with yaw angle fluctuating in phase with its velocity of lateral translation, are studied for different positions of the yawing axis. The wasted energy in the wake has a sharp minimum when that axis is at the 'three-quarter-chord point', but rate of working increases somewhat for axis positions distal to that. Something like an optimum regarding efficiency, thrust and the proportion of thrust de-

rived from suction at the section's rounded leading edge is found when the yawing axis is along the trailing edge.

This leads on the present over-simplified theory to the suggestion that a hydro-mechanically advantageous configuration has the leading edge bowed forward but the trailing edge straight. Finally, there is a brief discussion of possible future work, taking three-dimensional and non-linear effects into account, that might throw light on the commonness of a trailing edge that is itself slightly bowed forward among the fastest marine animals.

## 1. Introduction

Lighthill (1969) gave a general survey of the hydromechanics of aquatic animal propulsion, composed with the help of zoological colleagues to be as far as possible equally restrained in its presuppositions of hydromechanical and of zoological knowledge in the reader. An important concept in that survey was the hydromechanical efficiency  $\eta$  of an animal's propulsive flexural movements. This has a definition similar to that of the Froude efficiency of a propeller; in fact,

$$\eta = U\bar{P}/\bar{E}, \quad (1)$$

where  $U$  is the mean forward velocity,  $\bar{P}$  is the mean thrust required to overcome what viscous drag the animal would sustain for forward velocity  $U$  if it remained rigid and symmetrical, and  $\bar{E}$  is the mean rate at which the flexural movements do work against the surrounding water.

The survey includes a discussion of aquatic propulsion in some twenty classes within the animal kingdom, as well as a more extensive investigation within one class, that of the fishes. The most prominent method of propulsion is the undulatory mode, in which a transverse wave, normally of increasing amplitude, passes backwards along the body from head to tail. This propulsive method, together with various modifications of it, has been successful for motion at both high and low values of the Reynolds number,  $R$ , based on the animal's length and forward velocity. Alternative methods, such as ciliary propulsion at low  $R$ , and propulsion by jet reaction at high  $R$ , have been found more limited in scope and application.

From another point of view, the survey suggests a division of modes of aquatic propulsion into two categories: high-efficiency and low-efficiency, with the separation between them occurring at around  $\eta = 0.5$ . Essentially, this is a division within the undulatory mode (together with its modifications), because propulsive modes limited to low  $R$  like ciliary propulsion seem to be necessarily of low  $\eta$  on such a criterion, while propulsion by the reaction of a jet of velocity  $U_J$  has a low Froude efficiency if  $U_J$  is a substantial multiple of  $U$ , as seems generally to be found in aquatic animals.

Even within the field of undulatory propulsion at relatively high  $R$ , a distinction between 'good hydromechanical shapes' with high  $\eta$  and 'bad' ones with low  $\eta$  was already made in the survey. Typical 'bad hydromechanical shapes' for undulatory propulsion at high  $R$  include those with roughly circular cross-sections (like most terrestrial snakes) while probably for low to moderate  $R$  (say  $< 10^3$ ) all shapes are in this sense 'bad'.

Some excellent theoretical studies of the undulatory mode of propulsion in bad hydromechanical shapes have been made, notably by Taylor (1952) and by Gray & Hancock (1955). Lighthill (1969, §13) summarizes these in an approximate, simplified form. Their essential assumption is that the instantaneous force between the water and a section of the animal's body is the same as when that section moves steadily through the water at the same relative vector velocity. This assumption is particularly good at low  $R$ , when viscous forces dominate over inertia.

It is often assumed that this 'quasi-static' theory of undulatory propulsion is equally good at high  $R$ . This, however, is not necessarily the case, since it neglects those additional forces required in unsteady motion to accelerate water close to the body section, the mass of water which effectively must be accelerated being that usually described as the 'virtual mass'. Actually, the 'good hydromechanical shapes' seem to be precisely those which are able to improve greatly their propulsive efficiency  $\eta$  by utilizing this virtual-mass effect.

For 'elongated' animals (those whose length very greatly exceeds their other dimensions) Lighthill (1969), using theoretical ideas developed by Lighthill (1960), showed that it is the animals (almost exclusively vertebrates) that have developed transverse compression at the posterior end which can use the virtual-mass effect to improve propulsive efficiency. 'Compression' here means the flattening of the cross-section into a posterior edge or 'trailing edge', and 'transverse compression' signifies compression occurring in that transverse direction in which are made the undulatory displacements that are propagated backwards as the propulsive wave. It is of course lateral (side-to-side) undulations that typical aquatic vertebrates, including most fishes, amphibia and reptiles, use for propulsion, and the characteristic lateral compression at the posterior end of many of these improves hydromechanical efficiency, but, where undulations are dorso-ventral (that is, vertical) as in cetacean mammals, dorso-ventral posterior compression is required.

Posterior transverse compression is important because it permits the virtual mass for the transverse motions of the body section in the water to have a substantial value even at the posterior end, without that very large hydromechanical resistance that would be associated with any blunt backward-facing base. The propulsive benefits of such a substantial trailing-edge virtual mass are given a rather simple mechanical interpretation in §4 of Lighthill (1969), but the point is so important that it must be further developed in the present paper. In particular, those other conditions besides transverse compression which must be satisfied, essentially by the form of the undulatory mode itself in various types of animal, require further study.

Characteristic propulsive modes in the great majority of fishes (for some of the principal exceptions see Lighthill (1969, §§10 and 11)) are broadly divided into two classes: anguilliform and carangiform. Fishes using the anguilliform mode have the whole body flexible, and the propulsive wave travelling from head to tail has an amplitude which, although increasing posteriorly, is significant all along the fish's length. Carangiform propulsion, by contrast, means propulsion in which the amplitude of undulation becomes significant only in the posterior

half, or even one-third, of the length of the fish; the remainder of the fish's body is relatively inflexible.

The mechanics of anguilliform propulsion is, perhaps, simplest in those eels (including *Anguilla* from which the mode's name is derived) and other animals that possess long continuous dorsal and ventral fins whose cross-sectional depth (taking body and both fins into account) either increases or remains substantially constant all the way to the (vertical) trailing edge. It is with this group of animals that theoretical study of anguilliform propulsion begins in §2 below. As elsewhere in this paper, the exposition is confined to cases of only moderate amplitude of undulation, permitting the simplification of Lighthill (1960, 1969) which considers the only significant relative motions of different parts of the fish to be in the lateral direction. Not only the mean forces produced by the undulations but also fluctuations about those means are estimated, and related to the lateral water motions produced adjacent to the body and in the vortex wake.

Anguilliform propulsion is additionally found, however, in several groups of fishes, such as catfishes (see Lighthill (1969) for a fuller enumeration), which possess discrete dorsal and ventral fins behind which the total cross-sectional depth falls to considerably reduced values before increasing again at the tail to a caudal-fin trailing edge of, approximately, the same depth again. In these fishes there are vortex sheets shed already behind the dorsal and ventral fins, and the forces referred to above are modified in an interesting way (investigated in §3) by the free development of those vortex sheets. Mean propulsive force and its efficiency can, it is shown, both be increased under such circumstances, even though the vorticity becomes 'bound' once more on to the caudal fin before it is discarded into the wake.

The same elongated-body theory that is suitable for studying anguilliform motion is useful also for studying the fundamental modification of that motion (with the undulations confined to a tail region) that we call carangiform. This propulsive mode is, in the main, investigated in §4, which demonstrates why the modification is advantageous from the point of view of propulsive efficiency.†

This advantage, however, is not reaped unless there is a very large reduction of depth (that is, a 'necking' in lateral view) in the region, just anterior to the caudal fin, where the undulation amplitude has a steeply increasing gradient. Farther forward still the total depth of dorsal fin, body and ventral fin can, and indeed should, be as large as that of the trailing edge, with comparable depths maintained for a considerable length of fish body. Complications such as those mentioned above, with a vortex sheet shed from the dorsal and ventral fins, are absent however because those fins do not significantly take part in carangiform undulation.

The calculations in §§2–4 are limited to a fish pursuing a straight path to which no part of it is inclined at more than a moderate angle (say, less than 30°), so that to good approximation the important relative motions of different parts of the fish are in the lateral direction. One advantage of setting out calculations on this basis in careful detail is that they can be conducted either in the form

† The exception is a particular variant of the carangiform mode, to which elongated-body theory is inapplicable, and which is treated in §5.

of a rigorous, but complicated, perturbation expansion in which the forces acting between the body and the water are obtained by integrating surface pressure distributions, as in the appendix to Lighthill (1960), or as here by some rather more tentative, but simpler, considerations of bulk momentum and energy, and the two methods give identical results, which adds to one's confidence in both.

A less restrictive analysis is desirable, however, partly because amplitudes used by fishes swimming at their top speed are in reality considerably greater, with tails inclined by as much as  $60^\circ$  to the direction of motion. Furthermore, it is of great interest to be able to analyse turning manœuvres. With these objects in mind, a new version of elongated-body theory is being prepared for future publication, based on similar considerations of bulk momentum and energy to those whose reliability is verified here, but using far less restrictive geometrical assumptions.

The crude idea expounded so far, that for the 'bad hydromechanical shapes' quasi-static values of resistance to the motion of body sections may properly be used, as in Taylor (1952) and Gray & Hancock (1955), while for 'good hydro-mechanical shapes' the virtual-mass effects of elongated-body theory dominate and resistance forces are negligible by comparison, is obviously exaggerated. It would be most desirable to work out a theory combining the two effects. Only a beginning is here made on this problem; see especially § 4, where an idea of Lighthill (1960) for effecting such a combination is pursued somewhat further.

The last section of this paper is devoted to investigations arising from the principal conclusion which Lighthill (1969) draws (see especially his figure 6): namely that all the fastest marine animals, and in particular a certain group of bony fishes that includes the tunnyfishes, a certain group of unusually fast sharks, and also most of the cetacean mammals, have adopted an essentially carangiform mode of propulsion but with tails greatly modified along essentially identical lines. These are high-aspect-ratio tails of crescent-moon shape, and such a tail is often called 'lunate' whether it is a vertical fish caudal fin or a horizontal cetacean tail-fluke. Lighthill gives reasons for supposing that these fast animals from very different lines of evolution have 'converged' upon the lunate tail (see also Kramer 1960, figure 18) primarily because it possesses hydromechanical advantage, and indicates ways in which its hydromechanics may be analysed.

It is impossible to use elongated-body theory to study lunate-tail hydro-mechanics, because elongated-body theory assumes that the water is set into motion by body actions that, essentially, are distributed along the direction of motion, so that each vertical slice of water perpendicular to that direction is influenced primarily by body actions close to the slice. By contrast, the action of the propulsive lunate tail is spread out at right angles to the direction of motion; vertically in the case of the fishes. The large vertical extent of the fish's tail makes the assumptions of elongated-body theory untenable, because changes in the body action on neighbouring water slices perpendicular to the direction of motion are altogether too abrupt for their influences to be regarded as acting independently.

On the other hand, the action of the lunate vertical fin's lateral motion on different *horizontal* slices of water is far more gradually varying, and can more

reasonably be regarded as mutually independent to a first approximation. This suggests treating the lunate tail by the two-dimensional theory of oscillating aerofoils. Lighthill (1969, §§8 and 9) proposes as a first approximation this method, taking into account only the cross-stream (in the case of fishes, vertical) components of wake vorticity, although he suggests as a more accurate (but more difficult) method of calculation a 'lifting-line' theory that would also take the streamwise components into account. The latter theory would give a lower, more accurate, estimate of efficiency through not ignoring parts of the wake energy associated with the streamwise vorticity.

It seems clear that, if indeed two-dimensional oscillating-aerofoil theory is of any value as a first approximation in fish hydrodynamics, then it is in this application to lunate-tail propulsion. By contrast, any attempt to study propulsion in the more normal elongated fishes by investigating an 'analogous two-dimensional fish' cannot give answers that are even approximately reliable because the basic assumption of two-dimensional theory, that body actions are spread out at right angles to the direction of motion, is the complete opposite of the truth for an elongated fish. Accordingly, the values of an aerofoil frequency parameter  $\omega c/U$  (based on radian frequency  $\omega$  and forward speed  $U$ ) that are relevant to fish propulsion must involve a chord  $c$  equal not to the fish length  $l$  but to the lunate tail's dimension in the direction of motion. Such values of  $\omega c/U$  are typically less than 1, even though  $\omega l/U$  is normally around 10.

A general characteristic of the carangiform mode (Lighthill 1969, §5) is that the oscillations of the caudal fin's angle of attack and lateral velocity are in phase with one another. Accordingly, Lighthill made lunate-tail calculations, of which he gave the results in his figure 9, based on assuming that the postulated two-dimensional aerofoil sideslips with oscillating velocity  $W \cos \omega t$  and yaws with the in-phase oscillating angle of attack  $(\theta W/U) \cos \omega t$ . In these calculations the axis of the yawing oscillation was taken for simplicity as the central (or 'half-chord') axis of the aerofoil; the parameter  $\theta$  was described as a 'proportional feathering'. There is no thrust when  $\theta = 1$ , but although thrust increases as  $\theta$  falls away from 1 the fin generates what thrust there is with greatest efficiency if  $\theta$  is not too far below 1. When  $\theta$  falls to zero, efficiency is reduced (even on this two-dimensional theory, neglecting streamwise vorticity, it is down to 0.65 for  $\omega c/U = 1$ ) but thrust itself is then at a maximum.

This case  $\theta = 0$  is one of pure sideslip without yawing, so that none of the thrust can come from any difference of pressure between the two sides of the fin. Under these circumstances (when actually thrust is greatest though it is not generated with great efficiency), the whole thrust comes from leading-edge suction; that is, from the action on the rounded leading edge of the reduced pressure in the water swirling round it. When  $\theta$  is non-zero, not all but a large part of the thrust comes from leading-edge suction, and to realize it a bluntly rounded leading edge, such as is indeed found in all the animals with lunate tails, is necessary.

In §5 below, the calculation just described is repeated with a more general position of the yawing axis. The dependence of efficiency on this position is predicted as unexpectedly sensitive. For given amplitudes the wasted wake energy

has a minimum when the axis is at the three-quarter-chord point (that is, three times as far from the leading edge as from the trailing edge). Rate of working increases somewhat for axis positions distal to that point, and an optimum for both thrust and efficiency places the axis very close to the trailing edge.

This new result seems to have definite relevance to the question of why the lunate tail should be hydromechanically efficient. If a caudal fin were yawing as a whole about a single axis, and if two-dimensional theory could be applied to each section, we could infer that good thrust with good efficiency would best be obtained if the trailing edge ran straight along the axis of yaw. For a tapered fin shape this would mean that the trailing edge was straight although the leading edge was considerably bowed forward. This degree of departure from 'straight wing' conditions (when the leading edge would be bowed forward but the trailing edge would be bowed backward to an equal extent) is in the right direction, although it does not go far enough; possibly a fully three-dimensional theory might explain the fact that the trailing edge is even slightly bowed forward.

The most promising type of three-dimensional theory may be a sort of lifting-line theory in which the local flow around each cross-section is taken as a two-dimensional flow but with a local angle of attack influenced by the whole pattern of wake vorticity. The form of the two-dimensional theory as explained in §6 has been selected to facilitate its possible use in three-dimensional calculations of this kind.

In the meantime, the general conclusion concerning the trailing edge may, perhaps, find application to other aspects of animal locomotion. For example, if the flapping of a bird's wings were approximated on a two-dimensional theory as a combination of a pitching oscillation (superimposed, as in figure 7 of Lighthill (1969), on the uniform angle of attack required for weight support) and a heaving motion whose velocity oscillates in phase with the pitch angle, then the predicted propulsive efficiency would be greatest if the pitch axis were along the trailing edge.

It will be seen that §§2 to 5 of this paper form a sort of mathematical appendix to the generally non-mathematical discussion of Lighthill (1969) and of this introduction.

## 2. The pure anguilliform mode of propulsion

Fishes may be said to adopt a 'pure' anguilliform mode of propulsion when, like *Anguilla vulgaris*, they possess long continuous dorsal and ventral fins whose cross-sectional depth (taking body and both fins into account) is maintained, or even increases, all the way to the posterior end; there, such taper as may appear in lateral view is so abrupt that to a close approximation the fish body may be regarded as terminating in a vertical trailing edge. Such a fish, at its lower speeds of movement, often propels itself by undulations that pass only along the fins, but at its higher speeds makes a close approximation to the pure anguilliform mode, in which a single lateral undulation, involving the body as well as the fins, passes backwards from the head to the tail. This mode is analysed by

'elongated-body theory' (an adaptation of what is called 'slender-body' theory in aeronautics) in the present section.

We first set it out, as Lighthill (1960) did, in a frame of reference moving with the mean speed  $U$  of the fish, so that in this frame of reference the fish is making undulatory movements in order to remain in the same average position in a stream whose undisturbed velocity is  $U$ , directed along the  $x$ -axis. We describe the fish's undulations by reference to a 'stretched straight position' in which the body is perfectly symmetrical about the vertical plane  $z = 0$ . To hold the fish in the stream in the stretched straight position would require a force equal to the viscous drag  $D$  (whose magnitude we can infer from observations of deceleration when a fish of known mass glides rigidly). We investigate how, without any external force, the fish may perform undulations about the stretched straight position which maintain its mean position in the stream, so that they may be said to generate a net thrust  $P$  that exactly balances the drag  $D$ .

In the stretched straight position, the cross-section of the fish at a distance  $x$  from the front end is called  $S_x$  by Lighthill (1960), who further supposes that, during undulations, the cross-section is displaced by an amount  $h(x, t)$  in the  $z$ -direction (in other words laterally). Lighthill (1969) used the notation  $W$  for the lateral velocity  $\partial h / \partial t$  of a cross-section. We follow him in this, as also in the use of

$$w = \partial h / \partial t + U \partial h / \partial x, \quad (2)$$

rather than the  $V$  of Lighthill (1960), to signify the velocity of lateral pushing of a vertical water slice (perpendicular to the  $x$ -axis) by the successive cross-sections past which it sweeps with velocity  $U$ .

Lighthill (1969, §4) explained in as simple terms as possible why  $w$  is considerably smaller than  $W$  at the trailing edge if in its neighbourhood the undulation has developed into a travelling wave of velocity  $V$  only moderately greater than  $U$  and of constant amplitude. In the language of partial derivatives this is because  $\partial h / \partial t + V \partial h / \partial x$  is then zero, and so equation (2) gives

$$w = W(V - U)/V. \quad (3)$$

It was noted that  $w/W$  should be relatively small for good efficiency, but must not be too small if thrust is to be adequate to overcome the viscous drag  $D$ .

Associated with the lateral pushing of a water slice at velocity  $w$  is a certain lateral momentum in the water slice. When such pushing begins, this momentum rises *immediately* to a value  $mw$  per unit length (more strictly, that momentum is communicated to the water by a signal propagated at the speed of sound, which may however be taken infinite in the present context), where  $m$  is the virtual mass per unit length of fish. Continued pushing will produce a certain amount of increase above this value, through the action of viscous forces, especially in the production of streamwise vorticity shed in the cross-stream flow, and this is investigated in §4, but here the momentum present without any such augmentation is first studied.

Elongated-body theory as developed by Lighthill (1960), assuming that rates of change of cross-sectional dimensions along the fish's length are small, and that undulatory movements do not much distort the streamwise distribution of fish

mass and are not too abruptly varying, arrives at the conclusion that for a particular cross-section  $S_x$  we may estimate  $m$  as the virtual mass per unit length (here written  $\rho A$ ) for an infinite rigid cylinder  $C_x$  with the same cross-section. This quantity is easily obtained from the complex-variable theory of two-dimensional irrotational flow for a variety of cross-sectional shapes relevant to different groups of fish.

An interesting conclusion from such calculations is that  $m$  depends to a rough approximation only on the depth  $s$  of the cross-section. Indeed, if we write

$$m = \frac{1}{4}\pi s^2 \rho \beta, \quad (4)$$

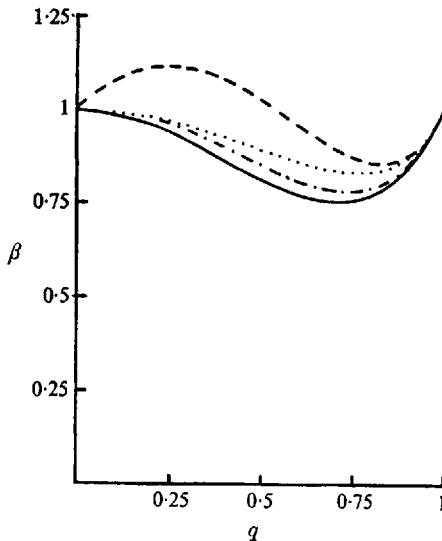


FIGURE 1. Virtual-mass coefficients  $\beta$ , defined in equation (4), for cross-sections such that a fraction  $q$  of the total depth  $s$  is occupied by fish body and the remainder by dorsal and/or ventral fins. ...., elliptical body cross-section (axis-ratio 2) with equal dorsal and ventral fins. Curves for circular body cross-section: —, equal fins; - - -, fin depths in ratio 3:1; - - - - , one fin only.

where  $\rho$  is the water density, then the non-dimensional parameter  $\beta$  varies little from 1. It is exactly 1 for an elliptic cross-section of any eccentricity, ranging from the nearly circular section of the anterior portions of many eels, through the highly eccentric elliptical section of for example a herring, to the quite flat section characteristic of a trailing edge. Figure 1 shows the extent to which  $\beta$  departs from 1 for some other typical cross-sections.

These include cross-sections in which a fraction  $q$  of the total depth  $s$  is occupied by a body whose section is a circle, or an ellipse with major axis vertical, and the remainder by two vertical fins (dorsal and ventral) of negligible thickness. When the fins are of equal depth (that is,  $\frac{1}{2}(1-q)s$  each) the plain line shows that  $\beta$  is 1 for  $q = 0$  or 1 but falls to a minimum of 0.75 for  $q = 0.7$  in the case of a circular body, while departing much less from 1 when the body is elliptical (dotted line). Results (chain-dotted line) for fins of unequal depths  $\frac{3}{4}(1-q)s$  and  $\frac{1}{4}(1-q)s$ , and also (broken line) for a single fin of depth  $(1-q)s$ , show that even for a

circular body the departure of  $\beta$  from 1 is reduced in these cases, while corresponding curves for an elliptical body (not shown) involve, as would be expected, still smaller departures from 1.

The importance of confidently knowing the virtual mass per unit length  $m$  makes it desirable, however, to probe further the inference from elongated-body theory that, at the section  $S_x$ , it is the same as for an infinite rigid cylinder  $C_x$  with the same cross-section. How far does this need correction due to two facts: (i) that the fish body is flexible so that different cross-sections push the water with different velocities, and (ii) that its cross-section is non-uniform? Fairly simple answers to these questions can be obtained in the special case when cross-sections are circular; these answers indicate that corrections are then not large provided that significant harmonic components of cross-section shape or velocity along the fish have wavelengths  $\lambda$  of at least  $5s$ .

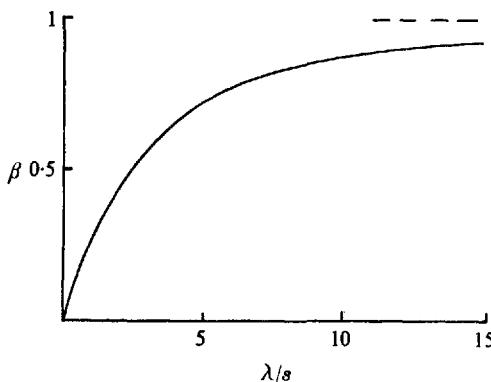


FIGURE 2. Virtual-mass coefficient  $\beta$  as a function of the ratio  $\lambda/s$  of wavelength to depth of fish cross-section (assumed circular).

Such answers are obtained from solutions of Laplace's equation in cylindrical polar co-ordinates  $(x, r, \theta)$  proportional to  $K_1(kr) \cos \theta \cos kx$ , where  $k = 2\pi/\lambda$ . If, for example, a cylinder of uniform circular cross-section pushes the water in the direction of the axis  $\theta = 0$  (corresponding to the  $z$ -axis in Cartesian co-ordinates) with a velocity varying like  $\cos kx$ , these solutions allow us to calculate the virtual mass per unit length in the form (4) with

$$\beta = K_1(\pi s/\lambda)/[-(\pi s/\lambda) K'_1(\pi s/\lambda)]. \quad (5)$$

Figure 2 shows that this  $\beta$  is within 25 % of 1 for  $\lambda > 5s$ . Similar calculations on the effect of a particular harmonic component of variation of the cross-section depth, under conditions of uniform pushing, show that the greatest departure of  $\beta$  from 1 is then the same as equation (5) would give but further reduced by the ratio of the amplitude to  $\frac{1}{2}s$ , a ratio which inherently must be less than 1.

These indications for circular cross-sections  $S_x$ , that the virtual mass is to a rough approximation the same as for a rigid cylinder  $C_x$  with the same cross-section if wavelengths of significant variations in the streamwise direction exceed  $5s$ , can probably be relied on also for various more flattened (that is, laterally

compressed) types of cross-section. These, indeed, should be *more* able to 'get a grip on' the water near them in a way such that the lateral water motion simply reflects that of the local body cross-section. Therefore, since in pure anguilliform motions a typical wavelength of undulation (see, for example, figure 4 of Lighthill 1969) is considerably more than 5s, it is reasonable to treat them all by elongated-body theory, founded on the above approximations to virtual mass.

The z-component of the force with which the fish acts on a water slice, per unit distance in the x-direction, must equal the rate of change of z-component of momentum of the water slice as it sweeps past the undulating fish at velocity  $U$ . This is

$$Z = \left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) [m(x) w(x, t)], \quad (6)$$

where  $m(x)$  is given by equation (4) essentially in terms of the depth  $s(x)$  of the cross-section  $S_x$  (since  $\beta$  is close to 1) and  $w(x, t)$  by equation (2). In (6), departures of the x-component of water-slice velocity from  $U$  can reasonably be neglected since they are of the second order in the undulations (being of the order of products of transverse velocities  $w$  with local angles of body twist out of the stretched straight position).

The rate,  $E$ , at which the fish is doing work through its transverse movements with velocity  $\partial h/\partial t$ , exerting forces  $Z$  per unit length, is

$$E = \int_0^l Z \frac{\partial h}{\partial t} dx, \quad (7)$$

and by equation (6) for  $Z$  and equation (2) for  $w$  this can be written

$$\begin{aligned} E &= \int_0^l \left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \left( mw \frac{\partial h}{\partial t} \right) dx - \int_0^l mw \frac{\partial w}{\partial t} dx \\ &= \frac{\partial}{\partial t} \int_0^l \left( mw \frac{\partial h}{\partial t} - \frac{1}{2} mw^2 \right) dx + U \left[ mw \frac{\partial h}{\partial t} \right]_{x=l}. \end{aligned} \quad (8)$$

The integrated term in (8) (last term on the right-hand side) involves a contribution only from the posterior end  $x = l$ , where the lateral compression makes  $m$  non-zero; by contrast,  $m$  becomes zero at the anterior end  $x = 0$ .

This last term in (8) is particularly important because only through it can the fish exert any mean power. The first term, by contrast, is merely the time derivative of a fluctuating quantity so that its time mean is zero.

Lighthill (1969) gives a simple mechanical interpretation of this important last term, which is  $UmwW$  in the notation he uses. Here,  $mw$  is the lateral momentum per unit length in the water near the trailing edge, and so  $Umw$  is the rate of shedding of that momentum into the vortex wake per unit time, which, effectively, is a lateral force with which the trailing edge acts on the wake, doing work at a rate  $UmwW$  because its lateral velocity is  $W$ . Equation (3) shows this rate of working to be always positive and to have a mean value simply proportional to the mean square of the trailing-edge lateral velocity  $W$ .

We need to ask, however, why this simple mechanical argument yields correctly only the mean value of the rate of working, and not the instantaneous value given by (8), including the first term on the right-hand side. The reason is

that the undulating fish does not do work only on the wake. Part of its rate of working involves exchange of energy with water that has not yet reached the wake. The total energy of such water ahead of the trailing edge fluctuates up and down, so that its rate of change is not instantaneously zero, although it has a zero mean rate of change over a long time (or over a single period of oscillation).

This interpretation of equation (8) can be verified if it is recast, using (2), as

$$E - U[mwW]_{x=l} = \frac{\partial}{\partial t} \int_0^l \left( \frac{1}{2} mw^2 - Umw \frac{\partial h}{\partial x} \right) dx. \quad (9)$$

This equates the rate of working by the fish, minus the rate at which it does work on the wake, to the rate of change of an energy integral. In this integral the first term represents the energy of water motions in the  $y$ - and  $z$ -directions. In the frame of reference used, however, there are also comparable fluctuations in the energy of water motions in the  $x$ -direction.

Specifically, when unit mass of water makes a small change in its velocity from  $U$  to  $U+u$ , there is an energy change which to a first approximation is  $Uu$ . Therefore, when a large body of water moving with velocity  $U$  suffers a non-uniform small perturbation of velocity, the energy change is  $U$  times the change in the  $x$ -component of momentum. Where the water momentum per unit length due to the fish's transverse motion is  $mw$ , but not exactly in the  $z$ -direction because the fish body is sloped at an angle  $\partial h/\partial x$  to the stretched straight position, there will to first approximation be an  $x$ -component of momentum  $-mw\partial h/\partial x$  (again per unit length), with a corresponding energy change  $-Umw\partial h/\partial x$  which explains the second term inside the integral in (9).

Here we have used an important extension of the basic idea that water momentum, due to relative motion in the  $z$ -direction with velocity  $w$  of a section  $S_x$  of fish, is that of a cylinder  $C_x$  with the same cross-section. The extended idea is that if  $S_x$  is sloped at a non-zero angle to its stretched straight position then the corresponding cylinder  $C_x$  must be thought of as so sloped, with the consequence that the associated momentum, necessarily at right angles to the generators of the cylinder, is sloped back from the  $z$ -direction. This extension is fully supported by the arguments in favour of the virtual-mass concept given earlier.

The rate of working  $E$  can also be expressed in terms of the thrust  $P$  (whose value can therefore be obtained by comparing the two expressions), if we use a different frame of reference; one in which the water far from the fish is at rest. We can then write

$$E = UP + U[\frac{1}{2}mw^2]_{x=l} + \frac{\partial}{\partial t} \int_0^l (\frac{1}{2}mw^2) dx, \quad (10)$$

where the first term on the right is the work done by the fish in simply moving at velocity  $U$  in the direction of the thrust  $P$ . The second term represents the rate at which kinetic energy of water movements, namely  $\frac{1}{2}mw^2$  per unit length of fish, is shed to the wake at the trailing edge. This 'wasted' energy was emphasized by Lighthill (1969, §4). The third term in (10) represents the rate of change of a fluctuating quantity, namely the total kinetic energy (in the new frame of reference) ahead of the trailing edge.

From (9) and (10), by subtraction, we obtain the thrust

$$P = [mw(W - \frac{1}{2}w)]_{x=l} - \frac{\partial}{\partial t} \int_0^l \left( mw \frac{\partial h}{\partial x} \right) dx. \quad (11)$$

Equations (9) and (10) can also be used directly to estimate the hydromechanical efficiency  $\eta$ , given by (1) as  $U\bar{P}/\bar{E}$ . Using the fact that the mean value of the time derivative of each integral is zero, we obtain

$$\eta = 1 - (\bar{E} - U\bar{P})/\bar{E} = 1 - \frac{1}{2}[w^2]_{x=l}/[wW]_{x=l}. \quad (12)$$

Provided that  $w$  and  $W$  at  $x = l$  are related by equation (3), this implies that  $\eta = 1 - \frac{1}{2}(V - U)/V$ . This reinforces the argument of Lighthill (1969) that good efficiency requires  $w$  to be considerably smaller than  $W$  (and hence  $V$  not too much greater than  $U$ ), although for substantial thrust, as (11) shows,  $w$  must certainly not be too small.

Equation (11), however, gives not only the mean, but also the instantaneous value of the thrust  $P$ . This value is obtained more easily by the above subtraction method than by any direct physical argument. The latter is possible, however, and is based on the following modified form of (11):

$$P + [Umw\partial h/\partial x - \frac{1}{2}mw^2]_{x=l} = \frac{\partial}{\partial t} \int_0^l (-mw\partial h/\partial x) dx, \quad (13)$$

a form which, we may note, agrees with that derived in a completely independent way (through integrating the distribution of surface pressures obtained by a careful perturbation expansion) in the appendix (see equations (A 21), (A 26) and (A 27)) to Lighthill (1960).

To interpret (13) physically, we note that the right-hand side is the rate of change of the  $x$ -component of water momentum in the region anterior to a plane  $\Pi$  through the trailing edge perpendicular to the  $x$ -direction. This momentum component changes partly due to the action of the force  $P$  between the fish surface and the water, and partly due to momentum transport across  $\Pi$ , represented by the expression in square brackets. In this expression, loss by convection at velocity  $U$  of momentum  $-mw\partial h/\partial x$  per unit length appears first. The second term represents a loss due to the resultant pressure force acting over the plane  $\Pi$ .

A direct physical argument to derive  $P$  is made more difficult by the fact that momentum transfer occurs additionally through the agency of this non-vanishing pressure force. That its value is  $\frac{1}{2}mw^2$  follows most easily from the unsteady form of Bernoulli's equation in a frame of reference in which the fluid velocity at large distances is  $-w$  in the  $z$ -direction; in this frame of reference the fish motion relative to the water is zero at the trailing edge. Then the quadratic terms in Bernoulli's equation give pressures

$$\frac{1}{2}\rho[w^2 - (\partial\phi/\partial y)^2 - (\partial\phi/\partial z)^2] = \rho w(\partial\phi/\partial z + w) - \frac{1}{2}\rho[(\partial\phi/\partial y)^2 + (\partial\phi/\partial z + w)^2], \quad (14)$$

whose resultant can be expressed in terms of the total momentum  $mw$  and kinetic energy  $\frac{1}{2}mw^2$  per unit length in the field  $(\partial\phi/\partial y, \partial\phi/\partial z + w)$  as

$$w(mw) - \frac{1}{2}mw^2 = \frac{1}{2}mw^2. \quad (15)$$

(On the other hand the  $-\rho \partial\phi/\partial t$  term in the pressure gives no resultant because for a section symmetrical about  $z = 0$  the potential  $\phi$  is an odd function of  $z$ .)

The presence of such a resultant pressure force  $\frac{1}{2}mw^2$  across the plane  $\Pi$  is inferred again by vorticity considerations in § 3, through consideration of how the  $x$ -component of momentum in the wake is changing otherwise than by the shedding of momentum from the trailing edge. Other refinements of the theory of this section, specially designed to throw light on evolutionary deviations from pure anguilliform motion, are given in both the subsequent sections.

### 3. Vortex sheets shed by fins

In this section, we precede a discussion of vortex shedding by fins in general with a discussion of the form of the water flow just behind the trailing edge in pure anguilliform motion. This is the flow carrying those quantities of momentum and kinetic energy predicted in § 2 as shed into the wake. It will be evident to certain readers, especially those with aeronautical experience, that this must take the form of a vortex sheet, and specifically (since we are dealing with the low-aspect-ratio limit) one with a distribution of vorticity corresponding to an 'elliptic distribution' of velocity potential (minimizing, per unit length of wake, the kinetic energy for given lateral momentum and span).

To see how this follows from the theory of § 2, we note that when the cross-section  $S_x$  is located at a vertical trailing edge the corresponding cylinder  $C_x$  is an infinitely long, infinitesimally thin strip of depth  $s$ . The velocity potential  $\phi$  for motion of such a strip with velocity  $w$  in the  $z$ -direction takes values

$$\phi = \mp w[\frac{1}{4}s^2 - (y - y_0)^2]^{\frac{1}{2}} \quad (16)$$

on the sides of strip  $|y - y_0| < \frac{1}{2}s$ ,  $z \geq 0$  respectively. Equation (16) exhibits already the elliptic distribution of trailing-edge velocity potential. It means that the vertical velocity component  $\partial\phi/\partial y$  changes by an amount

$$\Delta(\partial\phi/\partial y) = 2w(y - y_0)[\frac{1}{4}s^2 - (y - y_0)^2]^{-\frac{1}{2}} = -\Xi, \text{ say,} \quad (17)$$

between the  $z < 0$  and  $z > 0$  sides of the trailing edge.

Such a discontinuity in velocity is equivalent to a vortex sheet, a thin layer in which the  $x$ -component of vorticity  $\xi$  takes large values, whose integral across the layer is  $\Xi$ ; or, since in reality there are boundary layers on both sides of the solid strip, this must actually be the integral of  $\xi$  across the two boundary layers. It follows that a perfectly consistent picture of the flow can be built up in which equation (16) represents the cross-flow around the posterior end of the fish (more strictly, around the boundary layers attached to its surface), whereas behind the trailing edge there is a free vortex sheet with the same strength  $\Xi$ , with the same values (16) of the velocity potential on the two sides of it, and inducing the same cross-flow around it.

This is a consistent model because (i) such continuity at the trailing edge between the upstream and downstream cross-flows means that variation in cross-flow is gradual there as well as around the main part of the body, which justifies (as in § 2) the local use of the two-dimensional form of cross-flow (that around

$C_x$ ); (ii) continuity in the integrated  $x$ -component of vorticity,  $\Xi$ , at the trailing edge is consistent with the solenoidal character of the vorticity field and with the convection of vortex lines with the fluid.

In a frame of reference in which the undisturbed fluid is at rest, the vorticity shed into the wake remains unchanged in magnitude and continues to exhibit the same bodily motion, with velocity  $w$  in the  $z$ -direction, that is implied by the surface values (16) of velocity potential. At each point in the wake the local value of  $w$  (which is linked through equation (17) with the local value of the vortex-sheet strength  $\Xi$ ) is equal to the value that  $w$  had at the trailing edge when the trailing edge passed that point. This implies a certain non-uniformity in  $w$  along the wake, reflecting the variability with time in the trailing-edge value of  $w$ .

The vorticity in the sheet is rotated about the  $y$ -direction by any variation of  $w$  with  $x$ , the angular velocity of rotation of vortex lines being  $-\partial w/\partial x$ . As this happens the direction of the associated momentum,  $mw$  per unit length, also rotates; this momentum is in a direction close to the  $z$ -direction, but the twisting causes  $x$ -momentum to appear in addition, at a rate  $-mw\partial w/\partial x$ . The total rate of change of wake momentum due to this cause is

$$\int_l^\infty (-mw\partial w/\partial x) dx = [\frac{1}{2}mw^2]_{x=l}, \quad (18)$$

since the fluid is undisturbed at infinity. Equation (18) confirms the value obtained by quite different considerations of pressure distribution in §2 (see especially (15)).

There is no special point in pursuing farther the discussion of the wake in the pure anguilliform motion of animals with continuous dorsal and ventral fins. However, considerations analogous to those just described are relevant also to understanding the wakes from discrete dorsal and ventral fins, with more or less unswept trailing edges, carried by animals performing an essentially anguilliform mode of undulation. In the neighbourhood of a cross-section including such a fin, the velocity potential takes different values on the fin's two sides, just as in (16), corresponding to a vortex sheet of strength  $\Xi = -\Delta(\partial\phi/\partial y)$  as in (17). Equations (16) and (17) would only represent their values to good approximation if the body were thin in relation to its depth; corresponding results for rounder bodies can, however, be calculated by complex-variable theory.

Immediately behind the posterior edge of the fin, a vortex sheet of the same strength  $\Xi$  must be found, for exactly the reasons described above. In a frame of reference in which the undisturbed water is at rest, this vortex-sheet strength remains unchanging, and has at any point a value proportional to the value of  $w$  at the fin's posterior edge when the latter passed the point. In the frame of reference of §2, with the undisturbed water moving at velocity  $U$ , the vortex-sheet strength is accordingly a fixed function of  $t - (x/U)$  and  $y$ .

We apply these considerations first to cod, and those other members of the family Gadidae in the order Anacanthini, which both dorsally and ventrally possess a succession of closely spaced fins, and swim with an essentially anguilliform motion. Behind the anterior dorsal fin such a vortex sheet will be found, filling the gap between it and the second dorsal fin, and producing a far greater continuity in the cross-flow than would be expected from the geometry.

At the second dorsal fin, this vortex sheet gets mixed with the boundary layer. Such incident streamwise vorticity should not significantly influence the cross-flow, or the overall vortex-sheet strength over that fin, since those are uniquely determined by the local value of  $w$ . Accordingly the presence of the detached vortex sheet between the two fins is purely a local modification, which influences little the flow behind it. Other gaps between fins are similarly filled in by vortex sheets.

When, as in many of the Gadidae, these gaps are small (for the case of large gaps, see below), the dynamical effect of the vortex sheet is practically the same as if the dorsal fin were continuous. Actually any difference between the cross-flow around a body section  $S_x$  *including* part of the vortex sheet and that which would be present if the vortex sheet were replaced by solid fin can only be due to: (i) differences between the body section shape where the vortex sheet was shed and at  $S_x$ ; or to (ii) differences between the value of  $w$  where the vortex sheet was shed and at  $S_x$ . If both differences (i) and (ii) were absent, then by the condition determining the strength  $\Xi$  of the vortex sheet it is the same vortex sheet as that attached to a solid fin, and hence by the uniqueness theorem of hydrodynamics the flows are the same.

For small gaps both differences (i) and (ii) are small. Note that because the body undulation propagates at a speed  $V$  not very much greater than  $U$ , the value of  $w$  at  $S_x$  has only a slight phase advance over that obtaining where the vortex sheet was shed, although there may also be a slight amplitude increase at the posterior section. The whole analysis applies to ventral just as much as to dorsal fins, and suggests in either case that with anguilliform motion a succession of closely spaced fins is dynamically equivalent to a long continuous fin, as stated by Lighthill (1969; last sentence of §4).

In certain fishes, by contrast, including most cat-fishes of the sub-order Siluroidea (order Ostariophysi), anguilliform motion is retained although there is a large gap between the main dorsal fin and the caudal fin (any second dorsal fin being relatively insignificant). Analysis of their motion must take into account the extensive region of vortex sheet between the dorsal and caudal fins, where the above-noted differences (i) and (ii) may be sufficiently important to make the sheet behave very differently from a solid fin.

These differences make it necessary to define at each body section  $S_x$  a separate virtual mass  $\tilde{m}(x)$  per unit length for the vortex sheet in the presence of the body section. To do this, we bear in mind that the velocity potential on the two sides of the vortex sheet has a certain value proportional to

$$w_F = w(x_F, t - (x - x_F)/U), \quad (19)$$

the value of  $w$  at the fin's posterior edge  $x = x_F$  at the time when water now at  $S_x$  was passing that edge. The coefficient of proportionality depends on the fish's cross-sectional shape at  $x = x_F$ , and is given by (16) with  $s = s(x_F)$  in the special case of a fish thin in relation to its depth.

We continue to define  $m(x)$  as the virtual mass per unit length associated with the motion in the  $z$ -direction of an infinite rigid cylinder  $C_x$  with cross-section  $S_x$ ,

in the absence of any vortex sheet. For motion of  $C_x$  with velocity  $w$  relative to the water, the cross-flow potential  $\phi$  is that which satisfies

$$\partial\phi/\partial n = wn_z \quad \text{on } C_x, \quad \phi = 0 \quad \text{on vortex sheet}, \quad (20)$$

where  $n_z$  is the  $z$ -component of a unit vector normal to  $S_z$ . The momentum of the flow is  $m(x) w$  per unit length.

On the other hand, we define  $\tilde{m}(x)$  as the virtual mass associated with the vortex sheet in the presence of a completely stationary cylinder  $C_x$ . The associated potential  $\tilde{\phi}$  is that which satisfies

$$\partial\tilde{\phi}/\partial n = 0 \quad \text{on } C_x, \quad \tilde{\phi} = \text{given on vortex sheet}, \quad (21)$$

the given value being proportional to  $w_F$  (see equation (19)), and the momentum of the flow is taken to be  $\tilde{m}(x) w_F$  per unit length. The complete cross-flow, resulting from the motion of  $S_z$  in the presence of the vortex sheet, is defined, evidently, by the potential  $\phi + \tilde{\phi}$ , and its momentum is therefore

$$m(x) w + \tilde{m}(x) w_F. \quad (22)$$

The kinetic energy of this combined flow also has a simple form, because it follows from Green's theorem that the two cross-flows  $\phi$  and  $\tilde{\phi}$  are orthogonal in a space with kinetic energy as norm. (This requires that  $\oint \phi(\partial\phi/\partial n) ds$ , taken over a contour consisting of  $C_x$  and the two sides of the vortex sheet, is zero, which is obvious from equations (20) and (21).) Hence the kinetic energy is

$$\frac{1}{2}m(x) w^2 + \frac{1}{2}\tilde{m}(x) w_F^2. \quad (23)$$

The correctness of these formulas (22) and (23) is easily verified for particular cross-sectional shapes; thus, for a thin cross-section  $S_x$  we obtain them by direct calculation with

$$\tilde{m} = \frac{1}{4}\pi(\bar{s}^2 - s^2)\rho, \quad m = \frac{1}{4}\pi s^2\rho, \quad (24)$$

where  $\bar{s}$  and  $s$  are the local depths of cross-section with and without the vortex sheet included. The formulas (24) make clear that at  $x = x_F$  (where the vortex sheet starts) there is no discontinuity between the values of  $m(x)$  for  $x < x_F$  and of  $m(x) + \tilde{m}(x)$  for  $x > x_F$ . Actually, this is a general property for all shapes of body cross-section, because at  $x = x_F$  we have  $w = w_F$  by (19) and so by (22) the momentum (which as explained earlier must be continuous) is  $[m(x) + \tilde{m}(x)] w$ .

We may now apply the above results to the estimate of forces between the body and the water. We can no longer write the sideforce  $Z$  as equation (6), but instead (using (19) and (22)) as

$$\begin{aligned} Z &= \left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) \left[ m(x) w(x, t) + \tilde{m}(x) w \left( x_F, t - \frac{x - x_F}{U} \right) \right] \\ &= \left( \frac{\partial}{\partial t} + U \frac{\partial}{\partial x} \right) [m(x) w(x, t)] + U \tilde{m}'(x) w \left( x_F, t - \frac{x - x_F}{U} \right), \end{aligned} \quad (25)$$

provided that we remember that  $\tilde{m}$  jumps practically discontinuously at  $x = x_F$  from 0 to  $\tilde{m}(x_F)$ , so that the  $\tilde{m}'(x)$  in (25) includes a delta-function term  $\tilde{m}(x_F) \delta(x - x_F)$ , while according to the previous paragraph  $m'(x)$  includes an equal and opposite delta-function term.

Equation (7) for the rate  $E$  at which the fish is doing work through its transverse movements with velocity  $\partial h/\partial t$ , exerting forces  $Z$  per unit length, can now be transformed into a form similar to (8) but with an interesting extra term. The first term on the right of (25) gives precisely the terms on the right of (8), the argument being unmodified by the fact that  $m(x)$  is discontinuous at  $x = x_F$ . The term in  $\tilde{m}'(x)$  on the right of (25), on the other hand, contributes two terms to (7), one from the delta function and one from the remainder of  $\tilde{m}'(x)$ , so that we have

$$\begin{aligned} E = & \frac{\partial}{\partial t} \int_0^l \left( mw \frac{\partial h}{\partial t} - \frac{1}{2} mw^2 \right) dx + U \left[ mw \frac{\partial h}{\partial t} \right]_{x=l} \\ & + U \tilde{m}(x_F) w(x_F, t) \frac{\partial}{\partial t} h(x_F, t) \\ & + \int_{x_F}^l U \tilde{m}'(x) w \left( x_F, t - \frac{x-x_F}{U} \right) \frac{\partial}{\partial t} h(x, t) dx. \end{aligned} \quad (26)$$

The other method (10) of writing  $E$ , from which the thrust  $P$  can be obtained by subtraction, calculates rates of change of energy in the frame of reference in which the undisturbed water is at rest. With the modified form (32) of the kinetic energy per unit length, it becomes

$$E = UP + U \left[ \frac{1}{2} mw^2 + \frac{1}{2} \tilde{m} w_F^2 \right]_{x=l} + \frac{\partial}{\partial t} \int_0^l \left( \frac{1}{2} mw^2 + \frac{1}{2} \tilde{m} w_F^2 \right) dx. \quad (27)$$

The fishes with which we are here concerned have the depth of their caudal fin practically as large as (or larger than) that of the dorsal fin and body combined. These, then, are fishes which reabsorb the vortex sheet from the dorsal fin on to the caudal fin, where it gets mixed with the boundary layer, so that the vortex sheet finally shed from the caudal fin, being determined by that fin's motion relative to the water, is practically uninfluenced by the incident sheet. At the trailing edge  $x = l$ , then, no separate vortex sheet from the dorsal fin remains, which means that  $\tilde{m} = 0$  and that the second term on the right of (27) takes the same form as in (10); in other words, rate of shedding of kinetic energy of water motions into the wake is unaffected by the vortex sheet in the gap between the dorsal and caudal fins.

Admittedly, the last term on the right of (27) is modified by this vortex sheet, but this is still the rate of change of a fluctuating quantity and so has mean value zero. Hence the mean rate of wastage of energy  $\bar{E} - UP$  is unaffected, according to the approximations here used, by the vortex sheet in the gap. This makes it particularly interesting to enquire whether the total mean rate of working  $\bar{E}$  given by (26) can be increased by the new vortex-sheet terms, because if so then all the extra power should be effective propulsively and the fish should achieve simultaneous improvements in thrust and efficiency.

This does in fact seem to be possible, and for reasons which can be given a clear physical interpretation. The first term in (26) has zero mean, and the second is the usual  $UmwW$  term, interpreted earlier as rate of shedding of lateral momentum from the caudal fin times trailing-edge lateral velocity. This term can, as we know, have a positive mean. The third term has an identical interpretation in relation to

the dorsal fin and can for the same reason (essentially because  $V > U$ ) have a positive mean.

We must ask therefore whether the last term in (26), in which the factor  $\tilde{m}'(x)$  must on the average be negative as  $\tilde{m}$  decreases from the positive value  $\tilde{m}(x_F)$  to the zero value  $\tilde{m}(l)$ , must necessarily cancel the third term in mean value. Certainly we can interpret this last term as a negative rate of working, equal to minus an integral over the caudal fin's leading edge of the product of its rate of picking up of lateral momentum from the vortex sheet,  $U(-d\tilde{m})w_F$ , with the lateral velocity of the caudal-fin leading edge  $\partial h/\partial t$ . That term is actually negative however only if the quantity  $w_F \partial h/\partial t$  occurring in it possesses a positive mean value.

This might be thought inevitable if  $V > U$ , and the mean value in question might be supposed greater even than the value at the dorsal-fin trailing edge, because of posterior increase in amplitude. This is not necessarily true, however, if there is a substantial phase difference between  $w_F$  and  $\partial h/\partial t$  at the caudal-fin leading edge. If this phase difference exceeds  $\frac{1}{2}\pi$ , their mean product might be negative.

In terms of a mean position  $x = x_C$  of the caudal-fin leading edge, this phase difference has a mean value

$$\frac{2\pi(x_C - x_F)}{\lambda} \left( \frac{V}{U} - 1 \right), \quad (28)$$

where  $\lambda$  is the wavelength of the undulation. This is because  $2\pi(x_C - x_F)/\lambda$  is the phase difference in  $\partial h/\partial t$  between  $x = x_F$  and  $x_C$ , while the phase difference in  $w_F$ , as (19) shows, is greater by a factor  $V/U$  owing to vorticity convection at a speed  $U$  slower than the speed  $V$  of the body undulation. With a gap  $(x_C - x_F)$  of at least half a wavelength, as is common, and typical values of  $U/V$  around  $\frac{2}{3}$ , the phase difference (28) can well be  $\frac{1}{2}\pi$  or more.

The above argument seems to indicate that a large enough gap between the dorsal and caudal fins can improve the anguilliform mode by increasing total power output without any increase in the power wasted in creating a vortex wake. This is because momentum shedding from the dorsal fin, in phase with its lateral motion, causes mean power to be exerted but the rate of annihilation of that momentum, after it has reached the caudal fin at a speed  $U$  less than the propagation speed  $V$  of the body undulation, is out of phase (or is even in antiphase) with the caudal fin's lateral velocity, and so does not produce any balancing reduction in power output.

It is conceivable that the arguments of this section might have some bearing on the problem of the evolution of discrete dorsal fins. Primitive jawless vertebrates of the class Cyclostomata, such as lampreys and hagfishes, possess continuous dorsal fins. We have suggested that their replacement at some stage of fish evolution by a sequence of closely spaced discrete fins would not be disadvantageous as regards thrust, while its effect on drag reduction might be beneficial. Later on, reinforcement of the first dorsal fin and reduction of the others might be favoured by additional (phase-related) considerations of thrust advantage in the first place, although advantages relating to stability and control in yaw would soon emerge.

#### 4. Mechanics of the carangiform mode

Still later, further improvements in thrust and efficiency might result from gradual conversion to the carangiform mode, in which the dorsal fin no longer sheds a vortex sheet, playing a different, although still essential, role. Mechanical considerations relating to this further development must now be set out. We shall see that the confinement of undulations to a reduced fraction of the fish's length in the neighbourhood of the caudal fin produces a gain in efficiency, which however might well have been cancelled out by losses due to 'recoil' effects if adoption of carangiform motion had not been accompanied by morphological changes tending to minimize these.

To understand the gain, we must further refine the theoretical considerations of §2 by taking into account the fact (noted already in the paragraph following equation (3)) that pushing a slice of water with a body of virtual mass  $m$  gives the water not only an immediate access of momentum  $mw$  but also a subsequent gradual further increase of momentum. This is associated with the vortex-force on the body, which is a direct result of vorticity progressively shed by the body as the pushing continues. At the same time the water's kinetic energy increases above the energy  $\frac{1}{2}mw^2$  of the irrotational flow (by an amount equal, as Kelvin's 'minimum energy theorem' shows, to the energy that the vortex system would possess with the body stationary).

Lighthill (1960) suggested very tentatively that the theory of §2 could be extended as follows to take into account changes (if they could be estimated) in the water's momentum  $M$  and kinetic energy  $T$  per unit length from the values  $mw$  and  $\frac{1}{2}mw^2$ . Note that various checks made on equations (9) and (10) in §2 suggest that the physical ideas leading directly to those equations may possibly be reliable in more general cases. If these are used to predict only mean values  $\bar{P}$  and  $\bar{E}$  of thrust and rate of working, they yield

$$\bar{E} = \overline{U[MW]}_{x=l}, \quad (29)$$

$$\bar{E} - U\bar{P} = \overline{U[\bar{T}]}_{x=l}, \quad (30)$$

the differences being in each case the mean value of a rate of change of kinetic energy of the water motions adjacent to the body, in a frame of reference in which either the body, or the water respectively, is at rest. Such kinetic energy immediately adjacent to the body in each case fluctuates periodically, so that its time rate of change has mean value zero.

For mean thrust and efficiency, equations (29) and (30) imply

$$\bar{P} = [\bar{MW} - \bar{T}]_{x=l} = \{V/(V-U)\} \overline{Mw} - \bar{T}]_{x=l}, \quad (31)$$

$$\eta = 1 - \{(V-U)/V\} [\bar{T}]_{x=l} / [\bar{Mw}]_{x=l}, \quad (32)$$

equations in which equation (3) is still used to express  $W$  in terms of  $w$ . These suggest that parts of the momentum  $M$  at the trailing edge  $x = l$  which are badly correlated with the local pushing velocity  $w$  may be ineffective for producing thrust, while the associated parts of the kinetic energy  $T$  might add very considerably to the wasted energy in the vortex wake.

The momentum per unit length of fish,  $M$ , in a water slice normal to that length that has just reached the trailing edge, has been changed from the virtual-mass value  $mw$  by an amount equal to the vortex-force applied by different sections of the fish which it passed, integrated with respect to time. Unfortunately, it is impossible to attempt an accurate estimate of vortex-force on the water slice in the exceedingly unsteady conditions while it is pushed at varying velocities by cross-sections of varying shapes, and it is even harder to estimate the addition to  $\frac{1}{2}mw^2$  at the trailing edge, namely the local value of the vortex-system energy. However, very crude estimates suffice to make plausible the substantial reduction in efficiency of anguilliform motion through these effects, and the virtual disappearance of such reduction resulting from a modification into carangiform motion.

The crude estimate used here, for vortex-force per unit length due to pushing at velocity  $w$ , is based on two simplifying approximations. First, a drag depending only on  $w$  and on the cross-sectional depth  $s$ , and characteristic of steady motion of a long thin strip at right angles to its plane (Flachsbart 1935) is used; this is  $\rho sw^2 \operatorname{sgn} w$ , where the  $\operatorname{sgn} w$  (+1 if  $w > 0$  and -1 if  $w < 0$ ) expresses that the drag force is in the same direction as the velocity. Secondly, an approximate linearization of this is used. If  $w$  varies between  $-w_0$  and  $+w_0$ , then the r.m.s. absolute error in replacing  $w^2 \operatorname{sgn} w$  by the regression line  $\frac{3}{4}w_0 w$  is only  $0.11w_0^2$ . This linearization is reasonable in the present context where drags at different velocities contribute linearly to the total momentum  $M$  (so that only absolute errors, not the relative errors which are far bigger for small  $w$ , are relevant).

This leads to the value  $\frac{3}{4}\rho sw_0 w$  for vortex-force, which used together with the virtual-mass contribution to momentum is like representing the water reaction to motion of fish cross-sections as a combination of a linear resistance and an inheritance. The excessive momentum at the trailing edge  $x = l$  at time  $t$  is now represented as an integral with respect to time  $t_1$  of a vortex-force proportional to the value at that time of the product  $sw$  at the point  $x = l - U(t - t_1)$  where the water slice then was:

$$\begin{aligned}[M]_{x=l} &= [mw]_{x=l} + \int_{t-U/l}^t \frac{3}{4}\rho w_0 [sw(l - U(t - t_1), t_1)] dt \\ &= [mw]_{x=l} + \int_0^l \frac{3}{4}\rho w_0 [sw(x, t - (l - x) U^{-1})] dx/U.\end{aligned}\quad (33)$$

We can compare the phase relationships between  $M$  and  $w$  at  $x = l$  for different rates of growth of wave amplitude with  $x$  by evaluating (33) with

$$sw(x, t) = s_0 w_0 e^{\kappa(x-t)} \cos [\omega(t + (l - x) V^{-1})], \quad (34)$$

where  $V$  is the wave speed relative to the body and  $\kappa$  is small for anguilliform motion but large for carangiform. If  $m$  is approximately represented (see § 2) as  $\frac{3}{8}\rho s^2$  this gives

$$[M]_{x=l} = \frac{3}{4}\rho s_0^2 w_0 \cos \omega t + \frac{\frac{3}{8}\rho s_0 w_0^2 [\kappa \cos \omega t + \omega(U^{-1} - V^{-1}) \sin \omega t]}{U[\kappa^2 + \omega^2(U^{-1} - V^{-1})^2]}, \quad (35)$$

where the small contribution to the integral in (33) from the lower limit has been neglected. Hence

$$[\overline{Mw}]_{x=l} = \frac{3}{8}\rho s_0^2 w_0^2 + \frac{\frac{3}{8}\rho s_0 w_0^2 \kappa}{U[\kappa^2 + \omega^2(U^{-1} - V^{-1})^2]}. \quad (36)$$

These results show that for small  $\kappa$  the phase lag relative to  $[w]_{x=l}$  in the part of  $[M]_{x=l}$  due to momentum gain from anterior sections reduces considerably the value (36) of their mean product. The value of  $[T]_{x=l}$  is not so affected. The part due to vortex-force can be estimated as an integral with respect to time  $t_1$  of  $w$  times the vortex-force, giving

$$[T]_{x=l} = [\frac{1}{2}mw^2]_{x=l} + \int_0^l \frac{3}{4}\rho w_0 [sw^2(x, t - (l-x)U^{-1})] dx/U. \quad (37)$$

Now taking  $sw^2(x, t) = s_0 w_0^2 e^{r(x-l)} \cos^2[\omega(t + (l-x)V^{-1})]$ , (38)

which with  $s$  taking its maximum value  $s_0$  at the trailing edge implies  $\kappa \leq \tau \leq 2\kappa$ , equation (37) becomes

$$[\bar{T}]_{x=l} = \frac{3}{16}\rho s_0^2 w_0^2 + \frac{3}{8}\rho s_0 w_0^3 U^{-1} \tau^{-1}. \quad (39)$$

We conclude from (39) that the energy flow into 'wasted' wake motions is increased through vortex-force effects on body cross-sections by a factor about  $1 + 2w_0/U\tau s_0$ . Here  $2w_0/U$  is of order 1 for a wide range of fish motions, so the increase in wasted energy is significant if  $\tau$  is of order of magnitude around  $s_0^{-1}$  or less, as in anguilliform motion. This increase is not balanced by a corresponding increase of thrust because for such values of  $\tau$  (and corresponding values of  $\kappa$  down to one-half as much) (36) shows  $[\bar{M}]_{x=l}$  to be insignificantly increased† essentially because the addition to  $[M]_{x=l}$  is poorly correlated with  $[w]_{x=l}$ , when, as is normal, the term  $\omega(U^{-1} - V^{-1})$  in (36) is itself of order  $s_0^{-1}$ .

For carangiform motion, on the other hand, there is a rapid increase of amplitude of motions near the caudal fin, and  $\tau$ , the relative rate of increase of  $sw^2$ , is significantly larger than  $s_0^{-1}$ , generating a far smaller increase in wasted energy. Thus, carangiform motion represents the logical dénouement of the process begun by posterior lateral compression. The utilization of the virtual-mass effect (the water's 'inertance' as opposed to its 'resistance'), which lateral compression made possible, becomes in carangiform motion almost the exclusive basis of thrust production.

With the transition to carangiform motion, however, there arises a possibility of one other source of thrust reduction through imperfect correlation of  $[M]_{x=l}$  and  $[W]_{x=l}$ , namely the 'recoil' effect (Lighthill 1960). This can lead to additional vortex-force, and also to a possible departure from equation (3), which states that  $W$  and  $w$  are perfectly in phase; such departure may decrease the rate of working, and reduction in the ratio of their amplitudes may at the same time increase wasted energy. The derivation of that equation, and others, assumed that a fish can by muscular action cause a wave of lateral movement specified by some definite equation for  $h(x, t)$  to pass down its body. This ignores the fact that the lateral movement of the fish is subject to two overriding laws of motion:

† The estimate (36) for  $\bar{E}(V-U)/VU$  is based on the tentative equation (29), which may overestimate the extent to which the effect of vortex-force wakes is modified by posterior motions. However, an opposite assumption, that the increase in  $\bar{E}$  due to vortex-forces is exactly equal to the extra work that they locally do through pushing at velocity  $W$ , would replace the second term in (36), as an equation for  $\bar{E}(V-U)/VU$ , by  $\frac{3}{8}\rho S_0 w_0^3 U^{-1} \tau^{-1}$ , and the relative increase in work done would still be only half as much as the relative increase in wasted energy.

(i) the rate of change of lateral momentum of the fish must equal the total side-force with which the water acts on it; (ii) the rate of change of its angular momentum about a fixed axis (which we take as the axis  $x = l$ ,  $z = 0$ ) must equal the moment of the sideforces with which the water acts on it.

Actually the fish's muscular contractions can only determine changes in its shape relative to the centre of gravity; at the same time, translations and rotations of that shape must in general accompany those changes, and must be such that conditions (i) and (ii) above are satisfied. For example, an attempt to generate an undular lateral displacement

$$h = H(x) \cos [\omega(t + (l - x) V^{-1})] \quad (40)$$

will in practice generate the displacement

$$h = H(x) \cos [\omega(t + (l - x) V^{-1})] + [h_1 + h_3(l - x)] \cos \omega t + [h_2 + h_4(l - x)] \sin \omega t, \quad (41)$$

where the constants  $h_1$ ,  $h_2$ ,  $h_3$  and  $h_4$  must be determined by applying conditions (i) and (ii) above and equating the coefficients of  $\cos \omega t$  and  $\sin \omega t$  on both sides.

We carry out this calculation, with vortex-force neglected, as follows. We first obtain the 'net' total sideforce (by 'net' we mean after correction for rate of change of the fish's own momentum) associated with the simple motion (40), say

$$\int_0^l (Z + m_f \partial^2 h / \partial t^2) dx = Z_1 \cos \omega t + Z_2 \sin \omega t, \quad (42)$$

where  $m_f(x)$  is fish mass per unit length, and the 'net' moment of sideforces about the line  $x = l$ ,  $z = 0$  (which is the undisturbed position of the trailing edge),

$$\int_0^l (l - x) (Z + m_f \partial^2 h / \partial t^2) dx = Z_3 \cos \omega t + Z_4 \sin \omega t. \quad (43)$$

Then we find what full expression of the form (41) would make such additions to both the total sideforce and the moment of sideforces that conditions (i) and (ii) would be satisfied. This requires the linear terms in (41) to be such that the integrals in (42) and (43) when calculated for them alone take the same values but with the sign changed.

The calculation is done in two parts like this because different terms are dominant in the quantitative expressions for elements in (42) and (43) which arise from the sinusoidal term (40) in  $h$  and from the linear terms added to  $h$  in (41). For example, the sinusoidal term in carangiform motion (with  $H(x)$  significant only rather near  $x = l$ ) produces a negligible fish-mass element in (42) and (43), because posterior lateral compression makes fish mass negligible compared with virtual mass of water near  $x = l$ , while corresponding virtual-mass elements, on the other hand, are dominated by the tendency, expressed in equation (3), for values of  $w$  associated with this term to be considerably smaller than corresponding values of  $W$ .

By contrast, the linear terms in (41) produce fish-mass elements in (42) and (43) that involve the fish's whole inertia, as well as its first and second moments about  $x = l$ . Furthermore, for these linear terms the  $U \partial / \partial x$  parts in (2) and (6) are not nearly so important as the  $\partial / \partial t$  parts for the actual frequencies  $\omega$  that are used (with  $\omega l/U$  of order 10). As a first crude approximation, neglecting the  $U \partial / \partial x$  parts, those equations give  $Z = m(x) \partial^2 h / \partial t^2$ , so that the integrals in (42) and (43)

can be calculated as for a rigid body subject to no external forces whose mass distribution is  $m(x) + m_f(x)$  per unit length.

This first crude approximation, that the fish's recoil to given sideforces takes the form of an inertial response based on the combined inertia of the fish and the water, is quite a useful one. In terms of

$$M = \int_0^l (m + m_f) dx, \quad ML = \int_0^l (l - x) (m + m_f) dx, \quad I = \int_0^l (l - x)^2 (m + m_f) dx, \quad (44)$$

where  $L$  is distance of the centre of inertia from the trailing edge, and  $I - ML^2$  is the moment of inertia about that centre of inertia, it gives

$$\begin{pmatrix} M\omega^2 & ML\omega^2 \\ ML\omega^2 & I\omega^2 \end{pmatrix} \begin{pmatrix} h_1 + ih_2 \\ h_3 + ih_4 \end{pmatrix} = \begin{pmatrix} Z_1 + iZ_2 \\ Z_3 + iZ_4 \end{pmatrix}, \quad (45)$$

with the simple solution

$$\frac{[(I/M)Z_1 - LZ_3, (I/M)Z_2 - LZ_4, Z_3 - LZ_1, Z_4 - LZ_2]}{\omega^2(I - ML^2)} \quad (46)$$

for  $[h_1, h_2, h_3, h_4]$ . The exact form of the matrix that should appear on the left-hand side of (45) can, actually, be calculated as

$$\begin{pmatrix} M\omega^2 + m(l)Ui\omega & ML\omega^2 - M_wUi\omega + m(l)U^2 \\ ML\omega^2 + M_wUi\omega & I\omega^2 + M_wU^2 \end{pmatrix}, \quad (47)$$

where  $M_w = \int_0^l m dx$  is the total virtual mass (inertia of the water alone for lateral rigid-fish movements), and the departures of the associated exact solutions from those in (46) are not great.

These considerations on recoil are potentially rather important for carangiform motion (with  $H(x)$  in (40) significant only in a short posterior portion of the fish) because the associated total sideforce (42) is potentially rather large. By equation (2), the  $w$  associated with (40) is

$$w = UH'(x) \cos [\omega(t + (l - x)V^{-1})] - (V - U)V^{-1}H(x)\omega \sin [\omega(t + (l - x)V^{-1})]. \quad (48)$$

This formula emphasizes the fact that the small value of  $w$  relative to  $W$  implied by equation (3) is confined to that small region near the trailing edge where the amplitude  $H(x)$  of undulation has reached a plateau with  $H'(x)$  negligible. Much greater values of  $w$  are to be expected in the anterior region where  $H(x)$  is rapidly increasing.

If the virtual mass  $m(x)$  continues in this region to take large values (such as were shown in § 2 to be necessary at the trailing edge itself) then the fluid momentum  $m(x)w(x, t)$  rises to large peak values, and so the associated sideforces (6) are very big. In fact, their total integrated value

$$\int_0^l Z dx = \frac{\partial}{\partial t} \int_0^l mw dx + [Umw]_{x=1} \quad (49)$$

consists of two terms, one associated with fluctuations in the lateral momentum of water movements anterior to the trailing edge, and one associated with

momentum being transferred to the wake. The latter is important in relation to mean rate of working (§2), but the former constitutes a much bigger fluctuating sideforce when  $m_w$  possesses a large peak.

If the rapid rise of  $H(x)$  from 0 to its trailing-edge value  $H(l)$  is centred around  $x = x_r$ , where  $m$  takes values that are not depressed to compensate for that rapid rise, then a rough estimate of (49) in which the integral is estimated solely by the contribution from the rise, taken as instantaneous, is

$$\int_0^l Z dx = -Um(x_r) H(l) \omega \sin [\omega(t + (l - x_r) V^{-1})] \\ - Um(l) H(l) (V - U) V^{-1} \omega \sin \omega t. \quad (50)$$

The two terms differ only moderately in phase, and the first term greatly augments the whole through its lack of the  $(V - U) V^{-1}$  factor. Similar augmentation is absent with anguilliform motion, where spatial variation of the cosine term in (48) greatly reduces the fluctuations in the integral (49).

It is clear from these considerations that, unless the adoption of carangiform motion is accompanied by a substantial reduction of cross-section depth (with associated large reduction in  $m(x)$ , as (4) shows) in the neighbourhood of  $x = x_r$ , where the rapid rise in wave amplitude occurs, total sideforce will be large. Under those circumstances, recoil amplitudes will also be large, as equation (46) implies. Actually the  $Z_1$  and  $Z_2$  terms are the important ones there, and the moment of sideforce about  $x = l$  is less important (because the effective action of the sideforce is near  $x = l$ ). The lateral recoil is roughly obtained, therefore, by multiplying (50) by

$$I/[M(I - ML^2)\omega]. \quad (51)$$

Here  $I - ML^2$  is the moment of inertia of fish plus water around its centre of inertia and can be written  $Mk^2$ , where  $k$  is a radius of gyration and can be expected to be substantially less than  $L$ . The recoil amplitude associated with the first term in the total sideforce (50) would therefore be

$$[(L^2 + k^2)/k^2] [m(x_r)l/M] (U/\omega l) H(l), \quad (52)$$

which is a by no means negligible fraction of  $H(l)$  itself. Although the  $U/\omega l$  factor is of order  $10^{-1}$ , the factor  $(L^2 + k^2)/k^2$  can easily be of order 4 or 5, and the factor  $m(x_r)l/M$  is of order 1 unless  $m(x_r)$  is artificially reduced.

Substantial angular recoil (given by  $h_3$  and  $h_4$ ) is also produced, and large lateral motions are generated all along the length of the fish. This wipes out the whole potential advantage of carangiform motion, namely its confinement of lateral motion to a short posterior portion. Water arriving at the trailing edge has been subjected to vortex drag, not just while it traversed such a short posterior portion, but for a much longer time, and has acquired additional vortex motions with substantial energy whose momentum is badly correlated with the trailing-edge velocity (both results being harmful, by (29) and (30)).

Accordingly, it is not surprising that the adoption of carangiform motion goes together (Lighthill 1969) with a large reduction of depth of body and fin in the region of abrupt amplitude increase. Then near  $x = x_r$ , the virtual mass  $m(x)$  is so small that there is no peak in lateral momentum, and hence no associated large sideforces.

Some sideforce still remains, however; smaller, admittedly, owing to the  $(V - U) V^{-1}$  factor in the second term of (50), but still capable of producing observable recoil, since  $m(l)$  must be large for adequate thrust production. It is most important, therefore, that the multiplying factor (51) is kept small, so that the residual sideforce does not generate too much recoil. This requires a large fish depth in an extensive length of the anterior part of the fish.

To see this, suppose that the caudal-fin shape were specified (primarily by thrust considerations) and had total inertia  $M_C$ , and consider different specifications for the anterior portion of fish, with total inertia  $M_A$ , and centre of inertia at  $x = l - l_A$ , about which its moment of inertia is  $M_A k_A^2$ . Then  $M = M_A + M_C$ , and approximately  $ML = M_A l_A$ ,  $I = M_A(l_A^2 + k_A^2)$ . It follows that the multiplying factor (51) can be written

$$\omega^{-2} [M_C + M_A k_A^2 / (l_A^2 + k_A^2)]^{-1}. \quad (53)$$

For this to be small requires above all (since  $k_A^2$  is necessarily a good deal smaller than  $l_A^2$ ) that the part of the fish anterior to the caudal fin possess a large moment of inertia  $M_A k_A^2$  about its centre of inertia, taking inertia of both fish and water into account.

Morphologically, this means that for efficient carangiform motion a fish needs a long anterior section of body with its mass so disposed that a large virtual mass of water is drawn into participation in its lateral movements. A degree of lateral compression now becomes advantageous in the anterior part of the fish. Additionally, a large overall depth  $s$  between dorsal and ventral fin is most valuable for increasing  $m(x)$  (in proportion, by equation (4), to  $s^2$ ), and the greater the proportion of the fish's length in which such an increased depth is found the more effectively is the recoil reduced.

Among the disadvantages of recoil, the above discussion has emphasized the associated vortex-force losses, but any substantial recoil would also bring other difficulties. At  $x = l$ , the full equation (41) implies that

$$W = \partial h / \partial t = -\omega [H(l) + h_1] \sin \omega t + \omega h_2 \cos \omega t, \quad (54)$$

and the equations for the recoil coefficients show that  $h_1$  is negative, which reduces the amplitude of  $W$  at the trailing edge (for given fish deformations) and so tends to reduce the mean thrust

$$\bar{P} = [m(\bar{w}\bar{W} - \frac{1}{2}\bar{w}^2)]_{x=l}. \quad (55)$$

(Specifically, equation (50) or any other estimate of (49) shows not only that  $Z_2$  is large and negative, but also that  $Z_1$ , the coefficient of  $\cos \omega t$ , is negative; the moments (43) are smaller but it turns out that  $Z_3$  is positive, and both results combine in (46) to make  $h_1$  (as well as  $h_2$ ) negative. If the more exact matrix (47) is used, the phase of  $h_1 + ih_2$  is reduced slightly, which makes  $h_1$  still more negative.)

Next, the perfect correlation between  $w$  and  $W$  expressed by equation (3) is in general destroyed at the trailing edge by recoil effects, with harmful effects on the mean thrust (55) and on the efficiency (12). At  $x = l$ , equations (2) and (41) give

$$w = -[\omega(V - U)V^{-1}H(l) + \omega h_1 + Uh_4] \sin \omega t + [UH'(l) + \omega h_2 - Uh_3] \cos \omega t. \quad (56)$$

In the absence of recoil it is important that  $H'(l)$  be negligibly small in (56) (i.e. that just before the trailing edge wave amplitude cease to increase fast) so that  $w$  and  $W$  are well correlated. However, if  $H'(l)$  is negligible, then recoil (which makes both  $\omega h_2$  and  $-Uh_3$  negative) produces a greater coefficient of  $\cos \omega t$  in  $w$  than in  $W$ , with harmful effects on the efficiency (12). Further, adoption of a positive value of  $H'(l)$  such that the recoil coefficients just made the coefficient of  $\cos \omega t$  in (56) small and negative (to restore the correlation with  $W$ ) might involve too fine an adjustment to be useful as a practical way of getting good efficiency.

To sum up, it is not surprising that fishes adopting carangiform motion possess features suitable for diminishing recoil—particularly, a greatly reduced depth of body in the region immediately anterior to the caudal fin where wave amplitude is increasing rapidly, and, farther forward still, a long region characterized by greatly augmented depth.

## 5. Two-dimensional theory of the lunate tail

It is possible to trace, from the hydromechanical point of view, three distinct stages in the further development of carangiform propulsion in teleosts (bony fishes), aimed at improving speed and efficiency. First, there is the 'scooping out' of a central posterior portion of the caudal fin (as in the herring *Clupea*) so that the fin becomes geometrically like a pair of highly sweptback wings. Provided that the posterior part of the caudal fin is moving as a rigid whole, the discussion at the beginning of §3 above suggests that no thrust would be lost by this modification, because the vortex-sheet strength in the gap would be the same as its strength on solid parts of fin. The decline in surface area, on the other hand, could produce some slight reduction of resistance and associated gain in speed.

Secondly, there is the reduction of sweepback of the two wing-like surfaces then composing the caudal fin (as in the horse-mackerel *Caranx*). This increases trailing-edge span  $s$  (with important thrust benefits, by (4) and (55)), without increasing surface area, and a still greater speed improvement should result. Where, as with *Caranx*, the degree of sweepback can be varied by muscular action, the animal acquires a valuable extra measure of control.

At some point during the process of decreasing sweepback, the elongated-body theory of §§2–4 becomes inapplicable, as explained in §1. Certainly, it is out of the question to apply it to the third type stage of development of carangiform propulsion, when the caudal fin acquires high aspect-ratio  $s^2/S$  (where  $S$  is its surface area) and the 'lunate' form described in the introduction. This seems to be some sort of culminating point of the process of improvement of speed and efficiency in the teleosts, a supposition that is borne out by the fact that a similar lunate tail was acquired by the fastest sharks and by the cetacean mammals (as well as by *Ichthyosaurus*) through quite different evolutionary processes.

Without discussing these pathways of convergent evolution further, the rest of this section is devoted to a first approximate theory of the carangiform motion of animals with lunate tails. For reasons described in §1, such a first

approximate theory must be a two-dimensional theory, in which the caudal fin (in the case of the fishes) is considered to act independently on each separate horizontal slice of water. This theory must overestimate efficiency because it takes into account only the cross-stream elements of wake vorticity. Here it is set out, however, in a form designed to facilitate later extension to a lifting-line theory that would also take streamwise wake vorticity into account.

In the two-dimensional theory, the section of the caudal fin (in its symmetrical, undisturbed position) by a horizontal water slice  $y = \text{constant}$  is taken as stretching from  $x = -a$  to  $x = +a$  (so that the origin of  $x$  is shifted to the half-chord position), and the undisturbed fluid flow has velocity  $U$ . The section's lateral displacement is taken as

$$z = [h - i\alpha(x - b)] e^{i\omega t}, \quad (57)$$

where  $h$  and  $\alpha$  are real numbers signifying the amplitude of the sideslip and yawing motions respectively, and  $x = b$ ,  $z = 0$  is the yawing axis. A  $90^\circ$  phase difference between the sideslip and yawing motions is assumed, but note that any other phase difference, represented by giving an imaginary part to  $h$  in (57), is simply equivalent to a change in  $b$ .

The velocity potential  $\phi e^{i\omega t}$  of flow disturbances satisfies boundary conditions which, on linearization, can be applied on  $z = 0$  and take the following form. For  $-a < x < a$ ,

$$(\partial\phi/\partial z)_{z=0} = i(\omega h - U\alpha) + \omega\alpha(x - b), \quad (58)$$

which equates lateral velocity of fluid to the rate of change,  $\partial z/\partial t + U \partial z/\partial x$ , of the section's lateral displacement (57) relative to a particular water slice. Symmetry shows also that  $\phi$  must be an odd function of  $z$ , and this means on  $z = 0$  that  $\phi$  vanishes for  $x < -a$  (upstream of the leading edge). This argument fails for  $x > a$ , downstream of the trailing edge, where a vortex sheet must be present for the same reasons as were adduced in §3 (but in this two-dimensional flow including only cross-stream vorticity). Its strength is unchanging relative to the fluid, so that on  $z = 0$ ,  $x > a$  the quantity

$$\Phi = i\omega\phi + U \partial\phi/\partial x \quad (59)$$

is a continuous as well as an odd function of  $z$  and therefore vanishes.

It is convenient, following Possio (1938) and also Wu (1961), to use  $\Phi$  as a new dependent variable, particularly since it is continuous throughout the fluid and vanishes on  $z = 0$  for both  $x < -a$  and  $x > +a$ . Also, it has important physical significance, since departures of the pressure  $p$  from its hydrostatic value  $p_0$  take the form

$$p - p_0 = -\rho\Phi e^{i\omega t} \quad (60)$$

on linearized theory, and thus it is the continuity of pressure across the vortex sheet which is essentially being used. For  $-a < x < a$ , we have by (58) and (59)

$$(\partial\Phi/\partial z)_{z=0} = B + Cx, \quad \text{where } B = 2U\omega\alpha - \omega^2(h + iab), \quad C = i\omega^2\alpha, \quad (61)$$

an equation which relates lateral pressure gradient to lateral acceleration of a fluid element resulting from the fin's displacement (57).

Solutions of  $\nabla^2\Phi = 0$  satisfying (61) are easily obtained, but we must consider carefully what singularities can be permitted, or are necessary, so that such a

solution corresponds to a value of  $\phi$  satisfying the boundary conditions. As in many steady-oscillation problems, it is convenient to think of  $\omega$  as possessing a small negative imaginary part (which is later allowed to tend to zero), representing a slow build-up of the oscillation to its present level. Then equation (59) can be solved for  $\phi$  as

$$\phi = U^{-1} \int_{-\infty}^x \Phi(x_1, z) e^{i\omega(x_1-x)/U} dx_1, \quad (62)$$

which automatically vanishes for  $z = 0, x < -a$  (as it should) if  $\Phi$  does.

The singularities of  $\phi$  needed to represent flow near the leading and trailing edges are well known. Near a sharp trailing edge  $x = a$ , where the Kutta-Zhukovskii condition of no flow around the edge is applied,  $\phi$  has a three-halves power singularity (so that  $(\phi)_{z=0}$  is of order  $(a-x)^{\frac{3}{2}}$ ). This means that  $\Phi$ , and the pressure

$\Phi$	$\Phi_1$	$\Phi_2$	$\Phi_3$
$\Omega + i\Phi$	$1 - [(\zeta - a)/(\zeta + a)]^{\frac{1}{2}}$	$\zeta - (\zeta^2 - a^2)^{\frac{1}{2}}$	$\frac{1}{2}\zeta^2 - \frac{1}{2}\zeta(\zeta^2 - a^2)^{\frac{1}{2}}$
$(\Omega)_{z=\pm 0} \begin{cases}  x  < a \\  x  > a \end{cases}$	$1$	$x$	$\frac{1}{2}x^2$
	$1 - [(x-a)/(x+a)]^{\frac{1}{2}}$	$x - (x^2 - a^2)^{\frac{1}{2}} \operatorname{sgn} x$	$\frac{1}{2}x^2 - \frac{1}{2} x (x^2 - a^2)^{\frac{1}{2}}$
$(\Phi)_{z=\pm 0} \begin{cases}  x  < a \\  x  > a \end{cases}$	$\mp[(a-x)/(x+a)]^{\frac{1}{2}}$	$\mp(a^2 - x^2)^{\frac{1}{2}}$	$\mp\frac{1}{2}x(a^2 - x^2)^{\frac{1}{2}}$
	$0$	$0$	$0$
$Z = 2 \int_{-a}^a (\Phi)_{z=-0} dx$	$2\pi a$	$\pi a^2$	$0$
$Q = 2 \int_{-a}^a (-x)(\Phi)_{z=-0} dx$	$\pi a^2$	$0$	$-\frac{1}{8}\pi a^4$

TABLE 1. This specifies the properties of three solutions  $\Phi_1$ ,  $\Phi_2$  and  $\Phi_3$  of  $\nabla^2\Phi = 0$ , each tending to zero as  $x^2 + z^2 \rightarrow \infty$ , and each defined in terms of the behaviour of a function  $\Omega + i\Phi$  of  $\zeta = x + iz$ , regular outside a cut from  $\zeta = -a$  to  $\zeta = +a$ . Here the square roots of  $(\zeta - a)/(\zeta + a)$  and  $\zeta^2 - a^2$  mean those 'branches' which behave like 1 and  $\zeta$  respectively for large  $|\zeta|$ . Square roots of functions of the real variables  $x$  mean the positive square roots. Across the cut  $z = 0, |x| < a$ ,  $\Omega$  is continuous but  $\Phi$  is discontinuous. At all points

$$\partial\Phi/\partial z = \partial\Omega/\partial x$$

(60), have square-root singularities. The leading edge  $x = a$ , on the other hand, permits good hydrodynamic performance *only* if it is well rounded so as to permit flow around the edge, and this corresponds to a square-root singularity in  $\phi$  (so that  $(\phi)_{z=0}$  is of order  $(x+a)^{\frac{1}{2}}$ ), and an inverse-square-root singularity in  $\Phi$ . (More accurate analysis places the singularity a short distance behind the leading edge, equal to half the section's radius of curvature (Lighthill 1951), but the calculations placing it at  $x = -a$  give a good approximation when overall forces are being calculated.)

Those considerations specify the required solution of  $\nabla^2\Phi = 0$  satisfying (61) as

$$\Phi = A\Phi_1 + B\Phi_2 + C\Phi_3 \quad (63)$$

in terms of functions  $\Phi_1$ ,  $\Phi_2$ , and  $\Phi_3$  specified in table 1. Here the coefficient  $A$

is arbitrary, and the term in  $\Phi_1$  must be included to allow for the expected singularities just referred to.

In order to determine  $A$ , we must require that the  $\phi$  specified by (62) and (63) satisfies condition (58), just as it was earlier shown to satisfy the other conditions on  $\phi$ . By table 1, this requires, for  $|x| < a$ , that

$$\begin{aligned} U^{-1} \int_{-\infty}^{-a} e^{i\omega(x_1-x)/U} \{Aa(x_1+a)^{-1} + Bx_1 + C(x_1^2 - \frac{1}{2}a^2)\} (x_1^2 - a^2)^{-\frac{1}{2}} dx_1 \\ + U^{-1} \int_{-\infty}^x e^{i\omega(x_1-x)/U} (B + Cx_1) dx_1 = i(\omega h - U\alpha) + \omega\alpha(x - b), \end{aligned} \quad (64)$$

where values of  $\partial\Phi/\partial z$  equal to  $\partial\Omega/\partial x$  have been used. The improper integrals are to be interpreted in the generalized-functions sense (Lighthill 1958), or (in other language) in the Abel sense as  $x_1 \rightarrow -\infty$  (equivalent to  $\omega$  having a negative imaginary part that is allowed to tend to zero), and in the Hadamard sense as  $x_1 \rightarrow -a$ . This is because  $\partial\Phi_1/\partial z$  is the  $x$ -derivative of a function  $\Omega$  whose limit as  $z \rightarrow 0$  includes the singular term  $-[(x-a)/(x+a)]^{\frac{1}{2}}$  for  $x < -a$  but not for  $-a < x < a$ , and so the corresponding term in (64) would be correctly evaluated through integration by parts with the ‘infinite’ integrated term replaced by zero.

It is easily verified that the part of (64) written on the second line is separately an identity, as could have been predicted since its left-hand side is the  $(\partial\phi/\partial z)_{z=0}$  that would correspond to (61) if the latter gave  $(\partial\Phi/\partial z)_{z=0}$  for all  $x$ , and this must be (58) by the way that (61) was derived. Hence the first line of (64) vanishes separately, and after cancelling the  $U^{-1}e^{-i\omega x/U}$  factor this gives a single condition to determine  $A$ .

When the integrals have been evaluated as Bessel functions of

$$r = i\sigma, \quad \text{where } \sigma = \omega a/U \quad (65)$$

(and  $r$  should be thought of as having a small positive real part), this condition becomes

$$Ar[K_0(r) + K_1(r)] - BaK_1(r) + \frac{1}{2}Ca^2[K_0(r) + 2r^{-1}K_1(r)] = 0. \quad (66)$$

We show the evaluation only in the (doubly improper) case of the coefficient of  $A$ , which on the substitution  $x_1 = -a \cosh u$  becomes

$$\begin{aligned} - \int_0^\infty e^{-r \cosh u} (\cosh u - 1)^{-1} du &= [e^{-r \cosh u} \coth \frac{1}{2}u]_0^\infty \\ &\quad + r \int_0^\infty e^{-r \cosh u} (1 + \cosh u) du, \end{aligned} \quad (67)$$

an integration by parts in which the rules mentioned above require the integrated term to be replaced by zero, while the remaining term is  $r[K_0(r) + K_1(r)]$ . In terms of the Theodorsen function (Garrick 1957),

$$F(\sigma) + iG(\sigma) = K_1(i\sigma)/[K_0(i\sigma) + K_1(i\sigma)], \quad (68)$$

whose values are displayed in figure 3, equation (66) with (62) and (65) gives

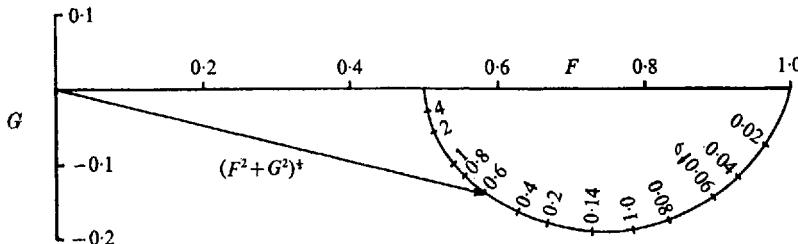
$$A = -U\{[\omega\alpha(b - \frac{1}{2}a) + i(U\alpha - \omega h)](F + iG) + \frac{1}{2}\omega\alpha a\}. \quad (69)$$

The sideforce (in the  $z$ -direction) per unit span,  $\rho Z e^{i\omega t}$  say, and the yawing moment per unit span acting about the half-chord position  $x = z = 0$  in the positive sense (that is, tending to turn the leading edge in the  $z$ -direction),  $Q e^{i\omega t}$  say, are given by (60) and (63) and table 1 as

$$Z = 2 \int_{-a}^a (\Phi)_{z=-0} dx = 2\pi a A + \pi a^2 B, \quad (70)$$

$$Q = 2 \int_{-a}^a (-x) (\Phi)_{z=-0} dx = \pi a^2 A - \frac{1}{8} \pi a^4 C. \quad (71)$$

From these can be deduced the mean rate of working per unit span,  $\rho \bar{E}$  say. We use two principles for this: (i) rate of working in combined sideslip and yawing equals sideforce times rate of sideslip of centroid, plus yawing moment about centroid times rate of yaw; (ii) if two quantities are each expressed as the real part of complex exponentials,  $a e^{i\omega t}$  and  $b e^{i\omega t}$ , their mean product is  $\frac{1}{2} \Re(a\bar{b})$  where  $\Re$  means real part and  $\bar{b}$  the complex conjugate of  $b$ .



**FIGURE 3.** Argand-diagram representation, adapted from Garrick (1957), of the Theodorsen function  $F + iG$  as a function of the frequency parameter  $\sigma = wa/U$  (where  $a = \frac{1}{2}c$  is the half-chord length of the aerofoil).

These principles give

$$\bar{E} = \frac{1}{2}\mathcal{R}[Z\omega(-\alpha b - ih) + Q(-\omega\alpha)], \quad (72)$$

since (57) implies a rate of sideslip of centroid  $\omega(-ab + ih)e^{iat}$  and a rate of yaw  $-\omega x e^{iat}$ . From equations (70), (71) and (72), with (61), we obtain

$$\bar{E} = -\pi a \omega \alpha (b + \frac{1}{2}a) \mathcal{R}A + \pi a \omega h \mathcal{I}A - \pi a^2 U \omega^2 \alpha^2 b, \quad (73)$$

where  $\mathcal{I}$  means imaginary part.

The mean thrust per unit span is  $\rho \bar{P}$ , where

$$\bar{P} = \pi a U^{-2} |A|^2 + (\pi a x \mathcal{J} A - \frac{1}{2} \pi a^2 \omega^2 \alpha^2 b). \quad (74)$$

The term in parentheses here represents the mean resultant  $\frac{1}{2}\rho[Z(-i\alpha)]$  of a sideforce  $\rho Z e^{i\omega t}$  given by (70) acting on a surface (57) inclined backwards at an angle  $i\alpha e^{i\omega t}$  to the  $x$ -axis. An important additional thrust, however, is derived from the first term, representing the mean suction force acting on the rounded leading edge due to the fast flow around it. Instantaneously, the suction force takes the well-known steady-flow value obtained by Blasius's theorem (see, for example, Robinson & Laurmann 1956, p. 126). This is associated only with the quadratic terms  $-\frac{1}{2}\rho[(\partial\phi/\partial x)^2 + (\partial\phi/\partial z)^2]$  in the pressure (because the linear

term (60), being an odd function of  $z$ , has zero resultant), and the mean suction force is  $\frac{1}{2}\pi|K|^2$ , where  $K(x+a)^{-\frac{1}{2}}$  is the asymptotic form as  $x \downarrow -a$  of  $(\partial\phi/\partial x)_{z=+\infty}$ , which by (59) is that of  $U^{-1}\Phi$  (since  $\phi$  vanishes at  $x = -a$ ,  $z = 0$ ). Equation (74) follows since this  $K$ , by (63) and table 1, is  $-U^{-1}A(2a)^{\frac{1}{2}}$ .

The rate of energy wastage per unit span,  $\rho(\bar{E} - U\bar{P})$ , is deduced from (73) and (74) by simple algebra, using (69), as

$$\bar{E} - U\bar{P} = (U\pi a)[\omega^2\alpha^2(b - \frac{1}{2}a)^2 + (wh - U\alpha)^2](F - F^2 - G^2). \quad (75)$$

In this important formula all three factors are essentially positive, and the middle factor shows a clear minimum as a function of  $b$  when the yawing axis  $x = b$  is at the three-quarter-chord point  $b = \frac{3}{4}a$ .

More arduously, the formula (75) can be derived by showing that in the region of the vortex wake  $x > a$  equations (62) and (63), together with a variety of Bessel-function identities, imply that  $(\phi)_{z=\pm 0} = \mp\phi_w e^{-i\omega x/U}$ , where

$$\phi_w = \frac{\pi U[wh - U\alpha + i\omega\alpha(b - \frac{1}{2}a)]}{\omega[K_0(i\sigma) + K_1(i\sigma)]}. \quad (76)$$

The parts of such a vortex wake far beyond the influence of the solid body  $|x| < a$  have mean energy  $\frac{1}{2}\rho(\omega/U)|\phi_w|^2$  per unit length, and such energy increases at a rate  $\frac{1}{2}\rho\omega|\phi_w|^2$ . Agreement of this rate of energy wastage with (75) is then proved by another Bessel-function identity,

$$|K_0(i\sigma) + K_1(i\sigma)|^{-2} = (2\omega a/\pi U)(F - F^2 - G^2). \quad (77)$$

This much more complicated method is only useful as a check.

The efficiency  $\eta$  is still given by equation (1), although in this section both  $\bar{E}$  and  $\bar{P}$  are values per unit span, divided by the density. Hence, by (69), (73) and (75),

$$\begin{aligned} 1 - \eta &= (\bar{E} - U\bar{P})/\bar{E} \\ &= \frac{\{\omega^2\alpha^2(b - \frac{1}{2}a)^2 + (wh - U\alpha)^2\}(F - F^2 - G^2)}{\{\omega\alpha(b - \frac{1}{2}a)[\alpha(b + \frac{1}{2}a)F - hG - \frac{1}{2}\alpha a] + (wh - U\alpha)[hF + \alpha(b + \frac{1}{2}a)G]\}\omega}. \end{aligned} \quad (78)$$

The numerator of (78), proportional to the rate of energy wastage, shows the clear minimum as a function of  $b$  where  $b = \frac{3}{4}a$  already referred to. This minimum is particularly sharp for the higher values of both the frequency parameter  $\sigma = \omega a/U$  and the proportional-feathering parameter  $U\alpha/wh$ , which Lighthill (1969) called  $\theta$  (and which must be expected to be less than 1 for significant positive thrust).

The corresponding maximum in  $\eta$  is shifted to a value of  $b$  only slightly greater than  $\frac{1}{2}a$  by the behaviour of the denominator. There, the terms in square brackets that form the coefficients of  $\omega\alpha(b - \frac{1}{2}a)$  and  $(wh - U\alpha)$  are both positive near  $b = \frac{1}{2}a$ , so that total work done increases as  $(b - \frac{1}{2}a)$  increases. However, the first of these coefficients is considerably the smaller in most cases of interest and therefore, when the numerator has a sharp minimum, does not allow the associated maximum in  $\eta$  to depart far beyond  $b = \frac{1}{2}a$ . On the other hand, it means that thrust is somewhat increased for a value of  $b$  greater than  $\frac{1}{2}a$ , and this consideration may need to be balanced against considerations of efficiency.

These remarks are borne out by figure 4, which compares the dependence of thrust and efficiency on  $\theta$  and  $\sigma$ , calculated by this two-dimensional theory for four positions of the yawing axis (half-chord, three-quarter chord, full-chord or trailing edge, and five-quarters-chord). Of these four positions, the middle two permit better maintenance of good efficiency at the higher values of the frequency parameter  $\omega c/U = 2\sigma$  based on the 'chord'  $c = 2a$  (fore-and-aft dimension of fin section).

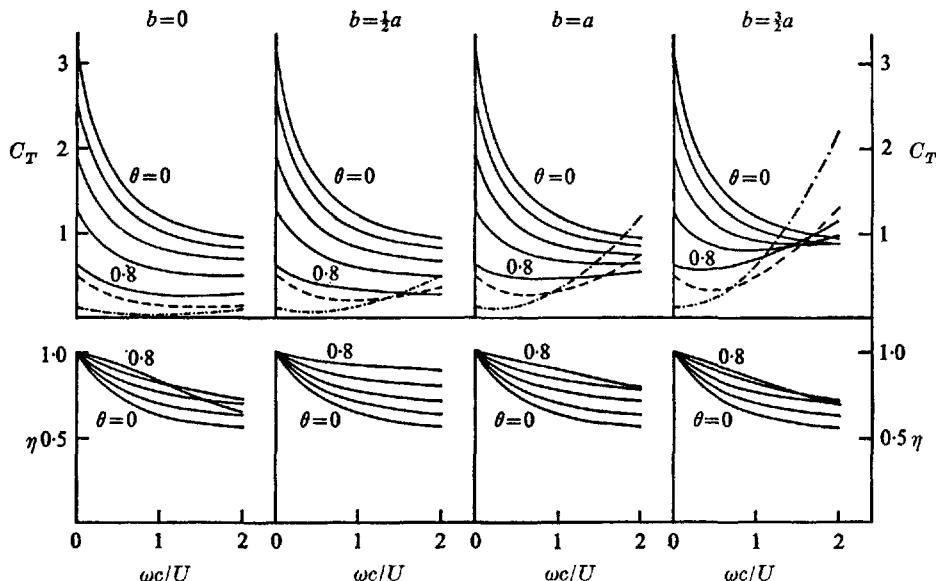


FIGURE 4. Thrust coefficient  $C_T$  and efficiency  $\eta$  predicted by two-dimensional aerofoil theory, for values 0, 0.2, 0.4, 0.6 and 0.8 of a feathering parameter  $\theta = U\alpha/\omega h$ , plotted as a function of  $\omega c/U$  (which is  $2\sigma$ ), for different positions  $x = b$  of the yaw axis (namely half-chord  $b = 0$ , three-quarter-chord  $b = \frac{1}{2}a$ , trailing-edge  $b = a$ , and a position 'five-quarters-chord',  $b = \frac{3}{4}a$ , beyond the trailing edge). ——, part of  $C_T$  for  $\theta = 0.6$  predicted as coming from leading-edge suction. - - - - -, same for  $\theta = 0.8$ .

The thrust is represented by a thrust coefficient  $C_T$  equal to thrust per unit fin area divided by  $\frac{1}{2}\rho(\omega h)^2$ ; thus,  $C_T = \bar{P}/\omega^2 h^2 a$ . (79)

This relates thrust to the amplitude  $\omega h$  of the fin's lateral motion, and the reduction in  $C_T$  as  $\omega c/U$  increases describes the decline in thrust from given fin movements at the lower forward velocities.

For the larger values of  $\theta$  (say, 0.6 and 0.8) which are best for maintaining good efficiency, the thrust values are greater for the positions of the yawing axis which are farther downstream, and such positions might be tentatively preferred (for example,  $b = a$  preferred to  $b = \frac{1}{2}a$ ) as maintaining thrust better at the lower forward velocities. Some caution is needed here, however. For  $\theta = 0.6$ , the broken line shows which part of the thrust comes from leading-edge suction, the remainder being due to the backwardly inclined component of side-force. The chain-dotted line does the same† for  $\theta = 0.8$ . (Note that for  $\theta = 0$ , by

† In the case of the  $b = 0$  diagrams, these are the only changes from figure 9 of Lighthill (1969).

contrast, there is no inclining of the fin and *all* the thrust predicted is due to leading-edge suction.)

Evidently, the increase of thrust as  $b$  increases, calculated for the larger values of  $\theta$  and of  $wc/U$ , is entirely due to a steep increase in the leading-edge-suction component, whereas the element of backwardly inclined sideforce falls off, and may even become negative. When we question how it can become negative, we soon find that it is due to the second term in parenthesis in (74), which arises from the sideforce (70) acquiring an imaginary part from the  $-i\omega^2 ab$  term in equation (61) for  $B$ . This can be thought of as a virtual-mass contribution to sideforce, due to rate of change of lateral momentum of water adjacent to the fin and swept from side to side by it, so fluid inertance plays a substantial role here as in elongated-body theory. The farther the yawing axis is behind the aerofoil centroid, the more this contribution to sideforce is negatively correlated with backward inclination of fin.

This tendency towards a reduced thrust contribution from backwardly inclined sideforce as  $b$  increases is more than made up, however, by the increased suction force, that is, the term in (74) proportional to  $|A|^2$ . It is the growth in the real part of  $A$  (see (69)) as  $b$  increases, associated with fluctuations in sideforce (70) out of phase with the fluctuations in angle of incidence, which increases this suction force.

These considerations indicate that attempts to achieve high thrust coefficients  $C_T$  from a two-dimensional aerofoil by choice of a yawing axis behind the trailing edge (such as the  $b = \frac{3}{2}a$  choice in figure 4) are probably inadvisable quite apart from considerations of efficiency. They depend too much on realizing a very high suction force at the rounded leading edge, and if owing to separation only part of that force were realized the total thrust (including a negative component from backwardly inclined sideforce) would be very greatly reduced. Probably an optimum from thrust considerations as well as from efficiency considerations lies somewhere between  $b = \frac{1}{2}a$  and  $b = a$ , which in practice means very close to the trailing edge.

This conclusion is, perhaps, relevant to why the lunate tail should be hydro-mechanically efficient. If a caudal fin were yawing as a whole about a single axis, with yaw angle in phase with its velocity of lateral translation, and if two-dimensional theory could be applied to each section, then good thrust with good efficiency would best be achieved if the axis of yaw were close to the trailing edge of each section. This requires that the trailing edge as a whole stretch almost straight along the axis of yaw.

Tapering of the fin must accordingly take place through the leading edge being bowed forward. Any bowing of the trailing edge should be small by comparison. We recover here the concept of a vertical trailing edge where angle of yaw and lateral velocity are perfectly in phase, which was a favoured idealization also in § 2.

This degree of departure from 'straight wing' conditions suggested as optimum by two-dimensional theory, with leading edge bowed forward but trailing edge practically straight, is in the right direction (see figure 6 of Lighthill 1969), although it does not go far enough; possibly a fully three-dimensional theory

might explain the fact that most lunate tails have also a trailing edge bowed forward (though not as much as the leading edge). In any case a properly worked-out two-dimensional theory is necessary as a preliminary to a more thorough three-dimensional study, using lifting-line techniques, of the advantages of a lunate-tail configuration.

Further discussion of three-dimensional aspects of the flow is not here attempted, but we conclude with some remarks about the appropriateness within two-dimensional theory of the linearization of boundary conditions, used in this section as well as in the earlier sections concerned with elongated-body theory. Extension of elongated-body theory to large-amplitude motions seems feasible, as mentioned in § 1, but the problems of carrying out such an extension for even a two-dimensional theory of the lunate tail appear far from straightforward.

This is not to say that linearized theory is useless for evaluating the large-amplitude lunate-tail motions that are actually found (Fierstine & Walters 1968). Aerofoil characteristics at amplitudes relatively large, but below those where catastrophic stalling occurs (which fishes may be presumed to avoid), are often indicated reasonably well by extrapolation of small-amplitude characteristics. Nevertheless, one would like both to understand any tendency for thrust and efficiency to continue to follow curves predicted by linearized theory, and to estimate deviations from them.

A theory might be attempted for well-feathered oscillations of a two-dimensional aerofoil, making large displacements but combining them with such variations in angle of incidence that disturbances to the uniform stream of velocity  $U$  could still be regarded as small. Then a pressure

$$p - p_0 = -\rho(\partial\phi/\partial t + U\partial\phi/\partial x) \quad (80)$$

linearly related to velocity potential (as in (59) and (60) but now with  $\phi$  not  $\phi e^{i\omega t}$  as velocity potential), could still be used, and  $p$  would satisfy Laplace's equation and be continuous even across the vortex wake. The normal derivative of  $p$  at all points of the aerofoil surface would be known at each instant as the rate of change following a fluid particle of the normal velocity specified by the boundary conditions. For a flat-plate aerofoil,  $p$  could then be deduced in a form constituting a generalization of (63), and involving one arbitrary constant  $A$  that now would be a general function of time. This function would have to be determined by specifying that the solution of equation (80) for  $\phi$ , got by integrating along lines on which  $y$ ,  $z$  and  $x - Ut$  are constant, satisfied the aerofoil boundary condition.

That last step, leading to the generalization of equation (66), would involve extremely complicated integrations, but the whole would be easier, perhaps, than a method which sought to calculate the velocity field induced by a sinusoidal vortex wake with amplitude approaching half a wavelength (Fierstine & Walters 1968). Methods of this latter kind cannot be ruled out, however, in any future efforts at achieving a three-dimensional non-linear theory, possibly using the idea (Lighthill 1969) that the lunate tail may generate a wake consisting of a succession of vortex rings. That ultimate development, however, would call for

something like the genius that Goldstein (1929) showed when he calculated the effect of the helicoidal vortex wake behind a propeller.

The author acknowledges with gratitude and pleasure the benefits of having sat for seven years at the feet of Sydney Goldstein.

#### REFERENCES

- FIERSTINE, H. L. & WALTERS, V. 1968 *Mem. S. Calif. Acad. Sci.* **6**, 1.  
FLACHSBART, O. 1935 *Z. angew. Math. Mech.* **15**, 32.  
GARRICK, I. E. 1957 In *Aerodynamic Components of Aircraft at High Speeds* (ed. A. F. Donovan & H. R. Lawrence), pp. 658–793. Princeton University Press.  
GOLDSTEIN, S. 1929 *Proc. Roy. Soc. A* **123**, 440.  
GRAY, J. & HANCOCK, G. J. 1955 *J. Exp. Biol.* **41**, 135.  
KRAMER, E. 1960 *Z. Wiss. Zool.* **163**, 1.  
Lighthill, M. J. 1951 *Aeron. Quart.* **3**, 193.  
Lighthill, M. J. 1958 *Fourier Analysis and Generalized Functions*. Cambridge University Press.  
Lighthill, M. J. 1960 *J. Fluid Mech.* **9**, 305.  
Lighthill, M. J. 1969 *Ann. Rev. Fluid Mech.* **1**, 413.  
POSSIO, C. 1938 *Aerotecnica*, **18**, 441.  
ROBINSON, A. & LAURMANN, J. A. 1956 *Wing Theory*. Cambridge University Press.  
TAYLOR, G. I. 1952 *Proc. Roy. Soc. A* **214**, 158.  
WU, T. Y. 1961 *J. Fluid Mech.* **10**, 321.

*This page intentionally left blank*

## CHAPTER 5

# Large-Amplitude Elongated-Body Theory of Fish Locomotion\*

### 1. INTRODUCTION

The undulatory motions that most elongated animals make to propel themselves through water have been investigated theoretically, with the aim of estimating their energy cost and the forces between the animal and the water which they generate, by two main methods. The older theory may be called 'resistive' in that the force between a small section of the animal and the water was regarded as a resistive force depending exclusively, though not necessarily linearly, on the instantaneous value of the velocity of that section relative to the water. An excellent feature of the more advanced papers (Taylor 1952; Hancock 1953; Gray & Hancock 1955) which expound the resistive theory is that the methods are so developed as to be applicable to undulatory motions of arbitrarily large amplitude.

A newer theory (Lighthill 1960) may be called 'reactive' in that it lays principal emphasis on reactive forces between a small volume of water and the parts of the animal's surface in contact with it. These forces, due to the inertia of the water and proportional to rate of change of the relative velocity of animal surface which a given volume of water feels, are neglected on the resistive theory. They can be particularly important when the cross-section of the animal is much thinner in the direction of the bodily displacements that it makes for swimming purposes than in a perpendicular direction; then, the 'virtual mass' of water which acquires momentum through such displacements far exceeds the associated animal mass.

A survey of aquatic animal propulsion (Lighthill 1969) indicated that the undulatory motions associated with elongated invertebrates are best studied by the resistive theory. This is obvious in the many cases when the Reynolds number is low (meaning that water inertia is relatively unimportant), but appears probable even for several cases of swimming, with the Reynolds number around  $10^3$ , in those worms whose cross-sectional form is not such as to enhance virtual-mass effects.

With the vertebrates, however, posterior lateral compression of body cross-

---

\* Reprinted from Proc. Roy. Soc. B., 179 (1971), pp. 125-138.

section appears. This has the advantage that it allows significant reactive forces on body cross-sections to contribute to propulsive effort, with less energy wasted in generating a vortex wake than in the case of resistive forces. In fishes, the undulatory mode of propulsion is called anguilliform, after the eel *Anguilla*, a good example of an animal which needs a propulsive method with low energy wastage because its life-cycle requires it to traverse huge distances.

Anguilliform propulsion of elongated fishes probably uses a combination of resistive and reactive forces (Lighthill 1970). The other main type of propulsive movement capable of producing fast forward motion in fishes seems, however, to have developed so as to make a more exclusive use of reactive forces. In this movement, known as carangiform, the front half of the body has lost its flexibility and flexural movements are confined to the rear half, or even the rear one-third, of the body length.

The characteristic wavelike feature, that the phase of lateral oscillations of any posterior section lags behind that of any anterior section, is retained in carangiform motion, which probably developed from the more obviously wavy anguilliform motion. In other respects, however, carangiform motion is very different: wave amplitude (which in anguilliform motion increases gradually from head to tail) shows such a steep increase towards the tail in the last one-third of the body length that the observer of carangiform motion cannot see anything like a whole wavelength at any one time. He is much more conscious, simply, of lateral tail oscillations with the posterior end lagging behind anterior sections.

The reactive forces dominate in carangiform motion because the acceleration of water passed by the fish takes place very fast, during the time required for a short posterior portion to pass and flick it into motion through the instantaneously acting virtual-mass effect. By contrast, little time is available for build-up of resistive forces through the usual process of vortex shedding in the cross flow (typically, the water feels a total lateral displacement of fish surface against it of only half a cross-sectional depth).

Lighthill (1970), in a paper constituting a sort of mathematical appendix to his 1969 survey, showed that these features give an advantage over anguilliform motion with respect to efficiency, provided that carangiform motion is combined (as is found in practice) with certain morphological features, including especially a pronounced reduction in depth of body cross-section just anterior to the caudal fin, whose effect is to minimize 'recoil' motions of yaw and sideslip due to unbalanced oscillations of sideforce. Although Lighthill (1960) had assumed that such recoil would result in too much loss of efficiency unless the swimming motions included practically a whole wavelength at any one time, Lighthill (1970) calculated that this local depth reduction makes the oscillations of sideforce much smaller, while maintenance of substantial depth of cross-section for a lengthy extent of fish around its mass centre reduces what yaw and sideslip does result by involving a large mass of water in any such motions.

The elongated-body theory that goes back to Lighthill (1960), and is based on

ideas from what aerodynamicists call 'slender-body theory', is the reactive theory appropriate to the great majority of fishes using carangiform propulsion: namely, those with caudal fins that are 'slender' in the sense that each individual bony ray of which the fin is made up makes only a moderately small angle (not more than say 30°) with the backbone. A different reactive theory is appropriate to that important minority which have acquired even greater speed by abandoning slender caudal fins in favour of the so-called 'lunate' or crescent-moon-shaped tails of high aspect-ratio. The right theory for them must be one based on wing theory: for example, that developed by Wu (1961) and others, and applied by Lighthill (1969, 1970) to the fishes with lunate tails as well as to the cetacean mammals. In this paper, however, those problems are put aside in order to concentrate upon the possible improvement of elongated-body theory.

The object is to extend reactive theory so that it can be applied to interpret observational data on animal movements of arbitrary large amplitude (as was earlier mentioned to have been achieved already with resistive theory). The existing theory is based on a perturbation expansion in powers of an amplitude parameter  $\epsilon$  (Lighthill 1960, Appendix) and derives thrust and energy consumption as multiples of  $\epsilon^2$  with terms of order  $\epsilon^4$  neglected. The inadequacy of such an expansion in powers of amplitude can be seen from the fact that in carangiform motion the component of caudal fin velocity at right angles to the path may rise to twice the forward velocity of the fish. A theory which takes into account the large fish flexures and lateral velocities used in propulsion is needed. Such a theory should have the additional advantage that it could be applied to study interesting questions like the mechanism of turning in fishes of the type here investigated (although no work on this application is included in the present paper).

Clues on how to obtain a large-amplitude theory may be suggested by studying the conclusions of the rather complicated perturbation theory and finding ways of interpreting them that suggest a physical basis on which a large-amplitude theory might be founded. Lighthill (1960, Section 2) gave one physical interpretation of the perturbation-theory results, in terms of rate of working and rate of energy wastage, that was further developed by Lighthill (1969, 1970), but that interpretation is not easy to extend to large-amplitude motions.

Lighthill (1970), however, finally obtained an alternative, quite different physical interpretation of the perturbation-theory results, directly in terms of the force of propulsion (see his equations (13) to (15)). It is expressed most satisfactorily in a frame of reference in which the water far from the fish is at rest, and makes use of three principles:

(i) Water momentum near a section of fish is in a direction perpendicular to the backbone and has magnitude equal to the virtual mass,  $m$  per unit length, times the component  $w$  of fish velocity in that direction.

(ii) Thrust can be obtained by considering rate of change of momentum within a volume enclosing the fish whose boundary at each instant includes a flat surface  $\Pi$  perpendicular to the caudal fin through its posterior end.

(iii) In the momentum balance it is necessary to take into account transfer of momentum across  $\Pi$  not only by convection but also by the action of the resultant  $\frac{1}{2}mw^2$  of the pressures generated by the motions within the plane  $\Pi$ .

Fortunately, principles (i) to (iii) are in a form suitable for direct use in making a reactive theory valid for arbitrary large amplitudes of motion. This statement is supported by detailed argument in §2, but one important aspect may be noted at the onset: even though individual sections of fish have large lateral velocity components, even in excess of the forward velocity, the velocity component  $w$  perpendicular to the backbone is usually much smaller (as, indeed, is required for good efficiency). We should note, furthermore, that the estimation of recoil, an important matter for carangiform propulsion, is facilitated by the new formalism, which, finally, is also useful for studying how far the detailed motions made in carangiform propulsion have the effect of reducing variations of the thrust with time about its mean value.

It goes without saying that it would be desirable to progress ultimately to a combined resistive-reactive theory for motions of large amplitude, which might be applied to elongated animals in general. That would be an extremely complicated task, however, and it is therefore natural to make a first effort, as here, on a purely reactive theory, and to relate it to the appropriate field of application: namely, the motion of fishes with slender caudal fins which use the carangiform mode of propulsion.

## 2. CALCULATION OF THE REACTIVE FORCE

The theory, as already remarked, is set out in a frame of reference in which the water far from the swimming fish is at rest. The fish is supposed to be swimming at a fixed horizontal level. The  $y$ -axis is vertical, and the  $x$ - and  $z$ -axes are horizontal axes at right angles whose directions may be chosen arbitrarily, although when there is a constant *mean* direction of swimming that might suitably be taken as that of the positive  $x$ -axis.

To describe large-amplitude motions of the fish's spinal column in a horizontal plane  $y = 0$ , we use a 'Lagrangian' coordinate  $a$ , which takes values from 0 to  $l$  (the length of the undistorted fish), and identifies a point on the spinal column by its distance from the posterior end of the fish (tip of the caudal fin) measured along the spinal column, whose extensibility is small and will here be neglected. The position coordinates  $(x, z)$  of that point at time  $t$  within the horizontal plane  $y = 0$  are regarded as functions  $x(a, t)$  and  $z(a, t)$  of  $a$  and  $t$ , as the point moves due (i) to swimming actions, (ii) to the resulting forward motion of the fish, and (iii) to any lateral recoil motions that may be involuntarily produced (figure 1).

The assumed inextensibility of the spinal column can be expressed mathematically by the equation

$$\left(\frac{\partial x}{\partial a}\right)^2 + \left(\frac{\partial z}{\partial a}\right)^2 = 1. \quad (1)$$

The horizontal velocity vector ( $\partial x/\partial t$ ,  $\partial z/\partial t$ ) then has a component

$$u = \frac{\partial x}{\partial t} \frac{\partial x}{\partial a} + \frac{\partial z}{\partial t} \frac{\partial z}{\partial a}, \quad (2)$$

directed tangentially to the spinal column in the forward sense, and a component

$$w = \frac{\partial z}{\partial t} \frac{\partial x}{\partial a} - \frac{\partial x}{\partial t} \frac{\partial z}{\partial a}, \quad (3)$$

directed perpendicularly to the spinal column, in a sense so that  $(u, w)$  make a left-handed system of axes if  $(x, z)$  do (figure 2).

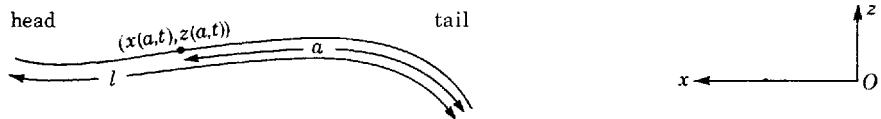


FIGURE 1. The coordinate system used to describe the spinal column's configuration in the plane  $y = 0$  at time  $t$ .

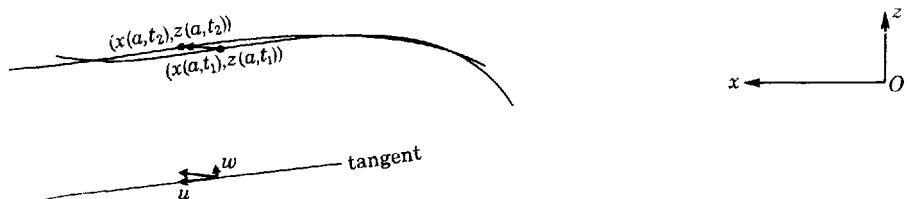


FIGURE 2. The spinal column's configuration at two successive instants  $t_1$  and  $t_2$ . The arrow indicating the displacement of a particular point on the spinal column between the two instants represents the velocity vector of that point multiplied by the time separation  $t_2 - t_1$ . The lower diagram shows this velocity vector resolved into components  $u$  tangential to the spinal column and  $w$  perpendicular to it.



FIGURE 3. Motion of the tip of the caudal fin. The upper diagram shows its position at two successive instants and defines the angles  $\theta$  and  $\alpha$ . The lower diagram shows how the tip's velocity vector, of magnitude  $V$ , may be resolved either into components  $u$  tangential and  $w$  perpendicular to the tip, just as in figure 2, or alternatively into a component  $W$  perpendicular to the direction of mean motion (see equation (10) below).

The essential property characteristic of an 'elongated body' is that the virtual mass,  $m$  per unit length, in respect of the  $w$  motions (perpendicular to the backbone) is large, whereas the  $u$  motions (tangential to the backbone) have negligible virtual mass. Those tangential motions, to be sure, are what actually contribute the viscous resistance that will have to be considered as balanced by the reactively produced thrust; one way to put this is to say that tangential motions will be treated resistively and perpendicular motions reactively. We begin that process in this Section, however, by simply determining the reactive force associated, on purely inviscid considerations, with the momentum of the  $w$  motions.

This momentum per unit length of fish is represented by the vector

$$mw(-\partial z/\partial a, \partial x/\partial a), \quad (4)$$

where the factor in parentheses is a unit vector in the  $w$ -direction (expressed in terms of its  $x$ - and  $z$ -components). Lighthill (1970) gives a full discussion<sup>†</sup> of the appropriateness of regarding  $m$  for a given cross-section of fish as a constant close to  $\frac{1}{4}\pi\rho s^2$ , where  $\rho$  is the water density, and  $s$  the depth of the cross-section; in the present notation this would normally make  $m$  a fixed, known function  $m(a)$  of the distance  $a$  from the posterior end of the fish. However, in cases like those described by Bainbridge (1963) where fishes with flexible caudal fins cause their depth to vary with time  $t$  during the swimming cycle,  $m$  needs to be written as  $m(a, t)$ , and we shall for generality's sake assume this in what follows.

It might be thought straightforward to use the momentum distribution (4) to infer the reactive force between fish and water. For this purpose, however, Lighthill (1970) showed that any method based on the rate of change of momentum of the whole water mass encounters a serious difficulty: that of calculating the rate of change of one element of that momentum, namely, the momentum in the wake (that consists of vorticity shed from the caudal fin's trailing edge). The approximate method for doing this given in equation (18) of that paper would be hard to extend to large-amplitude motions.

A much more satisfactory approach is based on a study of the rate of change of momentum in a certain part only of the fluid; a part  $V$ , that excludes the wake. To give a definition of  $V$ , at each moment of time, in a manner satisfactory for large-amplitude motions, we first imagine a vertical plane  $\Pi$  intersecting the tip of the caudal fin at right angles to the tip of the spinal column; this plane has to be thought of as swinging around as the fin moves. Then at each instant the part  $V$  of water whose momentum is considered is that half-space, bounded by the plane  $\Pi$ , which includes the fish and excludes the wake.

<sup>†</sup> Note that this includes discussion of the influence of cross-section non-uniformity. When that non-uniformity can be analysed into sine-wave components, the error is less than 25% for all components with wavelengths greater than  $5s$ . It may also be noted from aerodynamic slender-wing theory that the overall forces at small angle of incidence on a 'delta' shape (an isosceles triangle with apex forwards, such as one might indeed use to represent a caudal fin) are given to within 25% by using the above value of  $m$  if the total included angle at the apex is less than  $60^\circ$ .

The momentum† in  $V$  can be written as the integral of expression (4) from 0 to  $l$  with respect to  $a$ , and its rate of change can be written as the sum of three terms: (i) rate of change due to convection of momentum out of  $V$  across the plane  $\Pi$ ; (ii) rate of change due to the pressure force acting across  $\Pi$ ; (iii) minus the reactive force ( $P, Q$ ) with which the fluid acts on the fish.

To estimate (i), we note that convection across  $\Pi$  is produced to only a negligible extent by any motion of fluid in the  $u$ -direction perpendicular to  $\Pi$ , since only the  $w$ -motion of the fin produces significant fluid motion. Convection arises, however, from the motion of  $\Pi$  itself; this consists of a forward translation at the  $u$ -velocity (2) and a rotation; however, since fish cross-sections are laterally symmetric the water-momentum distribution has lateral symmetry and therefore no net rate of change of momentum within  $V$  results from the rotation of  $\Pi$ .

Lighthill (1970, equations (14) and (15)) gave a simple argument, directly applicable also to large-amplitude motions, proving that the pressure force in (ii), that is, the resultant over  $\Pi$  of the pressures associated with the  $w$ -motion of a laterally symmetric section with virtual mass  $m$  per unit length, is  $\frac{1}{2}mw^2$ . It acts in the tangential direction represented by the unit vector  $(\partial x/\partial a, \partial z/\partial a)$ . Using this with the results of the last two paragraphs, we obtain

$$\frac{d}{dt} \int_0^l mw \left( -\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a} \right) da = \left[ -umw \left( -\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a} \right) + \frac{1}{2}mw^2 \left( \frac{\partial x}{\partial a}, \frac{\partial z}{\partial a} \right) \right]_{a=0} - (P, Q), \quad (5)$$

where the terms on the right-hand side are written down in order as in (i), (ii) and (iii) above.

Inspection of the square bracket in (5) shows that it can be usefully simplified. In fact, if the  $\frac{1}{2}$  were absent, then equations (1) to (3) show that it would reduce to  $mw(\partial z/\partial t, -\partial x/\partial t)$ , as can easily be verified if  $mw$  is taken out as a factor. It follows that we can rewrite the actual equation (5) as

$$(P, Q) = \left[ mw \left( \frac{\partial z}{\partial t}, -\frac{\partial x}{\partial t} \right) - \frac{1}{2}mw^2 \left( \frac{\partial x}{\partial a}, \frac{\partial z}{\partial a} \right) \right]_{a=0} - \frac{d}{dt} \int_0^l mw \left( -\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a} \right) da. \quad (6)$$

Equation (5) and this simpler form of it (6) are, respectively, generalizations of equations (13) and (11) of Lighthill (1970), that extend them so as to give the force in vector form for the case of large-amplitude motions.

Note that, for periodic swimming movements, producing a mean motion in (say) the positive  $x$ -direction, there is no contribution to the mean thrust, that is, to  $\bar{P}$  (where the bar denotes a mean), from the integral term in (6), because its value fluctuates periodically between fixed limits, so that the time mean of its rate of change is zero. The mean thrust can therefore be written

$$\bar{P} = \overline{[mw(\partial z/\partial t - \frac{1}{2}w \partial x/\partial a)]}_{a=0}. \quad (7)$$

This suggests the same conclusion as was obtained in the small-amplitude case, that good thrust can be obtained if  $w$  and  $\partial z/\partial t$  have high positive correlation but

† More strictly, the momentum as calculated from the point of view of a reactive theory, and neglecting therefore any influence of the viscous boundary layer at the fish surface.

the latter has a substantially larger amplitude of oscillation. The requirement for  $w$  to have relatively smaller amplitude is reduced however if fluctuations of  $w^2$  are negatively correlated with those of  $\partial x/\partial a$  at  $a = 0$ , as characteristically occurs in carangiform motion, since  $\partial x/\partial a$  is the cosine of the angle between caudal fin and mean swimming direction, an angle which tends to be greatest when  $w$  is large.

It is noteworthy, moreover, that the  $x$ -component of equation (6) gives not only the mean thrust  $\bar{P}$  but also instantaneous values of  $P$ . Although the term in square brackets may have a substantial positive mean value, it falls to zero twice in each swimming cycle (when  $w$  passes through zero), so that the associated thrust is distinctly intermittent. Fortunately, the carangiform type of movement is such that fluctuations of the second term

$$\frac{d}{dt} \int_0^l mw \frac{\partial z}{\partial a} da \quad (8)$$

in  $P$  are often nearly opposite in phase, and help to smooth out very considerably the fluctuations of  $P$ .

To see this, note first that carangiform motion involves a rapid falling away in the amplitude of oscillation of  $z(a, t)$  as well as of  $mw$  as  $a$  increases from its tip value  $a = 0$ . The integral (8) might be approximated, crudely, therefore, as a product of the amount  $[-z]_{a=0}$ , by which  $z$  changes between the tip  $a = 0$  and an anterior section, and of a crude estimate  $[\frac{1}{2}mw]_{a=0}$  of an average of  $mw$  in that region. This admittedly rough approximation to the amplitude, and underestimate of the phase of the latter average, produces exact cancelling of the fluctuations of thrust if sinusoidal variations of  $z$  and  $mw$  at  $a = 0$ , say

$$z = z_0 \sin \omega t, \quad mw = m_0 w_0 \cos \omega t, \quad (9)$$

are assumed. For then (6) would give

$$P = m_0 w_0 z_0 \omega \cos^2 \omega t + \frac{d}{dt} (-\frac{1}{2} m_0 w_0 z_0 \sin \omega t \cos \omega t) = \frac{1}{2} m_0 w_0 z_0 \omega. \quad (10)$$

It is probable that even in practice some such smoothing of the fluctuations, although not so complete, does occur.

In addition, (6) is important as giving the fluctuations in the sideforce ( $-Q$ ) acting on the fish, which are needed for estimating recoil. Actually, the change in conclusions from those of small-amplitude theory is still more moderate in the case of the sideforce, which (6) gives as

$$-Q = \left[ mw \frac{\partial x}{\partial t} + \frac{1}{2} mw^2 \frac{\partial z}{\partial a} \right]_{a=0} + \frac{d}{dt} \int_0^l mw \frac{\partial x}{\partial a} da. \quad (11)$$

The differences from the expression (49) in Lighthill (1970) are (i) the mean forward velocity  $U$  is replaced by the fluctuating forward velocity  $\partial x/\partial t$  of the tip of the caudal fin; (ii) a third-order term  $\frac{1}{2}mw^2 \partial z/\partial a$  of very modest magnitude is included; (iii) the integral is modified by the factor  $\partial x/\partial a$ .

None of these differences is large, and the general conclusions of Lighthill (1970), that recoil is minimized by a shape which maintains a good depth of cross-section in the anterior portion of the fish and exhibits a substantial reduction in depth immediately anterior to the caudal fin can be drawn as before. To see in the present frame of reference that the latter feature is needed to minimize sideforce, note that where the amplitude of oscillation of  $z(a, t)$  is rapidly changing with  $a$ , the second term in (3) must have a large value not offset by a correspondingly large value for the first term. Fluctuations in the integral in (11) would therefore be large unless  $m$  were greatly reduced in the same region.

Finally, we may note how the method used for calculation of the reactive force can be adapted to determine the rate at which the swimming movements which produce that force generate waste energy in the wake. On inviscid theory the kinetic energy associated with  $w$ -motions is  $\frac{1}{2}mw^2$  per unit length, and the volume  $V$  loses such energy only through convection across the plane  $\Pi$  due to its forward translation at the  $u$ -velocity (2). (Rotation of  $\Pi$  again produces no net rate of change because the kinetic-energy distribution has lateral symmetry, and the pressure distribution over  $\Pi$  conveys negligible energy across it because velocity components perpendicular to  $\Pi$  are negligible.) The mean rate of energy wastage, therefore, is

$$\overline{[\frac{1}{2}mw^2u]}_{a=0}. \quad (8)$$

This predicted rate of energy wastage is greater than the value given by the small-amplitude approximation, on which  $u$  is replaced by the mean swimming speed

$$U = \overline{[u \partial x / \partial a]}_{a=0}, \quad (9)$$

with a resulting reduction in (8) since  $\partial x / \partial a < 1$ .

### 3. DISCUSSION OF THE THRUST-DRAG BALANCE IN RELATION TO OBSERVATIONS

In this paper we make only a first attempt to use the theory of §2 in the analysis of experimental data: an attempt confined to the problem of mean thrust and its balance against mean drag in motions with constant mean speed. Within the limitations of the reactive theory of thrust production, this requires us to study, for actual carangiform movements of fishes with slender caudal fins, the balance between the mean thrust, produced reactively by their  $w$  motions (perpendicular to the spinal column), and the mean drag, produced resistively by their  $u$  motions (tangential to the spinal column).

We shall see that not nearly enough data exist to make any accurate balance possible, but that even an approximate balance suggests conclusions perhaps interesting enough to stimulate experiments aimed at obtaining more complete data. Ideally one would like, for horizontal swimming in water known to be still, pictures taken by a vertical cine camera, at not less than 16 frames per swimming

cycle, of the complete cycle of swimming movements associated with each of a range of forward speeds in the same fish. To compare with the drag estimated from its measured  $u$  motions, it would be desirable to be able to infer also the drag incurred by the same fish when not making swimming movements, e.g. from motion pictures of its gliding deceleration under such conditions.

Available data are not nearly so complete for any single fish. Bainbridge (1958, 1960, 1963) obtained controlled water conditions in his 'fish wheel' (annular fish tank with antiswirling gate), and we analyse below his excellent motion pictures of a dace *Leuciscus* of 0.3 m length swimming at 0.48 m/s. Unfortunately this was only a small fraction of the animal's maximum speed and comparable records at higher speed were not obtained; also, no drag measurements on the animal are available. Excellent motion pictures obtained much earlier from various species by Gray (1933) are subject to the same reservations. Conversely, a few experimenters have determined the drag of a fish moving rigidly and symmetrically; for example, Sundnes (1963) in towing experiments found values for salmon and herring,  $C_D$  being around 0.01 in each case† for Reynolds number around  $10^6$ ; a drag value somewhat, but not excessively, above the expected smooth-body value (see below and also Hoerner 1965, chapter 6), the towing attachment being the probable cause of drag enhancement. Records of swimming movement for the same fish do not, however, exist.

Among such records of swimming movements as do exist, those involving carangiform motion with slender caudal fin represent only a small proportion, within which the clearest record seems to be that of Bainbridge (1963) on *Leuciscus*. He shows in his figure 1 the position and attitude of the caudal fin, seen from above, at successive instants 0.02 s apart, for a mean forward velocity of 0.48 m/s achieved by swimming movements at a frequency of about 2.5 Hz. From these data and from the variations in depth of the caudal fin depicted in his figure 3, it is possible to estimate the mean thrust  $\bar{P}$  as given on the reactive theory by equation (7).

Table 1 shows measured values of several parameters relating to the motion of the tip of the caudal fin at various times  $t$  in the swimming cycle. Here,  $V$  means its absolute speed relative to the water, the angle  $\theta$  is its inclination to the direction of mean motion, and the angle  $\alpha + \theta$  is the inclination of the path, travelled by the tip of the caudal fin, to that direction, these angles being taken positive to the right. Thus,  $\alpha$  is an angle of incidence which is positive when the sideforce on the caudal fin is directed to the left. In these terms we have (see figure 3)

$$w = V \sin \alpha, \quad W = \partial z / \partial t = V \sin (\alpha + \theta), \quad \partial x / \partial \alpha = \cos \theta, \quad (10)$$

quantities whose values are also shown in table 1, together with the variations in the tip value of  $m = \frac{1}{2} \pi \rho s^2$  obtained from Bainbridge's measurements of the tip depth  $s$ .

† Here and elsewhere,  $C_D$  is drag divided by  $\frac{1}{2} \rho U^2$  and by the total 'wetted area' of the fish.

TABLE 1. MEASURED VALUES OF SWIMMING PARAMETERS FOR *LEUCISCUS*  
OBSERVED BY BAINBRIDGE (1963, TEXT-FIGURES 1 AND 3)

(Note:  $1\text{N} = 1\text{kg ms}^{-2} = 10^5\text{ dyne.}$ )

quantity	... t	V	$\alpha + \theta$	$\alpha$	W	w	$\partial x/\partial a$	s	m	$mwW$	$\frac{1}{2}mw^2$	$\frac{1}{2}mw^2$
											$\partial x/\partial a$	$\partial x/\partial a$
unit	... s	$\text{ms}^{-1}$	deg	deg	$\text{ms}^{-1}$	$\text{ms}^{-1}$	l	m	$\text{kg m}^{-1}$	N	N	N
0.04	0.50	17	3	0.15	0.03	0.97	0.094	6.9	0.03	0.00	0.03	
0.08	0.60	45	3	0.42	0.03	0.74	0.087	5.9	0.07	0.00	0.07	
0.12	0.74	60	5	0.84	0.06	0.57	0.079	4.9	0.19	0.01	0.18	
0.16	0.68	48	16	0.50	0.19	0.85	0.083	5.4	0.51	0.08	0.43	
0.20	0.52	13	10	0.12	0.09	1.00	0.087	5.9	0.06	0.02	0.04	
0.24	0.62	-36	-8	-0.36	-0.09	0.88	0.088	6.1	0.20	0.02	0.18	
0.28	0.87	-60	-16	-0.75	-0.24	0.72	0.083	5.4	0.97	0.11	0.86	
0.32	0.67	-46	-10	-0.48	-0.12	0.81	0.084	5.5	0.32	0.03	0.29	
0.36	0.52	-25	-6	-0.22	-0.05	0.95	0.092	6.6	0.07	0.01	0.06	
0.40	0.48	-3	0	-0.02	0.00	1.00	0.093	6.8	0.00	0.00	0.00	

From these results are obtained the values of  $mwW$  and of  $\frac{1}{2}mw^2 \partial x/\partial a$  and of their difference, as given in the last three columns of table 1. The second term (physically due to the pressure force acting across the plane II) is found to make only a small negative contribution to this difference, a difference whose mean over the cycle is equal to the predicted mean thrust. This mean comes to 0.2 N, where the computation is not accurate enough to justify giving more than one significant figure.

This conclusion is interesting, however, because it implies a corresponding value for the mean resistance associated with the tangential motions of the surface, and this value is several times greater than would be expected for the gliding motion of a streamlined shape such as *Leuciscus* of length 0.3 m and velocity 0.48 m/s. The wetted surface area  $S$  of the fish can be inferred from Bainbridge's photographs to be about  $0.04\text{ m}^2$  so that the value of  $\frac{1}{2}\rho U^2 S$  is about 5 N. A drag of about 0.2 N implies therefore a drag coefficient based on wetted area of about  $C_D = 0.04$ , something like four times greater than would be expected at a Reynolds number based on length of  $R = 10^6$ .

This discrepancy is big enough to be insensitive to the many inaccuracies in the thrust calculation, as well as to uncertainties in the  $C_D$  value for gliding motion. Under the circumstances, it is sufficiently accurate to estimate this  $C_D$  value from known results for streamlined bodies of revolution. The maximum cross-section of the fish (see Bainbridge's Plate I) is an oval of major and minor axes 0.07 and 0.04 m, which from a viscous-resistance standpoint may be roughly equivalent to a circle of diameter 0.06 m (which would have a perimeter of approximately equal length). The fineness ratio of the equivalent body of revolution is thus 5 based on the complete length of fish, or 4 based on the length taken to the caudal peduncle (where the cross-sectional area first becomes very small). Results summarized by Hoerner (1965, chapter 6, figure 22) suggest that  $C_D$  is normally around 0.01 for

fineness ratios between 4 and 5 when  $R = 10^5$ . (It might be imagined that the turbulence which was probably present in Bainbridge's 'fish wheel' would have increased the appropriate value of  $C_D$ , but data in figure 23 of Hoerner's chapter 6 indicate that drag is negligibly increased by turbulence stimulation around  $R = 10^5$ .)

Evidently it may be rather important to investigate the discrepancy here apparent, by something like a factor of 4, between the simplest estimates of reactive thrust and resistive drag, even though the degree of importance of the discrepancy will not be clear until similar comparisons have been made for many other fishes. In the meantime a few comments must be made about various factors that may or may not be contributing to the discrepancy.

To begin with possible sources of error in the thrust estimate, the most obvious source of error is neglect of the resistive component, but probably this can be eliminated as an explanation of the discrepancy, which it would tend to increase if anything rather than decrease (Lighthill 1970). At the same time, error in the estimate of reactive thrust itself might arise because the estimate does essentially assume some small contribution from suction force at the leading edge of the caudal fin. Although this is a small part of the whole (of order  $w^2$  rather than  $wW$ ) it may be difficult to realize because of flow separation associated with the relative sharpness of the caudal fin's leading edge. It is on the other hand, possible, in view of the caudal fin's marked degree of flexibility, that the leading edge flexes sufficiently so that flow separation is significantly postponed. The flexing observed by Bainbridge (1963) is in the right direction for this, but even if the postponement does not occur the error is much too small to explain the discrepancy.

In Bainbridge's experiments the fish swam close to the bottom of the tank. It is not certain what the effect of this on the thrust-drag discrepancy would be, but the most obvious possibility is one of substantial thrust augmentation from ground effect. Estimation of the ground effect by regarding the ground as a reflexion plane suggests that the effective depth of the tip of the caudal fin and its reflexion taken together might be doubled, leading to a quadrupled virtual mass for the two taken together. That part of the associated momentum which attaches to water above the ground plane would on this argument be half the quadrupled water momentum associated with the pair of fins. There might accordingly be a doubling of lateral momentum associated with the caudal fin when it is near the ground, and a concomitant doubling of the thrust produced by given swimming movements.

All these considerations of possible refinements in the thrust calculation lead one with increasing conviction to the view, first expressed to the author by Dr Bone, that the viscous drag on the fish while it is swimming (that is, the drag needed to balance this reactive thrust) must for some reason be many times greater than that which would be associated with gliding motion. This conclusion is in the opposite sense to the famous 'Gray's paradox', but we must emphasize that it is being drawn here for a quite different group of animals.

In phrasing this conclusion, the term 'viscous drag' has been used to emphasize

that the resistance associated with tangential movements ( $u$ -movements) is involved. The wave-like character of the fish's movements ensures, by contrast, that the vortex-force associated with any perpendicular movements ( $w$ -movements) would be in the thrust direction. This is why, as already mentioned, any such force must make a positive resistive contribution to the total thrust.

Some other studies have shown how, when a fish is passively towed through the water, but lateral flutter of its body and fins is permitted, resistance similarly exceeds (and by a comparable factor) the value associated with rigid gliding motion. Webb (1970, p. 98) found this for freshly killed trout in a water tunnel, and obtained a resistance increasing with the degree of flutter permitted. His values of  $C_D$  with flutter practically restricted to the fins are about 0.015, agreeing well with values obtained for live trout in a measurement of gliding deceleration by Gray (1957). When the posterior portion of the dead trout, though stiffened by wires, was able to flutter significantly, Webb obtained values clustered about 0.03. It is uncertain whether these are primarily augmentations of viscous resistance, such as we postulate for drag augmentation by swimming movements, but the arguments that follow suggest this as a possibility.

Some independent evidence of viscous drag augmentation by swimming movements is provided by Smit (1965). He estimates the power output of the swimming muscles in goldfish (*Carassius auratus*) from their oxygen consumption when swimming in a fish wheel, and calculates (assuming 80% propulsion efficiency) that it exceeds by a factor 3.6 that needed to overcome viscous resistance in a rigid gliding motion (estimated approximately as above). This factor 3.6 obtained in a fish wheel, where relative velocity of fish and water is known rather accurately, is to be preferred to the factor that Smit obtained in water-channel experiments because fishes commonly use the slower flow in a channel boundary layer to reduce their required power output.

Dr Bone has suggested to the author one explanation of the augmentation of viscous drag through swimming movements, which seems quite the most promising at the present time. It is based on the idea that skin-friction over a large part of the surface is probably much greater than would be expected at a body length corresponding to a Reynolds number of  $10^5$ , because lateral movements prevent boundary layers from growing to anything like the thickness associated with such a body length.

The possibility that large drag augmentation may result from this is suggested by calculations of boundary-layer thicknesses on highly simplified assumptions. For example, a Blasius boundary layer (with uniform external stream velocity  $U$ ) has a 'frictional boundary-layer thickness'  $\delta$  (defined so that skin friction is  $\mu U/\delta$ ) equal to  $3\sqrt{vx/U}$  at a distance  $x$  from the leading edge. We may compare with this value a frictional boundary-layer thickness  $\delta$  for a flat section of depth  $s$  moving perpendicularly to itself with velocity  $w$ , which is  $0.6 \sqrt{vs/w}$  on the side towards which the section is moving.

Evidently, the two thicknesses differ by a factor of around 5, since the two square

roots would be close to each other in value for typical fish-swimming conditions. In practice, recoil causes there to be some significant  $w$ -velocity all along the swimming fish, for which therefore the effective boundary-layer thicknesses may be very substantially reduced below those associated with gliding motion.

On this view it would be the balance, existing between viscous diffusion of vorticity away from the fish's surface and convection of vorticity towards the surface by relative lateral motion between it and the water, that would determine the boundary-layer thickness, and hence also the skin-friction drag associated with the tangential  $u$ -motion. On parts of the surface that have a component of movement towards the undisturbed water, the skin friction might be increased over that for gliding flow by a large factor, around 5. This is the sort of difference that is needed to explain the thrust-drag discrepancy.

On this explanation, the relatively large amplitudes of tail motion exhibited by most swimming fishes must be interpreted in terms of the need to produce reactive thrusts sufficient to balance the greatly enhanced viscous drag that appears once any lateral movements commence. Steady swimming with small amplitudes of tail motion would on this hypothesis be rather futile, and indeed less efficient as a means of achieving a moderate rate of progress than production of thrust by those large-amplitude movements, interspersed with periods of gliding, such as are commonly observed among fishes in tanks.

In conclusion, it is necessary to emphasize the tentative nature of these suggestions about a probable thrust-drag discrepancy and its possible explanation. More data to reach greater clarity on the subject are urgently needed.

It is a pleasure to acknowledge the benefit of extensive discussions with Dr Quentin Bone, particularly on the topics of §3.

#### REFERENCES

- Bainbridge, R. 1958 *J. exp. Biol.* **35**, 109.
- Bainbridge, R. 1960 *J. exp. Biol.* **37**, 129.
- Bainbridge, R. 1963 *J. exp. Biol.* **40**, 23.
- Gray, J. 1933 *J. exp. Biol.* **10**, 88.
- Gray, J. 1957 *Scient. Am.* **197** (Aug.), 48.
- Gray, J. & Hancock, G. J. 1955 *J. exp. Biol.* **32**, 802.
- Hancock, G. J. 1953 *Proc. R. Soc. Lond. A* **217**, 96.
- Hoerner, S. F. 1965 *Fluid-dynamic drag*. Midland Park, N. J.: S. F. Hoerner.
- Lighthill, M. J. 1960 *J. Fluid Mech.* **9**, 305.
- Lighthill, M. J. 1969 *Ann. Rev. Fluid Mech.* **1**, 413.
- Lighthill, M. J. 1970 *J. Fluid Mech.* **44**, 265.
- Smit, H. 1965 *Can. J. Zool.* **43**, 632.
- Sundnes, G. 1963 *FiskDir. Skr. (Bergen)* **13**, 126.
- Taylor, G. I. 1952 *Proc. R. Soc. Lond. A* **214**, 158.
- Webb, P. 1970 Ph.D. thesis, University of Bristol.
- Wu, T. Y. 1961 *J. Fluid Mech.* **10**, 321.

## CHAPTER 6

# Aquatic Animal Locomotion : A Survey of Recent Theoretical Developments\*

### 1. Introduction

To this Congress of Theoretical and Applied Mechanics I want to report about an aspect of Zoologically Applied Mechanics: namely, the Theoretical Mechanics of Interactions between an aqueous medium and the external surfaces of a totally immersed animal swimming through it.

Generally speaking, progress in Zoologically Applied Mechanics can be made only through close collaboration between a zoologist and a mechanician. The animal kingdom is vast, and the problem of discerning and describing what types of animal motion have in different parts of it been significant, above all in the evolutionary sense, is one for the zoologist alone. At the same time, study of the complicated interactions between those motions and a surrounding aqueous medium is one for the hydrodynamicist alone. It is therefore when a zoologist and a hydrodynamicist have got to know each other well enough to be able to talk together about the problems, and gradually to learn enough of each other's language so as to be able to communicate effectively, that collaborative progress involving hydrodynamically sound analysis of zoologically significant motions becomes possible.

Especially significant for aquatic animal locomotion was such a collaboration which started around 1950 between Sir Geoffrey Taylor with his enormous mechanics experience and Sir James Gray whose zoology school at Cambridge had for over 20 years been a prime source of knowledge, and of research workers, in almost every aspect of animal locomotion (Gray 1968). It was in fact on *aquatic* animal locomotion and its hydrodynamics that Taylor and Gray particularly talked, and this led to important papers by Taylor (1951, 1952a, 1952b) on cases of such motion both at very low Reynolds numbers and at high Reynolds numbers.

These, like all Sir Geoffrey Taylor's papers in hydrodynamics, were so interesting that various younger workers on reading them were impelled to initiate work in imitation or extension of them. I was one of those both in respect of my own work (Lighthill 1952) and that of my research student Hancock (1953). However, that work of ours at Manchester on motions at very low Reynolds numbers failed to be influential because we had not learned the art of communicating with zoologists: either to discover from them the problems of zoological significance or to indicate

---

\* General Lecture to the International Congress of Theoretical and Applied Mechanics, Moscow 1972.

to them our mechanically significant conclusions. Fortunately, G. J. Hancock subsequently went to work with Sir James Gray and this collaboration produced some really significant results (Gray & Hancock 1955).

I personally did not meet James Gray until 1960 when I had the good fortune to spend a fascinating day with him discussing aquatic animal locomotion at high Reynolds number. This convinced me that it would be valuable to work out the application of hydrodynamical slender-body theory to fish swimming movements. I did this rather hastily and published it (Lighthill 1960) but I failed to maintain the collaboration with the Cambridge zoologists which would have been necessary to follow through the significance of the work, partly because duties as Director of the Royal Aircraft Establishment left little time for personal studies and partly because I still did not appreciate how essential it was to immerse myself in the zoological literature and the language of zoologists if I was to grasp the true nature of their problems or to indicate to them the meaning of my hydrodynamical investigations. Those were followed up by parallel studies in the application of two-dimensional aerofoil theory carried out with great hydrodynamical brilliance by Dr. T. Y. Wu (1961) but again without clear identification of the areas of zoological relevance.

Much later, in 1968, I returned to the subject in response to a proposal by the editors of Annual Reviews of Fluid Mechanics to include a survey article on it by me in their first volume. By then I had at last found out what kind of closeness of communication was needed between biologists and hydrodynamicists, through several combined studies in the physiological fluid dynamics of circulation and respiration: studies which I have recently described in summary form (Lighthill 1972). This pertinent experience, together with a reading of the literature of aquatic animal locomotion which exposed the general failure of communication between the two bodies of expertise evident in most previous work, convinced me that if I was to help in clarifying the hydrodynamics of aquatic animal locomotion I must talk to zoologists and go on talking to them; read their works and go on reading them: study their collections (in museums and aquaria) and go on studying them!

Finally with a great deal of help not only from Sir James Gray but from his colleague Dr. R. Bainbridge as well as from Dr. N. B. Marshall of the British Museum of Natural History and Dr. Quentin Bone of the Marine Biological Association at Plymouth, I produced (Lighthill 1969 and Chapter 2 of this book) a survey that included a paragraph as follows: 'It is likely that most readers may possess specialised knowledge in either zoology or the mechanics of fluids, but not both. The survey is written as far as possible to be equally intelligible to either kind of reader.' In fact I tried to produce, with the help of all that talking and reading and study of collections, a conducted tour through the animal kingdom in respect of the methods used for aquatic animal propulsion. It called upon voluminous work by zoologists on the data and their evolutionary interpretations, and attempted to relate to them such findings in the hydrodynamicists' papers as were of true zoological significance, identifying areas within the animal kingdom where each of those was specially relevant. That Review article tried to identify also the areas where further hydrodynamical researches would be particularly valuable.

Several of the researches there proposed have since been carried out and described in new papers of my own (Chapters 4 and 5), in new work by Dr. Wu (1971a, 1971b) and in work by two of my new research colleagues Dr. J. R. Blake and Dr. D. Weihs. It is this new work which makes up the main subject matter of this lecture. It has been a pleasure to see how effective my colleagues Dr. Blake and Dr. Weihs have been in learning to collaborate with zoologists and to use their language; much quicker than I had been! I shall be quoting from their work, including some still unpublished material, with their much appreciated permission.

This has been work where a considerable amount of observational data by Sir James Gray and other British zoologists gathered over the past 40 years has for the first time been analysed by hydrodynamic theory. At the same time zoologists are beginning to publish new observational material which they are themselves analysing by the new hydrodynamical methods (see, for example, Webb 1971).

The plan of this lecture assumes then, as stated in my Abstract available in advance, that the material in my survey article in *Annual Reviews of Fluid Mechanics* (Chapter 2) may be taken as a starting point. Those of you who have not read this but would like to appreciate more clearly how what I tell you about the hydrodynamics today fits into the zoological context, may find out something about it in that Review article, which is readily accessible. In fact I must urge you to do so, since the material is given importance and significance only by its zoological context. I feel justified, however, in leaving it out in a lecture to this 13th Congress of Theoretical and Applied Mechanics, since including the zoology in detail would leave no time for any mechanics, and also since it is my duty to concentrate on work done since the 12th Congress; in other words, on the new work that has appeared since that Review article was written.

## **2. The subdivisions of hydrodynamic theory relevant to aquatic animal locomotion**

I shall mention, however, before describing that work in detail, the main kinds of distinction that exist between the different types of hydrodynamic theory relevant to aquatic animal locomotion (Figure 1). The distinction whose importance is most obvious to any hydrodynamicist is that already mentioned, based on the value of the Reynolds number. The locomotion of animals whose size is of the order of 1 mm or less is a phenomenon where inertial effects are negligible because Reynolds number is very low (of order 1 or less). Theories of them have to find distributions of pressure and viscous stress satisfying *equilibrium* equations, leading to equations for velocities in a biharmonic form. The hydrodynamicist's problems for larger animals are, however, harder since fluid inertia is important as well as fluid stresses, and the equations of motion are fundamentally nonlinear.

One consequence of this distinction is a further subdivision that appears only at high Reynolds number: between on the one hand those types of aquatic animal propulsion on which I shall here concentrate, types that can be viewed as various kinds of developments from the *undulatory* methods of propulsion normal at very low Reynolds number, and on the other hand quite different types of propulsion using the inertial reaction of a jet. Little research work on these jet-propulsion

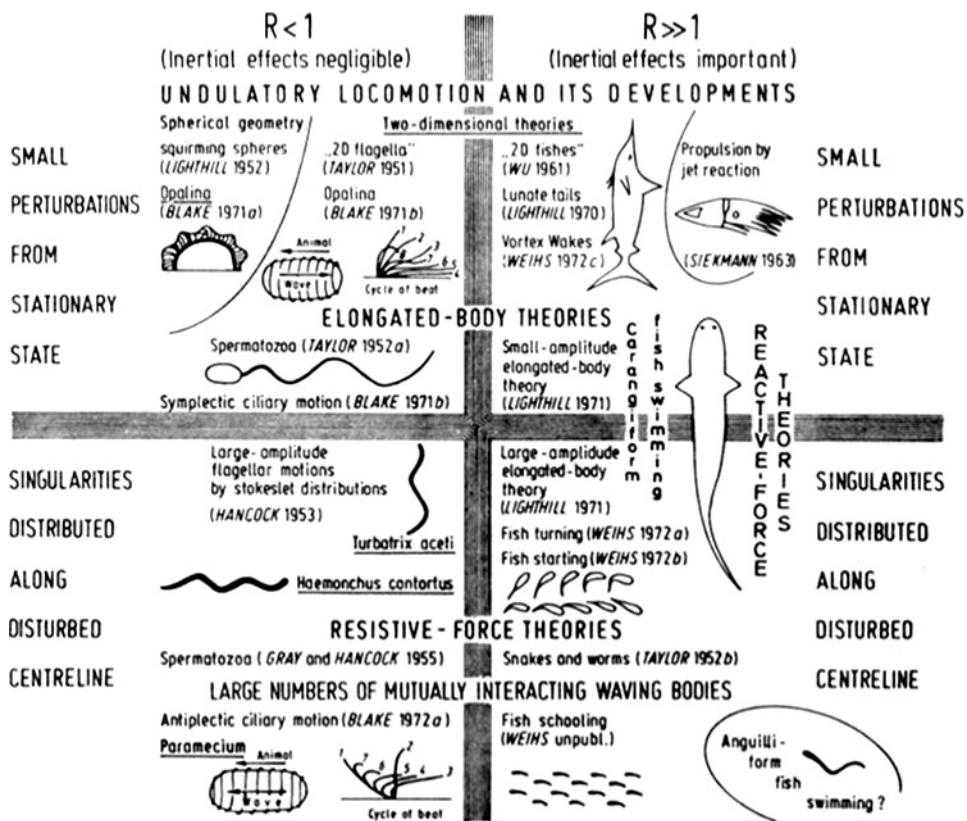


FIG. 1. Illustrating subdivisions of hydrodynamic theory relevant to aquatic animal locomotion.

mechanisms seems to have emerged since my 1969 Review. They are in many groups of invertebrates available as an escape reaction which depends on a sudden muscular contraction expelling a high-velocity jet of water so that the animal shoots off in the opposite direction. Much less widespread are the animals using such contractions in a regular periodic sequence to produce a 'pulse-jet' type of propulsion, although I might remark that those include not only various squids and other animals noted by Lighthill (1969) but also certain bivalve molluscs that can make rather startling progress by repeated sharp closures of the gap between their two shells!

Within the topic of undulatory methods of propulsion and developments from them, there are two more distinctions, each important both at very low Reynolds number and at high Reynolds number; distinctions apparent already in the papers of the early nineteen-fifties. They are both concerned with the kinds of method used for handling the complex geometry of the animal's motions, and the fact that those motions involve three space-dimensions and one time-dimension.

As an instance of one of the distinctions, Taylor (1952a) when he analysed the propulsion of a microscopic single-celled organism by a waving tail or flagellum,

modelled it as an elongated body of revolution, thus improving on an earlier study (Taylor 1951) with a much cruder two-dimensional model in which the third dimension of the space in which the animal swims is completely suppressed! Most subsequent theories have been based either on two-dimensional analysis or on some sort of elongated-body theory: the latter is certainly to be preferred when the body really is elongated and moves in three dimensions. Some aquatic animals, however, do (as we shall see) possess propulsive surfaces of genuinely high aspect-ratio (like the marlin's tail in Figure 1) to which an approximate two-dimensional analysis is quite appropriate.

Both these early papers possessed at the same time one *common* feature: they further approximated the effects of the animal's complex undulatory motions by regarding them as a small perturbation of the stationary stretched-straight state. That sort of small-perturbation approach has subsequently been much used also at high Reynolds number: especially by Lighthill (1960) and Wu (1961) for elongated-body theory and two-dimensional theory respectively. Obviously, however, it is not ideal since rather large perturbations are involved in most really interesting swimming motions.

For example, the Taylor theory made the velocity of a swimming spermatozoon increase indefinitely as the square of the amplitude of its swimming motions. My research student Hancock (1953) used a quite different method from the small-perturbation approach: one in which the fluid velocity field was modelled as that of a line-distribution of singularities but the singularities were distributed along the actual distorted position of the animal's centreline (line of cross-section centres). We gave at this time a special name to the most important kind of singularity that characterizes these solutions of the Stokes equations of low-Reynolds-number flow: we called this singularity (first used by Stokes 1851) the *stokeslet*.

Hancock's solutions involved distributions of stokeslets and ordinary source doublets along the instantaneous centreline of the organism at each instant. For determining their strengths he derived an integral equation, which it was feasible to solve numerically although very arduous (with the aids then available), and whose solutions gave a reasonable levelling-off of spermatozoon speed as the amplitude of swimming movements become large. He applied his work also to nematode worms, of which two are illustrated in Figure 1. However, the whole analysis was mathematically far too subtle and intricate to have any chance of appealing to zoologists and the calculations inconveniently complex, perhaps because the level of accuracy attempted was too high in relation to the accuracy of the observations that can be made in such a field.

It was the collaboration between Dr. Hancock and Sir James Gray that was finally successful in hammering out a method of adequate accuracy which is easily comprehensible to all concerned and gives quantitative results relatively easily (Gray & Hancock 1955). The approximation was relatively crude in that in a suitably chosen form of the integral equation the integral was suppressed altogether, but the results derived are in reasonable agreement with observations on spermatozoa. Physically, the approximation meant regarding the interaction between a short segment of the elongated animal and the surrounding aqueous

medium as consisting of a localised resistive force depending only upon the relative velocity between them, but with components of that relative velocity tangential to the line of centres resisted less than components normal to the line of centres.

The general idea of that approximation had been suggested earlier by Taylor (1952b) in work on motions at high Reynolds number of animals like worms and snakes whose body shapes are not very well adapted to aquatic locomotion. He similarly proposed the use of simple resistive forces with tangential or normal component depending on the tangential or normal component of relative velocity respectively. The only difference was that the dependence was linear at very low Reynolds number but quadratic at high Reynolds number, while perhaps at intermediate Reynolds number it ought to be given an intermediate form. I propose to call all theories 'resistive' if they make the local instantaneous force per unit length of animal depend, either linearly or nonlinearly, on the instantaneous velocity of that section of animal relative to the water.

Sir James Gray subsequently developed the resistive-force theories still further and showed in his book *Animal Locomotion* (Gray 1968) what a wide range of phenomena they will adequately account for. Here I shall take just a few minutes to expound them in by far their simplest form when the resistance relationships (whatever the Reynolds number) are approximated by linear laws and the cosine of the angle between the animal's line of centres and its direction of motion is approximated by 1.

Figure 2 shows a short segment *BEC* of the centreline of an elongated animal supposed to be performing undulatory motions from side to side in a stream of

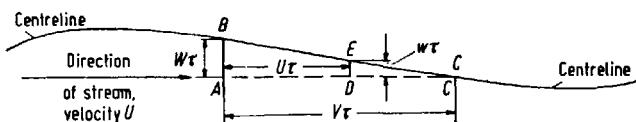


FIG. 2. *The connection between lateral velocities  $W$  relative to animal and  $w$  relative to water in undulatory motion.*

velocity  $U$  in which those swimming motions just stop it slipping backwards. (Thus  $U$  is its mean swimming speed relative to the stream.) The undulation passes backwards at a different speed  $V$ . The lateral velocity is  $W$  so that in time  $\tau$  the centreline moves sideways a distance  $W\tau$  from  $A$  to  $B$ , while the wave moved backwards a distance  $V\tau$  from  $A$  to  $C$  and therefore  $C$  is at the lateral position occupied by  $A$  a time  $\tau$  earlier. During the time interval  $\tau$  the water moves a distance  $U\tau$  so that relative to the water the centreline moves a distance  $DE = w\tau$ , where  $w$ , its normal velocity relative to the water, is

$$w = W \left( 1 - \frac{U}{V} \right). \quad (1)$$

This conclusion is as in equation (2) and Figure 5 of Lighthill (1969); note also an

alternative, analytical way of appreciating it using  $h(x, t)$  for lateral displacement as in Lighthill (1960) and the equations

$$W = \partial h / \partial t, \quad w = \partial h / \partial t + U \partial h / \partial x, \quad 0 = \partial h / \partial t + V \partial h / \partial x. \quad (2)$$

We take force per unit length of centreline resisting this normal velocity  $w$  as  $K_N w$ , where  $K_N$  is a normal-force coefficient, and multiply that by the slope  $W/V$  (see Figure 2) to obtain the corresponding component of forward *thrust* per unit length  $K_N w(W/V)$ . The same slope means however that the lateral velocity  $W$  includes a tangential component  $W(W/V)$  which with  $K_T$  as the tangential-force coefficient *reduces* the thrust per unit length by  $K_T W(W/V)$  to the net amount

$$P = (K_N w - K_T W)(W/V). \quad (3)$$

This thrust per unit length (3) averaged along the whole centreline must balance the drag per unit length  $K_T U$ .

Equation (3) is a formula of which two uses can be made. The useful thrusting work  $UP$  done per unit length can be compared with the rate  $W(K_N w)$  of mechanical energy expenditure by the lateral motions with velocity  $W$  against the resisting normal force  $K_N w$  per unit length. The ratio is an efficiency

$$\eta = \frac{UP}{W(K_N w)} = \left(1 - \frac{w}{W}\right) \left(1 - r \frac{W}{w}\right), \quad (4)$$

where

$$r = K_T / K_N \quad (5)$$

is the ratio of tangential to normal force coefficients. Maximum efficiency is achieved when

$$w/W = r^{1/2}, \quad \text{giving} \quad U/V = 1 - r^{1/2} \quad \text{and} \quad \eta = (1 - r^{1/2})^2. \quad (6)$$

This agrees with observations that at a Reynolds number of 1 or less when  $r$  is close to 0.5 (Gray & Hancock 1955) we have the ratio  $U/V$  of animal speed to wave speed around 0.3; note that maximum efficiencies are low (around 0.1) at these Reynolds numbers. A grass-snake (*Tropidonotus natrix*) with Reynolds number around  $10^6$  might however have  $r$  as low as 0.1, and thus  $U/V$  around 0.7 again as observed (Gray 1939); maximum efficiency might then be greater (around 0.5). Alternatively, as Taylor (1952b) pointed out, a polychaete worm with numerous parapodial attachments as in Figure 3 might have tangential-force coefficient greater than normal-force coefficient, for example,  $r = 1.5$ . This would give  $U/V$  around -0.2, again as observed; such polychaete worms adopt a vigorous forward-travelling undulation, although probably with maximum efficiency less than 0.05.

Furthermore, in all these cases the resulting swimming speed  $U$  can be deduced from a balance between the mean of the thrust per unit length  $P$  and the drag per unit length  $K_T U$ . This gives

$$U^2 = \overline{W^2} \left(1 - \frac{w}{W}\right) \left(\frac{w}{W} - r\right) \frac{1}{r}, \quad (7)$$

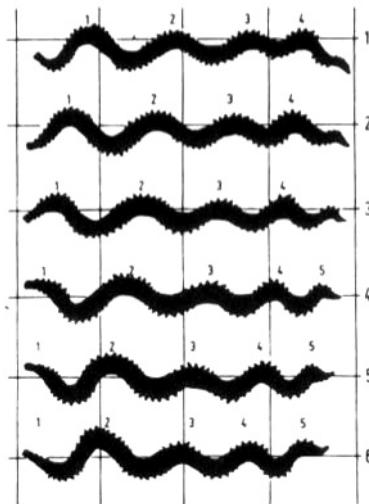


FIG. 3. Successive positions, at intervals of 0.05s, of a polychaete worm (*Nereis diversicolor*) swimming to the left. Numbers attached to moving 'crests' show that the wave travels in the same direction as the worm, although considerably faster (Gray 1939).

and maximum swimming speed is achieved when

$$w/W = \frac{1}{2}(1 + r), \quad \text{giving} \quad U^2/\bar{W}^2 = (1 - r)^2/4r. \quad (8)$$

This varies from 2 for  $r = 0.1$  (grass-snake) through 0.12 for  $r = 0.5$  (nematode) to 0.04 for  $r = 1.5$  (polychaete worm) in general agreement with the trend of ratios of swimming speed squared to mean square lateral velocity. Actually in the maximum-efficiency condition (6) the swimming speed is given by

$$U^2/\bar{W}^2 = (1 - r^{1/2})^2/r^{1/2}, \quad (9)$$

which is not much less than the maximum given by (8) even when  $r = 0.1$ , perhaps explaining why the higher value of  $w/W$  is not observed in such a case.

These very simple inferences from the resistive theory in the most straightforward approximation of all can be improved upon, either by not replacing cosines by 1 or by not linearising resistance laws, and yet involve calculations that are still quite manageable. The method is, however, essentially a method for organisms that propel themselves by means relatively wasteful of energy, as a hydrodynamicist will appreciate from the relatively low values of 'Froude efficiency' quoted.

Considerable improvements in the efficiency of the undulatory mode at high Reynolds number were achieved by the appearance in the fishes and other vertebrates of a deep, laterally compressed tail. Those advantages accrued partly because such a shape of cross-section was well adapted to interact with the water, not primarily through a resistive force depending on instantaneous relative velocity,

but rather with a *reactive* force depending on rate of change of that velocity. At high Reynolds number such a reactive force arises from the rate of change of momentum of water due to the rate of change of either the velocity or the virtual mass of surfaces in contact with it.

Anguilliform swimming, that is, the swimming of creatures like eels that maintain a markedly undulatory waveform all the way from head to tail so that considerably more than a whole wavelength is recognizably present, but where the mean swimming thrust is augmented by the mean rate of shedding of momentum into the wake at the laterally compressed tail, can be regarded as utilising both resistive and reactive forces. The development of a different mode of swimming, the carangiform swimming which appears in a wide variety of groups of fishes and other vertebrates, may however derive from the fact that it makes more *exclusive* use of the reactive forces which inherently can operate at a better Froude efficiency.

Carangiform swimming retains that feature of a backward-moving wave which requires the oscillations of the posterior end to lag behind those of an anterior section, but the amplitude is far more variable, being practically zero over the front half or even two-thirds of the length of the fish, and rising very steeply towards the posterior end. A fish using such swimming motions to move forwards produces very sudden accelerations in the water that it 'flicks into motion', and the resulting thrust is derived principally from the reactive rate of change of momentum.

The reactive theory for swimming of animals with elongated bodies was given by Lighthill (1960, 1969, 1970) in a series of papers based on a small-perturbation expansion about the stationary state. This theory is a development of the slender-body theory in aerodynamics, but 'elongated body' seems to be a phrase more readily understood in the world of zoology. The special feature that arises in the zoological application to carangiform motion is that any unbalanced sideforces in the posterior region of the fish may produce wasteful, undesired 'recoil' motions of sideslip of the fish's mass centre, and yaw about it. Lighthill (1969, 1970) showed that two morphological features which are almost universally found together with carangiform motion greatly minimise these undesired recoil motions; first, the body *depth* is large for an extensive region around the mass centre so that any recoil effects involve a large virtual mass of water and are consequently reduced in amplitude, and secondly there is a major reduction in body depth and hence of virtual mass just before the caudal fin, which is where especially large sideforces would otherwise result from the large spatial gradient in oscillation amplitude.

The ultimate development in speed and efficiency of aquatic animal locomotion was found in the combination of carangiform movement with a tailfin of high aspect-ratio as in certain scombrid fishes, in certain sharks, and in cetaceans. This crescent-moon-shaped or 'lunate' tail with its well-rounded leading edge has many of the excellent qualities of aircraft wings (though it possesses their horizontal arrangement and oscillates up and down only in the cetaceans, that is, whales and dolphins, and of course is vertical and oscillates from side to side in the relevant sharks and scombrid fishes). The thrust and efficiency of these animals' posterior propulsive units can be approximately analysed on a two-dimensional theory, which here comes into its own (Chapter 2, § 8).

In the last two sections of this lecture I shall be reporting upon two developments in the theory of carangiform locomotion. First, for the slender or elongated animals, there is the extension (Chapter 5) of the 'reactive' theory so that it no longer relies on a small-perturbation expansion but works like the developed forms of the resistive theory by placing the necessary singularities along the actual disturbed position of the animal's centreline. I developed this large-amplitude elongated-body theory to be able to improve the accuracy of thrust prediction in normal carangiform swimming, but Dr. Weihs (1972a, 1972b) went on to add more accuracy to the method and then apply it to problems where no small-perturbation approach would have been possible: problems of turning and starting. (The sequence of 10 pictures in Figure 1 shows a goldfish *Cyprinus auratus* simultaneously starting and turning!) Swimming fishes, indeed, have succeeded in developing refined methods of turning through (say) 90° in a distance of half their length with practically no loss of kinetic energy, and it is interesting to find out what movements they make to achieve this and why those movements are mechanically effective.

Finally I shall give some recent developments in two-dimensional theory that are relevant to lunate tails and their vortex wakes. They include studies of my own (Chapter 4) on why it should have been this lunate form of tail on which strains so different as whales, sharks and tuna should have 'converged' in a common search for speed and propulsive efficiency and also some recent studies on vortex wakes by Dr Weihs (1972c) which are relevant to wakes of individual fishes with lunate tails and also to the problems of assemblages of slender fishes that are grouped together in schooling patterns. Here we have an extension of animal locomotion theory at high Reynolds number to a study of large numbers of mutually interacting waving bodies.

It is this last-named feature which will be paralleled in the middle section of my lecture, where I describe the important recent extensions by Dr. J. R. Blake of theories of motion of individual flagella at very low Reynolds number to the case of ciliary propulsion, where the movements of large numbers of attached cilia in something close to synchronism produce relative movement between an animal and the surrounding fluid. Two different types of theory were found appropriate for this study of large numbers of mutually interacting waving bodies at low Reynolds numbers: one derived from small-perturbation theory applied to symplectic ciliary motion as in the organism *Opalina*, and one derived from large-amplitude resistive-force theory applied to antiplectic ciliary motion as in *Paramecium*.

### 3. Hydrodynamics of ciliary propulsion

The Ciliophora are a class of single-celled creatures which propel themselves by the nearly synchronised motion of large numbers of hair-like organelles attached to the cell surface. These organelles, the cilia, wave not in exact synchronism but in so-called *metachronism* with the phase of one row of cilia lagging by a certain amount behind that of an adjacent row. The effect is that of a wave of movements passing across the bank of cilia. The motion of each individual cilium

consists of a fast 'effective stroke' in the direction opposite to the animal's motion, with the cilium almost rigid during this effective stroke, and a slower 'recovery stroke' with the cilium relatively limp. The metachronism is called symplectic when the wave propagates approximately in the same direction as the effective stroke. This is reminiscent of the undulatory mode of propulsion in general since the wave is then moving in the direction opposite to the animal's motion.

A typical animal exhibiting symplectic metachronism is *Opalina*, a disk-shaped multinucleate protozoon about 200  $\mu\text{m}$  wide and about 20  $\mu\text{m}$  thick. Individual cilia are 10  $\mu\text{m}$  long or a little more and are arranged in rows 3  $\mu\text{m}$  apart in which the cilia may be as close together as 0.3  $\mu\text{m}$ . The tips of the cilia never become far apart from one another. This is particularly because during the fast effective stroke which might separate a cilium from its neighbours it is moving towards a cilium at an earlier stage in the cycle and away from one at a later stage. The general appearance is of 'a field of waving corn' with a clear undulating envelope embracing the tips of all the cilia. It is as if a disk-shaped body with a highly flexible surface were propelling itself by passing undulations in that surface backwards so as to exercise the necessary traction on the fluid.

Amusingly enough, that was exactly the hydrodynamical problem I had tackled 20 years ago (Lighthill 1952) using spherical geometry to simplify the necessary mechanics. At that time in my naïve way I had thought, 'Ah well! Taylor is working out how cells can propel themselves if they have tails; but let's see if they could get along at all without tails by some sort of squirming motion . . . ', whereas of course the only sensible thing to do was to talk to zoologists and find out how protozoans without a flagellum do propel themselves! After my 1969 Review it was obvious that it was to symplectic motion in Ciliophora that such a method of analysis should really be applied, and accordingly the first piece of work that Blake undertook after I suggested as a subject for his Ph.D. thesis the hydrodynamics of ciliary propulsion was the improvement and extension of my old spherical-geometry theory and its application to motions like those of *Opalina* (Blake 1971a).

He quickly had good success with this, and an encouraging feature of his results was that, with amplitudes of motion corresponding to ciliary lengths just quoted, the organism was predicted as able to achieve a velocity of about 100  $\mu\text{m}/\text{s}$ , close to those actually achieved. The rate of working needed could be expressed, interestingly enough, as an efficiency of under 0.03 when compared with the rate of working by a constant external force propelling the organism at the same speed. (This corresponds to 0.1 at low Reynolds number estimated in § 2 for a waving elongated body.) Furthermore, not all waveforms were calculated to produce such large velocities, or even positive velocities of swimming, though it was encouraging that those with relatively pointed crests and shallow troughs like those observed in nature were predicted to be the effective ones.

Another model used by Blake (1971b) for *Opalina* was an infinite plane sheet carrying a moving envelope of ciliary tips, his reasoning being that the real finite disk might have properties intermediate between the finite sphere and the infinite plane sheet. Comparing the two models with the behaviour of the animal he came to the conclusion that 'planeness is more important than finiteness', showing

clearly that cilia attached to a plane surface parallel to the direction of desired motion were more effective than cilia arranged all round a spherical shape. In nature most organisms employing ciliary propulsion do seem to have the cilia attached to surfaces parallel to the direction of motion. Note that the predicted speed of swimming would not be expected to change as the disk area increased to infinity since the thrust and the drag would equally increase in proportion to area, and so the infinite-sheet case may really be useful.

The plane model has several other advantages. It enables a deeper study of the relative merits of different waveforms, showing that a given swimming speed is achieved in symplectic motion with lowest rate of working if the tips of the cilia trace out the same kind of pattern as do the surface particles in a water wave. The motion of the tips comes fairly close to this in *Opalina*. Another advantage of the plane model is that the small-perturbation expansion can be carried to higher orders, and Blake (1971b, 1972b) was able to show by those means that even with a wave of the maximum possible amplitude corresponding to cusped crests the ratio of swimming speed to wave velocity (what I called  $U/V$  in § 2) could not exceed about 0.5. This agrees with maximum values observed in symplectic metachronism.

I leave that subject now with a mere note of a third model used by Blake (1971b) based on a body of revolution, and pass to antiplectic metachronism. This is different in important ways: the swimming velocities produced are larger relative to either the wave velocity or the square root of the rate of working. Primarily this is because the individual cilium has greater freedom of movement during its effective stroke than in the symplectic case. Now in its effective stroke it is moving towards a cilium at a later stage in the cycle which is already stretched forward away from it, and its neighbour behind is similarly stretched backward away from it (being at an earlier stage). Thus, the critical act by each cilium is performed in splendid isolation, and a model which regards the tips of the cilia as if they were a flexible solid surface ceases to have any potential value. *Paramecium*, another multinucleate protozoan, is a typical animal exhibiting antiplectic metachronism, being an elongated prolate spheroid about  $250 \mu\text{m}$  long with cilia  $10 \mu\text{m}$  long and a metachronal wavelength of  $20 \mu\text{m}$ , but its typical velocity is  $1000 \mu\text{m/s}$ , an order of magnitude greater than for *Opalina* (and  $U/V$  for *Paramecium* even exceeds 1).

Blake (1972a) developed a completely new type of theory for antiplectic metachronism although building it up on the sound foundations laid by Gray & Hancock (1955). Thus he regarded each cilium as a distribution of stokeslets along its centreline corresponding to the distribution of the force of interaction between the cilium and the fluid. He had, however, to make two important steps forward beyond the Gray & Hancock model.

First, he had to consider the effect of each stokeslet in the presence of the nearby solid wall of the organism. He could treat this problem as that of a stokeslet near an infinite plane wall and showed (Blake 1971c) that the solution to the Stokes flow equations in this case may elegantly be expressed in terms of an image system. The image system is different for stokeslets corresponding to force components

parallel to and perpendicular to the boundary, as illustrated in Figures 4 and 5 respectively. In the first case the image system consists of an equal and opposite stokeslet plus a stokes-doublet (consisting of two equal and opposite stokeslets oriented as shown) plus a source doublet. The far field is substantial, being of the combined stokes-doublet form shown with velocities falling off like  $r^{-2}$ . In the second case the type of stokes-doublet in the image system is different and cancels out the equivalent stokes-doublet provided by the original stokeslet and its equal and opposite image. The far-field in consequence is much weaker, falling off like  $r^{-3}$ . For ciliary propulsion this means that relative motion between the water and the wall is contributed to principally by ciliary forces *tangential* to the boundary

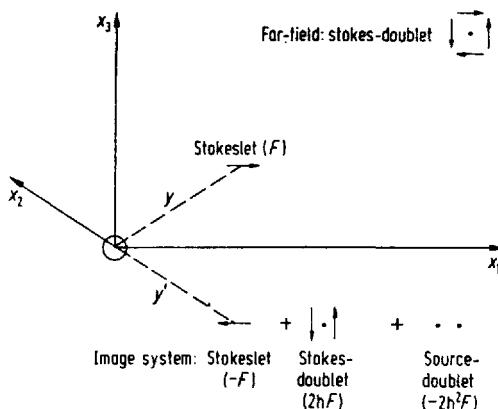


FIG. 4. Image system for a Stokeslet of strength  $F$  directed parallel to, and at a distance  $h$  from, the solid plane wall  $x_3 = 0$  (Blake 1971c).

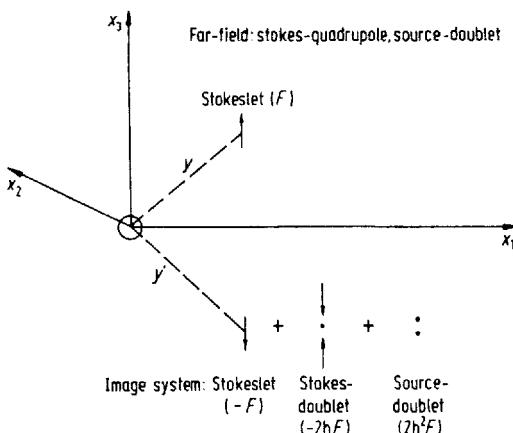


FIG. 5. Image system for a Stokeslet of strength  $F$  directed perpendicular to, and at a distance  $h$  from, the solid plane wall  $x_3 = 0$  (Blake 1971c).

rather than by forces perpendicular to it. Blake subsequently found that the image system had previously been discovered by the great Lorentz (1896), but no one seems to have made any effective use of it till Blake came along!

Secondly, he had to take into account that each cilium was acting in the presence of a velocity field brought about by the actions of all of the other cilia. This statistical element of the effect of the large numbers of cilia involved was obviously the hardest aspect of the problem.

The method of solution chosen was similar to that used in atomic theory to study the structure of atoms of high atomic number; that is, containing large numbers of electrons. For this purpose the atomic theorists use a 'self-consistent field' approximation where the radial distribution of electrostatic field resulting from the electronic distribution is regarded as the key unknown function, to be determined by the fact that Schrödinger's equation, with that radial field distribution, must possess such solutions for the distributions of electrons as are *consistent* with the electrostatic field postulated. Exactly such a self-consistent field model was found to give good results in Blake's problem.

There, the distribution of the velocity field parallel to the wall as a function of distance from it is regarded as the key unknown function  $U(x_3)$ ; this represents velocities relative to the wall, and so its limit as  $x_3 \rightarrow \infty$  is the swimming velocity of the organism while its lower values as  $x_3$  reduces to its wall value of zero describe the shearing that occurs in the ciliary sublayer. The unknown function  $U(x_3)$  is now determined as a self-consistent field as follows. Gray & Hancock (1955) give the force with which each cilium acts on the fluid in terms of the velocity of the cilium *relative* to the fluid, here taken as moving with the velocity  $(U(x_3), 0, 0)$ . The effect of this force is represented by a stokeslet together with its image system discussed earlier and the velocity fields of all the cilia are thus written down and summed up, the sum being simplified by use of Poisson's summation formula. Then the self-consistency condition requires of course that the velocity field so obtained is the one that we first thought of! This leads to an integral equation for  $U(x_3)$  which lends itself to numerical solution by iterative means.

Figure 6 shows some results of such calculations with  $x_3$  nondimensionalised in terms of the length  $L$  of a cilium and  $U$  nondimensionalised in terms of the product of  $L$  with the radian frequency  $\sigma$ . The method was carried out for various values of a frequency parameter  $\kappa$ , which is the ratio of  $\sigma L$  to the wave velocity  $c$ , and of  $\gamma$  which is the ratio of normal-force coefficient to tangential-force coefficient so that it is the reciprocal of the  $r$  used in § 2. Note the much higher values of velocity produced in *Paramecium* with its antiplectic, than in *Opalina* with its symplectic, metachronism. The third animal *Pleurobrachia* is one of the ctenophores, small organisms related to the jellyfishes which propel themselves by eight 'combs' of cilia, waving in antiplectic metachronism.

Blake (1972a) uses one more parameter  $\tau$ , proportional to the density of the cilia per unit area. The data in Figure 6 are for  $\tau = 150$ , whereas Figure 7 shows how  $U^\infty$ , the velocity of fluid relative to the organism outside the ciliary sublayer, varies as a function of this density parameter  $\tau$ . Time does not permit reference to any more results from this paper which include calculations of many

of the quantities that the zoologists most need to know: particularly, how the forces and bending moments applied by the animal to a cilium vary during the cycle, as well as of overall rate of working. I will remark, however, that elsewhere Blake (1972b) shows good agreement, for *symplectic metachronism only*, between the envelope method and this much more comprehensive method of analysis.

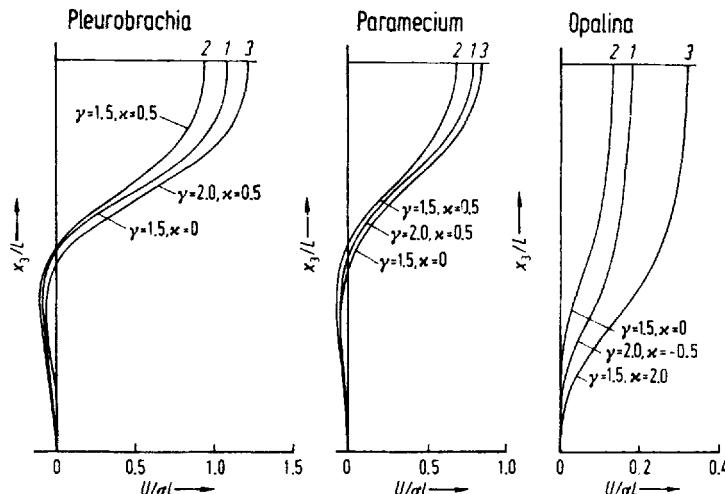


FIG. 6. Velocity distributions  $U(x_3)$  in the ciliary sublayer for motions of cilia of length  $L$  at radian frequency  $\sigma$  when the density parameter  $\tau = 150$  (Blake 1972a).

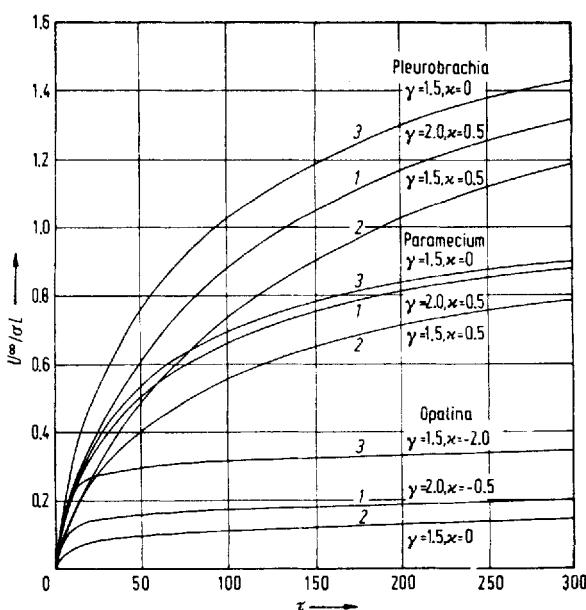


FIG. 7. Variation of  $\lim_{x_3 \rightarrow \infty} U(x_3) = U^\infty$ , the swimming speed of the organism relative to fluid at a large distance, as a function of the density parameter  $\tau$  (Blake 1972a).

#### 4. Large-amplitude elongated-body theory

Now I must return to motions at high Reynolds number and reactive-force theories, and describe my recent work (Lighthill 1971) in developing these so that they can be applied to motions of elongated bodies with arbitrary amplitude. This new method is based, like the successful low-Reynolds-number methods, on distributing the disturbed centreline those singularities that are relevant: here, the momentum-carrying dipoles (source-doublets). The only difference is that since the present method is applied exclusively to fishes I shall call the centreline the spinal column or backbone!

After many searches for the right approach, I recognized finally that what I wanted could be obtained by use of the following three principles:

- (i) Water momentum near a section of fish is in a direction perpendicular to the backbone and has magnitude equal to the virtual mass,  $m$  per unit length, times the component  $w$  of fish velocity in that direction.
- (ii) Thrust can be obtained by considering rate of change of momentum within a volume  $V$  enclosing the fish whose boundary at each instant includes a flat surface  $\Pi$  perpendicular to the caudal fin through its posterior end.
- (iii) In the momentum balance it is necessary to take into account transfer of momentum across  $\Pi$  not only by convection but also by the action of the resultant  $\frac{1}{2}mw^2$  of the pressures generated by the motions within the plane  $\Pi$ .

One advantage of considering only rate of change of momentum within the volume  $V$  which excludes the complicated vortex wake is that all the complex changes of momentum due to movements of vortices in that wake, each under the influence of the others, can be ignored. A simple hydrodynamical theorem equates the resultant pressure force on such a plane, due to motions in the plane with virtual mass  $m$  per unit length, to the expression  $\frac{1}{2}mw^2$ .

To turn these principles into a method for calculation, we use a coordinate system as defined in Figure 8, where a point on the spinal column is defined by a Lagrangian coordinate  $a$ , measured along the spinal column from its tail. The  $y$ -axis is vertical and the  $x$ - and  $z$ -axes horizontal and otherwise arbitrary except that

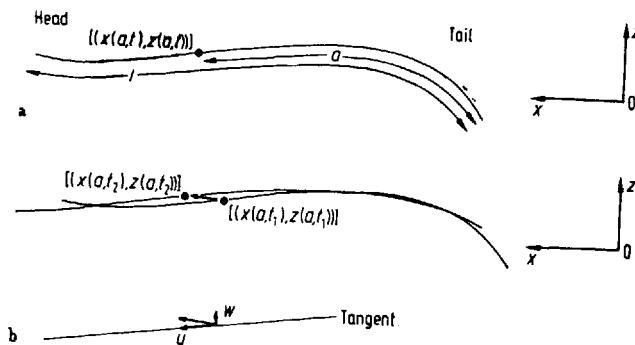


FIG. 8. Notation used by Lighthill (1971) to describe large-amplitude motions of fishes in a frame of reference  $(x, y, z)$  that is stationary relative to fluid at a large distance.

in cases of steady swimming we take the  $x$ -axis as the direction of the mean swimming velocity. The velocity component perpendicular to the spinal column is  $w$  as already stated while the tangential component is  $u$ .

The spinal column is effectively inextensible which in analytical terms means

$$\left(\frac{\partial x}{\partial a}\right)^2 + \left(\frac{\partial z}{\partial a}\right)^2 = 1. \quad (10)$$

Thus unit vectors tangential and perpendicular to the backbone in the horizontal plane  $y = 0$  have components

$$\left(\frac{\partial x}{\partial a}, \frac{\partial z}{\partial a}\right) \quad \text{and} \quad \left(-\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a}\right) \quad (11)$$

respectively, from which we deduce that

$$u = \frac{\partial x}{\partial t} \frac{\partial x}{\partial a} + \frac{\partial z}{\partial t} \frac{\partial z}{\partial a}, \quad w = \frac{\partial z}{\partial t} \frac{\partial x}{\partial a} - \frac{\partial x}{\partial t} \frac{\partial z}{\partial a}. \quad (12)$$

Now we can write down the equation for the instantaneous reactive force  $(P, Q)$  due to the momentum  $mw$  per unit length brought instantaneously into being by the fish's perpendicular motions. This reactive force  $(P, Q)$  must on the average provide a thrust to balance the viscous drag resisting the fish's tangential motions. In other words, the whole idea of this method is that tangential motions will be treated resistively but perpendicular motions reactively.

This reactive part  $(P, Q)$  of the total force, then, is determined by the equation

$$\begin{aligned} \frac{d}{dt} \int_0^l mw \left( -\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a} \right) da \\ = \left[ -umw \left( -\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a} \right) + \frac{1}{2} mw^2 \left( \frac{\partial x}{\partial a}, \frac{\partial z}{\partial a} \right) \right]_{a=0} - (P, Q). \end{aligned} \quad (13)$$

The left-hand side is the rate of change, in a volume  $V$  that includes the fish but not the wake, of a total water momentum whose value per unit length has magnitude  $mw$  and direction (see (11)) perpendicular to the backbone. The right-hand side adds up three elements contributing to this rate of change: (i) loss by convection across the plane  $\Pi$  with velocity  $u$ ; (ii) gain by action of the pressure force  $\frac{1}{2}mw^2$  acting across  $\Pi$ ; (iii) loss due to the water acting with a force  $(P, Q)$  on the fish.

A minor rearrangement using (12) enables us to write (13) as

$$\begin{aligned} (P, Q) = & \left[ mw \left( \frac{\partial z}{\partial t}, -\frac{\partial x}{\partial t} \right) - \frac{1}{2} mw^2 \left( \frac{\partial x}{\partial a}, \frac{\partial z}{\partial a} \right) \right]_{a=0} \\ & - \frac{d}{dt} \int_0^l mw \left( -\frac{\partial z}{\partial a}, \frac{\partial x}{\partial a} \right) da. \end{aligned} \quad (14)$$

In the case of periodic swimming movements the mean thrust  $\bar{P}$  can be calculated using the fact that the mean of the time-derivatives must be zero as

$$\bar{P} = \overline{\left[ mw \frac{\partial z}{\partial t} - \frac{1}{2} mw^2 \frac{\partial x}{\partial a} \right]}_{a=0}, \quad (15)$$

where the bar denotes a mean. The time-derivative term in (14) is important, however, because without it the thrust would fluctuate enormously, falling to zero whenever  $w = 0$  at  $a = 0$ , whereas Lighthill (1971) shows that the time-derivative term acts to smooth out those fluctuations. The expression for  $Q$  is important also in leading to the conclusions regarding recoil mentioned in § 2.

The same method gives an expression for the mean rate of energy wastage

$$\bar{T} = \overline{[\frac{1}{2} mw^2 u]}_{a=0} \quad (16)$$

by shedding of kinetic energy  $\frac{1}{2} mw^2$  per unit length at velocity  $u$  across  $\Pi$  into the vortex wake. Thus all the terms in the Froude efficiency

$$\eta = \bar{U}\bar{P}/(\bar{U}\bar{P} + \bar{T}), \quad (17)$$

including the mean swimming speed

$$\bar{U} = \overline{\left[ u \frac{\partial x}{\partial a} \right]}_{a=0}, \quad (18)$$

can be written in terms of mean values of quantities evaluated at the posterior end  $a = 0$ . This makes it easy to infer how Froude efficiencies in the 80% region can be expected, in sharp contrast to the values around 50% (see § 2) expected at high Reynolds number from a resistive interaction with perpendicular motions.

In balancing these reactively produced thrusts against the resistive reactions to tangential motions, as Lighthill (1971) does in detail for the dace *Leuciscus*, it must be remembered that boundary layers tend to remain very thin towards the back of a swimming fish due to the lateral movements bringing the fish in contact with continually new fluid, and accordingly the drags are greater once the fish is swimming than when it is gliding. The relatively large amplitudes of tail motion exhibited by most swimming fishes may thus be interpreted in terms of the need to produce reactive thrusts sufficient to balance the greatly enhanced viscous drag that appears once lateral movements commence. Steady swimming with small amplitudes of tail movement may be rather futile and indeed less efficient as a means of achieving a moderate rate of progress than production of thrust by those large-amplitude movements, interspersed with periods of gliding, such as are commonly observed among fishes in tanks.

These theoretical approaches were found suitable also for the work of Weihs (1972a, 1972b) on fish turning and fish starting, of which I shall describe the turning studies here. Sir James Gray had taken some excellent films of fishes turning in 1933 but they were only analysed from a hydrodynamical point of view when Dr.

Weihs came to work with me in Cambridge in 1971. Weihs made an improvement to my method in the details of his treatment of fins: thus, he used the virtual-mass approach for forces associated with the fish body but used slightly improved aerodynamic coefficients for forces on fins which he felt were not sufficiently 'slender' for the fish to be treated accurately enough by 'slender-body theory'. He was able to calculate the forces and moments at each of the three stages in the observed turning process and show how they relate to the observed motion of the fish's mass-centre and to the rotation of the fish about it.

Figure 9 shows a sequence of shots from Sir James Gray's film of a turning rudd *Cyprinus erythrophthalmus*. The fish's success in making the turn in an extremely short distance with little kinetic energy loss is most noticeable. It seems to be attributable according to Dr. Weihs's analysis to the fact that the earliest big movement is a *nose turn* and that vigorous activity with the tail fin only follows after this to complete the job. The benefits of this *alternate* use of fore and aft control surfaces seem very clear.

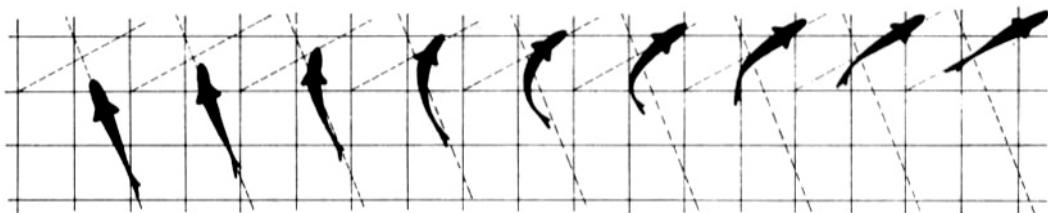


FIG. 9. Successive positions at intervals of 0.04s (time sequence from left to right) of a rudd (*Cyprinus erythrophthalmus*) making a turn to the right (Weihs 1972a).

Figure 10 shows successive positions of the spinal column of the turning rudd measured up carefully from the photographs. The circles indicate successive positions of the mass centre. The first stage of the turning process (shots 1, 2 and 3) is one in which the mass centre continues in a straight line because a substantial turn of the nose to the right produces momentum-change forces (like the  $d/dt$  terms in equation (14)) that are balanced by the inherently more effective momentum-shedding forces from fins due to smaller movements of the posterior part of the body to the left. The total side force is zero and the net effect is a growing couple permitting rotation about the mass centre.

The second stage of the turning process (shots 4, 5 and 6) is one in which the mass centre itself follows a highly curved path, indicating a substantial force component towards the centre of turn. This is contributed by the increasing dominance of the forces provided by the substantial turning of the posterior parts, and the forces calculated from that observed turning agree with the observed centripetal acceleration.

The third stage (shots 7, 8 and 9) is one in which the tail is doing all the work and the hydrodynamic 'lift' (really sideforce) on the caudal fin is accelerating the fish along its new path to balance any loss of momentum due to friction during the

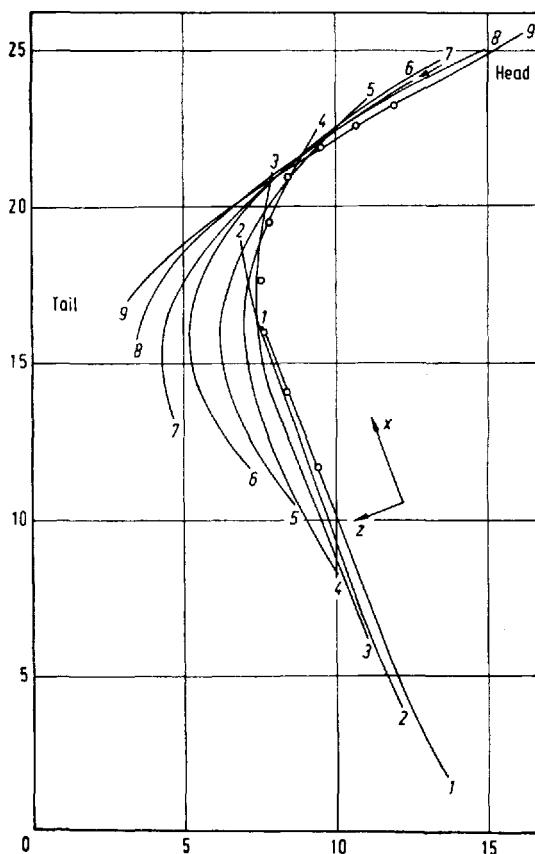


FIG. 10. *Tracings of position of the backbone of the rudd during the turning process. Numbers refer to frames in Figure 9 and distances are marked in cm (Weihs 1972a).*

turn. At the same time the moment of that force about the mass centre is such as to stop the rotation of the fish and leave it moving in a straight line.

All the detailed calculations of forces and moments made by Dr. Weihs bore out this analysis except that the calculated moments in stages 2 and 3 were enough to bring about a much *greater* retardation of rate of turn than seemed necessary. Then he noticed an important feature of the film record, namely, that at the onset of stage 2 the fish extends and twists its right pectoral fin to exert a brake action on the right-hand side, reducing speed and increasing rate of turn. This explains the rather enhanced rate of turn at the beginning of stage 2 and the fact that larger moments than would otherwise be expected are required to damp it down. The minor speed reduction reduces the radius of turn generated by a given sideforce in proportion to the square of the speed reduction and then that final tail flick gets back most of the lost speed relatively easily! The airbrakelike action of that pectoral fin on the side towards which the fish is turning comes as a sudden movement between stage 1 (when nose turning dominates the scene) and stage 2 with its

extensive tail deflection and nose straightening. Finally stage 3 with a sharp tail straightening completes the process. It is intriguing to speculate whether ships will ever do as well!

### 5. Vortex wakes

Time permits now only a very brief discussion of my final topic: vortex wakes, which appropriately come last because they are what aquatic animal locomotion leaves behind! At this series of Congresses it is a pleasure to begin this last discussion with a contribution by Karman and Burgers (1934): Figure 11 is adapted in



FIG. 11. *Illustrating vortices cast off by the caudal-fin trailing edge as the fish moves to the left, and also the jet-like streamline pattern induced by those vortices (Lighthill 1969, after Karman & Burgers 1934).*

fact from a diagram in their famous joint article in the series *Aerodynamic Theory*.

We see here Karman and Burgers emphasizing that the famous Karman vortex street can also work in reverse: it can, with the sense of rotation of all the vortices altered, be responsible for thrust instead of drag. Between the vortices with that altered sense of rotation a jet-like motion is induced, and the momentum it carries is related to the thrust on the fish. So after all there is rather more relationship between the undulatory mode and jet propulsion than first appeared!

I shall mention a few ways in which this simple picture might be improved on and two ways in which the idea might have zoological application. The picture as it stands shows a purely two-dimensional motion so that it would at first sight be most nearly applicable to the wakes of deep high-aspect-ratio lunate tails (see § 2), although Dr. Weihs suggests that also fishes swimming in formation one above another might produce such two-dimensional wakes and would be able in this way to assure the advantage of high aspect-ratio just when it suited them!

Be that as it may, Weihs (1971c) has studied the behaviour of these hitherto neglected thrust-generating vortex streets when the motion is strictly two-dimensional, and found that they have interesting tendencies to change their form with distance downstream, becoming wider by a factor around 1.5 and somewhat altering also their longitudinal spacing.

Somewhere, of course, this behaviour must be modified by interaction with the drag wake though for a fish with lunate tail that would be more concentrated and differently located (around the locus of the fin *centre*). The changes it may bring to the vorticity distribution must, however, at some downstream location, reduce to zero the momentum carried by the wake, since the average values of thrust and drag cancel.

I want to mention now three-dimensional effects and their possible relation to why the fastest marine animals from widely different lines of evolution should

have converged on to the lunate tail as their method of propulsion. My detailed mathematical analysis of its merits by purely two-dimensional theory (Chapter 4) indicated merely that something like a *straight* trailing edge would be optimal: that is, a leading edge convex forwards accompanying a straight trailing edge. This suggests that any advantage of the trailing edge itself being somewhat convex forwards must stem from three-dimensional considerations.

Such three-dimensional considerations involve taking streamwise vorticity shed by a finite-span lunate tail into account alongside the cross-stream vorticity shown in Figure 11. They then indicate that Figure 11 must be viewed as a section by a horizontal plane of a sequence of vortex rings that are batted backwards diagonally alternately to the right and to the left by the animal. Actually the typical frequency parameters used by animals with lunate tails are such that those would be approximately *circular* rings, and evidently the curved trailing edge may assist in forming such circular rings, which in general are very good devices for carrying as much momentum as possible in relation to their energy.

Initially the vortex rings would move diagonally backwards but when the effect of the drag wake had mingled sufficiently with them they would be left with only a lateral motion. Dr. McCutcheon mentioned to me that he had seen very recently such a pattern of laterally moving vortex rings forming the wake behind a scombrid fish. I continue to like the vortex-ring argument for lunate-tail optimality and am also continuing to seek ways of checking it more satisfactorily.

Finally, I come to fish schooling. I mentioned already that the wake in Figure 11 might be closely two-dimensional if several fishes were swimming in formation one above the other and that it would be to their advantage in respect of the power needed to swim at a given speed if they did so. It is also clear that fishes swimming diagonally behind the fish shown would particularly benefit from its vortices helping to drag them forward. This suggests a diamond pattern of fishes (Figure 12) as having a particular advantage in respect of reduced power consumption to swim at a given speed, and Dr. Weihs in unpublished work has been calculating the advantages which are in fact obtained.

There is a certain body of observational evidence that such diamond-shaped

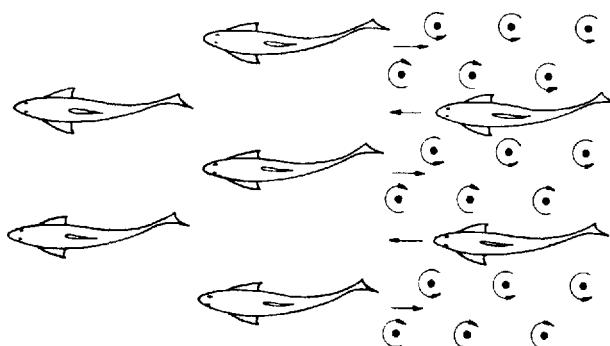


FIG. 12. Diamond-shaped fish schooling pattern.

lattice patterns are in fact taken up by fish schools when moving fast, and the question arises whether *passive* forces bring the pattern into play or whether very elaborate control mechanisms act to maintain the lattice pattern. Dr. Weihs can see certain mechanisms whereby passive forces might possibly be effective. I have jokingly suggested to him that the whole phenomenon may indeed be like the way in which the disordered molecules of a liquid form into a crystal lattice when the temperature is lowered: in other words, as the speed of the fish school increases, the *energy* advantage of the fishes dropping into this diamond-shaped lattice formation is suddenly able to outweigh the loss of entropy!!!

## REFERENCES

- BLAKE, J. R. 1971a *J. Fluid Mech.* **46**, 199–208.  
 ———, 1971b *Ibid.* **49**, 209–222.  
 ———, 1971c *Proc. Camb. Phil. Soc.* **70**, 303–310.  
 ———, 1972a *J. Fluid Mech.* **55**, 1–23.  
 ———, 1972b Ph.D. thesis, University of Cambridge.  
 GRAY, J. 1939 *J. Exptl. Biol.* **16**, 9–17.  
 ———, 1968 *Animal Locomotion*, Weidenfeld & Nicholson, London.  
 GRAY, J. & HANCOCK, G. J. 1955 *J. Exptl. Biol.* **32**, 802–814.  
 HANCOCK, G. J. 1953 *Proc. Roy. Soc. A*, **217**, 96–121.  
 KARMAN, T. VON & BURGERS, J. M. 1934 *General Aerodynamic Theory: Perfect Fluids*, vol. 2 of *Aerodynamic Theory* (ed. Durand, W. F.), Springer, Leipzig.  
 LIGHTHILL, M. J. 1952 *Comm. Pure Appl. Math.* **5**, 109–118.  
 ———, 1960 *J. Fluid Mech.* **9**, 305–317.  
 ———, 1969 *Ann. Rev. Fluid Mech.* **1**, 413–446. (Chapter 2.)  
 ———, 1970 *J. Fluid Mech.* **44**, 265–301. (Chapter 4.)  
 ———, 1971 *Proc. Roy. Soc. B*, **179**, 125–138. (Chapter 5.)  
 ———, 1972 *J. Fluid Mech.* **52**, 475–497. (A version of Chapters 10 and 14.)  
 LORENTZ, H. A. 1896 *Zittingsverlag Akad. v. Wet.* **5**, 168–187.  
 SIEKMANN, J. 1963 *J. Fluid Mech.* **15**, 399–418.  
 STOKES, G. G. 1851 *Trans. Camb. Phil. Soc.* **9**, pt. II, 8–106.  
 TAYLOR, G. I. 1951 *Proc. Roy. Soc. A*, **209**, 447–461.  
 ———, 1952a *Ibid.* **211**, 225–239.  
 ———, 1952b *Ibid.* **214**, 158–183.  
 WEBB, P. W. 1971 *J. Exptl. Biol.* **55**, 489–540.  
 WEIHS, D. 1972a *Proc. Roy. Soc. B*, **182**, 59–72.  
 ———, 1972b Submitted to 1st Internat. Congr. Biorheology. (Published in its final form as the last reference in Chapter 7.)  
 ———, 1972c *J. Fluid Mech.* **54**, 679–691.  
 WU, T. Y. 1961 *Ibid.* **10**, 321–344.  
 ———, 1971a *Ibid.* **46**, 337–355.  
 ———, 1971b *Ibid.* **46**, 521–568.

*This page intentionally left blank*

## CHAPTER 7

# Some Current Investigations of Aquatic Animal Motions

### 1. Developments at low Reynolds numbers

Concluding our account of aquatic animal locomotion, we refer briefly in Chapter 7 to some current investigations that follow naturally out of work surveyed in Chapter 6, beginning with developments at low Reynolds number arising out of Dr. Blake's work there described.

First we briefly note that certain of Dr. Blake's recent researches (Blake 1973a, Blake & Sleigh 1974) form a bridge between external and *internal* biofluidynamics, since he has used his model of the ciliary sublayer, developed for the study of ciliary propulsion, to analyse internal fluid motions in ducts *within* animals, induced by the motions of cilia lining those ducts. Although internal flows induced by cilia are widely found throughout the animal kingdom, we here mention *only* two kinds that are important in human physiology: the movement of mucus in the airways of the lung and the movement of semen in the ductus efferentes of the male reproductive tract.

Because the metachronal coordination in both cases is thought to be antiplectic, Blake's ciliary sublayer is the only suitable fluiddynamic model known. Figure 1

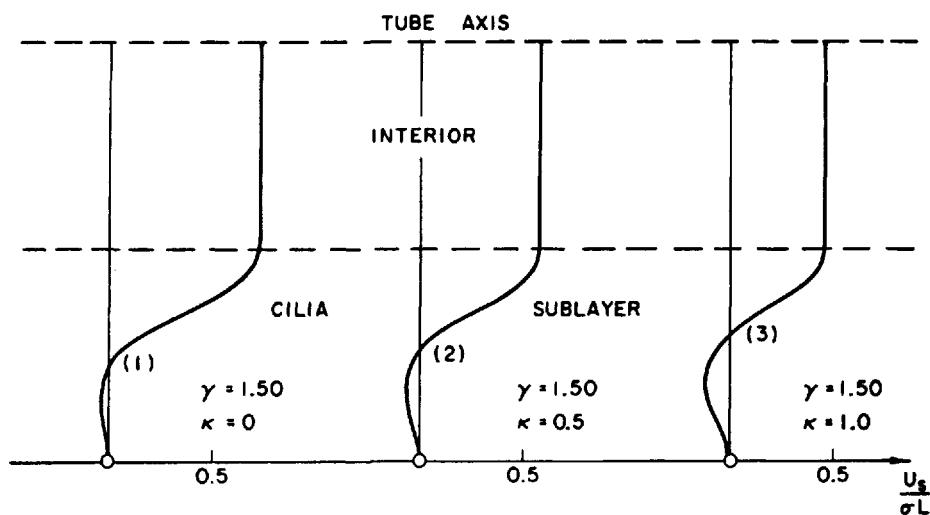


FIG. 1. *Blake's calculations of the distribution of the velocity of semen  $U_s$  in the ductus efferentes.*

shows results using this model to calculate the velocity distribution for semen in the ductus efferentes using this model with  $\gamma = 1.5$  (corresponding to a value  $\frac{2}{3}$  for the resistance-ratio  $r_K$  of Chapter 3) and with different values of  $\kappa = \sigma L/c$  (where  $\sigma$  is radian frequency,  $L$  is cilium length and  $c$  the velocity of the metachronal wave).

The problems for movement of mucus in lung airways are different: evidently, the mucus does not fill the whole of the tube; in addition, the mucus is a *thixotropic* fluid. In other words, its long-chain mucopolysaccharide molecules under relatively quiescent conditions build up a structure that gives mucus a high effective viscosity, but under conditions subjecting it to a high rate of shear a marked reduction in effective viscosity results from the break-up of that larger-scale structure.

These considerations indicate the special importance of the type of velocity distribution indicated in Figure 1 (or in Figure 6 of Chapter 6), with its region of high shear in the outer portion of the ciliary sublayer. Blake's calculations for antiplectic metachronism had indicated maximum shear rates of about  $2\sigma$  in this region (where  $\sigma$  is the radian frequency of the ciliary motion). This suggests (Blake 1973b) a maximum shear rate of about  $200 \text{ s}^{-1}$  for mucus in the lung airways, which would be sufficient to give it a 'watery' consistency (greatly reduced viscosity) in that outer part of the ciliary sublayer, allowing 'thick' (much more viscous) mucus outside it to slip over it relatively easily. Various observations tend to confirm this view of mucus movement.

Returning to external biofluidynamics, we consider a possible additional use of Blake's image system for a Stokeslet in a solid plane boundary. It is argued that this could be applied to determine how flagellar propulsion is modified in the close proximity of a solid plane boundary, such as the glass of a microscope slide, along general lines described in the remainder of this section.

Consider an organism making swimming motions with its centreline remaining solely in the plane  $x_3 = h$  parallel to and at a distance  $h$  from the solid wall  $x_3 = 0$ . We represent the flow around the organism as in Chapter 3 by means of normal and tangential Stokeslets distributed along the centreline and find that the boundary conditions can be satisfied provided that:

- (i) the normal Stokeslets are accompanied by dipoles satisfying equation (28) of Chapter 3;
- (ii) both the normal and the tangential Stokeslets, being directed parallel to the wall, are accompanied by image systems as in Figure 4 of Chapter 6 to satisfy the wall boundary conditions;
- (iii) those boundary conditions, on the other hand, are negligibly disturbed by the dipoles (i) whose influence at the wall is insignificant, on the assumption that the ratio  $h/a$  is substantial;
- (iv) on the same assumption, the field of the *image system* (ii) varies negligibly over any cross-section from its value on the centreline.

Then the calculations in Chapter 3 leading, on the flagellum boundary, to the velocity field (29) for normal Stokeslets accompanied by their appropriate dipoles, or to the velocity field (30) for tangential Stokeslets, are modified only by the integrated effect of the image system, calculated in accordance with (iv) above on the

centreline itself. We comment first on the *sign* of this effect: it is a reduction in both velocity fields, since the principal element in a stokeslet's image system (and the only element with a  $1/r$  term in its far field, leading by equation (24) of Chapter 3 to a logarithmic form of integrated effect) is an equal and *opposite* stokeslet. Secondly, a larger proportional reduction is made to the lesser velocity (29) than to the greater velocity (30). It follows that the resistance-ratio  $r_K$  is decreased for flagellar motions in planes parallel to the wall, with consequences clear from discussions in Chapters 2 and 3: in particular, the swimming-velocity ratio  $U/V$  must be increased.

We can verify that the effect is indeed substantial by a rough calculation assuming that the length  $b + c$  within which the flagellum is effectively a straight cylinder is large compared with the distance  $2h$  to its image system.

Then the integral along that image system of the reciprocal distance  $1/r$  to a point on the flagellum can be approximated as in equation (24) of Chapter 3, but with  $2h$  (the perpendicular distance from the image system) replacing  $a$  in the logarithm  $\log(4cb/a^2)$  to convert it into  $\log(cb/h^2)$ . The flagellar velocity for stokeslets of strength  $\mathbf{f}$  per unit length, calculated in Chapter 3 as

$$\frac{\mathbf{f}}{8\pi\mu} \left( 1 + \log \frac{4cb}{a^2} \right) \quad \text{or} \quad \frac{\mathbf{f}}{8\pi\mu} \left( -2 + 2 \log \frac{4cb}{a^2} \right) \quad (1)$$

for normal or tangential stokeslets respectively, is then reduced in either case by the amount

$$\frac{\mathbf{f}}{8\pi\mu} \left( 1 + \log \frac{cb}{h^2} \right) \quad \text{or} \quad \frac{\mathbf{f}}{8\pi\mu} \left( -2 + 2 \log \frac{cb}{h^2} \right), \quad (2)$$

respectively. Here the logarithmic term comes from the image stokeslet, which indeed is responsible for the whole expression in the tangential case. In the normal case, the image stokes-doublet produces an addition of 2 to it which is reduced to 1 by the counteracting effect of the source-doublet or dipole of strength  $2h^2\mathbf{f}/\mu$  (with, of course, a far more significant velocity field at distances of the order  $h$  or  $2h$  than the already neglected field of the dipole of strength  $\mathbf{f}a^2/4\mu$  associated with any normal stokeslet).

The resistance-ratio  $r_K$ , which is the ratio of the first velocity in (1) to the second (that is, the normal velocity associated with a given normal force, divided by the tangential velocity associated with an equal tangential force) is on this simple approximate analysis (with the expressions (2) subtracted from the expressions (1)) reduced, from values as high as 0.7 indicated in Chapter 3 as appropriate to flagellar motions in unbounded fluid, to the well-known value 0.5 used by Gray & Hancock (1955) in their classic work! Obviously, this is a crude estimate, which nevertheless gives a preliminary indication that the effective value of  $r_K$  can be reduced significantly for flagella very close to a wall.

The above comments were confined to *planar* motions of a flagellum *parallel* to a solid wall: a case with two simplifying features, that all points on the centreline

are at the *same* distance  $h$  from the wall and that only the image system in Figure 4 of Chapter 6 needs to be used. Analysis of more general motions will be considerably complicated by the need to take into account variations in  $h$  along the flagellum. It must also involve stokeslets perpendicular to the wall with their different image system (that of Figure 5).

Here we note only one feature of the more general case: study of this image system for stokeslets perpendicular to the wall shows that the flagellar velocity (1) is reduced as a result of the image system not by the first term in (2) but by the greater term

$$\frac{\mathbf{f}}{8\pi\mu} \left( 3 + \log \frac{cb}{h^2} \right).$$

Admittedly, this is a very crude estimate based as before on supposing the images spread along a straight cylinder of length  $c + b$  parallel to the wall: an even less appropriate assumption when the flagellar motions are not in planes parallel to the wall. It suggests, however, a possibility worth investigating: in certain modes of flagellar propulsion along a plane solid wall, motions normal to the centreline might have stronger components perpendicular to the wall than tangential motions, in which case an enhanced reduction of flagellar velocity for perpendicular force components would be more applicable to the normal than to the tangential motions, leading to still greater reductions in the ratio  $r_K$ .

This interesting possibility makes it worth noting why the bracketed expression in (3) includes an extra term 2 as compared with the first bracketed expression in (2): the extra term 2 results entirely from a stronger effect of the image stokeslet itself. As shown in Chapter 3, § 1, a stokeslet generates, at a given distance  $r$ , a bigger motion at positions in the direction of the stokeslet itself than in a perpendicular direction; an effect which entirely accounts for the excess of (3) over (2), while the effects of both the stokes-doublet and the source-doublet elements in the image system turn out to be unchanged. For any stokeslets, then, whether parallel or perpendicular to the wall, the dominant feature reducing flagellar velocity associated with a given force applied by unit length of flagellum is the simple effect of an equal and opposite image stokeslet. Dominance of this simple effect may be found to facilitate future investigations of flagellar locomotion near solid surfaces that may take into account much fuller geometrical details of the motion.

## 2. Developments at high Reynolds numbers

The opportunities for new developments in mathematical biofluidynamics are widespread also in relation to aquatic animal locomotion at high Reynolds numbers. In this section we note several such opportunities before outlining certain current developments that arise out of work described in Chapter 6.

One general feature of aquatic animal locomotion has been much less effectively analysed at high than at low Reynolds numbers. This is the one emphasized in Chapter 3, § 1: that an animal's motile activity determines only how its external surfaces move *relative to* its mass center, whereas its total motion is a combination

of these relative movements with a certain translation of the mass centre and a rotation about that mass centre. At high Reynolds numbers we cannot neglect inertia, so that to determine the velocity  $\mathbf{U}$  of that translation and the angular velocity  $\boldsymbol{\Omega}$  of that rotation the resultant force on the animal must be set equal, not to zero, but to the animal's rate of change of momentum, while the resultant couple must be set equal to the animal's rate of change of angular momentum.

There have been relatively few investigations taking these principles into account for the *oscillatory* components of the resultant force and couple on a swimming animal. One such was given, however, for carangiform motion in Chapter 4, § 4 on the basis of small-amplitude reactive-force theory. This fairly extensive investigation, from equation (40) onwards, exemplifies methods of estimating the effect of various morphological features on 'recoil' motions of sideslip and yaw in carangiform swimming of elongated fishes, but there is considerable scope for improving it: both as applied there to small-amplitude motions, and through extension to large-amplitude motions.

An earlier part of Chapter 4, § 4 also represents a first attempt in a field that needs thorough investigation: the use of both reactive and resistive forces in a combined theory that might describe swimming motions of a generally anguilliform nature. Between two *extreme* theories that are linear even for large-amplitude motions (resistive-force theory for low Reynolds numbers, with solutions of the biharmonic equation obtained by linearly combining the fields of stokeslets distributed along an arbitrary distorted centreline of the animal, and reactive-force theory for high Reynolds numbers, with solutions of Laplace's equation obtained from a similar distribution of dipole fields) we are concerned here with an area more difficult because the phenomena are inherently nonlinear. This difficulty was bypassed in Chapter 4, § 4 by a very crude linearisation of the cross-sectional resistive-force law (replacement of  $\rho sw^2 \operatorname{sgn} w$  by  $\frac{3}{4}\rho sw_0 w$  in the interval  $-w_0 < w < w_0$ ).

Other difficulties, mentioned there in a footnote, arise from uncertainty about how much the effect of vortex-force wakes generated in the cross-flow around anterior sections may be modified by posterior motions. It would be valuable to work out an extreme theoretical model neglecting any such modification, as suggested in that footnote. This would take into account purely local resistances  $\rho sw^2 \operatorname{sgn} w$  per unit length to normal motion with velocity  $w$ , alongside the reactive forces calculated by the methods of Chapter 4 or (better) by the large-amplitude theory of Chapter 5.

Similar uncertainties arise with another type of vorticity aligned with the direction of swimming; namely, the trailing vorticity shed by any anterior fins. In Chapter 4, the part of § 3 from equation (19) onwards is devoted to a small-perturbation analysis of the effect of this vorticity, assumed shed from fin trailing edges at right angles to the backbone and unmodified after shedding. One interesting feature of the calculation is that an anterior fin is predicted to generate an additional thrust without additional *wasted* energy provided that a certain phase relationship holds. This is that the lateral momentum (or 'sidewash') in the trailing vorticity from the anterior fin is at least  $90^\circ$  out of phase with the lateral motion of

the caudal fin reabsorbing it. Since the body motions are propagated along the fish at a wave speed  $V$  while the vorticity is carried relative to the fish at the lower swimming speed  $U \doteq \frac{2}{3}V$ , the two can acquire a phase difference of  $90^\circ$  over a distance between fins of about half a wavelength (equation (28) of Chapter 4).

Newman & Wu (1972) and Newman (1973) have extended that theory in various ways. They allow for modification of the vorticity shed by anterior fins due to convection along the streamline pattern associated with flow over the stretched-straight body of the fish. This can be important for configurations typical of submarines, as they demonstrate in the case of a body of revolution at constant angle of attack with rigid fins, the tail fin having a considerably lower span than the anterior fin. A fish body, on the other hand, with its substantial lateral compression, should not cause so much distortion of the vorticity trailing from dorsal and ventral fins; furthermore, that vorticity's interaction with a caudal fin is much reduced under conditions, not of constant angle of attack, but of undulation leading to the phase difference just described. Newman & Wu indicate another interesting generalisation by proposing the type of analysis that would have to be used for fins with slant trailing edges.

Section 5 of Chapter 4 gave a two-dimensional small-disturbance theory of the action of a lunate tail. This has now been extended to a three-dimensional finite-span theory by Dr. M. G. Chopra (1974), still assuming small disturbances. There would also be great interest in the extension of lunate-tail studies (initially in the two-dimensional case) to a truly large-amplitude theory, as exhibited for *elongated* bodies in Chapter 5.

A major lacuna in existing knowledge of elongated-body locomotion at high Reynolds number is highlighted in § 3 of Chapter 5: the study of swimming fishes has been unbalanced in that far more attention has been paid to thrust than to drag. Here is a marked difference from the low-Reynolds-number investigations of Chapter 3, where a single analysis evaluates both the drag production by a flagellum's tangential motions and the thrust production by its normal motions and balances the two. The inherent nonlinearity of the Navier-Stokes equations has prevented such a unified treatment at high Reynolds number. Rather, it has been assumed that drag results from viscous stresses, resisting the fish's tangential motions through the water and being developed in the thin laminar or turbulent boundary layer on the fish's body, whereas thrust results from forces generated by the fish's normal motions, associated with a distribution of water momentum mainly outside the boundary layer.

When a fish is gliding forward in a stretched-straight condition, it is subject only to a steady-flow resistive drag, whose value can be determined as the rate at which the fish loses momentum. Evidently for a fish to maintain a constant speed it must make oscillatory swimming movements generating a reactive thrust at least as large as will counteract that steady-flow drag. It was suggested in Chapter 5, however, that the thrust required is in fact considerably *greater* because the oscillatory-motion resistive drag for elongated fishes may considerably exceed the steady-flow resistive drag. (In this suggestion there was of course no implied view that the same distinction holds for animals with lunate tails including the

cetaceans, for which Gray's well-known studies had suggested a different conclusion.)

Evidence presented to this general effect in Chapter 5, § 3 has since been supplemented by the studies of Webb (1971), particularly those estimating a trout's rate of working from its rate of oxygen consumption. The indication from Webb's work is that oscillatory-motion resistive drag may be about 3 times greater than steady-flow. This is in general accord with the factor of 4 derived for dace in Chapter 5 prior to deductions due to likely overestimates of reactive thrust by elongated-body theory. It can be understood in terms of a possible fivefold increase in skin friction on the side of a fish towards which its normal velocity  $w$  is directed, due to something like a fivefold reduction in thickness of that boundary layer of newly attached fluid which is constantly being formed on that side.

The increasing acceptance of a substantial ratio, of order 3, between the oscillatory-motion and steady-flow resistive drags has led Dr. Weihs to investigate some of its possible biological implications. One implication for fishes of the same density as water is that a given distance can be covered with about half the energy consumption needed for constant-speed motion by periods of swimming alternated properly with periods of rigid gliding, and a paper going into details of this optimisation study and relating it to some observational data is in process of preparation. This has been preceded by a paper (Weihs 1973a) on fishes heavier than water, which may also be able to halve their energy requirements for a given journey by alternating periods of climb with periods of downward gliding!

Dr. Weihs has, in addition, been pursuing further that aspect of his paper on the turning of fishes (Weihs 1972) which was not specially highlighted in Chapter 6: the study of a goldfish *Carassius auratus* simultaneously starting and turning. This study had showed interesting parallels with the turning of an already swimming rudd described in Chapter 6: the total change in orientation of the fish was once more about  $90^\circ$  and the turning process was divided into three stages in a broadly similar fashion.

Figure 2 reproduced from Weihs (1972) set out schematically his conclusions regarding the mechanics of the combined start and turn. Just as with the rudd, the first stage (frames 1, 2 and 3) is one involving no net force on the fish, whose centre of mass therefore remained fixed (unlike that of the rudd, which, naturally, continued in a straight line). The fish's muscular activity produces a general bending during this stage, and the dominant forces are momentum-change forces associated with rates of change of water momentum normal to the body. These forces are indicated for frames 1 and 2 in the lower part of the diagram; they are seen to be close to static equilibrium, generating no net rotation of the fish about its mass centre. By frame 3, however, more powerful momentum-shedding forces at the caudal fin have come into play, while the momentum-change forces have slackened off, and a substantial clockwise rotation of the fish begins. This large twisting couple given to the fish as the caudal fin moves from position 3 to position 4 replaces the action of the pectoral fin which produced the same effect (together with a desirable deceleration) on the already swimming rudd.

The second stage (as with the rudd) is a stage where turning continues and a

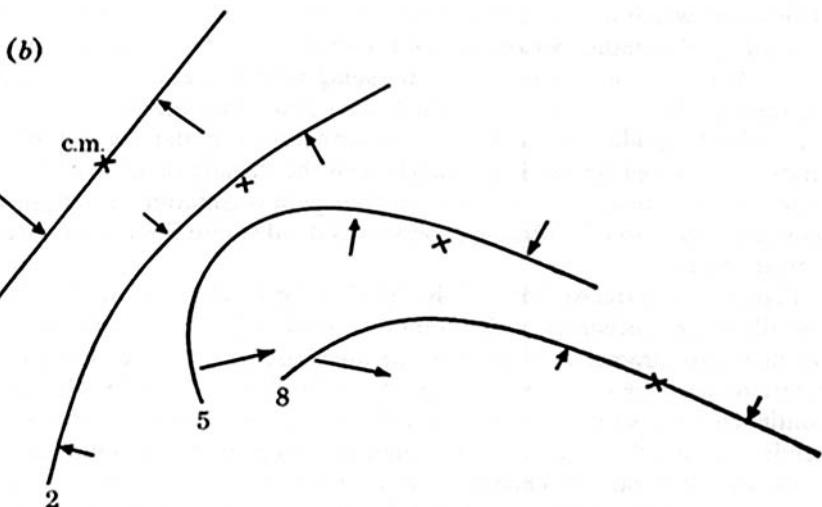
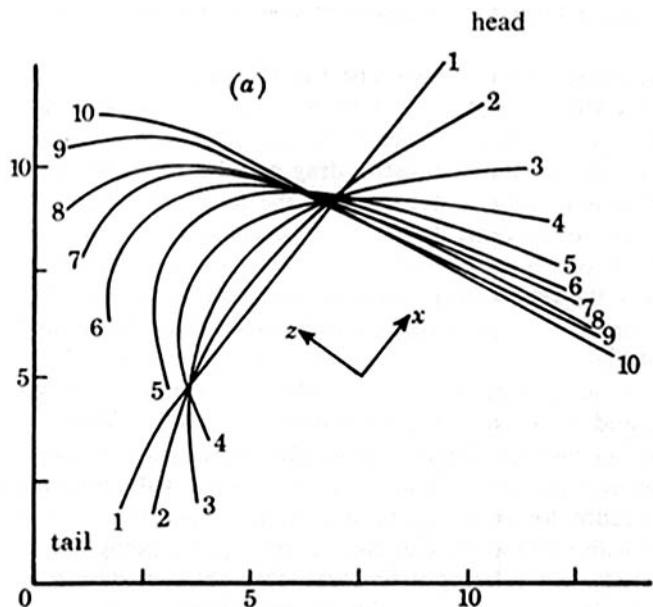


FIG. 2. This Figure 3 from Weihs (1972) illustrates the combined start and turn of a goldfish.

large force component tending to bring about the desired change of momentum of the fish comes into play. At frame 4 the caudal fin reversed the sign of its angle of attack and momentum-shedding forces began to give the centre of mass a substantial velocity component in the direction of ultimate motion, while no significant couple retarding the rotation about it was yet present owing to continuing momentum-change forces in the anterior part of the body. This is illustrated for frame 5

in the lower diagram. In the third stage, as with the rudd, the caudal-fin forces began to dominate entirely, completing the acceleration of the centre of mass and annihilating the rotation about it.

Dr. Weihs's current studies are bringing to light various evidence that for many fishes a rapid turn (whether from a swimming or from a standing start) is such a 90° turn. For example, a film sequence of a 180° turn proved it to be made up of two 90° turns with the fish straightening up in between! A recent paper (Weihs 1973b) is devoted to rapid starting of fishes, such as the pike *Esox lucius*, which specialise in a sudden lunging movement for predatory purposes, involving an acceleration of (say)  $50 \text{ ms}^{-2}$ . Interestingly enough, the natural rapid lunge of such a fish appears to include a turn through nearly 90°, and to take a form broadly similar to that in Figure 2. Dr. Weihs suggests that the only available forces big enough for a rapid lunge are caudal-fin forces which must be exercised in a movement of the caudal fin at right angles to the ultimate direction of motion at moderate angle of attack; such a movement is possible only after a general body bending, which tends to determine the whole operation as taking the form illustrated in Figure 2. The pike's capacity for specially rapid starts may derive from the action of dorsal and ventral fins that are positioned very far back on the body, close to the caudal fin where they can supplement its accelerating action.

By contrast, a trout stimulated to make a start in the direction in which it was originally pointing had to perform a sequence of motions somewhat like that in Figure 2 but with the use of its left pectoral fin as an asymmetric brake to turn the fish back towards its original orientation at the end of stage 2. The net rate of acceleration achieved was then less than when the same fish made a combined start and turn. Scope remains for far more widespread studies of fish manoeuvrability along these general lines.

## REFERENCES

- BLAKE, J. R. 1973a *Flow in tubules due to ciliary activity*, Bull. Math. Biol. **35**, 513–523.
- , 1973b *A note on mucus shear rates*, Resp. Physiol. **17**, 394–399.
- BLAKE, J. R. & SLEIGH, M. A. 1974 *Mechanics of ciliary locomotion*, Biol. Reviews **49**, 85–125.
- CHOPRA, M. G. 1974 *Hydromechanics of lunate-tail swimming propulsion*, J. Fluid Mech. **64**, 375–391.
- GRAY, J. & HANCOCK, G. J. 1955 *The propulsion of sea-urchin spermatozoa*, J. Exptl. Biol. **32**, 802–814.
- NEWMAN, J. N. 1973 *The force on a slender fish-like body*, J. Fluid Mech. **58**, 689–702.
- NEWMAN, J. N. & WU, T. Y. 1972 *A generalized slender-body theory for fish-like forms*, Ibid. **57**, 673–693.
- WEIHS, D. 1972 *A hydrodynamical analysis of fish turning manoeuvres*, Proc. Roy. Soc. B, **182**, 59–72.
- , 1973a *Mechanically efficient swimming techniques for fish with negative buoyancy*, J. Marine Res. **31**, 194–209.
- , 1973b *The mechanics of rapid starting of slender fish*, Biorheology **10**, 343–350.

*This page intentionally left blank*

## CHAPTER 8

### Animal Flight

#### 1. The conquest of the air

Within the field of external biofluidynamics, the aquatic aspects with which Chapters 1 to 7 have been concerned comprise the *earliest* developments in the evolution of animal locomotion, as well as many interesting later developments: at least  $6 \times 10^8$  years of animal evolution in the marine environment can be traced in the extensive fossil record from around the beginning of the Cambrian period to the present time. By contrast, the *aerial* aspects of external biofluidynamics, to which Chapters 8 and 9 are devoted, began their process of evolution much later: the 'conquest of the air' by flying animals is a considerably more recent event than the origin of swimming.

Nevertheless, the chronology of animal flight is longer, and the epoch of the conquest of the air has to be placed earlier, than many people would imagine. Already  $3 \times 10^8$  years ago, in the middle of the Carboniferous period, the problems of sustained powered flight had been solved by the insects. An excellent account of some of the processes involved in the evolution of animal flight has been given by Smart & Hughes (1972), whose diagrammatic summary is reproduced here as Figure 1.

They describe the invasion of the land by both plants and animals, including primitive Arthropoda such as myriapods, as having taken place during the Silurian period a little more than  $4 \times 10^8$  years ago, after which in the Devonian period arthropod evolution on land led to the six-legged form characteristic of the insects before any development of wings took place. Extensive fossil material from winged insects appears, however, by the middle of the Carboniferous period. It includes a remarkable diversity of forms: already the divergence had occurred between the palaeopterous orders such as the Odonata (dragon-flies) which cannot reflex their wings over the abdomen when resting and the neopterous orders, then including the Orthoptera (grasshoppers), which can. The evidence is that holometabolous insects (with the larva and imago stages very markedly different in form, and separated by a resting pupa stage) already existed.

Smart & Hughes (1972) relate this phenomenon of the development of insect flight in a wide diversity of forms during the first half of the Carboniferous period to the appearance at that time of tall terrestrial plants in great abundance in those dense 'forests' whose débris were ultimately compressed to become the coal measures. Plant life in the Carboniferous period, although very different from the characteristically angiosperm (flowering-plant) life of the Tertiary era, included 'trees' of the order of 30 m tall; lycopods, calamites and early gymnosperms, including trees with expanded leaf-laminae and many other food sources suitable

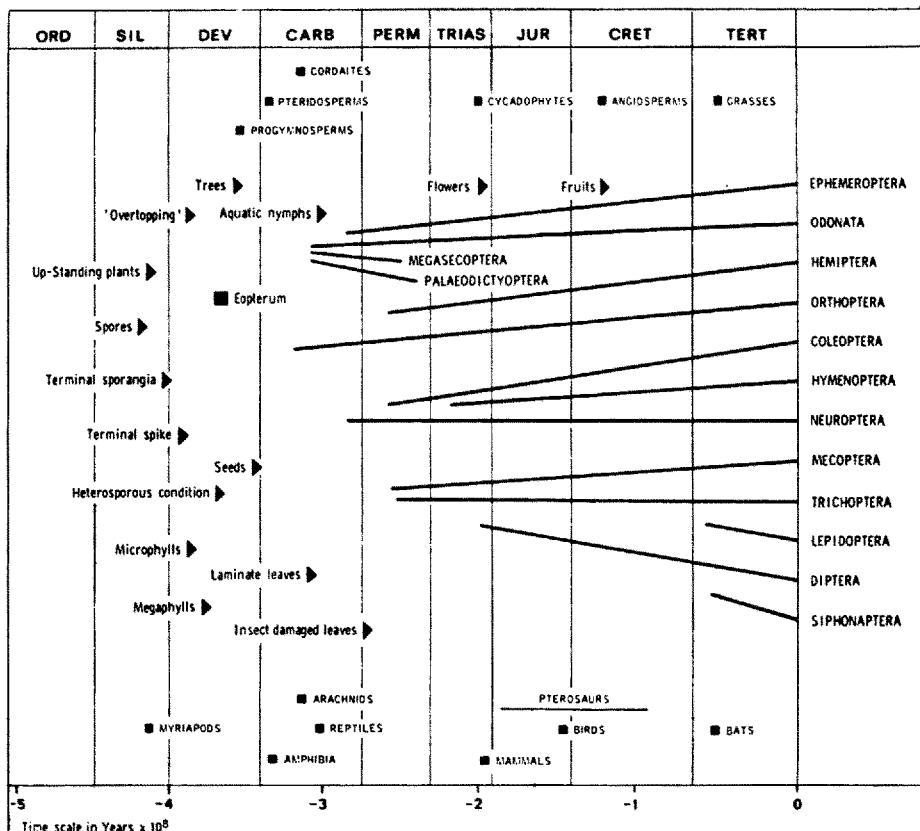


FIG. 1. This Figure 3 from Smart & Hughes (1972) summarises diagrammatically certain matters relevant to the evolution of animal flight.

for crawling insects. These are conditions which could have favoured the evolution of aerodynamic surfaces, initially to facilitate horizontal dispersal from elevated positions through gliding. Advantages in food-seeking and dispersal could then continue to accrue from a developing articulation of those surfaces through a series of stages leading to sustained winged flight.

Specialisations might then follow: for example, to predatory forms like the Odonata; and continued evolutionary development of flying in different insect groups must have resulted not only from advantages in respect of food-seeking and dispersal but also from the pressures brought about by such predators, as well as by non-insect predators such as the Amphibia (who at this period became the first terrestrial vertebrates) and a newly appearing class of arthropods: the Arachnida.

After the Carboniferous period, the class Insecta continued to diversify so that structures of wide variety including examples characteristic of most of the extant orders of insects were present already in the Triassic period, long before the appearance of flowering plants. Much later the development of the angiosperms

led to their extraordinarily successful symbiosis with the insects, which encouraged an enormous degree of specialisation in the Insecta and made them the class including a greater number of known species than all the others put together, while also allowing certain new orders of insects to appear: especially the nectar-feeding Lepidoptera (moths and butterflies).

In the meantime the conquest of the air by the insects had been succeeded by a sequence of quite independent achievements of sustained flight by other groups of animals. The reptiles, having evolved from the Amphibia, later flourished remarkably in the Jurassic and Cretaceous periods (roughly from  $2 \times 10^8$  to  $1 \times 10^8$  years ago), when they included an extensive group of flying animals: the pterosaurs. A separate acquisition of the power of sustained flight was then made by the birds (again evolving from wingless forms), while certain mammals, including bats about  $0.5 \times 10^8$  years ago, later followed yet another path to the same goal.

The plan of this book, as explained in Chapter 1, does not allow for any general comparative survey, like Chapter 2 for aquatic animal locomotion, in the subsequent main areas of inquiry. Accordingly, Chapters 8 and 9 are devoted almost entirely to aerodynamic analysis of the function of flight movements observed in selected species. That analysis, however, is preceded by a few comparative comments in the remainder of this section.

Many different capabilities are needed by flying animals. Sustained forward flight demands wing motions yielding both thrust and lift; in Chapter 2, a preliminary indication of the type of motion required was given diagrammatically (Figure 7) and related to the locomotion of heavier-than-water animals with wing-like flippers such as turtles (§ 13). Section 2 below describes the corresponding motions for a typical four-winged insect capable of fast forward flight (the desert locust *Schistocerca gregaria*, from the order Orthoptera), while § 3 does the same for two bird species (a gull and a pigeon). In spite of all the obvious differences, the essential mode of motion of each individual wing has much in common with the basic approach schematized in Chapter 2, Figure 7: combining thrust with lift by superimposing on to a thrust-generating oscillatory motion an adequate positive mean angle of attack to the animal's direction of relative motion through the air.

Most flying animals are capable also of hovering flight, either for extended periods or at least for short periods during take-off and landing.† Evidently, the problems of weight support need a completely different solution for hovering flight, without any relative motion between the animal and the air.

Weis-Fogh (1973) has recently made a general comparative study of modes of hovering flight in insects, birds and bats. This has demonstrated the widespread use of a mode of motion which he calls 'normal hovering flight' with the body in a head-up (almost erect) position and the wings achieving the necessary lift by beating back and forth in a horizontal plane, preceding each beat with a wing

† In aeronautical terms, the flying animals are 'V.T.O.L.', with a few exceptions like some aquatic birds that can use a stretch of water as a runway!

rotation that allows always the same leading edge to move forwards at an appropriate angle of attack. Section 4 briefly describes the main conclusions from Weis-Fogh's quantitative analysis of normal hovering flight, and indicates also the significant new discovery in biofluidynamics to which he was led by considering species exceptional in their modes of hovering: a mechanism of lift generation not previously described which we analyse mathematically in Chapter 9.

Between these extreme capabilities, of sustained flight at a steady speed and of hovering flight in still air, several others are commonly present and confer advantages on the animals possessing them. Some, which utilise natural air-current distributions, are described in § 3 on bird flight. Manoeuvring capabilities, and flight control generally, are also remarkably well-developed in many flying animals.

It is interesting that sustained forward flight in two-winged insects, in birds and in bats is commonly achieved without any stabilising action from posterior aerodynamic surfaces. For example, the tail-feathers of birds in fast flight are normally in an *unspread*, tapered, essentially non-lifting configuration. Animals with the very detailed control of a wing's angular attitudes in three degrees of freedom required for the modes of sustained and hovering flight just described seem able actively to control the position of the centre of lift on a beating wing to avoid undesired pitching moments. Birds do spread their tail feathers, however, in other flying conditions including sharp manoeuvres, hovering and gliding.

In commenting on these matters we should emphasize that flight arrangements conferring good manoeuvrability, whether in animals or man-made vehicles, necessarily lack any high inherent stability. We may also note that a general comparative survey of flight control methods in animals from a biofluidynamic point of view would be of the greatest interest.

In the meantime we may ask the question: what special features characterize the flight of the birds and of the insects, which have been the two most successful classes of flying animals? Probably the best answers to this question would not emphasize the biofluidynamical features, which are to some extent common among the different classes, but certain structural aspects.

The birds as a class possess many peculiar structures advantageous for flight: for example, the distribution of air sacs, besides being part of a highly efficient respiration system (Chapter 11), helps to combine lightness with structural stiffness. Above all, feathers appear to be most subtly adapted to the problems of load distribution in a light wing structure.

Some biofluidynamical advantages also are derived from feathers. In contrast to the impermeable wings of insects, pterosaurs and bats, a bird's wing with its variable permeability conferred by feathers allows for better aerodynamic performance at high Reynolds numbers (§ 3) through boundary-layer control by principles akin to the aeronautical use of leading-edge slots.

Attempting similarly to pick out a special advantage for flying purposes of the insects as a class over all the vertebrate classes, we may perhaps highlight their capability for high-frequency tuned oscillations. This avoids the need for a separate muscular effort in each wing beat to overcome wing inertia. Elastic forces do this

in the insects generally, and muscular effort only has to overcome aerodynamic damping.

All this was made possible through the evolution by the insects of an external skeleton consisting of a hard cuticle with good elastic properties, as well as of connecting tissues incorporating a high-quality elastomer, resilin (Weis-Fogh 1960), with something of the structural characteristics of vulcanized rubber. Resilin responds elastically, with very low internal damping, to large-amplitude oscillations at frequencies of the order 20 to 500 Hz characteristic of insect-wing oscillations.

The vertebrates, for all their numerous advantages, have been impeded in the development of high-frequency oscillatory movements for locomotion or other purposes by failure to evolve a suitably elastic skeleton or any elastomer of even approximately comparable qualities to those of resilin. The dominant elastomer in vertebrate physiology is the elastin which gives distensibility to mammalian arteries. We shall see in Chapters 12 and 13 that its elastic properties, though adequate at frequencies characteristic of pulse beats, are relatively poor at frequencies of the order of 100 Hz.

Vertebrate musculature, in general, has to expend energy to overcome inertia of limbs in motion, and this has militated against the development of high-frequency modes of locomotion in the vertebrates, although a few specialised groups such as hummingbirds (§ 4) or seahorses (Chapter 2, § 11) have overcome the difficulties sufficiently to be able to move wings or fins at frequencies of order 50 Hz. The much readier achievement of comparable or substantially higher frequencies through elastic systems primarily based on the properties of cuticle and resilin facilitated the development of remarkable flying and hovering capabilities in insects.

## 2. Sustained forward flight of an insect

We now describe the wing motions of the desert locust *Schistocerca gregaria* as observed and analysed in extensive detail by Weis-Fogh & Jensen (1956): a zoologist and an aerodynamicist working in an extremely effective collaboration. For their observations a wind-tunnel was found an invaluable tool, just as it is in engineering aerodynamics.

They suspended the animal in the wind-tunnel from a force balance, having selected a point of suspension avoiding interference with the motion of the thorax (to which the four wings are attached). A feedback control system, whereby the force measured by the balance controlled the speed of the wind-tunnel fan, allowed conditions representative of free sustained forward flight to be produced: the wind speed, once the insect was beating its wings regularly, was automatically adjusted so that the net horizontal force measured by the balance was zero. Thrust was then balancing drag as in free flight.

In each experiment the mean value of the net vertical force was also deduced from the readings of the force balance. Cases when this was zero represented the obviously interesting condition of steady horizontal flight, with the mean lift of the wings balancing the actual weight of the insect. Other interesting cases were

also studied, however, involving a mean net upward or downward force of up to 70% of the weight of the insect. Such cases could represent conditions with a vertical acceleration of up to 0.7 g. They could also represent wing motions necessary to support the same insect under different conditions (such as a heavy egg load).

A typical weight of a specimen was 2 g and a typical wing length 4 cm. A typical flying speed  $U$  was  $4 \text{ ms}^{-1}$ , corresponding to a Reynolds number  $Uc/v$  of around 2000 based on the forewing chord  $c$ . A typical wing-beat frequency was 20 Hz, corresponding to a frequency parameter  $\omega c/U$  (see Chapter 4) of around  $\frac{1}{4}$ : quite a low value.

Sustained forward flight both in insects and in birds shows this characteristic tendency for frequency parameters based on radian frequency  $\omega$ , wing chord  $c$  and speed  $U$  of forward motion to take values substantially lower than those values of order 1 found typical of lunate-tail marine propulsion (Chapters 2 and 4). Figure 2, showing the effective motion through the air of a single forewing chord

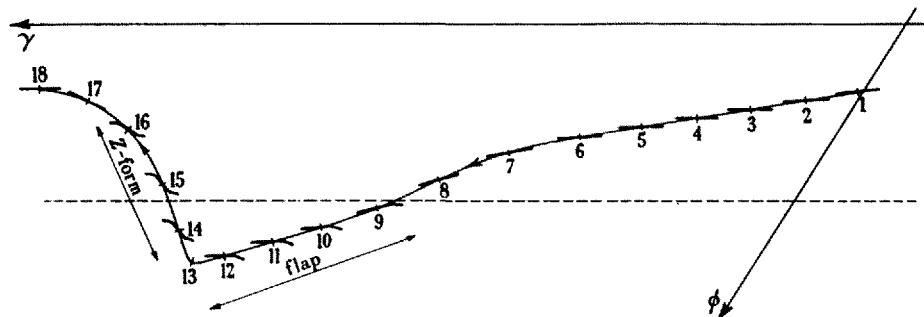


FIG. 2. This Figure III, 8 from Weis-Fogh & Jensen (1956) shows the effective motion through the air of a forewing chord in the forward flight of *Schistocerca gregaria*.

(section halfway towards the tip), makes evident the low value of this frequency parameter, effectively bringing about a quasi-steady aerodynamic response.

Note in particular the long downstroke with the chord nearly parallel to the direction of mean motion (very much as depicted in Figure 7 of Chapter 2). There is plenty of time during this stroke for the airfoil to build up its steady-state force, nearly at right angles to its instantaneous path through the air, and therefore including components of both lift and thrust. After the 'pronation' (in aeronautical terms, 'pitch-down') immediately before the downstroke, quasi-steady values of those force components are presumably attained in a distance of only a few chord lengths.

A rapid 'supination' (in aeronautical terms, 'pitch-up') marks the onset of an upstroke with the airfoil at a large positive angle of pitch to the direction of mean motion (again as in Figure 7 of Chapter 2). Note that at the same time its angle of attack to its instantaneous direction of motion relative to the air is *negative*, suggesting that, although departures from quasi-steady force conditions are

likely to be somewhat greater in this phase of motion, the main resulting force component is a modest element of additional thrust.

The wing twisting involved both in pronation and in supination is illustrated in Figure 3. This exhibits also the definition of angle of attack used by Weis-Fogh & Jensen (1956), in terms of the angle between the trailing edge and the relative wind (air motion relative to wing chord).

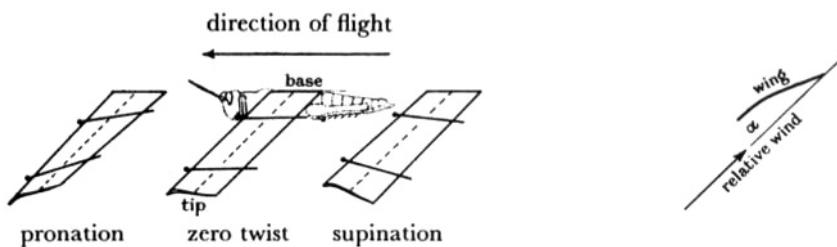


FIG. 3. These Figures I, 2 and I, 3 from Weis-Fogh & Jensen (1956) illustrate wing twisting in pronation and supination and the definition of angle of attack  $\alpha$ .

Figure 4 shows the standard wing stroke of *Schistocerca gregaria*, indicating the variation with time of a positional angle  $\gamma'$  for both the hindwing and the forewing. Here  $\gamma'$  is the angle between (i) a line stretching along the length of the wing and (ii) the plane of symmetry of the insect, and is taken as obtuse or acute according as the wing is lifted above or below its mean position.

Note how the motion of the forewings exhibits a *phase lag* behind that of the hindwings: this phase lag is characteristic of four-winged insects in general. Aerodynamic interference between the two wings was estimated by Weis-Fogh & Jensen (1956) as significant, but not of dominant importance: the effective angle of attack of the hindwing is reduced by about  $2^\circ$  through operation in the down-wash from the forewing, whose own effective angle of attack is increased by about the same amount due to the upwash field ahead of the hindwing.

Weis-Fogh & Jensen (1956) made separate determinations of the variation of lift coefficient and drag coefficient with angle of attack to a steady relative wind for both the forewing and the hindwing. These are shown in Figure 5, with corresponding curves for two standard wings used in aeronautics: the main difference, operating to the disadvantage of the insect wings, is a higher minimum drag coefficient  $C_D$ , around 0.05 compared with values around 0.03 for the aeronautical shapes (which on the other hand are inherently much heavier).

They used these steady-state data together with measurements of angle of attack (corrected as noted above for mutual interference) to obtain estimates of the variation of the vertical component of force from each wing with time (Figure 6, in which results are shown for two different experiments I and II, with mean lifts 97% and 111% of insect weight respectively). An important check on these results was the rather good agreement (3% and 7% error, respectively) between the total

mean lift obtained from each calculation and the values actually measured. Of that mean lift, about 70% is attributable to the hindwing and 30% to the forewing. As already suggested, the lift is developed primarily in the *downstroke*.

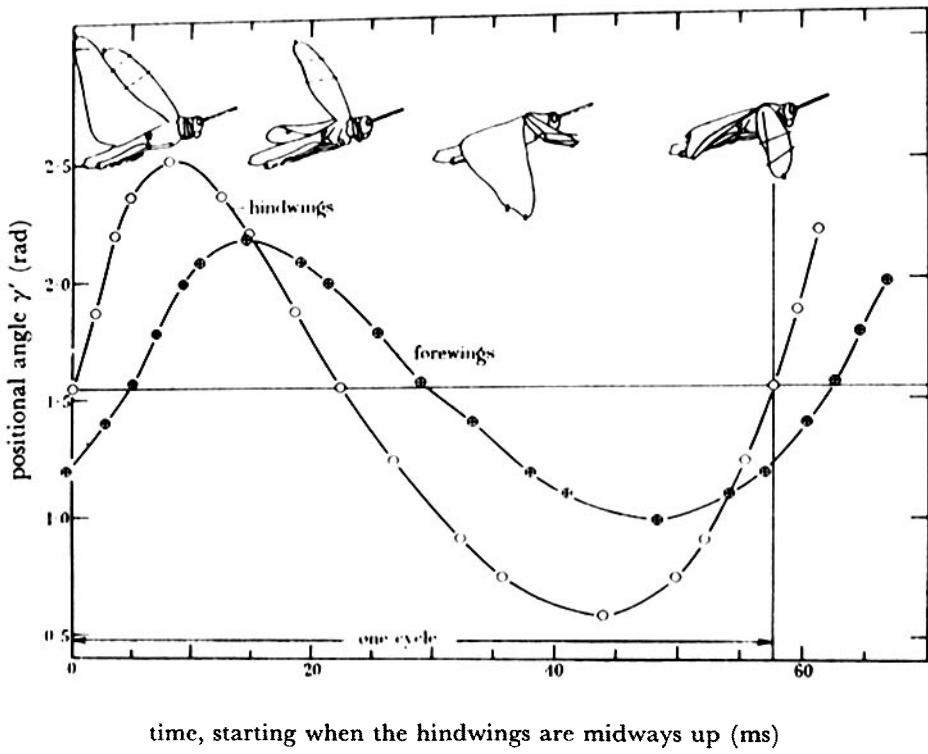


FIG. 4. This Figure II, 20 from Weis-Fogh & Jensen (1956) exhibits the hindwing and forewing motions in the forward flight of *Schistocerca gregaria*.

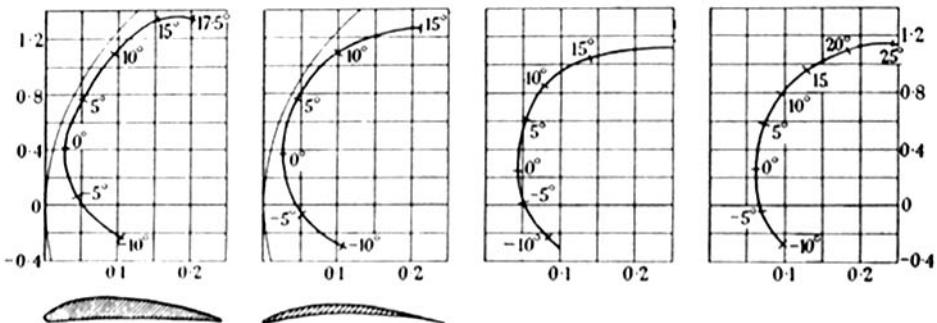


FIG. 5. This Figure I, 5 from Weis-Fogh & Jensen (1956) compares graphs of lift coefficient  $C_L$  versus drag coefficient  $C_D$  for artificial and for natural wings (locust forewing and hindwing).

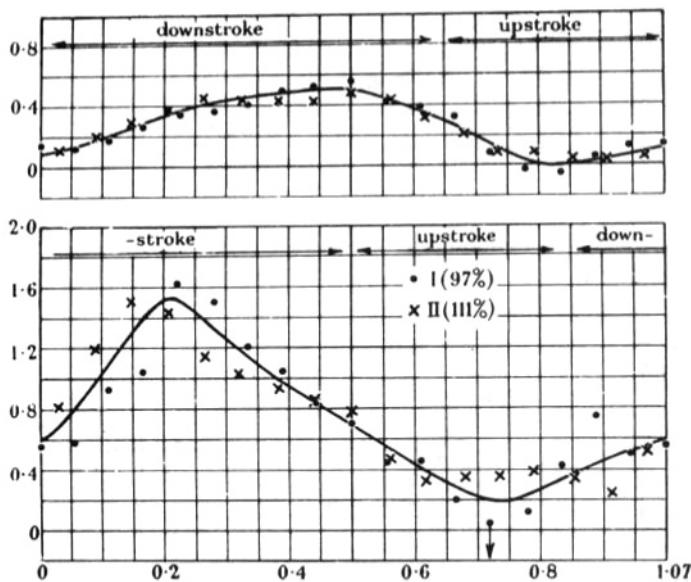


FIG. 6. This Figure III, 16 from Weis-Fogh & Jensen (1956) depicts the variation with time of the lifting force on the forewings and hindwings in the forward flight of *Schistocerca gregaria*.

Weis-Fogh & Jensen (1956) made a similar analysis of the horizontal forces (thrusts) produced by the wings, checking the analysis this time by the fact that the mean wing thrust is required to balance body drag which was separately measured. Figure 7 depicts as a function of time the thrust of each pair of wings as a fraction of the total mean thrust developed by both pairs. Note, however, that this total mean thrust is only about 7% of the mean lift. Of this total mean thrust, about three-quarters comes from the hindwings. The total thrust has a variation with time (full curve in lowest diagram) with two peaks: a larger peak during the downstroke and a smaller one during the upstroke.

In this section we have given only a brief summary of some of the more aerodynamically significant conclusions of Weis-Fogh & Jensen (1956), whose own much fuller account may perhaps be regarded as a model for investigations of animal locomotion and as such will repay very detailed study. The main biofluid-dynamic conclusion, that insect forward flight proceeds at frequency parameters small enough to make a quasi-steady type of analysis satisfactory, is of great interest and has been generally confirmed in subsequent investigations.

### 3. Bird forward flight

In § 3, by first describing classic data on the flapping flight of a gull and of a pigeon (Brown 1953), we establish certain major similarities, and minor differences, between biofluidodynamic mechanisms in the sustained forward flight of birds and of insects. Then we consider some power-consumption problems special to the larger birds, and a related matter: their acquisition of interesting techniques for power economization.

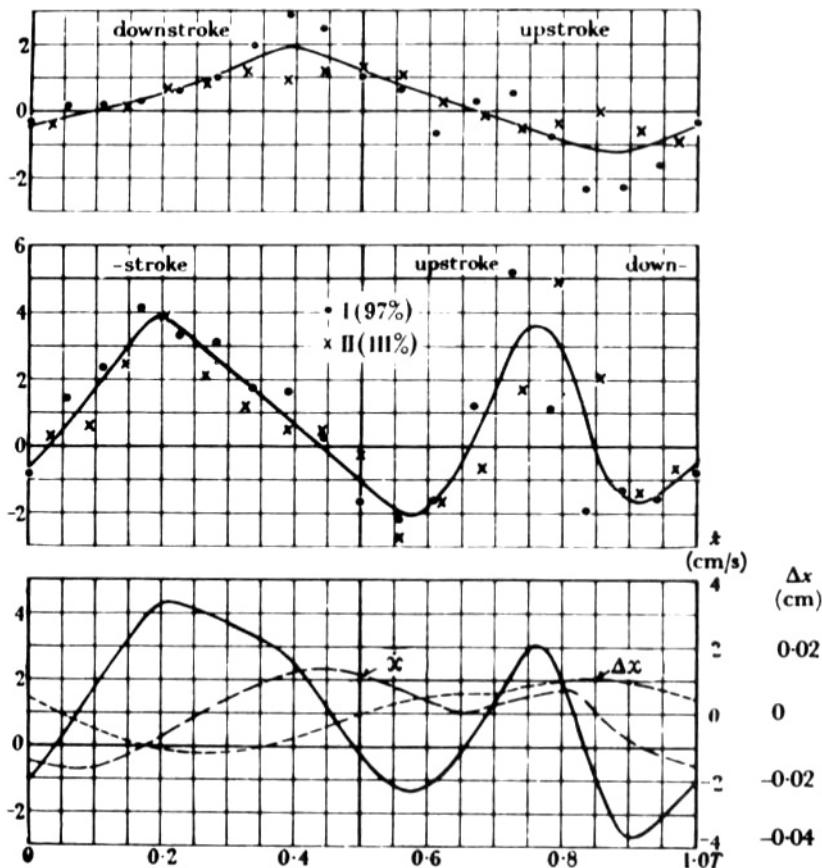


FIG. 7. This Figure III, 19 from Weis-Fogh & Jensen (1956) depicts the variation with time of forewing thrust, hindwing thrust and total wing thrust in the forward flight of *Schistocerca gregaria*.

Brown trained birds to fly straight along a passage 60 m long to a cage at one end. His high-speed camera system was set up halfway along the passage. Figure 8 shows a sequence of pictures so obtained for a pigeon *Columba* flying at around  $10 \text{ ms}^{-1}$ , corresponding to a Reynolds number of around  $10^4$  based on the wing chord  $c$ . The wing-beat frequency is about 5 Hz, corresponding to a frequency parameter  $\omega c/U$  of around  $\frac{1}{2}$ . Similar values were found also in later experiments (Pennycuick 1968) with pigeons trained to fly freely in wind-tunnels (a technique introduced also by Tucker 1968). This mode of flight, then, involves aerodynamic forces departing from quasi-steady values more than in locust flight but still only moderately.

Drawings of the pigeon's position from three mutually perpendicular directions are shown in six successive phases of the wing-beat cycle, beginning with the highest wing position, in Figure 9. This emphasizes the similarity of the downstroke, with the wing chord practically horizontal, to that depicted for the locust in

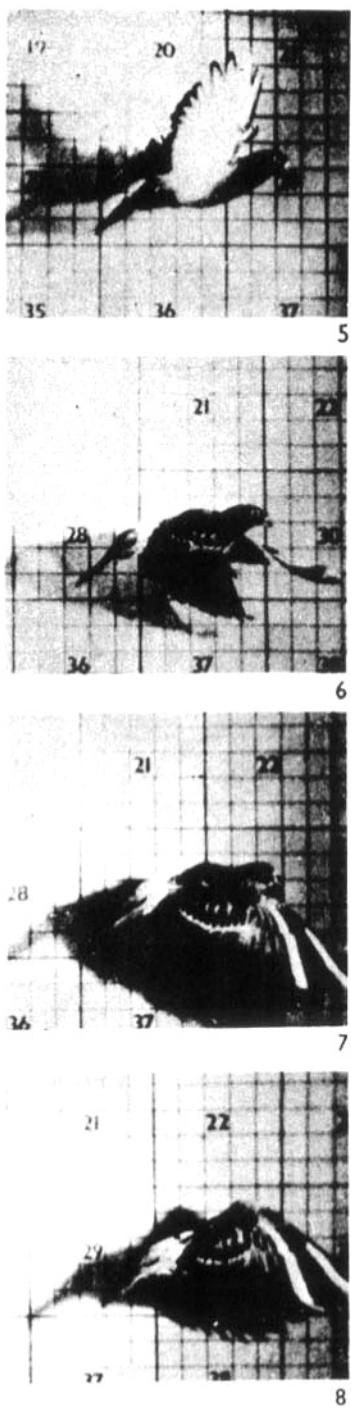
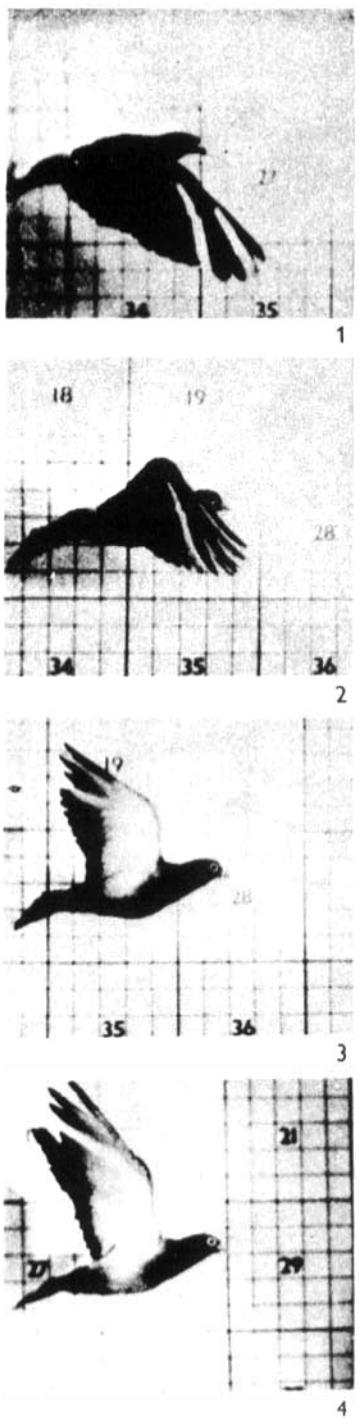


FIG. 8. Forward flight of a pigeon (Brown 1953).

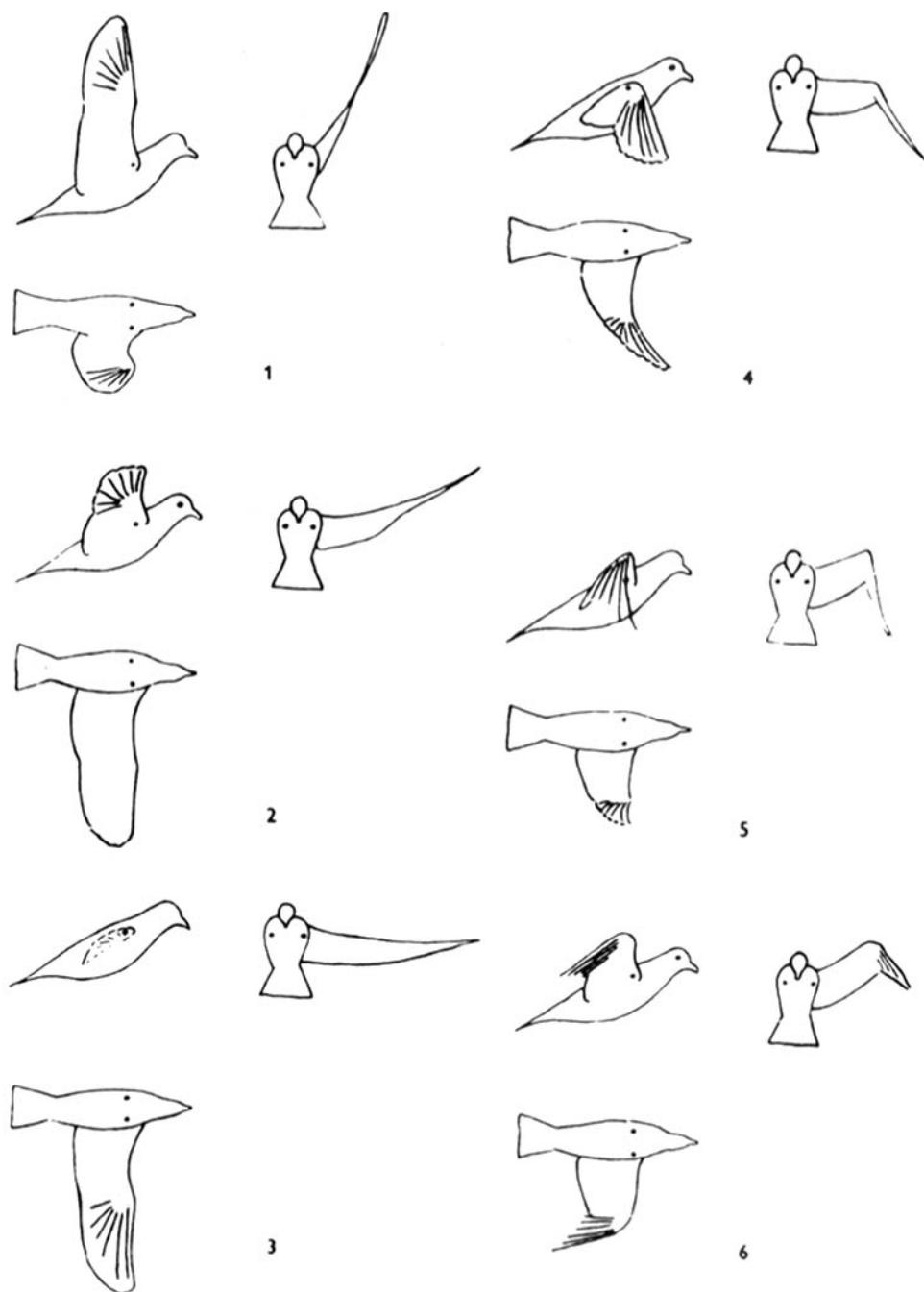


FIG. 9. These Text-figures 1-6 from Brown (1953) illustrate the pigeon's position from three mutually perpendicular directions in six successive phases of the wing-beat cycle.

Figure 2 or for 'flying' modes of aquatic animals heavier than water in Chapter 2, Figure 7. The rapid supination at the end of the downstroke, to produce a highly 'feathered' upstroke (with a large positive pitch angle to the direction of mean motion) is also just as marked as in Figure 2.

A new feature, on the other hand, is the flexure of the *wrist* which accompanies upward movement of the pigeon's wing. Probably this assists in the maintenance of significant wing thrust in the middle of the upstroke when the primary feathers (near the wing tip) are swung violently backwards.

Lift, on the other hand, is probably acquired mainly on the downstroke. Brown established the lift-enhancing role of the separation between the different primary feathers during this downstroke, seen in particular in frame 5 of Figure 8. He made wind-tunnel experiments on various typical arrangements of feathers in an inert holder, and demonstrated a 60% improvement in maximum lift for an arrangement with such a separation of the primary feathers (reminiscent of the lift-enhancing leading-edge slots used in aeronautics) over a 'solid' continuous arrangement of the feathers.

Although the pigeon uses to special effect those techniques for enhancing thrust and lift mentioned in the last two paragraphs, they are present to some extent in a wide variety of birds, as Brown (1953) emphasizes by comparative pictures for a gull *Larus*. In Figure 10 there is again a considerably greater degree of separation of the primary feathers on the downstroke (frames 1 to 5) than on the upstroke (frames 6 to 10). Figure 11 gives diagrammatic views from three mutually perpendicular directions for comparison with Figure 9: wrist flexure on the upstroke leading to a backward movement of the primary feathers is again very clearly present if a little less marked than before.

More important, however, than minor differences of biofluidynamic detail between sustained forward flight in insects and in birds are their rather bigger differences in power requirement per unit mass. Pennycuick (1968) estimated the specific power requirement for sustained pigeon flight as around 20 W/kg, amounting to over 96% of the total metabolic rate: in terms of oxygen consumption, over 130 mL/min were needed to sustain flight, compared with a resting metabolic requirement of only 5 mL/min. On the other hand, the sustained locust flight studied by Weis-Fogh & Jensen (1956) requires only about one-third as much muscular power output per unit mass of animal.

We postpone to Chapter 11 any discussion of the internal biofluidynamics of the specialised respiratory system which in birds facilitates such enormous variability of ventilation. With regard to the contrast between birds and insects, we add nothing to the discussion in § 1 on the power-economizing role of cuticle and resilin, the elastomer which permits tuned vibrations in the wings of insects. We do, however, emphasize what external biofluidynamics tells us about the necessary effects of scale changes on power requirements for flight.

As Pennycuick (1968) notes for birds, and as is familiar also to aeronautical engineers, the minimum *thrust* required for flight at mass *m* is closely proportional to *m*, being indeed equal to the weight *mg* divided by the maximum lift-drag ratio achievable with a structurally practicable aerodynamic design. The minimum



FIG. 10. *Forward flight of a gull* (Brown 1953).

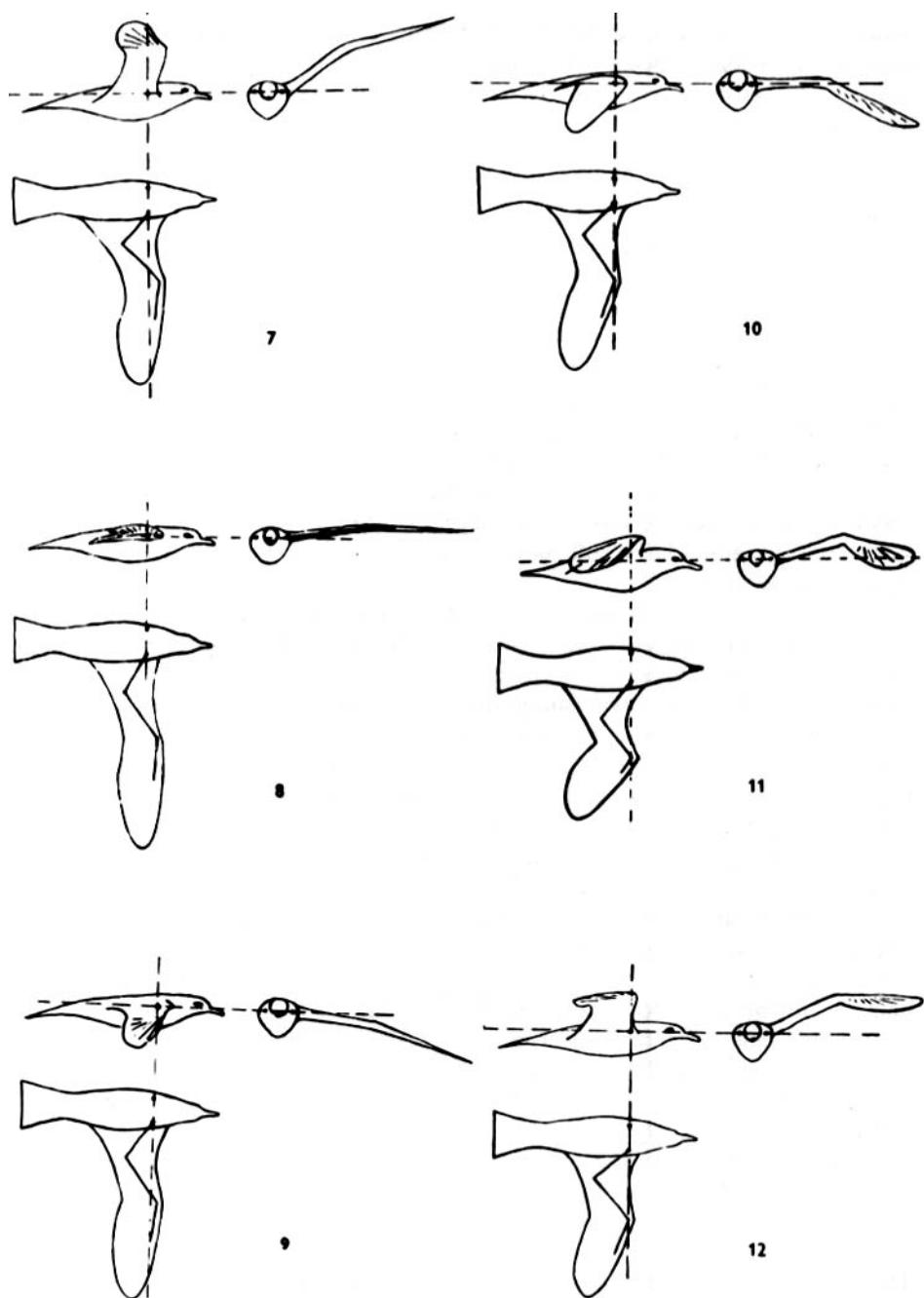


FIG. 11. These Text-figures 7-12 from Brown (1953) illustrate the gull's position from three mutually perpendicular directions in six successive phases of the wing-beat cycle.

power required to exercise this thrust, however, is thrust times velocity, which increases as a higher power of  $m$  than the first. This is because the velocity  $U$  required to satisfy a condition like maximum lift-drag ratio, which effectively fixes the lift coefficient

$$C_L = mg / (\frac{1}{2} \rho U^2 S), \quad (1)$$

varies as the square root of the wing loading  $mg/S$  (where  $S$  is wing area). For geometrically similar animals, the wing loading *increases* in proportion to the linear dimension  $l$ , since  $m$  increases as  $l^3$  and  $S$  as  $l^2$ . Hence this velocity  $U$  varies as  $l^{1/2}$ , and so does the minimum power consumption per unit mass of flying animal.

These crude considerations make it intelligible that a pigeon and a locust with wing semi-spans in ratio around 6 (say, 25 cm and 4 cm) and masses in ratio around  $6^3$  (say, 450 g and 2 g) may have typical velocities in ratio around  $6^{1/2}$  (say, 10 ms<sup>-1</sup> and 4 ms<sup>-1</sup>) and specific power consumptions in around this same ratio before allowance for special advantages like power economisation by the insect's elastic system or like the superior aerodynamic properties of the bird's wing. The same considerations have still more important consequences, however, for the flight of *larger* birds.

In fact, fundamental properties of muscle limit the maximum power that can be produced per unit mass of animal, so that if the minimum requirement for flight increases as  $l^{1/2}$  (or, what is the same thing, as  $m^{1/6}$ ) weight support by sustained forward flight becomes increasingly difficult to achieve as the linear dimension  $l$  (or the mass  $m$ ) increases. At a certain level of size, it must cease to be possible.

Observation suggests strongly that the maximum animal mass for a mode of life that includes a requirement for weight support by sustained forward flight is around 12 kg: this greatest observed mass for flying birds is approximately attained in the largest members of four different orders (Falconiformes, Anseriformes, Pelecaniformes and Gruiformes). This value, on the other hand, is much greater than the maximum mass for weight support in sustained *hovering* flight, indicated in § 4 as being only about 20 g.

The increasing 'cost' of forward-flight weight support with increase in size helps us, furthermore, to understand the fact that the larger birds have acquired remarkable capacities for power economization through 'free' use of natural upcurrents and wind-shears (Gray 1968 and references therein). We end this section by describing briefly the biofluidynamics of 'gliding' and 'soaring' birds.

A glider of mass  $m$ , whether natural or artificial, glides downwards through still air at an angle  $\theta$  to the horizontal against a drag  $D$  opposing the motion directly and a lift  $L$  at right angles to the motion, where

$$L = mg \cos \theta, \quad D = mg \sin \theta, \quad \theta = \tan^{-1}(D/L). \quad (2)$$

The angle  $\theta$  takes a minimum value achieved under the same condition of maximum lift-drag ratio noted above as minimizing thrust for horizontal flight. The aerodynamics of gliding is, however, considerably simpler than the aerodynamics of flapping flight and may here be outlined so that the results can be applied not only

to gliding motion in still air but also to cases where the glider's velocity is the vector sum of such a relative downward motion at angle  $\theta$  and an undisturbed wind velocity possessing an upward component.

Understanding of gliding behaviour needs to take into account the usual distinction between two elements in the drag  $D$  on a body developing an aerodynamic lift  $L$ : a frictional element  $D_f$  whose reaction on the air pulls a frictional wake behind the body, and an induced drag  $D_i$  whose rate of working on the air generates the energy of that growing 'trailing-vortex' wake, of air movements perpendicular to the direction of motion, whose rate of growth of momentum is required to balance the perpendicular force  $L$ . We may expect that, for larger birds at Reynolds numbers exceeding  $10^4$ , the frictional drag varies as a power of velocity not much below 2, and therefore write it as  $\frac{1}{2}\rho U^2 S C_{Df}$  where the frictional drag coefficient  $C_{Df}$  for wings and body combined is only slowly varying. The trailing-vortex wake, of perpendicular motions with momentum  $L/U$  per unit length, has mass per unit length proportional to  $\rho b^2$  where  $b$ , the semi-span of the wing, is an approximate radius for that wake. Its kinetic energy per unit length,  $D_i$ , varies in proportion to the square  $(L/U)^2$  of the momentum per unit length, divided by this mass per unit length, so we may write it as  $K L^2 / (\frac{1}{2} \rho U^2 b^2)$  where  $K$  is approximately constant. It is important to note here the inverse dependence not only on velocity  $U$  but also on the wing semi-span  $b$ , a quantity which birds are able to vary during gliding. Aeronautical measurements for a wide variety of wing shapes give values of  $K$  clustered around 0.10, somewhat greater than a theoretical minimum value of  $(1/4\pi)$ .

For analysing *gliding at a fixed angle  $\theta$* , we may take the lift  $L$  as fixed since it must balance the weight component  $mg \cos \theta$ . Then for given wing span  $b$ , the total drag

$$D = \frac{1}{2}\rho U^2 S C_{Df} + K L^2 / (\frac{1}{2} \rho U^2 b^2) \quad (3)$$

is a function of the speed  $U$  that increases either when  $U$  becomes large or when  $U$  becomes small and takes a minimum at some intermediate minimum-drag speed  $U_{md}$ . For example, on the crude approximation which ignores any variations of  $C_{Df}$  or  $K$ ,  $U_{md}$  takes that value,

$$U_{md} = \left( \frac{4K}{\rho^2 C_{Df}} \frac{L^2}{Sb^2} \right)^{1/4}, \quad (4)$$

which makes both terms in (3) equal. When  $U > U_{md}$ , the drag *increases* with  $U$  (essentially because frictional drag dominates over induced drag). Stable gliding is therefore possible, since any rise in speed would be countered immediately by the consequential increase in drag. When  $U < U_{md}$ , however, a rise in speed *reduces* the drag (because induced drag dominates over frictional drag) so that stable gliding is impossible.

We must therefore expect to observe birds gliding stably only at velocities

$$U > U_{md} \quad (5)$$

greater than the minimum-drag speed. On the approximation (4) for  $U_{md}$  obtained by taking  $C_{Df}$  and  $K$  constant this is equivalent to the condition that *induced drag must be less than half of total drag*, which demands in turn that the wing semi-span  $b$  must be large enough:

$$b > \left( \frac{KS}{C_{Df}} \right)^{1/2} \left( \frac{L}{\frac{1}{2}\rho U^2 S} \right). \quad (6)$$

As the gliding velocity  $U$  is reduced (while still, however, exceeding the *stalling* velocity at which the lift coefficient  $L/(\frac{1}{2}\rho U^2 S)$  reaches its peak possible value), a bird necessarily spreads its wings more widely, so as to continue to satisfy this condition (6).

The drag  $mg \sin \theta$ , resisting the gliding at angle  $\theta$  of a bird of mass  $m$  that can adjust its wing semi-span  $b$  subject to the above condition, can take values only between 1 and 2 times the frictional drag  $\frac{1}{2}\rho U^2 SC_{Df}$ . The speed  $U$  at which it is able to glide must therefore lie

$$\text{between } \left( \frac{mg \sin \theta}{\rho SC_{Df}} \right)^{1/2} \text{ and } \left( \frac{2mg \sin \theta}{\rho SC_{Df}} \right)^{1/2}, \quad (7)$$

and also must exceed the stalling speed. Evidently, the limits become wider as  $\theta$  increases, and could be widened further by a bird that could bring about a controlled increase in  $C_{Df}$  (rather as a pilot uses dive brakes).

For many large birds it is advantageous to be able, with minimal expenditure of energy by flight muscles, to maintain a substantial altitude above the ground for an extended period of looking for prey. Gliding offers an excellent method of achieving this aim. For example, on the windward side of a cliff or hill such a bird can often find natural winds inclined upwards at an angle  $\theta$  to the horizontal and blowing at a speed  $U$  within that range which allows the bird to glide stably downwards at the same angle  $\theta$  and speed  $U$  relative to the wind, and thus to remain stationary relative to the ground.

In hot climates, the localized regions of vertically upward airflow known as 'thermals' are similarly used, not only by glider pilots, but also by gliding birds. For example, the vultures (Aegypiinae) are expert in a gliding motion that follows an approximately *circular* horizontal path so as to remain within the thermal. If their speed and angle of descent *relative* to an upcurrent of velocity  $v$  are  $U$  and  $\theta$ , then  $\sin \theta = v/U$  which allows condition (7) to be satisfied if  $U$  lies

$$\text{between } \left( \frac{mgv}{\rho SC_{Df}} \right)^{1/3} \text{ and } \left( \frac{2mgv}{\rho SC_{Df}} \right)^{1/3}. \quad (8)$$

As  $v$  increases, the circling motion becomes possible as soon as speeds in this range (8) exceed the stalling speed.

Finally, we analyse the famous 'soaring' motion of the albatross *Diomedea* by which it varies its height above the ocean in a periodic cycle with minimal expenditure of energy by utilizing wind *shear*: that is, the variation of wind speed with

height. The essence of the required analysis was given already by Rayleigh (1883), but I am indebted to Professor S. Corrsin for the following simple way of expressing it in modern terms.

If the mean natural wind velocity is in the  $x$ -direction, with magnitude  $\bar{u}(z)$  varying with altitude  $z$  above the ocean, and if the velocity of the albatross has components

$$(\bar{u}(z) + u', v', w'), \quad (9)$$

then for stalling to be avoided the kinetic energy

$$\frac{1}{2}m(u'^2 + v'^2 + w'^2) \quad (10)$$

of the albatross's motions *relative* to the local wind must be steadily maintained against the dissipative action of aerodynamical resistance. But in the theory of turbulence we learn that any fluctuating motion with velocity (9) (whether of an eddy or of an albatross!) extracts from the *mean* motion total energy (kinetic and potential) at a mean rate

$$- \overline{mu'w'} \frac{d\bar{u}}{dz}, \quad (11)$$

where the bar denotes a mean value.<sup>†</sup> But  $\overline{u'w'}$  can be large and negative, making (11) large and positive and thus maintaining the mean value of (10), provided that the relative motion of the albatross is always upwind when upwards and always downwind when downwards: precisely the condition satisfied by its observed soaring movements in a path like a spiral around a downwind axis but greatly sheared backwards so that the upward motions relative to the air are upwind.

#### 4. Hovering flight

We have seen that forward flight of an animal at velocities (relative to the air) above its stalling speed can admirably support its weight, although larger animals may need to exert large powers per unit mass to maintain that relative velocity unless they can utilise for the purpose energy from upcurrents or wind shears. Enormously greater demands for power consumption are made, however, on an animal of given mass in *hovering flight*: that is, in motions that can support its weight without any significant horizontal movement (relative to the air) of the animal as a whole. Indeed, the hovering motions described in this section can be *continuously* sustained (under conditions of oxygen balance) only by small animals of which the largest (such as the hummingbird *Patagona gigas*) are of mass around 20 g. The same hovering motions are of wider importance, however, because larger animals are able anaerobically to develop the power required to hover thus for the very short times that are involved in take-off and landing.

In his general review of sustained hovering flight, Weis-Fogh (1973) has identified an essentially common pattern of what he has named 'normal hovering' in

---

<sup>†</sup> Essentially, (11) is the mean rate of working by an *inertial force* equal to  $(-m)$  times the acceleration ( $w' d\bar{u}/dz, 0, 0$ ) of that frame of reference in which the albatross velocity is  $(u', v', w')$ .

the hummingbirds generally (Trochilidae) and in insects from eleven different orders, of which four (the Coleoptera, Lepidoptera, Hymenoptera and Diptera) are particularly discussed below. Biofluidodynamically, the motions of 'normal hovering' may be recognised as an adaptation of those flapping motions that are used for forward flight. Lift for weight support requires horizontal movements of wings relative to the air: therefore, in the absence of any horizontal movement of the animal as a whole relative to the air, the wings must themselves beat horizontally back and forth. Evidently wings that for forward-flight purposes beat in a plane perpendicular to the horizontal body axis are readily adapted to this requirement if the body axis becomes erect (that is, practically vertical) in hovering flight.

Figure 12 depicts normal hovering flight in the sphingid moth *Manduca sexta* viewed from above. As just suggested, the body axis is practically vertical, which

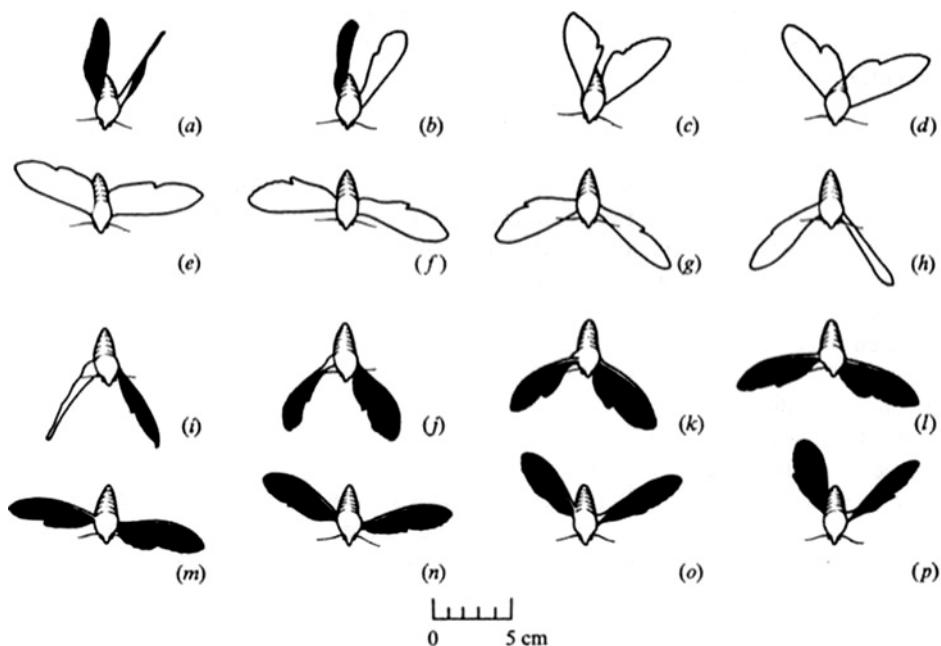


FIG. 12. This Figure 8 from Weis-Fogh (1973) depicts normal hovering in *Manduca sexta*.

facilitates beating of the wings in an approximately horizontal plane. In order to ensure that the wing leading edge moves forward always at an effective angle of attack within the range for good lift-drag ratio, the twisting movements depicted for flapping flight in Figure 3 are necessarily much enhanced: the usual pronation at the start of the downstroke is greatly intensified (frames (a) and (b)), leading to a horizontal wing movement (frames (c) to (g)) with the upper surface uppermost at a geometrical angle of attack of less than 30°. (The effective angle of attack is

still smaller because the wings move in the general downdraught produced by the animal's hovering activity.) Next, the usual supination at the start of the upstroke is greatly intensified (frames (h) and (i)), so that the wing with its *underside* uppermost (and thus with the *same* leading edge leading!) follows in reverse the previous horizontal path at approximately the same angle of attack.

This very widely found 'normal hovering' is a rather logical adaptation to hovering requirements of the wing capabilities needed in flapping forward flight. An actual transition between the two modes is interesting to watch in large birds as they come in to land: while the body becomes more erect and the tail-feathers are spread to provide braking, the stroke plane of the wing-beat changes from vertical to horizontal and the twistings of the wing are intensified so that the bird's weight just before landing is being supported by the lift on the horizontally beating wings. In *take-off*, which requires substantial *thrust* for horizontal acceleration, normal hovering is almost immediately superseded by an interesting blend with the forward-flight mode: here the stroke plane is diagonal so that even at a low forward speed the wings can develop both thrust and lift without stalling. The slow forward flight of a pigeon described by Brown (1948) exhibits this blend very clearly.

In the rest of this chapter, however, we omit further reference to modes of motion intermediate between fast forward flight and hovering flight. For hovering flight, furthermore, we do not describe those subtle internal arrangements of musculature and skeleton that permit the execution of complicated hovering motions like those of Figure 12 at wing-beat frequencies of order 30 Hz for such moths, or for typical hummingbirds, and indeed at frequencies an order of magnitude greater for small Hymenoptera and Diptera. We concentrate, rather, on biofluidodynamic problems of how the forces developed, whether in 'normal hovering' or in various modifications thereof, are able to support the animal's weight.

Weis-Fogh (1973) makes a wide survey of the observational data on animals capable of sustained hovering flight. His conclusion is that the hummingbirds and most of the insects adopt the motions of 'normal hovering' illustrated in Figure 12, with the following principal exceptions:

- (i) certain very small insects (of which the best documented is the chalcid wasp *Encarsia formosa*) which in a modified form of 'normal hovering' clap their wings together dorsally once per beat, with effects discussed in detail below and in Chapter 9;
- (ii) within the Lepidoptera, the true butterflies (Papilionoidea), whose flight movements appear still more subtle and complex, while again including the power to clap the wings together (ventrally as well as dorsally);
- (iii) the Odonata (dragon-flies) which as is well known hover with an almost horizontal body axis making a complicated coordinated motion of their four wings;
- (iv) within the Diptera, the subfamily Syrphinae or 'true hover-flies' with a similar habitat to the dragonflies and a similar capability of hovering with an almost horizontal body axis but with a quite different complicated motion of their single pair of 'sweptback' wings.

Among the very extensive data analysed by Weis-Fogh (1973) in the process of arriving at these conclusions, a substantial body of direct cinematographic observation on hovering motions was supplemented by some more circumstantial evidence: namely, calculations of the mean lift coefficient  $\bar{C}_L$  which wings of the shape observed would need to develop to support the animal's weight in 'normal hovering' at the wing-beat frequencies observed acoustically and the stroke-angle amplitudes that are visible to the naked eye. Figure 13 shows some of the results of these calculations, but with asterisks attached to insects falling within those exceptional groups (i) and (ii) above (as well as to bats, not treated in this chapter at all) for which figures calculated on the assumption of 'normal hovering' are not expected to be relevant and are therefore shown only in brackets.

	Average coefficient of lift, $\bar{C}_L$	Reynolds number, ( $Re$ )
Bats: * <i>Plecotus auritus</i>	(1.3)	14000
Birds: hummingbirds, Trochilini		
<i>Archilochus colubris</i>	1.9	6400
<i>Amazilia fimbriata fluviatilis</i>	2.0	7500
<i>Patagona gigas</i>	1.8	15000
Coleoptera: beetles		
<i>Melolontha vulgaris</i>	0.6	4700
<i>Amphimallon solstitialis</i>	0.7	3000
<i>Helicopris</i> sp.	0.6	17000
<i>H.</i> sp.	0.5	23000
<i>Cetonia aurata</i>	0.5	4300
Cerambycid species	1.1	1600
Lepidoptera: butterflies, moths		
<i>Pieris napi</i> **	(2.2)	1400
<i>P. brassicae</i> **	(0.9)	4000
<i>Sphinx ligustri</i>	1.2	6300
<i>Manduca sexta</i>	1.2	6700
<i>M. sexta</i>	0.9	6100
<i>Macroglossum stellatarum</i>	1.1	2800
<i>Amathes bicolorago</i>	1.6	1600
Hymenoptera: wasps, bees		
<i>Vespa crabro</i>	0.8	4200
<i>V. vulgaris</i>	0.8	1600
<i>Bombus terrestris</i>	1.2	4500
<i>B. lapidarius</i>	0.9	3700
<i>Apis mellifica</i>	0.8	1900
<i>Encarsia formosa</i> (Chalcid wasp)***	(2.4)	18
<i>E. formosa</i> (Chalcid wasp)***	(3.2)	15
Diptera: crane flies, mosquitoes & flies		
<i>Tipula</i> sp.	0.8	770
<i>Theobaldia annulata</i>	0.7	480
<i>Aedes aegypti</i>	0.6	170
<i>Eristalis tenax</i>	0.9	2000
<i>Calliphora erythrocephala</i>	1.3	1000
<i>Drosophila virilis</i> ***	1.0	210

FIG. 13. Calculations of average lift coefficient  $\bar{C}_L$  required by various species for weight support in normal hovering flight, and of a characteristic Reynolds number based on wing chord for the associated horizontal wing motions (Weis-Fogh 1973).

Outside those exceptional groups, Figure 13 gives strong support to the view that the motions of 'normal hovering' can support the weight of each of the animals concerned without requiring the production of improbably high mean lift coefficients. Among the insects, even with their impermeable wings, it is quite reasonable to expect a mean lift coefficient of 1.2 as a specimen of *Manduca sexta* of mass 2 g needs for the flight pattern illustrated in Figure 12. A similar value suffices for the normal hovering flight of a large bumble bee (specimen of *Bombus terrestris* of mass 0.9 g) although previous authors had questioned whether its hovering flight could be explained on ordinary aerodynamic principles. Rather smaller values suffice for the Coleoptera (only 0.5 for the biggest elephant-dung beetle studied: a specimen of *Helicocoris* of mass 13 g) and also for Diptera (including, at the opposite end of the mass scale, only 0.6 for the yellow-fever mosquito: a specimen of *Aedes aegypti* of mass 1 mg). Hummingbirds need higher mean lift coefficients, around 1.9; we may note, however, that their arrangement of primary feathers is eminently suitable for achieving that enhancement of maximum lift described in § 3 as resulting from a separation of the primary feathers that slow-motion pictures reveal them as actually adopting in flight: the 60% enhancement of  $C_L$  there quoted would indeed raise 1.2 up to 1.9.

The estimations of  $\bar{C}_L$  given in Figure 13 demanded for each animal a computation of what will here be called  $S_2$ , the second moment of the wing area about its hinge axis. Indeed, for given angular velocity  $\Omega$  of wing beat the lift developed per unit area of wing can be written as  $\frac{1}{2}\rho U^2 C_L$ , where  $U$ , the horizontal velocity of a small area of wing through the air, is  $\Omega$  times its distance from the hinge axis. Hence the total lift can be written as  $\frac{1}{2}\rho\bar{\Omega}^2 S_2 \bar{C}_L$ , where  $\bar{C}_L$  is a weighted mean of the sectional lift coefficient  $C_L$ . But for given frequency and angular amplitude of wing beat, assumed to follow a sine curve with respect to time, a time-average  $\bar{\Omega}^2$  is readily derived. Hence a value of  $\bar{C}_L$ , involving now a further weighted averaging with respect to time, is inferred from the force-balance equation

$$\frac{1}{2}\rho\bar{\Omega}^2 S_2 \bar{C}_L = mg. \quad (12)$$

Putting aside other aspects of 'normal hovering' of great interest studied quantitatively by Weis-Fogh (1973), including structural implications of the associated bending moments at the wing root and a detailed comparative study of the heavy power-consumption requirements, we conclude this chapter with some further reference to the exceptional modes of hovering numbered (i) to (iv) above. Weis-Fogh brilliantly elucidated (see below) the role of the 'clap' and 'fling' processes in *Encarsia formosa*: that is, the clapping of the wings 'behind the back of' the insect and their subsequent 'flinging open'. Now the way is open to research on the biofluidynamics of butterflies, which on take-off certainly fling their wings apart from a clapped-together position, and in flight may perhaps achieve important enhancements of lift involving 'clap' and 'fling' motions: note that *Pieris napi* appears according to Figure 13 as if it can achieve enhancements of  $\bar{C}_L$  to as high as 2.2.

Crude estimates of effective  $\bar{C}_L$  for the very different hovering motions in Odonata and in Syrphinae are just as high. Certain very interesting suggestions regarding special lift-enhancing mechanisms in the Syrphinae, involving a postulated 'flip' mechanism not unrelated to the 'fling' in *Encarsia formosa*, are made by Weis-Fogh (1973); their further study will undoubtedly be of great interest. Beyond that lies the challenging field of the Odonata, with profound interactions between the movements of their four wings, probably requiring further complexities of biofluiddynamic analysis.

*Encarsia formosa* is a small parasitic wasp widely used in the biological control of greenhouse aphids: Figure 14 indicates its general morphology and also its

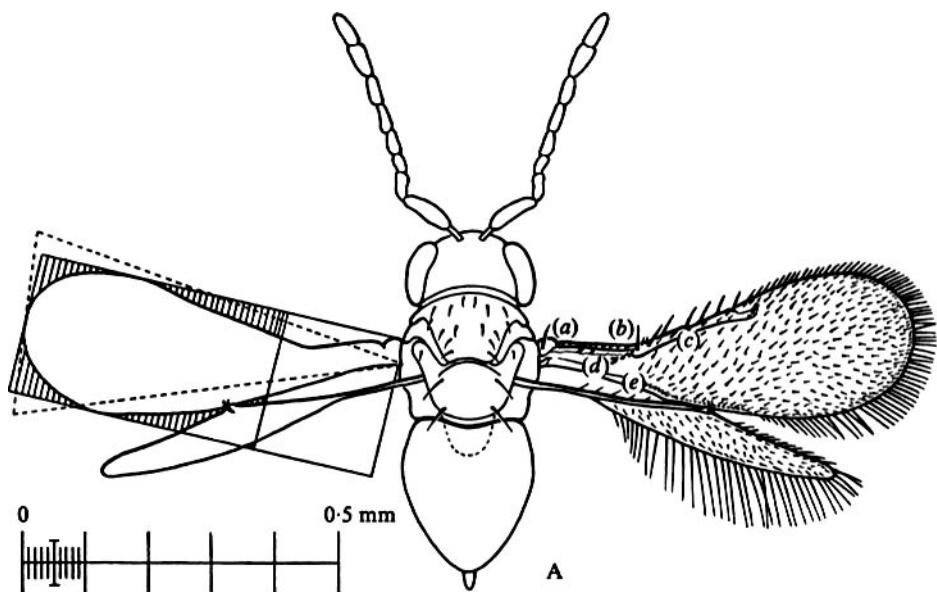


FIG. 14. This Figure 13(A) from Weis-Fogh (1973) indicates the morphology of *Encarsia formosa*.

very small wing semi-span of 0.7 mm. Note the 'hooklets' which make a sliding connection between the forewing and hindwing and ensure that in the principal flying motions of this insect those wings beat as one. Note also the 'bristles' on the wing surfaces: these probably prevent the wings from *adhering* to opposite wings when they are clapped together; they may also assist in promoting the boundary-layer reattachment whose importance is stressed in Chapter 9, § 3.

Professor Weis-Fogh's remarkable slow-motion pictures of the hovering motion of *Encarsia formosa*, taken at over 7,000 frames per second, show what a considerably more complicated wing-beat cycle even than that of Figure 12 this insect succeeds in performing 400 times a second. Figure 15 shows tracings from successive frames of that film. Note again the erect body of the insect during hovering;

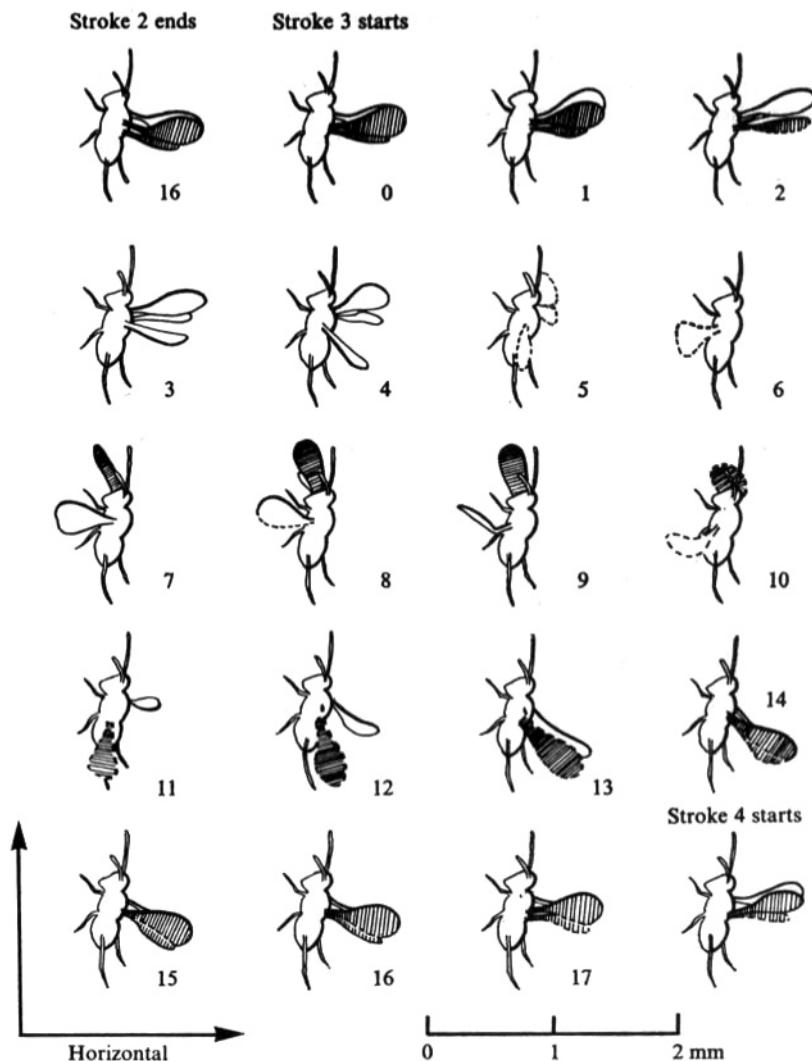


FIG. 15. This Figure 14 from Weis-Fogh (1973) shows tracings from successive frames of a slow-motion picture of *Encarsia formosa* in hovering flight.

note also the positions of the legs which indicate that in these pictures the right-hand side, where the insect claps its wings, is the *dorsal* side of the insect's erect body.

The sequence begins (frames 16 and 0) with the wings temporarily at rest in the clapped-together position. The 'fling' then occurs (frames 1 and 2): a rapid rotation of the effectively single pair of wings (see above) about a common axis along their trailing edges. The rest of the motion (frames 3 to 13) is as in 'normal hovering' (Figure 12) until the next 'clap' occurs (frames 14 to 16).

These motions have the property (see Figure 13) that when analysed as 'normal hovering' they appear to require mean lift coefficients around 3. Furthermore, a representative Reynolds number based on the wing's chord and on its velocity during the horizontal 'normal hovering' phase of motion is under 20, although the Reynolds number based (Chapter 9, § 1) on the speed of the leading edge during the 'fling' is 30. These are extremely low values for wings requiring to achieve exceptionally high lifts. Figure 16 shows the variation with time of the positional

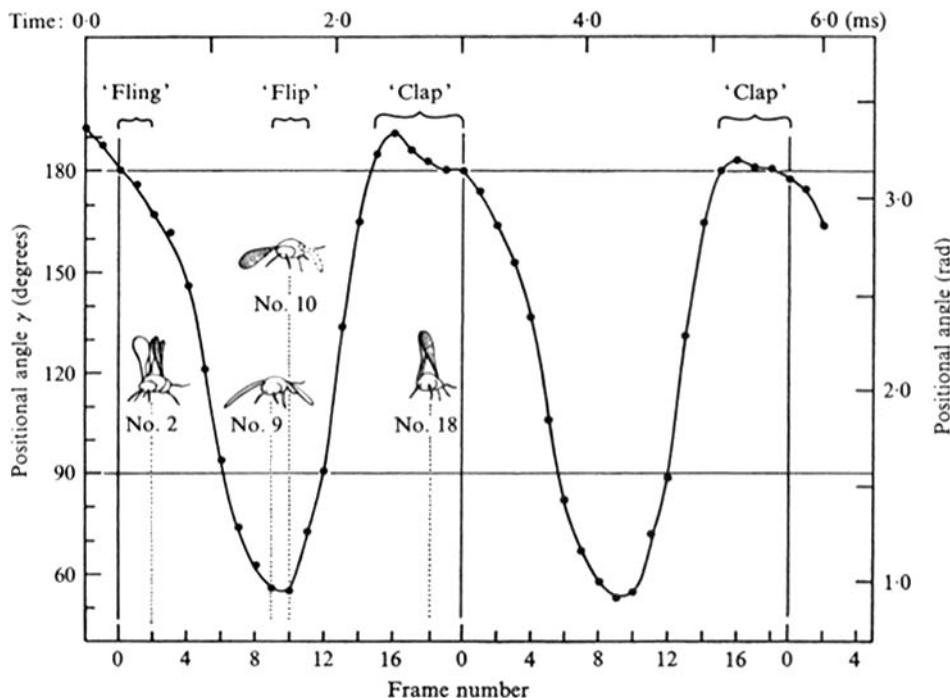


FIG. 16. This Figure 15 from Weis-Fogh (1973) shows the variation with time of the wing's positional angle  $\gamma$  in the hovering flight of *Encarsia formosa*.

angle  $\gamma$  similar to the  $\gamma'$  of Figure 4. It indicates the durations of the 'clap' and 'fling', and of the subsequent roughly sinusoidal horizontal motions of 'normal hovering', interrupted by the 'flip': a very rapid supination similar to that in Figure 12.

Weis-Fogh (1973), after considering in depth the implications of the 'clap' and 'fling' motions, proposed a fundamentally new biofluidodynamic mechanism by which they may operate to generate instantaneously an exceptionally large wing lift. The Weis-Fogh mechanism of lift generation is of such fundamental interest, and potential importance in other animal-flight problems (for example, the biofluidynamics of butterflies), that Chapter 9 below is devoted to a detailed mathematical analysis of its mode of operation. In the meantime, we conclude Chapter 8

with Weis-Fogh's own diagrammatic summary (Figure 17) of how the 'fling' may be able to generate instantaneously the large circulation required for a high level of lift generation: essentially, by means of the inward rush of air into the opening gap.

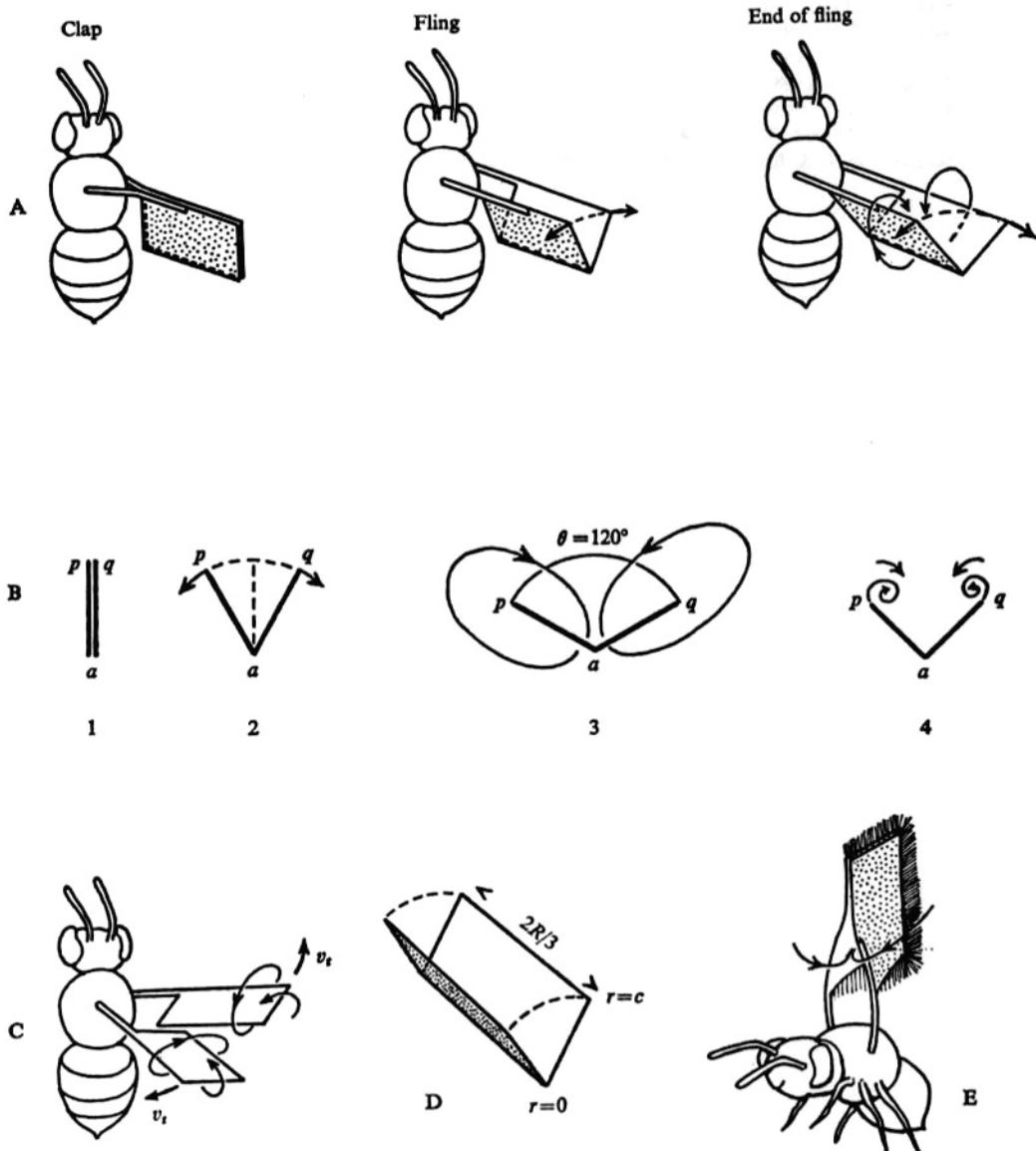


FIG. 17. This Figure 21 from Weis-Fogh (1973) summarises diagrammatically his interpretation of the 'fling' mechanism in *Encarsia formosa*.

## REFERENCES

- BROWN, R. H. J. 1948 *J. Exptl. Biol.* **25**, 322–333.  
\_\_\_\_\_, 1953 *Ibid.* **30**, 90–103.
- GRAY, J. 1968 *Animal Locomotion*, Weidenfeld & Nicolson, London.
- PENNYCUICK, C. J. 1968 *J. Exptl. Biol.* **49**, 527–555.
- RAYLEIGH, LORD 1883 *Nature* **27**, 534–535.
- SMART, J. & HUGHES, N. F. 1972 In *Insect/Plant Relationships*: Sympos. R. Entomol. Soc. Lond. no. 6, 143–155.
- TUCKER, V. A. 1958 *J. Exptl. Biol.* **48**, 67–87.
- WEIS-FOGH, T. 1960 *Ibid.* **37**, 889–907.  
\_\_\_\_\_, 1973 *Ibid.* **59**, 169–230.
- WEIS-FOGH, T. & JENSEN, M. 1956 *Proc. Roy. Soc. B*, **239**, 415–584.

## CHAPTER 9

### On the Weis-Fogh Mechanism of Lift Generation\*

Weis-Fogh (1973) proposed a new mechanism of lift generation of fundamental interest. Surprisingly, it could work even in inviscid two-dimensional motions starting from rest, when Kelvin's theorem states that the total circulation round a body must vanish, but does *not* exclude the possibility that if the body breaks into two pieces then there may be equal and opposite circulations round them, each suitable for generating the lift required in the pieces' subsequent motions! The 'fling' of two insect wings of chord  $c$  (figure 1) turning with angular velocity  $\Omega$  generates irrotational motions associated with the sucking of air into the opening gap which are calculated in § 2 as involving circulations  $-0.69\Omega c^2$  and  $+0.69\Omega c^2$  around the wings when their trailing edges, which are stagnation points of those irrotational motions, break apart (position ( $f$ )). Viscous modifications to this irrotational flow pattern by shedding of vorticity at the boundary generate (§ 3) a leading-edge separation bubble, and tend to increase slightly the total bound vorticity. Its role in a three-dimensional picture of the Weis-Fogh mechanism of lift generation, involving formation of trailing vortices at the wing tips, and including the case of a hovering insect like *Encarsia formosa* moving those tips in circular paths, is investigated in § 4. The paper ends with the comment that the far flow field of such very small hovering insects should take the form of the exact solution (Landau 1944; Squire 1951) of the Navier-Stokes equations for the effect of a concentrated force (the weight  $mg$  of the animal) acting on a fluid of kinematic viscosity  $\nu$  and density  $\rho$ , whenever the ratio  $mg/\rho\nu^2$  is small enough for that jet-type induced motion to be stable.

#### 1. Introduction

Weis-Fogh (1973), in his analysis of the hovering motions of the chalcid wasp *Encarsia formosa* (an economically important parasite used in the biological control of greenhouse aphids), concluded that its performance is markedly superior to that of most hovering animals as a result of lift generation by a mechanism of considerable fundamental interest not previously studied by aerodynamicists. Normal hovering animals beat their wings back and forth in a horizontal plane, preceding each lift-beat with a wing rotation that allows always the same leading edge to move forwards at an angle of incidence appropriate to a relatively high lift coefficient. Building up that lift coefficient is, however, delayed by the Wagner effect: that is, the time required for vorticity shedding from the trailing edge to generate the necessary circulation around the wing

\* Reprinted from J. Fluid Mech., 60 (1973), pp. 1-17.

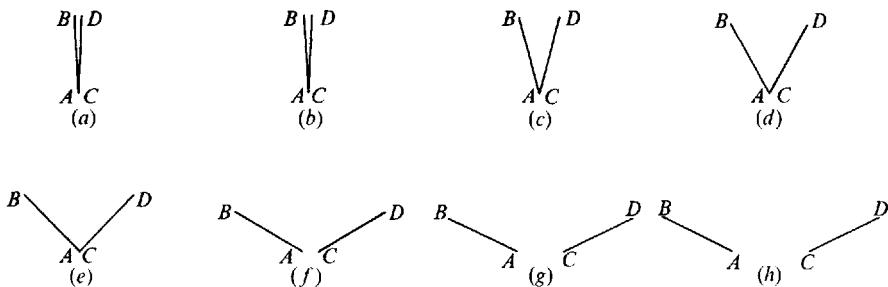


FIGURE 1. Sequence of motions of the wings of *Encarsia formosa*, shown in section by a mid-span vertical plane on the dorsal side of the insect's erect body. In positions (a) and (b) the wings are momentarily at rest after the 'clap'; positions (c)–(f) exhibit the 'fling' motion; the wings break apart at (f) and in (f), (g) and (h) exhibit normal flight movement.

(Wagner 1925). *E. formosa* precedes each beat with a special movement (the 'clap and fling') which, as Weis-Fogh argues, may cause the necessary circulation to be generated immediately and avoid any delay in the build-up of maximum lift. High-speed photography was needed to observe the details of these movements, at a wing-beat frequency around 400 Hz.

The Weis-Fogh mechanism, like the Wagner effect, can be described in terms of a purely two-dimensional flow: a description reasonably appropriate as each of the animal's wings has an aspect ratio around 5. This paper gives a quantitative analysis of the Weis-Fogh mechanism in terms of two-dimensional flow theory, but also offers qualitative comments on three-dimensional aspects of the flow patterns in the concluding section, § 4.

A particularly remarkable feature of the mechanism is that Weis-Fogh (1973) was able to describe it approximately in terms of a purely *inviscid* two-dimensional flow. This flow is calculated in § 2 below prior to the investigation of viscous-flow corrections in § 3. It is surprising that a fundamentally new mechanism of lift generation in inviscid two-dimensional flow should be discovered six decades after the work of Prandtl, Zhukovski, Kutta and Lanchester.

The Weis-Fogh mechanism works, furthermore, for a fluid of zero viscosity: not simply in the limit of vanishing viscosity when thin Prandtl boundary layers shed Lanchester vortices at sharp trailing edges where the Kutta-Zhukovski condition is satisfied. To be sure, for inviscid two-dimensional flow the doctrines of Helmholtz, Stokes and Kelvin tell us that a body starting to move in fluid at rest retains always the same zero circulation of fluid around it, preventing the generation of lift on the body (or of any forces except those associated with virtual-mass effects). Those doctrines do not, however, rule out the possibility that when the body *breaks into two pieces* there may be equal and opposite circulations round them, each suitable for generating the lift required in the pieces' subsequent motions!

Figure 1 illustrates in terms of two-dimensional flow the sequence of wing movements comprising the Weis-Fogh mechanism of lift generation. We view 'in elevation' successive positions taken up by sections *AB* and *CD* of the two wings in a mid-span vertical plane. The body of the hovering insect is erect with

the head uppermost, and the vertical plane of figure 1 is to the dorsal (back) side of it, halfway between the body itself and the wing tips.

The last three positions (*f*), (*g*) and (*h*) in figure 1 show a motion characteristic of normal hovering flight with both wings moving horizontally at a positive angle of incidence. It will be explained why this motion is enabled by the preceding 'clap and fling' movements to generate maximum lift from the outset; before describing that, however, we note that the normal horizontal motion of the wings is shown as a rectilinear motion in figures 1(*f*), (*g*) and (*h*), although the wings really move in a horizontal circle around the erect body. Positions showing the horizontal motion at the end of figure 1 proceeding considerably further could properly be depicted only if figure 1 were regarded as a section not by a vertical plane but by a cylindrical surface with the insect's vertical body as axis.

At a certain phase in each wing-beat cycle the insect performs the 'clap': its wings are 'clapped together behind its back' into the contiguous situation shown in the first two positions of figure 1, with leading edges vertically above the trailing edges. We may regard position (*b*) with the surrounding fluid undisturbed as the initial condition for the operation of the Weis-Fogh mechanism, on the grounds that residual eddy motions generated in the clapping process should have been blown far enough away from the wings by then to be uninfluential.

The sequence from figure 1(*b*)–(*g*), then, shows two wings *AB* and *CD* which in the first four positions are touching (with the points *A* and *C* coinciding) to form effectively a *single body*. Figure 1(*f*) shows them breaking apart, while they are fully separate in figure 1(*g*). The Weis-Fogh mechanism depends on the idea that the rotary movements of *AB* about *A* and of *CD* about *C* (that is, the 'fling') depicted in positions (*b*)–(*f*) generate a fluid motion which at the moment of figure 1(*f*) when the wings break apart involves substantial equal and opposite circulations, in the negative sense around *AB* and in the positive sense around *CD*, of magnitudes close to those required for generating maximum lift at once in the subsequent horizontal motions. This is the idea that is quantitatively evaluated on inviscid theory in § 2, while modifications due to viscous effects are estimated in § 3. The necessity of such viscous considerations is particularly evident from the low Reynolds numbers involved: around 30 based on a wing chord of 0.22 mm and a leading-edge velocity of  $2.2 \text{ ms}^{-1}$ . Finally some considerations regarding the idea's application to the fully three-dimensional motions of the insect's wings are sketched in § 4.

## 2. Two-dimensional inviscid-flow theory

In the two-dimensional model of the 'fling' process depicted in figure 2, the broken line *EF* represents a plane of symmetry while the lines *AB* and *CD* each have length *c*, the 'chord' of the wings which they represent. They each make an angle  $\alpha$  with *EF* and the 'fling' process is one in which  $\alpha$  increases from zero, with the points *A* and *C* stationary and coincident, until when  $\alpha$  takes a value  $\alpha_0$  (around  $\frac{1}{2}\pi$ ) the wings break apart.

The flow field at each instant of the fling process is, according to inviscid-flow theory, simply the irrotational flow associated with the instantaneous angular

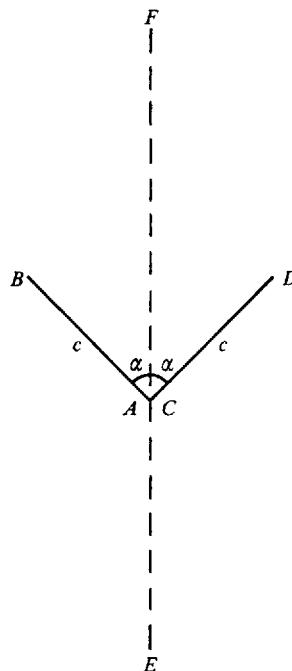


FIGURE 2. Two-dimensional model of the 'fling' process. The points  $A$  and  $C$  are stationary and coincident and the wings  $AB$  and  $CD$  rotate about them with angular velocity  $\Omega = d\alpha/dt$ . The whole motion is symmetrical about the line  $EF$ .

velocity  $\Omega = d\alpha/dt$  of rotation of the wings. In the present section we calculate this flow, but may note in advance two features of it: there is a stagnation point (zero fluid velocity) on both sides of the corner, and there is a circulation

$$\Gamma = \Omega c^2 g(\alpha) \quad (1)$$

around each wing (in the negative sense around  $AB$  and in the positive sense around  $CD$ ), where  $g(\alpha)$  is a computed function of  $\alpha$ .

It follows that if the wings break apart when  $\alpha = \alpha_0$  and  $\Omega = \Omega_0$  the flow field, not involving any motion at the corner where the break occurs, remains unchanged to a close approximation. The circulation around each wing is then  $\Omega_0 c^2 g(\alpha_0)$  and on two-dimensional inviscid-flow theory must continue to take that value in the subsequent motion. According to unsteady aerofoil theory this permits substantial lifts on both wings without the need for any vortex shedding. We postpone till § 3 considerations of possible effects modifying this irrotational-flow description, resulting from vortex shedding whether from the leading edges during the fling or from the trailing edges after they break apart, and concentrate here on determining  $g(\alpha)$ .

In the irrotational flow associated with the motion of figure 2, the line of symmetry  $EF$  is a streamline and we investigate the flow to the left of it due to the motion of  $AB$  and calculate the resulting circulation  $\Gamma$ . The investigation is carried out (figure 3) in the upper half complex  $z$  plane cut from the origin  $A$  to

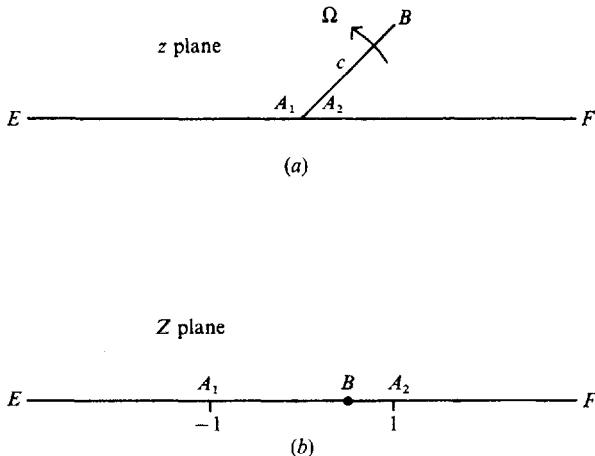


FIGURE 3. Illustrating the complex planes used: the  $z$  plane (a), representing the left-hand half of the flow field of figure 2 turned through  $90^\circ$ , is mapped conformally into the upper half  $Z$  plane (b).

the point  $B$ , where  $z = c e^{i\alpha}$ , with  $EF$  the real axis. We distinguish the two corners at  $A$  between the wing and the line of symmetry with the designations  $A_1$  and  $A_2$ .

A Schwarz-Christoffel conformal mapping from the *cut* upper half  $z$  plane to the *uncut* upper half  $Z$  plane (figure 3) is defined by

$$\frac{dz}{dZ} = K \left( \frac{Z-1}{Z+1} \right)^{\alpha/\pi} \frac{Z-a}{Z+1} \quad (2)$$

with the term in brackets given argument zero at  $Z = \infty$ . The points at infinity correspond; the points  $A_1$  and  $A_2$  are mapped into  $Z = -1$  and  $+1$  respectively, while the need for their positions in the  $z$  plane to coincide requires that the integral of (2) from  $Z = -1$  to  $Z = +1$  vanishes, giving

$$a = 1 - (2\alpha/\pi) \quad (3)$$

as the value of  $Z$  at the point  $B$ . The quantity  $z e^{-i\alpha}$  on the wing  $AB$ , which represents distance from the origin  $A$ , may be written as

$$z e^{-i\alpha} = K f(Z), \quad \text{where } f(Z) = \int_{-1}^Z \left( \frac{1-Z}{1+Z} \right)^{\alpha/\pi} \frac{a-Z}{1-Z} dZ, \quad (4)$$

and the condition that it takes the value  $c$  at  $B$  (where  $Z = a$ ) determines  $K$  as

$$K = c f_{\max}^{-1}, \quad \text{where } f_{\max} = f(a) \quad (5)$$

is the maximum of  $f(Z)$  for  $-1 < Z < 1$ .

The stream function  $\psi$ , which is the imaginary part of the complex potential  $w$ , can be taken as zero on the streamlines  $EA_1$  and  $A_2F$ . It is non-zero, however, on the wing  $AB$ , along which the normal velocity is  $\Omega$  times the distance  $z e^{-i\alpha}$  from  $A$ . The rate of change of  $\psi$  with this distance is minus this normal velocity, giving

$$\operatorname{Im}(w) = \psi = -\frac{1}{2}\Omega(z e^{-i\alpha})^2 \quad (6)$$

as the boundary condition on both  $A_1B$  and  $A_2B$ .

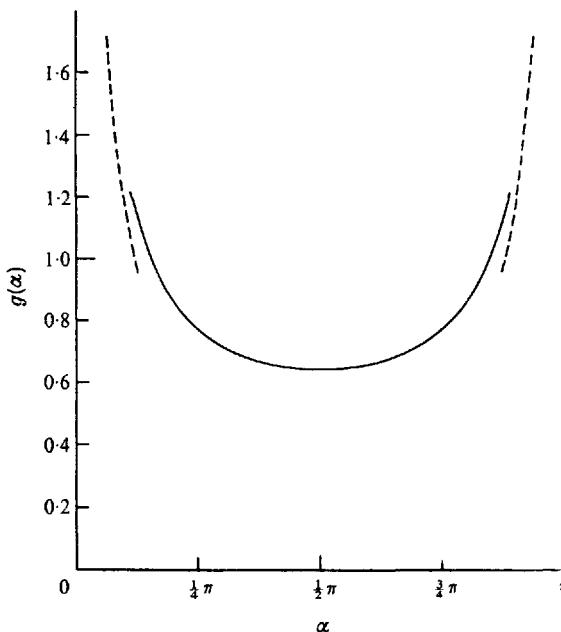


FIGURE 4. The coefficient  $g(\alpha)$  in equation (1) for the circulation  $\Gamma$  around  $AB$ , as a function of the semi-angle  $\alpha$  between the wings given by equation (10), with  $f$  and  $f_{\max}$  defined in (4) and (5). The broken line represents the approximate formula (12) for small  $\alpha$ ; a similar approximate form for small  $\pi - \alpha$  has been plotted, although it is probably of no practical interest.

The corresponding boundary conditions in the  $Z$  plane state that for real  $Z$

$$\text{Im}(w) = \begin{cases} -\frac{1}{2}\Omega K^2 f^2(Z) & (-1 < Z < 1), \\ 0 & (\text{otherwise}). \end{cases} \quad (7)$$

The complex potential  $w(Z)$  satisfying these boundary conditions is

$$w = \frac{\Omega K^2}{2\pi} \int_{-1}^1 \frac{f^2(t) dt}{Z-t}. \quad (8)$$

Now the circulation  $\Gamma$  round the wing  $AB$  in the negative sense is the change in velocity potential  $\phi$  (the real part of the complex potential  $w$ ) as we move once round  $AB$  from  $A_1$  to  $A_2$ , which in the  $Z$  plane is from  $-1$  to  $+1$ . Hence

$$\Gamma = (\Omega K^2/\pi) \int_{-1}^1 (1-t^2)^{-1} f^2(t) dt, \quad (9)$$

which with (5) specifies the coefficient  $g(\alpha)$  in (1) as

$$g(\alpha) = \left[ \int_{-1}^1 (1-t^2)^{-1} f^2(t) dt \right] (\pi f_{\max}^2)^{-1} \quad (10)$$

This function is easily computed, with results shown in figure 4.

An interesting feature of  $g(\alpha)$ , not unexpected from its expression (10) as a sort of ‘weighted mean’, is the flat nature of its graph in the central half of the interval

of possible values of  $\alpha$ : thus,  $g(\alpha)$  lies between 0.64 and 0.77 when  $\frac{1}{4}\pi < \alpha < \frac{3}{4}\pi$  (so that  $f(t)$  attains its maximum between  $-\frac{1}{2}$  and  $\frac{1}{2}$ ). In fact  $0.64 < g(\alpha) < 0.77$  whenever the angle  $2\alpha$  between the wings is obtuse, and  $g(\alpha)$  is varying particularly slowly once that angle has risen to the value  $120^\circ$  ( $\alpha = \frac{1}{3}\pi$  with  $g(\alpha) = 0.69$ ) at which the separation of the wings is observed.

This is significant not only because it indicates that the circulation  $g(\alpha_0)\Omega_0 c^2$  is insensitive to the exact value  $\alpha_0$  of  $\alpha$  when the wings break apart, but also because it suggests that the pressure difference across the gap is negligible. This difference in pressure between the stagnation points  $A_1$  and  $A_2$  is the difference in values of  $-\rho\partial\phi/\partial t$ , which is

$$-\rho d\Gamma/dt \quad (11)$$

since  $\Gamma$  is the difference in the values of  $\phi$  at  $A_1$  and  $A_2$ . On the flat part of figure 4 this quantity should be small.

It may be of interest to note that for very small values of  $\alpha$  the quantity  $g(\alpha)$  becomes large, approximately like

$$g(\alpha) \sim \frac{1}{4\alpha} + \frac{1}{2\pi} \log \frac{\pi}{\alpha}, \quad (12)$$

of which the graph is shown as a broken line in figure 4. This behaviour can be deduced analytically from (10); alternatively, its physical significance can be seen as follows.

At the start of the 'fling' process, the flow field is dominated by the inrush of air to fill the opening gap between the wings. The volume of air per unit span at a distance less than  $r$  from  $A$  is  $\alpha r^2$  and the rate of increase of this, namely

$$\Omega r^2, \quad (13)$$

must be achieved (figure 5) by inflow across an arc of small length  $2\alpha r$  at a mean speed  $\Omega r/2\alpha$ . The integral of this from 0 to  $c$  is  $\Omega c^2/4\alpha$ , which explains the leading term in (12) as the large contribution to circulation made by the integral of velocity over the region between the wings where the inflow (13) is spread over a narrow arc.<sup>†</sup> Outside the region the 'sink' flow into the opening predominates, with an inflow per unit span  $\Omega c^2$  coming equally into  $BD$  from all directions. In a distance  $c$  from  $A$  to  $B$  the area per unit span over which this inflow is spread drops from  $2\pi c$  to  $2\alpha c$  (figure 5), and the corresponding change

$$(\Omega c^2/2\pi) \log (\pi/\alpha)$$

in velocity potential accounts for the second term in (12).

Strictly speaking, these questions of what is the flow and the consequent circulation for  $\alpha$  small are not important for an inviscid-flow model, on which the circulation at a later instant when  $\alpha = \alpha_0$  is determined only by the motion of the boundary and resulting flow at that instant. They have been mentioned here, however, because the modifications by viscous effects studied in § 3 give the

<sup>†</sup> For the energetics of the 'fling' process, see Weis-Fogh (1973): the essential point is that the combination of elastic system and musculature involved is adapted to operations in which most of the energy expenditure precedes most of the displacement, as is required for small  $\alpha$  (though hairs on the wings prevent  $\alpha$  from becoming exactly zero) to generate the velocity distribution  $\Omega r/2\alpha$ .

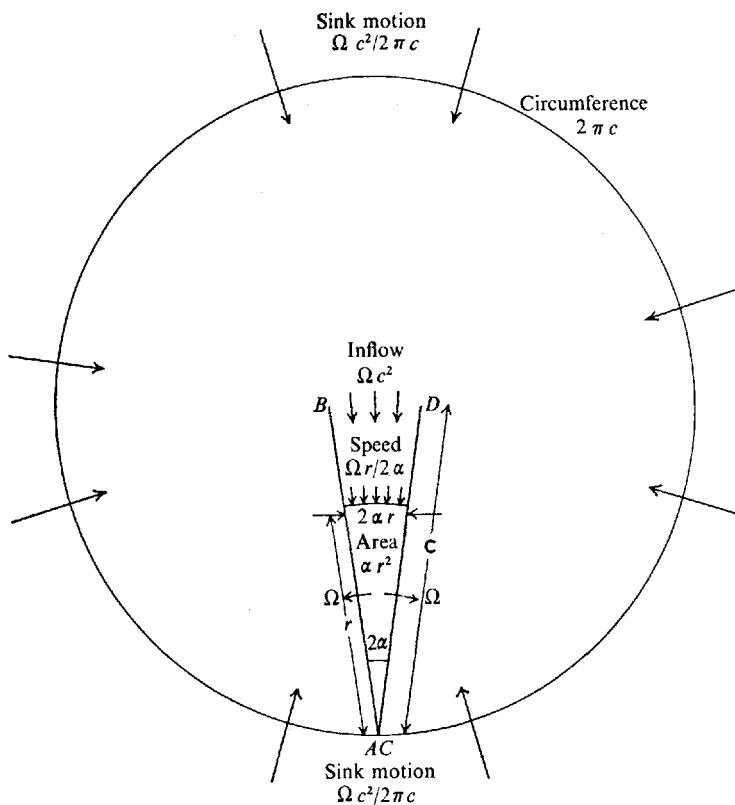


FIGURE 5. The initial stage of the 'fling' process. The area  $\alpha r^2$  of a sector of radius  $r$  and semi-angle  $\alpha$  increases at a rate  $\Omega r^2$ , which must be balanced by inflow across an arc of length  $2\alpha r$  at a mean speed  $\Omega r/2\alpha$ . The external motion, on the other hand, is that due to a sink of strength  $\Omega c^2$  representing the inflow into  $BD$ .

model 'memory': in fact, it is modified by a distribution of vorticity that has been convected and diffused since first being shed from the boundary at a rate depending on how the boundary was moving the fluid at that time.

### 3. Modifications due to viscous effects

The inviscid-flow theory of § 2 predicts that, after the wings break apart, the fluid continues to move irrotationally (that is, without vorticity) but with circulations  $\Omega_0 c^2 g(\alpha_0)$  around  $AB$  and  $CD$  in the negative and positive senses respectively. In this section we consider modifications to this conclusion, still on a two-dimensional model, due to shedding of vorticity from the wings both before and after they break apart.

If vorticity is shed from the boundary in a motion like that depicted in figure 2, the total flow field is the sum of (i) the irrotational flow (calculated in § 2) compatible with the boundary's motion normal to itself, and (ii) the flow field that the vorticity distribution would induce if the boundary were at rest. The modifying flow field (ii) can often be calculated by the method of images: in fact, when

the vorticity distribution is mapped into the  $Z$  plane of figure 3 with an infinite straight-line boundary, the associated flow field is that of the vorticity distribution plus that of a mirror-image distribution of equal strength and opposite sign.

The vorticity distribution in a two-dimensional flow is determined by the principles (i) that vorticity is subject to convection by the local fluid velocity and diffusion with diffusivity  $\nu$  (the kinematic viscosity), and (ii) that the boundary is a source of vorticity (Lighthill 1963), which appears from it at the rate required to maintain the no-slip condition. In the problem of figure 2, for example, flow changes occurring in a short time interval leave the boundary condition on normal velocity undisturbed: the irrotational component changes at once to that associated with the new attitude  $\alpha$  and angular velocity  $\Omega$  of the boundary, while the previously existing vorticity distribution adopts after convection and diffusion a new configuration but, in combination with its image system, continues to make no alteration in the normal velocity at the boundary. All these changes, by contrast, perturb the tangential velocity at the boundary, so that the no-slip condition can be maintained only if new vorticity appears at the boundary in the form of a vortex sheet of strength equal to that perturbed tangential velocity: a vortex sheet which itself begins at once to be convected and diffused.

In certain circumstances its *convection* may be essentially tangential so that all the vorticity generated remains close to the boundary, filling after time  $t$  a boundary layer with thickness proportional to a diffusion length  $(\nu t)^{\frac{1}{2}}$ . This is particularly the case wherever the irrotational component of flow involves accelerating motions close to the boundary, so that the source of vorticity must always be of one sign: that which permits the slip across the boundary layer to increase. For example, on the lower surfaces of the wings in figure 2, such an accelerating motion is found. Initially, when  $\alpha$  is small (figure 5), it is an accelerating 'sink' flow into the opening orifice  $BD$ . For all  $\alpha$ , however, the velocity calculated by the irrotational-flow theory of § 2 increases along the lower surface of each wing even for fixed angular velocity  $\Omega$ , while the fluid's acceleration is even more pronounced when  $\Omega$  is increasing with time.

This is important because, when  $\alpha$  has risen to values around  $\frac{1}{2}\pi$  where the wings break apart, about half of the total circulation  $\Omega c^2 g(\alpha)$  predicted in § 2 (with  $g(\alpha)$  in the 'flat' region of figure 4) for the circulation round each wing comes from the underside. This part, as a contribution to the circulation round the wing and the attached boundary layer, is not significantly modified, then, by viscous effects.

By contrast, wherever the irrotational component of flow involves decelerating motions close to the boundary, the vorticity required to allow its slip over the boundary decreases, which demands the generation at the boundary of vorticity of opposite sign. When that becomes sufficient to counteract diffusion of previously existing vorticity towards the surface, a reversed flow results near the surface which combines with forward flow in the region of accelerating motions to cause movement of fluid away from the surface. Such flow separation may convect vorticity to a much greater distance from the surface in a given time  $t$  than pure diffusion could.

Irrational motions involving substantial flows around a wing leading edge that is sharp or has relatively small radius of curvature are classic cases where the tangential velocity accelerates up to the edge and then decelerates, commonly leading to reversed flow beyond the edge. Then the flow may separate from the edge, the associated convection of vorticity being known as leading-edge vortex shedding. Cases exist when that vorticity moves far away from the surface, facilitating a complete flow separation with fluid on rounding the leading edge moving far from the surface (as in a stalled aerofoil flow). In other cases it is forced back onto the upper surface, enclosing a 'leading-edge bubble' of separated flow of relatively modest dimensions. This can happen in steady flow, and happens still more commonly in starting flows generated by a wing's movement over only a moderate distance compared with its chord.

A leading-edge bubble is particularly to be expected in the flow of figure 2. At the leading edge of each wing the tangential velocity rises to a maximum and then sharply decelerates, making flow separation practically certain. In the early phase depicted in figure 5, however, the equation of continuity demands that the total fluid motion be sucked into the opening gap, carrying with it the vorticity. This convective effect greatly limits the potentiality for shed vorticity to move far from the boundary. At the same time diffusive effects in that opening region may reduce the strength of the vorticity of opposite sign to the right and left of the line of symmetry by diffusive flux across it.

When the wings break apart, the leading edges have each travelled about one chord length through the fluid, and vorticity shed in the early stages has been sucked into the opening gap, some of this being destroyed by diffusive action. This makes it probable that most of the shed vorticity which remains is confined to a bubble-shaped region attached to the leading edge, enclosing a relatively short region of reversed flow. The low Reynolds number (around 30) of the motions of *Encarsia formosa* would certainly help to promote bubble reattachment by increasing diffusive effects: at high Reynolds number it is known that laminar-separation bubbles from the leading edges of aerofoils are particularly prone to reattachment when transition to turbulence takes place in the separated boundary layer, essentially because that transition enhances diffusion; the analogous enhancement of diffusion at Reynolds numbers around 30 can be expected to have a similar effect.

Such a leading-edge bubble would not impede the operation of the Weis-Fogh mechanism. When the wings move apart as in figures 1(f), (g) and (h) the immediate development of a high lift coefficient would indeed be facilitated by the effective rounding of the leading edge to a 'good aerofoil section' provided by the bubble.

Furthermore, the circulation around the effective aerofoil consisting of wing and bubble would be enhanced, if anything, by the bubble's presence. Figure 6 sketches possible shapes (whose accuracy or otherwise is not important) for the vortex sheet round the bubble in both the  $z$  plane and the  $Z$  plane of figure 3. In the  $Z$  plane the equal and opposite image vorticity in the straight-line boundary is also shown: the complex potential  $w$  in that plane is equal to the irrotational-flow value (8) plus the potentials of the vorticity distribution and its image

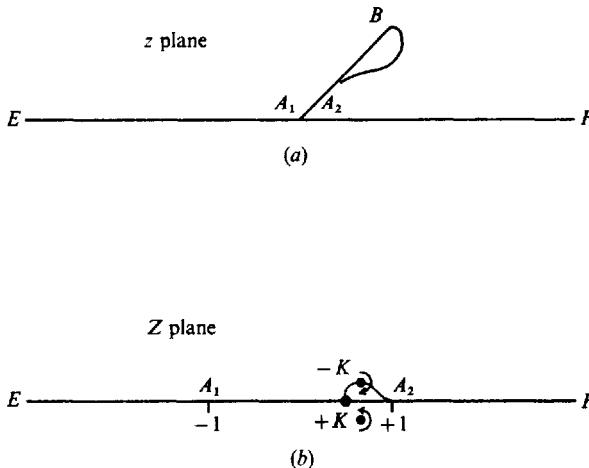


FIGURE 6. Possible shape of a vortex sheet enclosing a leading-edge bubble in (a) the  $z$  plane and (b) the  $Z$  plane (compare figure 3). In the  $Z$  plane the position suitable for a line vortex used to estimate its effect on  $\Gamma$  is also shown, together with the image position.

distribution. The circulation round the wing and the bubble from  $A_1$  to  $A_2$  is the change in velocity potential  $\phi$  from  $Z = -1$  to  $Z = +1$  round a path lying above all those vortices, and is increased above its irrotational-flow value (1) by a slightly larger contribution from the negative vorticity in the upper half-plane than from the positive image vorticity in the lower half-plane.

To see that this effect, if present, would however be small, we may consider the influence of a single vortex of strength  $-K$  at the point  $Z = l + im$  (figure 6). The irrotational-flow velocity in the  $Z$  plane has in  $-1 < Z < 1$  an average value  $\frac{1}{2}\Gamma$  since  $\Gamma$  is its integral from  $-1$  to  $1$ , so that the postulated vortex at  $Z = l + im$  and its image at  $Z = l - im$  are of the right order of magnitude to reverse the flow if  $K/\pi m = \frac{1}{2}\Gamma$ . But the additional circulation round a contour going from  $-1$  to  $+1$  above the vortex is

$$\frac{K}{\pi} \left[ \tan^{-1} \left( \frac{m}{1-l} \right) + \tan^{-1} \left( \frac{m}{1+l} \right) \right], \quad (14)$$

which is of order of magnitude

$$\Gamma m^2 / (1 - l^2) \ll \Gamma. \quad (15)$$

Similar conclusions result from consideration of vorticity distributed in a sheet around a bubble. They are analogous to the classical prediction of a slight enhancement of circulation round a two-dimensional aerofoil in steady flow at given angle of incidence due to increased aerofoil thickness; and they might like that be cancelled out in practice by other effects.

We also consider briefly any vortex shedding after the wings move apart. Note first that such trailing-edge vortex shedding as results in unsteady aerofoil theory from pressure inequalities at the trailing edge is not expected initially in the flow of figure 1(f), where the pressures on the two sides of the opening gap are approximately equal.

Initially, then, the flow over the upper surface of the wing  $AB$  can continue along lines depicted in figure 6(a) as an ordinary aerofoil flow with a leading-edge bubble. Continued reattachment of the upper-surface flow is influenced favourably by translational acceleration of the wing surface. Admittedly, on the lower surface, the cessation of wing rotation and adoption of a translational mode of motion imply a change in sign of the surface's tangential motion relative to the external fluid. A new boundary layer typical of flow over the lower surface of an aerofoil must at once be formed. Its positive vorticity must more than overcome the already existing negative vorticity in the boundary layer associated with the external flow calculated in § 2, to such an extent that the *total* vorticity takes the positive value that permits the required slip between the lower surface and the external fluid. Actually, no special difficulties arise in boundary layers associated with solid surfaces whose tangential motion relative to the fluid changes sign in this way (compare the well-known Stokes layers involving periodic changes in sign): the diffusion helps to produce gradually a cancelling of negative and positive vorticity but at each instant it is in any case only the total vorticity in the layer which significantly influences the external flow.

As the translational motion proceeds, say with velocity  $U$ , the circulation  $\Gamma$  may change owing to gradual vortex shedding at the trailing edge, but it will stay constant if  $\Gamma = \frac{1}{2}UcC_L$ , where  $C_L$  is the value of the lift coefficient for wing motion through the ambient fluid at the angle of incidence in question† with zero net vortex shedding (the viscous-flow generalization of the Kutta-Zhukovski condition). This determines a value of  $U$  such that the full steady-flow lift per unit span

$$\rho U \Gamma = \frac{1}{2} \rho U^2 c C_L \quad (16)$$

can be realized from the outset. This possible implication, a total absence of the Wagner effect, is further explored in § 4.

#### 4. Conclusion

Weis-Fogh (1973) described the zoological implications of his mechanism of lift generation regarding which aerodynamical details have been worked out in the present paper. We may conclude with a broader aerodynamic perspective of its method of working.

In terms of a purely two-dimensional flow (figure 1) the 'fling' allows the wings immediately after they break apart to experience something close to maximum lift, essentially because in the language of the Prandtl-Wagner theory each acts as a 'starting vortex' for the other: one of full strength, indeed, rather than the half strength of Wagner's classical starting vortex (Wagner 1925). The concentration of fluid vorticity into two 'bound vortices' (that is, the vorticity attached to a wing in boundary layers, leading-edge bubbles, etc., all adding up to the circulation round it) implies that the fluid impulse downwards, which is the moment of the distribution of horizontal vorticity, has a constant rate of increase, generating by reaction a constant lift.

† The effective angle of incidence is, however, less than the geometrical angle of each wing to the horizontal because (§ 4) each wing moves in the 'downwash' from the other.

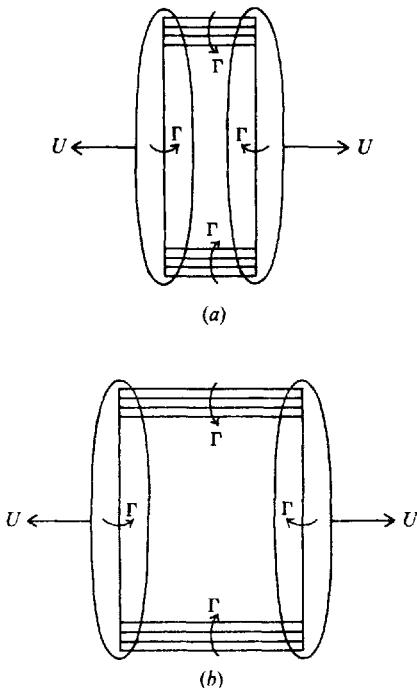


FIGURE 7. Schematic diagram of the Weis-Fogh mechanism for wings of finite span in rectilinear motion. The instantaneous generation of the circulation  $\Gamma$  round each wing by the 'fling' process is necessarily accompanied as in position (a) by the generation of tip vorticity with the same total circulation round it, which then grows by the usual process of tip-vortex shedding as in position (b).

The problem of fitting the local, approximately two-dimensional motions studied in §§ 2 and 3 into a fully three-dimensional model of the flow around real wings of finite span can be tackled by the classical methods of Prandtl (1918). We indicate this first in a case as close as possible to those studied by Prandtl, considering parallel straight wings of finite span in *rectilinear* motion away from one another, which at each cross-section along the span takes the form illustrated in figure 1.

Then the mechanism studied in §§ 2 and 3 for generating circulation around each wing cross-section works to produce bound vorticity of opposite signs on the two wings of figure 1: 'into the paper' on the left-hand wing and out from it on the right. The full three-dimensional pattern of vorticity must, however, be solenoidal: that is, the vortex lines must close up, which as in Prandtl's theory requires the presence of tip vortices (figure 7). These must be generated simultaneously with the circulations and be of sufficient strength to carry all the bound vorticity on the left-hand wing round through  $180^\circ$  at the wing tips and back into the right-hand wing.

The flow near the wing tips as the circulation and this associated tip vorticity are being formed can be inferred from the directions of the tip vorticity vectors in figure 7 (a) as a flow *inboard* from the tips in the region above the wings, and

a corresponding flow outboard to the tips in the region below them. The flow inboard can be regarded in the early stages of the 'fling' motion as a spanwise inflow into the opening gaps. Of course, the opening of mid-span wing sections can suck in air most readily in the plane of those sections as in figure 5; near the tips, however, the sucking can produce significant inboard motion that represents part of the local flow pattern associated with the generation of tip vortices.

Figure 7(b) shows as the wings move apart the equal and opposite bound vorticity on each joined continually by tip vortices so as to keep the whole vorticity field solenoidal. They become longer by the usual mechanism that causes lifting wings to shed tip vorticity: pressure excess on the lower surface pushes fluid outboard near the tips, while pressure defects on the upper surface suck fluid inboard, to generate flow twist near the tips about the wing's direction of motion. In the meantime the fluid impulse downward, which is the moment of the distribution of horizontal vorticity (for example, circulation times area in the use of a single closed line vortex), has a constant rate of increase, generating by reaction a constant lift.

The three-dimensional flow pattern of figure 7 can be expected to modify in some degree the two-dimensional motions around mid-span sections through essentially the same mechanism as in Prandtl's theory: the effective motion of each section is not through 'otherwise undisturbed' fluid, but through fluid subjected to a 'downdraught' or 'downwash' induced by the full three-dimensional pattern of vorticity. As a result, each wing section possesses an effective angle of incidence less than its geometric angle of incidence by an amount equal to the ratio of downwash to wing speed.

Such a downwash correction, indeed, in *unsteady* aerodynamic problems (including the problem of this paper), is present even when they are treated two-dimensionally: for example, the Wagner effect can be thought of as due in part to reduction of the effective angle of incidence through downwash induced by the starting vortex. The Weis-Fogh mechanism involves a similar reduction, as remarked in a footnote at the end of § 3, because each wing moves in the downwash field of the other wing's bound vorticity. There is, however, no resulting loss of lift in this latter case, since the circulation about the wing section has been fixed independently by the 'fling' motion. The reduction means only that relatively high geometric angles of incidence (for example, 30° according to figure 1) are appropriate as the wings first move apart (and are achievable without the wings stalling). Note, furthermore, that the three-dimensional vortex pattern of figure 7, acting together perhaps with some vorticity generated in earlier wing beats, somewhat increases the induced downdraught at each wing section above the value suggested on two-dimensional flow theory.

The wings of *Encarsia formosa*, of course, do not adopt the simple rectilinear motion of figure 7: their tips describe, as explained in § 1, a circular path around the erect body. This must bring into being a tip vortex in the form of a growing circular arc to close the vortex lines. This important consequence of the circulation set up round the wings is shown in figure 8 together with the 'inboard vortex': a much shorter circular-arc vortex close to the body, required to close the vortex lines in that region. Figure 8 shows a linear growth in the area enclosed by vortex

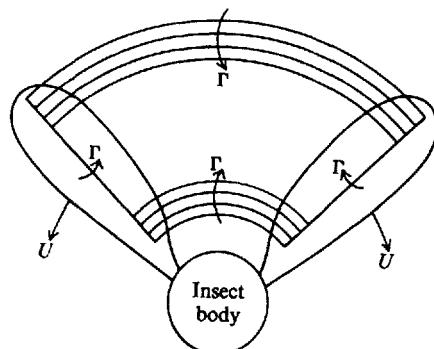


FIGURE 8. Schematic diagram of the outboard and inboard circular-arc tip vortices in the hovering flight of an insect moving its wings in a horizontal circle about its erect body.

lines and hence in the downward impulse of the vorticity distribution, corresponding to a constant lift as before. Its value is independent of induced down-wash effects, which again influence only the local angles of incidence.

It is arguable that the fullest possible utilization of the Weis-Fogh mechanism would be one in which the animal could continue the movement of both wing tips circumferentially as in figure 8, but through the total angle of  $180^\circ$ , making the full 'clap and fling' motion twice per beat instead of once as in *Encarsia formosa*. Then when the circular-arc tip vortex of figure 8 had grown into a complete downward-moving vortex ring the 'clap' would take place and blow the loose ends downwards clear of the body,† to level up the whole vortex ring (whose diametrically opposite side had begun its descent earlier). Then another 'fling' would occur, generating new circulation of opposite sign around the wings, which as they moved back in the opposite direction would create a further complete vortex ring of the same sign as before.

This picture of an 'ideal' utilization of the Weis-Fogh mechanism leading to the support of an animal's weight by momentum generation in the form of a sequence of downward-moving circular vortex rings, two per wing beat, is of some theoretical interest because the circular shape of vortex line carries the greatest possible momentum (proportional to the area enclosed) per unit kinetic energy. We may note incidentally that, far below the animal, viscous dissipation of energy flux without change of momentum flux would gradually convert that motion into the classical 'laminar round jet' solution of the equations of motion (Landau 1944; Squire 1951) corresponding to the action of a point source of momentum in a viscous fluid. Indeed, the flow field far below any small enough hovering animal of mass  $m$  must take the form of this similarity solution, depending on just one parameter: the ratio between the force  $mg$  with which the animal acts on the fluid and the quantity  $\rho\nu^2$  of the same dimensions formed from the fluid's density  $\rho$  and kinematic viscosity  $\nu$ . The ratio (a sort of Reynolds

† Note that the air motion during the 'clap' is not the reverse of that during the 'fling': evidently, the outflow from the closing gap, far from being an irrotational source-type motion (the reverse of the sink-type motion of figure 5) is a separated efflux in the form of a downward-pointing jet.

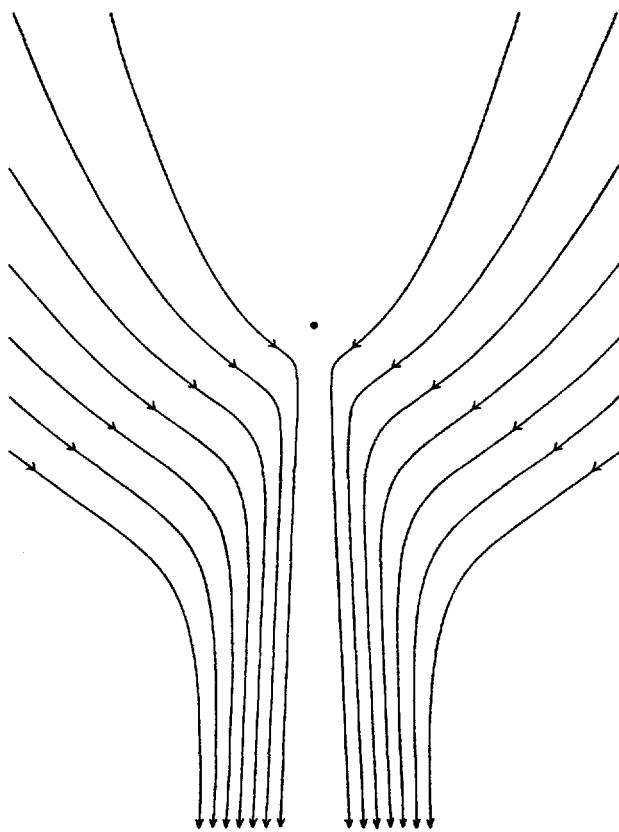


FIGURE 9. Streamlines in the far flow field due to the delivery by *Encarsia formosa* of downward momentum  $mg$  per unit time at the marked position to air of kinematic viscosity  $\nu$  and density  $\rho$ , where  $mg/\rho\nu^2 = 900$ . Streamlines are plotted at equal intervals of Stokes's stream function.

number squared) is around 900 for *Encarsia formosa* (with  $m = 0.025 \text{ mg}$ ): quite small enough for the associated downward jet-type far-field flow (see figure 9) to be extremely stable.

*Encarsia*, admittedly, is not able to move its wings through the full angle of  $180^\circ$  so as to perform a true 'clap and fling' twice (as suggested in the last paragraph but one) per beat: they can move, in fact, through only about  $130^\circ$ . Subject to this limitation, however, it can be argued that its motions approximate as closely as is feasible to those described above. It does a complete 'clap and fling' at one extreme of every wing beat, and makes at the other extreme a broadly similar 'flip' motion (with the same total magnitude of wing angular movement), harder to analyse and without bringing the two wings together. Quite possibly the combination may give advantages not too far from what the 'ideal' motion would achieve!

## REFERENCES

- LANDAU, L. D. 1944 *Dokl. Akad. Nauk. SSSR*, **43**, 286.
- LIGHTHILL, M. J. 1963 In *Laminar Boundary Layers* (ed. L. Rosenhead), chap. 2, §1.7. Oxford University Press.
- PRANDTL, L. 1918 Tragflügeltheorie. *Nachr. Ges. Wiss. Göttingen*, pp. 107, 451.
- SQUIRE, H. B. 1951 *Quart. J. Mech. Appl. Math.* **4**, 321.
- WAGNER, H. 1925 *Z. angew. Math. Mech.* **5**, 17.
- WEIS-FOGH, T. 1973 *J. Exp. Biol.* (to appear).

*This page intentionally left blank*

**PART II**

**INTERNAL BIOFLUIDDYNAMICS**

*This page intentionally left blank*

## **CHAPTER 10**

# **Physiological Fluid Dynamics: A General Survey\***

### **1. Introduction**

Work in physiological fluid dynamics needs very close and intimate collaboration between specialists in physiological science and specialists in the dynamics of fluids. The necessary collaboration has to be preceded by a process of mutual education sufficiently prolonged to bring about on each side an adequate understanding of the other side's language and modes of expression, as well as recognition of which are the main areas where the other discipline has developed a particularly extensive and intricate body of knowledge and skills which can be called upon when required. After this, real communication between the different specialism becomes possible, and can lead to effective research progress.

In this lecture I shall indicate features of this collaborative research that have particularly struck me, as a specialist in fluid dynamics, during the six years in which I have been engaged in it with several colleagues among whom the leader on the physiological side was Dr. Colin Caro. The first feature I want to emphasise is the richness of the field from the fluid dynamicist's point of view. This will not, I believe, come as a surprise to anyone with even an elementary knowledge of the complex organisation of the human body and in particular of its respiratory tract and cardiovascular system, to say nothing of other flow systems like those for transporting lymph or urine. It is easy to believe that wide experience in the very extensive and long established discipline of fluid dynamics can help someone to make a contribution to research in these fields, after he has acquired enough elementary anatomy and physiology to be able to understand the words that his colleagues are using when they put problems to him.

The second feature I want to emphasise is of an exactly opposite character. Many of the most important problems in this field, even if the fluid dynamics in them is considered in isolation, are found to raise questions which, in the whole preceding history of research on fluid dynamics, have totally failed to be answered, or in some cases even to be asked! It is quite humbling to notice how often those questions in fluid dynamics, suggested by study of a physiological problem, turn out to be questions never tackled during all the vast development of knowledge in fluid dynamics aimed mainly at engineering applications. From another point of view, however, this adds richness to the field: it is by no means merely advice that the fluid dynamics specialist needs to give; constantly he is forced to initiate quite new researches on basic problems of fluid flow, and many of these are as interesting

---

\* Lecture (1971) to International Centre for Mechanical Sciences, Udine, Italy.

as the most interesting investigations in the field that have been suggested by engineering needs.

You may ask, 'Why do fluid flows within the human body raise problems so very different from those raised by the huge assemblage of engineering fluid flows?' To this question I can attempt an answer in the form of a list of five main reasons:

(i) *Unusual range of Reynolds number*: both in the larger airways of the lung and in the larger arteries the Reynolds number takes values from a few hundred to several thousand, and in this range the study of internal flows has been largely neglected—particularly at the high-Reynolds-number end of the laminar-flow regime. Engineers have concentrated mainly upon fully developed turbulent flows (except in relation to high viscosity liquids, where they have been interested in quite low Reynolds numbers); in the body, by contrast, only very sporadic or highly localized 'bursts' of turbulence occur (§ 10). On the other hand, laminar internal flows gain greatly in complexity at these high Reynolds numbers because an enormously long so-called 'entry region' is required to establish any simple flow field like Poiseuille flow in a straight tube (§ 3), and also because centrifugal action in curved tubes produces intense secondary-flow effects (see § 2).

(ii) *Unusual multiplicity of tube branchings*: both the lungs and the cardiovascular system are extremely intricate and complex branched networks of tubes, whereby the initial flow, of air entering the trachea (windpipe) or of blood leaving the heart, is subdivided after perhaps 20 to 30 separate branchings into an enormous number (many hundreds of million) of small individual flows, involving inflation of an alveolus to a diameter of a few hundred microns, or passage of blood through a capillary of diameter less than ten microns. In the earlier stages of this branching process the flow pattern suffers a major distortion at each bifurcation, from which, for reasons given under heading (i), it may well not recover before the flow branches yet again (§ 7). In the later stages, where the Reynolds number is too small for this difficulty to arise, the nature of the flow becomes hard to analyse owing to the vast number of tubes in parallel and the insufficiency of information on distribution of numbers by length and by diameter.

(iii) *Unusual distensibility properties of containing vessels*: the networks of vessels containing both the airflows and the blood flows exhibit rather complicated distensibility relations. For example, the way in which a given degree of muscular action to expand the chest cage increases the volume of different parts of the lung is subtly influenced by the ways in which they are separately deformed by the action of gravity on that very flexible structure. Again, the distensibility of the arteries is of great importance in matching the heart pump's reciprocating action to the steady perfusion of the peripheral capillaries, but it is complicated by viscoelastic features, which play a significant role in attenuating the pulse wave as it travels outward from the heart, and by nonlinear features due mainly to fibres of two very different materials in the arterial wall (elastin and collagen) being involved in resisting different degrees of arterial distension, not to speak of the effect of smooth muscle linings possessing various important control functions, particularly (in the case of the arterioles; see Chapter 14) control of the rate of perfusion of different peripheral tissues.

(iv) *Unusual fluid properties*: these are most marked in the case of blood, a suspension of some 40 to 50% by volume of small deformable bodies, mainly the red blood cells which are highly flexible disk-shaped bodies of diameter about 8 microns, in a colourless fluid, the blood plasma. Although the plasma itself has flow properties close to the familiar ones described by Newton's viscosity law, the application of viscometry to whole blood yields so-called values of effective viscosity that show a substantial increase with decreasing rate of strain, due to increased formation of various aggregations of red blood cells at the lower strain rates. Actually, data at the higher strain rates occurring in arteries indicate that errors from treating blood in those as a Newtonian fluid with constant viscosity may not be too serious. In the meantime some excellent research on the flow properties of suspensions is in progress; although this does not particularly encourage us to use different continuum models of blood with unusual rheology, it does emphasise how in small vessels with diameters a few times that of a red blood cell the migration of cells may produce a most unequal distribution of concentration across the tube, with (see Chapter 14) quite a complicated resulting effect on resistance. Incidentally, the air we breathe is also, regrettably enough, a suspension of dust particles and one of the key problems in respiratory fluid mechanics is to study what determines the proportion of particles of different sizes which the airflow may cause to be deposited at different levels in the bronchial tree.

(v) *Unusual pulsatility*: the regular flow reversals in pulmonary inspiration and expiration need careful study, e.g., to determine those departures from exact reversibility at different levels which alone can permit particle deposition. In the circulatory system of normal subjects, valves prevent the total blood flow integrated across a cross-section becoming negative, although some reversed flow localised near an arterial wall may occur. Furthermore, the variation of integrated flow against time in the ascending aorta, where the systemic blood flow leaves the heart, takes the form of a strong surge lasting for less than half of the cycle followed by a far weaker flow during the remainder. If this pulse form is Fourier-analysed the first few harmonics are comparable in amplitude to the steady state component. It is tempting to study the propagation of different Fourier components of the pulse separately and assume that they can be linearly superposed, but evidence for significant nonlinear interaction between such components is accumulating: this can arise from the 'entry region' phenomena and from the 'nonlinear elasticity' phenomena noted under headings (i) and (iii) respectively. On the other hand, attenuation of the pulse wave makes such pulsatility far less important in the microcirculation and in the veins.

In the remainder of Chapter 10, I indicate several lines of research in progress that arise from these various difficulties and that happen to interest me personally, as follows.

## 2. Steady secondary flows

Centrifugal forces acting as a result of curvature of the streamlines of a 'primary flow' produce 'secondary flow' velocities at right angles to those streamlines, with

motion of central fluid towards the outside bend and a return flow towards the inside bend near a wall. The effect, which occurs both in curved tubes and at bifurcations, is strongest in the high-Reynolds-number laminar-flow regime in which it has been least studied. We know now (Barua 1963, McConalogue & Srivastava 1968) that flow in a curved pipe in this regime under the action of a steady pressure gradient involves practically uniform motion towards the outside bend of the central fluid, which is being accelerated by the pressure gradient, whereas the return flow, retarded by viscous stress, is confined to a relatively thin boundary layer. This seems to explain (Lighthill 1970, p. 114) why the critical Reynolds number for transition to turbulence increases from a typical 2000 for straight pipes to a typical 6000 for curved pipes; these values are based on pipe diameter, whereas the Reynolds number based on boundary-layer thickness changes far less. In the meantime it was shown experimentally by Caro (1966) that these secondary flows do act to promote lateral mixing of injected substances, mixing which for substances of low diffusivity in the absence of turbulence would normally be very slow.

### 3. Entry regions

For high-Reynolds-number steady flow in a straight pipe, it is well known (for an excellent recent review see Fargie & Martin 1971) that the length of the 'entry region' required for an approximately uniform initial velocity distribution to come within 5% of a Poiseuille distribution is about  $(0.03R)d$  where  $R$  is Reynolds number based on the diameter  $d$  and initial velocity  $U_0$ . This expression suggests that any one of the larger arteries is almost all entry region, where it is certainly not correct to apply the law of Poiseuille (although he was a physiologist!). Note that the steady pressure drop in an entry region exceeds the Poiseuille value by an amount often referred to as 'the kinetic energy correction'; it is required mainly to accelerate the central flow to the more peaked Poiseuille distribution with its mean kinetic energy higher by  $\frac{1}{2}\rho U_0^2$  but the total correction, allowing also for extra resistance in the extra thin boundary layer near entry, is about  $0.67\rho U_0^2$ . When we consider the total pressure drop across a system with very numerous bifurcations like the arterial system, we can find that even though a single kinetic-energy correction is small in relation to that total, the summation of a large number of them in series may not be negligible.

Work to combine these last two topics is in progress, mainly because the very first entry region in the systemic circulation is itself a highly curved tube, the great arch of the aorta. We cannot expect a fully developed secondary flow in this entry region, into which the blood probably enters with small initial vorticity if the aortic valve opens properly, so that the vorticity of both the primary and secondary flows may be confined to an entry-region boundary layer. The central flow would then have the characteristic irrotational property of velocities greatest on the *inside* bend (the opposite of that for a fully developed secondary flow), although secondary motions in the boundary layer would scour the *outside* bend and cause the rates of shear and of mass of transport to be greatest there. Preliminary work by associates of myself and Dr. Caro (especially Dr. Scarton on the theoretical

side and Dr. Seed on the experimental side) is tending to support this rather simple description.

#### 4. Incipient atheroma

You may wonder why so much emphasis is being placed upon precise distributions of flow velocity or of rates of shear and mass transport. There are several reasons, but one of the most important is in relation to the onset of atheroma, a degeneration of the arterial wall which leads to so-called hardening of the arteries. Actually the rates of shear and of mass transport at a wall have generally similar distributions in flow systems, and there are two exceedingly interesting lines of research which link these distributions with the distribution within the circulation of sites of incipient atheroma. Unfortunately the two lines of research, described for example by Fry (1968) and by Caro, Fitz-Gerald & Schroter (1971), are emphasising opposing trends. The Ciba Foundation has arranged a symposium in 1972 where I have to act as an impartial chairman and try to help reach a consensus regarding the relevance of the two types of work, so I must be careful not to prejudge the issue. I will say on the one hand that Fry (1968) is able to show that extremely high rates of shear are able to produce actual mechanical damage to the endothelial lining of the arterial wall, and that this can foster the onset of incipient atheroma. On the other hand the claim of Caro, Fitz-Gerald & Schroter (1971) is that these very high rates of shear, required for mechanical damage, are not found in the actual circulation. They study the distribution of sites of incipient atheroma in large numbers of human and canine subjects and appear to exhibit a close correspondence between these sites and sites where there are low rates both of shear and of mass transport; that is, of transport of some substance away from the arterial wall due to scouring by the flow. They have a complicated explanation of atheroma onset in such 'dead-water' and other low-shear regions in terms of a failure of the blood flow to transport away from the arterial wall at a sufficiently high rate the cholesterol synthesised inside it. Without expressing any personal view on these two extremely interesting and extensive research programmes, I will remark that since atheroma is one of the big killing diseases of modern society they both in their different ways suggest how important it will be to obtain a good knowledge of the distribution of rate of shear in the circulatory system.

#### 5. Distribution of shear in branched systems

We may continue for just a little longer with comments based on steady-flow fluid dynamics; in fact, by studying steady flow through a branched system. The relation of such studies to real flow in the cardiovascular system or in the lung airways may be that they can indicate typical distributions at instants of *local maximum* flow, where effects of rate of change of flow are smallest. Alternatively, for the blood flow, they may indicate something of the distribution of the *steady-state component* of velocity (except in so far as this may nonlinearly interact with higher Fourier components).

The rate of shear at a particular point in a particular tube of a branching system in steady flow (Caro, Fitz-Gerald & Schroter 1971) depends on two factors: (i) a

flow factor depending on the total flow  $Q$  in that particular tube; this factor is  $32Q/\pi d^3$ , the wall shear for simple Poiseuille flow at volume flow rate  $Q$  in a tube of diameter  $d$ : (ii) a geometrical factor depending on position in the tube, and representing any local departure of the rate of shear from its Poiseuille value. These departures take forms already discussed: thus in any entry length (§ 3) the geometrical factor is considerably greater than 1 upstream, but it decreases downstream; in a curved tube (§ 2) it is greater on the outside bend than on the inside; at a bifurcation, the geometric factor is large on the central 'flow divider' where a new boundary layer tends to form, but is small on opposite parts of the tube walls, which are effectively 'inside bends' for the flow curving round into the daughter tubes.

To these remarks I will add brief comments on how the flow factor varies in a branching system, using the concept of the area ratio  $\beta$  at a bifurcation; here,  $\beta$  is the ratio of the combined area of the daughter tubes to the area of the mother tube. This area ratio is important both in the cardiovascular system and in the lungs, because it is necessary to achieve a large *total* increase in area through both these networks: two orders of magnitude from the aorta to the whole assemblage of systemic capillaries, and still more from the trachea to the alveoli.

If a fluid dynamicist were asked to design such a branching system, he would accept the need to make  $\beta$  consistently greater than 1 at each bifurcation, but he would be very cautious about making  $\beta$  considerably greater than 1 in at any rate the earlier bifurcations. This is because the mean flow velocity will be divided by  $\beta$ , and this retardation may cause flow separation if  $\beta$  is greater than about 1.2. Such flow separation may be undesirable either (i) because it increases resistance, or (ii) for some other reason such as a postulated harmful effect of a 'dead-water' region upon the arterial wall.

However this may be, the measured values of  $\beta$  at early stages of both networks are small: less than 1 for the main aortic bifurcation (Caro, Fitz-Gerald & Schroter 1971) and not more than 1.2 for other bifurcations. We note the consequences of this here in only the simple special case of a symmetrical bifurcation, with half of the mother-tube flow going into each daughter tube.

Then the main velocity is multiplied by  $\beta^{-1}$ , the tube diameter by  $(\frac{1}{2}\beta)^{1/2}$ , and the Reynolds number by  $(2\beta)^{-1/2}$ . This reduction in Reynolds number is important because after enough bifurcations the Reynolds number will be reduced to values of 10 or less for which flow separation is not expected even for quite large values of  $\beta$ , which there would become acceptable. Finally, the 'flow factor' in the rate of shear, namely  $32Q/\pi d^3$ , is multiplied by  $(2/\beta^3)^{1/2}$ . We now see that values of  $\beta$  in the early bifurcations are small enough so that this quantity is greater than 1, and conclude therefore that this 'flow factor' increases downstream. In general, then, rates of shear are least in 'proximal' arteries (those nearest the heart) and, within them, in those sites where the 'geometrical factor' is smallest.

## 6. Distribution of resistance in branched systems

We note also how the different parts of a branched system may contribute to the total pressure drop in a steady flow through it. To do this we make the crude

assumption that as tube diameters are reduced the associated lengths of tube are reduced approximately in proportion, which appears to be broadly speaking true.

The pressure drop predicted by Poiseuille's law exhibits a simple behaviour: to obtain the value in the daughter tubes, we multiply the value in the mother tube by the same factor  $(2/\beta^3)^{1/2}$  as for the shear. We need to remember that the true pressure drop will exceed that predicted by Poiseuille's law in entry-length regions, and for other reasons at high Reynolds number (see § 7 below, and also heading (i) above), but can immediately observe that while the shear is increasing distally (that is, towards the periphery, and for the reason that  $\beta$  does not much exceed 1) the Poiseuille contribution to pressure drop will also increase.

The cardiovascular system and lung airways differ in the details of this. It appears that in the former the values of  $\beta$  are thus kept down for very many generations; certainly to tube diameters less than 0.1 mm. This is consistent with the 'fluid dynamicist's plan' to wait until Reynolds number is small before allowing higher values of  $\beta$  that might otherwise permit flow separation. By contrast, in the lung the ratio  $(2/\beta^3)^{1/2}$  falls below 1 soon after the fifth generation, where Reynolds number is still several hundreds. Substantial flow separations in the immediately subsequent generations can be expected but for various reasons may be less harmful in the airways than in the blood-stream.

Associated with this is the fact that in the blood-stream the greatest pressure drop occurs in small vessels of diameter less than 0.1 mm (some features of this are discussed in Chapter 14); whereas all the significant pressure drop in the airways network occurs in the first ten generations, with diameter exceeding 1 mm (Macklem & Mead 1966). Put in a different way, the arterial resistance is a low-Reynolds-number problem but the bronchial resistance is a high-Reynolds-number problem.

## 7. Bronchial resistance

Accordingly, the relatively high-Reynolds-number considerations of Chapter 10 are relevant to pressure drop, and its enhancement beyond the Poiseuille prediction, mainly in the context of bronchial resistance. With this in mind, Pedley, Schroter & Sudlow (1970) studied that enhancement in detail for branching geometries typical of the first ten generations of lung airways, and came to some interesting conclusions.

They showed that the secondary flows present in a bifurcation are so powerful that the initial velocity distribution in a daughter tube is distorted in shape to a quite extraordinary degree, with a very high peak near the inside bend; and that very little of this distortion is lost through the whole length of the daughter tube. In consequence, the ratio of viscous dissipation to that predicted by Poiseuille's law is *much* more than in an ordinary entry length with uniform entrance conditions; in fact, it is  $0.33(R d/l)^{1/2}$ , where Reynolds number  $R$  is based on diameter  $d$  and mean velocity, while  $l$  is tube length.

This means that the contribution to pressure drop of the first few generations where  $R$  is greatest is still further enhanced. Furthermore, it implies a total pressure

drop in this high-Reynolds-number laminar-flow region proportional to flow to the  $(\frac{3}{2})$ th power. Actually the flow in the trachea itself is normally turbulent, but this is in a Reynolds number range around 10,000 where the turbulent resistance law happens to be also close to a  $(\frac{3}{2})$ th power law! Pedley, Schroter & Sudlow (1970) conclude, therefore, that pressure drop through the *whole* bronchial tree follows such a law, and describe some confirmatory experimental data.

### 8. Velocity distributions in pulsatile flow

Some consequences of the pulsatile character of arterial blood flow, to which I must now urgently turn, will at first be considered without taking arterial elasticity into account. Probably this is not too bad an approximation if we are interested in how the pulsatile local distribution of velocity across a systemic artery responds to the pulsatile variations of local pressure gradient. (Note that typical variations of arterial diameter as a pulse passes are only about  $\pm 2\%$ ; note also that the pulse wave velocity is at least five times the maximum blood velocity so that if we neglect elasticity and thus make the effect of pressure changes propagate infinitely fast there is not such an enormous exaggeration involved!)

The velocity distribution associated with a single Fourier component of the blood pulse, with radian frequency  $\omega$ , depends critically on the quantity called  $\alpha$  by Womersley (1955), which is a ratio of the tube radius  $a = \frac{1}{2}d$  to the oscillating-boundary-layer thickness  $(v/\omega)^{1/2}$ , where  $v$  is kinematic viscosity (say,  $4 \text{ mm}^2 \text{s}^{-1}$  for blood flow in large arteries). With  $\omega = 8 \text{ s}^{-1}$  (typical value for the first Fourier component in normal humans) this thickness is about 0.7 mm, while for the  $n$ th Fourier component this figure should be multiplied by  $n^{-1/2}$ .

In the large arteries, then,  $\alpha$  is of order 10 and the response to sinusoidal pressure gradient is predicted to be a velocity distribution uniform across the tube except for a boundary layer occupying about 10% of the tube radius. Almost all the pressure gradient goes into acceleration, that is, into combating the inertia of the fluid, rather than into combating viscous resistance, so that the flow lags by almost  $90^\circ$  behind the pressure gradient.<sup>†</sup> Within the boundary layer, however, the lag is less; only  $45^\circ$  for the rate of shear at the wall itself.

### 9. Pulse propagation

The simple classical theory of pulse propagation in an elastic thin-walled tube with Young's modulus  $E$  and wall thickness  $h$  gives the basic value  $(Eh/2\rho a)^{1/2}$  for the wave velocity  $c$ . Much study has been devoted to estimation of departures from this value due to fluid viscosity or due to complicated wall properties (thickness; compressibility; 'tethering'; viscoelasticity; nonlinearity) but surprisingly little change in the result is predicted for the real arteries. Attenuation of the wave seems to be underestimated by effects of fluid viscosity alone, but significantly augmented by arterial viscoelasticity. Wave velocity may perhaps be very slightly

---

<sup>†</sup> This is true even though viscous resistance is as much as twice the Poiseuille prediction (owing to the thin boundary layer) for  $\alpha = 10$ . It is only for  $\alpha < 3$  that inertial reactance becomes less important than viscous resistance, which in *this* region (arteries of diameter  $< 4 \text{ mm}$ ) is predicted to follow Poiseuille's law rather accurately.

smaller for the first Fourier component (with greater inertia coefficient owing to the thicker boundary layer) than for higher ones.

Measurements show that typical wave velocities increase peripherally to values around  $10 \text{ ms}^{-1}$  from values for the aorta around  $5 \text{ ms}^{-1}$ , consistently with a much lower modulus of elasticity observed for the material of the aorta. These measurements are made by timing the passage of an artificially induced sharp pressure disturbance.

By contrast, the natural human pulse is *not* purely a wave travelling outwards from the heart. It includes a number of reflexions from bifurcations. The theory of wave reflexion at a bifurcation makes it depend on the 'admittance'  $Y$ : this is the reciprocal of 'impedance', and is the ratio of the amplitudes of the volume flow fluctuations and the pressure fluctuations, being given for a travelling wave in a tube of cross-sectional area  $A$  as  $Y = A/\rho c$ . A positive pressure pulse produces a *positive* reflected pulse (of reduced amplitude) at any bifurcation where the sum of the admittances of the daughter tubes is less than that of the mother tube. This is the case at the main aortic bifurcation (where we have seen that the total area actually decreases, while  $c$  increases) and at some other early bifurcations in the arterial tree.

This is why the waveform in the aorta and a few other large arteries includes a *positive* reflected-wave component, and is thus intermediate in character between a travelling wave and a standing wave. These considerations make the largest arteries, and aorta in particular, able to act effectively as a so called *Windkessel*, storing by its capacitance the stroke volume from the cardiac output for delivery at a very steady rate through the peripheral circulation (see, for example, Taylor 1966).

Study of the transmitted wave at such a bifurcation explains also why pressure amplitude in one of the arteries in the leg may exceed, owing to reduced admittance, the pressure amplitude in the aorta. The smoothed form of the pressure-time curve at such a site may be attributed to greater attenuation for higher harmonics. The cause of another feature of it (phase advance of the second harmonic, leading to a more marked 'diastolic peak') is more in dispute; we note this as a point where either the influence of nonlinear effects or the smaller wave velocity of the first Fourier component could produce an effect in the direction observed.

## 10. Turbulence in the blood stream

Uncertainty that has long existed about whether turbulence appears in the human aorta has now been resolved by work using rapid-response velocity-measuring probes in at least three centres (work of Dr. Ling in Washington, Dr. Schultz in Oxford, Dr. Seed in London, together with several collaborators). The peak Reynolds numbers are such that the secondary flow in the arch of the aorta might under *steady* conditions just avoid the onset of turbulence (see § 2). Actually turbulence appears for quite a small fraction of each cycle, while the velocity is first decelerating (see, for example, Seed & Wood 1971).

I am personally inclined to attribute this turbulence to change in the boundary-layer velocity profile, whose response very near the wall to the retarding pressure

gradient has a phase advance of  $45^\circ$  (see § 8) over that of the central core fluid. This produces the type of velocity profile with a point of inflexion, which is known to be enormously more unstable than any other type. 'Bursts' of turbulence are commonly generated soon after such a point of inflexion appears locally near a wall. They may die away quite quickly, however, when the conditions change (see, for example, Lighthill 1970).

Turbulence is not common elsewhere in the circulation, except where jets form; for example, owing to valvular stenosis (failure to open completely); a circumstance that of course is of great clinical value. In the usual technique for blood pressure measurement using an arm cuff inflated to a known pressure, that is allowed to fall while the physician listens to the flow immediately downstream, the well-known Korotkov sound is heard once the cuff pressure is intermediate between systolic and diastolic. This, however, is a more musical sound than that produced by turbulence.

Recently the Korotkov sound was rather clearly explained in excellent work of a combined theoretical and experimental nature as due to a rather regular nonlinear limit-cycle oscillation which occurs while a tube is in the process of collapsing under excess external pressure. Very roughly speaking, the mechanism for this is that the rate of flow constriction increases to a point where the fluid dynamics demands unsteady pressures which temporarily halt or even reverse that rate, which then begins to increase again because the excess external pressure once more dominates (Conrad 1969, Katz et al. 1969).

## 11. Urinary tract

Flows of other fluids besides air and blood pose very interesting problems. In the kidney, remarkable mechanisms, including some similar to the chemical engineer's 'counter-current exchangers', form the urine out of a fraction of the blood plasma. This passes to the bladder through a ureter (one for each kidney) by means of a peristaltic pumping mechanism, whose mode of operation repays study for many reasons. An excellent review of peristaltic pumps has recently been given by Jaffrin & Shapiro (1971).

## REFERENCES

- BARUA, S. N. 1963 Quart. J. Mech. Appl. Math. **16**, 61.
- CARO, C. G., 1966 J. Physiol. **185**, 501.
- CARO, C. G., FITZ-GERALD, J. M. & SCHROTER, R. C. 1971 Proc. Roy. Soc. B, **177**, 109.
- CONRAD, W. A. 1969 IEEE Trans. Bio-Med. Eng. **BME-16**, 284.
- FARGIE, D. & MARTIN, B. W. 1971 Proc. Roy. Soc. A, **321**, 461.
- FRY, D. L. 1968 Circulation Res. **22**, 165.
- JAFFRIN, M. Y. & SHAPIRO, A. H. 1971 Ann. Rev. Fluid Mech. **3**, 13.
- KATZ, A. I., CHEN, Y. & MORENO, A. H. 1969 Biophys. J. **9**, 1261.
- LIGHTHILL, M. J. 1970 *Turbulence*. In *Osborne Reynolds and Engineering Science Today* (ed. McDowell, D. M. & Jackson, J. D.), pp. 83–146, Manchester University Press.
- MACKLEM, P. T. & MEAD, J. 1966 J. Appl. Physiol. **22**, 395.
- MC CONALOGUE, D. J. & SRIVASTAVA, R. S. 1968 Proc. Roy. Soc. A, **307**, 37.

- PEDLEY, T. J., SCHROTER, R. C. & SUDLOW, M. F. 1970 *Respiration Physiol.* **9**, 371.  
SEED, W. A. & WOOD, N. B. 1971 *Cardiovascular Res.* **5**, 319.  
TAYLOR, M. G. 1966 *Biophys. J.* **6**, 697.  
WOMERSLEY, J. R. 1955 *J. Physiol.* **127**, 553.

*This page intentionally left blank*

# CHAPTER 11

## Respiratory Flow Patterns

### 1. Introduction

In this chapter we pursue further (see § 2) the description of flow patterns in the larger airways of human lungs given in Chapter 10, § 7. These are high-Reynolds-number flow patterns in which the transport of oxygen and of carbon dioxide is by convection. Discussion of the low-Reynolds-number motions in alveoli and in pulmonary capillaries, and of the primarily *diffusive* transport of gas between them, is postponed to Chapter 14.

In the meantime it is of interest to contrast the respiratory flow patterns in mammals generally with those in other classes of animals. For this purpose we here select two classes that also figure prominently in Part I of this book, the fishes and the birds, while omitting discussion of a third such class, the insects, in which respiratory gas transport at least under resting conditions is essentially diffusive.

The Göttingen symposium on comparative physiology of respiration in vertebrates, reported at length in Volume 14 of *Respiration Physiology*, includes extensive detail on the different types of flow involved. It is the general architecture of the biofluiddynamic processes that varies greatly from class to class (and, to a lesser extent, within classes). Their end result, the bringing into close juxtaposition of red blood cells with oxygenated air or water, is achieved by generally similar geometrical arrangements in most cases: for example (Weibel 1972, Hughes 1972), Figure 1 shows a red blood cell in the alveolar septum of a human lung, while a rather similar configuration within a fish gill is illustrated in Figure 2 (electron-micrograph of a secondary lamella in a tench *Tinca tinca*); however, the thickness of the blood-air barrier in mammals takes somewhat lower values (0.3 to 0.5 µm) than the blood-water barrier's thickness in fishes (2 to 3 µm for tench, as much as 10 µm for dogfish and down to 0.6 µm only for extremely active fishes such as the scombrids).

Fishes, for the most part, use a double pumping action to ventilate their gills: they take in water in successive gulps that are forced through the gills by contractile action after the mouth closes, but they partially smooth out this intermittent 'buccal-pump' forcing by out-of-phase posterior suction from expansions of the opercular cavities. There are many exceptions (Marshall 1965, Chapter 6, gives a good general survey): scombrid fishes keep their mouths open, relying on the Bernoulli pressure due to their forward motion to force water through the gills; the eels lack effective opercular-pump action; and so on. Fish gills in general, however, are *continuously* ventilated, in sharp contrast to mammalian lungs.

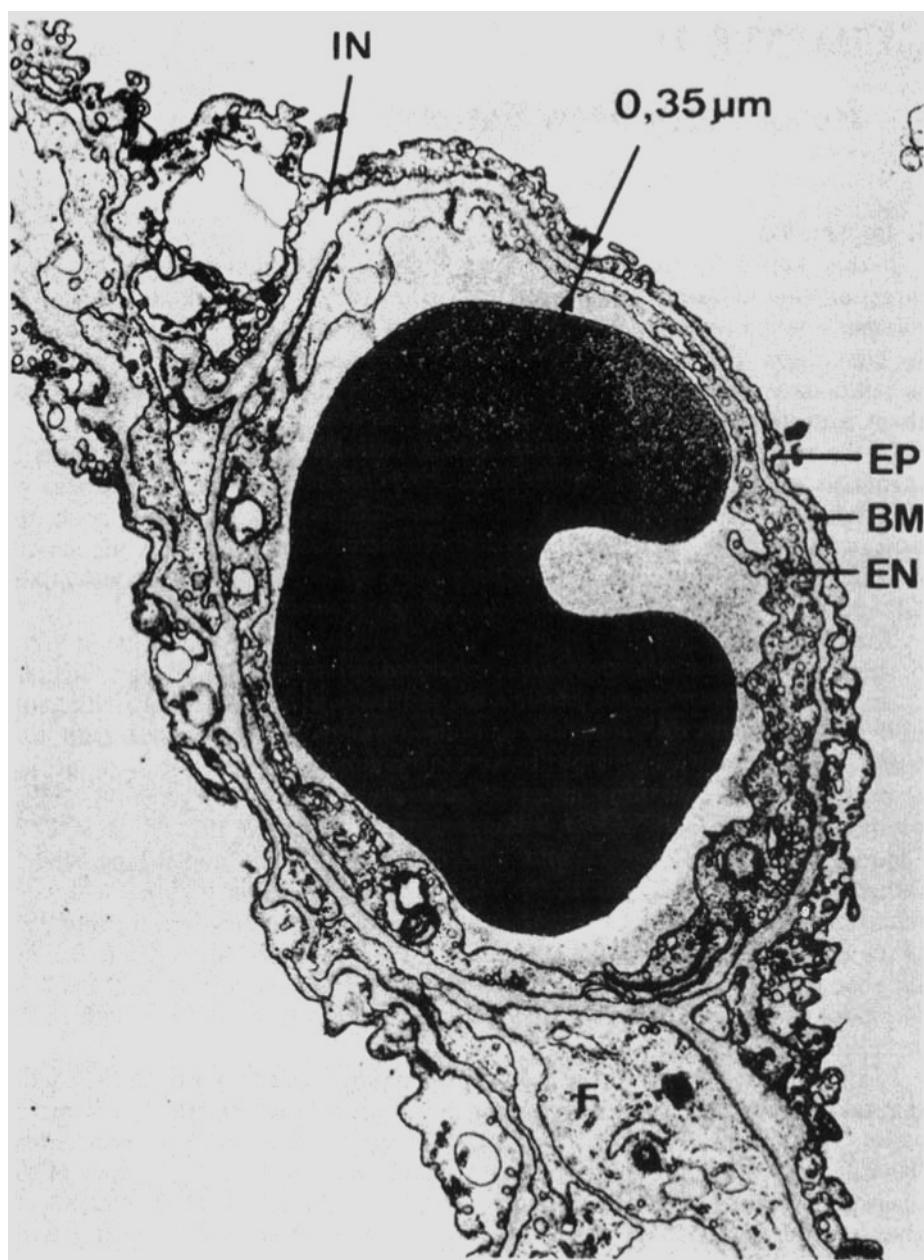


FIG. 1. This Figure 3 from Weibel (1972) shows a red blood cell within the alveolar septum of a human lung.

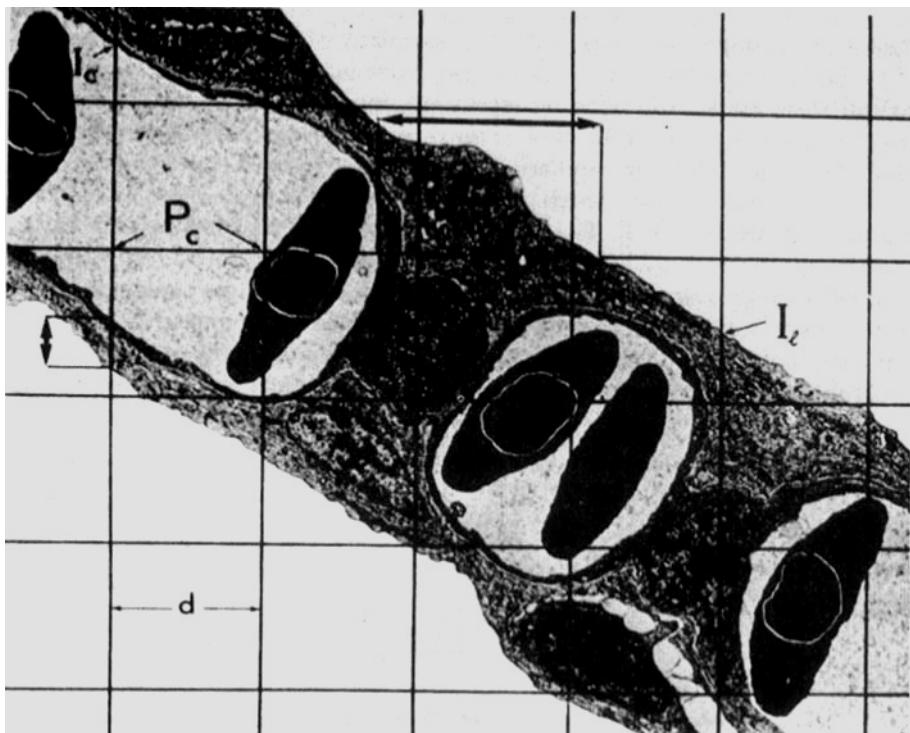


FIG. 2. This Figure 12 from Hughes (1972) shows red blood cells within a secondary lamella from a fish gill.

Gill flow patterns possess another special feature advantageous for gas exchanges: blood in the secondary lamellae flows in the opposite direction to the water, so that gills have the good efficiency of the 'counter-current' exchangers used by chemical engineers. Essentially, the whole exchange surface is effective because anterior water with high oxygen content exchanges with blood that has already been partly oxygenated whereas posterior water that has lost some of its oxygen exchanges with the least oxygenated blood which has just entered the capillaries.

A wide variety of fishes from many different orders have acquired, in addition, the ability to breathe air. Characteristically these are fishes living in stagnant tropical waters that may periodically dry up or become de-oxygenated. Many different adaptations, either of the gills or of the swimbladder (or even of the intestines), have facilitated these developments.

The numerous fishes in which the swimbladder has developed an alveolar or lung-like conformation suggest the type of respiratory system from which the lungs of amphibians evolved (from a group of fishes now extinct) in the Carboniferous period. In subsequent development of amphibian, reptilian and mammalian lungs, the gas exchange remains located at the terminal surface of blind air ducts, of which the smallest units are the alveoli. In all these tetrapod vertebrates *lung*

gas exchange (as opposed to that very variable proportion of respiration which takes place through the animal's skin) is essentially of an intermittent character.

The birds, by contrast, evolved auxiliary pumping units called air sacs which enabled their lungs to develop an essentially constant-volume form. While the air sacs expand and contract a flow of air through the gas-exchange region takes place along thin rigid 'air capillaries' of diameters 3 to 10  $\mu\text{m}$  (Duncker 1972) in close proximity to the blood capillaries: seemingly in a cross-current flow arrangement (see Figure 3). The constant-volume lung has some notable advantages: for example, an exchange surface area per unit volume of about  $250 \text{ mm}^{-1}$ , an order of magnitude greater than in mammalian lungs (whose expanded alveoli have diameters an order of magnitude greater than those quoted above). The distribution of air sacs in birds assisted them in developing flight capabilities both by permitting this respiratory improvement and by reducing their mean density.

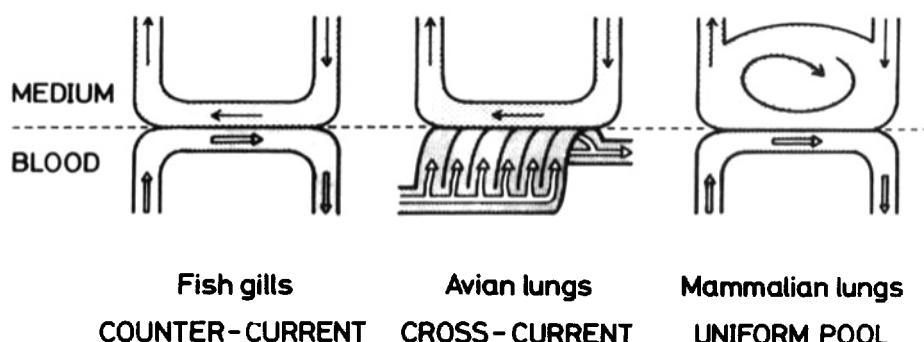


FIG. 3. This Figure 1 from Piiper & Scheid (1972) contrasts the types of respiratory gas transfer found in fishes, birds and mammals.

A more detailed account of respiration in birds is given in § 3, as a preliminary to the definition of an interesting field of research: the problem of how the pumping action of posterior and anterior air sacs can produce a rectified flow through the gas-exchange area in the absence of any valves. The opercular pumps of fish gills have valves, and so essentially does the fish's buccal pump for which the mouth acts as a valve. A continuous one-way flow, somewhat similar to that in fish gills, is produced in bird lungs, however, by expansions and contractions of upstream and downstream cavities which all lack means of shutting off either the airflow through some 'exit port' on expansion or that through an 'entry port' on contraction. This chapter is concluded by a tentative discussion of features in the geometrical conformation of lungs and air sacs which may for fluiddynamic reasons allow such rectification.

## 2. Flow patterns in human bronchi

Schroter & Sudlow (1969), regarding as somewhat too difficult the measurement of detailed velocity distributions in lung airways during inspiration, constructed perspex models of tube branchings with geometrical features typical of the first

ten generations of airways as measured in casts of human lungs. Figure 4 shows the types of bifurcation involved: symmetrical, with a total included angle of  $70^\circ$  and a diameter ratio of 0.78 (corresponding to a value 1.22 of the area ratio  $\beta$  defined in Chapter 10). Most important, the ratio of tube length to tube diameter, as in typical bronchi, is only 3.5, so that only a short distance is available for a velocity distribution as disturbed by flow through a bifurcation to settle down again before being similarly disturbed yet once more.

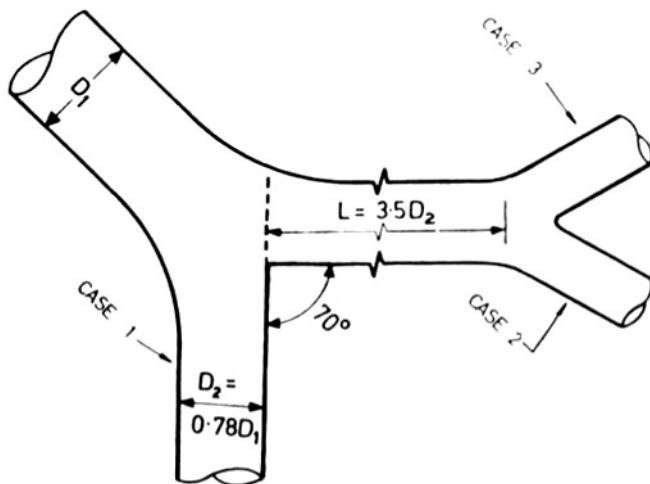


FIG. 4. This Figure 2 from Pedley, Schroter & Sudlow (1970) illustrates the geometry of the perspex models used in their flow experiments.

In fact, at the Reynolds numbers appropriate to these airways, an 'entry length' such as is required (Chapter 10, § 3) to reduce the distribution of velocity at entry to a Poiseuille distribution with the same mass flow is greater by an order of magnitude than such tube lengths. Actually, the departure from a Poiseuille distribution at entry is even greater than for a uniform distribution as used in the theory of the entry length. This comes about partly because on the flow divider a thin new boundary layer forms, where the velocity just outside the boundary layer is near the maximum value in the mother tube, and partly because secondary flows generated by curved motion of the primary flow into its new branch intensify such a peaking of the velocity on the 'outside bend' of that curved motion; that is, around the flow divider.

Figure 5 gives a typical set of velocity profiles in the daughter tube: these are velocity profiles at distances of 1, 2 and 3 diameters downstream of the bifurcation for a Reynolds number of 700 and a Poiseuille distribution in the mother tube. Note the very large departures from Poiseuille flow, with a strong peaking near the flow divider, that appear at once and persist for the whole length of the tube.

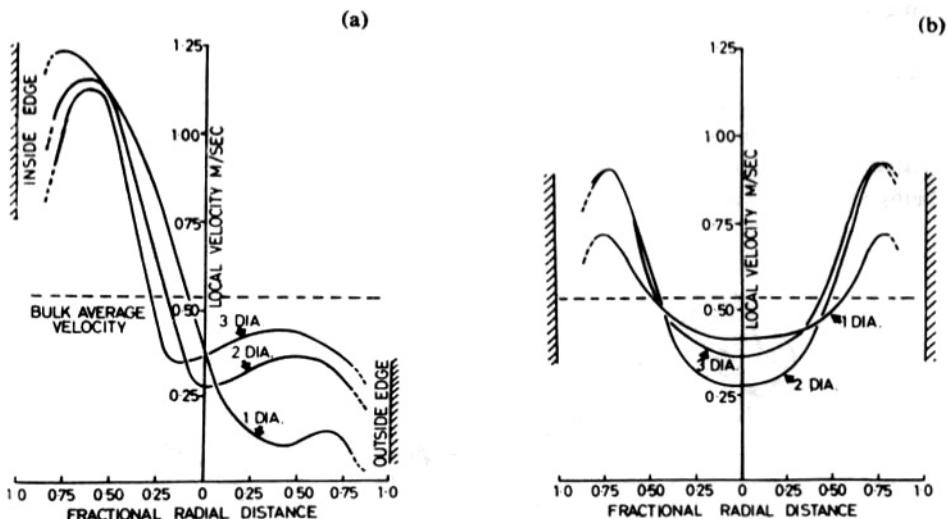


FIG. 5. This Figure 1 from Pedley, Schroter & Sudlow (1970) shows one particular set of velocity distributions which they obtained.

Pedley, Schroter & Sudlow (1970) show how to calculate the rate of viscous energy dissipation in a tube from the velocity profiles measured in this way. They carry out the calculation for a wide range of Reynolds number and for flows disturbed either by a single bifurcation or by two in series. The ratio between the dissipation rate calculated thus and on the assumption of Poiseuille flow is found to be sensitive mainly to the value of the Reynolds number  $R$  based on diameter  $d$  and on mean velocity  $U$ .

This dependence on Reynolds number can be understood in terms of the idea that energy dissipation in a flow with velocity peak very close to the wall is concentrated primarily in a boundary layer with small thickness  $\delta$  lying between the region of peak velocity and the tube wall. Velocity gradients in this region are of order  $(U/\delta)$  leading to an energy dissipation rate of order  $\mu(U/\delta)^2$  per unit volume in a boundary-layer cross-sectional area of order  $d\delta$ . The energy dissipation rate per unit length of tube is thus of order  $\mu U^2(d/\delta)$ , and exceeds that for Poiseuille flow by a ratio of order  $(d/\delta)$  which itself varies as  $R^{1/2}$ . If the length  $l$  of the tube may be a variable multiple of  $d$ , then  $d/\delta$  varies as  $(Rd/l)^{1/2}$ .

The calculations of energy dissipation rate carried out in some sixty cases by Pedley, Schroter & Sudlow (1970) all make its ratio to the dissipation rate for Poiseuille flow fairly close to

$$0.33(Rd/l)^{1/2}, \quad (1)$$

in agreement with these arguments. The significance of such a variation with Reynolds number can be appreciated from the fact that the Reynolds number in light breathing (10 L/min) varies from about 800 in the trachea through about

100 in the fifth-generation airways to about 10 in the tenth generation; with values ten times as great as these in heavy breathing (100 L/min).

The effect of these conclusions is to enhance the importance of the early generations of lung airways as the main contributors to loss of total pressure in inspiratory flows. Essentially, they make resistance proportional to the  $(\frac{1}{2})$ th power of velocity, and *loss of total pressure* (which equals resistance times velocity) proportional to the  $(\frac{3}{2})$ th power. (Here, total pressure can be defined in general as the sum of the kinetic and potential energies of the fluid per unit volume, plus the ordinary or 'static' pressure.)

In heavy breathing, Reynolds numbers exceed the 'critical' value (2,000) in the first 3 or 4 generations of lung airways, which raises the question of how far the above conclusions are modified at the low end of that range of Reynolds numbers where turbulent flow is expected. Pedley, Schroter & Sudlow (1970) argue that the formation of a new, predominantly viscous, thin boundary layer on the flow divider will not be significantly affected by the relatively large-scale turbulence expected at those Reynolds numbers, and that the energy dissipation in that layer will still be dominant, leading to the continued dependence of the total pressure drop on the  $(\frac{3}{2})$ th power of velocity. In any case this particular region of Reynolds number is one where the pressure drop due to turbulence itself varies approximately according to such a power law.

Figure 6 indicates the conclusions on the distribution of total pressure drop that follow from the assumption that its ratio to that in Poiseuille flow continues to take the value (1) wherever that exceeds unity, and is otherwise (that is, for Reynolds numbers less than about 40) equal to unity. Note how considerable is the predicted enhancement of total pressure drop above its Poiseuille value even in case (a) (light breathing at 10 L/min) while in case (b) (heavy breathing at 100 L/min) it becomes quite intense. This in turn enhances still more the dominance of the first ten generations in determining total pressure drop.

The entire drop in total pressure through all the generations is given as a function of flow rate in Figure 7 by the solid dots, with the open dots showing the linear dependence that would be predicted from Poiseuille's law. The experimental data of Hyatt & Wilcox (1963) and Ferris, Mead & Opie (1964) fit rather well the predictions from the solid dots.

We may conclude from the work of Pedley, Schroter & Sudlow (1970) that flow patterns in human bronchi are in general greatly distorted, with high velocity peaks occurring very near tube walls and most of the viscous energy dissipation occurring in the intervening high-shear boundary layers. Human lung resistance, both in its magnitude and in its dependence on flow rate, is determined primarily by this special characteristic of bronchial flow patterns that mainly results from their small ratios of the distance between bifurcations to the diameter of the intervening tube.

### **3. Lung flow patterns in birds**

The lungs of birds, as indicated in § 1, exhibit some very different patterns of flow; partly because of their connections with air sacs, whose expansions and

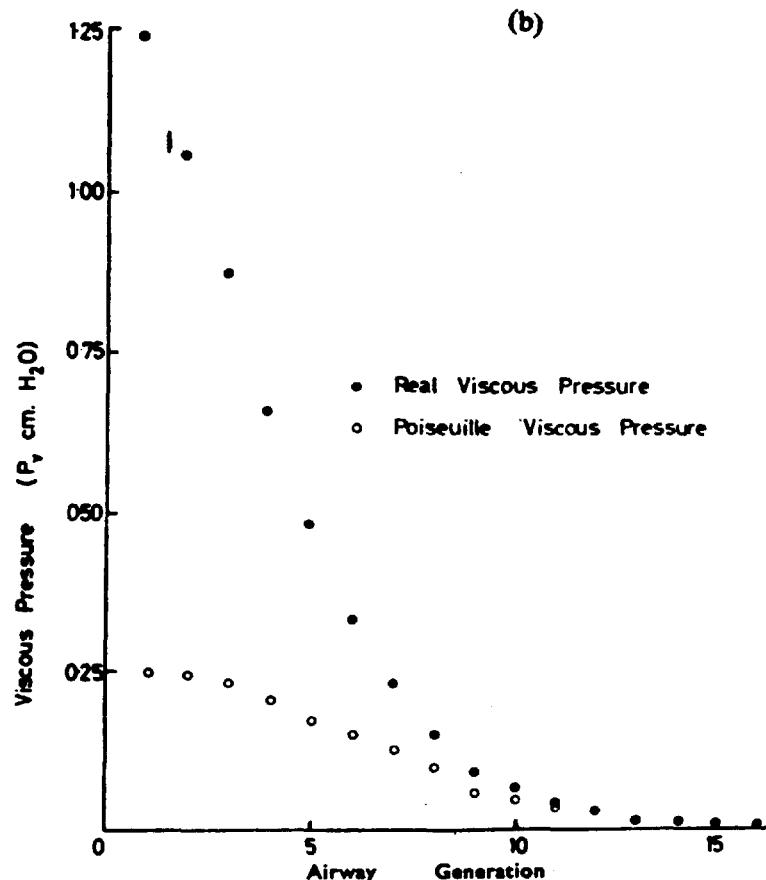
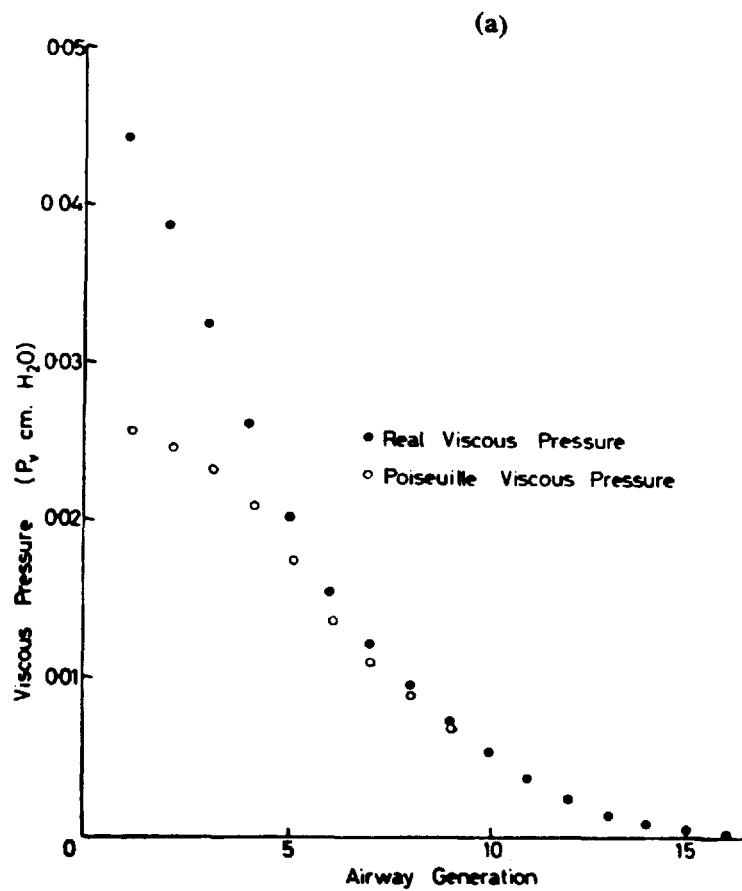


FIG. 6. Predictions by Pedley, Schroter & Sudlow (1970) of 'viscous pressure' (their abbreviated expression for the distribution of the drop in total pressure, due to viscous action, through the lung airways) in inspiration at (a) 10 L/min (b) 100 L/min.

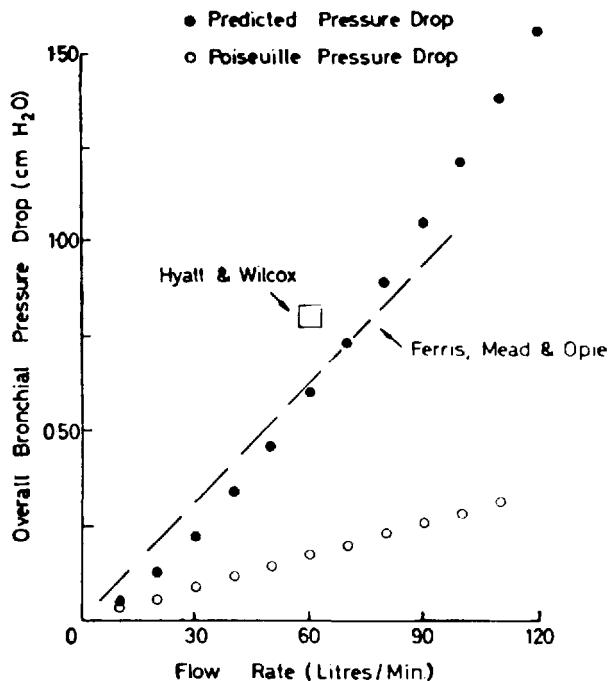


FIG. 7. This Figure 4 from Pedley, Schroter & Sudlow (1970) gives a comparison with experimental data.

contractions energize the flow through essentially constant-volume lungs; partly because junction geometries are very dissimilar; and above all because the branching arrangements, far from constituting (in the topological sense) a 'tree', contain numerous loops. As a consequence, we can speak of a 'circulation' of the air within the lung-air-sac-system, with the air sacs behaving as 'hearts' driving that circulation.

In this section we briefly outline the extensive store of anatomical data regarding bird lungs, the much more limited data on flow patterns within them, and some speculation on how those patterns should be interpreted; leaning partly on the 1971 Göttingen symposium (see all those references from Volume 14 of *Respiration Physiology*) and partly on an unpublished discussion on the subject held at Imperial College, London in March 1973. The major biofluidodynamic question that emerges can be crudely phrased, "How can 'hearts' work without valves?"—for the extensive anatomical data show that the air sacs possess no anatomical features of a valve-like character and yet the fluid dynamical work demonstrates a one-way flow through much of the lung, generated by their expansions and contractions. Hazelhoff (1943) who first postulated such a one-way flow put the question in the form, "Can there be an, as it were, *aerodynamic valving?*", and biofluidodynamicists should rise to the challenge of this question and look for sound aerodynamic interpretations of those observations that have since confirmed his hypothesis.

Duncker (1972) describes anatomical data which he obtained from making casts of the lung-air-sac system in birds of some 155 different species from 47 different families, using more than one specimen from all except 15 of those species. His work brings to light two types of gas-exchange region in bird lungs: one, which he calls the paleopulmo, is present in all families of birds, its broad structural features being remarkably similar in each, whereas that which he calls the neopulmo is absent in more primitive families such as Spheniscidae (penguins) and Dromiceiidae (emus), has an intermediate degree of development in families such as Ciconiidae (storks), Columbidae (pigeons), Anatidae (ducks) and Laridae (gulls), and a quite extensive development (where, however, it still occupies only 25 to 30% of lung volume) in families such as Phasianidae (fowls) and Turdidae (thrushes).

The paleopulmo, which his evidence points to as having been the original respiratory system in the birds, is the structure possessing the remarkable property above mentioned; namely, that expansion and contraction of the valveless air sacs can generate a one-way flow through its gas-exchange units: the *parabronchi*, with diameters of order 1 mm, out of which open the 'air capillaries' referred to in § 1. On the other hand, the air moves back and forth within the parabronchi of the neopulmo, where that is present.

Before describing the anatomy, we can make the mathematical point that such flow in the neopulmo is responding in an essentially linear manner to the pulsations of the air sacs, although the response of flow in the paleopulmo is so strongly nonlinear as to generate a rectification. A quadratic nonlinearity would, for example, be effective for this purpose and we shall find that the anatomy does indeed suggest a possible mechanism depending quadratically on the pulsations. A square-law dependence of air-capillary ventilation on amplitude of respiratory movements could be valuable in modes of life like those of birds where respiratory requirements may sometimes rise to very high values indeed, but it would perhaps have a minor disadvantage in that substantial amplitudes would still be needed even in resting conditions. Duncker suggests that a neopulmo, responding linearly to air-sac pulsation, may have developed to supply the respiratory requirements of birds at rest with minimal movements while the paleopulmonary ventilation, responding quite steeply to increases of amplitude at much higher levels, has continued to supply the birds' large respiratory requirements during flight.

We now attempt a brief synopsis of Duncker's detailed anatomical descriptions, concentrating mainly on the paleopulmo that raises the biofluidodynamically interesting question of the mechanism of rectification. Figure 8 is a drawing of a cast of the left lung and its five associated air sacs from a domestic fowl *Gallus domesticus*. The *posterior sacs* (the posterior thoracic air sac D and the abdominal air sac E) are those which principally excite respiratory flow, by their large expansions and contractions in response to muscular action producing coordinated motions of the breastbone and the posterior ribs. The *anterior sacs* (the cervical air sac A, the interclavicular air sac B and the anterior thoracic air sac C) make rather smaller pulsations in response to the same motions, while the movements of certain other muscles in antiphase with the above help to keep nearly constant the volume of the lung itself (see § 1).

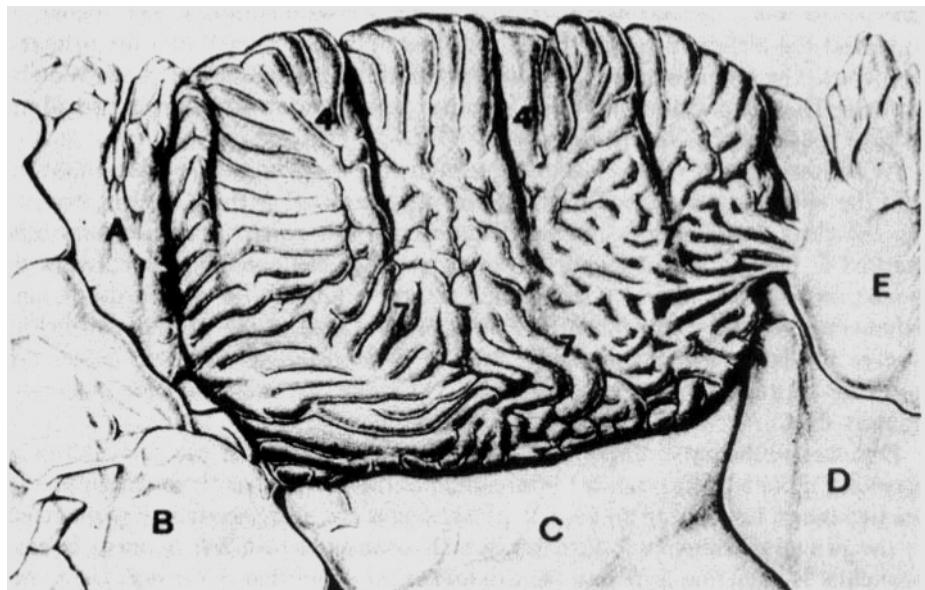


FIG. 8. This Figure 1 from Duncker (1972) introduces the discussion of bird lung anatomy.

To help in an initial appreciation of the internal structure of the lung and especially of the paleopulmo, we show in Figure 9 Duncker's drawing of the same cast as in Figure 8 but with most of the neopulmo removed. Tube 2 is the primary bronchus supplying the lung with all its air. At its anterior end (left on the diagram)

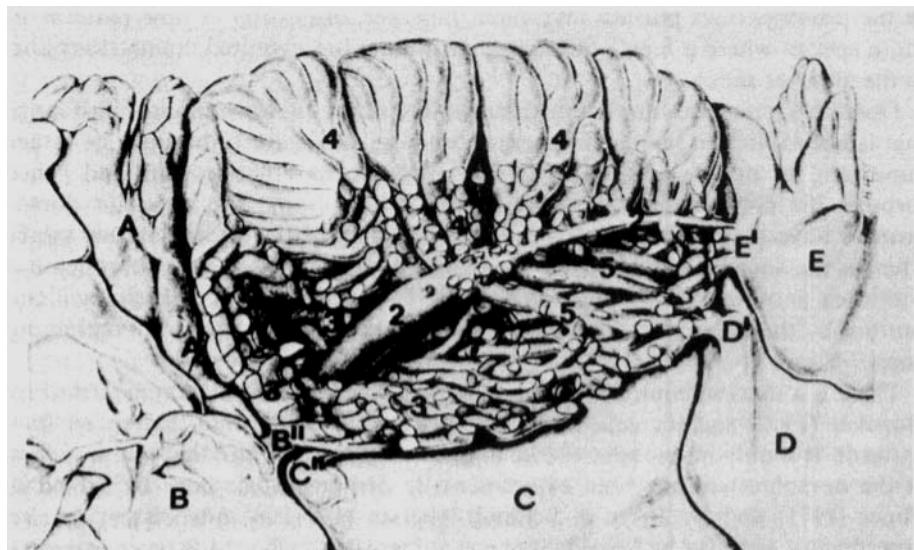


FIG. 9. This Figure 2 from Duncker (1972) shows more internal detail for the lung cast in his Figure 1.

emerge the four *ventrobronchi* marked 3: tubes whose function, it will appear, is to collect the airflow through the paleopulmo and pass it back *into* the primary bronchus. The anterior sacs A, B and C are each connected to one of the ventrobronchi. Their expansions and contractions may be quite significant in providing storage space for that returning flow.

At its posterior end (right on the diagram), before the primary bronchus exhausts into the abdominal air sac E (at the point E'), many other tubes emerge from it. On the dorsal side (upper on the diagram) are the seven to ten *dorsobronchi* marked 4. The gas exchange surface of the paleopulmo consists of a network of small tubes (the *parabronchi*), fed by the dorsobronchi and their ramifications, and exhausting into the ventrobronchi. On the lateral side of the primary bronchus emerge the laterobronchi marked 5: the *large laterobronchus* which enters the posterior thoracic air sac D (at the point D') is the one important for paleopulmonary flow.

Duncker's schematic diagram of the paleopulmo and its air sac connections is given in Figure 10. The trachea 1 bifurcates into the two primary bronchi supplying the two lungs. In the lung illustrated, pulsations of the posterior sacs, E (connected to the primary bronchus 2 directly) and D (connected to it by the large laterobronchus 5), generate flow into the dorsobronchi 4 and thence through the parabronchial network 6 into the ventrobronchi 3; this flow is assisted by smaller pulsations of anterior sacs (especially B and C) connected to ventrobronchi, and exhausts into the primary bronchus.

We briefly note here without detail or diagrams how a neopulmo where present (as it is in some degree in the great majority of bird species) fits into the above picture. Essentially, an additional parabronchial network (gas exchange region) is fitted in between the posterior part of the primary bronchus 2 and the posterior sacs D and E. This region is ventilated directly by air movements back and forth as the posterior sacs pulsate. (We omit, however, discussion of flow patterns in those species where a *highly* developed neopulmo has acquired connections also to the anterior sacs.)

Duncker emphasizes one geometrical feature of the anatomy of most bird lungs that is not explicit in the block-diagram topology of Figure 10 but may be rather important for maintaining one-way flow through the dorsobronchi (and hence through the parabronchial network of the paleopulmo). The posterior dorsobronchi have their entrances *facing in the direction* of the abdominal air sac E, whereas the anterior dorsobronchi are placed opposite the large laterobronchus 5 with their entrances facing towards it. These two features may, perhaps, facilitate capture by the dorsobronchi of a large part of the outflow from the contracting posterior sacs D and E.

There is a marked contrast between the extensive anatomical data described by Duncker (1972) and the much more limited extent of the information on flow patterns. It is only in one species (the duck *Anas platyrhynchos*) that one-way flow in the dorsobronchi has been experimentally demonstrated: both by Scheid & Piiper (1971) and by Bretz & Schmidt-Nielsen (1971) in independent *in vivo* experiments, and also by Scheid, Slama & Piiper (1972) who made more extensive

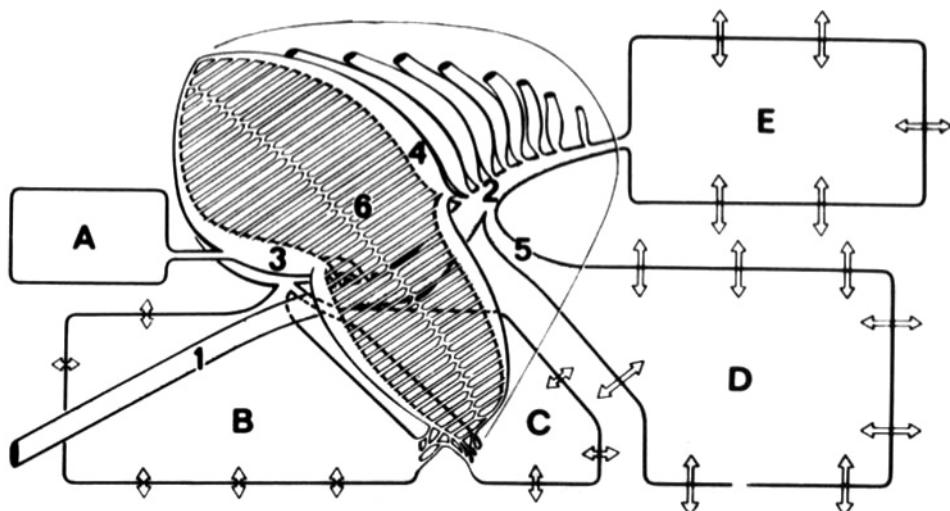


FIG. 10. This Figure 9 from Duncker (1972) gives his schematic diagram of the paleopulmo.

measurements in excised duck lungs and who give a good discussion of the evidence as a whole. At the same time they stress the likelihood that one-way flow can be expected to be general in the dorsobronchi of bird lungs because broad structural features of the paleopulmo show so much similarity throughout the class as a whole.

Figure 11 gives Scheid, Slama & Piiper's schematization of the paleopulmo: the lower diagram is at the same level of schematization as Figure 10 above (the primary bronchus being here called 'mesobronchus'). The upper figure lumps together elements in the block diagram still more drastically; lumping the action of the two posterior sacs (D and E in Figure 10) into that of one 'caudal air sac' and the action of the three anterior sacs into that of one 'cranial air sac'; and, similarly, representing both the dorsobronchi and the ventrobronchi in lumped-element form. The conclusions of all the flow-direction experiments are neatly summarized in this upper block diagram.

Note that during inspiration (black arrows) the air at the posterior end of the primary bronchus, besides being drawn into the expanding posterior sacs, flows out through the dorsobronchi into the parabronchial network of the paleopulmo. The less rapidly expanding anterior sacs store all the outflow from that network: none of it is returning into the primary bronchus, from which *also* there is indeed some airflow during inspiration into the anterior air sacs. In expiration (white arrows) the contracting posterior sacs force air through the dorsobronchi as well as through the primary bronchus. The ventrobronchi collect that air after it has flowed through the parabronchial network and exhaust it, together with the outflow from the anterior sacs, into the primary bronchus.

Beyond the substantial amount of data on flow *direction* so summarised, there is a shortage of *quantitative* information about the flow distribution and its

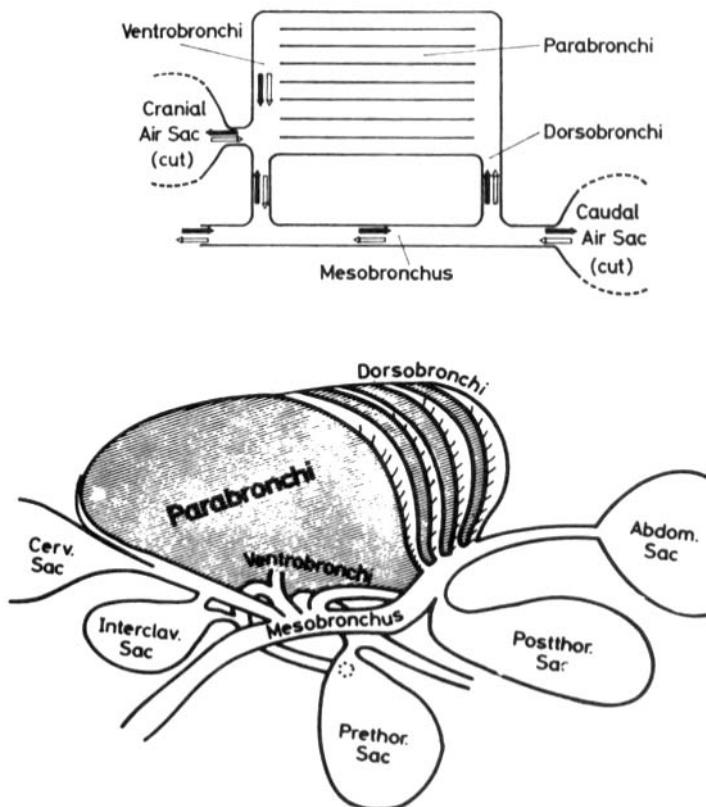


FIG. 11. These Figures 2 and 3 from Scheid, Slama & Piiper (1972) give a further schematization of the avian respiratory system, and a summary of their flow-direction data.

dependence on the amplitude of air-sac pulsation. This is an area where experimental data will be difficult and slow to obtain. It would be desirable in the meantime to hammer out a fluidodynamically sound theoretical framework to which limited data might be fitted so as to produce more widely applicable information on respiration in birds out of the combination of inferences from experiment and theory.

We conclude with some brief comments on the type of theory suggested as perhaps necessary by the facts summarized in this section. Certainly it would differ from those theories characteristic of internal biofluidynamics generally, owing to the absence of a tree-like topology. On the other hand, it might apply ideas culled from the theory of 'fluidic' devices, where there are similar considerations of the tendency of a jet-type flow to shoot across a gap and of the amount of suction needed to deflect it.

A simple experimental model along fluidic-device lines and exhibiting rectification was described by Professor Piiper at the Imperial College symposium.

Illustrated in Figure 12, it yields rectified (one-way) flow from Q to R to P on application of a pressure difference between A and B oscillating about a zero mean. (Here B can represent a posterior sac, although the model lacks any representation of an anterior sac.)

A crude interpretation of this result is that the flow when driven from A to B shoots past the opening at P but comes to rest at Q, making the difference between the total pressure at Q and the static pressure at P available to drive the flow from Q to R to P. This difference has a quadratic dependence on flow as discussed earlier; a more refined estimate taking into account 'entrainment' considerations at P might also be nearly quadratic. A substantial flow could result if the effective lumped resistance of QRP were not too great.

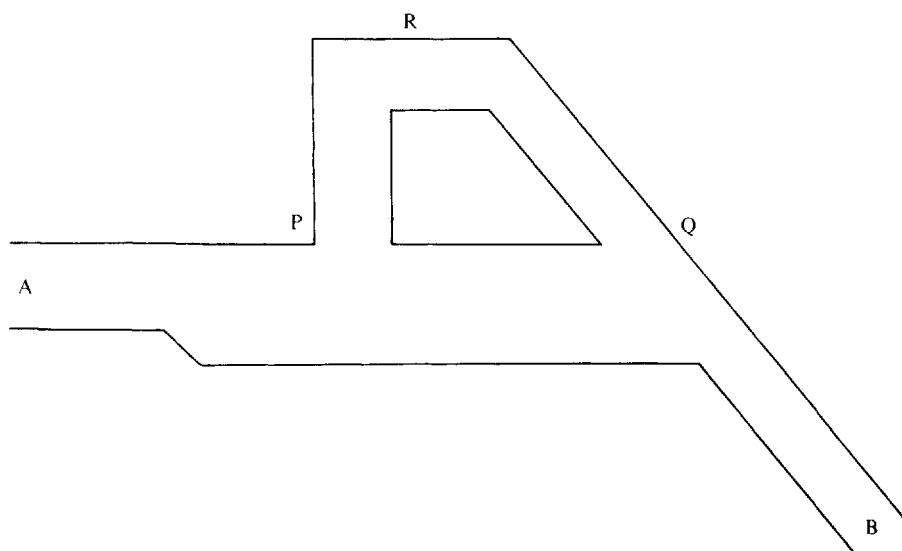


FIG. 12. *Experimental model used by Dr. P. Scheid and Professor J. Piiper.*

Similarly, the flow when driven from B to A would show a preference for shooting past the opening at Q and flowing on towards R. Crudely estimating the extent of this preference by imagining a high resistance inserted at R, we see that the flow through it would be forced by the difference between a total pressure in QR and the corresponding static pressure (or possibly slightly less due to entrainment or other effects) at P: an argument again suggesting a nearly quadratic response. This model of expiration may represent to some degree the effect of the strong flows exhausting from the posterior sacs of bird lungs being geometrically directed (as noted above) into the dorsobronchi; but the model of inspiration is more questionable as it neglects all the influence of the anterior sacs.

It is too soon to foretell what will prove to be the necessary features of a broadly satisfactory theoretical or experimental model of these motions. What does seem certain is that looking for such models will prove to be a particularly interesting

and rewarding new area of biofluiddynamic research work, preferably to be pursued in collaborative studies between zoologists and veterinary scientists on the one hand, and fluiddynamic experimenters and theorists on the other.

## REFERENCES

- BRETZ, W. L. & SCHMIDT-NIELSEN, K. 1971 J. Exptl. Biol. **54**, 103–118.  
DUNCKER, H. R. 1972 Respiration Physiol. **14**, 44–63.  
FERRIS, B. G., MEAD, J. & OPIE, L. H. 1964 J. Appl. Physiol. **19**, 611–622.  
HAZELHOFF, E. H. 1943 Versl. gewone Vergad. Afd. Natuurk. Kon. Ned. Akad. Wet. **52**, 391–400.  
English translation (1951): Poultry Sci. **30**, 3–10.  
HUGHES, G. M. 1972 Respiration Physiol. **14**, 1–25.  
HYATT, R. E. & WILCOX, R. E. 1963 J. Clin. Invest. **42**, 29–39.  
MARSHALL, N. B. 1965 *The Life of Fishes*, Weidenfeld & Nicolson, London.  
PEDLEY, T. J., SCHROTER, R. C. & SUDLOW, M. F. 1970 Respiration Physiol. **9**, 371–405.  
PIIPEL, J. & SCHEID, P. 1972 Ibid. **14**, 115–124.  
SCHEID, P. & PIIPEL, J. 1971 Ibid. **11**, 308–314.  
SCHEID, P., SLAMA, H. & PIIPEL, J. 1972 Ibid. **14**, 83–95.  
SCHROTER, R. C. & SUDLOW, M. F. 1969 Ibid. **7**, 341–355.  
WEIBEL, E. R. 1972 Ibid. **14**, 26–43.

# CHAPTER 12

## Pulse Propagation Theory

### 1. Introduction

The theory of pulse propagation in arteries is a special case of the general theory of one-dimensional waves in fluids. In that general theory the local accumulation of fluid can be resisted either by the elastic properties of a containing vessel as in pulse waves, or by gravity as in long waves in open channels, or by compressibility as in sound waves. In every case the character of the wave propagation is determined by the balance between the resisting force and the inertia of the fluid. Here we give the theory for the physiologically interesting case of an effectively incompressible fluid (blood) contained within distensible vessels while briefly noting how the results apply with very slight modifications to other systems.

We consider then blood of constant density  $\rho$  contained within the cardiovascular system of a stationary animal and denote by  $p_0$  the hydrostatic distribution of blood pressure. This hydrostatic value is the distribution that would be present in the absence of pumping action by the heart and is usually taken as equal to atmospheric pressure at the level of the aortic valve, while increasing at a rate  $\rho g$  per unit distance below that level. The excess pressure

$$p_e = p - p_0 \quad (1)$$

represents the effect of the heart's pumping action in generating at any point of the system a pressure in excess of that hydrostatic distribution.

The order of magnitude of the excess pressure in mammalian arteries is  $10^4 \text{ N m}^{-2}$  (or  $10^{-1}$  times a standard atmospheric pressure). This pressure increase, in normal physiological conditions, produces an increase in the arterial lumen  $A$  (internal cross-sectional area) of order 10%. The distensibility  $D$ , defined as the relative area increase per unit increase in excess pressure

$$D = A^{-1} dA/dp_e, \quad (2)$$

is thus of order  $10^{-5} \text{ N}^{-1} \text{ m}^2$ , although its actual magnitude shows substantial variation between individuals, and within one individual tends to take values that are greatest in the aorta, while decreasing towards peripheral arteries. Furthermore, while the complicated composite structure of the arterial wall includes fibres of an elastomer, elastin, that permit distensions of this order under moderate pressure changes, the associated arrangements of stiffer collagen and 'smooth-muscle' fibres substantially reduce the increment of area produced by pressure change around the highest values of  $p_e$ .

Pulse propagation can be influenced by variations along an individual artery, either in the distensibility  $D$  or in the undisturbed value of the lumen area  $A$ , as

well as by viscous effects, by the influence of arterial branchings, and by this dependence of  $D$  upon  $p_e$  itself. In this introduction, however, we outline only the simple linear theory, with viscous effects neglected, for propagation along a single artery with uniform undisturbed area  $A$  of pulses weak enough for the distensibility  $D$  to be regarded as a constant and for the convective term  $\mathbf{u} \cdot \nabla \mathbf{u}$  in the fluid acceleration to be neglected. Extensions to branching systems are given in § 3 and § 4, after the stress systems in the arterial wall giving rise to its distensibility have been described in § 2. The effects of variations along an individual artery, and of viscosity, are then described in § 5, while some nonlinear effects are indicated in § 6.

We use  $x$  as a coordinate specifying *longitudinal* position (along the length of the artery) and  $u$  as the associated velocity component. More strictly, we may define  $u$  at any one instant as the average value of the longitudinal velocity over a given arterial cross-section, although noting that since variations over the cross-section are due principally to blood viscosity such a strict definition of  $u$  can become important only in § 5.

Newton's second law of motion states that mass times longitudinal acceleration, which with convective acceleration neglected is  $\rho \partial u / \partial t$  per unit volume of fluid, equals longitudinal force. With viscous effects neglected, the longitudinal force per unit volume of fluid is  $-\partial p_e / \partial x$ , where only the gradient of the excess pressure (1) appears because the gradient of the hydrostatic pressure  $p_0$  is exactly balanced by the weight of the fluid. This linearized momentum equation,

$$\rho \partial u / \partial t = -\partial p_e / \partial x, \quad (3)$$

must be supplemented by a linearised equation of continuity,

$$D \partial A / \partial t = -\partial u / \partial x, \quad (4)$$

where the left-hand side, by the definition (2) of distensibility, is the relative rate of increase  $A^{-1} \partial A / \partial t$  of the cross-sectional area of fluid, which for an incompressible fluid must be minus its average relative rate of extension  $\partial u / \partial x$  at right angles to the cross-section.

Eliminating the velocity  $u$  from (3) and (4), we obtain the wave equation

$$\partial^2 p_e / \partial t^2 = c^2 \partial^2 p_e / \partial x^2, \quad (5)$$

where the wave speed  $c$  takes the value

$$c = (\rho D)^{-1/2}. \quad (6)$$

The order of magnitude of this predicted wave speed, with  $\rho = 10^3 \text{ kg m}^{-3}$  and  $D$  of order  $10^{-5} \text{ N}^{-1} \text{ m}^2$ , is  $10 \text{ ms}^{-1}$ , in agreement with typical observed values. Distensibilities elevated to around  $4 \times 10^{-5} \text{ N}^{-1} \text{ m}^2$ , on the other hand, may be observed in a healthy thoracic aorta, together with measured wave speeds (for artificially generated pressure signals) duly reduced to around  $5 \text{ ms}^{-1}$  (McDonald 1960).

We note briefly that the above theory applies without change to the theory of long gravity waves in open channels of uniform cross-sectional water area  $A$

and free-surface breadth  $B$ , for which a small excess pressure  $p_e$  raises the free surface a distance  $p_e/\rho g$  and so increases the cross-section by  $Bp_e/\rho g$ , giving by (2) and (6)

$$D = B/\rho g A \quad \text{and} \quad c = (gA/B)^{1/2}; \quad (7)$$

the familiar value of the wave speed for such long gravity waves. We also note that the compressibility  $K$  of the fluid, if taken into account, simply appears added to  $D$  on the left-hand side of (4). Equation (6) then becomes  $c = [\rho(D + K)]^{-1/2}$ . Of course, compressibilities of liquids, including blood, are less than  $10^{-9} \text{ N}^{-1} \text{ m}^2$ , so that this correction is totally negligible in the cardiovascular system. For fluids in metal pipes, by contrast,  $D$  is considerably smaller than  $K$  and the theory then coincides with the theory of sound propagation at the wave speed  $c = (\rho K)^{-1/2}$ .

In all cases the solution of equation (5) for  $p_e$  which represents an arbitrary waveform propagating in the direction  $x$  increasing can, with the associated solution of (3) for  $u$ , be written

$$p_e = f(t - c^{-1}x), \quad u = (\rho c)^{-1}f(t - c^{-1}x). \quad (8)$$

The corresponding solutions for a wave propagating in the direction  $x$  decreasing are

$$p_e = g(t + c^{-1}x), \quad u = -(\rho c)^{-1}g(t + c^{-1}x). \quad (9)$$

The difference of sign in the two equations for  $u$  is important. It can, for a pure travelling wave propagating in either direction, be interpreted by saying that the fluid velocity *in that direction* is everywhere equal to  $(\rho c)^{-1}$  times excess pressure. A general motion in the tube, however, is a linear combination of (8) and (9), for which the relationship between fluid velocity and excess pressure is evidently more complicated. Its implications will be pursued in § 3.

## 2. Wall stress systems associated with pulse propagation

In the meantime we discuss briefly the systems of stress with which the arterial wall reacts to excess pressures in the propagating blood pulse: that is, the systems that give the artery its distensibility. This excursion into 'biosolidmechanics' is not included for the purpose of inferring the value of the distensibility  $D$  (a quantity relatively easy to measure) from other material constants for the arterial wall (quantities that may be much *harder* to measure). Its main object, rather, is to influence analysis of the factors that may govern the health or disease of arterial walls under the influence of various types of movement of blood within their lumen: an analysis mentioned in Chapter 10 and pursued in greater detail in Chapter 13. The present discussion is placed before the main studies of pulse propagation, however, because a deeper feel for the nature of the wall's response can help in understanding those.

The ratio  $h/a_0$  of arterial wall thickness to the undisturbed radius  $a_0$  of the lumen lies typically in the range from 0.1 to 0.2, so that ideas from the theory of thin-walled elastic tubes can suitably be applied, provided that no unwarrantable assumptions of *isotropic* elasticity are made. The essential result that is required is

known to physiologists as 'Laplace's law': a piece of nomenclature that workers in theoretical mechanics, if they regard many different ideas of great importance to their subject as stemming from Laplace's work, may find surprising but, perhaps, may be well advised to learn!

Laplace's law states that an excess pressure  $p_e$  inside a tube of internal radius  $a_0$  must generate† a circumferential tension  $a_0 p_e$  per unit length of tube. This relatively unsurprising fact is perhaps most easily comprehended by imagining such a unit length divided into two equal parts, with semicircular cross-sections, pulled apart by the internal pressure  $p_e$  acting upon a resolved area  $2a_0$  with a resultant force  $2a_0 p_e$  balanced by the circumferential tension acting at both joins between the parts.

This circumferential tension  $a_0 p_e$  pulls on an area of wall material (again per unit length) equal to the wall thickness  $h$ ; it follows that the circumferential tensile stress  $\tau_1$  (force pulling on unit area normal to that force), averaged across the wall thickness, is

$$\bar{\tau}_1 = a_0 p_e / h. \quad (10)$$

In arteries and other thin-walled tubes with  $h/a_0$  small, the mean circumferential tensile stress (10) greatly exceeds the magnitude of the radial stress (which varies from  $p_e$  inside to 0 outside).

This suggests an interesting approximate view of the stress system in arterial walls under internal pressure, as a stress system entirely dominated by circumferential tension (sometimes called 'hoop tension'). Although we revise this approximate view a little later, we may pursue briefly the idea that the stress system in the arterial wall is exactly as in a simple tensile test, with the tension applied in the circumferential direction. The corresponding circumferential strain takes the value

$$(a - a_0)/a_0 \quad (11)$$

if the internal radius changes from  $a_0$  to  $a$ , since variations in this strain across a small enough thickness of wall are negligible (see below). Therefore, if the Young's modulus for the tensile test performed in the circumferential direction (isotropic elasticity is not assumed) is  $E$ , then  $\tau_1 = E(a - a_0)/a_0$  is the tensile stress at each point, and  $\bar{\tau}_1 = \bar{E}(a - a_0)/a_0$ , where  $\bar{E}$  is average of  $E$  across the tube thickness, is the average tensile stress (10).

The relative change in lumen radius is therefore

$$(a - a_0)/a_0 = a_0 p_e / \bar{E}h, \quad (12)$$

and the relative change in lumen area is twice as much, giving

$$D = 2a_0 / \bar{E}h \quad (13)$$

---

† The stress system 'generated' by an excess pressure  $p_e$  means the stress system when that excess pressure acts minus the stress system for zero excess pressure. This is what affects the distensibility  $D$  given by (2).

for the distensibility (2), and a pulse wave velocity (6) equal to

$$c = (\bar{E}h/2\rho a_0)^{1/2}. \quad (14)$$

This agrees with the classical Moens-Korteweg result provided that their homogeneous and isotropic Young's modulus  $E$  is replaced by an average across the wall thickness of the Young's modulus for the wall material in a *circumferential* tensile test.

The above approximate view of the stress system and its consequences is reasonably satisfactory for most purposes. Readers familiar with the theory of elasticity may at this point feel concerned, however, that no mention has been made of one well-known fact: extension in that direction in which the tensile test is applied is accompanied by *contraction* in any perpendicular direction, smaller by a factor which for isotropic materials is known as *Poisson's ratio*  $\sigma$ . Here we briefly analyse the implications of such contraction, but without assuming isotropic elasticity.

Such contraction in the *radial* direction, involving a negative radial strain smaller in magnitude than the circumferential strain (11), is unimportant: it produces a change in wall thickness less than the product  $h[(a - a_0)/a_0]$  of two small quantities. This supports the previous assumption that the circumferential strain (11) is approximately constant, since it implies, for example, that the relative changes in the external and internal radii of the tube differ only by a factor  $1 + O(h/a_0)$ .

Some more interesting problems are raised, however, by the *longitudinal* contraction of the arterial wall that would accompany simple circumferential tension if that were indeed the only response to internal pressure. If we retain the subscript 1 to denote circumferential tension or extension, and use subscript 2 for longitudinal tension or extension, and in the notation of Love (1927, Chapter 3) let the matrix element  $C_{ij}$  denote the  $i$ th component of strain generated by a unit increment in the  $j$ th component of stress, then  $C_{11}$  is the circumferential extension per unit circumferential tension, which is  $E^{-1}$  in the notation used above. On the other hand the quantities  $C_{12}$  and  $C_{21}$ , whose equality is necessitated by conservation of energy, represent firstly the circumferential extension per unit longitudinal tension and secondly the longitudinal extension per unit circumferential tension; in isotropic materials they take the value  $-\sigma E^{-1}$ . The quantity  $C_{22}$ , again, is the longitudinal extension per unit longitudinal tension and need not be equal to  $C_{11} = E^{-1}$ , although for a stable material it must satisfy the inequality

$$C_{11}C_{22} > C_{12}^2. \quad (15)$$

In simple circumferential tension  $\tau_1$ , the *longitudinal* strain in the arterial wall would be

$$C_{12}\tau_1 = C_{12}E(a - a_0)/a_0 = \frac{1}{2}C_{12}E(A - A_0)/A_0. \quad (16)$$

This may be compared with the longitudinal strain  $-(A - A_0)/A_0$  in the fluid itself: it is smaller only by a factor  $(-\frac{1}{2}C_{12}E)$  of order around 0.2 for a ratio  $C_{12}/C_{11}$  typical of rather incompressible biological materials. Such a conclusion, implying longitudinal movements of the arterial wall as much as 0.2 times those of the fluid, is extremely hard to accept.

In reality, the soft tissue within which an artery is embedded, although compliant enough for its pressure to respond negligibly to the artery's pulsations (as assumed in deriving Laplace's law), is probably quite effective for 'tethering' the arterial wall against such longitudinal displacements. If so, then its tethering action must supply a cancelling strain  $-C_{12}\tau_1$  equal and opposite to (16). It can do this by applying forces generating in the arterial wall an associated longitudinal stress, which must take the value

$$-C_{12}\tau_1/C_{22}, \quad (17)$$

since  $C_{22}$  is the longitudinal strain per unit longitudinal stress. From this longitudinal stress (17), however, is generated an additional circumferential strain equal to  $C_{12}$  times it, which must be added to the circumferential strain  $C_{11}\tau_1$  associated directly with the circumferential stress  $\tau_1$  to give the total circumferential strain

$$(a - a_0)/a_0 = C_{11}(1 - C_{12}^2/C_{11}C_{22})\tau_1. \quad (18)$$

Thus the constraint preventing longitudinal displacement of the wall has the effect of increasing the circumferential stress  $\tau_1$  from  $E(a - a_0)/a_0$  (where  $E = C_{11}^{-1}$ ) to  $E'(a - a_0)/a_0$ , where

$$E' = E/(1 - C_{12}^2/C_{11}C_{22}) \quad (19)$$

is an effective Young's modulus increased by a factor which is positive by (15) and in practice represents something like a 20% increase. This means that  $\bar{E}$  must be replaced by  $\bar{E}'$  in equations (12), (13) and (14): in particular, the application of this constraint (like any other) increases the wave velocity, even though by only about 10%.

Measurements of  $c$ ,  $D$  and  $E$  under identical conditions are broadly in accord (McDonald 1960) with equations (13) and (14), but insufficiently accurate to distinguish between them and the forms modified by substituting  $\bar{E}'$  for  $\bar{E}$ . The above discussion has been included not to lay emphasis on that minor theoretical modification of those equations but to emphasize that the stress system in an arterial wall associates with an excess pressure  $p_e$  not only the mean excess circumferential tension (10) but also a proportional excess longitudinal tension (17), generated by the surrounding connective tissue so as to prevent any substantial longitudinal displacements.

### 3. Propagation through a junction

It is now necessary to extend the theory of pulse propagation along a single artery as outlined in § 1 and § 2 so that propagation through a junction between arteries can be treated. In later sections the results of this treatment are used to study propagation through the branching arterial system.

The appropriate methods, as M. G. Taylor (1965, 1966) especially emphasized, have much in common with the methods for analysing electrical transmission-line networks. Those are dominated by the Kirchhoff laws at each node: that the electric potential varies continuously, and that the total current flows into and out of a node are equal. The corresponding rules at a junction in the cardiovascular system are that excess pressure is continuous across the junction, while the total volume flows of blood into and out of the junction are equal.

We can crudely estimate the relative error in these assumptions in terms of the linear dimension  $l$  of a junction and a characteristic radian frequency  $\omega$  of the pulse. Viewing any gradient of excess pressure by (3) as of characteristic magnitude  $\rho\omega u$ , where the fluid velocity  $u$  by (8) is of characteristic magnitude  $p_e/\rho c$ , we may estimate the change in  $p_e$  across the junction as  $l$  times such a gradient, or  $(\omega l/c)$  times a characteristic value of  $p_e$  itself. Similarly, the unbalance between the total volume flows into and out of a junction is equal to the rate of change of fluid volume within the junction region, which may be estimated as  $l$  times a characteristic rate of change of cross-sectional area  $ADdp_e/dt$  (by the definition (2) of distensibility  $D$ ). Viewing  $dp_e/dt$  as of characteristic magnitude  $\omega\rho cu$  and using (6) to replace  $D$  by  $\rho^{-1}c^{-2}$ , we may estimate the unbalance in volume flow as  $(\omega l/c)$  times a characteristic value  $Au$  of the volume flow itself.

The relative error in both cases is estimated as  $\omega l/c$ , which is very small for junctions whose dimension satisfies the condition:

$$l \text{ very small compared with } c/\omega. \quad (20)$$

That length  $c/\omega$  is as much as 1 m for typical values  $c$  of order  $10 \text{ ms}^{-1}$  (§ 1) and  $\omega$  of order  $10 \text{ s}^{-1}$  (namely,  $2\pi$  times a characteristic frequency in Hz), which suggests a very small relative error from assuming pressure continuity and volume-flow balance at junctions with dimension of order  $10^{-2} \text{ m}$ .

Before analysing junctions really typical of the cardiovascular system, we briefly raise a somewhat artificial question: what happens to a pulse travelling along one uniform tube when it reaches a junction with one other uniform tube of different distensibility and cross-sectional area? We may choose the  $x$ -coordinate so that the junction is at  $x = 0$ , and consider first whether a pulse propagating in the direction  $x$  increasing (as described in equations (8)) could travel straight through such a junction (adjusting its wave speed to the distensibility of the new tube) without generating any 'reflected' pulse travelling in the opposite direction.

To answer this first question we note that the excess pressure  $p_e$  and the volume flow  $J = Au$  in the direction  $x$  increasing in the travelling wave (8) take at  $x = 0$  the values

$$p_e = f(t), \quad J = Yf(t), \quad (21)$$

where we define

$$Y = A/\rho c \quad (22)$$

as the *admittance* of the tube: that is, the ratio of volume flow to excess pressure for a wave travelling in that direction in which the volume flow is measured.

Equations (21) imply that travelling waves in two tubes satisfy (i) the condition of continuity of excess pressure at the junction  $x = 0$  if and only if they have the same waveform  $f(t)$ , and (ii) the condition of balance between the volume flows into and out of the junction if and only if the tubes have the *same admittance*  $Y$ . Such ‘admittance matching’ (as when an area increase is *matched* by a wave-speed increase due to decreased distensibility) is the necessary and sufficient condition for a pulse to pass through the junction without generating any reflected pulse.

In a general case when the admittances are not matched, we use subscript 1 for values in the tube  $x < 0$  (with cross-section  $A_1$  and wave speed  $c_1$ , and therefore admittance  $A_1/\rho c_1 = Y_1$ ) and similarly subscript 2 for values in the tube  $x > 0$ . The general solution of the wave equation (5), allowing for the possibility of a reflected wave, is the sum of (8) and (9), which in the tube  $x < 0$  gives

$$\begin{aligned} p_e &= f(t - c_1^{-1}x) + g(t + c_1^{-1}x), \\ J &= Y_1[f(t - c_1^{-1}x) - g(t + c_1^{-1}x)]. \end{aligned} \quad (23)$$

Fluid in the tube  $x > 0$ , on the other hand, is supposed to be at rest until the pulse reaches the junction, after which disturbances can travel into that fluid *only* in the direction  $x$  increasing (from the junction), giving

$$p_e = h(t - c_2^{-1}x), \quad J = Y_2 h(t - c_2^{-1}x). \quad (24)$$

The resulting equations of pressure continuity and volume-flow balance at the junction  $x = 0$  are

$$f(t) + g(t) = h(t), \quad Y_1[f(t) - g(t)] = Y_2 h(t). \quad (25)$$

They can be used at least during the whole time taken for the pulse (24) to travel to the far end of the tube  $x > 0$  and for any reflected pulse generated at that far end to return to the junction and alter the right-hand sides of the equations; for methods that take into account such additional reflections, see later sections.

Dividing the second of equations (25) by the first we obtain an equation

$$Y_1 \frac{f(t) - g(t)}{f(t) + g(t)} = Y_2, \quad (26)$$

from which the ratio  $g/f$  between the reflected and incident waveforms can be deduced as

$$\frac{g(t)}{f(t)} = \frac{Y_1 - Y_2}{Y_1 + Y_2}, \quad (27)$$

whence, by a second use of (25),

$$\frac{h(t)}{f(t)} = \frac{2Y_1}{Y_1 + Y_2}. \quad (28)$$

Note that the left-hand side of (26) is a sort of ‘effective admittance’ at  $x = 0$  of the tube  $x < 0$ : that is, a ratio of volume flow to excess pressure taking into account as in (23) the presence of the reflected wave. Equation (26) matches this

effective admittance in the tube  $x < 0$  to the admittance  $Y_2$  of the tube  $x > 0$  without any reflected wave.

When  $Y_2 < Y_1$ , equation (27) shows that this matching requires a *positive* reflection ( $g$  of the same sign as  $f$ ). An extreme case of this is when the tube  $x > 0$  is closed ( $Y_2 = 0$ ), requiring  $g = f$ : *total positive reflection*. On the other hand, when  $Y_2 > Y_1$  (perhaps because the tube  $x > 0$  has larger cross-section or is more distensible), equation (26) requires *negative* reflection with  $g/f < 0$ .

In any travelling wave the *energy flow* is related quite simply to admittance. The rate of transfer of energy in the direction of propagation is the fluid *velocity* times a force equal to cross-sectional *area* times *pressure excess*; thus, it is volume flow times pressure excess, which in turn is the admittance times the *square* of the pressure excess. For example, in the incident wave it is  $Y_1 f^2(t - c_1^{-1}x)$ . Similarly, in the reflected wave it is  $Y_1 g^2(t + c_1^{-1}x)$  but in the transmitted wave it is  $Y_2 h^2(t - c_2^{-1}x)$ .

Conservation of energy at the junction  $x = 0$  can now be easily checked from (27) and (28): the fraction of the incident energy that is reflected is

$$g^2(t)/f^2(t) = (Y_1 - Y_2)^2/(Y_1 + Y_2)^2 \quad (29)$$

and the fraction transmitted is

$$Y_2 h^2(t)/Y_1 f^2(t) = 4Y_1 Y_2/(Y_1 + Y_2)^2. \quad (30)$$

These fractions are nonnegative numbers that add up to 1: all the energy that is not reflected is transmitted.

Note that if the admittances are matched only approximately ( $Y_2/Y_1$  being close, but not equal, to 1) there is still almost perfect transmission of energy: the fraction reflected, (29), is in this case about  $\frac{1}{4}(Y_2/Y_1 - 1)^2$ . This is 0.0006, for example, with a 5% matching error, suggesting that a succession of junctions with relatively slight changes of admittance, each reflecting such an exceedingly small fraction of the pulse energy (proportional to the *square* of the matching error) may to a close approximation transmit a constant energy flow: an idea usefully generalised in § 5 to the case of a tube with very gradually and continuously changing admittance.

In another interesting case, when though the tube  $x > 0$  is not closed  $Y_2$  is very small compared with  $Y_1$ , equation (27) understandably makes  $g/f$  close to 1 (its value for the closed tube with  $Y_2 = 0$ ) but (28) gives the rather surprising result that  $h/f$  is almost 2. Admittedly, because  $Y_2$  is small, the *energy flow*  $Y_2 h^2$  in  $x > 0$  is not large; in fact its ratio (30) to the incident energy flow is only about  $4(Y_2/Y_1)$ . Nevertheless, the transmitted pressure amplitude is nearly *twice* that in the incident wave; essentially because the nearly closed end of the tube  $x < 0$  produces a nearly total positive reflection and thus a nearly doubled pressure amplitude, to which the tube  $x > 0$  duly responds.

Propagation through a junction characteristic of the cardiovascular system, with, say, two or three arteries extending *distally* from the junction and receiving the transmitted pulse when a pulse from a single *proximal* artery is incident upon the junction, is treated by an exceedingly simple extension of the above

analysis. Using  $x$  to signify displacement in the distal direction (away from the heart) from the junction  $x = 0$ , we attach subscript 1 as before to the values in the single proximal tube  $x < 0$  but use subscripts from 2 to  $N$  for values in the distal tubes  $x > 0$  (see Figure 1 illustrating typical cases when  $N$  is 3 or 4). Equations (23) still represent the general motion in the tube  $x < 0$  (including both an incident and a reflected wave) but equations (24) for the transmitted wave need to be replaced by a separate set

$$p_e = h(t - c_r^{-1}x), \quad J = Y_r h(t - c_r^{-1}x), \quad \text{for } x > 0, \quad (31)$$

in each distal tube  $r \geq 2$ . Continuity of pressure at the junction  $x = 0$  requires, on the other hand, that in each such tube the waveform  $h(t)$  is the same, and satisfies

$$f(t) + g(t) = h(t) \quad (32)$$

as in (25).

Equating volume flow into the junction (given by (23) with  $x = 0$ ) to the sum of the volume flows out of it (given by (31) with  $x = 0$ ) we obtain also

$$Y_1[f(t) - g(t)] = \left( \sum_{r=2}^N Y_r \right) h(t). \quad (33)$$

The important point to notice is that equations (32) and (33) are *exactly* the same as equations (25) with the admittance  $Y_2$  of the tube  $x > 0$  replaced by the *sum*  $\sum_{r=2}^N Y_r$  of the admittances of all the distal tubes  $x > 0$ . As in the theory of electrical transmission lines the *admittances of tubes in parallel merely add up*, and the junction behaves just like a simple junction with one distal tube  $x > 0$  possessing an admittance equal to  $\sum_{r=2}^N Y_r$ .

For example, equations (27) and (28) for the reflected and transmitted waveforms as fractions of the incident waveform become

$$\frac{g(t)}{f(t)} = \frac{Y_1 - \sum_{r=2}^N Y_r}{Y_1 + \sum_{r=2}^N Y_r}, \quad \frac{h(t)}{f(t)} = \frac{2Y_1}{Y_1 + \sum_{r=2}^N Y_r}. \quad (34)$$

It can also be shown, since the energy carried away from  $x = 0$  by all the transmitted waves (31) adds up to  $(\sum_{r=2}^N Y_r)h^2(t)$ , that equations (29) and (30) similarly specify the fractions of the incident energy that are reflected and transmitted (provided that  $Y_2$  is replaced by  $\sum_{r=2}^N Y_r$ ) and again check conservation of energy.

A prominent branch in the cardiovascular system is the *iliac bifurcation* where the aorta divides, after passing down the abdomen, into the two iliac arteries. We briefly consider how a pulse wave from the heart would be affected by the presence of just this one bifurcation. Though such consideration ignores the real

complexity of the cardiovascular system (involving large numbers of interacting branches and some gradations of arterial properties between them) its results show some similarity to observed behaviour mainly because this is such a prominent branch.

In this case with  $N = 3$ , the sum of the admittances  $Y_2$  and  $Y_3$  of the two iliac arteries turns out to be less than the admittance  $Y_1$  of the aorta itself, partly because the sum of their cross-sectional areas is rather less (by around 20%, though the ratio takes widely varying values in different individuals) and partly because the aorta has greater distensibility and therefore a lower value of the wave speed  $c$  than do the iliac and other more peripheral arteries. Equations (34) imply positive reflection, therefore ( $g/f > 0$ ).

Observations of pressure fluctuations in the aorta are consistent with this qualitative conclusion of a positive reflected wave. The physiological significance of the resulting enhanced pressure changes in a relatively distensible aorta is that they facilitate storage of the blood volume expelled by the heart in each stroke for delivery at a relatively steady rate through the peripheral circulation.

Higher up, in the thoracic aorta, there are a number of junctions of an extremely different type, at each of which two narrow *intercostal* arteries draw away from the aorta a small quantity of blood to feed the region between two pairs of ribs. In this case (Figure 1 with  $N = 4$ ), the admittances  $Y_1$  and  $Y_2$  of the parts of the aorta proximal and distal to the branch are essentially equal and very large compared with the admittances  $Y_3$  and  $Y_4$  of the intercostal arteries. The sum  $\sum_{r=2}^N Y_r$  is therefore close to  $Y_1$  and our equations imply almost perfect transmission. This result, in no way surprising for the continued transmission down the aorta, indicates more interestingly the same fluctuations of pressure propagating along the intercostal arteries as in the aorta itself.

#### 4. Interaction between junctions in a branching system

It might be thought practically impossible to analyse pulse propagation in a complicated branching system representative of the human arterial system, because of the huge number of *interactions* between different junctions, all ignored in the theory of § 3. We find indeed that the interaction between just *two* junctions (studied in the present section) raises major new complexities, particularly in the sensitivity of the conclusions to frequency. Fortunately, however, it turns out that once this rather interesting interaction between a pair of junctions has been properly analysed, it is a fairly straightforward matter to extend the analysis to a quite intricate branching system (see especially § 6).

For studying a single interaction, we consider (Figure 2) a junction B *exactly* like one of those in Figure 1, except that a length  $l$  of the tube  $x < 0$  is itself one of the tubes receiving waves transmitted through a *proximal* junction A, where therefore our  $x$ -coordinate takes the value  $x = -l$ . The ratio  $g(t)/f(t)$  of reflected to incident wave at the junction B (where  $x = 0$ ) is again supposed given by (34).

We calculate the *effective admittance*  $Y_1^{\text{eff}}$  of the tube  $x < 0$ , still designated by the subscript 1, at the junction A (where  $x = -l$ ). This is defined as the ratio of volume flow  $J$  to pressure excess  $p_c$ , taking into account as in equations (23) the

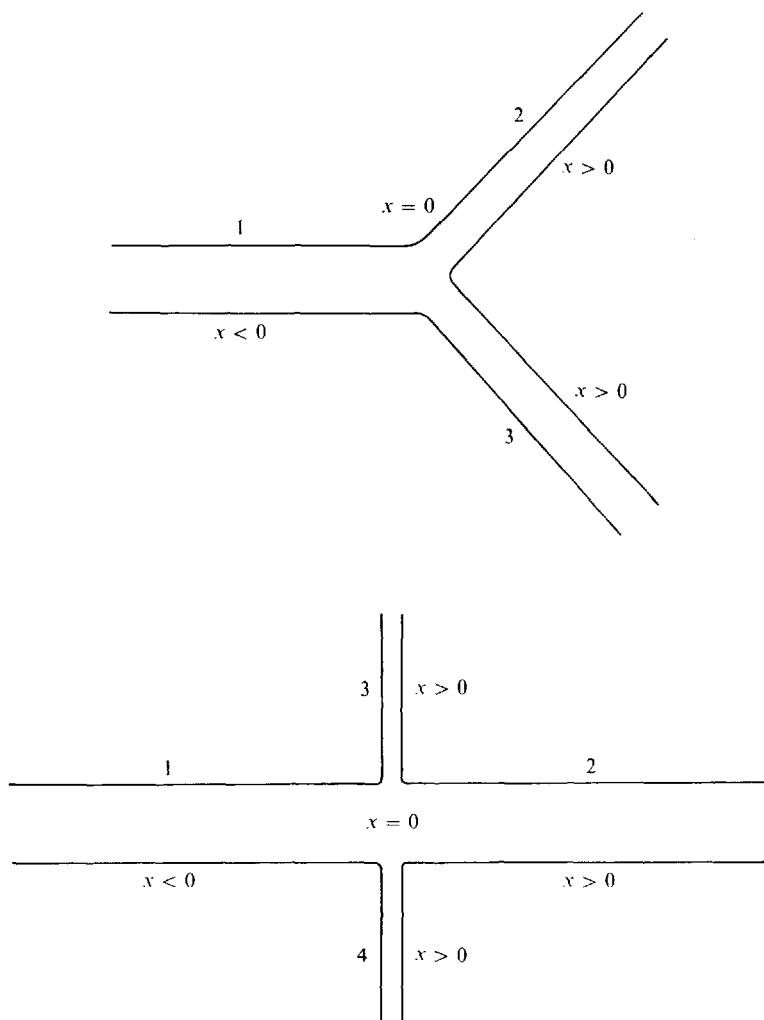


FIG. 1. *Diagrammatic illustrations of junctions in the cardiovascular system.*

presence of waves in *both* directions in the tube  $x < 0$  to give

$$Y_1^{\text{eff}} = Y_1 \frac{f(t + c_1^{-1}l) - g(t - c_1^{-1}l)}{f(t + c_1^{-1}l) + g(t - c_1^{-1}l)}. \quad (35)$$

If we can calculate once for all this ratio  $Y_1^{\text{eff}}$  of volume flow to pressure excess at junction A for the tube AB as influenced by the presence of the second junction B, then the analysis of junction A can proceed in exact accordance with the rules of § 3 for a single junction, with the admittance  $Y_1$  of that tube at A replaced by  $Y_1^{\text{eff}}$  in the required sum of admittances of all the tubes that receive pulse waves transmitted through A. Repeated use of this algorithm for a branching system, working inwards from the periphery, can complete its analysis.

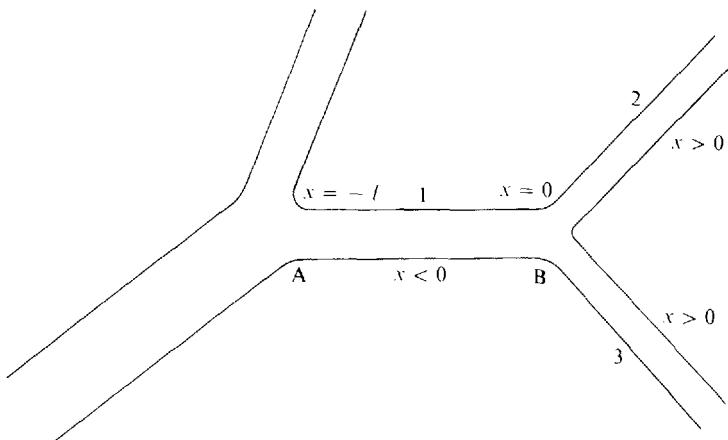


FIG. 2. Illustrating an interaction between two junctions A and B.

The qualitatively new feature of this problem is the time lag  $c_1^{-1}l$  between waveforms at A and at B which in (35) alters the f wave (travelling from A towards B) and the g wave (travelling from B towards A) in opposite senses. This is important because it prevents straightforward use of equation (34) for  $g(t)/f(t)$  to calculate (35).

The way round this difficulty, universally adopted in studies of propagation through branching systems, is Fourier analysis into complex-exponential waveforms proportional to  $e^{i\omega t}$ . On a linear theory, any 'forcing effect' that excites waves can be Fourier-analysed as a linear combination of  $e^{i\omega t}$  terms for different values of  $\omega$  and, if any feature of the system's response can be calculated separately for each such term, then the complete response must be the same linear combination of those separate responses. For example, a *periodic* but not sinusoidal forcing effect, like the variation of excess pressure at the aortic valve, can be analysed as a *Fourier series* of terms proportional to  $e^{i\omega t}$ , with  $\omega$  taking values that are integer multiples of  $2\pi/T$  where  $T$  is the period.

The fact that on linear theory a term proportional to  $e^{i\omega t}$  in the forcing effect elicits responses proportional to  $e^{i\omega t}$  everywhere in the system means that corresponding values of  $f(t)$  and  $g(t)$  in equations (34) and (35) are themselves proportional to  $e^{i\omega t}$ , which gives immediately

$$Y_1^{\text{eff}} = Y_1 \frac{\left( Y_1 + \sum_{r=2}^N Y_r \right) e^{i\omega c_1^{-1}l} - \left( Y_1 - \sum_{r=2}^N Y_r \right) e^{-i\omega c_1^{-1}l}}{\left( Y_1 + \sum_{r=2}^N Y_r \right) e^{i\omega c_1^{-1}l} + \left( Y_1 - \sum_{r=2}^N Y_r \right) e^{-i\omega c_1^{-1}l}}. \quad (36)$$

Reflection and transmission at junction A of waveforms proportional to  $e^{i\omega t}$  can accordingly be calculated taking into account the presence of junction B provided that the effective admittance of the tube AB is taken as this *complex number* (36) dependent on  $\omega$ . Such use of complex admittances dependent on

frequency exactly parallels a standard method of analysis of electrical transmission lines.

Equation (36) for the effective complex admittance of the tube AB at junction A in the presence of junction B is more useful when thrown into the simplified form

$$Y_1^{\text{eff}} = Y_1 \frac{\left( \sum_{r=2}^N Y_r \right) + i Y_1 \tan(\omega c_1^{-1} l)}{Y_1 + i \left( \sum_{r=2}^N Y_r \right) \tan(\omega c_1^{-1} l)}. \quad (37)$$

Although in § 5 a modification to (36) and (37) allowing for *viscous attenuation* of the pulse (essentially, by adding an imaginary part to  $c_1^{-1}$ ) is noted, this simple formula (37) indicates many important properties of the interaction between junctions when  $l$  is not too large.

In one extreme case when  $l$  is very small compared with  $c_1/\omega$ , equation (37) degenerates to

$$Y_1^{\text{eff}} = \sum_{r=2}^N Y_r, \quad (38)$$

exactly as if the tube AB were not there!—or, more strictly, as if it were just *part of the junction B*. This agrees with the earlier conclusion that the simple continuity conditions at a junction are satisfied if its dimension satisfies the condition (20).

Interestingly enough, the same ‘transparency’ of AB to propagation at frequency  $\omega$  is demonstrated by (37) also when  $\omega c_1^{-1} l$  is nearly equal to  $\pi$ : a tube AB whose length is about half a wavelength ‘hands on’ the wave to the branch B with a phase change of  $\pi$  but otherwise as if it had not been there. Values of  $l$  that are *multiples* of half a wavelength would indeed share this ‘transparency’ property, although modified by attenuation effects to a considerably greater extent.

Intermediate between the ‘transparency’ values of  $\omega c_1^{-1} l$  near 0 and near  $\pi$  is a range of values near  $\frac{1}{2}\pi$ , where the tube AB has length about a *quarter wavelength* and its admittance characteristics are *accentuated*, since

$$Y_1^{\text{eff}} \doteq Y_1^2 / \left( \sum_{r=2}^N Y_r \right). \quad (39)$$

Thus, if  $Y_1$  is *greater* than  $\sum_{r=2}^N Y_r$ , then  $Y_1^{\text{eff}}$  is greater still, while if  $Y_1$  is *less* than  $\sum_{r=2}^N Y_r$ , then  $Y_1^{\text{eff}}$  is less still.

In relation to a typical fundamental frequency of the heart beat, the aorta has length near a quarter wavelength in normal subjects. For such a frequency, accordingly, the effective admittance of the aorta at the aortic valve is *enhanced*: a matter of physiological importance since it allows the first Fourier component in the fluctuations of volume flow through the valve to be generated with only moderate fluctuations of excess pressure produced by the heart. We may note also that, for some higher Fourier components, wavelengths can be low enough for the *iliac* arteries to be around a quarter wavelength, accentuating the effect of their reduced admittance in generating positive reflection at the iliac bifurcation.

For more comprehensive information taking into account multiple interactions between junctions, and certain additional features, see § 6.

### 5. Amplitude gradations along a single tube

The introduction to branching systems in § 4 treats only systems where each separate tube is uniform and each travelling wave in it, whether directed towards  $x$  increasing or decreasing, retains a constant amplitude all along it. In this section, while remaining within the framework of a linear theory, we consider how amplitude *gradations* in a wave travelling along a tube may be produced, either by gradations in its cross-sectional area and distensibility or by energy-dissipation effects such as viscosity.

Very gradual changes in distensibility and cross-sectional area, in the absence of energy dissipation, can tentatively be investigated by a method foreshadowed in § 3 as a suggestion from equation (29). We approximate the tube of gradually changing properties by a sequence of uniform tubes, separated by junctions involving small changes. The travelling wave is then regarded as propagating at a wave speed  $c$  which varies from tube to tube according to equation (6), while the wave amplitude similarly varies so that the energy flow  $Yp_e^2$  (where again the admittance  $Y$  of the different tubes is different) remains constant as the wave travels. In the absence of energy *dissipation* this could be expected from the idea that  $n$  small changes of admittance each of order  $n^{-1}$  reflect at most a proportion of energy of order  $n^{-2}$  each, or  $n(n^{-2}) = n^{-1}$  altogether, tending to zero as  $n \rightarrow \infty$ .

In this limit, the excess pressure  $p_e$  and volume flow  $J$  for a wave travelling in the direction  $x$  increasing could on the above arguments be written, following Taylor (1965), as

$$p_e = [Y(x)/Y(0)]^{-1/2} f \left( t - \int_0^x c^{-1} dx \right), \quad J = Y(x)p_e, \quad (40)$$

where  $f(t)$  is the waveform at  $x = 0$  and  $\int_0^x c^{-1} dx$  is the time taken for the wave to travel from 0 to  $x$ . The corresponding equation for a wave travelling in the direction  $x$  decreasing, generalising (9) as (40) generalises (8), is

$$p_e = [Y(x)/Y(0)]^{-1/2} g \left( t + \int_0^x c^{-1} dx \right), \quad J = -Y(x)p_e, \quad (41)$$

where  $g(t)$  is its waveform at  $x = 0$ . Either equation implies that, as a wave propagates into (for example) a region of lower admittance, its pressure excess increases although the associated volume flow decreases and their product the energy flow remains constant.

It is natural to enquire just how gradual the changes in  $c$  and  $Y$  with  $x$  must be for equations (40) and (41) to be good approximate solutions to the appropriate linearised equations of motion. The answer to this question will be sketched only very briefly. It turns out that those linearised equations (generalisations of (3) and (4)) can be written

$$\partial J/\partial t = -cY \partial p_e/\partial x, \quad \partial p_e/\partial t = -cY^{-1} \partial J/\partial x. \quad (42)$$

Then, estimating the accuracy of, for example, equation (40) for  $p_e$  by substituting it into the second-order equation derived from (42) by eliminating  $J$ , we find that for good accuracy  $c$  and  $Y$  and also  $dY/dx$  must each vary by a small fraction of itself in a distance  $c/\omega$  (the same distance that arose in the quite different condition (20) above): then the square of that small fraction is a measure of the fractional discrepancy between the two terms in the equation for  $p_e$ .

In the cardiovascular system this means that, between junctions, only *very* gradual changes in tube properties (each changing with a gradient corresponding to only a small fraction of itself per *metre*) can be treated on the basis of (40) and (41) with energy reflection neglected and  $p_e$ ,  $J$  taken to vary as  $Y^{-1/2}$  and  $Y^{1/2}$  respectively. Allowance for such a gradual change is nevertheless worth making.

Note that the opposite extreme, describable perhaps in terms of *frequencies so low* that  $c/\omega$  greatly exceeds other relevant lengths, might allow the right-hand sides in (42) to be equated to *zero*. This would evidently reduce to the situation described within a general junction satisfying (20), where reflection may occur, with both  $p_e$  and  $J$  remaining approximately constant. Between our 'extremes', however, the situation in, for example, the aorta with its rather gradual reduction of admittance along its considerable length is thought to be nearer the former.

In any *branching* system equation (37), for the effective admittance at A (where  $x = -l$ ) of a tube AB taking into account reflections from a junction B (where  $x = 0$ ), can be readily extended to cases when the properties of the tube AB vary along it sufficiently gradually. This is especially because the same factor  $[Y(x)/Y(0)]^{-1/2}$  appears in the direct wave (40) and in the reflected wave (41), cancelling therefore in an expression like (35). In fact, the ratio  $Y_1^{\text{eff}}$  of the volume flow at A to the pressure excess, for a pulse specified as the sum of (40) and (41), becomes

$$Y_1^{\text{eff}} = Y_1^A \frac{f(t + c_{AB}^{-1}l) - g(t - c_{AB}^{-1}l)}{f(t + c_{AB}^{-1}l) + g(t - c_{AB}^{-1}l)}, \quad (43)$$

where  $Y_1^A$  is the admittance  $Y(-l)$  of the tube AB itself at A and  $c_{AB}^{-1}$  is the average reciprocal wave speed along AB, defined by

$$c_{AB}^{-1}l = \int_{-l}^0 c^{-1} dx. \quad (44)$$

On the other hand, the first of equations (34), determining the ratio of reflected waveform  $g(t)$  to incident waveform  $f(t)$  at  $x = 0$  (the junction B), applies with  $Y_1$  taken as the admittance  $Y_1^B = Y(0)$  of the tube AB at B. Accordingly, the arguments leading to equation (37) now give the slightly modified equation

$$Y_1^{\text{eff}} = Y_1^A \frac{\left( \sum_{r=2}^N Y_r \right) + i Y_1^B \tan(\omega c_{AB}^{-1}l)}{Y_1^B + i \left( \sum_{r=2}^N Y_r \right) \tan(\omega c_{AB}^{-1}l)}, \quad (45)$$

which can be used for calculating properties of branching systems of tubes with very gradually varying properties just as readily as (37) can for systems of uniform tubes. For example, there is a quarter-wavelength resonance condition, now defined by  $\omega c_{AB}^{-1}l = \frac{1}{2}\pi$ , and the resonant value, replacing (39), of the effective admittance at A is  $Y_A Y_B / (\sum_{r=2}^N Y_r)$ .

Other laws of gradation of pulse amplitude arise from the action of energy *dissipation*, due either to blood viscosity or wall visco-elasticity. Here we outline an appropriate treatment of the effect of blood viscosity and indicate the similar treatment required for wall visco-elasticity.

Blood flow in arteries under normal physiological conditions is subject to rates of shear of the order  $10^2 \text{ s}^{-1}$  or more. Although viscometer experiments on blood indicate quite anomalous flow properties at much *lower* rates of shear (that permit a significant degree of *aggregation* of red cells), experiments at shear rates around  $10^2 \text{ s}^{-1}$  or more can be interpreted to reasonable approximation by treating blood as a Newtonian fluid with viscosity  $\mu$  around  $4 \times 10^{-3} \text{ Nsm}^{-2}$  (about 2.5 times the viscosity of plasma, or 4 times the viscosity of water). The associated 'kinematic viscosity' or *diffusivity of momentum* is

$$\nu = \mu/\rho = 4 \times 10^{-6} \text{ m}^2 \text{ s}^{-1}. \quad (46)$$

For the blood pulse in an artery, the importance of viscosity is of course that it prevents the blood velocity from having a distribution across the lumen which is uniform and slips smoothly over the walls. Diffusion of momentum with diffusivity  $\nu$  causes exchange of momentum between blood at rest in immediate contact with the wall and uniformly moving blood farther out, the exchange region where the blood is moving at intermediate speeds being called the *boundary layer*.

This is yet another feature of pulse-wave propagation which, on a linear theory, is best analysed separately for each wave component proportional to  $e^{i\omega t}$ . The boundary layer on a wall associated with one such component has a thickness proportional to  $(\nu/\omega)^{1/2}$ , related to the distance of diffusion of momentum with diffusivity  $\nu$  in, say, a half-period  $\pi/\omega$ . The velocity distribution determined by Stokes (see Chapter 7 of Rosenhead 1964) is found in this boundary layer provided that the radius  $a$  of the tube (or for noncircular sections the wall radius of curvature) is large compared with the boundary-layer thickness: in other words, if the *Womersley parameter*

$$\alpha = a(\omega/\nu)^{1/2} \quad (47)$$

is large (in practice, greater than around 4).

In the aorta of larger mammals this condition is satisfied: even for the fundamental frequency of the heart beat,  $\alpha$  is around 20 in an ox, 15 in a man, 10 in a dog, 4 in a rabbit (McDonald 1960, Chapter 5), while  $\alpha$  is evidently larger still for any higher harmonic. In their smaller arteries, however,  $\alpha$  takes smaller values (typically, 3 for the femoral artery in man) and viscosity plays a bigger part, with boundary layers practically filling the whole lumen.

We give actual results for the case of large  $\alpha$  (though we indicate the trend of modifications for small  $\alpha$ ). The Stokes boundary layer driven by an oscillating gradient  $\partial p_e / \partial x$  of excess pressure proportional to  $e^{i\omega t}$  has velocity distribution

$$u = (\rho i\omega)^{-1} (-\partial p_e / \partial x) \{1 - \exp[-z(i\omega/v)^{1/2}]\}, \quad (48)$$

where  $z$  is the distance from the wall. Expression (48), as  $z$  becomes large, tends to the uniform velocity

$$u = (\rho i\omega)^{-1} (-\partial p_e / \partial x), \quad (49)$$

representing a purely *reactive* response (that is, by pure inertia, with a  $90^\circ$  phase lag) to the force per unit volume  $(-\partial p_e / \partial x)$ . For smaller values of  $z$ , within the viscous boundary layer, however, expression (48) includes not only such a pure imaginary multiple of  $(-\partial p_e / \partial x)$  but also a real positive multiple of it, representing a *resistive* response (without phase lag) to that force per unit volume.

The volume flow  $J$  associated with a *uniform* velocity distribution (49) over a cross-sectional area  $A$  satisfies

$$\partial J / \partial t = A i\omega u = -A \rho^{-1} \partial p_e / \partial x, \quad (50)$$

an equation that we have already written down as the first of the two equations (42). Viscous effects for large  $\alpha$ , however, being found in a thin Stokes boundary layer specified by equation (48), change this result to

$$\partial J / \partial t = -\{A - s(v/i\omega)^{1/2}\} \rho^{-1} \partial p_e / \partial x, \quad (51)$$

where  $s$  is the perimeter of the cross-section. Here, the second term in curly brackets represents the ‘volume-flow defect’ in the boundary layer associated both with an increase in effective inertia and with a resistive effect. For a circular cross-section of radius  $a$ , the ratio  $A/s$  is  $\frac{1}{2}a$  so that the right-hand side of (51) is that of (50) modified by a factor

$$\{1 - 2i^{-1/2}\alpha^{-1}\}. \quad (52)$$

Although viscous effects produce such a factor modifying the right-hand side of the *first* of equations (42), the *second* is unmodified by boundary-layer considerations. It states simply that the rate of change of fluid volume per unit length, namely  $\partial A / \partial t$  which is  $AD\partial p_e / \partial t$ , equals minus the gradient along the tube of the volume flow of fluid  $J$ ; where on the definitions (6) and (22) of wave velocity  $c$  and admittance  $Y$  the product  $AD$  is indeed  $c^{-1}Y$ .

We earlier saw that even in tubes with  $c$  and  $Y$  varying gradually enough, equations (42) possess to good approximation the propagating solutions (40) and (41), adding up to their general solution. We now observe that if the right-hand side of the first equation is modified by the factor (52) while the second equation is unmodified, the resulting equations are *the same as if both c and Y were multiplied by the square root of (52)*. The solutions (40) and (41) are therefore still approximately correct if in each of *them* both  $c$  and  $Y$  are multiplied by the square root of (52).

Actually a small change in  $Y$  makes little change in those solutions, but a small change in  $c^{-1}$ , which appears in the argument of the waveform function  $f(t)$  or  $g(t)$  (here proportional to  $e^{i\omega t}$ ) can be very significant. Using the linear binomial approximation to the inverse square root of (52) we can say that for large  $\alpha$  we should simply replace  $c^{-1}$  by

$$c^{-1}(1 + i^{-1/2}\alpha^{-1}) = c^{-1} \left(1 + \frac{1 - i}{\alpha\sqrt{2}}\right) \quad (53)$$

in these solutions.

For example, a factor

$$\exp \left[ i\omega \left( t - \int_0^x c^{-1} dx \right) \right] \quad (54)$$

in a wave (40) travelling in the direction  $x$  increasing becomes

$$\exp \left[ i\omega t - i\omega \int_0^x \left( 1 + \frac{1}{\alpha\sqrt{2}} \right) \frac{dx}{c} - \omega \int_0^x \frac{dx}{c\alpha\sqrt{2}} \right] \quad (55)$$

after this substitution. In (55) the second term implies a *reduced phase speed* associated with the increased effective inertia, while the third term implies an *amplitude decreasing* progressively by resistive action. The relatively modest phase-speed reduction is greatest for the fundamental and less for the higher harmonics with greater values of  $\alpha$ . On the other hand the rate  $\omega/c\alpha\sqrt{2}$  of exponential attenuation of amplitude with distance is *lowest* for the fundamental (by equation (47)) and *increases* like  $\omega^{1/2}$  for the higher harmonics of the blood pulse.

All the results obtained earlier in this section with viscosity neglected can be improved so as to apply in the presence of thin boundary layers if  $c^{-1}$  is replaced (in equation (44), for example) by expression (53). The fact that  $c_{AB}^{-1}$  is then a complex number reduces, needless to say, the sharpness of any quarter-wavelength resonance peak.

Similar results were obtained by Womersley (1957) for *general* values of his parameter  $\alpha$  by an analysis that starts from a replacement of equation (48) by an expression involving Bessel functions. In the limit for *small*  $\alpha$ , the instantaneous distribution of velocity across the tube becomes a Poiseuille parabolic distribution, and (51) is replaced by

$$\frac{4}{3} \hat{c}J/\hat{c}t + (8v/a^2)J = -A\rho^{-1} \hat{c}p_e/\hat{c}x, \quad (56)$$

where the  $\frac{4}{3}$  is an inertia factor by which the kinetic energy of a Poiseuille distribution is enhanced above that of a uniform distribution of velocity with the same volume flow  $J$ , and  $(8v/a^2)$  is the ordinary Poiseuille resistance. Adoption of equation (56) requires  $c^{-1}$  to be replaced by

$$c^{-1}(\frac{4}{3} - 8i\alpha^{-2})^{1/2} \quad (57)$$

and  $Y^{-1}$  to be changed by the same factor. As a matter of fact, the use of (53) for

$\alpha > 4$  and (57) for  $\alpha < 4$  gives results in quite good agreement with Womersley's general formulae.

We may conclude this section with a note concerning visco-elastic properties, if any, of the arterial wall. We do not go into any more biosolidmechanics here than was given in § 2 but remark that visco-elasticity effectively makes the distensibility  $D$  a complex number dependent on frequency; some measurements by Bergel (1961) in arteries suggest typical *phase lags* of  $5^\circ$  to  $10^\circ$  in distensibility. This, through equation (6), generates a new factor modifying  $c^{-1}$ : a factor with a negative imaginary part as in (53), leading to wave attenuation again as in (55). Effects of blood viscosity, of course, multiply this *modified*  $c^{-1}$  by the further modifying factor given in (53).

## 6. Some comparisons with experiment

In this section a selection of experimental observations is discussed in relation to the above theories of pulse propagation and their extensions.

Professor M. G. Taylor has been particularly active in theoretical and experimental investigations of the admittance of the aorta at the aortic valve. He pointed out (Taylor 1966) the feasibility of calculating the input admittance for a large complex branching system representative of the arterial tree by repeated use of an algorithm like (37) or (45), that calculates proximal from distal effective admittances, to work gradually back from peripheral values towards a value for the input admittance itself.

The nature of the dependence of this input admittance on the frequency  $\omega$  is of considerable interest, partly because the fundamental frequency of the heartbeat is indeed variable, but especially because in the complicated pulse waveform many harmonics of that fundamental frequency have significant amplitudes. It was noted at the end of § 4 that a typical value of the fundamental frequency may be in the region of quarter-wavelength resonance for the aorta itself, leading to increased effective input admittance, while for one of the higher harmonics a different resonance lowering the effective admittance of (say) the iliac arteries might similarly produce an increase in the effective input admittance of the aorta.

Taylor noted, however, that the effect of the great variety of different tube lengths between junctions in the arterial tree should be to smooth out individual resonance peaks, at any rate after the first, and allow a consistently high level of input admittance to be maintained at the higher frequencies. He demonstrated this (Taylor 1966) with a variety of models of branching systems, including for example that illustrated in Figure 3, including seven generations of branchings with decreasing area and gradually increasing wave speed, terminating in peripheral resistances characteristic of the arteriolar circulation.

Figure 4 gives Taylor's results for that branching system. It plots the input impedance (the reciprocal  $Z = Y^{-1}$  of the admittance): a quantity of particular physiological significance as specifying the pressure excess demanded of the heart to generate different Fourier components of volume flow. Note that the very high impedance at zero frequency, representing a pure resistance of the system to the *mean* volume flow through it, is followed by a very *low* minimum of the

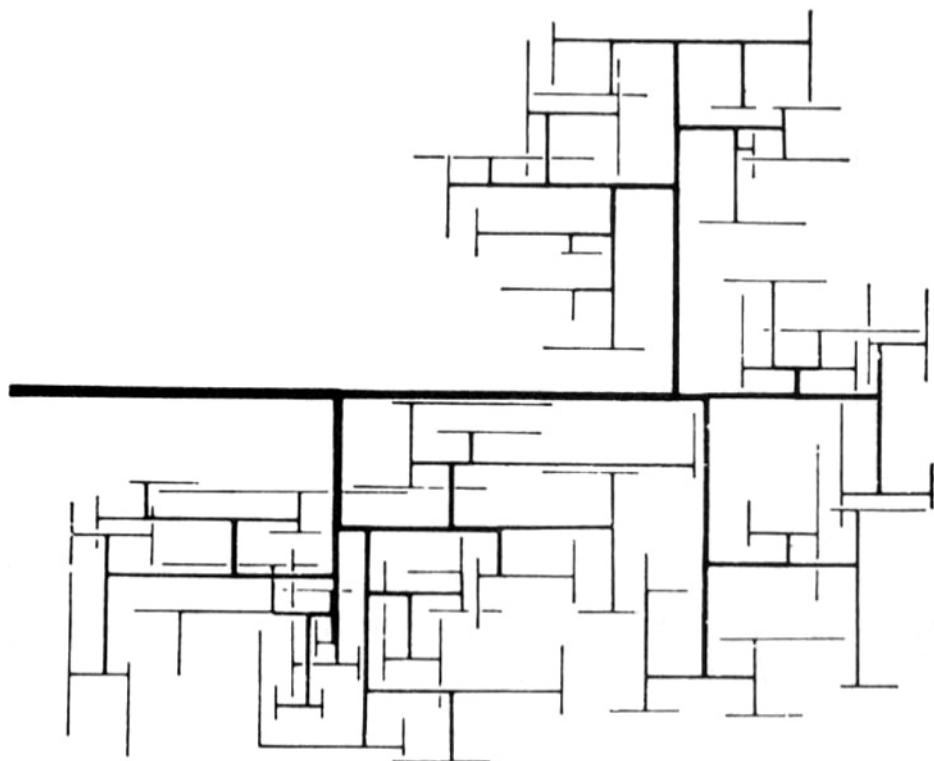


FIG. 3. This Figure 1 from Taylor (1966) gives a diagrammatic representation of the branching system which he principally analysed.

impedance (resonance peak of the admittance) at a typical fundamental frequency of the heart beat. Note also that at all frequencies either near or above that one (representing possible heart-beat frequencies and their harmonics) the impedance remains strikingly low compared with its values at zero frequency.

Such maintenance of a generally low level of impedance at all frequencies except extremely low ones is an important property of this model. Taylor (1964, 1965) summarized the evidence that the real arterial systems of mammals exhibit this feature, and benefit from it to such an extent that the external work performed by the heart in generating pulsatile volume flow exceeds by only 5% to 15% the work that would be needed to generate the mean volume flow at a steady rate. On the other hand, simplified models, lacking the wide variety of tube lengths shown in Figure 3, exhibit some much higher resonant peaks of impedance at frequencies in the neighbourhood of typical heart-beat harmonics.

Not only the modulus (or absolute value) of the impedance is shown in Figure 4, but also its phase angle. This is minus the phase angle of the admittance  $Y$ , which is thus seen to have a positive imaginary part increasing with frequency  $\omega$  (just as

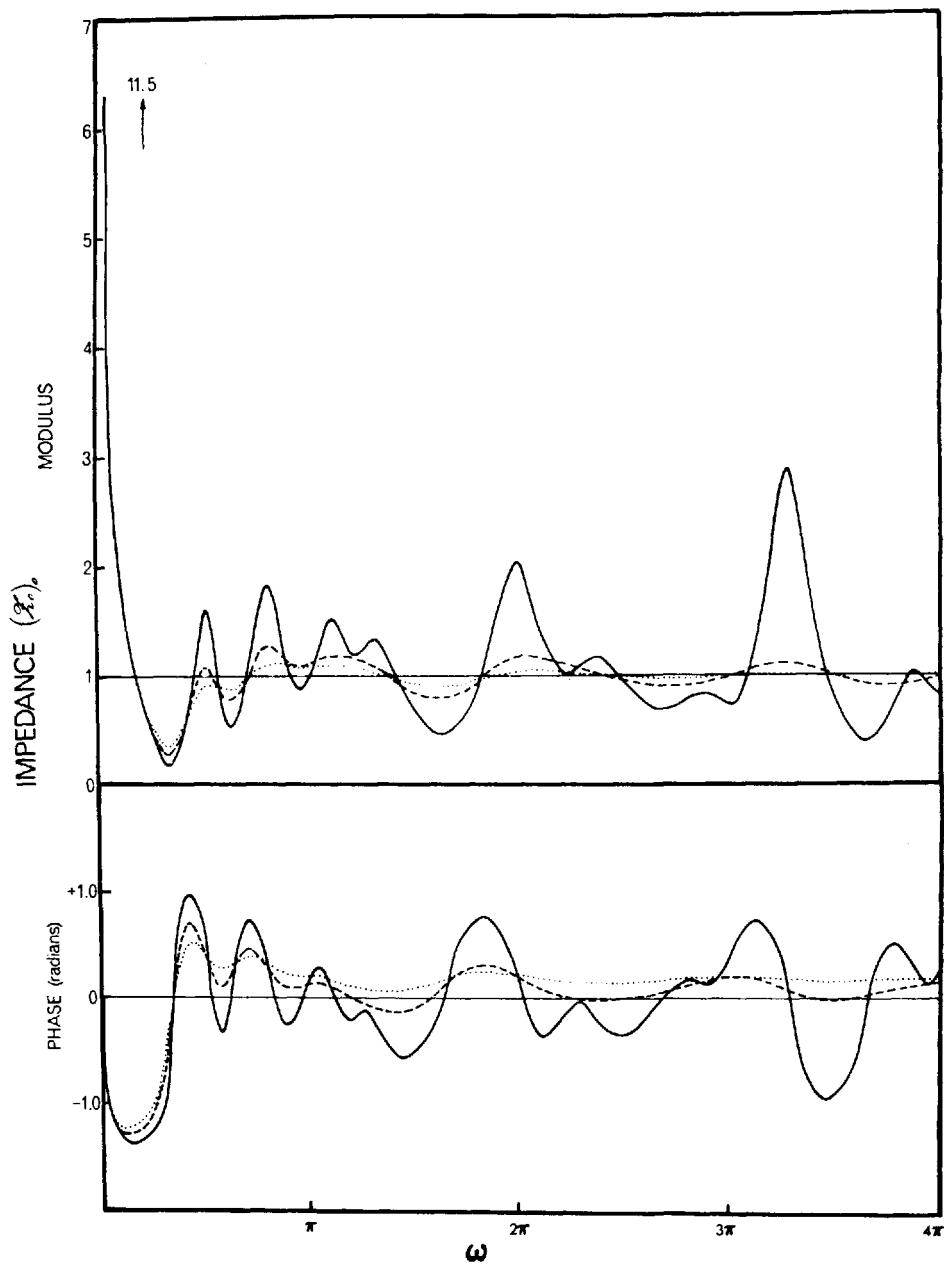


FIG. 4. This Figure 9 from Taylor (1966) gives his calculations of input impedance  $Y^{-1}$ .

for a capacitance  $C$ , with admittance  $Y = Ci\omega$  at very low frequencies. This represents the capacitance of the aorta: its storage property, sometimes referred to by the German word 'Windkessel' or 'bellows chamber'. Equation (37) already, with  $Y_1$  as an aortic admittance considerably greater than the combined distal

admittance  $\sum_{n=2}^N Y_n$ , suggested at low frequency just such a positive imaginary part for  $Y_1^{eff}$ , increasing in proportion to  $\omega$  for very small  $\omega$  and then falling back to zero at the quarter-wavelength resonant frequency. Figure 4 indicates, however, much less pronounced excursions of phase angle at higher frequencies.

The plain lines in Figure 4 represent calculations taking blood viscosity fully into account but no visco-elasticity of the arterial wall. The broken lines indicate something of an upper limit on the likely effect of visco-elasticity, as they assume a 10° phase lag in the distensibility of the wall. This, as might be expected from § 5, introduces modest changes at the lower frequencies but increasing changes at the higher frequencies, tending to smooth out still further the rises and falls in the impedance after its first strong minimum.

It is of interest to observe the associated changes in pulse waveform as it passes through the arterial tree. These are depicted for the dog in Figure 5, taken from McDonald (1960); Remington & Wood (1956) obtained similar results in human arteries. The most striking feature of such data is the peripheral increase in amplitude of the pressure variation.

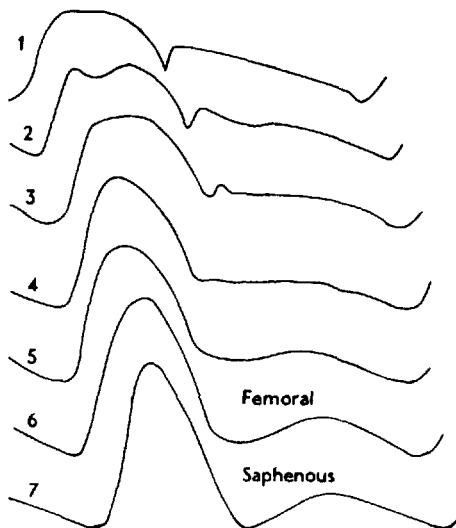


FIG. 5. This Figure 12.4 from McDonald (1960) indicates the changes in the pressure waveform as the pulse propagates towards the periphery.

For example, the root mean square pressure variation increases by over 50% from the proximal ascending aorta (curve 1) to the distal abdominal aorta (curve 4), and some further increase occurs in daughter tubes as far as the saphenous artery. This important observation is considered (Taylor 1965) to be mainly related to the peripheral decrease in admittance. A gradual decrease (associated with gradual reduction in both distensibility and cross-sectional area) was shown in § 5 to produce gradations in pressure amplitude in this sense, with correspondingly

decreasing flow amplitudes. Figure 6, due also to McDonald (1960), shows the measured flow velocities at the various sites falling as would be expected. Again, a *discontinuous* decrease in admittance at a junction produces such an increased pressure amplitude in the daughter tubes according to the second of equations (34).

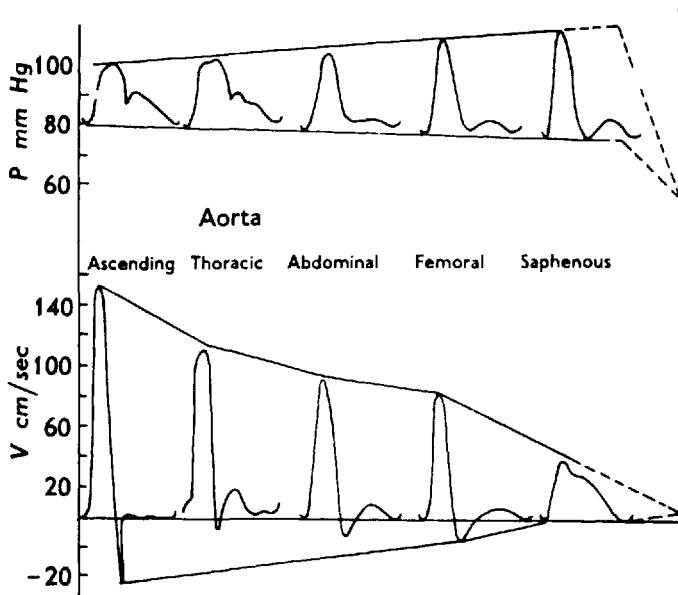


FIG. 6. This Figure 12.6 from McDonald (1960) contrasts the increase in pressure amplitudes and decrease in velocity amplitudes as the pulse propagates peripherally.

Another obvious feature of the waveform changes in Figure 5 is that small-scale irregularities of the proximal waveform are smoothed out peripherally, implying that higher harmonics in the waveform are quite rapidly damped out. This is just the effect that would be expected, according to § 5, from the action of blood viscosity probably coupled with wall visco-elasticity.

Yet another feature of the waveform changes in Figure 5 is the peripheral appearance of a *diastolic peak* pressure: that is, of a second smooth subsidiary pressure maximum about half a period after the main (systolic) peak. In terms of Fourier analysis this can be expressed by saying that the fundamental acquires a phase lag relative to its first harmonic. Linear theory gives more than one mechanism by which this could come about: for example, the fundamental has a lower value of the Womersley parameter  $\alpha$  than do its harmonics and therefore exhibits more of the reduction in phase speed due to viscous action emphasized in equation (55); again, transmitted waveforms depend on effective admittances such as (37) which are themselves complex numbers depending on frequency.

This means that we are not forced to look towards nonlinear effects to explain this waveform change. Nevertheless, such effects would also produce a change in the same direction. They predict that high-pressure parts of the waveform

propagate slightly faster (especially due to convection by the forward-moving fluid velocity in those parts) than do the low-pressure parts (with fluid velocity in the opposite direction). This transfers energy from the fundamental to its first harmonic with the right phase to explain the above effect.

Ling & Atabek (1972) gave a good review of nonlinear theories of pulse propagation, taking into account the distribution of velocity across an artery generated by blood viscosity. They compared the results with measurements of velocity profiles made in a reasonably realistic model of the canine aorta. This rather satisfactory comparison is shown in Figure 7.

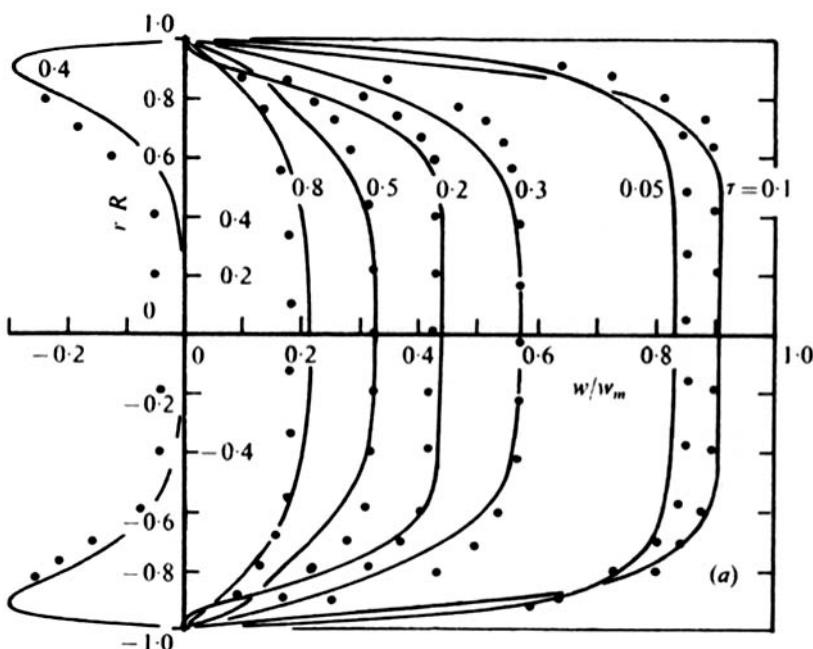


FIG. 7. Distributions of axial velocity  $w$  across a model of a canine aorta measured in experiments by Ling & Atabek (1972) with pulse period 0.88 when  $w$  takes a maximum value  $w_m = 0.77 \text{ ms}^{-1}$ . The curves depict their theoretical results.

In discussing this good agreement, however, they possibly lay too much emphasis on the nonlinear aspect of their calculations. They compare the results with a very inferior linear model which produces poor agreement while not underlining the fact that a slight modification of that model which leaves it still essentially linear improves the agreement enormously.

To make this point clear we may note that the calculations of viscosity effects in § 5 apply only to propagation of *sinusoidal* components of the pulse. They can be used satisfactorily both for the fundamental and for its harmonics, but a separate treatment is needed for the *mean flow*. A rather casual approach to the distribution of mean flow was adopted in most of the earlier versions of pulse propagation

theory: in fact, the error emphasised in Chapter 10, § 3 of ignoring entry effects for the mean flow and assigning it a Poiseuille distribution was widely made, and this type of model was used in Ling & Atabek's comparisons. Recognition that the whole aorta is an entry length for the mean flow, with a boundary-layer thickness not significantly greater than the Stokes boundary-layer thickness associated with fluctuating Fourier components, leads to linear-theory results in almost as good agreement with the experiments of Figure 7 as Ling & Atabek's nonlinear predictions. On the other hand, Anliker (1972) has produced interesting evidence that nonlinear effects, while of marginal significance under physiologically normal conditions, may begin to play a very significant role in pathological conditions such as aortic valve incompetence that can lead to an increase in the size of the heart and in the resulting gross ejection volume per stroke.

### REFERENCES

- ANLIKER, M. 1972 In *Biomechanics: its Foundations and Objectives* (ed. Fung, Y. C., Perrone, N. & Anliker, M.), Prentice-Hall, Englewood Cliffs, N.J.
- BERGEL, D. H. 1961 *J. Physiol.* **156**, 458–469.
- LING, S. C. & ATABEK, H. B. 1972 *J. Fluid Mech.* **55**, 493–511.
- LOVE, A. E. H. 1927 *A Treatise on the Mathematical Theory of Elasticity*, 4th Edition, Cambridge University Press.
- MCDONALD, D. A. 1960 *Blood Flow in Arteries*, Edward Arnold, London.
- REMINGTON, J. W. & WOOD, E. H. 1956 *J. Appl. Physiol.* **9**, 433–442.
- ROSENHEAD, L. (ed.) 1964 *Laminar Boundary Layers*, Oxford University Press.
- TAYLOR, M. G. 1964 In *Pulsatile Blood Flow* (ed. Attinger, E. O.), McGraw-Hill, New York.
- , 1965 *Phys. Med. Biol.* **10**, 539–550.
- , 1966 *Biophys. J.* **6**, 29–51.
- WOMERSLEY, J. R. 1957 *Phys. Med. Biol.* **2**, 178–187.

## CHAPTER 13

### Blood Flow and Arterial Disease

#### 1. Observations of turbulence in blood flow

This chapter illustrates the applications of internal biofluidynamics, by outlining work on certain interactions between local features of blood flow and the onset of arterial disease. The chapter ends with two long sections that develop in detail the comments on 'incipient atheroma' made in § 4 of Chapter 10, rather as § 8 and § 9 of that general introduction to internal biofluidynamics were developed in Chapter 12. We begin, however, by expanding the material of Chapter 10, § 10 concerning *turbulence* in blood flow, in preparation for two sections that describe pathogenic effects of turbulence on arterial walls (with relevance to problems of cerebral hemorrhage).

The work of Seed & Wood (1971), which described a short burst (see Figure 1 for a dog's aorta) of turbulent fluctuations of velocity during the early part of that period when the flow is decelerating, has since been extended by Nerem & Seed

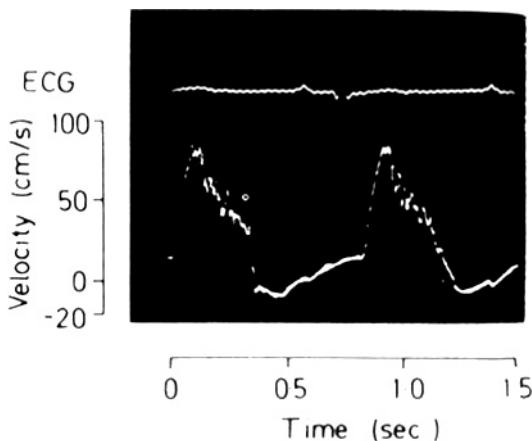


FIG. 1. *Velocity-time trace in a dog's aorta, plotted alongside the ECG signal indicating the instant of incipient contraction of the heart muscle (Seed & Wood 1971).*

(1972) and by Parker & Seed (1973). They developed a technique for obtaining an 'ensemble average' over (say) 500 pulse beats of the velocity-time trace measured locally by a hot-film anemometer, with the origin of time for each beat taken as the instant when the rising velocity passes through a certain positive value. The

turbulent velocity component in any one beat is then defined as the difference of the velocity-time trace in that beat from the ensemble average.

They distinguish three régimes of flow in the mammalian aorta:

- (i) *laminar*, with essentially identical velocity-time traces in different beats and therefore negligibly small 'turbulent' velocity components;
- (ii) *transitional*, with rather a short burst of significant turbulent velocity fluctuations in each beat, immediately following the peak average velocity;
- (iii) *turbulent*, exhibiting larger, more prolonged turbulent velocity fluctuations, with r.m.s. about 10% of maximum velocity, that extend as far as the low-flow conditions around diastole.

Experiments in man and in experimental animals under different conditions indicate that the Reynolds number  $R$  based on peak velocity and tube diameter is the parameter most influencing the choice of régime: laminar flows predominate with  $R$  around 4000 or less, turbulent flows with  $R$  around 8000 or more, and transitional flows in between (around 6000). Reynolds numbers in the aortas of normal human subjects span this whole range and all three conditions are found. In dogs, Reynolds numbers tend to be lower and flow is typically laminar but exceptionally transitional as in Figure 1. Reynolds numbers in horses are higher and flow is normally turbulent.

The choice of régime shows also some dependence on  $\alpha$ , Womersley's parameter  $a(\omega/v)^{1/2}$  (equation (47) of Chapter 12). As  $\alpha$  increases (for example, as an individual's pulse-rate frequency  $\omega$  increases) the critical Reynolds number  $R_c$  for onset of the 'transitional' flow regime increases somewhat. This is consistent with a general view that, when values of  $R_c$  in any non-Poiseuille pipe flow are considerably greater than the value 2000 typical of Poiseuille flow, it is because the vorticity is confined to a thin boundary layer and the Reynolds number based on that boundary layer's thickness is what principally affects transition. This point of view was expressed in § 2 of Chapter 10 for steady secondary flows, but is equally likely to apply to thin Stokes boundary layers of thickness proportional to  $(v/\omega)^{1/2}$  or to thin entry-length boundary layers. Nerem & Seed (1972) and Parker & Seed (1973) obtained values at around the onset of transition of a Reynolds number based on the thickness  $(v/\omega)^{1/2}$  of the order of magnitude 100, just as Collins (1963) obtained for visual observations of transition in the Stokes boundary layer on the bottom of a water channel over which small surface waves were travelling.

The mechanism by which laminar flow is disturbed is likely to be the rapid amplification of those small disturbances, which from a variety of causes are inevitably present in the blood flow, as soon as flow deceleration gives the velocity profile a point of inflexion. Lighthill (1970) summarises evidence for the rapid effectiveness of this mechanism at Reynolds numbers that are not too low, and for its power to generate large-scale disturbances far outside the region of the inflexion point itself.

The turbulence described above, as occurring within normal human aortas during the deceleration phase but before aortic valve closure, is not normally heard by a physician using a stethoscope. It is interesting to ask why this is so whereas the turbulence associated with jet formation behind a stenosed aortic

valve (one failing to open fully) is clearly audible. Parker & Seed (1973) used their catheter velocity probe to measure, for turbulence behind a stenosed aortic valve, r.m.s. velocities about twice as great as for turbulent motions in normal aortas. Associated pressure amplitudes might be expected to be four times as great, implying a twelve-decibel (12 db) enhancement of noise if the same mechanisms for attenuation of the pressure signal on transmission through the arterial wall and through the soft tissue beyond it apply in both cases. In fact, the discontinuity in properties of the arterial wall represented by the presence of the valve is expected to facilitate transmission of nearby pressure fluctuations, and the mechanical coupling of the jet fluctuations to the surface of the stenosed valve itself could further increase the enhancement as high as 20 db which would suffice to explain the difference between clear audibility and effective inaudibility.

The whole art of accurate measurement of turbulent blood-velocity fluctuations is, however, only in its infancy. Wide future use of the hot-film anemometer to measure turbulence *in vivo* will add greatly to our quantitative knowledge of its distribution in health and in disease.

## 2. Poststenotic dilatation

Pathogenic effects of the turbulence in the region distal to a stenosis were first clearly demonstrated by Roach (1963a, 1963b), who made a comprehensive study of the excess arterial dilation commonly observed behind stenosed aortic or pulmonary valves, or behind other constrictions in arteries (also called stenoses even though not associated with valves). Such 'poststenotic dilatation' had been regarded as 'a hemodynamic paradox', since de Vries & van den Berg (1958) confirmed conclusively by experiments in steady and pulsatile flow, in rigid and elastic vessels, in isolated and living aortas, what fluiddynamic theory would suggest: namely, that the pressure distal to a stenosis never exceeds, and is normally less than, the proximal pressure. It follows that poststenotic dilatation must be associated with locally increased distensibility of the arterial wall.

Roach (1963a) demonstrated in 50 experiments using the femoral and/or carotid arteries of dogs, exposed *in vivo* and subjected to different degrees of stenosis by external application of nylon bands, that poststenotic dilatation occurred when and only when turbulence distal to the stenosis could be detected by a microphone firmly held close to the artery. This happened in 39 stenoses with 23% to 76% decrease in lumen area. Average dilatation involved about a 20% distal increase in diameter. Microphone signals were at frequencies typically in the region 50 to 500 Hz, the corresponding sounds being described as 'thrill' and 'bruit'.

By contrast, no poststenotic dilatation whatever was found in 7 arteries with a minor degree of stenosis that did not generate any such indications of turbulence. Neither was it found in 4 arteries with more than 85% decrease in lumen area and the consequent formation of a sluggish thin jet, again without any indication of turbulence in the microphone signal.

Roach (1963b) obtained further evidence for both the influence of turbulence and the nature of its effect. She demonstrated significant changes in various other

elastic properties of a segment of artery, that occurred when (and only when) it was subjected *either* to turbulence *or* to vibrations artificially introduced in the same frequency band. The nature of the change was such as to make it most probable that the *elastin* (as opposed to the collagen) fibres changed their elastic properties. Such an effect of high-frequency fluctuations of pressure, and hence of stress in the wall, may be thought reminiscent of material fatigue in general. At the same time, the fluctuations may weaken connections throughout the 'matrix' of collagen and elastin fibres. Later work (Roach 1970) showed that normal reparative processes in dog arteries are, however, able to restore the diameter and the elastic properties of the dilated artery in due course once the stenosis has been removed.

The question of whether the lower level of turbulence described in § 1, associated not with jet formation behind an arterial constriction but with high-Reynolds-number decelerating flow in the thoracic aorta generally, may produce an influence on the elastic properties of the aortic wall, has not yet been investigated. It remains uncertain whether the associated amplitude of pressure fluctuations, estimated in § 1 as about a quarter of that behind a stenosis, is sufficient to produce a measurable effect, but such a possibility may need, perhaps, to be borne in mind.

### 3. Aneurysms in the cranial circulation

In this section we describe the work of Ferguson (1972) where the weakening of the arterial wall in response to fluctuating pressures associated with turbulence, as described in § 2, becomes a self-aggravating pathological condition with potentially serious consequences. The condition in question is the saccular aneurysm, which occurs (Figure 2) almost exclusively at the apices of bifurcations, usually in the cranial circulation.

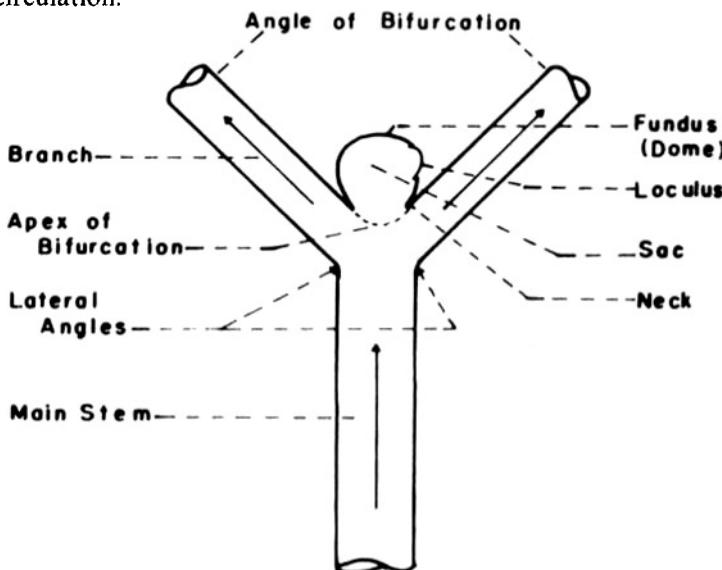


FIG. 2. This Figure 1 from Ferguson (1972) illustrates diagrammatically a typical saccular aneurysm.

Microphone observations were used by Ferguson (1972) to establish the association of turbulence with intracranial saccular aneurysms. In certain patients undergoing head surgery (craniotomy), opportunities for microphonic recording of flow sounds from the external surfaces of arteries and of aneurysms occurred. No sound was recorded from any bifurcations or other arterial sites in 4 patients without aneurysms. By contrast, 'bruits' in the frequency range 200 to 700 Hz were recorded in 12 out of 19 patients with aneurysms, and the amplitude distribution of the sound indicated its source to be in the aneurysm itself. Although among these 19 patients sound was absent in those whose mean arterial pressure had been reduced through drug action to very low values (around 60 mb or below), it was always substantial at more normally low values around 100 mb.

Ferguson (1972) used a series of model experiments to explain these results, and to indicate the likely role of turbulence in making the aneurysmic condition self-aggravating. Careful observations established the character of flow in typical bifurcations with an angle of 90° between the distal arteries. Away from the plane of symmetry, vortex lines *parallel* to that plane are stretched on entering the bifurcation to produce flow disturbances associated with a local intensification of vorticity across the apex. Figure 3(a) shows the motion of a dye-line strongly influenced by this vorticity in steady flow at a Reynolds number of 500, while Figure 3(b) shows how this complicated flow pattern has become turbulent downstream of the bifurcation at a Reynolds number of 1500. The mean transition Reynolds number was determined as 1300 in steady flow and 900 in pulsatile flow.

Other experiments including models of saccular aneurysms at the apex showed a marked interaction between the intense apical vorticity and the cavity, generating strong turbulence within the aneurysm even at  $R = 500$  as in Figure 3(c). The mean transition Reynolds number was determined as 500 in steady flow and 400 in pulsatile flow.

Ferguson (1972) argues that the effects of turbulent pressure fluctuations at the apices of arterial bifurcations may be greater in the cranial circulation than elsewhere for three reasons. Intracranial arteries have smaller ratios of wall thickness to diameter than other arteries; they have a lower elastin content; and they are not so well supported by surrounding tissue. Taking all this into account, he envisages the initiation and growth of intracranial saccular aneurysms, as influenced by the pathogenic effects of turbulence on the elastin fibres and on the integrity of the matrix of fibres within the arterial wall, as follows.

A localised degeneration of the internal elastic membrane at the apex of a major intracranial arterial bifurcation may gradually come about as a result of fluid-dynamically generated stresses of various kinds, including the effects of peak systolic pressures, and of the shear stresses associated with local intensification of vorticity, together with occasional transitions to turbulence under conditions of high cardiac output when the local Reynolds number may be elevated above 1000. The pulsatile pressure head in systole may then produce in the weakened apex the outpouching which represents the initiation of the aneurysmic condition.

Its self-aggravation then results from the far *more* frequent occurrence of turbulence, associated with a drop in critical Reynolds number from, say, 900 to 400

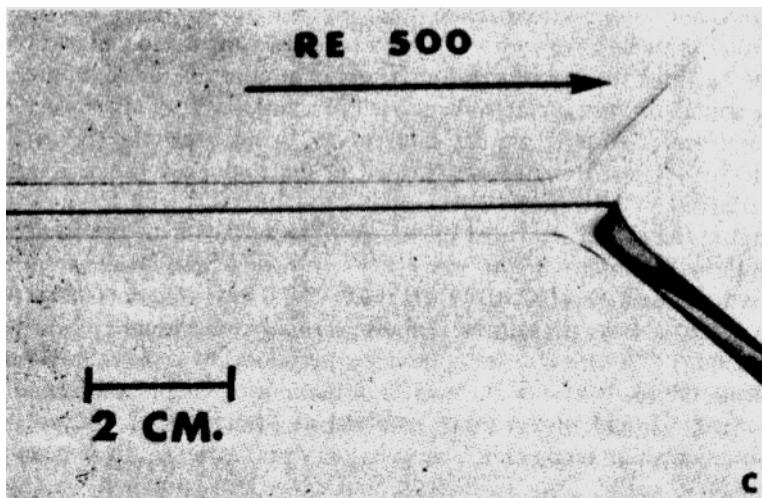
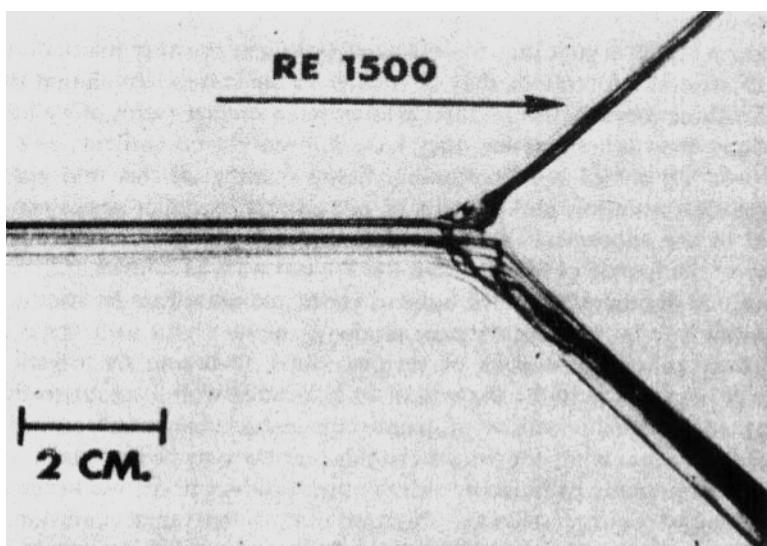
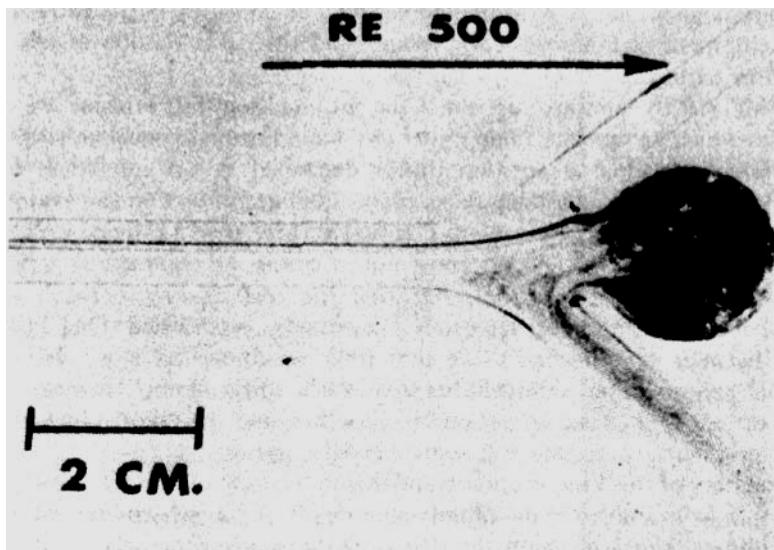


FIG. 3. *Flow visualisation by Ferguson (1972): (a) normal bifurcation with laminar flow;*

when an aneurysm is present. Degeneration of the elastin is expected, and was demonstrated by Ferguson to have occurred in samples of aneurysmal tissue under elastic test. Connections in the matrix of collagen fibres may also be weakened, allowing the aneurysm to grow progressively. Ultimate rupture of the aneurysm becomes more and more likely as it grows, since the maximum stress in the



(b) *normal bifurcation with turbulent flow;*



(c) model of saccular aneurysm with turbulence even at a Reynolds number  $R = 500$ .

aneurysmal wall is  $\frac{1}{2}ap_e/h$  where  $a$  is the radius,  $p_e$  is the excess pressure head in systole and  $h$  the minimum wall thickness. When the radius  $a$  has grown sufficiently, with an associated reduction in the minimum thickness  $h$ , rupture leading to cerebral haemorrhage becomes likely in any circumstance leading to unusually high systemic blood pressure.

The importance of the picture of aneurysmal development thus sketched out is its likely relevance both to problems of diagnosis and to assessment of the case for surgical interference by craniotomy including dissection and clipping of an aneurysm.

#### 4. Atherogenesis by intercellular lipid deposition

Among the arterial diseases involving a hardening or sclerosis, one is particularly widespread: atherosclerosis, which derives its name from the Greek word *athērē* meaning porridge. It involves a lesion of the arterial wall characterised by accumulation in the intima (wall region nearest the lumen) of lipids: a generic term that comprises fatty acids, sterols and their esters. Earlier stages of the lesion may have a 'mushy' or 'gelatinous' quality that explains the choice of name. Later stages exhibit a hardening by sterol-induced fibrosis, and by local disorders in the process called fibrinolysis which may give rise to blood-platelet adherence, fibrin encrustations, calcification and the formation of thrombi (clots).

We here omit descriptions of the many serious consequences of these later stages of hardening by atherosclerosis: not only thrombosis but the building up of atherosclerotic 'plaques', with progressive encroachment on the lumen and reduction in arterial distensibility, both leading to increasing demands on the heart musculature. We are concerned rather with *atherogenesis* and its relation to

biofluidynamics: that is, with the increasingly active study of the early stages in the development of atherosclerotic lesions and their distribution as affected by blood-flow patterns.

Lighthill (1973) summed up the Ciba Foundation Symposium on atherosclerosis in which he was chairman under two main headings: mechanisms starting with *intercellular* lipid deposition (lipids deposited in the interstices between cells) and mechanisms involving *intracellular* lipid accumulation (excess uptake of lipids within cells). This convenient distinction, here used to divide up the work described as between § 4 and § 5, must not of course be regarded as a rigid one. It should be recognised, furthermore, that the relationships between the two types of lipid accumulation have not been finally established. One broad difference between them seems to be that lipid accumulation is a relatively less *reversible* process when intercellular than when intracellular. However, the influence of one type of accumulation on the other, and the relative importance of the two in atherosclerosis, are still controversial questions.

Uncertainty of this kind is understandable in the light of the fact that a patient's arteries can be examined in detail only after death. When atherosclerosis has been a contributory cause of death the disease is usually too far advanced for such examination to throw much light on atherosclerosis. Early stages of lipid accumulation can, of course, often be examined in a subject after death from *other* causes, but without information on how those lesions would have developed had the patient lived.

The predominant lipids involved in both types of accumulation are the cholesterol esters, rather than free cholesterol or triglycerides or phospholipids. Among the cholesterol esters it is particularly the 'linoleate' ester that appears in the early development of fibrous plaques. This is regarded by Smith & Slater (1973) as an important clue to the probable importance for atherosclerosis of the intercellular deposition of lipids, observed in normal arteries to occur along fibres in the matrix of arterial collagen and elastin, since in these finely drawn out perifibrous deposits of lipid the linoleate ester normally predominates. By contrast, the 'oleate' ester predominates in intracellular lipid accumulation, possibly as a result of esterification within the cell.

Cholesterol and its esters, like most lipids, are exceedingly insoluble: lipids, in fact, are not transported around the body in solution. However, they readily form compounds with proteins and these compounds, the *lipoproteins*, are commonly soluble and present in solution in the blood plasma. Lipid accumulation in arterial walls involves, then, the penetration of lipoprotein into the intima and its subsequent dissociation into the lipid component and a remaining protein component called the *apoprotein*. The rate of accumulation may depend on many variables, including

- (i) average concentrations of lipoproteins in the blood plasma;
- (ii) factors which determine what concentration of a particular lipoprotein in the intima is in equilibrium with a given concentration in the plasma;
- (iii) the permeability of the endothelial wall of the artery, which determines the rate of diffusion of a lipoprotein from the plasma when its concentration

- in the intima is below the equilibrium value (or of diffusion into the plasma under conditions of excess intimal concentration);
- (iv) rate of dissociation by active uptake of the lipid through a cell membrane (a process where the apoprotein may remain outside the cell while the lipid combines with a new apoprotein inside);
  - (v) rate of dissociation by perifibrous deposition of the lipid component.

This last type of dissociation, which produces intercellular deposition of lipid, is viewed as being potentially a self-aggravating process in the longer term (Smith & Slater 1973). Perifibrous deposition over an extended period appears to generate a looser, thickened, more transparent, 'gelatinous' structure in the intima, which permits it to accommodate over twice as high a concentration of lipoprotein in equilibrium with a given plasma concentration.

The evidence that this gelatinous structure may be a precursor of an atherosclerotic plaque derives from its being frequently observed around the edges of actual fibrous plaques, as though the gelatinous edge were the growing point of the plaque. The enhancement of lipoprotein concentration within this gelatinous edge is found to be around *four* times that in normal intima, presumably accelerating further deposition of lipid. The developed plaque is even more penetrable.

Physico-chemical studies (Iverius 1973) suggest the significance for perifibrous deposition of a variety of protein mucopolysaccharides, called glycosaminoglycans, within the matrix of fibres making up the arterial wall. These long-chain molecules produce a general reduction in the space available for other large molecules, including lipoproteins, that tends to facilitate their precipitation. The looser-structured lipoproteins (molecules of typical size  $0.02\text{ }\mu\text{m}$  called LDL or 'low-density lipoprotein' by many biochemists and thought to be particularly responsible for the transport of cholesterol ester) are shown by Iverius to be rather readily bound to a glycosaminoglycan such as dermatan sulphate possessing negatively charged carboxyl and sulphate groups. This is significant because dermatan sulphate is found in close association with collagen fibres.

Fry (1973) has particularly studied biofluidodynamic factors that may modify the structure of the arterial wall in such a way as to facilitate lipoprotein penetration and intercellular deposition. Some of these factors are familiar to the reader already from the work of Chapter 12 and earlier sections of this chapter.

The discussion is potentially made ambiguous by the fact that the word 'stress' is used with quite different meanings in general biology (including medicine) where it describes abnormal demands on an animal by its environment, and in the mechanics of solids, where it describes force per unit area. Stress in the general medical sense might, perhaps, be expected to make arterial walls more vulnerable to lipid accumulation and has, indeed, been shown to be correlated with incidence of atherosclerosis. Some of the influence of 'stress' in this sense may be biofluidodynamic as we shall see, but major components of the influence are likely to be biochemical and therefore outside our scope.

From now on we use stress in the sense of the solid mechanics of the arterial wall as in § 2 of Chapter 12. Fry has experimentally demonstrated various responses

of the arterial wall to different types of mechanical stress, some of which alter the permeability emphasized under heading (iii) above.

It is particularly interesting in the light of those responses to turbulence described in § 2 and § 3 that he finds good correlation *in vivo* between positions of high transendothelial protein flux and positions with substantial high-frequency turbulent stress fluctuations. A fluiddynamicist might, perhaps, be inclined to postulate a connection between high flux and the reduced resistance to diffusion across the boundary layer which turbulence allows, but Fry shows that boundary-layer resistance to diffusion, even under laminar conditions, is quite negligible compared with the resistance to penetration of the endothelial wall, implying that changes within that wall due to turbulence, possibly of the general kind described in more extreme situations in § 2 and § 3, must be responsible for the correlation.

Among more steady types of stress distribution, Fry has successfully demonstrated that those producing *extension* of a piece of excised artery increase protein permeability. The *circumferential* extension emphasized in Chapter 12, § 2 as associated with excess pressure may therefore facilitate extra protein penetration each time the pressure reaches a peak value, particularly in patients with a high *mean* blood pressure (and indeed such patients are particularly prone to atherosclerosis). Other *in vitro* experiments showed that in the *absence* of extension the effect of excess transmural pressure on protein permeability was negligible; this rules out the concept of a forced-flow type of penetration and emphasizes the diffusional nature of the phenomenon.

Shearing stress also has been shown *in vitro* (Carew 1971) to increase protein permeability, by an amount proportional to shearing stress squared; such an even-function dependence on a directed quantity like shearing stress would be expected from considerations of invariance under direction. At the very high shearing stress of  $40 \text{ Nm}^{-2}$ , much higher than would be expected to occur physiologically, there is a yielding of the endothelium in the engineer's sense: it has what Dr. Fry calls a 'yield stress' producing gross damage at that level. It remains a matter for speculation whether exceptionally high shearing stresses *within* the range expected to occur in man may ultimately generate endothelial changes relevant in atherosclerosis: such changes are certainly suggested by experiments of Caro & Nerem (1973) described in § 5. The actual time for which the stress is applied, or the number of cycles of stress, may be considered very relevant, since Dr. Fry's experiments on yield stress were for quite short times of application, and since high-frequency application of stress was in § 2 and § 3 shown capable of producing gross changes.

In a certain series of experiments on dogs, Fry and his colleagues created an arteriovenous shunt in one of the two iliac arteries, leading to average Reynolds number of 70 and 800 in the 'control' and shunted arteries, respectively. Significant erosion of the endothelial wall occurred in the shunted artery. A shearing stress around  $3 \text{ Nm}^{-2}$  was calculated from the observed flow rate in that artery by assuming a Poiseuille distribution. On the other hand, the bifurcation may have produced enhanced shearing stresses of the type described in Chapter 11, § 2, or even (as § 3 above suggests) turbulence, which might have directly damaging consequences.

In arteries, of course, various reparative responses to stress, and to wear and tear generally, are continually taking place. Response to the characteristic *direction* of shearing stress has been demonstrated by Fry as occurring with a time constant of a few weeks and as involving an orientation of the cell nuclei into that direction, possibly acting to strengthen the endothelium against shear stress. A big change in the perfusion rate for a particular peripheral region might then alter the flow proportions at a bifurcation and bring about large new shearing stresses in a direction in which the endothelium was weaker.

Arterial response to high mechanical stress can also involve an intimal thickening described as fibromuscular hyperplasia, which has often been observed to promote lipid deposition. Another experiment with an arteriovenous shunt, this time in one *carotid* artery of a dog, was found by Fry to generate this type of response to the increased biofluidodynamic stresses.

There are many suggestions here of how detailed biofluidynamical flow patterns may affect atherogenesis. The connections are by no means finally established, partly because experimental knowledge on the flow patterns as well as on the lipid deposition remains inadequate; this is a spur to deeper study of both aspects separately, and of their connections.

### 5. Intracellular lipid accumulation

We end this chapter with an account of some factors that have been shown to influence accumulation of lipid *within* cells in general, and within intimal cells in particular, either to normal or to pathological levels. We also review evidence on the nature of the relationships between this type of lipid accumulation and atherosclerosis.

Cells require lipids: in particular, a lipid supply is needed both by the mitochondria and by the cell membranes. Bailey (1973) has described processes helping to regulate the lipid content of cells in general. He demonstrated by radioactive-tracer studies that a cell in equilibrium with a surrounding serum containing blood proteins and lipoproteins is continually *excreting* lipids, which bind with blood proteins such as alpha-globulins to form lipoproteins; and simultaneously *ingesting* lipids, typically leaving just such alpha-globulins outside the cell as dissociated apoproteins. His work suggests a highly *reversible* mechanism (not at present understood in detail) for transfer of lipid across the membrane.

When lipid is unavailable in the surrounding medium, it is produced in the cell by *de novo* synthesis (Bailey 1973) from glucose and other simple precursor molecules. On the other hand, significant lipoprotein content in the surrounding medium inhibits this biosynthesis almost completely after the time lapse of a few hours needed for significant transfer of lipid from that medium to the cells. Thus there are at least two mechanisms of control of cell lipid content: a certain 'threshold' level inhibits biosynthesis, while excretion supplies a further control mechanism that continues to operate beyond this level.

Certain types of cell, including the endothelial cells in the wall of arteries, are well endowed with 'plasmalemmal vesicles' that have been shown able, in the process known as 'pinocytosis', to entrap even such a large molecule as a lipoprotein and transport it into the cell. Stein & Stein (1973) regard this as an additional

mechanism of potentially dominant significance for lipid accumulation by intimal cells, although Bailey (1973) expresses the opposite view that such cells accumulate lipid principally from lipoprotein that has penetrated the intercellular spaces, and by a transfer process in which the apoprotein is excluded.

Whatever the mechanisms for synthesis, ingestion or excretion of lipid by intimal cells, there is some clear evidence that the whole combination of processes controlling excess lipid content in those cells can break down, at least when the ambient lipoprotein concentration is high. Lipid seems then to adopt new locations for accumulation within the cell (Peters, Takano & Duve 1973), and cells described as 'fat-filled' (containing a quite abnormal amount of lipid) can result.

The arteries of normal young adults, examined following accidental death, usually show a considerable number of 'fatty streaks': areas of arterial wall containing numerous fat-filled cells. It is a highly controversial question whether or not the distribution of these fatty streaks has a significant influence on the subsequent development of atherosclerotic plaques.

On the one hand, Mitchell & Schwartz (1965) studied the distribution of fatty streaks and atherosclerotic plaques in the aortas of 600 people who died from all causes. In young people, there was relatively little fatty streaking in the abdominal aorta; much more in the thoracic aorta, but with the lateral and anterior walls conspicuously spared and the mouths of the intercostal arteries relatively spared. In older people with atherosclerotic plaques, there were many more plaques in the abdominal than in the thoracic aorta, where in turn the anterior and lateral walls and the mouths of intercostal arteries were *not* spared. The total area of fatty streaks was not correlated with age or sex, whereas the area of plaques increased with age and was much greater for men than for women (and, incidentally, was correlated with high blood pressure).

Mitchell & Schwartz (1965) see these data on location as evidence against the idea that fatty streaks develop into fibrous plaques. Further evidence from Smith & Slater (1973) was noted in § 4: the linoleate ester of cholesterol predominates in fibrous plaques and in their 'gelatinous' precursors, whereas the oleate ester predominates in fat-filled cells.

On the other hand, Flaherty, Ferrans, Pierce, Carew & Fry (1972) emphasize interrelationships between fatty streaks and fibrous plaques, although regarding the possibilities of development of one into the other as strongly influenced by genetical and environmental factors, and by the biofluidodynamic and structural characteristics of different areas of arterial wall. Wahlgqvist, Day & Tume (1969) studied 'foam cells' (an extreme form of fat-filled cell) from human arterial lesions, and showed that isolated foam cells take up oleic acid from the environment in just such a way as to produce that predominance of the oleate ester of cholesterol observed in them. Foam cells may form a link between intracellular and intercellular lipid deposition, either because their extreme fragility makes them prone to disintegrate and deposit their lipid, setting up responses in the arterial wall (including the inflammatory responses) that may facilitate plaque development; or conversely because some foam cells may have been formed by incorporation of

intercellular lipid in some sort of 'phage' cell involved in the immunological responses.

Caro, Fitz-Gerald & Schroter (1971) emphasize that there can be two opposing kinds of biofluidodynamic influence on the location of arterial lesions. For example, we have seen that intracellular accumulation of lipid in the earlier stages of development of the fat-filled cell may be relatively reversible, and evidence that the fatty streaks may disappear completely on a low-fat diet was provided by their absence in the arteries of concentration-camp victims. Caro, Fitz-Gerald & Schroter show that any reversible type of lesion may well be *egress-controlled*, in the sense that those factors increasing protein permeability which we listed in § 4 may actually tend to diminish it. This may happen if there is a greater resistance to egress of the excreted lipid, bound to its protein carrier, than to the ingress of in general a different form of lipid bound to a different protein carrier (or conceivably even of simpler molecules that cells can synthesise into lipids); or if pinocytosis is a significant mechanism favouring ingress at the expense of excretion.

Among the factors influencing protein permeability (§ 4), Caro, Fitz-Gerald & Schroter (1971) particularly emphasize wall shearing stress. Caro (1973) summarizes the work of Caro & Nerem (1973) and other very recent evidence from his own laboratory using labelled lipoprotein which strongly indicates an increase of permeability with shear, just like the independent experiments from Dr. Fry's laboratory quoted in § 4. His recent work is also in agreement with Fry's in showing that changes in the rate of diffusion of protein across the boundary layer are totally insufficient to explain the degree of shear-dependence observed. Direct effects of shearing stress upon the arterial wall itself, or conceivably on the structure of an adsorbed layer (Miller, Graet & Frei 1973), must be responsible for increased transport.

One suggestion from this work is that general epidemiological evidence correlating a low level of physical activity with atherosclerosis might be interpreted in terms of low wall shearing stress failing to promote egress of lipid in earlier, more reversible phases of lipid accumulation. Another possible inference is that the more reversible types of lesion characteristically found in younger people might be expected to be found in low-shear regions where biofluidodynamic features promoting accelerated egress of lipid are absent. This inference in no way excludes the idea that development of advanced atherosclerotic *plaques*, considered to be a less reversible process, that is accelerated by absorption of further lipoprotein (see § 4), may preferentially occur where the biofluidodynamic stresses favour *increased* permeability, as in high-shear regions. Here is a possible explanation of the Mitchell & Schwarz (1965) data on different preferred locations for different types of lesion that by no means excludes the possibility of development of one type into the other.

Caro, Fitz-Gerald & Schroter (1971) bring forward considerable additional data on distribution of lesions in aortas of younger people examined after death which confirm and refine the conclusions of Mitchell & Schwarz (1965). They interpret the data in terms of preferential fatty-streak formation in low-shear regions, including the part of the thoracic aorta where secondary flow produces a marked

boundary-layer thickening, together with sparing in high-shear regions, including the 'flow divider' at any bifurcation, and including peripheral arteries generally. Fry (1973), incidentally, notices the fact that the areas of arterial wall with overall highest shear, namely the flow dividers, tend to be spared, although he associates this sparing with strengthening by a *reparative* response (involving collagen enrichment) to such extreme shearing stress.

Caro, Fitz-Gerald & Schroter (1971) argue that there must also be circumstances of extreme lipoprotein concentration in the blood plasma when a normally egress-controlled lesion must become ingress-controlled. They suggest that the 'experimental atherogenesis', produced by feeding experimental animals with an exceedingly high-cholesterol diet, may be of this type. A distribution of fatty streaks is observed which is quite opposite to that found in younger humans on normal diets: lipid accumulation is rapid in high-shear regions generally, and on flow dividers in particular. This interpretation of experimental fatty streaks as ingress-controlled in contrast to the juvenile fatty streak in humans suggests that questions of the relevance of such experiments to atherogenesis in man need to be viewed with caution. We may note, however, for what it is worth, that the idea that fatty streaks may develop into atherosclerotic plaques does tend to be supported by those experiments.

One aim in this chapter has been to give readers some idea, however incomplete, of the vast variety of non-biofluidodynamic factors that enormously complicate any interactions between blood flow and arterial disease. It will be evident that numerous different biological skills, including those of the physician, the physiologist and the biochemist, have to be represented in an effective team conducting research into such problems. A further complication derives from the quite sophisticated biofluidodynamic knowledge and skills that also have been indicated as relevant to such problems: the team needs a biofluidynamicist too! It may be quite hard to weld together effective interdisciplinary teams to attack the problems of arterial disease in relation to blood flow, but the great human importance of mounting such attacks should remind us that the effort needed is indeed worth making.

## REFERENCES

(Note. In the following list of references, the abbreviation Ath. stands for Atherogenesis: Initiating Factors (ed. R. Porter & J. Knight), Ciba Foundation Symposium 12 (new series), Associated Scientific Publishers (Elsevier, Excepta Medica, North-Holland), Amsterdam, 1973.)

- BAILEY, J. M. 1973 Ath., 63-88.  
CAREW, T. E. 1971 *Mechano-chemical response of canine aortic endothelium to elevated shear stress in vitro*, PhD thesis, The Catholic University of America, Washington, DC.  
CARO, C. G. 1973 Ath., 127-149.  
CARO, C. G., FITZ-GERALD, J. M. & SCHROTER, R. C. 1971 Proc. Roy. Soc. B, **177**, 109-159.  
CARO, C. G. & NEREM, R. M. 1973 Circulation Res., **32**, 187-205.  
COLLINS, J. I. 1963 J. Geophys. Res., **68**, 6007-6014.  
FERGUSON, G. G. 1972 J. Neurosurgery, **37**, 666-677.

- FLAHERTY, J. T., FERRANS, V. J., PIERCE, J. E., CAREW, T. E. & FRY, D. L. 1972 *Localizing factors in experimental atherosclerosis*. In *Atherosclerosis and Coronary Heart Disease* (ed. Likoff, W. et al.) Grune & Stratton, New York.
- FRY, D. L. 1973 Ath., 93-120.
- IVERIUS, P. H. 1973 Ath., 185-193.
- Lighthill, M. J. 1970 *Turbulence*. In *Osborne Reynolds and Engineering Science Today* (ed. McDowell, D. M. & Jackson, J. D.). Manchester University Press.  
, 1973 Ath., 277-282.
- MILLER, I. R., GRAFT, H. & FREI, Y. F. 1973 Ath., 251-262.
- MITCHELL, J. R. A. & SCHWARTZ, C. J. 1965 *Arterial Disease*, Blackwell Scientific Publication, Oxford.
- NEREM, R. M. & SEED, W. A. 1972 *Cardiovascular Res.* **6**, 1-14.
- PARKER, K. H. & SEED, W. A. 1973 *Turbulence in the Aorta*. unpublished lecture at Euromech Colloquium 32, Imperial College, London.
- PETERS, T. J., TAKANO, T. & de DUVE, C. 1973 Ath., 197-214.
- ROACH, M. R. 1963a *Circulation Res.* **13**, 537-551.  
— — —, 1963b *Amer. J. Cardiol.* **12**, 802-815.  
— — —, 1970 *Circulation Res.* **27**, 985-993.
- SEED, W. A. & WOOD, N. B. 1971 *Cardiovascular Res.* **5**, 319-330.
- SMITH, E. B. & SLATER, R. S. 1973 Ath., 39-52.
- STEIN, Y. & STEIN, O. 1973 Ath., 165-179.
- DE VRIES, H. & VAN DEN BERG, J. 1958 *Cardiologia* **33**, 195-211.
- WAHLQVIST, M. L., DAY, A. J. & TUME, R. K. 1969 *Circulation Res.* **24**, 123-130.

*This page intentionally left blank*

# CHAPTER 14

## The Microcirculation\*

### 1. Introduction

This concluding chapter is concerned with blood flow in vessels of diameter less than 0.1 mm (that is, 100 microns), and especially with flow in the arterioles, capillaries and venules, with diameters ranging from at most 50 microns down to values as low as 5 microns. This whole peripheral part of the circulation, called the *microcirculation*, was seen already in Chapter 10, § 6 to be where most of the drop in mean pressure between the arteries and veins, amounting to between 100 and 200 mb in normal human beings, occurs.

Chapter 14 is particularly concerned with quantitative aspects of peripheral resistance: that is, of the dependence of this pressure drop upon flow rate, a dependence which turns out to be interestingly nonlinear. Peripheral resistance is important because it determines the rate of working demanded of the heart, and also because the body's very needful ability to bring about large selective variations in the rate of perfusion of different parts of the periphery (with flow rates changing by an order of magnitude) depends on the power of the smooth muscle lining the arterioles to make large changes in peripheral resistance by expansion or contraction.

Another essential characteristic of the microcirculation is the function of the capillaries (vessels without a smooth muscle lining, with diameters around 10 microns or less) in various exchange processes. Capillaries in the lung are separated from alveoli by exceedingly thin membranes which readily admit exchange of O<sub>2</sub> and CO<sub>2</sub> between the air in an inflated alveolus and the blood in a perfused capillary. Within the blood, a similar exchange process occurs between the plasma and the haemoglobin, which of course is packed within the red blood cells surrounded by an even thinner cellular membrane. In the systemic capillaries the same exchange processes occur (though in the opposite direction) and in addition there is exchange of various nutrients between the plasma and the surrounding tissue. Some of this is by diffusion, and some by convection as small quantities of plasma (less than 1 % of the total flow) squeeze through gaps (less than 0.1 micron wide) in the endothelial wall of the capillary (Guyton 1966). Much of this fraction of the plasma is ultimately sucked back into the low-pressure end of some capillary, but the rest is drained away through the quite separate lymphatic circulation, to rejoin the blood flow only in one of the largest veins.

The feature which distinguishes the fluid flows in the microcirculation most markedly from the flows in larger vessels (see Chapters 10 and 12), and also from

---

\* Lecture (1971) to International Centre for Mechanical Sciences. Udine, Italy.

most flows that are generally familiar, is that fluid inertia is totally negligible. We are concerned with flows whose pattern is determined solely by the need for pressure gradients and viscous forces to be exactly in balance, because inertial effects (accelerations times densities) are too small to disturb that balance at all. Essentially this is because the Reynolds number is always considerably less than 1 in vessels of diameter less than 0.1 mm and falls as low as 0.001 in the narrowest capillaries, while also the Womersley parameter  $\alpha$  (Chapter 10, Section 8), which measures inertial effects due to *pulsatility* against viscous effects, is considerably less than 1. In this world of the microcirculation, this 'extreme-low-Reynolds-number world', where inertia is totally negligible, we have to forget most of what we learnt in studying the larger vessels and concentrate on some completely different considerations.

Let me list just what we ought to take pains to forget, in these circumstances when pressure gradients are purely balanced by viscous forces. We must forget about Bernoulli's equation: there is no measurable difference between static and dynamic pressures. We must forget about centrifugal forces: fluid can now negotiate sharp bends without any difficulty at all, and without setting up any kind of secondary flow. Generally speaking, in fact, motions are much less sensitive to vessel geometry: there is practically no tendency to flow separation. Indeed, in a certain subset of cases (fluid satisfying Newton's viscosity law flowing in rigid vessels) the flow is completely 'reversible'.

The insensitivity to geometrical detail is further shown by the virtual absence of any 'entry region'. If fluid satisfying Newton's viscosity law enters a tube at a large angle to the axis, for example, then the difference between the motion and a Poiseuille motion with the same total flow is reduced by an order of magnitude already in an axial distance of half a radius (Fitz-Gerald 1971): an exceedingly quick adjustment to Poiseuille flow. Study of the microcirculation is facilitated by these simplifications, and by the possibility of taking over ideas from parts of low-Reynolds-number fluid dynamics which were developed in an engineering context: for example, from the hydrodynamic theory of lubrication.

To be set against these advantages, however, is the major difficulty mentioned already in the introduction to Chapter 10, under heading (iv): we are constantly forced to take into account *interactions between the red blood cells and the wall* in any of these tubes whose diameters are at most a few times those of the red blood cells. Accordingly, much of this chapter (from § 4 onwards) is devoted to those special features of the microcirculation resulting from the fact that blood is a highly concentrated suspension of red cells (not to mention many other constituents, which however are probably far less influential upon its dynamics) in plasma.

## 2. Vasomotor control of peripheral perfusion

A remarkable feature of the systemic circulation, commented on already in Chapter 10, is the vasomotor control of peripheral perfusion: that is, the power of smooth muscle to alter the diameter of small vessels in such a way as to vary flow rate through the local microcirculation by at least an order of magnitude.

Although muscle is absent in the walls of the capillaries themselves, there is a large body of evidence that vasomotor control is exercised almost entirely in the smallest pre-capillary vessels with diameters substantially less than 0.1 mm.

It would be natural to suppose that, in an intricate branched system, control of local resistance must be exercised in that group of vessels which makes the biggest contribution to total resistance, and there is substantial evidence that this is the case. In Chapter 10, we noted regretfully under heading (ii) that the evidence on distribution of tube numbers by length and diameter is scanty, but indicated in § 6 the way in which any such knowledge that exists can be used to estimate the distribution of pressure drop through a branched system on the assumption of Poiseuille's law. Kraemer (1967) carried out such a calculation which suggests that under conditions of vasodilatation about half the total pressure drop in the circulation occurs in precapillary vessels of diameter less than 50 microns (what in his classification are described as the arterioles and the smallest group of arteries): about a quarter occurs in larger vessels and about a quarter in capillaries. (Note that in this section all 'pressures' mentioned are corrected for the purely hydrostatic component so as to become 'equivalent pressures at the height of the heart'.)

Such a conclusion, that even before vasoconstriction about half of the total resistance in the cardiovascular system is confined to vessels in quite a small size bracket (say, 15 to 50 microns), which indeed are the vessels within which vasomotor control of peripheral perfusion is thought to be exercised, is quite striking. Nevertheless, Kraemer has to assume an extremely large vasomotor increase in the resistance of those vessels to explain the large observed changes in perfusion rates, since half of the unperturbed total resistance is not under control.

It is natural to try to probe this slight difficulty by considering the effect of departures from Poiseuille's law. One type of departure, the high-Reynolds-number type, discussed in Chapter 10, would accentuate the difficulty, but to an insignificant extent since it augments the contributions of only those rather wide tubes where quite small pressure drops are in fact found. Another type characteristic of the microcirculation, the Fahraeus-Lindquist effect discussed below in § 4, was actually allowed for in Kraemer's calculations by assuming a 40% reduction in pressure drop below the Poiseuille value in tubes of diameter less than 25 microns. More recently another possibility was suggested (Lighthill 1968, p. 116), that a nonlinear resistance property of the microcirculation, resulting from features of the flow in the narrowest capillaries, may augment the effectiveness of vasomotor control; these features will be described in detail in § 5.

It would be out of place to discuss here the many different types of stimuli that produce vasomotor action; most of them (for example, temperature as stimulating vasodilatation, or a wide variety of chemical stimuli) have little to do with fluid dynamics. It is however appropriate to note that marked vasomotor response to a rapid change of internal fluid pressure was established by Caro, Foley & Sudlow (1970). Their work suggests that a sudden pressure reduction, due to deflation of a cuff that had been producing congestion in the veins of the forearm leading to an excess venous pressure of 40 mb, may produce, after a delay time of 2 to 3 seconds, a sudden vasomotor increase in forearm blood flow to almost three times the

normal resting value although during the brief delay before vasomotor action occurred the flow rate was above that resting value (possibly due to passive dilatation by the excess pressure) by only 25 %. This example of an active response to pressure change emphasizes the need for the specialist in the mechanics of fluids and of passively deformable tubes to approach the study of physiological phenomena only with the greatest respect and caution!

### 3. Pulmonary perfusion and ventilation

The microcirculation in the lungs exhibits several features that contrast very sharply with those for the systemic microcirculation summarised in § 2. There is no vasomotor activity in pre-capillary vessels, or elsewhere, and most of the resistance to flow appears to lie in the capillaries themselves (Fowler, West & Pain 1966). However, perfusion of the capillaries can be extremely unequal as between different parts of the lung, though we shall see that this is essentially a passive effect in which gravity plays a leading role. Since the lung's effectiveness as a gas-exchange system depends upon a reasonably good match being reached between the distributions of perfusion and of ventilation across different parts of it (Farhi 1966), it is worth observing that gravity also acts to promote inequalities of ventilation of the alveoli which go some (though not all) of the way towards matching the inequalities of perfusion (West 1966).

Typical capillaries in the lung pass through the thin septum that separates two alveoli, and are themselves separated from alveolar gas by the still thinner pulmonary membrane. Accordingly they are very easily distorted whenever the alveolar pressure  $p_A$  exceeds the blood pressure in an adjacent capillary. This distortion is found by West, Glazier, Hughes & Maloney (1969) to involve very large reductions in mean width, from 6.5 to 2.3 microns; even for a fluid satisfying Newton's viscosity law this would involve an enormous increase of resistance, and consequent flow reduction, but for real blood, for reasons (see § 5) involving the size of individual red cells, it brings about a complete stoppage of flow.

It is necessary to clarify immediately the meaning of the two pressures here compared, starting with the alveolar pressure  $p_A$ . In lung inflation, involving expansion of hundreds of millions of alveoli to a diameter of a few hundred microns, an enormous increase of essentially wet surface area occurs; it is like the formation of a foam! Thus, a very large effort would be required for inflation if the liquid wetting the surface were water with its high surface tension. Fortunately a certain protein is present in the liquid which operates rather like the detergents used to facilitate foam formation; this enormously reduces surface tension (Pattle 1966). It means also that the gas pressure  $p_A$  in an alveolus is essentially what determines closure of an adjacent capillary, any surface-tension correction to it being rather insignificant.

As for the blood pressure in the adjacent capillary, this must take values intermediate between the pressures  $p_a$  and  $p_v$  on the arterial and venous side of the pulmonary circulation, the pressure fall between these values being balanced by capillary resistance. The heart pump produces a mean excess pressure in the pulmonary arteries less by an order of magnitude than that in the aorta: for

example, about 12 mb during resting, although during exercise this can easily be doubled. Such figures are comparable with the hydrostatic pressure difference between the top and bottom of the lung, and in this section we cannot simplify as we did in § 2 by ignoring the hydrostatic component, since pressures in the circulation are to be compared not merely with each other but also with the gas pressure  $p_A$  in an alveolus.

Thus  $p_a$  and  $p_v$  show a hydrostatic decrease with height of 1 mb per cm. Actually there is *some* decrease in the alveolar pressure  $p_A$  with height, at a rate around 0.3 mb per cm. Fully to relate this to the mechanics of the lung hanging under gravity, supported through the thin layer of pleural fluid, requires lengthy discussions outside the scope of fluid dynamics; but the outcome is that the highly deformable lung structure essentially has some properties in common with a fluid, including just such a gradient of internal pressure close to the hydrostatic value for the mean specific gravity 0.3 of lung tissue. West (1966) indicates how this gradient is responsible also for certain variations of ventilation with height which as already remarked go some way towards matching the bigger variations of perfusion now to be discussed.

With all reservations regarding  $p_A$ , we have seen that  $p_a$  and  $p_v$  show a much bigger vertical variation. In general, then, the lung can be divided into three zones, defined (West 1966) as follows:

Zone 1 is the highest zone where  $p_a < p_A$ .

Zone 2 is the intermediate zone where  $p_v < p_A < p_a$ .

Zone 3 is the lowest zone where  $p_A < p_v$ .

In Zone 1, any section of capillary (wherever its pressure may be between  $p_a$  and  $p_v$ ) collapses under the alveolar pressure  $p_A$  and is without flow, as confirmed experimentally by West, Glazier, Hughes & Maloney (1969). In Zone 3 the tubes run fully open; low down in Zone 3 they are observed to bulge slightly, which may by mechanisms discussed in § 5 be responsible for the very low pressure drops  $p_a - p_v$  there found.

It might be thought sufficient to say that Zone 2 is a transitional zone, but its behaviour is of some theoretical interest and has attracted much attention. For thin collapsible tubes carrying homogeneous fluid between reservoirs at pressures  $p_a$  and  $p_v$  an intermediate pressure  $p_A$  external to the tube produces (Permutt, Bromberger-Barnea & Bane, 1962) a partial closure at the distal end with a resulting local high resistance balancing an abrupt drop in fluid pressure from  $p_A$  to  $p_v$ . The flow times the normal resistance of the main tube equals the pressure drop  $p_a - p_A$  up to this distal constriction. This implies the flow property often described as a 'Starling resistor' (a clearer designation than the above authors' name 'waterfall' with its inertial connotations): flow is proportional to  $p_a - p_v$  till  $p_v$  drops below  $p_A$  and thereafter is proportional to  $p_a - p_A$ . The distal partial collapse is inevitable for homogeneous fluid since full flow in the tube must cause tube closure wherever pressure falls to below  $p_A$ , but complete closure would allow the proximal pressure  $p_a$  to penetrate the whole tube and force it open.

The situation is more complicated for whole blood, since flow can cease in passages not fully closed and support a pressure gradient through dry friction of

red blood cells against walls (see § 5). Observations indicate that the Starling resistor remains a good description of pressure-flow relations but that the mechanism underlying it changes (West, Glazier, Hughes & Maloney 1969): briefly, it involves closure of an increasing number of passages as  $p_v$  falls below  $p_A$ , and recruitment of additional passages as  $p_v$  rises again.

Note that during exercise  $p_a$  is increased so that Zone 1 (without flow) may be absent; indeed, all the zone boundaries are then raised. Note also that in this brief account of the pulmonary circulation we have not implied that the capillary passages have the same tubular geometry as in the systemic circulation. Recent opinion (see, for example, Fung & Sobin 1968) tends to a view of the different capillary passages in the septum as being like the alternative routes for threading one's way across a hall crowded densely with thick pillars! However this may be, most pulmonary capillaries are passages which tend to squeeze red cells smaller than their undistorted diameter of 8 microns. This facilitates gas transfer between the pulmonary membrane and the haemoglobin, but it does raise special questions of lubrication and resistance in very narrow capillaries, which are taken up in § 5 after the problems of resistance in vessels not quite so narrow are first discussed in § 4.

#### 4. Axial concentration

To complement discussions in earlier chapters on several kinds of departure from Poiseuille's law within large vessels, the rest of Chapter 14 is devoted to those quite different kinds of departure that are found in very small vessels, and to their implications for the microcirculation. These are departures observed in passages whose width is comparable with the undistorted-red-cell diameter of 8 microns, and they take a rather surprising form: as the diameter of tube falls below about 30 microns, the pressure drop starts to take values substantially less than those indicated by Poiseuille's law using viscosity figures found adequate for blood flow in large tubes. This phenomenon, discovered by Fahraeus & Lindquist (1931), is discussed in § 4. However, the reversed tendency (to be discussed in § 5) is found in the narrowest capillaries, with widths around 6 microns or less. Accordingly, an 'apparent viscosity' inferred from Poiseuille's law falls initially (by about 50%) as the tube diameter decreases, reaching a minimum at a diameter of around 12 microns, and then rises to well beyond the large-tube value as the diameter falls further.

These phenomena are not significantly dependent on any variation of whole-blood viscosity as a function of rate of shear, such as was mentioned under heading (iv) in Chapter 10; actually, in the microcirculation as in large vessels, we are concerned with rates of shear too large for this variation to be substantial. It is essentially the interaction of suspended red cells with the vessel wall that produces the Fahraeus-Lindquist effect, to be discussed in § 4; considerations relevant to the reversal of the effect, occurring in passages so narrow that red cells need to suffer substantial deformation even to enter them, are postponed to § 5.

Fahraeus & Lindquist (1931) found, in fact, a statistical tendency for red cells to shun the wall and to be concentrated preferentially toward the axis. We

momentarily postpone discussing the subtle fluid mechanics involved in this observed 'axial concentration' phenomenon, in order to give first the rather simple explanation of how it brings about the reduced apparent viscosity.

The kinematic viscosity of the blood plasma at body temperature is about  $1.5 \text{ mm}^2 \text{ s}^{-1}$ , while that of whole blood increases with increasing haematocrit (that is, concentration of red cells by volume) slowly at first but at an increasing rate: that is, the curve of kinematic viscosity against haematocrit is concave upwards, passing through values around 4 at normal haematocrits of 0.4 to 0.5, and then rising to much higher values. Under conditions of axial concentration, therefore, the viscosity varies greatly with radial position, from a value near 1.5 in the almost cell-free region within a few microns of the wall to far larger values near the axis.

On the other hand, the elementary theory of laminar axisymmetrical flow with viscosity  $\mu$  dependent on distance  $r$  from the axis tells us that at given pressure gradient the equilibrium of a cylinder of radius  $r$  requires the rate of shear to vary as  $\mu^{-1}r$ ; in the situation just described both factors combine to make the shear very much greater near the wall than elsewhere, as is observed. The total flow can be written as an integral with respect to  $r$  of the product  $\pi r^2$  times the rate of shear, which is proportional to  $\mu^{-1}r^3$ . This  $r^3$  factor weights still more the high values of  $\mu^{-1}$  found where  $r$  is greatest: that is, near the wall. In fact we conclude from the analysis that the reciprocal apparent viscosity  $\mu_{ap}^{-1}$  (proportional to flow divided by pressure gradient) is the weighted mean of  $\mu^{-1}$  with respect to  $r$ , using  $r^3$  as the weighting factor.

This very high weighting given to the high values of  $\mu^{-1}$  present in any almost cell-free layer near the wall makes it easy to see why an apparent viscosity as low as 2 to  $2.5 \text{ mm}^2 \text{ s}^{-1}$  is typically observed for tubes of 10 to 30 microns diameter. Whitmore (1968) points out that even the extreme hypothesis of effectively infinite viscosity for values of  $r$  up to 80% of the tube radius, and constant finite viscosity beyond, gives an apparent viscosity only 1.7 times that constant value. This model might represent blood of high haematocrit filling a 30 micron tube except for a cell-free layer 3 microns thick near the wall. Alternatively, for a 10 micron tube it might represent the effect of individual red cells passing down the tube in single file, shunning the wall (see below) to form an 'axial train' of diameter 8 microns within which plasma shearing is effectively prevented.

Postponing to § 5 the question of what happens at still lower tube diameters when plasma layer thicknesses are far more restricted in value, we next note other consequences of axial concentration besides the Fahraeus-Lindquist effect. First, it means that the 'average speed of red cells' exceeds the 'average speed of plasma' in these vessels in the 10 to 30 micron size bracket. Hence, on the average, a red cell spends a smaller proportion of its time within such vessels than does an average speck of plasma! It follows (Whitmore 1968) that the haematocrit within such vessels is substantially less than that for the body as a whole. Such haematocrit reduction may be important in an organ like the kidney which has a process to carry out (Chapter 10, § 11) upon the blood plasma alone. Indeed, in the kidney circulation, a further reduction of haematocrit is thought to occur (Pappenheimer

& Kinter 1956) by 'skimming' off plasma from near the walls of arterioles into the fine capillary network, which is then by-passed by the remaining cell-rich fluid.

The rest of this section is devoted to the fluid dynamics of 'why red cells shun walls'; that is, to the mechanism of axial concentration, not at all an easy subject. Several errors must be avoided: explanations relying upon inertial effects (including 'Magnus effect'), which are totally negligible as explained in § 1; explanations valid only for dilute solutions; and explanations that ignore the red cell's shape and its deformability.

Mason & Goldsmith (1969) give a good review of this difficult field, though they are obliged to admit that no proper account avoiding all these errors has been given. In dilute suspensions at low Reynolds number, disk-shaped bodies like red cells spend most of their time almost edge-on (that is, presenting a small angle of incidence) to a sheared flow, but they 'flip over' (rotate quickly so as to present the opposite edge to the flow) at regular intervals of the order of the reciprocal of the rate of shear. If they do this very near a wall, there is no actual collision but high pressures in a thin lubricating layer of plasma very near the wall (compare § 5 below) may bring about some repulsion of the cell away from it.

If the disks are rigid, then no axial concentration occurs except for this avoidance of immediate proximity to the wall. Nevertheless, even in dilute suspensions there is an additional mechanism of axial concentration if the particles are deformable. During the 'flip-over' process, the fluid stresses on the disk tend to stretch it in the direction of the distal edge when that is in the higher-velocity fluid but to compress it when that is in the lower-velocity fluid. Radial movements inwards in the former situation then exceed corresponding movements outwards in the latter, and so migration towards the axis can occur.

At normal haematocrits it is far from clear whether this mechanism can operate, since each cell is endlessly jostled by neighbouring cells. Most experimenters have observed some degree of axial concentration *additional* to the existence of a relatively cell-free layer a few microns thick associated with wall-shunning, but its cause remains uncertain. It seems possible that if the statistical effect on a red cell of many close interactions with neighbouring cells were better understood we would see that in an equilibrium situation (statistically speaking) the neighbouring cells on the higher-shear side (where we may suppose that mean cell rotation is faster or alternatively that 'flip-over' is more frequent) produce a relatively intenser effect that has to be balanced by a greater density of neighbouring cells on the lower-shear side. With this hypothetical remark we must leave the disputed question of the mechanism of axial concentration.

## 5. Lubrication problems in very narrow capillaries

We conclude Chapter 14 by discussing blood flow through passages so narrow that the shapes of red cells must be substantially deformed even to enter them. For this discussion we need to begin with a more detailed description than

before of the undeformed shape of the red blood cell (erythrocyte) and of its deformability.

The erythrocyte is a cell of an unusually simplified kind; it is without a nucleus, consists essentially of a more or less fluid mass of haemoglobin surrounded by a thin flexible cellular membrane, and appears to respond to mechanical stimuli in a completely passive manner (in contrast to phagocytes, larger 'white blood cells' with a nucleus and exhibiting a range of highly active responses to stimuli, including responses involving enormous changes of shape). The undeformed red cell, described earlier as a disk of diameter about 8 microns, is actually 'dimpled' on both sides near the centre of the disk; in other words, it is concave outwards in that region. Discussion of how the convexo-concave shape of the undeformed red cell comes about would be out of place here, but we may note that such a shape, being far easier to deform without change of volume or surface area than a purely convex shape, facilitates greatly the cell constrictions needed in the very narrow capillaries of the body.

Nevertheless, external stresses are required to deform the cell from its equilibrium shape. From the work of Rand & Burton (1964) and of Fung (1966) we can give 0.5 mb as the order of magnitude of stress required (in the form, for example, of differences in the external pressures applied to different parts of the surface) to produce really substantial degrees of deformation, comparable with the cell's dimensions. Note that this stress is far smaller than the excess pressures required to distend capillaries: even in the lung, the work of West, Glazier, Hughes & Maloney (1969) indicates the corresponding figure to be two orders of magnitude greater, while in systemic capillaries the discrepancy is bigger still (Fung 1966).

When blood flow in a very narrow systemic capillary is made visible under the microscope, we observe what these figures would lead us to expect: the red cells moving in single file along the tube accommodate to its dimensions not by distending the tube but by themselves being deformed. Large deformations are observed, and of several different kinds: particularly a 'parachute' shape in capillaries that are not so extremely narrow (this is an almost axisymmetrical configuration, curved convexly to the direction of motion), and an 'edge-on' configuration bent or even folded along a diameter to pass through still narrower tubes as 'crêpe suzette'! Similar results are obtained in glass tubes of similar diameters (Hochmuth, Marple & Sutera 1970). No doubt a variety of deformed erythrocyte shapes appear also in flow through the so-called 'capillary marsh' in the lung, described at the end of § 3, and we must suppose in all these cases that stresses of the order of 0.5 mb act to effect the larger deformations (and rather less for the smaller deformations).

Lighthill (1968, 1969) pointed out the relevance to these flows of 'hydrodynamic lubrication theory': a theory developed for engineering purposes of analysing how one surface may be enabled to slide over another while a substantial force is transmitted between them, through the presence of a very thin oil film. The high velocities in engineering applications are matched by the high viscosities of the oil so that we have a low-Reynolds-number situation (as in the microcirculation);

indeed, it is the stress distribution due to viscous action in the oil that transmits the force, and normally the oil-film thickness adjusts passively to the value required for such transmission. Only when the 'required value' is 'too small' in some sense (for example, to avoid asperities on the two surfaces from engaging with one another, or to avoid some kind of breakdown of the fluid properties of the oil) can a kind of 'seize-up' or suspension of sliding occur, with 'hydrodynamic lubrication' replaced by 'dry friction'.

Lighthill argued that the stress which the wall of a very narrow capillary passage needs to transmit to a red cell sliding past it so that the latter may be sufficiently deformed to pass through must, similarly, be transmitted by viscous action in a thin lubricating layer or plasma; that normally the plasma-film thickness must adjust passively to the value required for such transmission; and that suspension of sliding may occur when this required value is 'too small' in some sense. His work was taken considerably further by Fitz-Gerald (1969a, 1969b).

Most of the stress required to deform the cell is a radial stress, that must be transmitted by means of a distribution of excess pressure along the lubricating layer (a distribution associated with its thickness distribution), although some supplementary deformation may be provided by axial viscous stress (Fitz-Gerald 1969a) to cells passing in the 'parachute' configuration through tubes not so extremely narrow. It is of course purely viscous considerations that bring about a variation of pressure along the lubricating layer, with a gradient which simple dimensional arguments (supported by lubrication theory) suggest should be proportional to  $\mu_p U/h^2$ : that is, proportional to the plasma viscosity  $\mu_p$ , and to the velocity of sliding  $U$ , and inversely proportional to the square of the plasma-film thickness  $h$ . Actually, film thicknesses of a few tenths of a micron are thus predicted as necessary to cause large deformations of red cells, at typical capillary flow speeds of 0.5 mm/s, but the most significant conclusion from the theoretical observations is that the film thickness  $h$  should vary roughly as  $U^{1/2}$  (so that  $\mu_p U/h^2$  has a roughly constant effect).

Evidently a substantial viscous force resisting red-cell motion, and given by Newton's law of viscosity as proportional to  $\mu_p U/h$ , is to be expected wherever  $h$  is very small. This demands, if  $h$  is proportional to  $U^{1/2}$ , an excess pressure drop in a capillary also proportional to  $U^{1/2}$ . The nonlinearity of these predictions, that both plasma-film thickness and the excess pressure drop that it produces vary as  $U^{1/2}$ , makes them interesting: the latter pressure drop, for example, must exceed any normal linear component (proportional to  $U$ ) when  $U$  is small enough. A variety of detailed mathematical models developed by Lighthill and Fitz-Gerald show that these conclusions continue to be predicted by hydrodynamic lubrication theory under a wide range of assumptions (taking into account, for example, departures from axisymmetrical motion and also loss of fluid through gaps in the capillary wall), although none of the models allow for anything like an adequate detailed mechanics of the red-cell deformation.

Although ordinary microscopic studies of red-cell motion in capillaries do not give sufficient resolution to determine the thickness of the lubricating plasma film, Hochmuth, Marple & Sutera (1970) made some very careful observations of blood

flow in glass micro-capillaries from which plasma-film thicknesses were obtained in generally good accord with the above considerations. For example, their Figure 5 indicates that, in tubes of diameters (a) 4.5, (b) 6.7 and (c) 8 microns, the plasma film thickness at low velocities takes values the shape of whose dependence on  $U$  is close to  $U^{1/2}$  (multiplied, if  $U$  is measured in mm/s, by (a) 0.6, (b) 1.0 and (c) 1.3 microns respectively), but that the curve flattens off to a constant thickness above about  $U = 1$  mm/s (rather as if further increase of film thickness were prevented by a large increase in resistance by the cell to the extra deformations it would demand).

These results encourage one to study observations on peripheral resistance *in vivo* and look for a component of total pressure drop proportional to the square root of flow rate; such a component might represent an extra low-velocity effect of the blood's passage through very narrow capillaries, and be additional to the main term directly proportional to flow (possibly together, at very high flow rates, with a term proportional to flow squared associated with entry-region effects). Actually the literature of peripheral resistance is full of such data, in which the pressure-flow curve comes parabolically into the origin with a vertical tangent (Green, Rapela & Conrad 1963). In the past the data were interpreted differently (for example, in terms of increased dilatation of the microcirculation at increased perfusing pressure), but it seems more and more probable that enhanced pressure drop at low flows due to thinning of lubricating plasma films in the narrowest capillaries is the true explanation.

Thus, as tube diameter falls, the Fahraeus-Lindquist effect should be thought of as predicting only a local minimum in apparent viscosity (see § 4), followed by a rise to well above the normal whole-blood viscosity at small tube diameters. This same result was observed by Dintenfass (1968) for flow of blood of haematocrit 0.49 through a very narrow gap between parallel plates (see his Figure 7).

Pulmonary capillary flow (see the end of § 3) may be intermediate in character between flow in tubes and such flow in parallel plates. Certainly, many of its characteristics can be readily interpreted by lubrication theory, including stoppage of flow in particular pathways when hydrodynamic lubrication breaks down. Recently, Dr. J. B. West carried out computer simulations of capillary networks (unpublished) which indicated how the uneven filling of different pathways in the pulmonary circulation observed (see § 3) by him and his colleagues would be expected to be very marked on the assumption of the nonlinear resistance law here postulated.

Some confirmation that the pressure-flow relations in the lung are directly related to the deformability of red cells is provided by unpublished experiments of Dr. Reginald Greene. Using red cells that had been made much less readily deformable by a heat treatment, he showed that the square-root-of-flow component in the overall pressure drop, present already for normal cells, was considerably increased for the treated cells, consistently with the idea from lubrication theory that the less readily deformable cells would need a much higher value of  $\mu_p U/h^2$ .

In the systemic circulation, the nonlinear resistance law for very narrow capillaries may increase the effectiveness of vasomotor control of peripheral perfusion

(see § 2), essentially since an increase of arteriolar resistance reduces the flow rate and thus automatically increases the capillary component of resistance. Note also that, in systemic capillaries, some 'bunching' of red cells is observed. This may be interpreted (see Whitmore (1968)) in terms of slight variations in red-cell size or constrictability, leading to slight variations in plasma-film thickness and thus in rate of leakback of plasma through the film. Plasma then tends to accumulate behind any unusually small or easily constrictable cell.

Detailed models of the plasma-flow streamlines between successive red cells in a very narrow capillary have been obtained, and applied to estimation of the combined effects of convection and diffusion of particular solutes (see § 1) by Aroesty & Gross (1970) and by Fitz-Gerald (1971). They conclude that diffusion is dominant for dissolved gases like O<sub>2</sub> and CO<sub>2</sub> but that convection by a toroidal circulation in what Prothero & Burton (1961) called the 'bolus' of viscous fluid between two cells, as well as in the flow leaking back through the lubricating plasma film, may be quite significant for transfer of low-diffusivity macromolecules in the systemic circulation.

All this work on lubrication problems in the microcirculation is at a very early stage, but it seems to be a body of ideas likely to find increasing application to problems of real physiological interest as time goes on.

## REFERENCES

- AROESTY, J. & GROSS, J. 1970 *Convection and Diffusion in the Microcirculation*, Rep. RM-6214-NIH, The Rand Corporation, Santa Monica, California.
- CARO, C. G., FOLEY, T. H. & SUDLOW, M. F. 1970 J. Physiol. **207**, 257.
- DINTENFASS, L. 1968 *The viscosity of blood* . . . In *Hemorheology* (ed. Copley, A. L.), pp. 197–210, Pergamon, London.
- FAHRAEUS, R. & LINDQUIST, T. 1931 Amer. J. Physiol. **96**, 562.
- FARHI, L. E. 1966 *Ventilation-perfusion relationship and its role in alveolar gas exchange*. In *Advances in Respiratory Physiology* (ed. Caro, C. G.), pp. 148–197, Edward Arnold, London.
- FITZ-GERALD, J. M. 1969a Proc. Roy. Soc. B, **147**, 193.
- , 1969b J. Appl. Physiol. **27**, 921.
- , 1971 J. Fluid Mech. **51**, 463.
- FOWLER, K. T., WEST, J. B. & PAIN, M. C. F. 1966 Resp. Physiol. **1**, 88.
- FUNG, Y. C. 1966 Federation Proc. **25**, 1761.
- FUNG, Y. C. & SOBIN, S. S. 1968 Ibid. **27**, 578.
- GREEN, H. D., RAPELA, C. E. & CONRAD, M. C. 1963 *Resistance and capacitance phenomena in terminal vascular beds*. In *Handbook of Physiology, Section 2: Circulation* (ed. Hamilton, W. F. & Dow, P.), pp. 935–960, American Physiological Society, Washington.
- GUYTON, A. C. 1966 *Textbook of Medical Physiology*, W. D. Saunders, London.
- HOCHMUTH, R. M., MARPLE, R. N. & SUTERA, S. P. 1970 Microvascular Res. **2**, 409.
- KRAEMER, K. 1967 Arch. f. Kreislaufforschung **52**, 79.
- LIGHTHILL, M. J. 1968 J. Fluid Mech. **34**, 113.
- , 1969 *Motion in narrow capillaries from the standpoint of lubrication theory*. In *Circulatory and Respiratory Mass Transport* (ed. Wolstenholme, G. E. W. & Knight, J.), pp. 85–104, J. & A. Churchill, London.
- MASON, S. G. & GOLDSMITH, H. L. 1969 *The flow behaviour of particulate suspensions*. In *Circulatory and Respiratory Mass Transport* (ed. Wolstenholme, G. E. W. & Knight, J.), pp. 105–129, J. & A. Churchill, London.

- PAPPENHEIMER, J. R. & KINTER, W. B. 1956 Amer. J. Physiol. **185**, 377.
- PATTLE, R. E. 1966 *Surface tension and the lining of the lung alveoli*. In *Advances in Respiratory Physiology* (ed. Caro, G. G.), pp. 83-105. Edward Arnold, London.
- PROTHERO, J. & BURTON, A. C. 1961 Biophys. J. **1**, 565.
- RAND, R. P. & BURTON, A. C. 1964 Ibid. **4**, 115.
- WEST, J. B. 1966 *Regional differences in blood flow and ventilation in the lung*. In *Advances in Respiratory Physiology* (ed. Caro, C. G.), pp. 198-254. Edward Arnold, London.
- WEST, J. B., GLAZIER, J. B., HUGHES, J. M. B. & MALONEY, J. E. 1969 *Pulmonary capillary flow, diffusion, ventilation and gas exchange*. In *Circulatory and Respiratory Mass Transport* (ed. Wolstenholme, G. E. W. & Knight, J.), pp. 256-276. J. & A. Churchill, London.
- WHITMORE, R. L. 1968 *The dynamics of blood flow in capillaries*. In *Hemorheology* (ed. Copley, A. L.), pp. 77-87. Pergamon, London.