AD-A270 052

KFW1664-HN-02
DAJA45-92-M-0289

Theoretical and Applied Mechanics 1992

edited by

S.R. Bodner
J. Singer
A. Solan
Z. Hashin

IUTAM

DTIC
ELECTE
SEP 30 1993
S D
E

[redacted]STAT
Approved for public rele[redacted]
Distribut[redacted]

North-Holland

THEORETICAL AND APPLIED MECHANICS 1992

Accesion For	
NTIS CRA&I	X
DTIC TAB	☐
Unannounced	☐
Justification	
By	
Distribution /	
Availability Codes	
Dist	Avail and / or Special
A-1	

DTIC ELECTE SEP 3 0 1993 S E D

DTIC QUALITY INSPECTED 1

[redacted] STATEMENT A
Approved for public release;
Distribution Unlimited

INTERNATIONAL UNION OF THEORETICAL
AND APPLIED MECHANICS

AD NUMBER	DATE	DTIC ACCESSION NOTICE
1. REPORT IDENTIFYING INFORMATION		REQUESTER:
A. ORIGINATING AGENCY University of Liverpool England UK		1. Put your mailing address on reverse of form. 2. Complete items 1 and 2. 3. Attach form to reports mailed to DTIC. 4. Use unclassified information only. 5. Do not order document for 6 to 8 weeks.
B. REPORT TITLE AND/OR NUMBER Structural Crashworthiness & Failure		
C. MONITOR REPORT NUMBER R&D 6917-AN-02		
D. PREPARED UNDER CONTRACT NUMBER DAJA45-93M-0150		
2. DISTRIBUTION STATEMENT Approved for Public Release - Distribution Unlimited (FP)		DTIC: 1. Assign AD Number. 2. Return to requester.

DTIC Form 50 Mar 91 PREVIOUS EDITIONS ARE OBSOLETE

THEORETICAL AND APPLIED MECHANICS
1992

Proceedings of the XVIIIth International Congress of Theoretical and Applied Mechanics, Haifa, Israel, 22-28 August 1992

Edited by

Sol R. BODNER
Josef SINGER
Alexander SOLAN

Technion - Israel Institute of Technology
Haifa, Israel

and

Zvi HASHIN

Tel Aviv University
Tel Aviv, Israel

93-22612

1993
ELSEVIER
AMSTERDAM • LONDON • NEW YORK • TOKYO

ELSEVIER SCIENCE PUBLISHERS B.V.
Sara Burgerhartstraat 25
P.O. Box 211, 1000 AE Amsterdam, The Netherlands

ISBN: 0 444 88889 6

© 1993 IUTAM. All rights reserved.

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior written permission of the publisher, Elsevier Science Publishers B.V., Copyright & Permissions Department, P.O. Box 521, 1000 AM Amsterdam, The Netherlands.

Special regulations for readers in the U.S.A. - This publication has been registered with the Copyright Clearance Center Inc. (CCC), Salem, Massachusetts. Information can be obtained from the CCC about conditions under which photocopies of parts of this publication may be made in the U.S.A. All other copyright questions, including photocopying outside of the U.S.A., should be reffered to the publisher, Elsevier Science Publishers B.V., unless otherwise specified.

No responsibilty is assumed by the publisher or by IUTAM for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein.

This book is printed on acid-free paper.

PREFACE

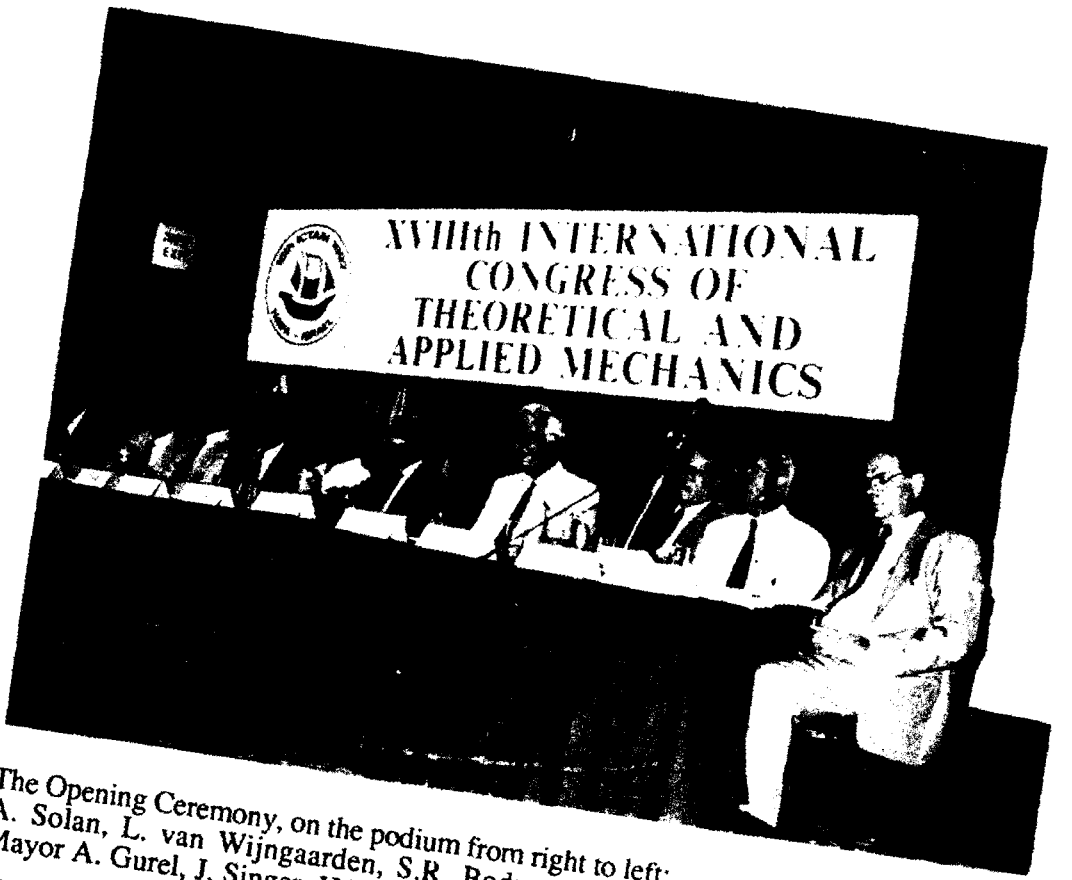
This book contains the Proceedings of the XVIIIth International Congress of Theoretical and Applied Mechanics, held at the Technion, Israel Institute of Technology, Haifa, August 22-28, 1992. The Congress was held under the auspices of The International Union of Theoretical and Applied Mechanics (IUTAM) by invitation of The Israel Society for Theoretical and Applied Mechanics and Technion, Israel Institute of Technology and under the sponsorship of the Israel Academy of Sciences and Humanities.

The full text of the two General Lectures, of introductory lectures of the three minisymposia and of sectional lectures, according to the list on pages xi, xii, are included in this volume. The contributed papers are listed by author and title; most of them will be published in appropriate scientific journals.

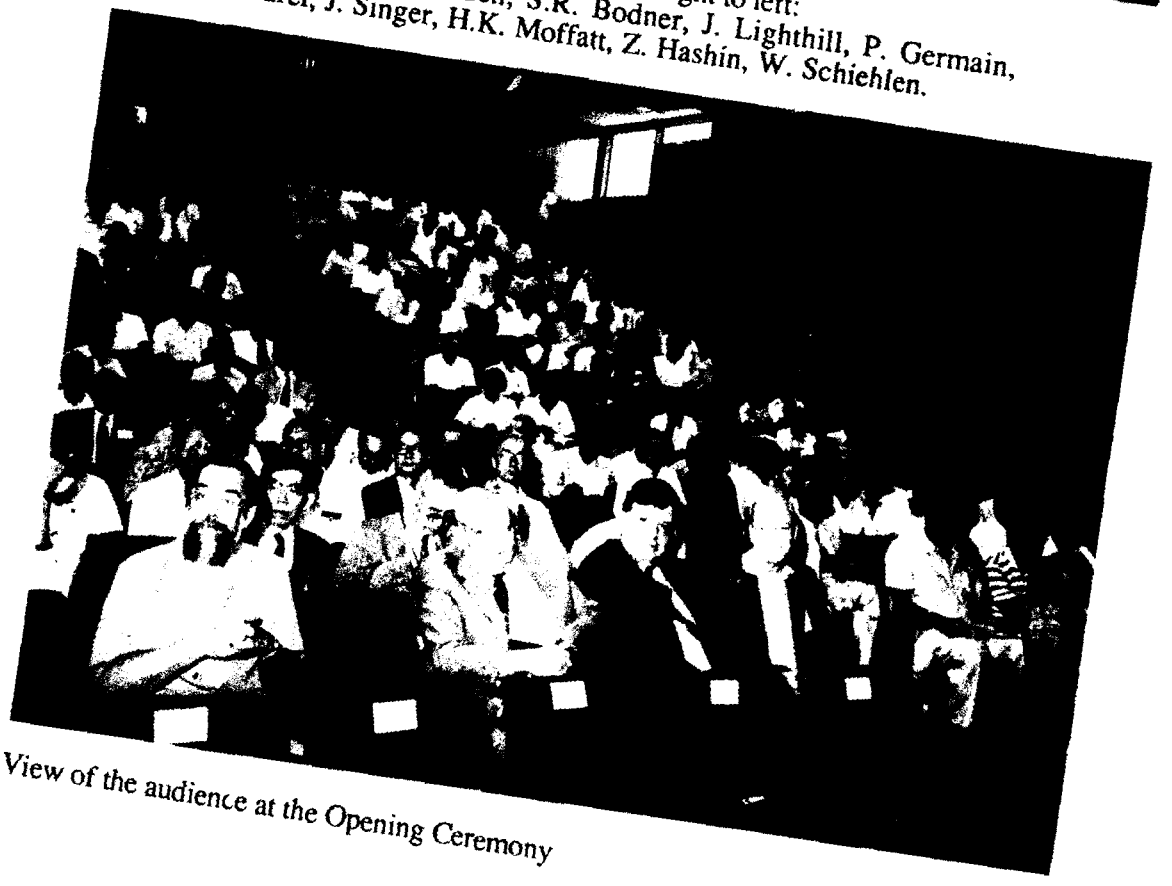
The publication of these Proceedings has been handled promptly and very capably by Elsevier Science Publishers B.V. and their editors to whom we are very grateful.

Josef Singer Sol R. Bodner Alexander Solan Zvi Hashin

Haifa
December 1992



The Opening Ceremony, on the podium from right to left:
A. Solan, L. van Wijngaarden, S.R. Bodner, J. Lighthill, P. Germain,
Mayor A. Gurel, J. Singer, H.K. Moffatt, Z. Hashin, W. Schiehlen.



View of the audience at the Opening Ceremony

Professor J. Singer, Chairman of the Local Organizing Committee

Professor P. Germain, President of IUTAM

H.K. Moffatt and M. Kiya

L. van Wijngaarden, Z. Hashin and M. Sayir

Mrs. N. Bodner, G.I. Barenblatt, S.R. Bodner

J.T. Stuart and T. Tatsumi

SPONSORING ORGANIZATIONS AND COMPANIES

The following academic, public, and industrial organizations have provided financial support:

- The Israel Academy of Sciences and Humanities
- Technion, Israel Institute of Technology
- Ben Gurion University of the Negev
  and the Pearlstone Center for Aeronautical Engineering Studies
- Tel Aviv University
- Zurich Chapter of the Swiss Technion Society

- Israel Ministry of Defense, (MAFAT)
- Israel Ministry of Science and Technology
- Israel Ministry of Trade and Industry
- Israel Ministry of Tourism
- Haifa Municipality

- European Research Office, United States Army

- American Israeli Paper Mills, Ltd.
- El-Al Israel Airlines
- Israel Electric Corporation, Ltd.
- I.B.M. Israel
- Israel Military Industries
- Klil Industries, Ltd.
- Ormat Turbines (1965) Ltd.
- Israel Aircraft Industries
- Rafael, Israel Armament Development Authority
- Oil Refineries, Ltd.
- Paz Oil Co., Ltd.

Support for the

MINISYMPOSIUM ON INSTABILITIES IN SOLID AND STRUCTURAL MECHANICS

was provided by the

- European Office of Aerospace Research and Development, United States Air Force

CONTENTS

Preface	v
Sponsoring Organizations and Companies	x
Congress Committees	xiii
List of Participants	xiv
Report on the Congress	xx

OPENING AND CLOSING LECTURES

A. ROSHKO : Instability and turbulence in shear flows	1
G.I. BARENBLATT : Micromechanics of fracture	25

INTRODUCTORY LECTURES OF MINISYMPOSIA

"Instabilities in Solid and Structural Mechanics"

R. ABEYARATNE : Material instabilities and phase transitions in thermoelasticity	53
S. KYRIAKIDES : Propagating instabilities in structures (Summary)	79
Y. TOMITA : Computational approaches to plastic instability in solid mechanics	81

"Sea Surface Mechanics and Air-Sea Interaction"

W.K. MELVILLE : The role of wave breaking in air-sea interaction	99
O.M. PHILLIPS : Extreme waves and breaking wavelets	119
P.G. SAFFMAN : Effect of wind and water shear on wave instabilities	133

'Biomechanics"

S.C. COWIN : Nature's structural engineering of bone on a daily basis	145
R.D. KAMM : Liquid layer dynamics in pulmonary airways	161
R.McN. ALEXANDER : Energy-saving mechanisms in animal movement	177

SECTIONAL LECTURES

H.H. BAU : Controlling chaotic convection 187

C.R. CALLADINE : Application of structural mechanics to biological systems 205

Y. COUDER : Viscous fingering as a pattern forming system (Summary) 221

G. GRIMVALL : Mechanics in sport 225

T. KAMBE : Aerodynamic sound associated with vortex motions: observation and computation 239

A. LIBAI : Nonlinear membrane theory 257

M.E. McINTYRE : On the role of wave propagation and wave breaking in atmosphere-ocean dynamics 281

J.B. MARTIN : Computational aspects of integration along the path of loading in elastic-plastic problems 305

S. MURAKAMI : Constitutive modeling and analysis of creep, damage, and creep crack growth under neutron irradiation 323

Q.S. NGUYEN : Stability and bifurcation in dissipative media 339

A. PROSPERETTI : Bubble mechanics: luminescence, noise, and two-phase flow 355

M.B. SAYIR : Wave propagation in non-isotropic structures 371

K.R. SREENIVASAN and G. STOLOVITZKY : Self-similar multiplier distributions and multiplicative models for energy dissipation in high-Reynolds-number turbulence 395

A.A. VAN STEENHOVEN, J.D. JANSSEN, and D.H. VAN CAMPEN : Cardiovascular fluid mechanics 407

J. ZIEREP : Trends in transonic research 423

CONTRIBUTED PAPERS

List of Contributed Papers presented at the Congress 437

CONGRESS COMMITTEE OF IUTAM

P. Germain* (France), Chairman
H.K. Moffatt* (UK), Secretary

J.D. Achenbach (USA)
A. Acrivos* (USA)
L. Bevilacqua (Brazil)
B.A. Boley (USA)
C.R. Calladine (UK)
C. Cercignani (Italy)
G.G. Chernyi (Russia)
I.F. Collins (New Zealand)
A. Crespo (Spain)
D.C. Drucker (USA)
Z. Hashin (Israel)
M.A. Hayes (Ireland)
N.J. Hoff (USA)
J. Hult (Sweden)
I.Imai (Japan)
A. Yu. Ishlinsky (Russia)
S. Kaliszky* (Hungary)
Y.H. Ku (USA)
S. Leibovich (USA)
J. Lemaitre (France)
M.J. Lighthill* (UK)
Z. Mroz (Poland)
E.-A. Müller (Germany)
R. Narasimha (India)
F.I. Niordson (Denmark)
N. Olhoff* (Denmark)
J.R. Philip (Australia)
W. Schiehlen (Germany)
B. Tabarrok (Canada)
L. van Wijngaarden (Netherlands)
R. Wang (China)
Zheng Zhemin (China)
F. Ziegler (Austria)

* Members of Executive Committee (1988-1992).

LOCAL ORGANIZING COMMITTEE

J. Singer*, Chairman
S.R. Bodner*, Co-Chairman
A. Solan*, Secretary

S. Abarbanel
D. Givoli
H. Elata
Z. Hashin*
D. Shilkrut
A. Silberberg
A. Yarin

* Members of Executive Committee

LIST OF PARTICIPANTS

AUSTRALIA (4 participants)

Asokanthan S.F.
Borgas M.S.
Karihaloo B.
Muhlhaus H.B.

AUSTRIA (3 participants)

Kluwick A.
Perktold K.
Ziegler F.

AZERBAIJAN (1 participant)

Iskenderov I.A.

BELGIUM (5 participants)

Bao H.
Boucher S.
Nielens H.
Verdonck P.
Verlinden O.

BRAZIL (2 participants)

Musafir R.E.
Zachariadis D.C.

BULGARIA (5 participants)

Bontcheva N.
Dabnichki P.A.
Karagiozova D.
Markov K.
Zapryanov Z.

CANADA (18 participants)

Bourassa P.A.
Cohen H.
Davies H.G.
Dickstein P.
Dost S.
Dziedzic M.
Floryan J.M.
Graham G.A.C.
Haddow J.B.
Han R.P.S.
Leutheusser H.J.
Mioduchowski A.
Rimrott F.P.J.
Smith S.H.
Szyszkowski W.
Tabarrok B.
Wegner J.L.
Yeh K.Y.

CHINA (BEIJING) (8 participants)

Guo Z.H.
He Y.
Hwang K.C.
Wang G.T.
Wang R.
Zhang R.J.
Zheng Z.
Zhou X.C.

CHINA (TAIPEI) (6 participants)

Chang C.C.
Chen F.
Chu C.C.
Pao Y.H.
Su F.C.
Wu E.

CZECHOSLOVAKIA (3 participants)

Brilla J.
Horacek J.
Klimes F.

DENMARK (9 participants)

Byskov E.
Jensen H.M.
Larsen P.S.
Mikkelsen L.P.
Niordson F.I.
Olhoff N.
Pedersen P.T.
Thomsen O.T.
Tvergaard V.

FINLAND (1 participant)

Holopainen P.

FRANCE (39 participants)

Berest P.
Berthaud Y.
Blanze C.
Boehler J.P.
Caulliez G.
Charru F.
Chaskalovic J.
Cochelin B.

Couder Y.
Dias F.
Djeridi H.
Drochon A.
Etay J.
Fressengeas C.
George J.
Germain P.
Grandidier J.C.
Iooss G.
Jullien J.F.
Klepaczko J.R.
Lagarde A.
Lemaitre J.
Lespinard G.
Magnaudet J.
Maigre H.
Maugin G.A.
Molinari A.
Nguyen Q.S.
Nouailhas D.
Pelissier R.
Potier-Ferry M.
Resch F.J.
Reynier M.
Rittel D.
Roseau M.
Rougee P.
Sulem J.
Thebaud F.
Waldura H.

GERMANY (36 participants)

Altenbach H.
Altenbach J.
Buhler K.
Chen X.N.
De Boer R.
Dinckelacker A.
Faciu C.
Furta S.
Gersten K.
Gross D.
Kazimierczyk P.
Koenig M.
Kowalewski T.A.
Krause E.
Kuhn G.
Labisch F.K.
Le K.C.
Lenz J.
Meyer L.W.
Mitra N.K.
Muller E.A.
Noack B.R.
Ohle F.
Olszok T.
Pfeiffer F.
Pingel T.
Rozvany G.
Schiehlen W.
Schirm W.
Schumacher A.
Soeller C.
Stein E.
Stumpf H.
Wierzbicki T.
Zhang C.
Zierep J.

GREECE (4 participants)

Georgantopoulos G.A.
Kounadis A.N.
Sotiropoulos D.A.
Vardoulakis I.

HONG KONG (1 participant)

Hui W.H.

HUNGARY (2 participants)

Kaliszky S.
Tarnai T.

IRELAND (1 participant)

Hayes M.A.

ISRAEL (85 participants)

Adan M.
Agnon Y.
Altus E.
Arcan M.
Ayzenberg M.
Bar-On E.
Barron A.
Baruch M.
Ben David D.
Ben Haim Y.
Benveniste Y.
Bergman R.
Bershadskii A.
Bodner S.R.
Branover H.
Broday D.
Burde G.I.
Chiskis A.
Dmitriuk W.
Drozdov A.D.
Durban D.
Elad D.
Elata D.

Elida D.
Fang D.
Frankel I.
Galper A.
Gandelsman M.L.
Gershtein M.S.
Ginzberg I.
Givoli D.
Glozman M.
Gringauz M.
Gurevich B.
Hashin Z.
Hetsroni G.
Khain A.
Khen R.
Kimmel E.
Lanir Y.
Lesin S.
Levich E.
Libai A.
Litovsky Y.
Markus S.
Mickulinsky M.
Miloh T.
Mizrahi J.
Moshaiov A.
Nekhamkina O.
Nezlina Y.
Nir A.
Parnes R.
Perlman E.
Partom Y.
Rom J.
Rosen A.
Rotem A.
Rubin M.B.
Ryvkin M.
Sabag M.
Salganik R.L.
Shalman E.
Shapiro M.
Sheinman I.
Shemer L.
Shilkrut D.
Shoavi E.
Shtark A.
Shtilman L.
Singer J.
Solan A.
Spivakovsky V.
Stavsky Y.
Stiassnie M.
Tirosh J.
Tsitverblit N.
Vigdergauz S.
Vilenkin B.
Weinstein M.
Weiss M.P.
Wiener Z.
Wygnanski I.
Yarin A.
Zvirin Y.

ITALY (11 participants)

Amadio C.
Bigoni D.
Cercignani C.
Franzese P.
Guglielmino E.
Martelli F.
Nappi A.
Perego U.
Ruta G.C.
Tatone A.
Zannetti L.

JAPAN (22 participants)

Amari T.
Ashida F.
Fukuyu A.
Hasimoto H.
Horii H.
Imai I.
Kambe T.
Kawata K.
Kitagawa H.
Kiya M.
Kobayashi S.
Kondo K.
Kumagai T.
Murakami S.
Sakao F.
Sumi Y.
Takahashi K.
Tatsumi T.
Tomita Y.
Uetani K.
Yamamoto Y.
Yoshida H.

KOREA (3 participants)

Jeong J.J.
Pak C.H.
Youm Y.

MEXICO (1 participant)

Sabina F.J.

NETHERLANDS (21 participants)

Arbocz J.
Bakker P.G.

Besseling J.F.
Braat G.F.M.
Coene R.
Dieterman H.A.
Huyghe J.
Janssen J.D.
Kouhia R.J.
Leroy Y.M.
Meijaard J.P.
Meijers P.
Menken C.M.
Shlyapobersky J.
Sluys L.
Van Beek P.
Van Campen D.H.
van der Giessen E.
Van Steenhoven A.A.
van Wijngaarden L.
Zhang G.Q.

NEW ZEALAND (1 participant)

Collins I.F.

NORWAY (1 participant)

Tyvand P.A.

POLAND (21 participants)

Bauer J.
Dolinski K.
Duszek-Perzyna M.K.
Gawinecki J.A.
Golos K.
Grabacki J.K.
Gutkowski W.
Kleiber M.
Kurnik W.
Mroz Z.
Nowak Z.
Osinski Z.
Pecherski R.B.
Peradzynski Z.
Perzyna P.
Petryk H.
Rakowski J.
Skrzypek J.J.
Szafranski W.
Tylikowski A.
Zorski H.

PORTUGAL (5 participants)

Arantes e Oliveira E.R.
Camotim D.R.Z.
Faria L.
Martins J.A.C.
Trabucho L.

ROMANIA (1 participant)

Cristescu N.D.

RUSSIA (32 participants)

Barenblatt G.I.
Belsky V.G.
Blekhman I.I.
Borodich F.M.
Chashechkin Y.D.
Chernyi G.C.
Elkin A.I.
Fabrikant A.L.
Freidin A.
Goldenveizer A.L.
Goldstein R.
Gregoryan S.S.
Ivanov V.
Kanaun S.K.
Lyubimov G.A.
Makarov S.O.
Manevitch L.I.
Matasov A.
Mikhailov G.K.
Mirkin M.
Nezlin M.V.
Nuller B.
Obraztsov I.F.
Ryaboy V.M.
Sedov L.I.
Shukhman I.
Slepyan L.I.
Smirnov B.I.
Stein A.A.
Talipova T.
Yanovsky Y.G.
Yumasheva M.A.

SOUTH AFRICA (2 participants)

Jekot T.
Martin J.B.

SWEDEN (7 participants)

Essen H.
Grimvall G.
Hult J.
Larsson P.L.
Neumeister J.M.
Sjodin B.
Storakers B.

SWITZERLAND (10 participants)

Cuche D.
Dual J.
Hazanov S.
Herrmann G.
Juengling D.F.
Meyer Matievic M.
Pellegrini O.
Sayir M.B.
Spirig T.H.
Staudenmann M.

TRINIDAD (1 participant)

Ramkissoon H.

TURKEY (1 particpant)

Kaykayoglu C.R.

UK (35 participants)

Alexander R.McN.
Brandt R.
Calladine C.R.
Carpenter P.W.
Cowley S.J.
Crighton D.G.
Dunne F.P.E.
Earles S.W.E.
Fleck N.A.
Guest S.D.
Hardy S.J.
Haughton D.M.
Healey J.J.
Lighthill J.
Lucey A.D.
McIntyre M.E.
Moffatt H.K.
Nagata M.
Ogden R.W.
Pedley T.J.
Pellegrino S.
Reid S.R.
Ricca R.L.
Richards K.J.
Rogers T.G.
Shi J.
Slaughter W.S.
Soldatos K.
Spence D.
Stronge W.J.
Stuart J.T.
Sutcliffe M.
Ursell F.
Vassiliev D.
Willis J.R.

UKRAINE (4 participants)

Guz A.N.
Levitas V.I.
Rassokha A.
Shupikov A.

USA (107 participants)

Abeyaratne R.
Achenbach J.D.
Acrivos A.
Agarwal R.K.
Aifantis E.C.
Akylas T.R.
Bammann D.J.
Batra R.C.
Bau H.H.
Bazant Z.P.
Berdichevsky V.
Berkooz G.
Bert C.W.
Boley B.A.
Brenner H.
Brock L.
Brodsky N.
Budiansky B.
Chaudhuri R.
Cherepanov G.
Cherkaev A.V.
Christensen R.M.
Chudnovsky A.
Clifton R.J.
Cohen B.
Cowin S.C.
Crandall S.H.
Dagan Z.
Drucker D.C.
Dvorak G.J.
Dvorkin J.
Elishakoff I.
Fernando H.J.
Finkelstein I.
Fish J.
Folias T.
Freund L.B.
Gao H.
Glenn L.
Goddard J.D.
Graham A.
Halpern D.
Hanagud S.V.
Haythornthwaite R.M.
Hetnarski R.B.
Hoff N.J.
Holmes P.J.
Huang T.C.
Hulbert G.M.

Hutchinson J.W.
Iacob A.
Jasiuk I.
Joseph D.
Kachanov M.
Kamm R.D.
Kapoor B.
Keer L.M.
Kestin J.
Knowles J.
Kohn R.V.
Krajcinovic D.
Kraynik A.M.
Kyriakides S.
Leibovich S.
Ling F.
Lurie K.
Mac Sithigh G.P.
Maity A.K.
Majerus J.
Mal A.K.
Markenscoff X.
Mauri R.
Meirovitch L.
Melville W.K.
Mendelsohn D.
Metzner A.W.K.
Mondy L.
Needleman A.
Ostrach S.
Phillips O.M.
Ponte-Castaneda P.
Prosperetti A.
Reshotko E.
Roshko A.
Rubinstein A.A.
Rumschitzki D.
Sackman J.L.
Saffman P.G.
Sen M.
Shaw S.
Shillor M.
Shtern V.
Sreenivasan K.R.
Steele C.R.
Steif P.S.
Triantafyllou G.
Truskinovsky L.
Vakakis A.F.
Vaynblat D.
Voloshin A.S.
Wei K.
Weinbaum S.
Weitsman Y.J.
Weng G.J.
Wnuk M.P.
Yang Z.
Yu Y.Y.

VIETNAM (1 participant)

Anh N.D.

YUGOSLAVIA (5 participants)

Cveticanin L.
Dzodzo M.B.
Hedrih (Stevanovic) K.
Ruzic D.
Vujicic V.A.

REPORT ON THE CONGRESS

Josef Singer

The decision to accept the invitation from Israel to hold the XVIIIth International Congress in Haifa was taken by the Congress Committee of IUTAM during its meeting in Grenoble in August 1988. It was decided that the format of the Congress would follow the one adopted for the recent successful Congresses in Lyngby and Grenoble. The Congress would cover the entire field of mechanics with special emphasis on three selected topics to constitute the so-called minisymposia.

The Congress Committee selected the following three topics for these minisymposia:

1. Instabilities in Solid and Structural Mechanics.
2. Sea Surface Mechanics and Air-Sea Interaction.
3. Biomechanics.

The Congress Committee also selected two general lecturers: Professor Anatol Roshko (USA) to present the Opening Lecture and Professor G.I. Barenblatt (Russia) to present the Closing Lecture; as well as 15 Sectional Lectures. The chairmen of the Minisymposia further selected 9 Introductory Lectures for their symposia.

As in Lyngby and Grenoble, the contributed papers were presented in parallel sessions, either as 25 minute lectures or in poster-sessions which were scheduled separately from the lectures. There were lively discussions both after the lectures and in the poster-sessions, where the second half of the sessions were devoted to general discussions guided by the chairmen.

The Opening Session of the Congress was held in the Churchill Auditorium of the Technion at 10 o'clock on Sunday, 23rd August, 1992. The Session was opened by Professor Josef Singer, Chairman of the Local Organizing Committee of *ICTAM* '92, with the following words:

"Distinguished Members of the Dais, Ladies and Gentlemen,

It gives me great pleasure to open the 18th International Congress of Theoretical and Applied Mechanics and to greet this outstanding gathering of scientists and engineers which will ensure the success of the Congress.

Welcome to Israel, to Haifa and to the Technion, Israel Institute of Technology.

Let me also greet you in Hebrew (welcome to participants).

I think you will understand me better if I do not continue in Hebrew. Maybe by the time we come to the Closing Session, your Hebrew will be good enough to permit me to address you all in our language.

So let me introduce you to the Dais:

I will start at the far end of the Head Table. There are the Deans of the two Technion Faculties which are most active in Mechanics:

Professor Blech, Acting Dean of the Faculty of Mechanical Engineering, and Professor Shinar, Dean of the Faculty of Aerospace Engineering.

Next, the Secretary-General of IUTAM, Professor Schielen of Stuttgart University, who is entrusted with the many IUTAM Symposia and other IUTAM activities.

To his left, the President of the Israel Society for Theoretical and Applied Mechanics, Professor Hashin of the Faculty of Engineering of Tel Aviv University, an active member of our Organizing Committee.

Next, the Secretary of the IUTAM Congress Committee, Professor Moffatt of Cambridge University, whose guidance and tireless efforts have been essential to the success of the scientific program and the preparations of the Congress.

To his left, the Mayor of our beautiful city, Haifa, Mr. Gurel, who graciously supports our meeting and many other international activities.

Then our IUTAM President, Professor Germain, of the French Academy of Science, whom many of you also remember as the Organizer of the successful Congress in Grenoble in 1988.

To his left, Professor Paul Singer, Senior Vice President of Technion, who will soon bring to you the greetings of the Technion President.

Then the IUTAM Vice President, Sir James Lighthill, the past IUTAM President and former Vice Chancellor of University College London.

To his left, Professor Bodner of our Faculty of Mechanical Engineering, the Co-Chairman of our Local Organizing Committee, who gave his time and knowledge untiringly and thus contributed so much to our Congress.

Then the Treasurer of IUTAM, Professor van Wijngaarden of the University of Twente in The Netherlands, whose importance is self-evident.

And last, but certainly not least, the wonderful Secretary of our Local Organizing Committee, Professor Alex Solan, Technion Vice President for Academic Affairs, with whom most of you have already corresponded and without whose outstanding and never ending work we could not have carried out the preparations for the Congress.

ICTAM Participants, as I look around me I see many friends, the top scientists of the international mechanics community, the international mechanics family. I am sure that our deliberations will not only advance our field, but will also weave many new collaborations and friendships.

For the benefit of the delegates who are new to ICTAM, I would like to ask some of the former IUTAM Presidents to please rise: Professor Frithiof Niordson of Denmark, and Professor Daniel Drucker of U.S.A.

To finish the introductions, it gives me great pleasure to see here among us, my former teacher and good friend, Professor Nicholas Hoff of Stanford University, who taught me the meaning of theoretical and applied mechanics and much more, and whom many of you will remember as the President and Organizer of our 1968 Congress at Stanford. Please rise Nicholas.

Friends, I may have taken up some time with these introductions, but I remember that at the first ICTAM I attended in Stresa in 1960, seeing the great names in the flesh was, indeed, a great experience!

Ladies and Gentlemen, we Israelis are honored and happy to host the 18th International Congress of Theoretical and Applied Mechanics in our old-new Land.

Though the ancient Israelis are better known for their Monotheism than for their achievements in mechanics, one finds that they tried their hand in some large scale experiments in mechanics.

For example, in the Book of **Exodus** we hear that in their flight from Egypt, the Children of Israel (with some assistance from the Almighty) experimented with sea-air interactions, for we read in Exodus 14/21,22:

"And Moses stretched out his hand over the sea and the Lord caused the sea to go back by a strong east wind ..... and made the sea dry land and the waters were divided".

"And the Children of Israel went into the midst of the sea upon the dry ground, and the waters were a wall unto them on their right and on their left".

Would this not fit into our second Minisymposium?!

Or, in the Book of **Joshua**, we hear of an early experiment in the dynamics of sound waves, for we read in Joshua 6/5:

"And it shall come to pass, that when they make a long blast with a ram's horn, and when ye hear the sound of the trumpet, all the people shall shout, and the wall of the city shall fall down flat".

So, there was some mechanics activity here 3,500 years ago!

In modern times, Israel has a small but very active mechanics community, whose presence and achievements are known internationally, and whose participation in IUTAM Congresses and Symposia has been very significant.

At the Technion we were fortunate to have had Professor Marcus Reiner, well known as one of the Fathers of Rheology, active here from 1947-1976. He built a thriving department of mechanics, that later diffused into other departments.

In 1961 Professor Reiner organized an IUTAM Symposium on "Secondary Effects in Elasticity, Plasticity and Fluid Dynamics", in Haifa (by the way, my wife worked with him on the organization of that symposium) a symposium which was the occasion of the first visit to Israel for some of you.

In 1985 Professors Bodner and Hashin organized another IUTAM Symposium on "Mechanics of Damage and Fatigue", at Technion, which was also very successful.

Hence, an ICTAM at Haifa seem to follow logically. Well, some of us in the audience have aged a bit since the first time Israel offered to host ICTAM many years ago, but I am happy that we all made it finally!

We have, I believe, an excellent scientific program thanks to you the contributors and to the excellent work of our International Papers Committee. As you know, the choice was difficult and the 600 papers chosen out of 1,183 submitted originated in 48 different countries. However, as we had many and continuous changes, even in the number of countries, the program you have could be finalized only two weeks ago.

The flags on the platform represent all the countries from which papers were accepted. The flags indicate origins of papers, but we are all here as individuals, contributing to theoretical and applied mechanics. Individuals whose friendship and collaboration will be reinforced by this Congress and who will assist in bringing our nations closer and improve their relations.

Before closing, I would like to thank the IUTAM Congress Committee, and in particular its Executive Committee, the IUTAM Bureau, the International Papers Committee, as well as the different National Committees. I will express my gratitude to my local collaborators, who worked so hard to prepare the Congress, in more detail at the Closing Session. For now, only

sincere thanks to you, my colleagues of the Local Organizing Committee and to the devoted staff of the Technion and Kenes teams who will also look after us in the coming days.

I would also like to thank the Israel Academy of Sciences and Humanities, the various Israeli government ministries, our municipality, and the Haifa Tourist Board, the universities and industrial companies, whose names appear on the program and, in particular, the U.S. Air Force European Office of Aerospace Research and Development, and the U.S. Army European Research Office, for their generous support.

As you know, the Israel Academy is one of the sponsors of our Congress. Its President, Professor Joshua Jortner, who is abroad, sent us his greetings which I would like to read to you:

"The Israel Academy recognizes the important contribution international meetings make to the advancement and excellence of research, by providing a forum for direct contact between scientists and engineers engaged in high quality scientific endeavor.

I am confident that the 18th International Conference in Theoretical and Applied Mechanics will contribute to the advancement and the enhanced cooperation in these exciting and important fields of scientific and technological research. The integration of science and technology which is so well reflected in your Conference is of prime importance for Israel and for the international community at large.

Please convey my compliments, on behalf of the Israel Academy of Sciences and Humanities, to all the participants.

I wish you a fruitful and stimulating meeting."

Ladies and Gentlemen, I concur with Professor Jortner, and wish us all, delegates and accompanying persons, a fruitful and enjoyable Congress, and a very pleasant stay in Israel. Thank you."

Professor Singer then called on Professor Solan, the Secretary of the Local Organizing Committee, to introduce the speakers who followed.

Mayor Gurel then brought the greetings of the Municipality of Haifa to the Congress participants and wished them successful deliberations and a pleasant stay in Haifa and Israel.

Since the President of the Technion was abroad, the Senior Vice President, President Paul Singer, greeted the assembly with the following words:

"Distinguished Guests and Friends,

It is my privilege and honor to welcome you, participants of the International Congress of Theoretical and Applied Mechanics to the Technion - Israel Institute of Technology, with the traditional Hebrew greeting - "Blessed be those who came". - This Congress is the XVIII th in a series which started in Delft, Netherlands and carries with it by now a most famous tradition. We are therefore *very proud* and consider it a great distinction, that for the first time the Congress convenes in Israel, and the Technion Management is especially proud that our scientists have been entrusted with the task of organizing and hosting this event.

But, there is another "FIRST" which I would like to point out to you. A perusal of the previous sites shows that you are meeting for the first time at an altitude of several hundred meters above the sea level, - I am confident this will bring to this Meeting "Scientific heights" previously unknown. You have certainly chosen a very unique location: Mount Carmel (on one of its hills we are here) - was mentioned already a few thousand years ago. For instance, in the Book of Jeremiah one finds "To the CARMEL by the sea, - one comes". This mountain is perceived and mentioned from very, very old times as a "SAFE SHORE". The ancient

Egyptians used to call it "The Holy Summit" and the famous Egyptian Emperor (Pharaoh) Ramses the Second, nicknamed it in the 13th Century B.C., "The Mountain of Strength". Possibly the most famous personality associated with it is the Prophet Elijah who used to prophesize and build his favorite models and theories and indoctrinate his followers right here on these hills. I mention this, since in the Ninth Century B.C., during the reign of King Ahab of Israel, a very famous Assembly was organized by the Prophet Elijah on Mount Carmel in the presence of the King, possibly the precursor of the subsequent scientific conferences held on this mountain. About 850 people attended that famous Assembly, not very different from the attendance of this Congress, at which some ideological differences (theories or approaches in our present language) were confronted. Well, here the parallel stops, since meetings in those times could have very violent endings as happened in the Assembly convened by King Ahab.

Interestingly, the location of Prophet Elijah's home-cave is determined to be in two different places, some 300 meters apart, by the Christian and Jewish faith. Given the distance of nearly 3000 years on the time scale, this is pretty good accuracy.

Referring anew to your Congress series, I was struck by an interesting historical coincidence. The first ICTAM took place in DELFT in 1924, and in the very same year our University, the Technion, opened its gates for business, enrolling its first 14 students with 7 Faculty members (a pretty good ratio we did not succeed in maintaining) - as the first University in modern Israel - it was followed less than a year later by the Hebrew University of Jerusalem. So, the date of your first Congress is also a cornerstone in the history of Modern Israeli Culture.

Today, we have 19 Faculties and Departments, and a number of Research Institutes, in most fields of Engineering and Sciences, as well as Architecture and Medicine. There are now over 10,000 students at the Technion, among them nearly 3,000 postgraduate students researching and studying for higher degrees. We have about 700 full-time Faculty members and many hundreds of associated scientists and adjunct teachers. I hope you will have the chance to visit and meet some, though this is the week of our annual vacation.

As you know, Israel is a country of small size and limited natural resources. As such, we put special emphasis on Science and Technology, the prime component in our strive to economic independence and the building of an advanced society. In this context, we considerably value international cooperation of which scientific congresses are an important form.

The recent political changes in the world have opened new avenues for international cooperation and for us in Israel they have significant implications. A sizable part of the scientific world was practically unreachable for us; this has changed completely during the last couple of years, - as the composition of this assembly, I believe, shows. Moreover, the energetic pace of European unification and the related developing of special technological and scientific collaboration schemes of international impact is also of great relevance for us. As a result, we are undergoing here a process of orientation towards a multitude of new opportunities, and I am confident that the outcome will benefit all those involved.

Maybe I should also mention that on historic scales (I already mentioned history before) the direction of international exchanges has gone through cycles.Today, there is no doubt that the leading Mecca's for Science and Technology are mostly in Europe and the United States of America. In his Nobel acceptance speech in 1979, the Pakistani physicist, Abdus Salam, from Imperial College, London, described through individual examples the clear direction of technology transfer in the early centuries of this millennium. He mentioned the cases of Michael the Scott and the Danish physician Henrik Harpestraeng who travelled in the 13th Century from the underdeveloped countries of the North to the flourishing Universities of the South, Toledo, Cordoba. In these places the Arabic and Hebrew Scholarship was then at its peak, and scholars were attracted to these and other Southern Universities from both the developed countries of the Middle East and Middle Asia, as well as from the developing northern countries.

It is my wish that we here do the utmost, so that again we may confidently say "That Scholarship will emerge from Zion". Meanwhile, on the way to this goal. I wish that your Congress will contribute significantly to the increase of knowledge and will stimulate and inspire you in your future work. I hope you have an interesting, enjoyable, fruitful and pleasant week at the Technion in Haifa."

Professor Zvi Hashin, President of the Israel Society for Theoretical and Applied Mechanics, then brought his greetings:

"Members of the Presidium, Ladies and Gentlemen,

I have the honor and privilege to bring you the greetings of the Israel Society for Theoretical and Applied Mechanics. Our society is one of the many national societies which are federated into the International Union of Theoretical and Applied Mechanics (IUTAM). The Israel Society was founded in 1950, only two years after the establishment of the State of Israel in 1948, by my dear and esteemed former teacher and advisor Prof. Markus Reiner. He also was the first President of the Society, until he passed away in 1976. It has been my great privilege to replace him as President of the Society from 1976 until the present. Professor Bodner and myself are the delegates of Israel to the IUTAM General Assembly. Professor Singer was our member in the IUTAM Congress Committee until 1988. Since 1976 we have consistently extended invitations to IUTAM to hold the International Congress of Theoretical and Applied Mechanics (ICTAM), in Israel and we are gratified and honored to have been selected during the last ICTAM in Grenoble 1988 to host this major event in 1992. We would like to believe that it is a recognition of the growth and maturity of the mechanical sciences in our country.

The Israel Society for Theoretical and Applied Mechanics is grateful to Professor Josef Singer for having accepted its invitation to chair the Local Organizing Committee, and to all the members of the Committee. In particular, we wish to express special gratitude to Professor Sol Bodner, Co-Chairman, and to Professor Alexander Solan, Secretary, for the enormous amounts of time and effort they have generously contributed to create ICTAM 18.

Thank you very much and I wish all of us a very successful and pleasant Congress."

Professor Sol R. Bodner, Co-Chairman of the Local Organizing Committee, then presented his greetings:

"I would like to make a few remarks on this occasion.

Exactly forty years ago, as a young graduate student working with Professor Nicholas Hoff at the Brooklyn Polytechnic Institute, I attended the 8th International Congress of Mechanics in Istanbul. It was a wonderful affair - the scientific level of the presentations, the extraordinary hospitality of the hosts, and the marvellous sights of the city. Present there were the famous personalities of Mechanics - Prager, who as partial host gave a brief talk in Turkish, von Mises, Courant, - I believe von Karman and G.I. Taylor were there, and many others. My impression from that Congress was that Mechanics was an excellent field in which to make a career.

Now, forty years later, the Congress has returned to the Eastern Mediterranean region and we are again gathered at its shore. I still consider Mechanics as a field that offers continuing possibilities for creative and intellectually demanding and satisfying work. It is impossible to recreate the atmosphere of Istanbul of 1952, but we shall do our best to offer the hospitality of the region.

The logo of the Congress, the ship, is taken from the emblem of the port city of Haifa, the fancy sail of the ship is the Hebrew word Chai which means 18 and also "life", or more literally "alive".

I would like to add my personal welcome to all of you who came to Haifa for this Congress."

Finally, Professor Paul Germain, President of IUTAM addressed the gathering:

"It is my privilege, as President of IUTAM, to close this official opening ceremony and to declare open our *XVIII th International Congress of Theoretical and Applied Mechanics*. but before I do it, I intend to say a few words, to thank our hosts, to recall the objectives and the responsibility of IUTAM and to emphasize the role and the importance of such a Congress for the modern development of Mechanics.

We are invited today by a country and a people who, at the same time are very old and very new. Thank you, Mayor Gurel for your presence this morning and for your welcome. You represent here both of them. A country which has to face many difficult problems, which must fight every day to solve them, which hopes to get and which has to get, full recognition by all its neighbours and then, which has to reach and to win the peace, a true peace, a peaceful peace. A land to which so many in the world are so deeply related because they feel that, here, lie some of their deepest roots. A people who, more than 3000 years ago, has begun to give the world an incomparable source of moral and spiritual life in which so many in the world today continue to draw inspiration, stimulation and strength. A people who was dispersed during centuries, who has too often suffered unjustly and especially during the first half of the century with this incomprehensible and inadmissible Shoah. A people who is today engaged with great success in modern life and in particular the scientific life. One of the evidences is this Technion, this very famous Institute of Technology. Our best thanks go to its President who has accepted that our Congress be held here. Many years ago, as it was recalled, our colleagues from the *Israel Society of Mechanics* asked our Congress Committee to hold one of our international congresses in their country. Despite the high standard of their achievements in the field of Mechanics, it was not found possible to accept their invitation on account of the international situation until four years ago at our last meeting in Grenoble. Many unpredictable events happened in the world since that time. But one must recognize today that the choice which looked a little risky was a reasonable one. No special difficulty arose for the attendance from the choice of the location. With the help of Technion and of the civilian authorities, the *Local Organizing Committee* under the friendly supervision of our *Congress Committee* was able to do the necessary work to hold this Congress in very good conditions. Without waiting for the closing session which is the most appropriate time to express our recognition, we may already express to Professor Singer and to his colleagues our gratitude.

Dear colleagues and friends, at the present time, the interactions of science and technology with the cultural, social economical, ecological problems are very strong. As scientists we are, each one of us, concerned by such a responsibility. Our Union provides a good tool to assume this responsibility at the international level, by its own initiatives first and second through ICSU, the *International Council of Scientific Unions*. As you know IUTAM is one of them. ICSU with its various specialized Committees has a role which is every day more important, in particular, as a partner on account of its scientific expertise of the United Nations and of its numerous programs and organizations. The fears and the expectations of the people cannot be ignored. The understanding and the appreciation of science have to be improved. I am glad to tell you that IUTAM has significantly increased its involvement in the ICSU work, by encouraging the operations in which the scientific content plays the principal role and by participating in those when its own expertise is very high. The best example is the scientific program of the *International Decade for Natural Disaster Reduction* run by a group of experts, coming from various Unions, chaired by Sir James Lighthill, our former President, with a remarkable success. The influence of IUTAM and of its voice inside ICSU is a function of the support it received from the mechanicians all over the world. And one measure of this support is the presence and the participation of many of them at the International Congress. It is one first reason to express to all of you my thanks on behalf of the *Bureau* and of the *General Assembly* .

Congress Statistics

Country	Papers						Participants
	S	AL	AP	L	P	IL	
Australia	6	2	2	1	2	-	4
Austria	5	3	-	2	-	-	3
Azerbaijan	*	*	*	-	1	-	1
Belgium	4	1	3	1	2	-	5
Brasil	10	1	2	-	2	-	2
Bulgaria	28	3	6	3	1	-	5
Canada	28	8	15	7	9	-	18
China - Beijing	114	3	25	3	8	-	8
China - Taipei	26	4	6	3	2	-	6
Czechoslovakia	11	1	3	-	3	-	3
Denmark	10	6	3	6	2	-	9
Estonia	1	-	1	-	-	-	-
Finland	1	-	1	-	-	-	1
France	65	25	19	22	11	2	39
Germany	46	13	18	13	13	1	36
Greece	5	2	2	1	1	-	4
Hong Kong	1	1	-	1	-	-	1
Hungary	4	-	2	-	1	-	2
India	8	-	3	-	-	-	-
Ireland	1	-	1	-	1	-	1
Israel	91	20	28	20	22	1	85
Italy	18	4	7	4	3	-	11
Japan	28	13	12	10	5	3	22
Kenya	1	-	1	-	-	-	-
Korea	1	-	-	-	-	-	3
Latvia	3	-	2	-	-	-	-
Lithuania	6	-	3	-	-	-	-
Mexico	1	1	-	1	-	-	1
Netherlands	16	8	7	8	7	1	21
New Zealand	2	1	-	1	-	-	1
Norway	1	1	-	1	-	-	1
Poland	37	6	14	5	10	-	21
Portugal	5	3	1	2	1	-	5
Romania	28	2	-	2	-	-	1
Russia	304	32	58	25	12	1	32
Singapore	1	-	1	-	-	-	-
South Africa	4	-	1	-	1	1	2
Sweden	10	2	2	2	1	1	7
Switzerland	7	2	4	2	4	1	10
Trinidad	-	-	-	-	-	-	1
Turkey	10	2	1	1	-	-	1
Ukraine	*	*	*	4	-	-	4
United Kingdom	37	14	15	13	13	3	35
USA	184	72	67	59	28	11	107
Venezuela	1	1	-	-	-	-	-
Vietnam	2	-	1	-	1	-	1
Yugoslavia	11	-	5	-	4	-	5
Total	1183	257	342	223	171	26	525
			599		394		
						420	

But our main task, our unique task this week, concerns mechanics and its development; mechanics an old science, but a- always young one. Those who like me attended many such International Congresses can tell you how fruitful these meetings are. They give to each participant as in any other scientific meeting the opportunity to communicate directly with the colleagues working in the same area, by letting them know their new results and by learning from them the latest progress. But you find more in such a Congress: the possibility to devote one week to your own on-going education, to broaden your interest to the neighbouring disciplines, to become aware to the new discoveries, to see the evolution of ideas, the creation of new concepts, the fantastic improvements in theory and applications. Our two *General Lectures*, our invited *Sectional Lectures*, our three *mini-symposia* with their lectures of pedagogical character meet precisely these purposes. And one discovers that the contributions selected by the *International Papers Committee* represent stones which are necessary to build this wonderful monument of mechanics, not yet achieved, always more beautiful, more impressive, more useful, 300 years after Newton. Yes, may all of us, and especially the younger ones, enjoy their participation; may they draw the greatest possible benefit of this week. For those who have prepared this meeting, that is their main wish; that would represent the best reward.

And now, it is time to start. I declare open the *XVIII th International Congress of Theoretical and Applied Mechanics* and leave the chair to Sir James Lighthill who, within a few minutes, after a few announcements by Professor Solan, will introduce the speaker invited to deliver the opening *General Lecturer*. Thank you for your kind attention."

Sir James Lighthill then introduced Professor Anatol Roshko who presented the Opening Lecture "Instability and Turbulence in Shear Flows".

The scientific program of the Congress was presented from Sunday afternoon till Thursday evening in 5-6 parallel sessions. The poster sessions took place in 8-9 adjoining rooms on three mornings, at times when no other lecture sessions were in progress. This resulted in good attendance and very lively discussions.

The detailed statistics of the Congress are presented on the following page. It is noted that the proportion of acceptances to submissions was almost exactly 1/2. The proportion of actual presentations to acceptances was lower than anticipated which seems to be primarily due to current difficulties in a number of countries for obtaining travel funds. As a consequence, the overall attendance was also somewhat lower than expected but was more than adequate for a very invigorating Congress.

The meaning of the symbols used in the following listing by countries is as follows:

S - Submitted abstracts
AL - Accepted as Lecture
AP - Accepted as Poster
L - Presented as Lecture
P - Presented as Poster
IL - Invited Lecture
* - The numbers for S, AL, AP for Azerbaijan and Ukraine and other countries of the former Soviet Union are included in Russia.

The Social Program included a Welcome Reception on Saturday evening, a Dinner Reception on Monday evening, both at the Churchill Auditorium Plaza; the Congress Excursion on Tuesday afternoon to Rosh Hanikra and the Crusader City, Acco; a city tour of Haifa and the Druze villages on Mount Carmel on Thursday, and other optional tours for accompanying persons on Sunday, Monday and Wednesday as well as pre-congress and post-congress tours to other parts of Israel. The Congress Banquet was held at the Dan Carmel Hotel on Thursday evening.

The Closing Session of the Congress was held in the Churchill Auditorium at 5.30 p.m. on Thursday, 27 August 1992, immediately after the Closing Lecture by Professor G.I. Barenblatt on "Micromechanics of Fracture". The closing ceremony was opened by Professor Josef Singer, Chairman of the Local Organizing Committee, with the following words:

"Ladies and Gentlemen,

All good things must come to an end, and the 18th ICTAM is no exception. I will not repeat in detail my thanks to the various IUTAM committees and to the different national committees for their excellent work and continuous assistance, nor to the Israeli sponsoring organizations and others for their generous support.

But I would like to express, on behalf of the Executive Committee of the Local Organizing Committee (Professors Bodner, Hashin, Solan and myself), our appreciation to our local collaborators:

To Professor Alexander Yarin and Dr. Dan Givoli for their very significant contributions and also to the other members of the LOC for their support.

To Dalia Sarid, the administrator of the Faculty of Mechanical Engineering, and to Bernice Hirsch and Edna Gal, who assisted us so ably in solving the many detailed problems and last minute rush jobs, as well as to Mottie Fein who did much of the computer programing. Also to the Superintendents of the Technion buildings, of the Churchill Auditorium and especially of the Lady Davis Complex, to the Technion electricians and to the staff of the Technion dormitories. To all of them, for their help to make Technion facilities available to us during the annual vacation week, when Technion is usually closed.

Our thanks also to the efficient Kenes staff for their excellent professional operation during the Congress that provided us with the daily technical support for our deliberations and other activities.

Ladies and Gentlemen,

We have had a very successful Congress, and this is primarily due to you, the lecturers, contributors, chairmen and participants.

As you know, clouds of uncertainties hung over our preparations caused by the difficult economic situation of the countries in Eastern Europe. The IUTAM Bureau and Executive Committee of Congress Committee made available special funds for partial support to participants and this helped to disperse some of those clouds. The attendance from these countries was significantly augmented, and we have nearly the same number of participants from Eastern Europe as attended in Grenoble and even more from the countries of the former Soviet Union.

Friends,

We were honored and happy to be host to the 18th Congress. We hope that you not only enjoyed the Congress, but that you will take with you fond memories of Israel and its people.

Here are a few more words for your Hebrew vocabulary: *toda raba*, which means thank you very much, *shalom*, which you know, and *lehitraot* , which means *au revoir*."

Professor Sol R. Bodner, Co-Chairman of the Local Organizing Committee, then remarked:

"As one of the group that worked very hard the past few months arranging the details of the Congress, I am very happy that the Closing Ceremony has come about. I have the impression that the Congress events proceeded fairly smoothly and properly. From what I saw in the various halls, the lecture and poster sessions were well attended with some very intensive discussions.

For most of you, the participants, this Congress was probably the occasion for your first visit to Israel. To all of you, we wish to invite you to come again whenever you can to visit the various universities and the many sights of interest. We greatly value and appreciate the continuous cooperation of the international Mechanics community and have been honored to be your hosts on this occasion. The word in Hebrew is *lehitraot* - meaning, see you again."

Then Professor H. Keith Moffatt, Secretary of the Congress Committee, presented his report:

"Mr President, Dear Colleagues,

I think you will agree that there is something quite unique about the International Congress of Theoretical and Applied Mechanics, or ICTAM as it is affectionately known. It is not just another Conference or Symposium to add to the many that are held annually; either nationally or internationally. Its uniqueness lies in the great historical tradition that goes back to the first Congress of Delft in 1924, and that has been maintained and strengthened through the succeeding decades. On your behalf, the Congress Committee strives to maintain the standard of excellence established in earlier years, and I believe that this XVIIIth Congress in Haifa has been quite exceptional in the quality and diversity of the lectures delivered and the posters displayed and defended. Our Opening and Closing Lecturers and our invited Sectional Lecturers have given us a brilliant survey of work in the fields of fluid and solid mechanics, and I wish to thank them collectively for the great effort they have devoted in preparing their lectures and delivering them here in Haifa.

The three mini-symposia have been equally successful and the introductory lectures have been extraordinarily stimulating and instructive. I wish to thank the Chairmen and Co-Chairmen of these mini-symposia for their efforts in constructing these programs, and the Lecturers for their willing cooperation.

I wish to say a particular word about the Poster-Discussion Sessions. The Congress Committee has always maintained that papers accepted for these Sessions must be at least equal in quality to the standard that we set for contributed lectures. The high standard of Posters at this Congress and the lively discussions that they have provoked testify to the success of this policy. The poster-discussion sessions promote the development of informal contacts that can lead to future research collaboration, and I believe that they form an important, integral and indispensable part of our Congress structure. Participants enjoy poster-discussion sessions precisely *because* of their participatory character, and time is available for discussion of controversial issues, in a way that is generally not possible within the strict time constraints of conventional lecture sessions.

The number of participants at this Congress has been low, although what we have lacked in *quantity* we have made up in *quality* . Nevertheless, the decline in numbers is a matter of great concern to the Congress Committee, and we believe that this trend can be reversed only by a vigorous campaign in all the 40 countries of the Union to raise awareness of the importance of the Congress, and to raise funds at National level to support increased participation. I invite you all to advertise the Congress to your colleagues and your graduate

students, so that they may already set their sights on participation in the XIXth Congress in August 1996, in a location that will soon be announced by the President.

The Congress Committee is eager to be responsive to any constructive criticisms or suggestions that participants may wish to make, which may influence the planning of future Congresses. If you have any suggestions, please write to me and I shall make sure that the suggestions receive careful consideration.

Now I have to report the impending retirement of nine members of the Congress Committee and the election of nine new members. The retiring members are:

Professors Bevilacqua, Collins, Drucker, Hult, Imai, Ishlinsky, Lighthill, Müller and Niordson.

I wish to thank all for their cooperation in the work of the Committee, and particularly Professor Dan Drucker, Sir James Lighthill and Professor Frithiof Niordson who, as past Presidents of IUTAM, have served also with great distinction as Presidents of the Congress Committee.

The new members, elected by the General Assembly are, in alphabetical order:

Professors Aref (USA), Bodner (Israel), Engelbrecht (Estonia), Hutchinson (USA), Lundberg (Sweden), Gert Meier (Germany), Pedley (UK), Sayir (Switzerland) and Tatsumi (Japan).

I wish to welcome these new members whose election to the Committee takes effect on 1 November 1992

The new Executive Committee was appointed by the Congress Committee this morning and is as follows:

Professor van Wijngaarden - President
Professor Olhoff - Secretary
Professor Acrivos, Professor Kaliszky, Professor Moffatt, Professor Bodner.

The Congress Committee is honoured to have two members who have served for many years and who now serve without limit of tenure. These are Professor Nicholas Hoff, who was President of the Stanford Congress in 1968 and who we have been very happy to see so well and active here in Haifa throughout the week of the Congress; and Professor Y.H. Ku, formerly of the National Chengchi University, Nanking, China, now Emeritus Professor at the University of Pennsylvania, USA. Professor Ku celebrated his 90th birthday last year, and his collected scientific papers of the last twenty years have just been published by the Shanghai Jiaotong University Press. I am sure you would wish to join me in congratulating him, and wishing him well for the future.

Finally, President, I wish to thank Josef Singer and the Local Organising Committee and particularly its Secretary, Alex Solan, for their devoted efforts in making this Congress run so smoothly; it is a Congress from which we have all gained intellectual stimulus and enlightenment, and on which we will all look back with great warmth and affection."

The President of IUTAM, Professor Paul Germain then said:

"Dear Colleagues and friends,

You have heard the report of the Secretary of the Congress Committee, Professor Moffatt. I strongly approve his statements: from the scientific point of view, the XVIIIth ICTAM was a very good Congress. I am sure that all of you agree with me when I express our best thanks to the Congress Committee and in particular to Keith Moffatt, to the International Papers

Committee which had to select the papers to be presented, a very difficult task and a great responsibility indeed, to the Local Organizing Committee, in particular Professor Solan and Professor Bodner, who had the duty to organize the programme taking into account all the constraints and requirements. and, last but not least, to the contributors who have delivered their best results and made us discover new problems, new methods and new points of view.

The facilities, very well managed by the Local Organizing Committee, were also of high standard. The outside temperature was pretty high. But the air-conditioning was running so well that the poor Europeans who are not used to such an equipment were somewhat freezing and they found very nice to go outside from time to time in order to get some warm air. Only one disappointment: we were expecting a more numerous attendance.

I have now the very pleasant duty to conclude the present Congress by announcing who receives the two IUTAM Bureau prizes. I remind you that such a prize is offered to a young scientist for an *outstanding presentation* of a good paper. Two prizes are distributed, one in fluid mechanics and one in solid mechanics. The fluid mechanics prize is given to Dr. Chen Xie Nung for an outstanding poster presentation of a paper, co-authored by Dr. Sharma, entitled *"Slender Ship Moving at Near-Critical Speed in a Shallow Channel"* . It is shown by a theoretical and a numerical analysis how the steady motion of the ship produces the unsteady generation of a train of solitary waves in front of it. The Solid Mechanics prize is given to Mr. S.D. Guest, a graduate student of the University of Cambridge for the very clear presentation of the paper *"Propagation of Destabilizing Waves in the Folding of Faceted Tubes"*, co-authored by Dr. S. Pellegrino. The lecture was of high quality. The non linear complex problem dealing with wave propagation and the motion of bifurcations was clearly outlined and excellently illustrated by two very well chosen models.

We have now to look at the future. I have the pleasure to announce that the Congress Committee after many discussions held during two meetings has decided that the XIXth ICTAM will take place in Kyoto. The Local Organizing Committee will be chaired by Professor Tatsumi. We are all confident that this first ICTAM in Asia will be the occasion of a deep renewal of IUTAM. We expect a numerous attendance from Far East Asia. Professor Moffatt did not want to continue to be Secretary General after 8 years of very hard and very efficient work devoted to ICTAM. Our thanks again to Keith Moffatt. Professor Olhoff will be his successor. He is well prepared to the job with the organization of the XVIth ICTAM in 1984 and his experience of the International Papers Committee in 1988 and 1992.

The General Assembly has chosen 19 symposia which will be held in 1994 and 1995, and has elected the new Bureau. Professor Leen van Wijngaarden will be our next President, Bruno Boley the Treasurer, Franz Ziegler the General Secretary. The other members will be Professors Chernyi, Moffatt, Schiehlen and Tatsumi. I want to express my deep gratitude after these 4 years during which I had the great honour to be the President of IUTAM. Needless to say that I will always keep the best recollection of these 4 years I have spent with colleagues and friends, all devoted like me to the promotion and to the development of Mechanics, all over the world.

Thank you all of you for your dynamic participation. Good luck to everybody. I hope that all of us will be in Kyoto in 1966."

Professor Germain then closed the 18th International Congress of Theoretical and Applied Mechanics.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

INSTABILITY AND TURBULENCE IN SHEAR FLOWS

A. Roshko

California Institute of Technology, MS 105-50, Pasadena, California 91125, U.S.A.

Abstract

Increasing attention is being paid to the large scale structure of turbulence and to the so-called "coherent" vortical structures which have been disclosed and studied for a number of turbulent shear flows in laboratory experiments and in numerical simulations. The coherent structures develop from the instability waves which create the flow; they portray the genesis of the turbulence in the primary instability of the global vorticity distribution. This is not in the sense of classical laminar instability, which *initiates transition* to turbulence, but as the *driving instability* in the *fully developed* turbulence. That instability provides the link to the amplitude of the turbulent motion, which in classical turbulent modelling must be calibrated empirically as a fundamental step or steps in the closure of the Reynolds averaged equations of motion. It also rationalizes the dependence on various parameters, such as compressibility, clarifies response to external disturbances, and suggests the possibility of "turbulence control". The global instability is being incorporated into new models of turbulent shear flow development and the coherent structures form the basis for new, Lagrangian models of chemical mixing and reaction in these flows.

1. INTRODUCTION

Our knowledge of turbulent flow is still mainly empirical. Feynman has called turbulence "the greatest puzzle in classical physics". It has been a prominent topic at these Congresses of Theoretical and Applied Mechanics. That it continues to attract the attention and energy of impressively large numbers of researchers attests to its technological importance and, just as importantly, to its intellectual challenge.

Since the recognition over a century ago of its importance in engineering, the problem of turbulence has developed in many directions and presents different points of view and different goals for various schools of research. Still the ultimate objective for applied mechanics is the original one, to model and "predict" turbulent fluxes of momentum and of passive scalars such

as heat. These fluxes are at the heart of practical technological problems such as the need to predict aerodynamic drag or to compute mixing rates in propulsive devices, in chemical engineering processes and in the upper atmosphere, to name just a few examples. Equally important as the urgent technological motivation has been an intense, continuing interest to understand turbulence in a fundamental, scientific way; and each has benefitted from the other. The pursuit of these two aspirations has been largely within the context of two dominant approaches to the turbulence problem, namely Reynolds-equation modelling and the statistical theory of turbulence.

The equations obtained by Reynolds in 1895 may be written in their simplest form for flow which is incompressible (density $\rho = \text{const}$) and in which the *mean* values, denoted by an overbar, ($\overline{\quad}$), are steady. Derived in an Eulerian frame, the equations are

$$\text{Continuity} \qquad \frac{\partial \bar{u}_j}{\partial x_j} = 0 \tag{1}$$

$$\text{Momentum} \qquad \rho \bar{u}_j \frac{\partial \bar{u}_i}{\partial x_j} = -\frac{\partial \bar{p}}{\partial x_i} + \frac{\partial}{\partial x_j}(\bar{\tau}_{ij} + T_{ij}) \tag{2}$$

The crucial result of "Reynolds averaging" is to introduce into these mean-flow equations the term $T_{ij} \equiv -\rho\overline{u_i' u_j'}$, which denotes mean momentum flux resulting from correlated fluctuation of momentum $\rho u_i'$ and velocity u_j' at a point in the flow. From its juxtaposition with the viscous stress $\bar{\tau}_{ij}$ in the momentum equation, T_{ij} is called the "Reynolds stress". For a Newtonian fluid, the viscous stress is related to the velocity field by a linear relation, $\bar{\tau}_{ij} = \mu\left(\frac{\partial \bar{u}_i}{\partial x_i} + \frac{\partial \bar{u}_j}{\partial x_i}\right)$, where μ is the coefficient of viscosity. For a turbulent flow, to relate the Reynolds stress to the velocity field is the great problem of "closure" for the Reynolds equations.

Other equations of this type include one for turbulent transport of a passive scalar c:

$$\rho \bar{u}_j \frac{\partial \bar{c}}{\partial x_j} = \frac{\partial}{\partial x_j}(\bar{q}_j + Q_j)$$

where $q_i = -D\frac{\partial c}{\partial x_i}$ and $Q_i \equiv \overline{c' u_i'}$. Still others have been developed over the years. For example it is possible to write an equation for the turbulent kinetic energy, $\overline{k'} \equiv \frac{1}{2}\rho\overline{u_k' u_k'}$, in the form

$$\rho \bar{u}_j \frac{\partial \overline{k'}}{\partial x_j} = T_{ij}\frac{\partial \bar{u}_i}{\partial x_j} + (\text{RDT}) - \mu\overline{\frac{\partial u_i'}{\partial x_j}\frac{\partial u_i'}{\partial x_j}} \tag{3}$$

The first term on the right hand side denotes the rate of *production* of turbulent energy by the Reynolds stress while the last one is the rate of viscous *dissipation*. The redistribution terms (RDT) contain new correlations such as $\overline{p' u_i'}$ and $\overline{k' u_i'}$. It is also possible to write a corresponding equation for the Reynolds stress itself:

$$\rho \bar{u}_\kappa \frac{\partial T_{ij}}{\partial x_\kappa} = \left(-\rho\overline{u_\kappa' u_j'}\frac{\partial \bar{u}_i}{\partial x_\kappa} - \rho\overline{u_\kappa' u_i'}\frac{\partial \bar{u}_j}{\partial x_\kappa}\right) + (\text{RDT}) - 2\mu\overline{\frac{\partial u_i'}{\partial x_\kappa}\frac{\partial u_j'}{\partial x_\kappa}} \tag{4}$$

with additional new correlations, e.g. $\overline{\rho u_k' u_i' u_j'}$, appearing in the RDT. Thus, introducing such higher order equations escalates the number of correlations which have to be "modelled" for closure, but this may still be advantageous because now the fundamental quantity T_{ij} is *not* modelled; its evolution is computed from Eq. (4) in conjunction with Eqs. (1) and (2) (Launder, 1990). We shall refer to the hierarchy of such equations as the "Reynolds-averaged equations" or, symbolically, "$\overline{Re}$ equations" and, correspondingly, to "$\overline{Re}$ modelling".

The statistical theory of turbulence, introduced by Taylor in 1935 and developed by Batchelor and others, is for *homogeneous* fields, thus does not address the turbulent fluxes and the all important question of the turbulence production. Instead it sets out to develop a rational description and theory of a simpler turbulent problem, through equations for *two-point* correlation functions. For example, $f(r) \equiv \overline{u_1'(x)u_1'(x+r)} \Big/ \overline{u_1'(x)u_1'(x)}$ correlates the fluctuating velocity component $u_1'(t)$ at two points separated by a distance r. This approach has produced a very useful language and framework for developing ideas and concepts in turbulence and has led to important results, for example Kolmogorov's theory of an inertial sub-range of turbulence dynamics.

In $\overline{Re}$ modelling as well as in the statistical theory, *scaling methods* have played a very important role. One might even say that they have given us most of our insights into turbulence and have provided the best vehicle for organizing our empirical knowledge of it. At its simplest level, scaling makes use of dimensional analysis and similarity concepts to make predictions about the scaling parameters in particular canonical flows.

An instructive illustration of this is the turbulent jet, an example of which is shown in Figure 1. It is a photo of a rocket exhaust plume on a test stand. The thrust F of the rocket is quoted as "3 million pounds". With that we can infer that the jet Reynolds number in the far field, $Re \equiv \sqrt{\rho F}/\mu$, is 2×10^9. At this value of Reynolds number the magnitude of the mean viscous stress $|\bar{\tau}_{ij}|$ in Eq. (2) is only about 10^{-7} compared to the Reynolds stress and cannot directly affect the development of the mean flow. Thus it is assumed that the viscosity μ drops out of the group of governing parameters. This assumption that, at high values of Re, turbulent free-shear flow fields do not depend on the viscosity has support from various experimental evidence. It implies that T_{ij} is independent of Re and that, in Eq. 4, the dissipation term is also independent of Re; i.e. the velocity gradients adjust themselves, with changing μ, to balance the dissipation required by the Re-independent production term. (In *wall flows*, the viscous terms cannot be omitted even in Eq. 2, because near smooth walls they become dominant.) It also follows that sufficiently far from the origin of the jet, in the "far field" where the initial geometrical details should no longer be important, the jet development depends only on the distance x from its origin and on global parameters, namely the thrust F and the ambient density ρ. Dimensional analysis then tells us that the jet thickness $\delta(x)$ must be proportional to x and the jet grows at a constant rate $d\delta/dx$ which must be the same for all free jets. Thus laboratory measurements at $Re \doteq 10^4$, where $|\bar{\tau}_{ij}| \sim 10^{-2}|T_{ij}|$, "predict" that for all higher Reynolds

Fig. 1. Turbulent jet from a rocket on a test stand. (From *Los Angeles*, Lane Magazine and Book Co., Menlo Park, California.

numbers the turbulent jet cone, made visible by some marker as in Fig. 1, will have an angle of about 23°. An important property of the jet is the rate at which it *entrains* surrounding fluid. It literally sucks or induces fluid into itself, so that its mass flow rate increases continually along its axis: the *entrained* mass flow rate is $\dot{m}_e(x)$. Again from dimensional analysis, this entrainment rate can be related to the thrust and the surrounding fluid density by the formula $d\dot{m}_e/dx = \text{const.}\sqrt{\rho \mathbf{F}}$. From an experiment by Ricou and Spalding (1961) at Re up to 10^5, the value of const. is known to be 0.28. With this we can "predict" that at 3 million pounds of thrust the jet in Fig. 1 was entraining air at the rate 1100Kg/sec/meter. (This result will be modified in early parts of the jet, where the gases are still hot and the density ρ is not uniform, and near the top of the jet, where buoyancy may play a role.) Such entraining capability is the basis for jet pumps, ejectors and thrust augmentors.

Similar and more sophisticated applications of dimensional and similarity arguments (c.f. Monin and Yaglom, 1971; Tennekes and Lumley, 1972; Barenblatt, 1979) have led to a large number of useful results and insights into various turbulent flows. Indeed, the use of "eddy viscosity" and "mixing length" formalisms in self-similar turbulent flows can perhaps be better justified as examples of scaling rather than "gradient-diffusion" methods. A notable result from this approach is von Karman's use of "mixing length theory" to derive the log law for turbulent boundary layers and to introduce his famous constant κ.

One impact of $\overline{Re}$ averaging and of the statistical theory was to tend to encourage a view of "fully turbulent" flow as too complicated and disorganized to contain structural features that could be usefully incorporated into any model. By design, $\overline{Re}$ modelling and the classical statistical theories deal only with *mean* quantities; the effects of turbulent motion appear in the statistical correlations. In the $\overline{Re}$ equations these are correlations at a point, containing no information about spatial or temporal structure of the turbulence. In the statistical theory, *spatial* correlation functions (and space-time correlations) do provide some information about turbulent structure. Such spatial correlation functions can, in principle, also be introduced into the $\overline{Re}$ equations (Batchelor, 1953; Lesieur, 1986). In recent work (e.g. Lundgren, 1982; Pullin and Saffman, 1993), efforts are being made to design mathematical models of physical vortical turbulent structure for special regimes such as Kolmogorov's universal equilibrium regime. Generally however, with a few such exceptions, arguments in turbulence modelling and theory make much use of imagery — "eddies"; "transfer of energy between eddies"; "interfaces"; "large and small structure", etc. — without the force of true, physically appropriate models. There is continuing need for such models, either for helping guide classical modelling or for discovering new approaches.

2. ORGANIZED STRUCTURE IN TURBULENCE

The preceding outline of the two dominant approaches to turbulence provides background for discussing in the main part of this talk another trend, possibly going back to Leonardo, in which it is recognized that patterns of organized structure exist in in turbulent shear flows.

After about 1940 experimental work on turbulent flow was largely steered in the two main directions of $\overline{Re}$-modelling and statistical theory. The hot-wire anemometer which had been perfected was able to obtain accurate time histories of velocity fluctuations $u_i'(t)$ and thus to provide the various correlations needed to model the $\overline{Re}$ equations as well as the correlation functions and spectra needed in the statistical theory. Highly successful in these tasks, the hot wire did not at first reveal anything that could be called organized structure. For one thing, there was little incentive to examine or analyze the velocity data for anything but the mean quantities and, in any case, even if examined in detail the fluctuation history at a single point in any turbulent flow is not revealing, nor are the *single-point* correlations used in the $\overline{Re}$ equations. But *two point* correlations of the fluctuation histories at two different points do give

a start on defining spatial structure. They were first used by Townsend (1956) and his school, e.g. Grant (1958), to find evidence of organized vortical structure in far wakes and boundary layers. Space-time correlations, introduced by Favre, Gaviglio and Dumas (1958), revealed moving patterns with defineable celerities, scales and lifetimes in boundary layers and in free shear layers (Fisher and Davies, 1964). All these were the first intimations of "organized" or "coherent" structure to be extracted from hot wire measurements in fully developed turbulent flows.

Actually, examples of coherent structure in flows *with turbulence* but with a deterministic scale were abundant. Most prominent is the example of vortex shedding from bluff bodies and the corresponding Karman "vortex street". For a circular cylinder, vortex "shedding" occurs from $Re \doteq 50$ to the highest values observed in the laboratory ($Re \doteq 10^7$), and to still higher values in (less well controlled) geophysical flows. Turbulent motion, superimposed on the organized, periodic motion, appears already at $Re \sim 10^2$. But the periodic creation and shedding of vortical structures is dominant, at a well defined frequency f_S that is related to the body dimension and the velocity of the cylinder, expressed non dimensionally in the Strouhal number, $S \equiv f_S d/U_\infty$. Therefore this was not usually considered to be a "fully turbulent flow", which by implication is (statistically) featureless. The power spectrum of velocity fluctuation in the near wake of a bluff body has a sharp peak at the Strouhal frequency, superimposed on a broad background, in contrast with the featureless spectra of homogeneous turbulence.

Far downstream of the position of the cylinder, in the "far wake", the sharp spectral peak at f_S disappears. It was in that region of "fully developed turbulence" that Townsend and Grant found evidence of defineable structure. Taneda (1959), using flow visualization, found that the original vortex street breaks down at some distance downstream, to be replaced by one with larger scale but similar in its side view appearance. Quoting him, "After this, however, the wake shows a strong tendency to rearrange itself again into the configuration of the Kármán vortex street. In this way the formation and the breakdown of the regular vortex street occur alternately as the distance from the obstacle is increased, the wave length becoming larger and larger with every transition from one vortex street to the next. Even when the regular vortex street is not formed, the dimension of the predominant discrete vortices which appear in the wake always increases as the distance from the body increases." The Reynolds numbers in Taneda's experiments were so low (< 300) that they were probably not seen as examples of turbulent flow but later experiments by Cimbala (1984) at higher values of Re (~ 5000), using flow visualization and hot-wire measurements, showed similar structure appearing far downstream where there was no Strouhal spectral peak in the spectrum. Indeed, the wake shown in Fig. 2 was downstream of a porous flat plate (a strip of screen placed normal to the stream) which did not shed vortices at all. At $x/d = 10$ the wake contains only small scale turbulence in fluid which has gone through the screen. Between $x/d = 50$ and 70 there is a burst of organized structure, quite similar to those observed by Taneda. While in this edge view it looks like a Karman "street", views normal to the plane of the wake show it to

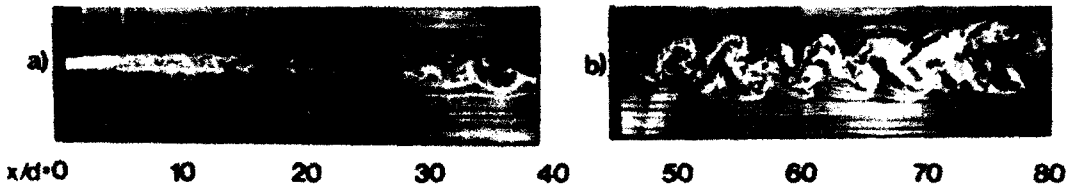


Fig. 2. Wake of porous flat plate with solidity 47%. $Re = 6000$. Visualized by a smoke wire placed at $x/d = 0$ and $x/d = 44$, resp. (Cimbala, 1984)

be three dimensional, suggestive of the mean, correlation structure inferred by Townsend and Grant. At other times, at the same location, the wake may be more disorganized, or contain a burst of coherent structure with a different scale. Thus, no fixed frequency shows up in spectral measurements of velocity at a fixed position but, rather, a *broad* spectrum with a well defined maximum peak which shifts continuously, to lower frequency with increasing distance downstream, i.e. with increasing wake scale.

These trends in Tuneda's and Cimbala's experiments illustrate general features of coherent structure in fully developed turbulent flow that had also emerged from experiments on free shear layers (also called "mixing layers"). In particular it was pictures like the one in Fig. 3,

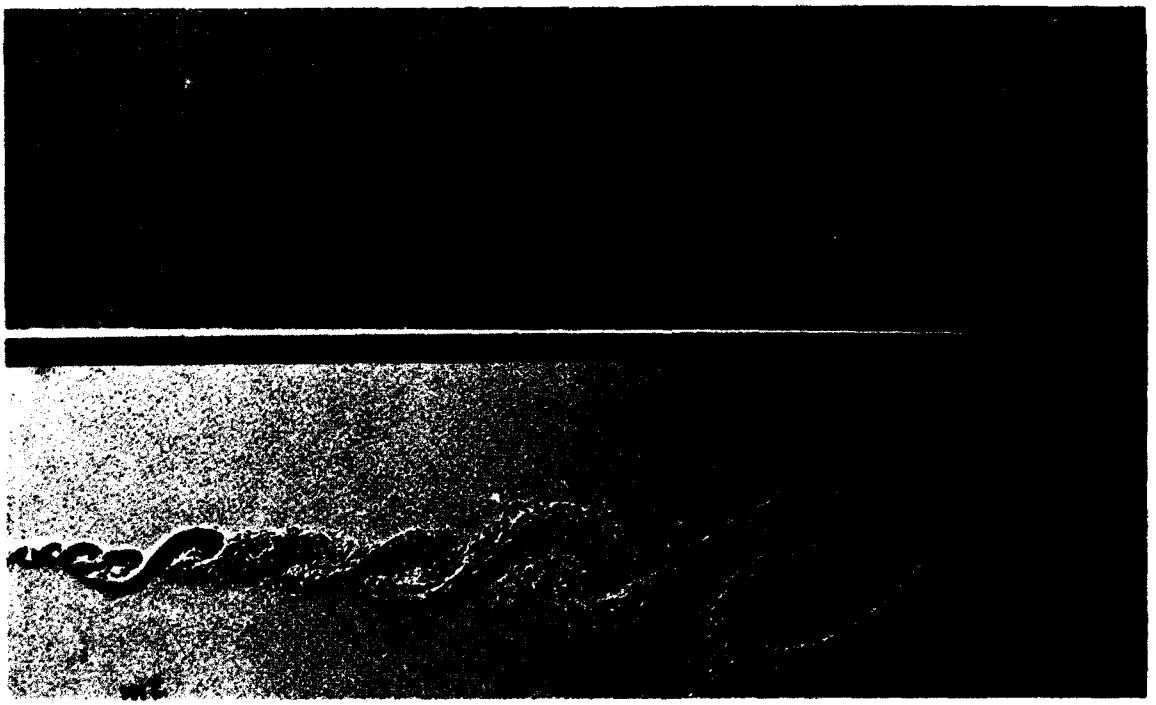


Fig. 3. Shadowgraphs of mixing layer with velocity ratio $U_2/U_1 = 0.38$. Density uniform. $(U_1 - U_2)\delta/\nu \doteq 5 \times 10^4$, based on scale of largest structures. Simultaneous edge and (partial) plan views. (Konrad, 1976).

showing vortical structures in a matrix of small scale turbulence, that helped confirm various earlier indications (Liepmann, 1952) that organized structures exist in "fully developed turbulent flows" and encouraged efforts to understand their significance and possible uses.

The following is a brief account of some of the developments from those efforts, with emphasis on the theme of this paper that the coherent structures in turbulent shear flow result from and exhibit the instability of the underlying, global mean flow. We discuss mainly jets, wakes and mixing layers, which are examples of *free* shear flows, contrasted with *wall-bounded* flows such as pipe flows and boundary layers, in which manifestations of coherent structure are also found (Kline et al., 1967, Robinson, 1991). The wall bounded flows are considerably more complicated than the free shear flows and identification of the coherent structures in them is more difficult. Wall flows have a two-scale structure while free shear flows have single-scale structure, characterized by some measure of the local thickness $\delta(x)$. At sufficiently high Reynolds number the development of free shear flows may be independent of viscosity, as indicated above. In wall flows, on the other hand, the no-slip condition at the wall ensures that viscosity is always "in the picture", creating a two-scale structure. The fundamental difference is also reflected in the stabilities of the underlying mean flow profiles. Free shear flows are variously described as having "inviscid", "dynamic", "inflectional-point" or "fast" instability; wall bounded flows are "inviscidly stable" or, at best, have "slow" instability. The no-slip condition, the two-layer structure and the slow instability in wall flows results in coherent structure which is quite different and more difficult to understand than that in the free-shear-layer class. In this paper only the latter is addressed.

In earlier paragraphs and in the following, reference is made to "fully turbulent flow". This is a term that has been used, usually in connection with the self-similar canonical shear flows (jets, wakes, mixing layers, boundary layers), to characterize the flow state at high Reynolds number and sufficiently far downstream to have "forgotten" the details of the initial conditions. It has also tended to suggest the loss of defineable structural features. To put this in perspective, it is important to make a distinction between the "near field" and the "far field", e.g. the near wake and far wake of a bluff cylinder. The near field depends on initial conditions and scales with the initial parameters, e.g. the vortex shedding frequency is locked in to the cylinder diameter. In contrast, the far field has forgotten the details of the initial conditions, depends only on global parameters such as the drag force, and scales with local parameters, such as the local thickness $\delta(x)$, which are evolving. These distinctions between near and far fields carry over to coherent structures. In both cases the structures have their genesis in the local instability of the flow, but they are more easily recognizeable in the near field because of fixed scales. In the near field the spectrum of coherent structure is sharp because the scales are sharply defined; in the far field the scales of coherent structures passing a given point are broadly distributed.

3. STRUCTURE IN FREE SHEAR LAYERS

A good introduction to the relation of instability to coherent structure in free shear layers is a picture, Fig. 4, obtained by Freymuth (1966), of the initial instability development in a laminar free shear layer. The initial part of this picture displays the expected Kelvin-Helmholtz (K-H) two dimensional "laminar" instability, whose beginnings are accurately predicted by Rayleigh's linear stability equation for *inviscid* flow. The exponential growth of the velocity and vorticity perturbations causes thickening of the mean shear layer, which eventually closes the instability window and cuts off the wave development, by which time the wave has steepened into a vortical structure; we call it a "coherent" structure. Its scale is related to the thickness of the initial laminar layer. The fluctuating velocity has a characteristic frequency, connected to the initial thickness and the initial velocity. These are *near-field* features.

Fig. 4. Kelvin Helmholtz instability, vortex formation and pairing in a free shear layer. (Freymuth, 1966).

The right hand side of the picture shows a remarkable phenomenon, the "pairing" of the vortices which form at the end of the first instability. This depicts a subharmonic response (Browand, 1966; Ho and Huang, 1982) of the thickened layer to available disturbances; the K-H instability in the thicker layer that has emerged from the initial instability is repeated at a lower wave number. That is, the K-H instability results in velocity fluctuations $u'(t)$ and $v'(t)$ which cause the layer to grow through the action of the Reynolds stress, $\overline{u'v'}$, that is developed. When the mean thickness has grown to approximately twice the initial thickness the "window" of instability for the initial wave number ends and the layer becomes receptive to a lower wave number. The second instability may be viewed either as the active interaction of two vortices from the first instability, resulting in "pairing", or as the "roll up" of (nearly passive) vortices by the subharmonic wave. This process is repeated and the growth of the shear layer is associated with repeated "pairing" (Winant and Browand, 1974) or amalgamation of the coherent structures into larger ones. After the second pairing the shear layer has largely lost memory of the initial conditions and is approaching a far-field, self-similar development, with scales independent of initial conditions.

More of the far field is shown in the picture in Fig. 3 obtained by Konrad (1976). Here the total length of shear layer is 50 λ_0, where λ_0 in the initial K-H wave length, and about 5 pairings or amalgamations have occurred, while Freymuth's picture spans 9 λ_0 and only 1 pairing has occurred. Most of Freymuth's picture is in the near field, while in Konrad's picture the near field is to the left of the "mixing transition" (Breidenthal, 1980) marked "mt" .

It is striking that the flow in the far field contains coherent structures that are similar to those in the near field. But there are fundamental differences. One of the most important is that, consistent with the far-field hypothesis, there is no *fixed*, external reference scale; the local *mean* scale of coherent structures, i.e. the vortex spacing or wave length is related to the local, *evolving scale*, i.e. the thickness $\delta(x)$. Since the mean thickness $\delta(x)$ increases smoothly, the implication is that the spectrum of scales (and corresponding frequencies) at given x cannot be sharp, since there is nothing special about any value of x. Indeed, measurements of the spacing $\lambda(x)$ (Bernal, 1981, 1988) or of frequency $f(x)$ (Winant and Browand, 1974) show a broad distribution about a maximum value. The implication (and observation) is that, at a given position, there are intervals of coherent wave trains, each with different instantaneous scale, interspersed with intervals of disorganized motion.

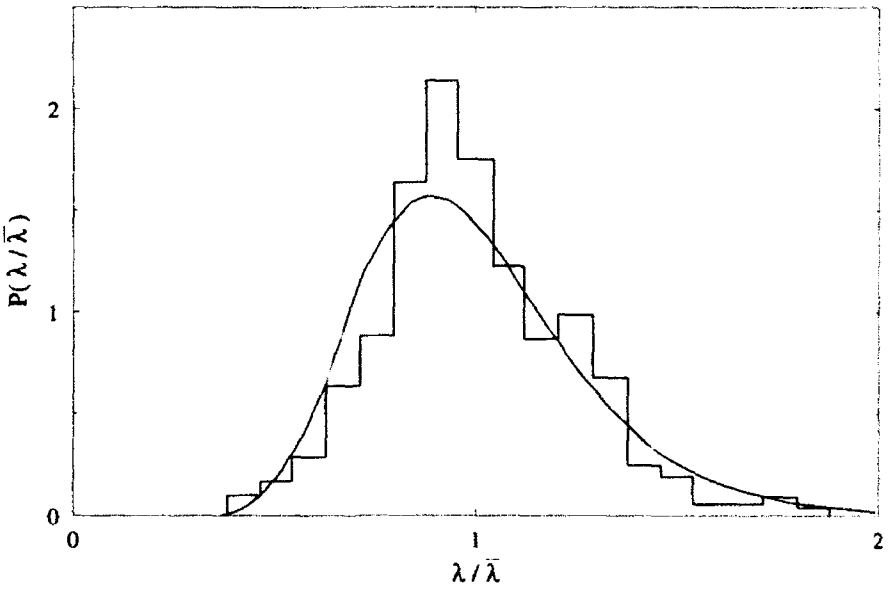


Fig. 5. Distribution of spacings of coherent structures in a turbulent free shear layer. $\bar{\lambda}(x)$ is the mean spacing at x. Vertical bars are from measurement. Curve is a log normal distribution. (Bernal, 1981, 1988).

How do the properties of these wave trains compare with those of classical K-H wave properties? First, as already mentioned, the spectrum of wave lengths or spacings λ of vortices passing a given position x are broadly distributed around a mean vlaue $\bar{\lambda}(x) \doteq 3.5\,\delta_\omega(x)$, where $\delta_\omega(x)$ is the evolving "vorticity thickness". That is, *although the large structure is coherent it is not deterministic*. The distribution (Fig. 5) has been measured by Bernal (1981, 1988), who constructed a statistical model, following the lead of Takaki and Kovasznay (1978), and found it to have a log normal distribution, which is characteristic of a fractionization process or its opposite, an amalgamation process.

Second, it is found that the most probable wave length in this distribution corresponds to the neutral point in the window of K-H amplification rates, when this is scaled with local parameters. Correspondingly, the frequency power spectrum, just as in the measurements of Cimbala in the wake, has a (broad) peak at the neutral point. The explanation of why these maxima appear at the local neutral frequency, rather than at the frequency corresponding to the peak in the amplification curve, is as follows (Ho and Huerre, 1984; Marasli, Champagne and Wygnanski, 1991). A wave in a spatially growing layer develops at fixed frequency. Its amplification rate depends on the local *dimensionless* frequency $\tilde{f} \equiv f\delta(x)/U_c$; the convective wave speed U_c is nearly constant but the thickness $\delta(x)$ is increasing. Thus $\tilde{f}(x)$ is decreasing and the wave eventually reaches a position x where $\tilde{f}$ is at the neutral value and the wave cannot grow further. It is now at its maximum amplitude and has developed into a "coherent" vortical structure. The corresponding frequency and the wave length, or vortex spacing, is the most probable one for that position. That is, the most probable vortex spacing at given x corresponds to the frequency with the greatest opportunity for amplification up to that point.

By that time the linear theory is no longer accurate. Gaster, Kit and Wygnanski (1985) made measurements of the velocity fluctuations in a mixing layer with local $Re \doteq 10^4$, using small forcing to provide a phase reference. They measured profiles across the shear layer of the velocity fluctuation amplitude, in effect the eigenfunction of the instability, and found that in the initial stages of amplification it agrees well with the theoretical shape, showing a single maximum and two side lobes. In the linear theory these lobes grow and become dominant at the neutral point; in the experiment the tendency is the same but quantitative correspondence is not as good as in the earlier stages. However their profiles of the phase distribution across the layer corresponds well with the theoretical ones up to the neutral point.

Similar measurements were made by Marasli *et al* (1991) in the plane wake of a flat plate at x/θ_0 up to 1400 and local $Re \sim 10^3$. "For the unforced flow, the peak in the measured spectrum of the cross-stream (turbulent) velocity fluctuations at any downstream location (in the far wake) corresponds to the local neutral frequency from linear, spatial stability theory for inviscid, parallel flow." That is, the trends were similar to those in the mixing layer and to those observed by Cimbala.

Another difference between the near and far fields of a free shear layer is that the layer emerges into the far field, after the mixing transition, with *three dimensional* structure. In

particular, a secondary, spanwise instability leads to the development of *streamwise* vorticity which organizes itself into a lacework of "ribs" between the primary rollers (Bernal, 1981; Hussain, 1984; Rogers and Moser, 1992).Three dimensional rib structure has also been observed in wakes in laboratory experiments and numerical simulations (Meiburg and Lasheras, 1990).

It may seem surprising that, in the presence of three dimensionality, the agreement with theoretical instability calculations for two dimensional layers and wakes could be as good as it is in the comparisons described above but, as noted, the agreement is best in the early, low-amplitude stages of the instability, where linear theory still applies. There the spanwise instability is still uncoupled from the primary instability. In the mixing layer, the spanwise instability is slower than the primary, thus the latter dominates and results in quasi two-dimensional rollers. In the wake, the spanwise instability may develop as fast as the primary (Robinson and Saffman, 1982), resulting in structures that are fully three dimensional (wavy or loopy); but in edge views the wake still looks like a "Karman-street" (Breidenthal, 1980; Cimbala, 1984).

With increasing distance downstream, the primary instability and the secondary instability both rescale themselves as may be seen in the edge and plan views in Fig. 3. In the mixing layer, rescaling of the secondary instability lags the primary; in the wake they appear to proceed together. This continuously rescaled large structure dominates the smaller structures which evolve from convectively developing chaos and from higher-order instabilities.

4. IMPLICATIONS AND APPLICATIONS OF COHERENT STRUCTURE

The developing knowledge of coherent structure has been useful in various aspects of turbulent shear flow, in ways that were usually not accessible within the context of traditional modelling. As indicated above, they have given qualitative and quantitative insights into traditional, measured correlations and spectra. Other phenomena, such as intermittency at the "turbulent interface" and the mechanics of entrainment have also been illuminated. In the following sections, we briefly discuss other developments, in particular the appearance of new directions in modelling.

4.1 Parametric Effects on Growth Rates.

The interpretation of coherent structure as a manifestation of the underlying primary instability suggested a connection between growth rates $d\delta/dx$ and instability amplification rates based on the mean velocity profile. A model of the direct connection took some time to work out (see section 4.4) but even earlier the effects of various parameters could be described. It was noted for example that density difference between the two streams of a mixing layer has an effect on the growth rate, $d\delta/dx$, which is qualitatively the same as the effect on amplification rate in linear stability theory; that is, lower density on the high speed side increases amplifi-

cation rate and growth rate, respectively; and *vice verse*. Using results from linear stability theory, Brown (1974) and Dimotakis (1986), obtained quantitative formulas for the effect.

Similarly, a connection with stability theory was made for the effect of compressibility on mixing layers between high speed flows of gases. The strong effect of compressibility, to decrease growth rate, correlates with a corresponding effect on amplification rate given by linear stability theory (Papamoschou, 1986). When the Mach number is defined in a frame which is convecting with the wave or, equivalently, with the coherent structure (Bogdanoff, 1982), it provides the proper reference for evaluating compressibility effects . The effects of this "convective Mach number" on the growth rates of mixing layers for a variety of parametric combinations of velocities (U_2, U_1) and gas properties (ρ, ρ_1, a_2, a_1) is shown in Fig. 6.

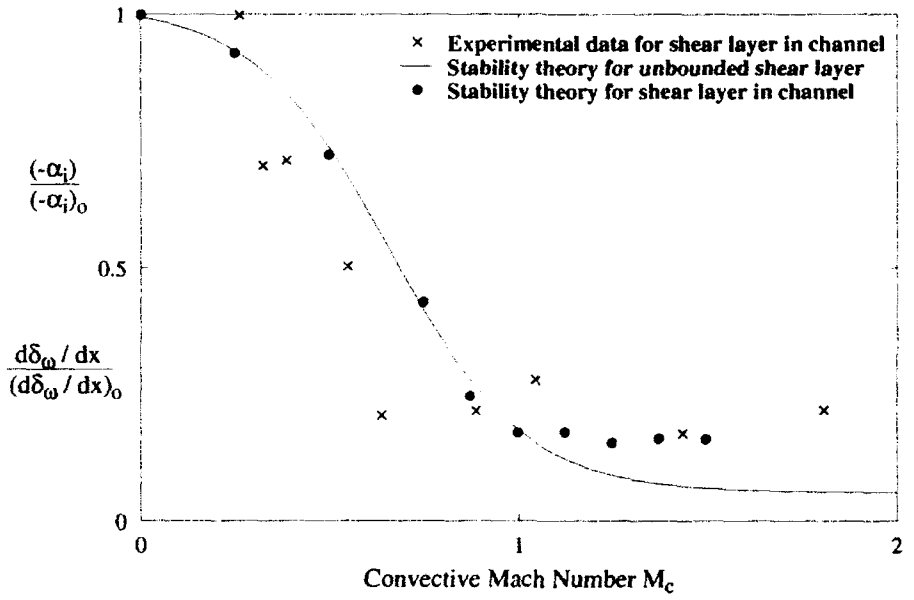


Fig. 6. Effect of compressibility on turbulent free shear layers. All data normalized at $M_c = 0$. Growth rates from experiment. Amplification rates from linear stability theory. (Papamoschou, 1986; Zhuang et al, 1990).

4.2 Turbulence Control and Sensitivity.

The existence of coherent structures suggested that they could be manipulated and the mean flow development thus modified, i.e. "controlled" in the current idiom. An example is the substantial reduction of the drag of a bluff cylinder that can be achieved by simply placing a splitter plate in its wake to interfere with the formation and shedding of vortices (Roshko, 1955). But the possibility for accomplishing such control in the far-field of a turbulent shear layer is a little more subtle because there is no fixed scale, as discussed earlier, and because

of possible difference in the nature of the underlying instability, "absolute" cf. "convective" (Huerre and Monkewitz, 1985).

Actually, control can be exercised in several ways. One is to change the stability scenario, as with the splitter plate in the near wake. Another is to change the external environment, i.e. to "change" the excitation or "forcing" to which the flow may be receptive. Such "active control" of a turbulent jet was demonstrated in the experiment by Crow and Champagne (1971) in which acoustic excitation at the jet origin enhanced growth rates when suitable frequencies, related to the "jet-mode" instability frequency, were used. The latter is analogous to the "Strouhal" frequency of vortex shedding from a bluff body.

For the far field of the mixing layer, where there is no fixed scale, Oster and Wygnanski (1982) showed that, by introducing a small, periodic oscillation at the trailing edge of the splitter plate, they could produce a strong effect on the growth rate, $d\delta/dx$. In a later experiment, Roberts (1985) produced similar effects by adding to the velocity of one stream a periodic variation of about 1%. Roughly, the effect is to increase the growth rate in the early part of the shear layer and then to stop it further downstream, at a value of x which increases with decreasing forcing frequency. After an interval of inhibited growth, the layer resumes its far-field growth rate. These effects and their relation to the shear layer parameters can be understood as the instability response of the developing shear layer to periodic excitation (Browand and Ho, 1983; Ho and Huerre, 1984) at an amplitude which, though small, takes control away from the broad band of other, available disturbances. The periodic disturbance forces the appropriate wave to emerge and grow until $\delta(x)$ reaches a cut-off value, after which the forcing tends to lock in the wave. Corresponding effects on Reynolds stress and mixing rate are quite spectacular.

This sensitivity to relatively small, imposed disturbances raises fundamental questions about how the turbulent mixing layer sustains itself, whether by self excitation or by external forcing from its environment. The ordinarily smooth, linear mean growth rate indicates that the shear layer is responding to a *broad* band of excitation. That excitation could come from disturbances in the free stream (which ordinarily have broad spectrum) or by Biot-Savart feedback to the origin from the developing turbulent layer itself. It is at the splitter-plate trailing edge, the origin of the shear layer, that the shear layer instability is most receptive to excitation. The presence of such feedback was noted by Dimotakis and Brown (1976); its possible role in establishing self excitation was studied by Kaul (1988). It has also been noted in connection with effects of "outflow boundary conditions" in numerical simulations. Also relevant to the question of how the turbulence sustains itself is the nature of the underlying instability, whether "convective" or "absolute".

4.3 Necessity for Lagrangian Viewpoints.

An important property of coherent structures, implicit in the name, is their coherence over significant lifetimes. This naturally makes for a Lagrangian view of them. Such a view

is indispensable in situations where unsteady data and their correlations are needed, as in meteorology. But even for problems in which only the mean fluxes of a stationary flow are required, as for the classical shear flows discussed above, it is sometimes advantageous or even crucial to first describe structures in a Lagrangian frame in order to correctly describe the processes leading to the final, mean values.

Such a case is the problem of determining the rate of production of chemical reaction products in turbulent mixing of reactant fluids, in mixing layers and jets. To account for some puzzling effects in such flows, and cognizant of the observed large, vortical entraining motions associated with the coherent structures, Broadwell and Breidenthal (1982) proposed a Lagrangian model of turbulent mixing in which the coherent structures are viewed as "reactors" into which the two fluids are entrained by the large scale motions, then stirred down to smaller scales and, finally, molecularly and chemically mixed. This approach makes it possible to rationally incorporate aspects of chemistry that are sensitive to how the reactants are brought together. In the final result, mean values of product fluxes are obtained, but in the Lagrangian approach some of the modelling is done before averaging. A review of the method and of some of its developments (Kerstein, 1992) and applications may be found in a recent paper by Broadwell and Mungal (1991).

Even in the basic problem of modelling the Reynolds stresses and the development of the mean flow, the limitations of the Eulerian reference frame are hidden when the shear flow is an equilibrium flow. In non equilibrium flows, e.g. relaxation from one equilibrium to another (Narasimha and Prabhu, 1972), the "memory" implicit in the lifetime of coherent structure may be important in the mechanics and for modelling.

4.4 New Directions in Modelling.

As stressed in the Introduction, a principal objective for a theory of turbulence, certainly for applied mechanics, are models which predict the turbulent fluxes in shear flows, in particular the turbulent stresses. At one extreme of possible approaches are the $\overline{Re}$ models, which are designed for broad applicability but contain empirical closure constants. At the other extreme is direct numerical simulation (DNS), i.e. the solution by numerical computation of the exact, unsteady Navier-Stokes equations. Such realization of turbulent flow in time-evolving detail in a computer is not really a "model" but rather an alternative to experimental realization in the laboratory, an alternative which has many attractive and useful features. It is limited at present in the values of Reynolds number ($\sim 10^3$) for which accurate simulation can be achieved. But even when and if it becomes possible to numerically simulate turbulent flows at will, there will still be a practical need and an intellectual incentive to seek the simplification and insight provided by models.

Between the extremes of $\overline{Re}$ modelling and DNS is the possibility for models which in some way incorporate the underlying, driving instability. Indeed, efforts to base turbulence models on hydrodynamic instability existed even before impetus was given to such efforts by

the observations of coherent structure. Examples include the marginal instability theories of Malkus (1956) for turbulent boundary layers and Lessen (1974, 1976) for free shear flows; and the wave-guide theory of Landhal (1967) for turbulent boundary layers. Theories which more explicitly sought to derive the production of Reynolds stress from the instability include those of Tam and Chen (1979); Zhang and Lilly (1981); and Liu (1988). Insights have also been obtained from related efforts to model acoustic noise radiation from turbulent jets, by associating the source of the radiation with the instability waves in the jet free shear layers (Mollo-Christensen, 1967; Liu, 1974; Ffowcs Williams and Kempton, 1978; Tam and Morris, 1980; Tam and Burton, 1984).

The present status of these modelling approaches is well represented in the recent papers of Morris, Giridhavan and Lilley (1990) and Liou and Morris (1992), who have constructed models, without empirical constants, for the turbulent development of free shear layers. Explicit use is made of the Rayleigh stability equations, which are for inviscid flow,

$$\phi = A(x)\ \mathbf{Rl}\ \phi(y) \exp\left[i(\alpha x + \beta z - \omega t)\right] \tag{5a}$$

$$(\bar{u} - \alpha/\omega)\left[\frac{d^2\phi}{dy^2} - \left(\alpha^2 + \beta^2\right)\phi - \frac{d^2\bar{u}}{dy^2}\,\phi = 0\right] \tag{5b}$$

for the eigenfunction ϕ of velocity fluctuations. The distribution and evolution of u', v' and thus of the Reynolds stress are computed, in coupled interaction with the developing mean flow equations (1)–(3), described earlier. In Eq. (5), α and β are streamwise and spanwise complex wave numbers, respectively, ω is the complex frequency and $\bar{u}(y)$ is the mean velocity profile. (Basically, the distribution of the *mean vorticity* $d\bar{u}/dy$ determines the instability.) Morris *et al.* explore several methods for coupling the equations. The simplest of these is an integral method, in which a shape function is assumed for $\bar{u}(y)$. A set of coupled equations is obtained for the shear layer thickness $\delta(x)$ and the fluctuation amplitude $A(x)$. For each wave mode $A(\omega, \beta)$, the equations have the form

$$\frac{1}{A^2}\,\frac{dA^2}{dx} = 2(-\alpha_i) + \frac{1}{\delta}\,\frac{d\delta}{dx} \tag{6}$$

$$\frac{1}{\delta}\,\frac{d\delta}{dx} = C(-\alpha_i)\,A^2 \tag{7}$$

where $(-\alpha_i)$, the imaginary part of α, is the amplification factor and C is a constant that depends, for each shear layer, on the velocity ratio U_2/U_1 and on the shape function assumed for $\bar{u}(y)$. If the shear layer contains more than one wave mode the right hand side of Eq. (7) becomes a summation, or integration, over the spectrum of wave numbers. The computation of the flow development is initiated at $x = 0$ (the "trailing edge") with an initial spectrum $A(0;\ \omega,\ \beta)$ of small amplitude, 0.01. Results for broad band as well as discrete spectra are obtained by Morris et al.

Equation (6) is derived from the energy equation (3) for the turbulent energy production, that is the wave energy production. At this point, modelling assumptions are introduced. The direct viscous dissipation and the energy transfer terms, except $\overline{p'u'}$, are omitted. The justifications, in particular for the treatment of the energy transfer from large to small scales, are based on insights from experiment (e.g. Ho and Huang, 1982) and on the earlier models. The result is a model with no empirical constants.

The values for growth rate $d\delta_\omega/dx$ computed from the model agree well with experimental data (Fig. 7). Results from a more general version of the model than that outlined above, which allows for effects of density ratio or of compressibility, are also obtained. Comparison with available experimental data for the effect of compressibility is also good. For example, it is found that spanwise instability, represented by wave numbers β, begins to play a significant role only at higher values of convective Mach number (~ 0.5), as had been found by Sandham and Reynolds (1990) in direct numerical simulations.

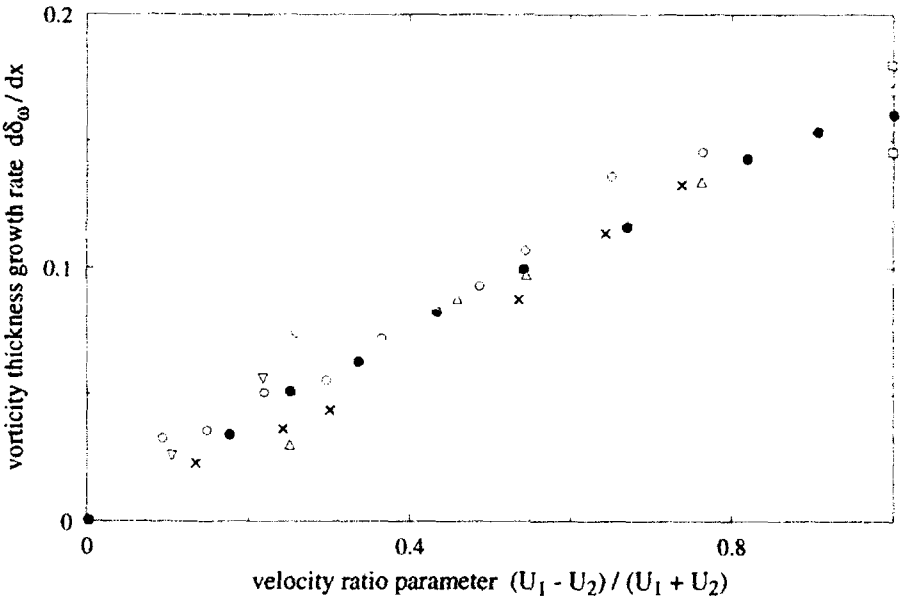


Fig. 7. Growth rate of free shear layer; dependence on velocity difference. ⊙ model of Morris *et al* (1990); ⊙ △ ▽ × experimental points compiled by Brown and Roshko (1972).

5. SUMMARY AND CONCLUDING REMARKS

The evidence that turbulent shear flows contain organized structure, and the idea that they are related to hydrodynamic instability in the mean turbulent flow, have been around for a long time, cf. the reviews by Liepmann (1952, 1962). Only recently have they fully come

together in the view that the primary instability due to the mean global vorticity distribution is the driving mechanism for the developing turbulent flow. The genesis of the turbulence is in that instability. Insightful theories should include that fundamental mechanism as the one from which other aspects of the turbulence derive and to which they relate: the indeterminancy and statistical nature of the coherent structures; the development of smaller and larger structure; and so on. Inputs from the theories of dynamical systems, chaos and fractals should be useful in illuminating various aspects of such theories for turbulent shear flow.

The evidence for the views presented is most obvious in the free shear layer, where the primary instability creates, basically, a row of vortex structures all of one sign, and it is not surprising that a successful model has been worked out first for that case. In plane wakes the basic configuration is two rows of structures of opposite sign, and the same is true for plane jets, which have not been explored as extensively. For axisymmetric wakes and jets, the coherent structure is more difficult to discern in the far field, because probably the primary instabilities are helical. The existence of large structure is not apparent in the far-field picture in Fig. 1 but it has been educed for similar large jets from computer enhanced video movies (Mungal *et al*, 1989, 1992). Morris has applied his model to the near field of round jets, i.e. to the annular free shear layer (Viswanathan and Morris, 1992). The far field is still to be addressed but, in time, the joining of instability theory with the mean flow equations will no doubt be extended to this and other, canonical shear flows. All these developments are still in the exploratory stage.

I wish to acknowledge the Office of Naval Research, United States Navy for support of research over many years. Ideas and points of view described here have developed from the research of students and colleagues at GALCIT and elsewhere and from discussions with them. The list of references from which I have benefited is many times longer than the one which is appended here.

6. SELECTED REFERENCES

Reviews, overviews and expositions shown in bold-face type

Barenblatt, G.I. 1979 *Similarity, Self-Similarity, and Intermediate Asymptotics*. Consultants Bureau, Plenum Publishers, 161–198.

Batchelor, G.K. 1953 *The Theory of Homogeneous Turbulence*. Cambridge University Press.

Bernal, L.P. 1981 The coherent structure of turbulent mixing layers. I. Similarity of the primary vortex structure, II. Secondary streamwise vortex structure, Ph. D. Thesis, California Institute of Technology. Also Bernal, L.P. and Roshko, A. 1986 *J. Fluid Mech.* **170**, 499–525.

Bernal, L.P. 1988 The statistics of the organized structures in a turbulent mixing layer. *Phys.*

Fluids **31**, 2533–2543.

Bogdanoff, D.W. 1983 Compressibility effects in turbulent shear layers. *AIAA J.* **21**(6), 926–927.

Breidenthal, R. 1980 Response of plane shear layers and wakes to strong three dimensional disturbances. *Phys. Fluids* **23**, 1929–34.

Breidenthal, R. 1981 Structure in turbulent mixing layers and wakes using a chemical reaction. *J. Fluid Mech* **109**, 1–24.

Broadwell, J.E. and Breidenthal, R.E. 1982 A simple model of mixing and chemical reaction in a turbulent shear layer. *J. Fluid Mech.* **125**, 397–410.

Broadwell, J.E. and Mungal, M.G. 1991 Large-scale structures and molecular mixing. *Phys. Fluids* **A3**, 1193–1206.

Browand, F.K. 1966 An experimental investigation of the instability of an incompressible, separated shear layer. *J. Fluid Mech.* **26**, 281–307.

Browand, F.K. and Ho, C-M. 1983 The mixing layer: An example of quasi two-dimensional turbulence. *J. de Mécanique Théorique et Appliqué, Numero Special*, 99–120.

Brown, G.L. 1974 The entrainment and large structure in turbulent mixing layers. Fifth Australasian Conference on Hydraulics and Fluid Mechanics, 352–359.

Brown, G.L. and Roshko, A. 1974 On density effects and large structure in turbulent mixing layers. *J. Fluid Mech.* **64**, 775–816.

Cantwell, J.B. 1981 Organized motion in turbulent flow. *Ann. Rev. Fluid Mech.*, 13, 457–515.

Cantwell, J.B. 1990 Future directions in turbulence research and the role of organized motion. *loc. cit. Lumley, J.L. 1990.*

Cimbala, J.M. 1984 Large structure in the far wakes of two-dimensional bluff bodies. Ph. D. Thesis, California Institute of Technology. Also Cimbala, J.M., Nagib, H.M. and Roshko, A. 1988 *J. Fluid Mech.* **190**, 265–298.

Cohen, J. and Wygnanski, I. 1987 The evolution of instabilities in the axisymmetric jet. *J. Fluid Mech.* **176**, 191–235.

Coles, D. 1985 Dryden Lecture: The uses of coherent structure. *AIAA-85-0506*.

Corcos, G.M. and Sherman, F.S. 1976 Vorticity concentration and the dynamics of unstable free shear layers. *J. Fluid Mech.* **73**, 241–264. Also 1984, **139**, 29–65.

Crighton, D.G. and Gaster, M. 1976 Stability of slowly diverging jet flow. *J. Fluid Mech.* **77**, 397–413.

Crow, S.C. and Champagne, F.H. 1971 Orderly structure in jet turbulence. *J. Fluid Mech.* **48**, 547–591.

Dimotakis, P.E. 1986 Two-dimensional shear-layer entrainment. *AIAA J.* **24**(11), 1791–1796.

Dimotakis, P.E. and Brown, G.L. 1976 The mixing layer at high Reynolds number: large-structure dynamics and entrainment. *J. Fluid Mech.* **78**, 535–560.

Dimotakis, P.E., Miake-Lye, R.C., and Papantoniou, D.A. 1983 Structure and dynamics of round turbulent jets. *Phys. Fluids* **26**, 3185–3192.

Favre, A.J., Gaviglio, J.J., and Dumas, R.J. 1958 Further space-time correlations of velocity in a turbulent boundary layer. *J. Fluid Mech.* **3**, 344–356.

Ferziger, J.H. 1977 Large eddy numerical simulations of turbulent flows. *AIAA J.* 15, 1261–1267.

Ffowcs Williams, J.E. and Kempton, A.J. 1978 The noise from the large-scale structure of a jet. *J. Fluid Mech.* **84**, 673–694.

Fiedler, H.E. 1988 Coherent structure in turbulent flows. *Prog. Aerospace Sci.* 25, 231.

Fiedler, H.E. and Fernholz, H.H. 1990 On management and control of turbulent shear flows. *Prog. Aerospace Sci.* 27, 305–387.

Fisher, M.J. and Davies, P.O.A.L. 1964 Correlation measurements in a non-frozen pattern of turbulence. *J. Fluid Mech.* **18**, 97–116.

Freymuth, P. 1966 On transition in a separated laminar boundary layer. *J. Fluid Mech.* **25**, 683–704.

Gaster, M., Kit, E., and Wygnanski, I. 1985 Large-scale structures in a forced turbulent mixing layer. *J. Fluid Mech.* **150**, 23–39.

Grant, H.L. 1958 The large eddies of turbulent motion. *J. Fluid Mech.* **4**, 149–190.

Gutmark, E. and Ho, C.M. 1983b Preferred modes and the spreading rates of jets. *Phys. Fluids* **26**, 2932–2938.

Ho, C.-M. and Huang, L.-S. 1982 Subharmonics and vortex merging in mixing layers. *J. Fluid Mech.* **119**, 443–473.

Ho, C.-M. and Huerre, P. 1984 Perturbed free shear layers. *Ann. Rev. Fluid Mech.* 16, 365–424.

Huerre, P. and Monkewitz, P.A. 1985 Absolute and convective instabilities in free shear layers. *J. Fluid Mech.* **159**, 151–168.

Huerre, P. and Monkewitz, P.A. 1990 Local and global instabilities in spatially developing flows. *Ann. Rev. Fluid Mech.* 22, 473–537.

Hussain, A.K.M.F. 1984 Coherent structures and incoherent turbulence. In: *Turbulence and Chaotic Phenomena in Fluids* (T. Tatsumi, ed.), 453–460, North Holland.

Hussain, A.K.M.F. 1986 Coherent structures and turbulence. *J. Fluid Mech.* 173, 303.

Jimenez, J. ed. 1981 ***The Role of Coherent Structures in Modelling Turbulence and Mixing. Lecture Notes in Physics No. 136*****, Springer Verlag.**

Jimenez, J. ed. 1991 ***The Global Geometry of Turbulence*****, Plenum Press.**

Kaul, U.K. 1988 Do large structures control their own growth in a mixing layer? An assessment. *J. Fluid Mech.* **190**, 427–450.

Kerstein, A.R. 1992 Linear-eddy modelling of turbulent transport. Part 7: Finite-rate chemistry and multi-stream mixing. *J. Fluid Mech.* **240**, 289–313.

Kline, S.J., Reynolds, W.C., Schraub, F.A., and Runstadler, P.W. 1967 The structure of turbulent boundary layers. *J. Fluid Mech.* **30**, 741–773.

Konrad, J.H. 1976 An experimental investigation of mixing in two-dimensional turbulent shear flows with applications to diffusion-limited chemical reactions, Ph. D. thesis, California Institute of Technology.

Laufer, J. 1981 Instability and turbulence in jets. ***loc.cit. R.E. Meyer, ed.*****.**

Launder, B.E. 1990 Phenomenonological modelling: Present and future. ***loc. cit. Lumley, J.L. 1990*****.**

Lepicovsky, J., Ahuja, K.K., Brown, W.H., and Burrin. R.H. 1987 Coherent large-scale structures in high Reynolds number supersonic jets. *AIAA J.* **25**, 1419.

Lesieur, M. 1989 (2nd ed. 1991) ***Turbulence in Fluids*****, Klewer Academic Publishers.**

Lessen, M. and Singh, J. 1974 Stability of turbulent jets and wakes. *Phys. Fluids* **17**(6), 1329.

Lessen, M. and Paillet, F.L. 1976 Marginal instability of turbulent shearing layers and the break point of a jet. *Phys. Fluids* **19**(7), 942.

Liepmann, H.W. 1952 Aspects of the turbulence problem. ***Z. Angew. Math. Phys.*** **3, 321–342, 407–426.**

Liepmann, H.W. 1962 Free turbulent flows. In: ***Mécanique de la Turbulence, A. Favre, ed. C.N.R.S., Paris*****, 211–227.**

Liepmann, H.W. 1979 The rise and fall of ideas in turbulence. ***American Scientist 67*****, 221–228.**

Liou, W.W. and Morris, P.G. 1992 Weakly nonlinear models for turbulent mixing in a plane mixing layer. *Phys. Fluids A* **4**(12), 2798–2808.

Liu, J.T.C. 1974 Developing large-scale wavelike eddies and the near jet noise field *J. Fluid Mech.* **62**, 437–464.

Liu, J.T.C. 1988 Contributions to the understanding of large-scale coherent structures in developing turbulent free turbulent shear flows. ***Adv. App. Mech.*** **26, 183–309.**

Liu, J.T.C. 1989 Coherent structures in transitional and turbulent free shear flows ***Ann.***

Rev. Fluid Mech. **21, 285–315.**

Lumley, J.L. (ed.) 1990 ***Lecture Notes in Physics: Whither Turbulence? Turbulence at the Crossroads.*** **Springer-Verlag, Berlin.**

Lundgren, T.S. 1982 Strained spiral vortex model for turbulent fine structure. *Phys. Fluids* **25**, 2193.

Malkus, W.V.R. 1956 Outline of a theory of turbulent shear flow. *J. Fluid Mech.* **1**, 521.

Marasali, B., Champagne, F.H. and Wygnanski, I.J. 1991 On linear evolution of unstable disturbances in a plane turbulent wake. *Phys. Fluids A* **3**, 665–674. Also *J. Fluid Mech.* **198**, 255–273.

Meiburg, E. and Lasheras, J.C. 1988 Experimental and numerical investigation of the three-dimensional transition in plane wakes. *J. Fluid Mech.* **190**, 1–37.

Mankbadi, R.R. 1992 Dynamics and control of coherent structures in turbulent jets. ***Appl. Mech. Rev. 45(6),*** **219–244.**

Meyer, R.E. (ed.) 1981 ***Transition and Turbulence.*** **Academic Press.**

Michalke, A. 1965 Vortex formation in a free boundary layer according to stability theory. *J. Fluid Mech.* **23**, 371–383.

Michalke, A. 1984 Survey on jet instability theory. ***Prog. Aerospace Sci.*** **21, 159–199.**

Mollo-Christensen, E. 1967 Jet noise and shear flow instability seen from an experimenter's viewpoint. *A.S.M.E. J. Appl. Mech.* **E89**, 1–7.

Monin, A.S. and Yaglom, A.M. 1971 ***Statistical fluid mechanics*** **(transl. from Russian ed. 1965). M.I.T. Press: Cambridge, Mass.**

Monkewitz, P.A. 1990 The role of absolute and convective instability in predicting the behavior of fluid systems. ***Eur. J. Mech., B/Fluids*** **9(5), 395–413.**

Morris, P.J., Giridharan, M.G. and Lilley, G.M. 1990 On the turbulent mixing of a compressible free shear layer. *Proc. Roy. Soc. A* **431**, 219–243.

Mungal, M.G. and Hollingsworth, D.K. 1989 Organized motion in a very high Reynolds number jet. *Phys. Fluids* **A1**, 1615–1623.

Mungal, M.G., Lozano, A., and van Cruyningen, I. 1992 Large-scale dynamics in high Reynolds number jets and jet flames. *Exp. in Fluids* **12**, 141–150.

Narasimha, R. 1990 The utility and drawbacks of traditional approaches *loc. cit. Lumley, J.L., (ed.) 1990.*

Narasimha, R. and Prabhu, A. 1972 Equilibrium and relaxation in turbulent wakes. *J. Fluid Mech.* **54**, 1–17.

Oster, D. and Wygnanski, I. 1982 The forced mixing layer between parallel streams. *J. Fluid*

Mech. **123**, 91–130.

Papamoschou, D. 1986 Experimental investigation of heterogeneous compressible shear layers. Ph. D. thesis, California Institute of Technology. Also Papamoschou, D. and Roshko, A. 1988 *J. Fluid Mech.* **197**, 453–477.

Pierrehumbert, R.T. and Widnall, S.E. 1982 The two- and three-dimensional instabilities of a spatially periodic shear layer. *J. Fluid Mech.* **114**, 59–82.

Pullin, D.I. and Saffman, P.G. 1993 On the Lundgren-Townsend model of turbulent fine scales. *Phys. Fluids A* **5**(1), 126–145.

Reynolds, W. C. 1990 The potential and limitations of direct and large eddy simulations *loc. cit. J.L. Lumley, ed. 1990* .

Ricou, F.P. and Spalding, D.B. 1961 Measurements of entrainment by axisymmetrical turbulent jets. *J. Fluid Mech.* **11**, 21–32.

Roberts, F.A. 1985 Effects of a periodic disturbance on structure and mixing in turbulent shear layers and wakes. Ph. D. thesis, California Institute of Technology.

Robinson, A.C. and Saffman, P.G. 1982 Three-dimensional stability of vortex arrays. *J. Fluid Mech.* **125**, 411–427.

Robinson, S.K. 1991 Coherent motions in the turbulent boundary layer. ***Ann. Rev. Fluid Mech.* 23, 601–639.**

Rogers, M.M. and Moser, R.D. 1992 The three-dimensional evolution of a plane mixing layer: the Kelvin-Helmholtz rollup. *J. Fluid Mech.* **243**, 183–226.

Roshko, A. 1955 On the drag and shedding frequency of bluff bodies. *J. Aero. Sciences* **22**, 124–132.

Roshko, A. 1976 Structure of turbulent shear flows: A new look. ***AIAA J. 14,* 1349–1357.**

Saffman, P.G. 1978 Problems and progress in the theory of turbulence, 101. In: ***Structure and Mechanisms of Turbulence II*, H. Fiedler (ed.), Springer-Verlag.**

Saffman, P.G. 1981 Vortex interactions and coherent structures in turbulence. ***loc. cit. R.E. Meyer, ed. 1981.***

Sandham, N. and Reynolds, W.C. 1990 Compressible mixing layer: Linear theory and direct simulation. *AIAA J.* **28**, 618–624.

Stuart, J.T. 1971 Nonlinear stability theory. ***Ann. Rev. Fluid Mech. 3,* 347–370.**

Takaki, R. and Kovasznay, L.S.G. 1978 Statistical theory of vortex merger in the two-dimensional mixing layer. *Phys. Fluids* **21**, 153–156.

Tam, C.K.W. and Chen, K.C. 1979 A statistical model of turbulence in two-dimensional mixing layers. *J. Fluid Mech.* **92**, 303–326.

Tam, C.K.W. and Burton, D.E. 1984 Sound generated by instability waves of supersonic flows. *J. Fluid Mech.* **138**, 249–271.

Taneda, S. 1959 Downstream development of the wakes behind cylinders. *J. Phys. Soc. Japan*, **14**, 843–848.

Tennekes, H. and Lumley, J.L. 1972 *A First Course in Turbulence*, M.I.T. Press, Cambridge, Massachusetts.

Thomas, F.O. 1991 Structure of mixing layers and jets. *Appl. Mech. Rev.* 44, 119–153.

Townsend, A.A. 1976 *The Structure of Turbulent Shear Flow*, 2nd edn. (1956, 1st edn.) Cambridge University Press.

Viswanathan, K. and Morris, P.J. 1992 Predictions of turbulent mixing in axisymmetric compressible shear layers. *AIAA J.* **30**(6), 1529–1536.

Winant, C.D. and Browand, F.K. 1974 Vortex pairing: the mechanism of turbulent mixing layer growth at moderate Reynolds numbers. *J. Fluid Mech.* **63**, 237–255.

Wygnanski, I. and Petersen, R.A. 1987 Coherent motion in excited free shear flows. *AIAA J. 25,* 201–213.

Yoda, M., Hesselink, L., and Mungal, M.G. 1992 The evolution and nature of large-scale structures in the turbulent jet. *Phys. Fluids A* **4**, 803–811.

Zhang, S. and Lilley, G.M. 1981 A theoretical model of the coherent structure of the turbulent boundary layer in zero pressure gradient. In *Proc. 3rd Int. Symp. on Turbulent Shear Flows.* Springer-Verlag, Berlin.

Zhuang, M., Kubota, T. and Dimotakis, P.E. 1990 Instability of inviscid, compressible free shear layers. *A.I.A.A. J.* **28**, 1728–1733.

Zhuang, M., Dimotakis, P.E. and Kubota, T. 1990 The effect of walls on a spatially growing supersonic shear layer. *Phys. Fluids A* **2**(4), 599–604.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
1993 IUTAM.

Micromechanics of fracture

G.I. Barenblatt

Theoretical Department, P.P. Shirshov Institute of Oceanology, Russian Academy of Sciences, Moscow, Russia*

Abstract

The characteristic property of the phenomena studied by micromechanics is the governing influence of the variations of the material microstructure on the macroscopic behavior of bodies. In the mathematical models of such phenomena, the macroscopic equations of mechanics and the kinetic equations of the microstructural transformations form a unified set that should be solved simultaneously.

Fracture occurs in the zones of high stress concentration where the phenomenon is often complicated by phase transformations, chemical transformations, and heat generation. This makes the application of the micromechanical approach to fracture phenomenon very natural.

At the beginning of the present lecture, a concise general outline of micromechanics is given. Applications of scaling, intermediate asymptotics, and renormalization groups to the micromechanics of fracture are discussed. Subsequently, an example of the application of the micromechanical approach to the study of damage accumulation is considered.

1. INTRODUCTION

Monsieur le Président, Ladies and Gentlemen: First of all I want to thank the Organizers of the Congress collectively and individually for the honour rendered to me by the invitation to present this lecture.

The construction of continuum mechanics, - the general mathematical model of motions and/or equilibrium of real deformable bodies usually proceeds in the following way. After proper definition of the basic concepts such as the observer, continuous medium, etc., the general covariance principle and conservation laws are introduced. According to the covariance principle, the laws of motion of the continuous media

*Present address: DAMTP, University of Cambridge, Silver Street, Cambridge CB3 9EW, U.K.

should be expressed via the equations equally valid for all observers. This principle of equality of all observers claims the invariant nature of the medium and its motions and therefore greatly simplifies the analysis. Conservation laws express the general fact that matter, momentum, angular momentum, energy, etc., do not arise spontaneously and do not disappear by themselves. In combination with the covariance principle, the conservation laws lead to such concepts as flux vector, stress tensor, energy flux, couple stresses, etc.

Conservation laws and the covariance principle taken alone are insufficient for designing the mathematical models of the motions of real bodies. To achieve this goal, it is necessary to provide the continuum with physical properties. This means that a certain apriori model of the continuum is proposed which is adequate for the class of motions and bodies under consideration, but is generally valid only for restricted classes of the motions of real bodies. Such models, the so-called constitutive equations, are represented by certain relations between the properties of the motion (or the state) of the continuous medium and the active internal forces.

Simplest examples are the classical models of continuous media such as Hooke's elastic solid. This model for an ideally linear isotropic elastic body is characterized by three constants: density ρ, and two elastic constants, e.g. Young modulus E and Poisson's ratio ν. For obtaining the mathematical model for the brittle fracture phenomenon, Hooke's model should be extended since fracture phenomena do not exist in the classical linear elasticity problem statement. Extension of the model of an ideal elastic solid for bodies with cracks assumes that the body is perfectly linearly elastic outside infinitely thin cracks. However, cohesive forces are acting on the crack surfaces near the edges of the cracks. The distribution of the cohesive forces over the crack surface in the mobile equilibrium state is autonomous, i.e. independent of the form and position of the outside edge regions of the cracks. In this extended model, an additional constant appears: fracture toughness or cohesive modulus K_0, which is an integral characteristic of the cohesive forces.

It as assumed in using classical models that the constants entering the constitutive equations are universal ones. This means that these constants, being determined from some special experiments, preserve their values for arbitrary motions described by the considered model.

An approach similar to the classical encounters essential difficulties when applied to many new materials entering modern research and high technology, or even traditional structural materials under extreme conditions characteristic of fracture regions. Researchers attempted at first to use modified classical models, especially for new materials. So, three, four, and even eight model constants appeared. This procedure, however, is unreliable. It is not only the growing complexity of the determination of new constants that arises, nor even the loss of a clear physical sense of these constants. Much worse is that these constants, so-to-say, "cease to be constants", i.e. the range of universality where the constants of the model can be considered as genuinely universal ones becomes so narrow that the models lose their predictive capability.

The reason for that is well known. Every material has its own relaxation time τ, i.e. a characteristic time when the shear stresses are

preserved under imposed fixed shear strain. In turn, every process introduces its own characteristic time τ. Therefore, for each process in each material a governing dimensionless parameter appears, the so-called Deborah number,

$$De=\tau/T \tag{1.1}$$

For "fluid-like" behavior of a material in a process, De should be $De<<1$, while for "solid-like" behavior, $De>>1$.

We now note that the normal duration of a human experiment encompasses about 18 decimal orders of magnitude: from 1ns = 10^{-9}s, to 10^{9}s ~ 30 years, so that $10^{-9}s<T<10^{9}s$ is the range of the usual human experiments. For water at normal conditions, $\tau\sim10^{-12}$s, while for steel at normal temperatures and not too large stresses, $\tau\sim10^{12}$s. Thus for the full range of human experiments at normal conditions, $De<<1$ for water, and $De>>1$ for steel. But that does not apply to steel near the edge of an extending crack where, due to the stress concentration and/or heat generation, τ is strongly reduced. It also does not apply for well known synthetic materials such as silly putty, for which $\tau\sim1$s under normal conditions. That material demonstrates all types of behavior ranging from perfectly elastic and brittle when it is struck or torn quickly, to perfectly viscous when it takes the form of a cup where it is placed for a period of less than an hour.

For describing such mixed behavior, an approach was advanced which seemed at first glance attractive: spatial and temporal non-locality was introduced to the models. It meant practically that functionals were introduced into the constitutive equations. For instance, for the so-called "simple body",

$$\underset{\sim}{\sigma} = \underset{\sim}{u}(\underset{\sim}{F}(\theta)) \tag{1.2}$$

(Here $\underset{\sim}{\sigma}$ is the stress in a given particle, $\underset{\sim}{F}(\theta)$ the strain gradient, $\underset{\sim}{u}$ the constitutive functional, and θ is the time over the whole range from the beginning of deformation to the actual current time t: $-\infty<\theta\le t$).

Without going into details, it is noted that the appearance of functionals in constitutive equations always means that the approach is insufficient. To demonstrate, let us consider the simplest example: gas flow in a tube with heat conducting walls. Imagine an obviously nonsensical case: we do not wish to introduce temperature as a consideration. Evidently the density of a certain particle will become a functional of pressure. It is not necessary to explain that in gas dynamics since another way is used (although not in the mechanics of polymers). The temperature is introduced into consideration and an energy equation is added to the mathematical model. The temperature, however, is nothing but the mean energy of microscopic molecular motion, and the energy equation can be interpreted as a kinetic equation for this directly observable characteristic of the microstructure. Introduction of the temperature simplifies everything since at least the functionals disappear from consideration.

2. MICROMECHANICS

In the last decades, and most intensively in recent years, an alternative approach has appeared more and more frequently in research practice. According to this approach, the properties of the material microstructure, directly or indirectly observable, are explicitly introduced into consideration. The equations of macroscopic motions and those of the kinetics of microstructural transformation are considered simultaneously.

In our opinion this is the very subject of micromechanics. Thus, in our understanding, micromechanics is the branch of mechanics studying the phenomena for which the variations of microstructure are of governing influence for the macroscopic behavior of bodies.

If the time scales of the processes under consideration are such that the variations of the microstructure can be considered either as instantaneous or as negligible, a classical approach neglecting microstructural variations and using finite constitutive equations - integrals of the kinetic equations - can be applied.

Concerning the very term "micromechanics", often, and in particular in the works of Professor J. Achenbach, Professor Z. Hashin, Professor V. Entov and other scientists, this term is used in a more narrow sense denoting the technique for determining macroscopic properties of bodies on the basis of certain models of their microstructure. From our viewpoint, this is only a certain part of the subject of micromechanics and we prefer the wider definition given above.

We do not have time and place here for including a comprehensive historical analysis of the origin and formation of micromechanics. It seems necessary, however, to mention several milestone works which are doubtless related to micromechanics in the sense just mentioned. At the same time, they belong to researchers whose reputation in mechanics is undisputable - they are generally accepted as outstanding representatives of the mechanical community. The last point is specially important: the mechanical community is rather conservative - in the good sense of the word - in matters that concern the subject of mechanical research. Therefore, the example offered by outstanding people within the mechanical community is important: to be modern in mechanics, the researchers should know micromechanics. The first to be mentioned is Th. von Kármán, one of the great founders of IUTAM and ICTAM's. In a series of papers, summarized to some extent in his Maryland lecture entitled "Aerothermochemistry" [1], Th. von Kármán formulated the general problem of aerothermochemistry as the problem of fluid dynamics with chemical transformations accompanied by heat generation. For the transformation rate W of an active component of the mixture, he used semi-empirical formulae of the Arrhenius type:

$$W \sim (1-\alpha)^{p} \exp(-U/kT) \qquad (2.1)$$

Here α is the mass concentration of an active component under consideration, U - a constant called the activation energy, k - the Boltzmann constant, p - "reaction order", - another constant. The heat generation rate is assumed to be proportional to the reaction rate.

It is instructive that Th. von Kármán did not intend to invent a new branch of science motivated only by the internal needs of science. Just the opposite: he was one of the first who understood, for example, that the problem of space vehicle reentry to the Earth's atmosphere cannot be solved without proper consideration of the chemical transformation and dissociation of air. Even clearer for him was the necessity to create this new branch of mechanical science in his attempts to design a rational theory of liquid fuel combustion in jet engines.

In the general lecture on the XIV ICTAM [2] devoted to "Microhydrodynamics", G.K. Batchelor summarized a series of works concerning micromotions of small bodies in fluids. Subsequent papers of G.K. Batchelor and his students, especially E.J. Hinch, applied the results and approaches of microhydrodynamical studies to create the modern hydrodynamics of suspensions.

In his papers [3,4] entitled "Micromechanics", B. Budiansky emphasized the growing attention of the mechanical community to "the mechanics of very small things". He considered several problems, seemingly special but in fact very important from the general viewpoint of the micromechanics of solids, such as transformation - toughening due to phase changes and particulate toughening.

It is instructive to look at the photographs of microstructural pictures from the mentioned papers by G.K. Batchelor [2] and by B. Budiansky [3] to better represent the objects which have recently entered the fields of interest of the mechanical community.

B. Budiansky writes [4]: "Micromechanics is the currently fashionable designation of what is really an old subject, but one that is receiving increasing attention from theoreticians in applied mechanics. Armed with their repertoire of analytical tools, they try to relate the overall deformation and strength properties of the materials to the behaviors and interactions of their microscopic constituents". Try to relate does not always mean the use of finite equations: they can be related by differential equations, and that is exactly the definition used here. A natural question arises: whether the coining of a new term is important and justified, thereby legitimizing a new branch of continuum mechanics and giving it a status of certain independence?

I suppose, yet, it is. The examples of the theory of oscillations which considers oscillations in a way irrelevant to their physical nature, functional analysis, cybernetics, and to a lesser extent cynergetics, confirm it. A general approach as well as a unified style of the analysis of new phenomena appears, which, to a certain extent, leads to a new general ideology. Seemingly uncoordinated results appear in unified form. The transfer of results from one subject to another becomes possible as well as the prediction of results based on the previous experience. Therefore the legitimization of micromechanics as a unified new branch of continuum mechanics seems to be well deserved and expedient.

For a more or less detailed further presentation, I had to make a choice. So, I selected the application of similarity, more exactly scaling (renormalization group) methods, in the micromechanics of fracture and the micromechanical analysis of damage accumulation.

3. APPLICATION OF SIMILARITY METHODS IN MICROMECHANICS OF FRACTURE

A. The systems of equations of micromechanical models are usually complicated, and analytical solutions of such sets may be obtained only as rare exceptions. Even obtaining numerical solutions involves serious problems. Therefore, the asymptotic methods and, in particular, similarity methods play an important role in the micromechanics of fracture.

The application of similarity methods in fracture mechanics has a long-time established tradition. Here we will speak about scaling (or similarity laws of the second type) which are closely connected with the renormalization group: a concept, very fashionable now in theoretical physics. We illustrate it firstly on a simple geometrical example. Consider two continuous curves of diameter D. One of them is a normal circle. We inscribe a regular n-gon with side length η in it. For the length of the perimeter of the polygon, L_η, dimensional analysis gives

$$L_\eta = D\Phi(\eta/D) \tag{3.1}$$

For sufficiently small η/D, the function Φ is arbitrarily close to its limit π, so asymptotically $L_\eta=\pi D$. It is trivial that this asymptotic is invariant with respect to a transformation group

$$L'_\eta = L_\eta,\ D'= D,\ \eta' = \lambda\eta \tag{3.2}$$

($\lambda<1$ is a group parameter). We refer to this case as the similarity of the first type, or complete similarity.

The second curve is a Mandelbrot fractal, for instance von Koch curve [5,6]. (It was B. Mandelbrot who coined the very term "fractal" and was able in his excellent monographs-essays [5,6] to show systematically the importance of such objects which entered mathematics and physics early in this century for modern natural philosophy in the whole.) Then, the asymptotic of the function Φ in (3.1) at small η/D is $\Phi\sim C(\eta/D)^{-\alpha}$, where α=Const, $0<\alpha<1$, so

$$L_\eta= \mathrm{Const}(D^{1+\alpha}/\eta^{\alpha}) \tag{3.3}$$

It is clear, that at $\eta/D\to 0$ there is no finite limit, $L_\eta\to\infty$. Meanwhile $L_\eta\eta^{\alpha}$ tends to a finite limit, and the asymptotic (3.3) is invariant with respect to the transformation group

$$L'_\eta= \lambda^{-\alpha}L_\eta,\ D' = D,\ \eta' = \lambda\eta \tag{3.4}$$

- the simplest example of the renormalization group mentioned above. The number $m=1+\alpha$ is called the fractal dimension of the curve under consideration. We refer to such a case as scaling, similarity of the second type, or incomplete similarity.

The natural step in micromechanics of fracture proposed by B. Mandelbrot and his colleagues [7] was to investigate the fracture surface of metals - whether the fracture surface is fractal and, if so, what information can be extracted from its fractal dimensions?

An excellent result of the perfect fractality of fracture surfaces was obtained by V.K. Horvath and H.J. Herrmann [8] for stress-corrosion cracks (Fig. 1). More typical and natural are the results demonstrated by R.H. Danskardt, F. Haubensack and R.O. Ritchie [9]: there exist several intervals of fractality between the natural material length scales. The non-trivial question is how the fractal dimension of the fracture surface is correlated to fracture toughness of the material or, in other words, how to use these results for quantitative strength calculations? Although this question has no definite answer yet, it is a good problem.

The application of the similarity approach combined with micromechanical analysis to fatigue crack studies is instructive. The classical result in fatigue fracture is the scaling law by Paris and Erdogan [10] for crack propagation under multi-cycle fatigue:

$$d\ell/dn = A(\Delta K)^m \tag{3.5}$$

Here $d\ell/dn$ is the fatigue crack velocity per cycle averaged over the cycle and ΔK is the stress-intensity factor amplitude. An example of such a "kinetic diagram" is presented in Fig. 2.

Let us consider [11-13] the kinetic diagram (3.5) from the viewpoint of the similarity approach. The average velocity $d\ell/dn$ can depend on the following quantities: ΔK, $R=K_{min}/K_{max}$ - the asymmetry of loading, h - the characteristic specimen size, f - frequency, K_{IC} - standard fracture toughness, σ_y - yield stress, t - time. Dimensional analysis gives

$$d\ell/dn = (\Delta K/\sigma_y)^2 \Phi(\Delta K/K_{IC}, R, z, ft) \tag{3.6}$$

where

$$z = \sigma_y \sqrt{h}/K_{IC} \tag{3.7}$$

is the basic similarity parameter. An asymptotic stage is considered where the influence of the argument ft disappears. The intermediate-asymptotic character of fatigue crack propagation is clearly confirmed by analysis of the fracture surface: a regular system of striations appears at this stage. Moreover, the parameter $\Delta K/K_{IC}$ is small. Then, two possibilities appear: complete similarity when the limit of Φ at $\Delta K/K_{IC} \to 0$ is finite or not. If it is finite, we would obtain the power law (3.5) with m=2 which is practically never the case. Let us assume, however, that the incomplete similarity (similarity of the second kind [14,15]) takes place at the intermediate asymptotic stage of fatigue crack extension, when

$$\Phi = (\Delta K/K_{IC})^{\alpha} \Phi_1(R,z) \tag{3.8}$$

and α, in principle, should also depend on the similarity parameter z.

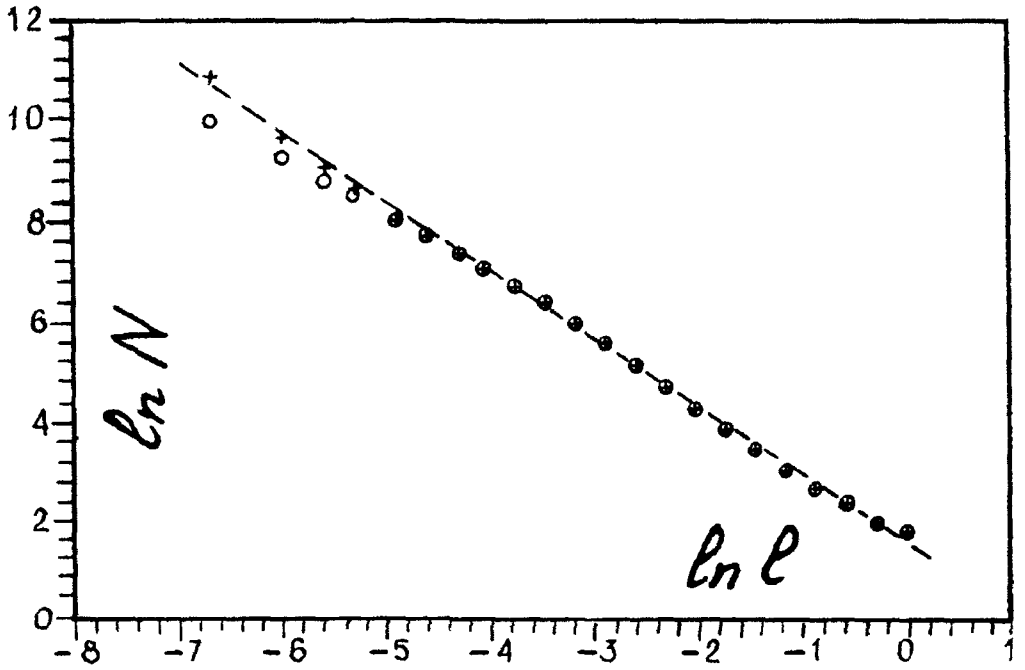


Figure 1. The straight line on the log-log plot of the number of "filled" boxes of the net; covering the section of the cracked body against the box size demonstrates the fractal character of the stress corrosion cracks in Inconel 600 in high temperature water (after [8]). (The circles and crosses denote various methods of data processing).

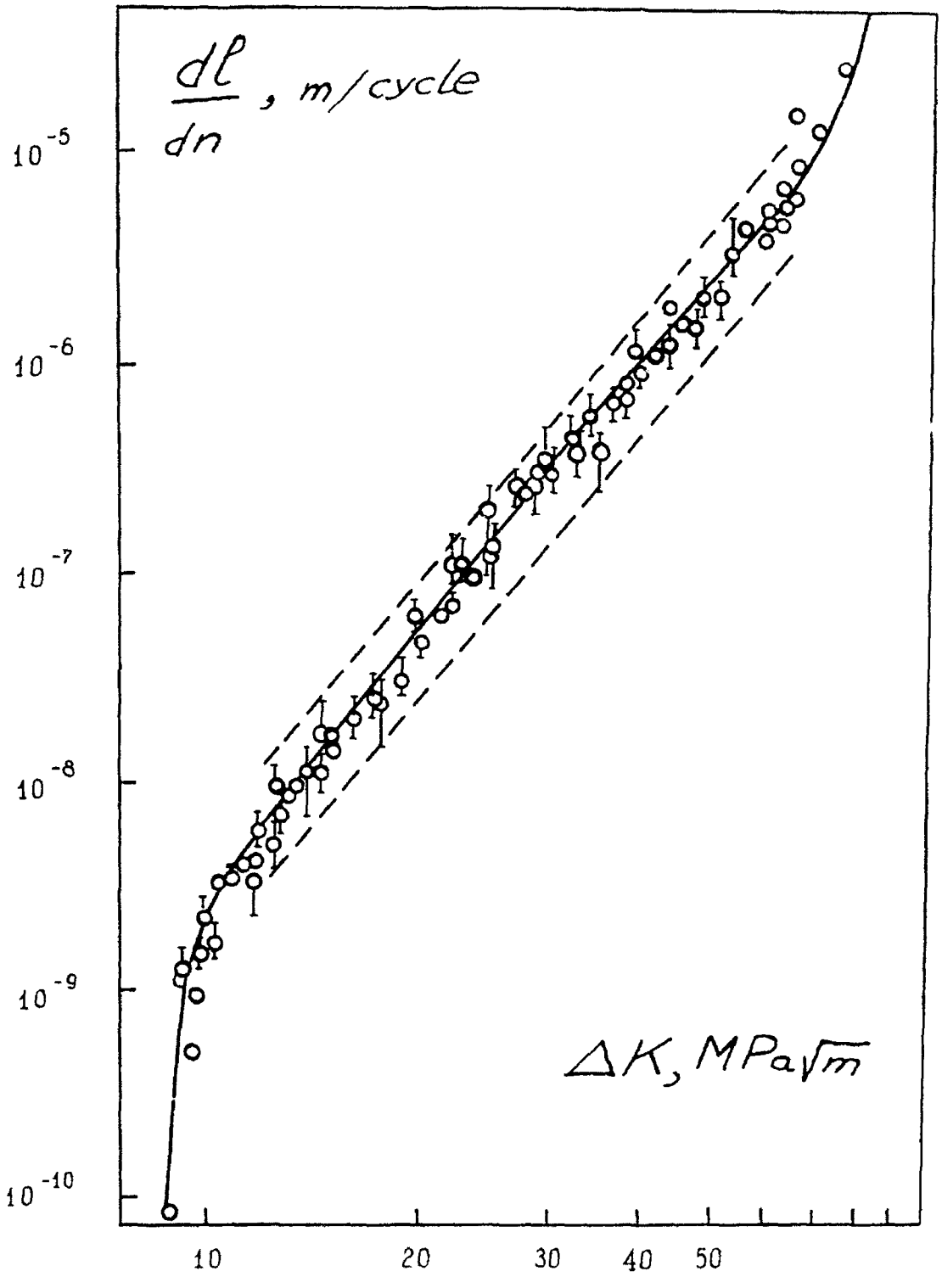


Figure 2. Kinetic diagram for fatigue crack growth in the aluminium alloy BT-3-1 [13].

Substituting (3.8) to (3.6) we obtain exactly the scaling law (3.5) with the result that m depends on z, i.e. the specimen size! Experiments (see Fig. 3) have shown that this dependence is often very strong, so the designers should be careful in using the results of the experiments with small specimens in predicting the life-time of large structures.

Two important notes should be added here. In fact, in the middle of the kinetic diagram the mechanism of fracture starts to change; traces of static fracture modes such as dimples appear on the fracture surface. It is connected quantitatively with reaching some critical stress intensity factor K_{GY} equal in order of magnitude to $\sigma_y\sqrt{d}$, where d is the microstructural length-size, e.g. the grain-size. Therefore, the Paris-Erdogan kinetic curve apparently consists of two sections corresponding to $\Delta K >> K_{GY}$ and $\Delta K << K_{GY}$.This was not properly identified apparently because the values of m corresponding to the two branches of the diagram were sufficiently close.

The second note is that sometimes (it was noticed with the aluminium alloys and required careful investigation with good statistics) the kinetic diagrams are not single-valued curves (Fig. 4), and there exists a rather large overlapping range (see [13]). Fracture surface analysis shows that the upper and lower branches of this non-single-valued kinetic curve correspond to different fracture mechanisms. Apparently these curves explain the remarkable effect noticed by Forsyth [16] - the alternating of smooth and rough regions at the surface of fatigue cracks.

B. The modern tendency of using high strength materials has advanced the problems of multiple fracture. The statistical approach cannot be avoided here, although it is not phenomenological statistics in the Weibull spirit of the late thirties. A combination of the statistical approach with modern fracture mechanics is needed here.

In some conceptual aspect, an analogy with developed turbulence was found to be fruitful. In the turbulence phenomenon we have the fluid instead of the deformable solid, the vortices instead of cracks and defects. The turbulent flow contains a cascade of interacting vortices of various length scales. The motion of a certain small number of vortices in the fluid can be studied more or less effectively as well as the evolution of a small number of cracks or defects in a solid under varying load or strain. Nothing can be performed effectively if there are multiple vortices in a fluid flow or multiple defects or cracks in a deformable solid. So, everything really achieved in turbulence, and the famous Kolmogorov-Obukhov theory first and foremost, is based on the statistical approach and similarity principles for the vortex cascade. In the studies of multiple fracture this approach is also followed.

We assume, therefore, the physical scheme of multiple fracture in the following way. There exists in a deformable solid a cascade of crack-like defects, pores, vacancies, dislocations, etc. which are interacting. The interaction of the defects could be, in principle, of two kinds: defects can either strengthen of weaken each other. Furthermore, the stress concentration near the tips of large defects stimulates the generation of new small defects. On the other side, small defects stimulate the extension and coalescence of large ones.

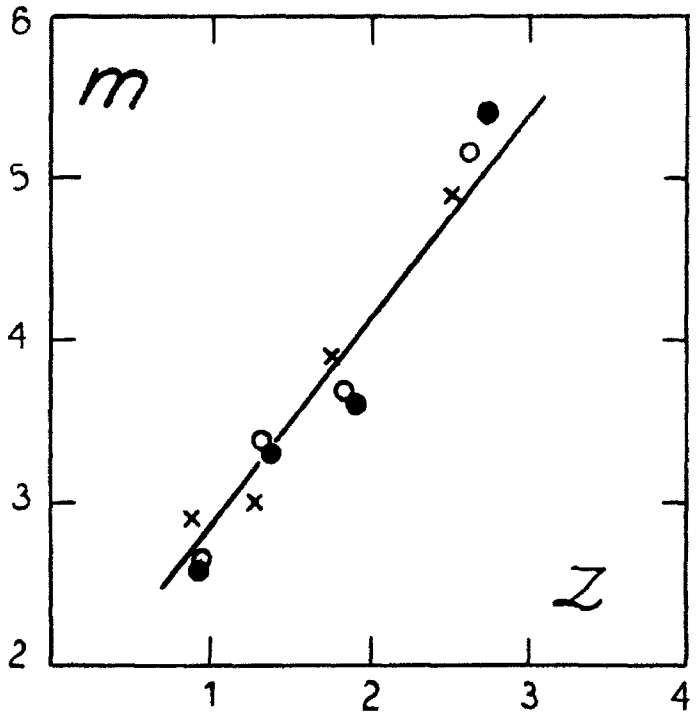


Figure 3. The dependence of the exponent in the Paris-Erdogan law (3.5) on the similarity parameter z for the steel 4340 for specimens of various orientation (A,L,T) with respect to the rolling direction: A(O), L(●), T(x) ([11] on the basis of data [15]).

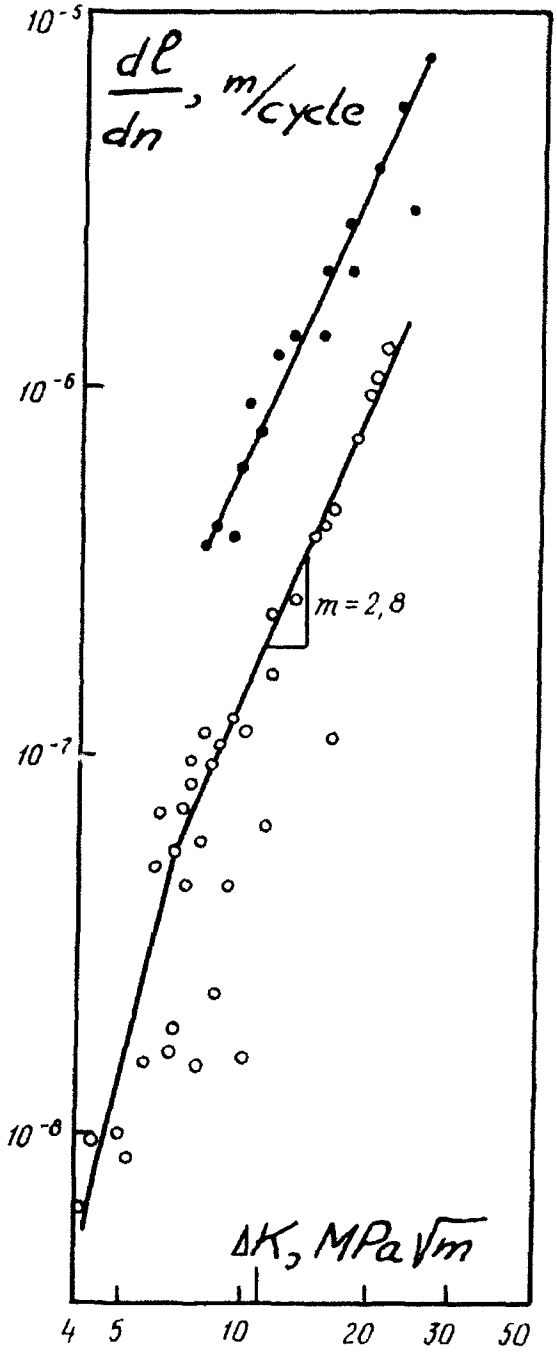


Figure 4. Kinetic diagram of fatigue fracture of Aluminium alloy B95 T1 for notched specimens. Two branches of the diagram with a large overlapping region are clearly seen ([13]).

The basic hypothesis of our approach, the similarity principle [18] is that in multiple fracture the process of the development of the cascade of defects is statistically self-similar. This principle means that in the development of the multiple fracture process, only the dimensional parameters vary with a statistically averaged defect size ℓ_{SC} and mean distance between the defects L_{SC}. The dimensionless statistical properties of the cascade remain invariable.

In particular, the distribution curve of the defects size exists for a cascade. Self-similarity means that when damage accumulation is proceeding, the form of this curve remains fixed and only the maximum frequency N_{SC} and the corresponding mean defect size ℓ_{SC} are varying. Therefore the size distribution can be represented in the following universal form:

$$N/N_{SC} = \Phi(\ell/\ell_{SC}) \tag{3.9}$$

Three basic questions appear in connection with this similarity principle:

I. Does self-similarity of the defects cascade evolution exist for the whole damage accumulation process or at least for a certain part of it?

II. If self-similarity is only observed on certain stages of the multiple fracture process, what is the reason of its loss and of the transition from one self-similar stage to the next one?

III. In general, does self-similarity of the cascade evolution have some practical and/or theoretical interest?

Let us answer these questions successively.

I. The data processing of the defects cascade evolution in damage accumulation has shown that the similarity principle does not contradict at least some parts of the fracture process in creep, fatigue, and static tension for materials of various classes. So, the pore size distribution for iron at various strain levels coincides upon being plotted in universal coordinates (Fig. 5).

Thereby, the pore size distributions at various stages of the damage accumulation process can be obtained, one from another, by a similarity transformation. Only the parameters N_{SC}, ℓ_{SC} are found to be time-dependent. Moreover, the universal size distribution curves of the defects obtained in the same way for different materials tested in creep, fatigue, or static tension turn out to be rather close in spite of the differences in material properties, as well as in the loading conditions, and especially in the length scales - from the atomic level to blocks of rocks the size of tens of miles (Fig. 6). This apparently means that there exists a certain universality in the damage accumulation process, at least for a sufficiently wide class of materials.

II. The multiple fracture process is multistage and multiscale. Apparently, transitions occur from one self-similar stage to the next one. The loss of self-similarity seems to be related to the loss of stability of cascade evolution at a certain critical value of the length size ℓ_{SC}. On the background of smaller defects, a new cascade appears with its own evolution laws and new self-similar stage.

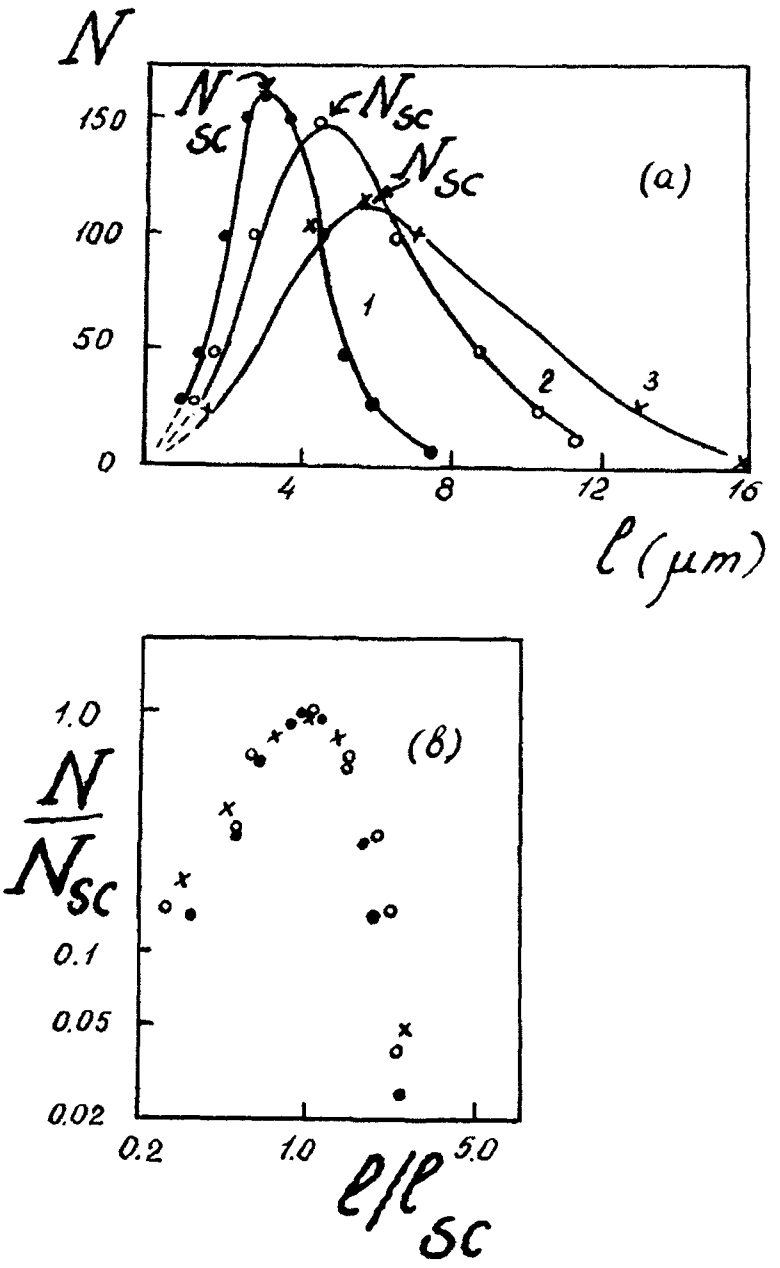


Figure 5. Pore size distributions in natural (a) [19], and universal (b) coordinates for crystalline iron specimens (0.006%C) tested in creep under 9.3 MPa tensile stress and 700°C temperature. 1: ε=2.1%, t=23 hours; 2: ε=6.2%, t=142 hours, 3: ε=9.3%, t=262 hours. N_{SC}: maximum frequency, ℓ_{SC}: corresponding pore size (after [13]).

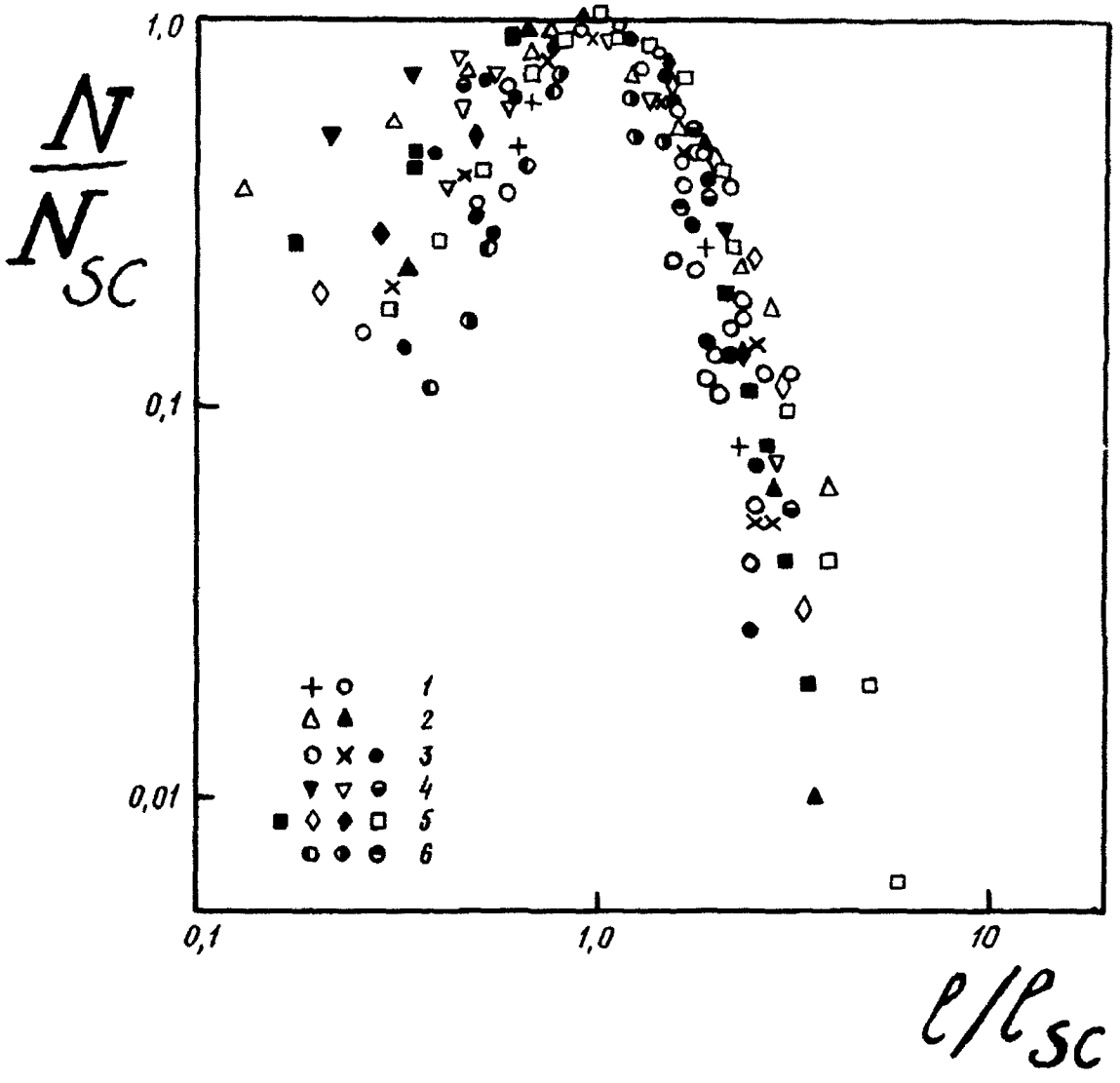


Figure 6: Defect size distribution for various materials under various types of loading. 1 - polycrystalline copper, fatigue; 2 - brass, static tension; 3 - iron, creep; 4 - steel 347, creep; 5 - steel 304, creep; 6 - rocks (after [13]).

III. For a symmetric isolated crack of the length ℓ, the condition for the critical value of the stress intensity factor K at the tip of the crack

$$K=K_0/\sqrt{\pi/2} \qquad (3.10)$$

gives, as is well known, a finite relation between the crack length and applied load. If there are multiple cracks in a body, then a set of 2n finite equations (n -the number of cracks) can be obtained by application of condition (3.10) at each crack tip. This set, however, becomes non-constructive with growing n, while for a rather small number like n=10-20 obtaining the solution becomes impossible even using modern computers. The statistical self-similarity of the cascade of defects returns us to the simplest case analogous to that of an isolated symmetric crack. Indeed, the finite relation exists between the average crack size in the cascade ℓ_{SC} and the applied load. Moreover, the principal possibility appears to be to estimate the loading capacity as well as the life-time of a structure using the parameter ℓ_{SC} and material properties.

The Paris-Erdogan scaling law (3.5), considered above in some detail, also follows ultimately from the self-similarity of the cascade of defects in the vicinity of the crack head. Moreover, from the self-similarity hypothesis of the cascade of defects, the Coffin-Manson equation for low-cycle fatigue, which is well known to experimentalists,

$$n_f[\Delta e](\Delta\varepsilon)^{\alpha} = \text{Const} \qquad (3.11)$$

can be obtained by a simple analysis [12]. Here $n_f[\Delta\varepsilon]$ is the number of cycles to failure at a given strain amplitude $\Delta\varepsilon$, α - a material constant.It is interesting that the well-known Palmer-Minor damage accumulation rule at variable strain amplitudes:

$$\int_0^n dn/n_f[\Delta\varepsilon(n)]=1 \qquad (3.12)$$

also follows rigorously from the self-similarity hypothesis formulated above.

It is also noted that the principle of statistical self-similarity of the cascade of defects could be a distinctive bridge between phenomenological strength studies and the microscopic studies of dislocations, vacancies, etc. which as yet have no constructive outcome to practical strength analysis. Remember again the analogy with turbulence. In the theory of turbulence, the methods of practical computation of turbulent flows are always based on the idea of the vortex cascade self-similarity (including various subgrid simulations in numerical models) which allows one to reduce the calculations to solving a rather reasonable system of the equations for turbulent energy density, average vortex size (or some related quantity), etc. A similar approach seems to be possible, in principle, in multiple fracture studies.

4. DAMAGE ACCUMULATION

The study of damage accumulation in a cylindrical specimen under tension, influenced by microstructural effects, is apparently one of the characteristic examples of the micromechanical approach to fracture studies. Here we will discuss a tentative mathematical model of this phenomenon.

The concept of damage was introduced to fracture mechanics by L.M. Kachanov [19,20], F.K.G. Odqvist and J. Hult [21,22]. According to this approach the damage factor ω is introduced, or, equivalently, the continuity factor $\psi=1-\omega$. In the virgin material $\omega=0$, $\psi=1$ and fracture corresponds to $\omega=1$, $\psi=0$.

A physical interpretation of damage at a point is the properly averaged relative amount of broken bonds between microstructural elements of the body. Under the natural assumption that every broken bond excludes a fixed volume of the material from carrying capacity, the simplest interpretation of damage is obtained as:

$$\omega = (S-S_r)/S \tag{4.1}$$

where S is the specimen cross-section area and S_r is the area of the part of the cross-section where the elements are able to carry the tensile load. A kinetic law of damage accumulation is assumed according to [19-22] which can be represented in the form

$$\partial_t\omega = (1/\tau)\, q(\omega,\sigma,T) \tag{4.2}$$

Here t is the time, τ - a certain constant characteristic time of the damage accumulation process, σ - the actual stress in the load-carrying part of the specimen cross-section which is related to the bulk tensile stress σ_0 and damage ω for the case of the simplest damage interpretation (4.1) by an equation,

$$\sigma = \sigma_0/(1-\omega)\ , \tag{4.3}$$

and T is the temperature. The dimensionless function q is assumed to be specified for the given material. We emphasize that stress and temperature dependence of damage accumulation rate is much stronger than the direct dependence on the current damage. As an example, the often used Arrhenius-type kinetic law with stress-dependent activation energy can be considered,

$$q=(1-\omega)^p\exp[-U-\gamma\sigma)/kT] \tag{4.4}$$

Here k is the Boltzmann constant, U - the zero-stress activation energy, p - a constant, "reaction order", γ - the other kinetic constant. The quantities U/kT, $\gamma\sigma_0/kT$ are large, of the order of ten or even more, so the stress and temperature dependence of the damage accumulation rate is strong. Power-type kinetic laws are also in use,

$$q=(1-\omega)^p(\sigma/\sigma_0)^n \qquad (4.5)$$

where p, n are dimensionless constants and n is assumed to be large.

For fixed constant temperature, the relation (4.2) can be considered as an ordinary differential equation for the damage ω as a function of time. An important point is that its solution exists only for a finite time. When the value ω becomes equal to one, the solution of the ordinary equation (4.2) ceases to exist which defines the specimen life-time, t_0. It is noted that due to initial damage inhomogeneity, the current damage distribution can be nonuniform over the specimen length. According to the approach formulated above, damage accumulation would proceed independently in each cross-section over the specimen length. Thus rupture and life-time of a specimen will be determined by the ultimate damage value $\omega=1$ being reached in the initially most damaged cross-section without any influence from the neighbouring cross-sections.

We will see that the influence of the material microstructure can drastically change the statement of this problem. Namely, a specific non-linear diffusion-type mechanism for damage transfer appears. Therefore, the ordinary differential equation (4.2) is replaced by a non-linear parabolic partial differential equation with a diffusion coefficient that is strongly stress dependent.

Indeed, real materials: metallic alloys, ceramics, composites, rocks, etc. always possess a particular microstructure. In polycrystalline metals, the microstructure is formed by the grains and intercrystalline matter, while a supramolecular structure of various orders exists in polymers, etc. This microstructure is microinhomogeneous; we have, so-to-say, "microstiffeners" and "microweakeners" distributed over the material. Damage accumulation consists in breaking bonds between the elements of microstructure and/or inside these elements and subsequent formation and coalescence of defects. An instructive example is the cavitation and the formation of micro-cracks at the grain boundary facets studied by V. Tvergaard [23,24]. Due to microinhomogeneity of the microstructure (different grain sizes, and grain properties), this damage accumulation process becomes microinhomogeneous. Moreover, the breakthrough of cracks at the grain boundary facets increases the local stresses arising under the action of the bulk tensile stress. At the same time there is a reduction of the local damage accumulation rate in stronger places (a compressive stress due to local "microstiffening"). Thus, in addition to the acting bulk tensile stress, a system of self-balanced microstresses appears.Therefore the stresses in an idealized microscopically homogeneous material differ from those really influencing the local damage accumulation rate.

Thus, consider a prismatic bar under static tension σ_0 and isothermal conditions. Microscopic damage distribution and microstress fields are stochastic. Both fields are considered as statistically homogeneous over the cross sections, although in general, non-homogeneous over the bar length and time-varying.

By averaging relation (4.2) which is valid for ideally homogeneous conditions over the cross-section, we obtain

$$\partial_t<\omega>=(1/\tau)[(1/S)\int_S q(\omega,\sigma,T)dS)] \ , \qquad (4.6)$$

where the brackets <>, which we will further omit, denote averaging of the ensemble. (Due to the ergodicity hypothesis which is natural here, we can replace averaging over the cross-section by ensemble averaging). However, the right-hand side cannot be replaced by the damage accumulation rate corresponding to the average stress $\sigma_0/(1-\omega)$ as it should be in the microhomogeneous case; there exists a certain additional microstress field and the stress dependence of the damage accumulation rate is strong.

To calculate the right-hand side of (4.6), we will use the approach often used after Khinchin and Kolmogorov in the theory of stochastic processes. Indeed, the right-hand side of (4.6) will be replaced by an integral over the bar length by introducing a certain weight function $\rho(x-\xi)$:

$$(1/S)\int_S q(\omega,\sigma,T)dS=\int_{-\infty}^{\infty} q(\omega,\sigma(\xi),T)\rho(x-\xi)d\xi \qquad (4.7)$$

Here the weight function $\rho(x-\xi)$ determines the relative amount of elements of the cross-section area where the damage accumulation rate corresponds to the actual average stress in a neighbouring cross-section having a certain displaced coordinate ξ; this always can be done. We used here essentially the fact that the average damage ω and consequently the average actual stress are the fields continuously varying over the bar length so that the actual stress at a certain point of the cross-section, increased or reduced by micro-stresses or microinhomogeneity, corresponds to the average stress in a certain neighbouring cross-section. Furthermore, under the assumption that the damage accumulation rate is strongly stress dependent, the weight function $\rho(x-\xi)$ is rapidly decreasing with growing modulus of its argument. Therefore, the limits in the right-hand side integral of (4.7) can be taken to be infinite.

Obviously, the weight function $\rho(x-\xi)$ satisfies the relation

$$\int_{-\infty}^{\infty}\rho(x-\xi)d\xi=1 \ , \qquad (4.8)$$

by consideration of an ideally microhomogeneous case. For the sake of some simplification, we also assume that the weight function is symmetric, so that

$$\int_{-\infty}^{\infty}(x-\xi)\,\rho(x-\xi)d\xi=0 \ . \qquad (4.9)$$

Furthermore, we introduce the microstructural length-scale λ by the relation

$$\lambda^2=(1/2)\int_{-\infty}^{\infty}(x-\xi)^2\rho(x-\xi)dx \qquad (4.10)$$

It is noted that we neglected the dependence of the weight function on the current damage for simplicity sake since otherwise λ would be a function of the current damage.

Now we expand the function $q(\omega,\sigma(\xi),T)$ under the integral sign in the right-hand side of (4.7) into the Taylor series near the point $\xi=x$:

$$q(\omega,\sigma,(\xi),T) = q(\omega,\sigma,(x,t),T) + (\xi-x)(\partial_\xi q)_{\xi=x} +$$

$$+\frac{(x-\xi)^2}{2}(\partial^2_{\xi\xi}q)_{\xi=x} + \ldots \qquad (4.11)$$

Neglecting the remaining terms which can be properly justified and substituting the expansion (4.11) to (4.7), using the relations (4.8)-(4.10), and substituting the result in (4.6), we come to the partial differential equation,

$$\partial_t\omega = (1/\tau)q(\omega,\sigma(x,t),T) + \partial_x[(\frac{\lambda^2}{t}\partial_\sigma q)\partial_x s] \qquad (4.12)$$

Obviously damage under ordinary conditions cannot heal, so that equation (4.12) is valid only if its right-hand side is not negative; otherwise it is replaced by

$$\partial_t\omega = 0 \qquad (4.13)$$

Using the relation (4.3) between the average actual stress and the average damage, we obtain from (4.12) an evolution equation for the damage ω:

$$\partial_t\omega =(1/\tau)q(\omega,\sigma_o/(1-\omega),T)+\partial_x[(\frac{\lambda^2}{\tau}\sigma_o\partial_\sigma q)\frac{1}{(1-\omega)^2}\partial_x\omega] \qquad (4.14)$$

Thus, instead of the ordinary differential equation (4.2), a non-linear diffusion-type equation is obtained for the damage ω, where the diffusion coefficient is strongly stress dependent. The resultant mathematical problem is currently fashionable among the constellation of mathematicians concerned with blow-up and quenching problems of non-linear partial differential equations (see the recent publications [25-28]). (Blow-up and quenching correspond to fracture; in fracture problems, however, an additional non-linearity related to condition (4.13) is added.)

In the simplest case of the Arrhenius-type damage kinetics of zero order, equation (4.14) assumes the dimensionless form

$$\partial_\theta \psi = \psi^2 e^\psi + k\psi^2 \partial^2_{\zeta\zeta} e^\psi = I(\psi) \qquad (4.15)$$

if $I \geq 0$, and $\partial_\theta \psi = 0$ if $I < 0$. Here

$$\theta = (t/\tau)\exp(-U/kT), \quad \psi = \mu/(1-\omega), \quad \mu = \gamma\sigma_o/kT$$

$$\zeta = x/L, \quad k = \lambda^2/L^2 \qquad (4.16)$$

where L is the characteristic length-size of the initial damage distribution over the bar length.

The initial condition takes the form

$$\psi(\zeta, 0) = \mu/(1-\omega_o(\zeta)) \qquad (4.17)$$

where $\omega_o(\zeta)$ is the initial damage distribution over the specimen length. The boundary conditions are taken in the form of vanishing damage flux at the boundaries:

$$\partial_\zeta \psi(0, \theta) = 0, \quad \partial_\zeta \psi(1, \theta) = 0 \qquad (4.18)$$

which corresponds to a periodic damage distribution of wavelength L or strongly reinforced specimen grips. The solution should be obtained in the region

$$0 \leq \zeta \leq 1, \quad 0 \leq \theta \leq \theta_o \qquad (4.19)$$

Here θ_o corresponds to the blow-up (fracture) time where the solution ψ at a point becomes infinite. The specimen life-time is determined by the equation

$$t_o = \theta_o \tau e^{U/kT} \qquad (4.20)$$

It is interesting to consider for demonstration purposes the results of the numerical integration of the problem (4.15), (4.17) (4.18)*, presented in Figs. 7,8.

In Fig. 7a, the process of damage accumulation for extremely low $k = 10^{-5}$ is presented. As we see, no damage diffusion-extension of the damaged region is observed. In Fig.7b ($k = 10^{-3}$), the extension of the damage region is clearly observed, and it is seen that damage accumulation in the initially most damaged cross-section commences only after noticeable extension of the damaged region which occurs at a time which is one order of magnitude larger than the life-time without

*Numerical calculations were performed by Dr. V.M. Prostokishin.

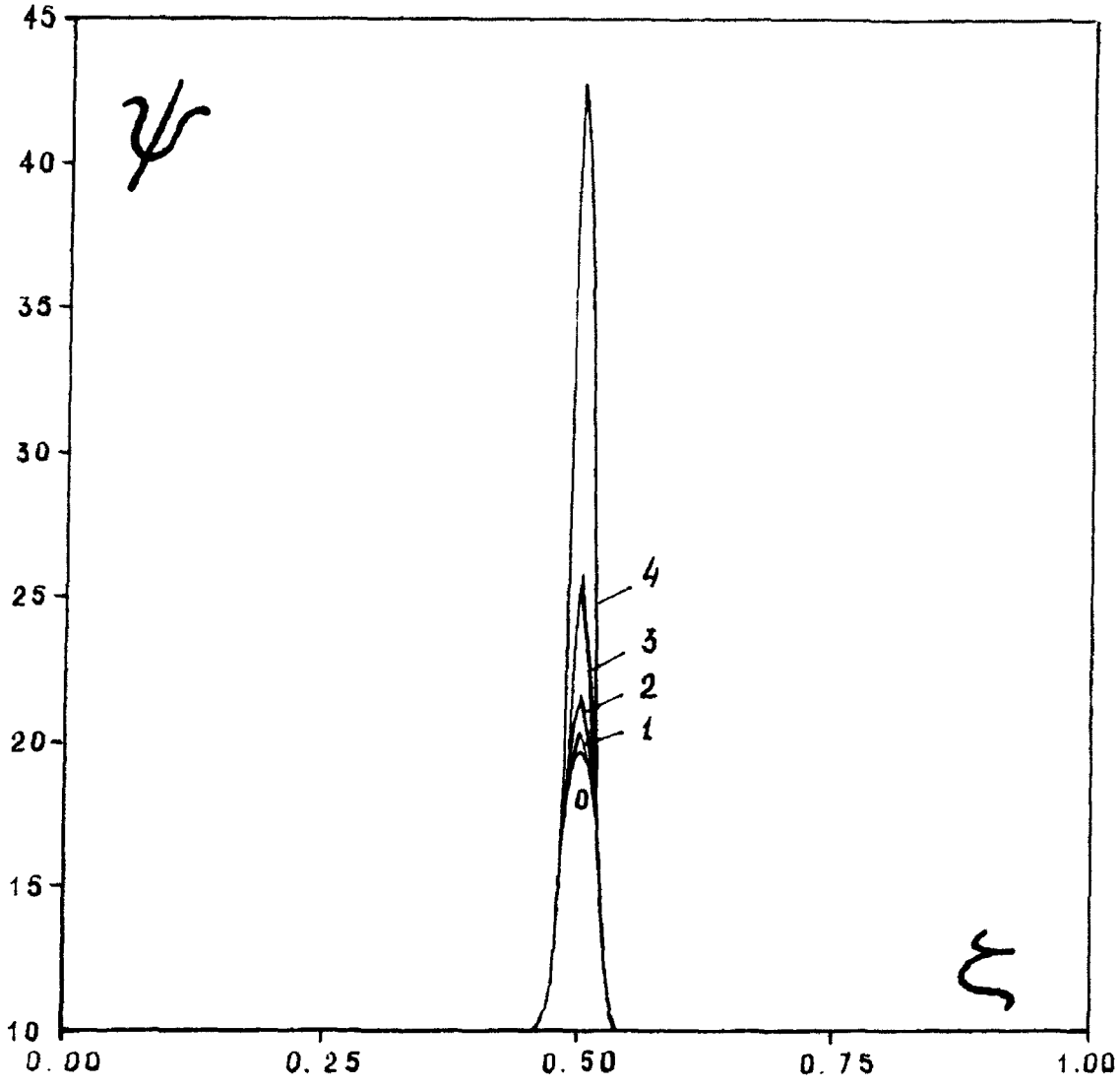


Figure 7a. Damage evolution at various values of dimensionless diffusion parameter k. [$\mu=\gamma\sigma_0/kT=10, \theta^*_0=4.7\cdot10^{-2}$ - dimensionless life-time at k=0.]; k=10^{-5}, (0)-θ=0; (1)-θ/θ^*_0=0.55; (2)-θ/θ^*_0=1.05; (3)-θ/θ^*_0=1.29; (4)-θ/θ^*_0=1.31.

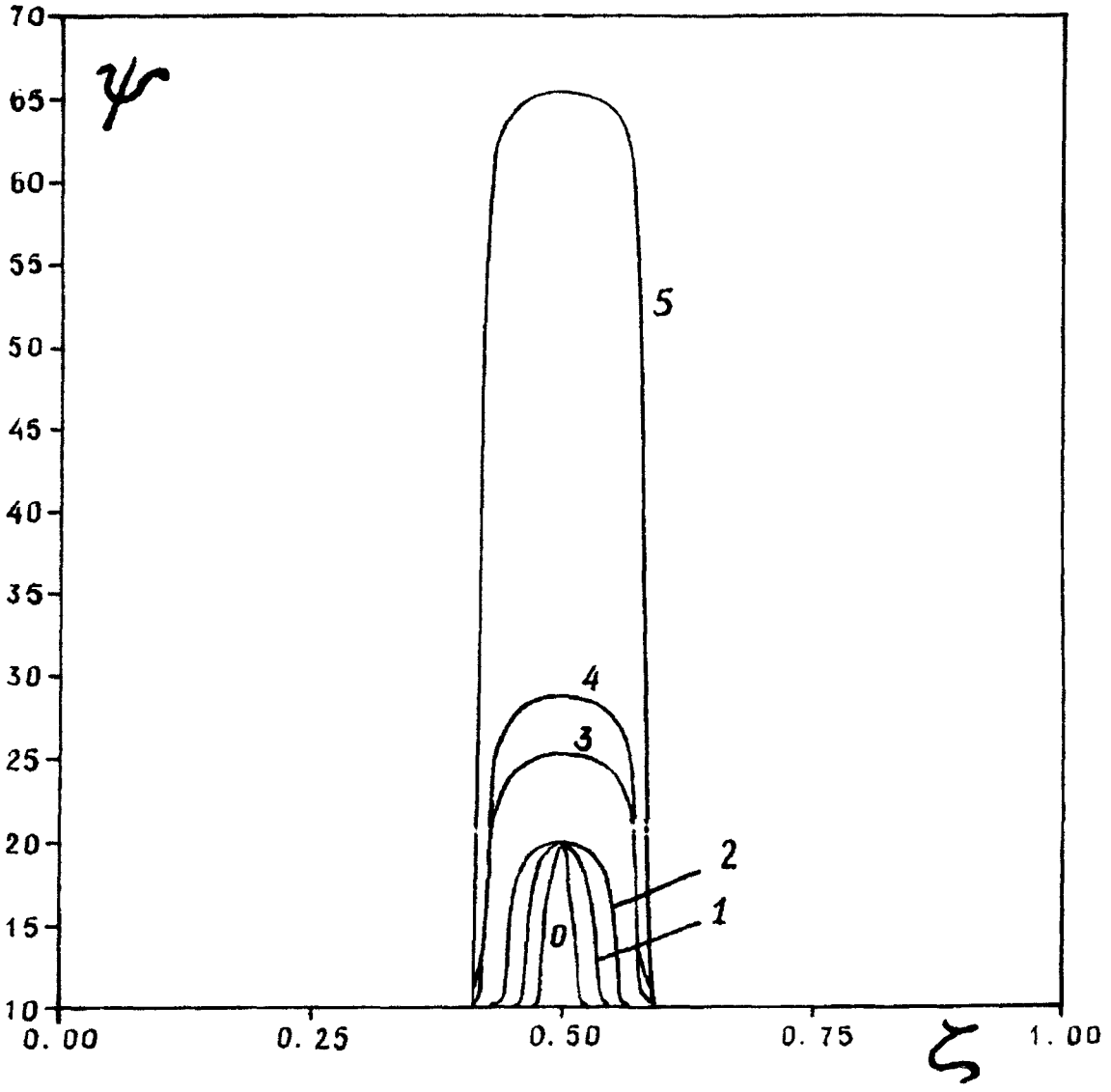


Figure 7b. Damage evolution at various values of dimensionless diffusion parameter k. [$\mu=\gamma\sigma_0/kT=10, \theta^*_0=4.7\cdot10^{-2}$ - dimensionless life-time at k=0.]; $k=10^{-3}$, (0)-$\theta=0$; (1)-$\theta/\theta^*_0=4.0$; (2)-$\theta/\theta^*_0=16.0$; (3)-$\theta/\theta^*_0=24.32$; (4)-$\theta/\theta^*_0=24.33$; (5)-$\theta/\theta^*_0=24.34$.

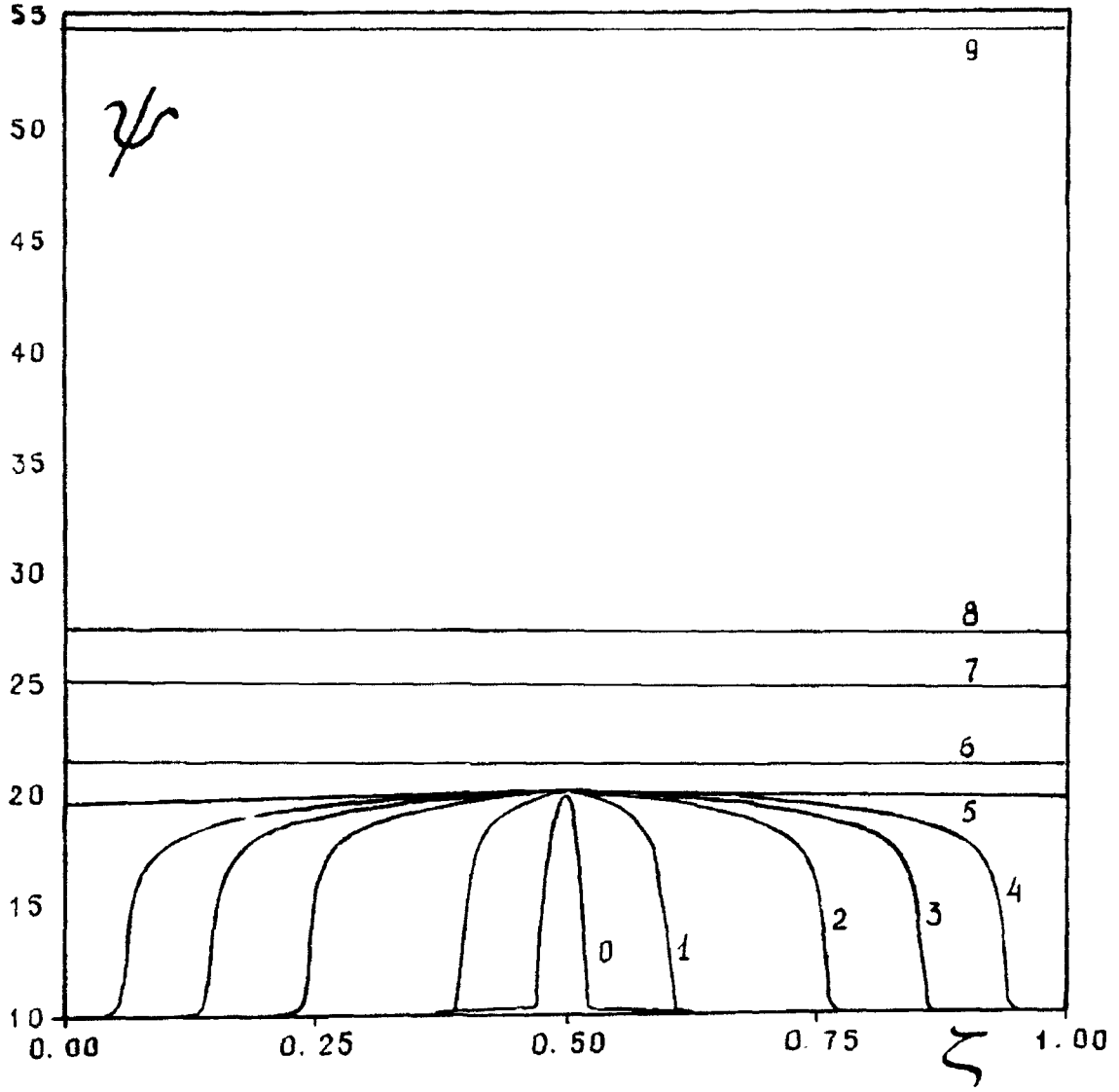


Figure 7c. Damage evolution at various values of dimensionless diffusion parameter k. [$\mu=\gamma\sigma_0/kT=10, \theta^*_0=4.7\cdot10^{-2}$ - dimensionless life-time at k=0.]; $k=10^{-1}$. (0)-$\theta=0$; (1)-$\theta/\theta^*_0=1.0$; (2)-$\theta/\theta^*_0=6.0$; (3)-$\theta/\theta^*_0=11.0$; (4)-$\theta/\theta^*_0=16.0$; (5)-$\theta/\theta^*_0=21.0$; (6)-$\theta/\theta^*_0=28.0$; (7)-$\theta/\theta^*_0=29.25$; (8)-$\theta/\theta^*_0=29.26$; $\theta/\theta^*_0=29.27$

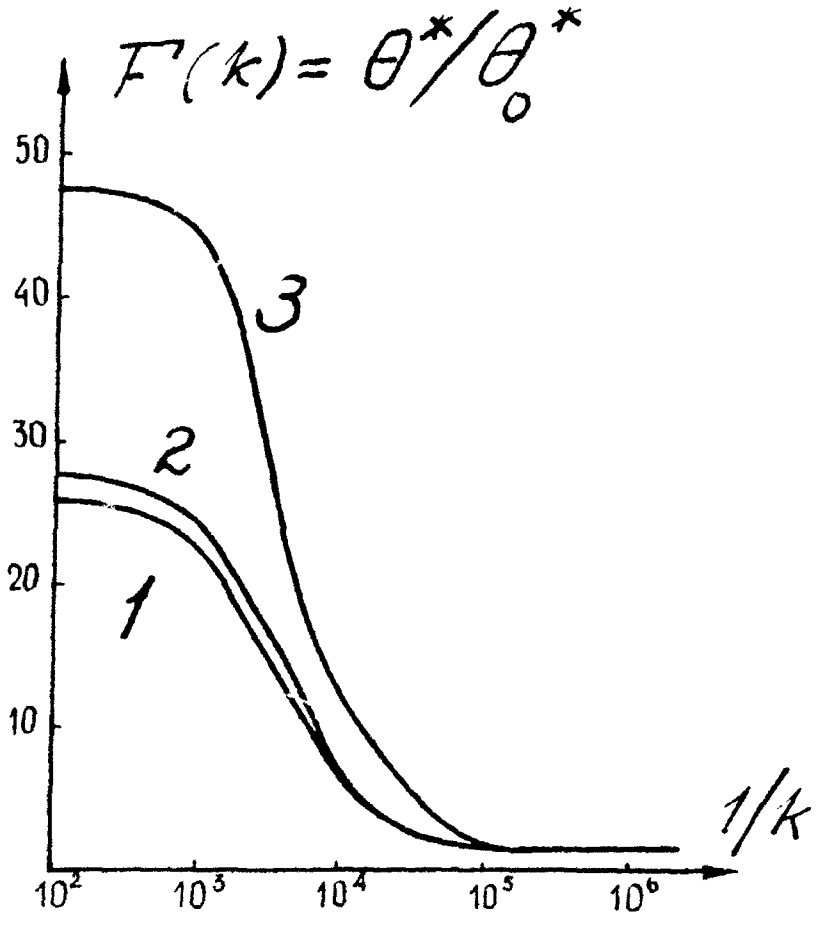


Figure 8. Graph of the function $\theta^*/\theta_0^*=F(k)$, dimensionless life-time ratio for positive k, and k=0. Existence of a critical value of k separating two regimes is seen. Curve (1) corresponds to p=1, $\mu=\gamma\sigma_0/kT=10$; curve (2) corresponds to p=0, $\mu=10$; curve (3) corresponds to $\mu=20$, p=0 and p=1 (the curves for both values practically coincide).

damage diffusion. Figure 7c ($k=10^{-1}$) is instructive: initially, damage uniformly covers the specimen, and then damage accumulation proceeds uniformly over the specimen. Note the damage wave propagation (curves 2-4). Figure 8 demonstrates the increase of the specimen life-time with increasing dimensionless diffusion factor k.

It is plausible that for large stress dependence of the accumulation rate there exists a critical value of the diffusion parameter k_{cr}. For $k<k_{cr}$, damage accumulation proceeds non-uniformly and rupture occurs in the cross-section that has the largest initial damage. For $k>k_{cr}$ damage accumulation proceeds in the following way: initially damage becomes uniformly distributed over the specimen and this process takes the major part of the specimen life-time. Upon reaching uniform damage distribution, damage accumulation continues uniformly over the specimen length subsequently leading to fracture.

The tentative model of damage accumulation was considered here mainly to demonstrate the approach and to discuss an apparently new phenomenon - stress-influenced damage diffusion. I would like to repeat here a sentence from a very comprehensive recent review by A. Needleman and V. Tvergaard [29] devoted to plastic flow localization in metals: "Now everybody loves a localization problem!" Localization of plastic deformation into shear bands in ductile metals is, indeed, a fascinating problem.The remarkable plenary lecture by J. Rice [30] at XIV ICTAM, the book of J. Hutchinson [31], the papers of the author of this review as well as those of other colleagues referenced here form really fascinating contributions. The approach, however, is based entirely on the constitutive equations. In principle, this problem is close to the problem of damage accumulation, and I sincerely hope that the love of localization problems will materialize also in considering them by the approach outlined above.

Ladies and Gentlemen, I come to the conclusion. You could have heard in recent years the opinion that fracture mechanics, while being so fashionable in former times, is now exhausted, that there is a crisis in fracture mechanics, that it is unable to answer important questions of engineers, etc.

I do not think so. Like a beautiful lady in her Balzac age, fracture mechanics has her past deserving to be remembered and her promising future. She attracts both young and mature people, and of course has problems of development. I think that the micromechanical approach which is now in development by many research groups throughout the world will help to answer the existing questions and will propose new fascinating challenges. Moreover, there is in this field some flavor of close discoveries, - discoveries important not only for science and technology but for everyday life of everybody.

REFERENCES

1 Th. von Kármán, *From Low Speed Aerodynamics to Astronautics*, Pergamon Press, Oxford, 1963, pp. 29-46.

2 G.K. Batchelor, Development in Microhydromechanics, in *Theoretical and Applied Mechanics*, W.T. Koiter (ed.), North-Holland, The Netherlands, 1976, pp. 33-55.

3 B. Budiansky, Micromechanics, Computers and Structures, **16**, No 1-4 (1981) pp. 3-12.

4 B. Budiansky, Micromechanics, II, *Proceedings of the 10th US National Congress of Applied Mechanics*, Austin, TX, (1986) pp. 1-8.

5 B. Mandelbrot, *Fractals: Form, Chance, and Dimension*, W.H.Freeman and Company, San Francisco, 1977.

6 B. Mandelbrot, *The Fractal Geometry of Nature*, W.H. Freeman and Company, San Francisco, 1982.

7 B. Mandelbrot, D.E. Passoja, and A.J. Paullay, Fractal character of fracture surfaces of metals, Nature, **308** (1984) pp. 721-722.

8 V.K. Horvath, H.J. Herrmann, The fractal dimension of stress corrosion cracks, Chaos, Solitons and Fractals, **1**, No. 5 (1991) pp. 395-400.

9 R.H. Dauskardt, F. Haubensack, R.O. Ritchie, On the interpretation of the fractal character of fracture surfaces, Acta Metall. Mater. **38**, No. 2 (1990) pp. 143-159.

10 P.C. Paris, F. Erdogan, A critical analysis of crack propagation laws, J. Basic Engineering, Trans. Am. Soc. Mech. Engrs. Ser. D, **85** (1963) pp. 528-534.

11 G.I. Barenblatt, L.R. Botvina, Incomplete self-similarity of fatigue in the linear range of crack growth, Fatigue of Engineering Materials and Structures, **3** (1981) pp. 193-212.

12 G.I. Barenblatt, L.R. Botvina, Application of the similarity method to damage cumulation and fatigue crack growth studies, In *Defects and Fracture*, Proc. Intern. Symposium of Defects and Fracture, Tuozno, ?., Sih, C. and Zorski, H. (Eds.), Nijhoff, Boston, 1982, pp. 71-79.

13 L.R. Botvina, The kinetics of fracture of structural materials, Nauka, Moscow, 1989 (in Russian).

14 G.I. Barenblatt, *Similarity, Self-Similarity, and Intermediate Asymptotics*., Consultants Bureau, New York, London, 1979.

15 G.I. Barenblatt, *Dimensional Analysis*, Gordon and Beach, New York, London, 1987.

16 F.A. Heiser, W. Mortimer, Effects of thickness and orientation on fatigue crack growth rate in 4340 steel, Met. Trans. **3** (1972) pp. 2119-2123.

17 P.J.E. Forsyth, Some observations and measurements on mixed fatigue-tensile crack growth in aluminium alloys, Scr. Met. **10**, No. 6 (1976) pp. 383-386.

18 G.I. Barenblatt, L.R. Botvina, Self-similarity of the fatigue fracture; the damage accumulation, Bull. USSR Ac. Sci., Mechanics of Solids, No. 2 (1983) pp. 88-92.

19 B.J. Cane, G.W. Greenwood, The nucleation and growth of cavities in iron during deformation at elevated temperatures, Metal Sci., **9**, No. 2 (1975) pp. 55-60.
20 L.M. Kachanov, On the life-time under creep conditions, Bull. USSR Ac. Sci., Division of Technical Sciences, No. 8 (1958) pp. 26-31.
21 L.M. Kachanov, Rupture time under creep conditions, In: *Problems of Continuum Mechanics*, J.R.M. Radok, (Ed.), SIAM, Philadelphia, 1961, pp. 202-218.
22 F.K.G. Odqvist, J. Hult, Some aspects of creep rupture, Arkiv for Fysik, **19**, No. 26 (1961) pp. 379-382.
23 F.K.G. Odqvist, *Mathematical Theory of Creep and Creep Rupture*, Clarendon Press, Oxford, 1966.
24 V. Tvergaard, Micromechanical modelling of creep rupture, Z. Angew. Math. Mech. **70**, No. 4 (1991) T23-T32.
25 E. van der Giessen, V. Tvergaard, A creep rupture model accounting for cavitation at sliding grain boundaries, Int. J. of Fracture, **48** (1991) pp. 153-178.
26 A.A. Samarksy, V.A. Galaktionov, S.P. Kurdyumov, A.P. Mikhailov, Blow-up regime in problems for quasilinear parabolic equations, Nauka, Moscow, 1987, (in Russian).
27 H. Brezis, F. Merle, Uniform estimates and blow-up behavior for solutions of $-\Delta u=V(x)e^{u}$ in two dimensions, Publication du Laboratoire d'Analyse Numerique, Université Pierre et Marie Curie, Paris, 1991.
28 Y. Giga, R.V. Kohn, Nondegeneracy of Blow-Up for Semilinear Heat Equations, Communications on Pure and Applied Mathematics, XLII, (1989), pp. 845-884.
29 H.A. Levine, Quenching, non-quenching, and beyond quenching for solutions of some parabolic equations, Annali di Matematica Pura ed Applicata (IV), CLV, (1989), pp. 243-260.
30 A. Needleman, V. Tvergaard, Analysis of plastic flow localization in metals, Applied Mechanics Reviews, **45**, S3-S18, 1992.
31 J.R. Rice, The localization of plastic deformation, In: *Theoretical and Applied Mechanics*, Koiter, W.T. (Ed.), North-Holland, The Netherlands, 1976, pp. 207-220.
32. J.W. Hutchinson, Micromechanics of damage in deformation and fracture, Published by the Technical University of Denmark, Lyngby, 1987, Ch. 6, pp. 51-64.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Material Instabilities and phase transitions in thermoelasticity

Rohan Abeyaratne

Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

Abstract

This paper reviews recent joint work with J. K. Knowles on the propagation of phase boundaries in a two-phase thermoelastic solid. We address some general thermo-mechanical issues, present an explicit model, compare the predictions of the model with experimental observations, consider the influence of inertia, and describe the role of viscous and strain-gradient effects. This paper is based on the work described in [2,3,4,7].

1. INTRODUCTION

Certain alloys can exist in more than one solid phase, each phase being associated with a distinct crystal structure. Typically, one phase is preferred under certain conditions of stress and temperature, while another is favored under different conditions. As the stress or temperature vary, the material may therefore transform from one phase to another, leading to a discontinuous change in the properties of the body. Examples of such materials include the shape-memory alloy $NiTi$, the ferroelectric alloy $BaTiO_3$, the ferromagnetic alloy $FeNi$ and the high-temperature superconducting ceramic alloy $ErRu_4B_4$. In each of these examples, the transition from one phase to another occurs without diffusion and one speaks of the transformation as being *displacive*.

Consider for instance the class of Cu-Al-Ni alloys described by Otsuka and Shimizu [36]. These alloys exist in a cubic phase (austenite) and an orthorhombic phase (martensite). Under stress-free conditions, the cubic phase is stable at temperatures above the transformation temperature θ_T, the orthorhombic phase is stable below it. When a stress-free specimen of austenite is cooled from above the transformation temperature, the martensite phase is nucleated at the "martensite start temperature" M_s ($\leq \theta_T$). The specimen now consists of a mixture of both austenite and martensite, and the phases are separated from each other by one or more interfaces - *phase boundaries*; one speaks of the phase boundaries as being coherent, implying that the deformation is continuous across them even though the deformation gradient is not. The phase boundaries that have nucleated now propagate, transforming particles from austenite to martensite as they pass through, eventually converting the entire specimen into martensite. If the specimen is now heated, the whole process is reversed, with the martensite to austenite transformation being initiated at the "austenite start temperature" A_s ($\geq \theta_T$). This is a *reversible*

or *thermoelastic* phase transition. The values of the transformation temperature θ_T and the nucleation temperatures M_s and A_s depend critically on the alloy composition and the heat treatment; they also vary with the application of stress, e.g. see Figure 10 in [36].

In the case of Cu-Al-Ni, there are in fact a number of different types of martensite, distinguished from each other by a difference in stacking order, see Figure 15 in [36]. Consider for example the two martensites γ_1' and β_1'. The transformation from austenite $\rightarrow \gamma_1'$ martensite occurs under either stress-free conditions or at small values of applied stress, while the austenite $\rightarrow \beta_1'$ martensite transformation occurs only at sufficiently large values of stress. More importantly, a phase boundary in the former case typically propagates *slowly*, while a phase boundary between austenite and β_1' martensite propagates *rapidly* at speeds which are of the order of the shear wave speed. Thus both slow and fast-moving phase boundaries occur, and both are of interest; see also Grujicic *et al.* [22] and Nishiyama [34].

Various continuum-level issues related to reversible phase transformations in crystalline solids have been successfully studied using the theory of finite thermoelasticity, e.g. Ericksen [17,18], James [25] and Pitteri [39]. For a thermoelastic material, the Helmholtz free-energy function ψ depends on the deformation gradient tensor $\mathbf{F}$ and the temperature θ: $\psi = \hat{\psi}(\mathbf{F},\theta)$; since the occurrence of diffusion is not envisaged here, it is not necessary to explicitly account for the dependency on alloy composition. If the stress-free material can exist in two phases, then the energy function $\hat{\psi}$ must have two disjoint energy-wells, each well corresponding to one phase. At $\theta = \theta_T$, the two minima have the same height; for $\theta > \theta_T$ the austenite minimum is smaller, while for $\theta < \theta_T$, the martensite minimum is smaller. In the presence of stress $\mathbf{S}$, the situation may be described in similar terms except that now one must consider the energy-wells of the potential energy function $G(\mathbf{F};\mathbf{S},\theta)$ where $\mathbf{S}$ denotes the first Piola-Kirchhoff stress tensor.

Much recent activity in continuum mechanical studies on thermoelastic phase transitions has been focussed on two basic issues: the first concerns *energy minimizing deformations* corresponding to the stable configurations of a body; the second is related to the non-equilibrium *evolution* of a body towards such stable configurations through intermediate states of metastability.

Ericksen [16] studied energy minimizing deformations for a two-phase material within the one-dimensional mechanical setting of a tensile bar. He showed, in particular, that for certain values of prescribed displacement, the stable equilibrium configurations of the bar involve *co-existent* phases. The analogous issue in higher dimensions is more complicated: typically, deformation gradient tensors $\mathbf{F}_a$ and $\mathbf{F}_m$ corresponding to the austenite and martensite energy minima are not kinematically compatible with each other, i.e. $\mathbf{F}_a - \mathbf{F}_m$ is not a rank-one tensor. Therefore an energy minimizing deformation cannot correspond to homogeneously deformed states of austenite and martensite separated by a phase boundary. In fact, an energy minimizer usually does not even exist, and one must contend instead with minimizing sequences. The deformation pattern associated with

such a sequence characterizes the underlying microstructure of the material; in Cu-Al-Ni for example, an austenite-martensite phase boundary separates a homogeneous state of austenite from a fine mixture of *twinned* martensite. These ideas were put forward by Khachaturyan and Shatalov [27] and Roitburd [41] using a geometrically linear theory, and by Ball and James [8] for the finite deformation theory; see Bhattacharya [10] for a comparison of these two theories. Ball and James [8] studied an austenite/twinned martensite interface in detail, and showed that the consequences of their theory are in agreement with the crystallographic theory of martensite which is a classical theory that is *not* based on the idea of energy minimization. Needle-like microstructures and self-accommodating microstructures have been explored by Bhattacharya [9] using similar ideas. The fineness of the microstructure is controlled by surface energy, though the effects of surface energy may be more subtle than this; e.g. see Gurtin and Struthers [21] and Kohn and Müller [29].

The usual continuum theory of thermoelasticity,though adequate for characterizing two-phase energy minimizers, does not, by itself, characterize quasi-static or dynamic *processes* of a body involving transitions from one phase to another. This is illustrated in a striking manner by the tremendous lack of uniqueness of solution to particular initial-boundary-value problems formulated on the basis of the usual theory; see Abeyaratne and Knowles [1,3].

Quasi-static or dynamic processes generally involve states that are merely metastable and so fall under the category of "non-equilibrium thermodynamic processes." Considerations pertaining to the rate of entropy production during such a process leads naturally to the notion of the *driving force* (or Eshelby force) f acting on a phase boundary, Abeyaratne and Knowles [2] and Heidug and Lehner [24]; see also Eshelby [19], Knowles [28], Rice [40]. The theory of non-equilibrium processes can then be used to argue for the need for a constitutive equation – a *kinetic law* – relating the propagation speed V_n of the phase boundary to the driving force f and temperature θ: $V_n = V(f, \theta)$. The kinetic law controls the rate of progress of the phase transition, as has long been recognized in the materials science literature, see e.g. Christian [13]; in fact, some, though not all, micro-mechanical models of kinetics lead to kinetic laws of the form $V_n = V(f, \theta)$.

The kinetic relation controls the progress of the phase transition once it has commenced. A separate *nucleation criterion* is required to signal the initiation of a transition. (This is analogous to the roles played by a flow rule and a yield criterion in continuum plasticity theory.) Explicit models of nucleation are more complicated to establish. A general discussion of nucleation theory in phase transitions, from a materials science point of view, may be found in Christian [13].

Thus a complete constitutive theory which is capable of modeling processes involving thermoelastic phase transitions consists of three ingredients: a Helmholtz free-energy function, a kinetic relation, and a nucleation criterion. The particular initial boundary-value problems referred to previously, which were ill-posed in the usual setting, were shown in Abeyaratne and Knowles [1,3] to be well-posed and to have unique solutions within this broadened setting.

A number of studies have been concerned with developing explicit constitutive models. For example, Ericksen [18] and Silling [45] have constructed three-dimensional Helmholtz free-energy functions for modeling certain crystals; Falk [20] has studied a one-dimensional polynomial free-energy function, see also Jiang [26]. Models of kinetic relations have been constructed, for example, by Müller and Wilmansky [32] using certain statistical considerations, and by Otsuka et al., [35] by assuming phase-boundary motion to be similar to dislocation motion.

In [7], we constructed a complete, explicit one-dimensional model for describing thermoelastic phase transformations. In constructing our model, we were motivated by a desire to maintain great simplicity, and to construct an analytically amenable model that could be used to gain helpful insights, especially for fast transformations, before undertaking computations based on more realistic assumptions. Our Helmholtz free-energy function has two energy-wells, corresponding to an austenitic phase and a single variant of martensite. Both phases are modeled as having the same constant elastic moduli, coefficients of thermal expansion and specific heats at constant strain; the phases are distinguished from each other by the presence of a transformation strain and a latent heat. We also constructed an explicit example of a kinetic relation based on the notion of thermal activation. Finally, the nucleation criterion considered assumes nucleation to occur at critical values of driving force, one value signalling the initiation of the austenite $\rightarrow$ martensite transition, and another, the reverse transition; this nucleation criterion can be stated equivalently in terms of temperature-dependent stress-levels (or stress-dependent temperature-levels). The quasi-static response of this model to various thermo-mechanical loadings was also studied in [7] and the results were compared qualitatively with some experimental observations as described in Section 4 below.

There is a considerable *mathematical literature* concerned with the issue of identifying meaningful solutions to a system of quasi-linear partial differential equations; e.g. see the review article by Dafermos [15]. In the one-dimensional purely mechanical setting of a tensile bar, the differential equations of interest stem from linear momentum balance and kinematic compatibility. A typical motion of the bar involves moving strain discontinuities corresponding to shock waves, and possibly phase boundaries, and each discontinuity is subject to a pair of jump conditions arising from the aforementioned physical considerations, as well as to the entropy inequality $fV_n \geq 0$ arising from the second law of thermodynamics. If the stress-strain relation is monotonically increasing, and in addition is either strictly convex or strictly concave, phase transformations cannot occur, and all strain discontinuities are shock waves; as in gas dynamics, the entropy inequality then serves to single out unique solutions to the Cauchy problem. The stress-strain curve associated with a two-well energy function is *not* monotonic (it consists of two rising branches that are connected by an intermediate declining branch), and the entropy inequality is no longer sufficient to deliver uniqueness. The standard approach to resolving this difficulty has been to replace the entropy inequality by a stronger "solution-selecting mechanism" or *admissibility criterion*. Among these, the most common in the setting of phase transformations are perhaps the viscosity-strain gradient criterion (Shearer [43], Slemrod [46], Truskinovsky [49]), the entropy-rate shock criterion (Dafermos [15], Hat-

tori [23]), and the chord criterion (Shearer [42], Pego [37]). An *alternative* approach for attempting to achieving uniqueness is to supplement the entropy inequality (which must hold at *all* discontinuities) with a kinetic relation (at subsonic phase boundaries) and a nucleation criterion. We have shown in [3] that within the context of the Riemann problem for a special material, the lack of uniqueness remaining after imposition of the entropy inequality is precisely that needed to accommodate these supplementary ingredients; see also LeFloch [31]. In addition, we have also shown that the three aforementioned admissibility criteria correspond precisely to three particular kinetic relations of the form $V_n = V(f)$, [4,5,6].

The solution to most multi-dimensional problems involving phase transitions must be obtained numerically. Collins and Luskin [14] and Silling [44] have studied energy minimizers using, respectively, finite element and finite difference based methods. Molecular dynamics simulations have also been carried out, e.g. Yu and Clapp [50].

The present paper reviews some of the aforementioned results concerning phase boundary propagation. In Section 2 we describe a general three-dimensional thermomechanical framework which includes the effects of inertia; this leads naturally to the thermodynamic notions of driving force acting on a phase boundary and a kinetic law controlling its propagation. We also describe some characteristics of the potential energy function associated with a two-phase material. From there on we restrict attention to a one-dimensional setting: In Section 3 we describe an explicit constitutive model. Then in Section 4 we describe the quasi-static response of this model in a number of thermomechanical loading programs, including one that displays the *shape-memory effect*. The predictions of the model are compared qualitatively with experimental observations. In Section 5 we consider the effects of inertia in the absence of thermal influences, and elucidate the precise roles of the entropy inequality, the kinetic relation, and the nucleation criterion in the context of a Riemann problem. Finally, in Section 6 we consider the effects of viscosity and strain-gradients; by considering a phase boundary to be a suitable limit of a travelling wave, we show that the presence of these effects leads, in a natural way, to a specific kinetic relation. A more detailed discussion of the material in the first part of Section 2, and all of Sections 3 and 4, may be found in [2,7]; other than for a slight difference in the underlying constitutive model, Sections 5 and 6 are based on [3,4].

2. SOME GENERAL CONSIDERATIONS

Consider a body which occupies a region $\mathcal{R}$ in a reference configuration and consider a motion of this body on some time interval. Let $\mathcal{S}_t$ be a propagating surface in $\mathcal{R}$ across which the deformation is continuous but the deformation gradient tensor $\mathbf{F}$ suffers a jump discontinuity; suppose that $\mathbf{F}$ varies smoothly on either side of $\mathcal{S}_t$. In the present setting, such a surface may represent either a shock wave or an interface that separates two distinct phases of the material. Note that $\mathcal{S}_t$ is a non-material surface, i.e. different particles of the body are located on $\mathcal{S}_t$ at different times.

Let $\mathbf{n}$ denote a unit normal vector and V_n the normal velocity of a point $\mathbf{x}$ on $\mathcal{S}_t$. We shall speak of the side into which $\mathbf{n}$ points as the positive side of $\mathcal{S}_t$. If $p(\mathbf{x}, t)$ is a generic

field quantity that varies smoothly away from $\mathcal{S}_t$ and suffers a jump discontinuity across $\mathcal{S}_t$, we let $\overset{+}{p}(\mathbf{x},t)$ and $\overset{-}{p}(\mathbf{x},t)$ denote the limiting values of p at a point $\mathbf{x}$ on $\mathcal{S}_t$ and write $[[p(\mathbf{x},t)]] = \overset{+}{p}(\mathbf{x},t) - \overset{-}{p}(\mathbf{x},t)$.

Propagation: We now list the various jump conditions that must hold on $\mathcal{S}_t$, using a Lagrangian description throughout. Continuity of the deformation across $\mathcal{S}_t$ implies that

$$[[\mathbf{v}]] + V_n[[\mathbf{F}]]\mathbf{n} = \mathbf{0} \qquad \text{on } \mathcal{S}_t, \tag{2.1}$$

where $\mathbf{v}(\mathbf{x},t)$ denotes particle velocity. If $\rho(\mathbf{x})$ is the mass density of the body in the reference configuration and $\mathbf{S}(\mathbf{x},t)$ is the first Piola-Kirchhoff stress tensor, the balance of linear momentum leads to

$$[[\mathbf{Sn}]] + [[\rho\mathbf{v}]]V_n = \mathbf{0} \qquad \text{on } \mathcal{S}_t. \tag{2.2}$$

Next, let $\mathbf{q}(\mathbf{x},t)$ represent the nominal heat flux vector and $\varepsilon(\mathbf{x},t)$ the internal energy per unit mass; then the first law of thermodynamics requires that

$$[[\mathbf{Sn}\cdot\mathbf{v}]] + [[\rho\varepsilon + \rho\mathbf{v}\cdot\mathbf{v}/2]]V_n + [[\mathbf{q}\cdot\mathbf{n}]] = 0 \qquad \text{on } \mathcal{S}_t. \tag{2.3}$$

Finally, the second law of thermodynamics leads to

$$[[\rho\eta]]V_n + [[\mathbf{q}\cdot\mathbf{n}]]/\theta \leq 0 \qquad \text{on } \mathcal{S}_t, \tag{2.4}$$

where θ is absolute temperature and η is entropy per unit mass. The various physical quantities just introduced are assumed to vary smoothly away from $\mathcal{S}_t$, and with the exception of θ, are permitted to suffer jump discontinuities across $\mathcal{S}_t$. The temperature θ is required to be continuous. (In the event that heat conduction is neglected and the process is modeled as being adiabatic, one must relinquish this assumption on temperature and permit θ to jump across $\mathcal{S}_t$, as in the classical theory of shock waves in gas dynamics.)

The jump conditions (2.1)-(2.4) impose restrictions on the field quantities on the two sides of the surface of discontinuity $\mathcal{S}_t$. When this surface corresponds to a shock wave, there are no further conditions to be imposed on $\mathcal{S}_t$. On the other hand if $\mathcal{S}_t$ describes a phase boundary, particles transform from one phase to another as they cross this moving surface and it is necessary to account for the *process* (i.e. kinetics) of this transition in a suitable manner; in order to do this within the continuum theory, one must supply further information on $\mathcal{S}_t$, in addition to (2.1)-(2.4). The need for such supplementary information can be made clear by considering the physics of phase transitions as described

in the materials science literature, e.g. [13]; or, by solving particular initial-boundary-value problems, and thereby demonstrating the tremendous lack of uniqueness of solution that exists in the absence of such information, [1,3]; or, by the constitutive formalism associated with non-equilibrium thermodynamic processes as outlined below.

Suppose for the remainder of this section that $\mathcal{S}_t$ represents a phase boundary. Consider a sub-region $\mathcal{D}$ of $\mathcal{R}$ and let $\Gamma(t;\mathcal{D})$ be the *rate of entropy production* at time t associated with the particles in $\mathcal{D}$. By using the basic field equations and jump conditions, one can express Γ in the form

$$\Gamma(t;\mathcal{D}) = \int_{\mathcal{D}} \frac{\rho\delta}{\theta} dV + \int_{\mathcal{D}} \frac{\mathbf{q}\cdot Grad\theta}{\theta^2} dV + \int_{\mathcal{S}_t\cap\mathcal{D}} \frac{fV_n}{\theta} dA. \tag{2.5}$$

Here $\delta(\mathbf{x},t)$ is the internal dissipation defined by $\delta = \theta\dot{\eta} + \mathbf{S}\cdot\dot{\mathbf{F}}/\rho - \dot{\varepsilon}$ (see [47]), and $f(\mathbf{x},t)$, defined by

$$f = [[\rho\psi]] - \frac{1}{2}(\overset{+}{\mathbf{S}} + \bar{\mathbf{S}})\cdot[[\mathbf{F}]] \qquad \text{on } \mathcal{S}_t, \tag{2.6}$$

is known as the *driving force* acting on the surface $\mathcal{S}_t$. In writing (2.6) we have introduced the Helmholtz free-energy per unit mass $\psi = \varepsilon - \eta\theta$. In the case of quasi-static processes, one can provide a purely mechanical interpretation for $f\mathbf{n}$ as a force per unit area exerted pointwise on $\mathcal{S}_t$ by the body; this suggests the alternative name *driving traction* for f.

The representation (2.5) for Γ decomposes the total rate of entropy production into the sum of three parts: the first arises from local dissipation in the bulk material away from $\mathcal{S}_t$, the second is the contribution due to heat conduction, and the third represents the entropy production rate arising from the moving phase boundary $\mathcal{S}_t$. Suppose from hereon that the material is *thermoelastic*, i.e. that it is characterized by a Helmholtz free-energy potential $\hat{\psi}(\mathbf{F},\theta)$ such that

$$\psi(\mathbf{x},t) = \hat{\psi}(\mathbf{F}(\mathbf{x},t),\theta(\mathbf{x},t)), \qquad \mathbf{S} = \rho\hat{\psi}_{\mathbf{F}}, \qquad \eta = -\hat{\psi}_{\theta}. \tag{2.7}$$

Then the internal dissipation δ, and therefore the associated entropy production rate, both vanish. If the body is in thermal equilibrium, $\theta(\mathbf{x},t) = constant$, the entropy production rate due to heat conduction vanishes; if the body is in phase equilibrium, by which we mean the driving force $f(\mathbf{x},t)$ vanishes, the entropy production rate due to the moving phase boundary vanishes. Thus, the temperature gradient $Grad\,\theta$ and the driving force f are the "agents of entropy production" when a thermoelastic body is not in thermal and phase equilibrium; one speaks of $(Grad\,\theta)/\theta^2$ and f/θ as thermodynamic "affinities" and the conjugate quantities $\mathbf{q}$ and V_n in the entropy production rate (2.5) as the corresponding "fluxes", e.g. Callen [12], Truesdell [47]. Note that while $\mathbf{q}$ represents heat flux, ρV_n represents the flux of mass crossing the propagating phase boundary.

The presence of non-vanishing affinities $(Grad\,\theta)/\theta^2$ and f/θ indicates a lack of thermal and phase equilibrium; according to the theory of non-equilibrium processes as described in the aforementioned references, the evolution towards a state of equilibrium is controlled by constitutive laws which relate each flux to its conjugate affinity, and possibly to other variables as well. In the present case, this suggests the need for two constitutive relations: $\mathbf{q} = \mathbf{Q}(Grad\,\theta, \mathbf{F}, \theta)$ in the bulk material and

$$V_n = V(f, \theta) \qquad \text{on } \mathcal{S}_t, \tag{2.8}$$

representing a *heat conduction law* and a *kinetic law* respectively. The entropy inequality (2.4) can be written equivalently as

$$fV_n \geq 0 \qquad \text{on } \mathcal{S}_t, \tag{2.9}$$

and so the constitutive function V must be such that

$$V(f, \theta)f \geq 0. \tag{2.10}$$

If V is continuous, this implies that $V(0, \theta) = 0$, indicating that the flux V_n also vanishes when the body is in phase equilibrium.

The basic principles of the continuum theory do not impose any further restrictions on, or provide examples of, the kinetic response function V. These must be supplied by appropriate constitutive modeling, two examples of which are given in Sections 3 and 6. In the special case of processes taking place close to phase equilibrium, it is natural to replace the relation (2.8) by its linear counterpart $V_n = \nu(\theta)f$ with ν positive.

Energy-wells: In general, a propagating surface $\mathcal{S}_t$ could be either a phase boundary or a shock wave. Shock waves are intimately connected with inertia and do not exist in the absence of such effects. In order to focus attention on features that are unique to phase boundaries, it is therefore useful to consider a quasi-static setting. We close this section with some discussion on the characteristics of the potential energy function G,

$$G(\mathbf{F}; \mathbf{S}, \theta) \equiv \rho\hat{\psi}(\mathbf{F}, \theta) - \mathbf{S} \cdot \mathbf{F}, \tag{2.11}$$

associated with a two-phase material. It should be noted that while the Gibbs free-energy $g(\mathbf{F}, \theta) = \rho\hat{\psi}(\mathbf{F}, \theta) - \rho\hat{\psi}_{\mathbf{F}}(\mathbf{F}, \theta) \cdot \mathbf{F}$ is distinct from the potential energy function G, their values coincide at extrema of $G(\cdot\,; \mathbf{S}, \theta)$.

Let $\mathcal{S}_t$ represent a *quasi-statically* propagating phase boundary, and consider, and fix, an instant t and a point $\mathbf{x}$ on $\mathcal{S}_t$. At this $(\mathbf{x}, t)$, let $\mathbf{n}$ be a unit vector normal to $\mathcal{S}_t$, let θ be the temperature, and let $\overset{\pm}{\mathbf{F}}$ and $\overset{\pm}{\mathbf{S}}$ be the limiting values of the deformation

gradient tensor and the first Piola-Kirchhoff stress tensor respectively. Continuity of the deformation requires that $\bar{\mathbf{F}}=\overset{+}{\mathbf{F}}+\mathbf{b}\otimes\mathbf{n}$ for a suitable vector $\mathbf{b}$; continuity of traction requires that $\bar{\mathbf{S}}\,\mathbf{n}=\overset{+}{\mathbf{S}}\,\mathbf{n}$. In addition, suppose that the phases on the two sides of $\mathcal{S}_t$ are *metastable* in the sense that the free-energy $\hat{\psi}(\mathbf{F},\theta)$ is *strongly elliptic* at $\mathbf{F}=\overset{\pm}{\mathbf{F}}$; see [48] for a discussion relating strong ellipticity to infinitesimal stability. Finally, the formula (2.6) for the driving force simplifies, in the present quasi-static setting, to $f=\rho[[\psi]]-\overset{\pm}{\mathbf{S}}\cdot[[\mathbf{F}]]$.

We now consider the potential energy function $G(\mathbf{F};\overset{+}{\mathbf{S}},\theta)$ and examine some of its features. Similar results hold for $G(\mathbf{F};\bar{\mathbf{S}},\theta)$. Consider the set $\mathcal{A}=\mathcal{A}(\overset{+}{\mathbf{F}},\mathbf{n})$ of *all* nonsingular tensors which are kinematically compatible with $\overset{+}{\mathbf{F}}$:

$$\mathcal{A}=\{\mathbf{F}\mid\mathbf{F}=\overset{+}{\mathbf{F}}+\mathbf{a}\otimes\mathbf{n}\quad\text{for all vectors }\mathbf{a}\},\tag{2.12}$$

and consider the restriction of the function $G(\cdot;\overset{+}{\mathbf{S}},\theta)$ to the set $\mathcal{A}$. On making use of the traction continuity condition and the constitutive relations $\overset{\pm}{\mathbf{S}}=\rho\hat{\psi}_{\mathbf{F}}(\overset{\pm}{\mathbf{F}},\theta)$, one finds that G, as a function of $\mathbf{a}$, has *extrema* at $\mathbf{a}=\mathbf{0}$ and $\mathbf{a}=\mathbf{b}$ (corresponding to $\mathbf{F}=\overset{+}{\mathbf{F}}$ and $\mathbf{F}=\bar{\mathbf{F}}$). Moreover, the requirement of strong ellipticity can be used to show that these extrema are *local minima*. Figure 1 shows a *schematic* graph of $G(\mathbf{F};\overset{+}{\mathbf{S}},\theta)$ versus $\mathbf{F}$ showing two local minima ("energy-wells") at $\mathbf{F}=\overset{\pm}{\mathbf{F}}$.

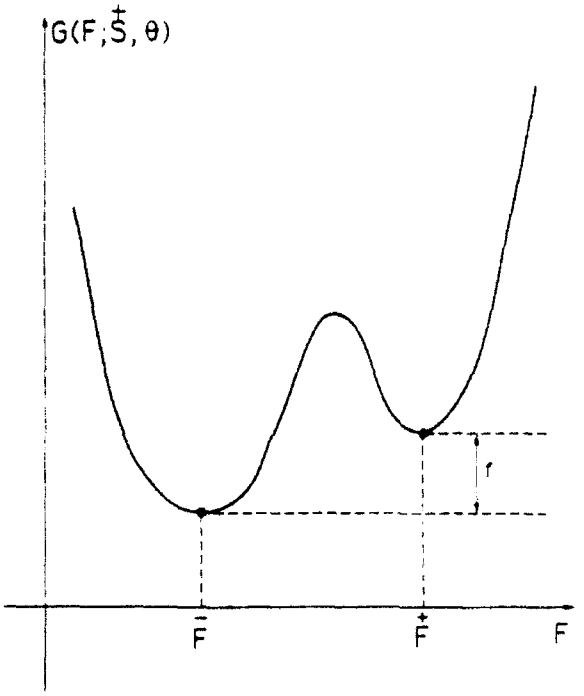


Figure 1. Schematic graph of potential energy function G versus deformation gradient tensor $\mathbf{F}$.

Finally, one sees that the difference between the values of G at the two minima $\mathbf{F} = \overset{\pm}{\mathbf{F}}$ is equal to the driving force: $f = \overset{+}{G} - \bar{G} = \overset{+}{g} - \bar{g}$. Suppose that the minimum at $\mathbf{F} = \bar{\mathbf{F}}$ is less than the minimum at $\mathbf{F} = \overset{+}{\mathbf{F}}$ as is the case depicted in Figure 1; then f is positive, so that by the entropy inequality $fV_n \geq 0$, one has $V_n \geq 0$. The phase boundary $\mathcal{S}_t$ therefore moves into its positive side, creating more material in the low-minimum phase at the expense of material in the high-minimum phase. Thus the phase corresponding to the lesser minimum of G is preferred. The same is true when f is negative. Thus we may speak of the phase associated with the lesser minimum of G as being *stable* (or at least as being more stable than the phase associated with the higher minimum).

3. A CONSTITUTIVE MODEL

From here on we restrict attention to a purely one-dimensional setting for a uniaxial bar. In the present section we describe a simple explicit constitutive model, consisting of a Helmholtz free-energy function, kinetic relation and nucleation criterion. Its implications will be discussed in the sections to follow.

Helmholtz free-energy function: Let θ_c be the critical temperature: when $\theta > \theta_c$ the material exists in only a single phase, no matter what the stress-level. On the other hand for $0 < \theta < \theta_c$, we assume that two metastable phases exist if the stress σ lies in a suitable range $\quad(\theta) < \sigma < \sigma_M(\theta)$, and that only one phase exists otherwise.

The potential energy function $G(\gamma; \sigma, \theta) = \rho\hat{\psi}(\gamma, \theta) - \sigma\gamma$ that characterizes the material must therefore be such that $G(\cdot; \sigma, \theta)$ has two local minima when $0 < \theta < \theta_c$, $\sigma_m(\theta) < \sigma < \sigma_M(\theta)$, and a single minimum otherwise.

This implies that the stress-strain-temperature relation $\sigma = \hat{\sigma}(\gamma, \theta) = \rho\hat{\psi}_\gamma(\gamma, \theta)$ must have the following properties: for $\theta > \theta_c$, $\hat{\sigma}$ must be a monotonically increasing function of γ. For $0 < \theta < \theta_c$, $\hat{\sigma}$ must first increase with γ on a range $-1 < \gamma < \gamma_M(\theta)$, then decrease over an intermediate range $\gamma_M(\theta) < \gamma < \gamma_m(\theta)$, and finally increase again for $\gamma > \gamma_m(\theta)$. We shall assume that $\hat{\sigma}$ has the *trilinear* form shown in Figure 2:

$$\sigma = \hat{\sigma}(\gamma, \theta) = \begin{cases} \mu\gamma - \mu\alpha(\theta - \theta_T), & -1 < \gamma < \gamma_M(\theta), \quad 0 < \theta < \theta_c, \\ \mu(\gamma - \gamma_T) - \mu\alpha(\theta - \theta_T), & \gamma > \gamma_m(\theta), \quad 0 < \theta < \theta_c, \end{cases} \tag{3.1}$$

where we have omitted displaying the equation of the declining branch. Though the figure has been drawn with $\sigma_m(\theta)$ negative, this need not be the case. The restriction $\gamma > -1$ arises from the requirement that the deformation be one-to-one. The relationships between $\sigma_M(\theta)$, $\sigma_m(\theta)$ and $\gamma_M(\theta)$, $\gamma_m(\theta)$ may be read from Figure 2. We take $\gamma_M(\theta)$ and $\gamma_m(\theta)$ to be linear functions of θ that coincide at $\theta = \theta_c$.

The two rising branches of the stress-strain curve are associated with distinct phases of the material which we refer to as the *low-strain phase* and the *high-strain phase*. In the simple model here, both phases have the same constant elastic moduli μ and coefficients

of thermal expansion α. The constant $\gamma_T(>0)$, representing the horizontal distance between the two rising branches of the stress-strain curve, is the *transformation strain*. We require θ_T to be the *transformation temperature*, so that both phases exist and both are stable when $\theta = \theta_T, \sigma = 0$; this means that $G(\cdot\,; 0, \theta_T)$ is to have two local minima, and that the values of G at these minima are to coincide.

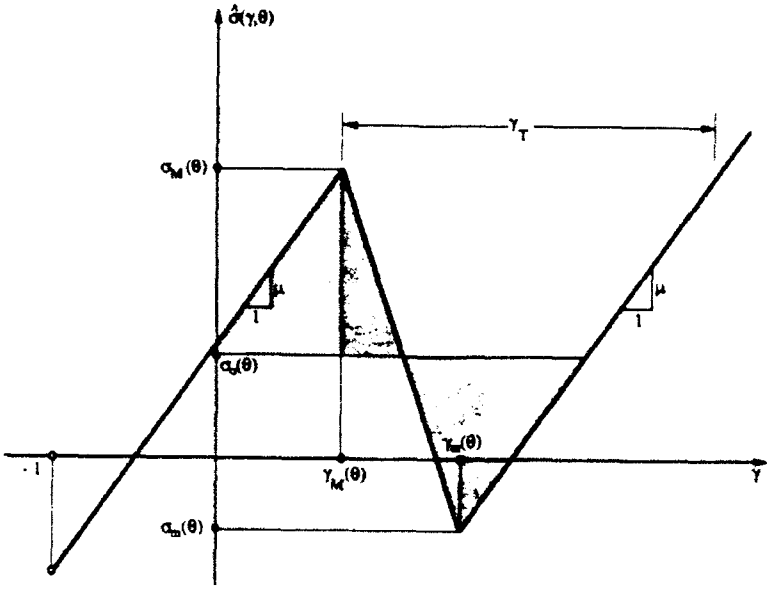


Figure 2. The trilinear stress-strain curve at a fixed temperature θ.

By integrating the relation $\rho\hat{\psi}_\gamma(\gamma,\theta) = \hat{\sigma}(\gamma,\theta)$ with respect to γ, and enforcing the requirements that *(i)* $\hat{\psi}(\gamma,\theta)$ be continuously differentiable for $\gamma > -1$, $0 < \theta < \theta_c$, *(ii)* that both phases have the same constant specific heat at constant strain $c = -\theta\hat{\psi}_{\theta\theta}(\gamma,\theta)$, and *(iii)* that θ_T be the transformation temperature, one finds

$$\rho\hat{\psi}(\gamma,\theta) = \begin{cases} \frac{\mu}{2}\gamma^2 - \mu\alpha\gamma(\theta-\theta_T) - \rho c\theta\, log(\theta/\theta_T), \\ \\ \frac{\mu}{2}(\gamma-\gamma_T)^2 - \mu\alpha(\gamma-\gamma_T)(\theta-\theta_T) - \rho c\theta\, log(\theta/\theta_T) + \frac{\rho\lambda_T}{\theta_T}(\theta-\theta_T), \end{cases} \tag{3.2}$$

where $(3.2)_1$ holds for the low-strain phase, $(3.2)_2$ for the high-strain phase. The constant λ_T can be shown to represent the latent heat of the transition at the transformation temperature θ_T. In view of the particularly simple manner in which the temperature θ enters into the present model, one would not expect it to be realistic when θ is far from the transformation temperature θ_T.

The stress-level $\sigma_o(\theta)$, defined by

$$\sigma_o(\theta) = (\rho\lambda_T/\gamma_T\theta_T)(\theta-\theta_T), \tag{3.3}$$

is known as the *Marwell stress* and has the geometric significance shown in Figure 2 where the two shaded areas are equal. One can readily verify that when $\sigma < \sigma_o(\theta)$, the low-strain minimum of the potential energy function $G(\cdot\,;\sigma,\theta)$ is smaller than the high-strain minimum, while the converse is true for $\sigma > \sigma_o(\theta)$. Thus the low-strain phase is stable when $\sigma < \sigma_o(\theta)$, the high-strain phase is stable when $\sigma > \sigma_o(\theta)$, and both phases are stable when $\sigma = \sigma_o(\theta)$. Observe from this and (3.3), that if the latent heat $\lambda_T > 0$, the low-strain phase is stable at high temperatures, while the high-strain phase is stable at low temperatures; thus λ_T would be positive when the low-strain phase is *austenite* and the high-strain phase is *martensite*. If the converse is true, then $\lambda_T < 0$.

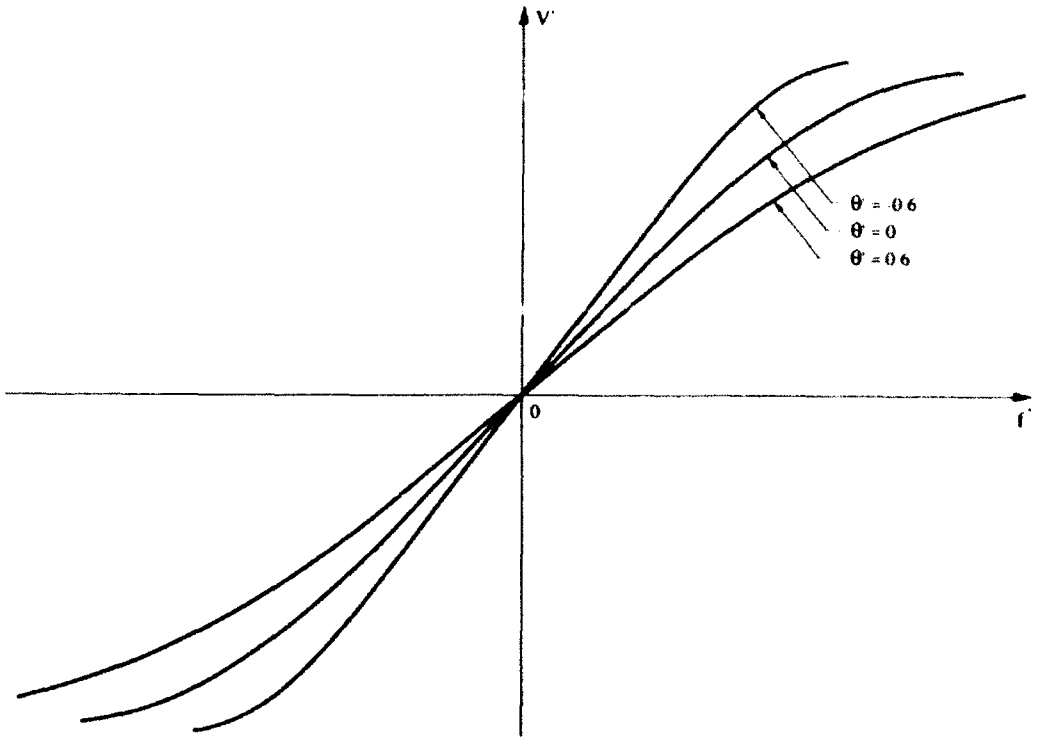


Figure 3. Phase boundary velocity as a function of driving force at various temperatures.

Kinetic relation: Let $x = s(t)$ denote the location of a phase boundary during a quasi-static process, and suppose that the particle at $x = s-$ is in the high-strain phase and the particle at $x = s+$ is in the low-strain phase. The driving force f on this phase boundary can be calculated by substituting (3.1) and (3.2) into $f = [[\rho\psi]] - \sigma[[\gamma]]$, thus leading to

$$f = \hat{f}(\sigma,\theta) = \gamma_T[\sigma - \sigma_o(\theta)]. \tag{3.4}$$

Since the bar is in a two-phase state, the potential energy function $G(\cdot\,;\sigma,\theta)$ has two local minima, one at $\gamma = \bar{\gamma}$, the other at $\gamma = \overset{+}{\gamma}$, and they are separated by a local

maximum; this is analogous to the case depicted schematically in Figure 1. As the phase boundary moves through the bar, the particle immediately ahead of it "jumps" from one local minimum of G to the other, and the underlying kinetic relation can be constructed by viewing this jumping process at the atomic scale. In particular, if this process is "thermally-activated", classical arguments based on the idea that atoms jump from one local minimum to the other when they possess sufficient energy to cross over the intermediate energy barrier, leads to the following kinetic relation:

$$\dot{s} = v(\sigma,\theta) = R\left\{exp\left[-\frac{\gamma_T}{2rk\theta}\frac{[\sigma-\sigma_M(\theta)]^2}{\sigma_M(\theta)-\sigma_m(\theta)}\right] - exp\left[-\frac{\gamma_T}{2rk\theta}\frac{[\sigma-\sigma_m(\theta)]^2}{\sigma_M(\theta)-\sigma_m(\theta)}\right]\right\}, \quad (3.5)$$

where R and r are positive material constants, k is Boltzmann's constant, and $\sigma_M(\theta)$ and $\sigma_m(\theta)$ have the meaning shown in Figure 2. One can use (3.4) to express (3.5) in the alternate form $\dot{s} = V(f,\theta)$. Graphs of $V' = V/R$ versus $f' = f/\mu\gamma_T$, at different values of temperature $\theta' = (\theta-\theta_T)/\theta_T$, are shown in Figure 3. One can show that $V(f,\theta)f \geq 0$ as required by (2.10), and also that V is an increasing function of f at fixed θ. The variation of V with temperature θ at fixed f is more complicated; see [7].

Nucleation criterion: Consider a process in which the entire bar is in a single-phase configuration for an initial interval of time and in two-phase states for subsequent times. The instant of time and the point in the bar at which the transition from the single-phase to the two-phase configuration is initiated are determined by a nucleation criterion.

Addressing first the question of location, in reality, the nucleation site is strongly influenced by inhomogeneities. Since our bar is uniform, the location of the nucleation site is rather arbitrary. In the case of a bar rendered inhomogeneous by means of a slight uniform taper with the small end at $x = 0$, the low strain phase to high-strain phase transition would necessarily commence at $x = 0$ and the reverse transition at $x = L$. We shall arbitrarily assume that this is the case in our uniform bar as well.

Turning next to the conditions under which nucleation occurs, we assume that each transition is initiated at a critical value of driving force, a value $f_*(>0)$ for the low-strain phase to high-strain phase transition and a value $f_{**}(<0)$ for the reverse transition. By making use of (3.4), one can restate this, for quasi-static processes, in terms of critical values of stress and temperature: the low-strain phase to high-strain phase transition is initiated when $\sigma = \sigma_{L\to H}(\theta)$ while the reverse transition is initiated at $\sigma = \sigma_{H\to L}(\theta)$ where

$$\sigma_{L\to H}(\theta) \equiv (\rho\lambda_T/\gamma_T\theta_T)(\theta-\theta_*), \quad \sigma_{H\to L}(\theta) \equiv (\rho\lambda_T/\gamma_T\theta_T)(\theta-\theta_{**}); \quad (3.6)$$

θ_* and θ_{**} denote material constants which may be readily related to the critical nucleation values f_* and f_{**} of driving force. Observe that when the stress vanishes, this nucleation criterion states that the low-strain phase to high-strain phase transition is initiated at $\theta = \theta_*$, the reverse transition at $\theta = \theta_{**}$. Thus when $\lambda_T > 0$, corresponding

to the case where the low-strain phase is austenite and the high-strain phase is martensite, θ_* and θ_{**} represent the "martensite start temperature" and the "austenite start temperature" respectively: $\theta_* = M_s$, $\theta_{**} = A_s$. The converse is true if $\lambda_T < 0$.

4. THERMO-MECHANICAL RESPONSE

Consider a bar composed of the material described in the preceding section. Let L be its length in the reference configuration and δ its elongation from this state. We consider *quasi-static* processes during which the bar is always at a uniform stress $\sigma(t)$ and temperature $\theta(t)$. Suppose that the temperature $\theta(t)$ and one of the two mechanical quantities $\delta(t)$ and $\sigma(t)$ are prescribed and that we wish to determine the remaining mechanical variable.

Suppose that initially, for $0 < t < t_*$, the entire bar is in the low-strain phase; by (3.1), the elongation, stress and temperature are then related by $\delta/L = \sigma/\mu + \alpha(\theta - \theta_o)$. At the instant t_*, the high-strain phase is nucleated at the left end of the bar so that by $(3.6)_1$,

$$\sigma(t_*) = (\rho\lambda_T/\gamma_T\theta_T)[\theta(t_*) - \theta_*]. \tag{4.1}$$

For some time thereafter, the bar is in a two-phase configuration with the high-strain phase on $0 < x < s(t)$, the low-strain phase on $s(t) < x < L$; by (3.1), one now finds that $\delta/L = \gamma_T s/L + \sigma/\mu + \alpha(\theta - \theta_o)$. Differentiating this with respect to time and using the kinetic relation (3.5) to eliminate $\dot{s}$ leads to

$$\dot{\delta}/L - \dot{\sigma}/\mu - \alpha\dot{\theta} = (\gamma_T/L)v(\sigma,\theta) \tag{4.2}$$

with $v(\sigma,\theta)$ given by (3.5). The response of the bar during this second stage is governed by the initial-value problem consisting of (4.2), (4.1). If at some subsequent time the phase boundary reaches one end of the bar, then the bar is again in a single-phase state, and its response can be determined directly from (3.1). If the bar is entirely in the high-strain phase, and the high-strain phase to low-strain phase transition is nucleated at some instant t_{**}, the initial condition (4.1) must be replaced by $\sigma(t_{**}) = (\rho\lambda_T/\gamma_T\theta_T)[\theta(t_{**}) - \theta_{**}]$.

In order to study the response of the model in such processes, we carried out some simple numerical computations for three types of loading. In all cases the bar was initially in the low-strain phase. The specific values of the various material parameters used may be found in [7]; in particular, we took $\lambda_T > 0$ corresponding to the low-strain phase being austenite and the high-strain phase being martensite. The non-dimensional quantities referred to in the results shown in Figures 4-7 are the stress $\sigma' = \sigma/\mu\gamma_T$, temperature $\theta' = (\theta - \theta_T)/\theta_T$, elongation $\Delta' = \Delta/L\gamma_T$ where Δ is the elongation from the *initial* configuration (rather than from the reference configuration), and the elongation-rate $q' = \dot{\delta}/R\gamma_T$. Consider the following three loading programs:

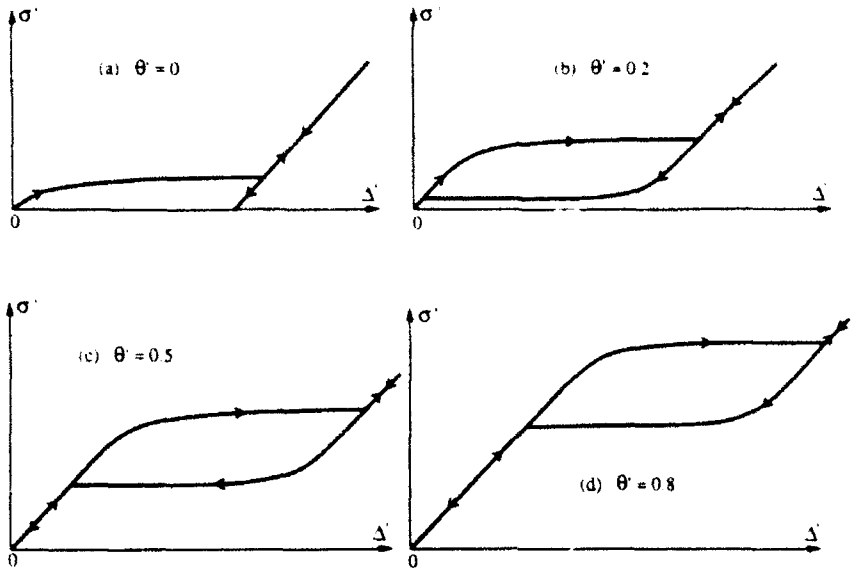


Figure 4. Mechanical cycling at various temperatures, all at the same elongation rate.

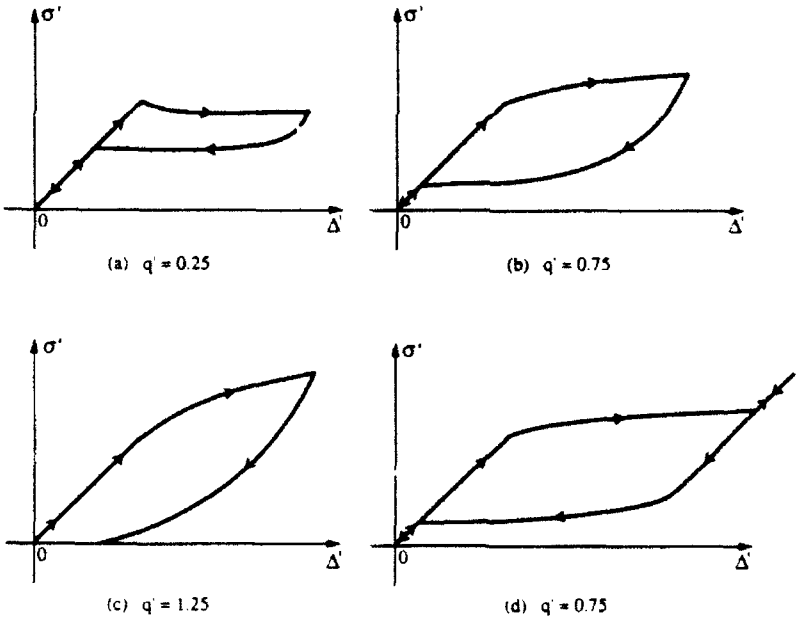


Figure 5. Mechanical cycling at various elongation rates, all at the same temperature.

(i) The initially stress-free bar is loaded isothermally at a constant elongation-rate $\dot{\delta} > 0$. At some subsequent time, loading is replaced by unloading at the rate $-\dot{\delta}$, still at the same temperature, until the stress returns to zero. Figure 4 shows the resulting response for different values of θ at the same value of $\dot{\delta}$. Observe that permanent deformation occurs at sufficiently low temperatures, but does not occur at sufficiently high temperatures. Note also that the principal effect of raising temperature is to shift the hysteresis loop in the direction of increasing stress, without significantly altering its vertical thickness. Figure 5 shows the effect of different elongation-rates $\dot{\delta}$ at the same value of θ. Observe that at sufficiently slow elongation-rates, the force decreases immediately after nucleation, whereas at higher rates the force increases. In Figures 5(a),(b),(c), unloading commenced *before* the entire bar transformed to the high-strain phase, while in Figure 5(d) unloading started *after* the low-strain to high-strain transition was complete. These observations are qualitatively similar to experimental observations on some shape-memory alloys, e.g. AgCd ([30]), NiTi ([36]), CuZnSi ([38]) and CuAlNi ([33]).

(ii) The bar is loaded thermally at constant stress σ by initially cooling it at a constant temperature-rate $\dot{\theta} < 0$ and later heating it at a rate $-\dot{\theta}$. Figure 6 shows the resulting response at different values of stress and the same cooling/heating rate $|\dot{\theta}|$. Observe that the effect of increasing the stress-level is to shift the hysteresis loop in the direction of increasing temperature without significant change in its thickness. This feature has been observed in InTl by Burkart and Read [11] in experiments conducted at *compressive* stress.

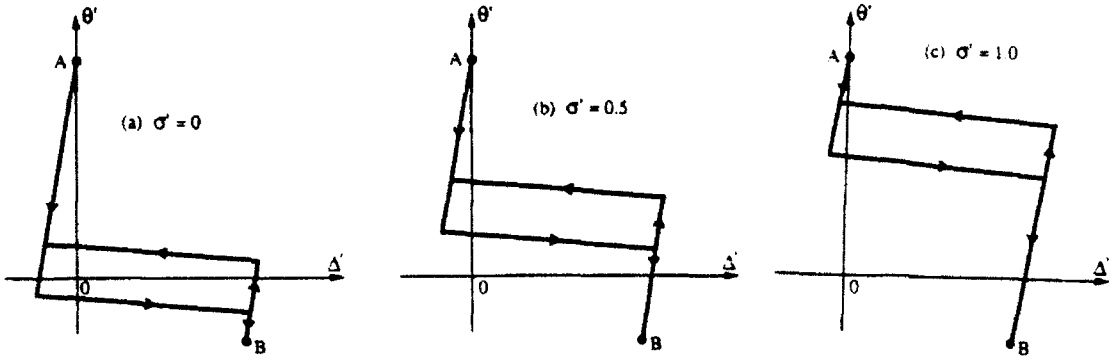


Figure 6. Thermal cycling at various stresses, all at the same cooling rate.

(iii) Consider a stress-free bar at the transformation temperature; thus initially, $\sigma = 0, \theta = \theta_T, \delta = 0$. The bar is now subjected to a four-stage thermo-mechanical loading process: first, it is stretched isothermally at a constant elongation-rate $\dot{\delta} > 0$ (leading to the curve OAB in Figure 7). After the bar has fully transformed to the high-strain phase, it is then mechanically unloaded back to zero stress, still at $\theta = \theta_T$, (curve BC). The bar is now at $\sigma = 0, \theta = \theta_T$, but it is permanently deformed $\delta > 0$. From hereon the stress in the bar is maintained at zero. In stage three, the bar is heated at a constant rate $\dot{\theta} > 0$ until it is fully transformed to the low-strain phase (curve CDE). Finally it

is cooled back to the initial temperature $\theta = \theta_T$ (curve EO). This returns the bar to its original state $\sigma = 0, \theta = \theta_T, \delta = 0$, demonstrating the so-called *shape-memory effect*.

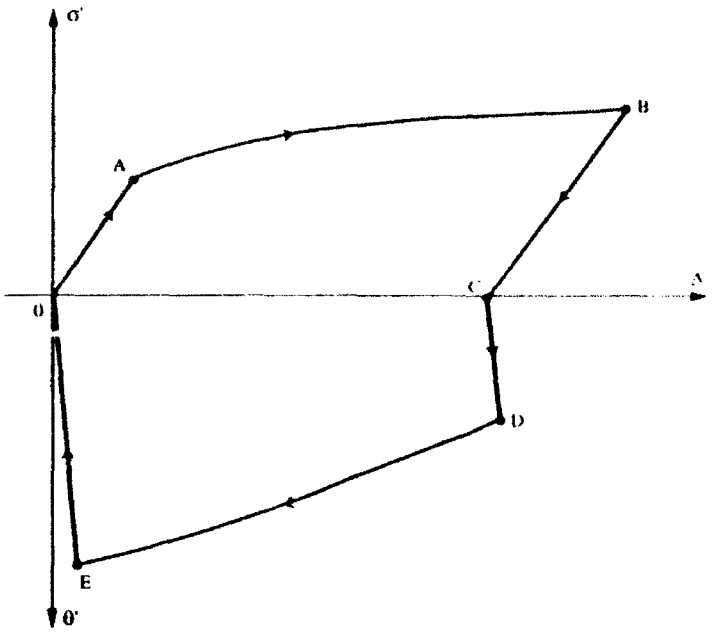


Figure 7. The shape memory effect.

5. INERTIAL EFFECTS

We now consider the effects of inertia, and for simplicity, make the unrealistic assumption that the processes of interest take place *isothermally*, $\theta(x,t) =$ constant $(< \theta_c)$. All thermal variables can now be suppressed, the energy field equation and jump condition being satisfied a posteriori by applying a suitable distribution of heat sources. The presence of inertial effects permit a discontinuity to be either a shock wave or a phase boundary: the strains $\bar{\gamma}$ and $\overset{+}{\gamma}$ on either side of a shock wave both belong to the same phase of the material, and so are associated with the same branch of the stress-strain curve; in contrast, a phase boundary is associated with two different phases, and therefore with two different branches of the stress-strain curve.

Consider a bar of uniform cross-sectional area whose stress-strain relation is $\sigma = \hat{\sigma}(\gamma)$ with $\hat{\sigma}$ having the trilinear form shown in Figure 2. Let $x = s(t)$ denote the location of a strain discontinuity during a motion of the bar. The one-dimensional counterparts of (2.1), (2.2), (2.6) and (2.9) must hold at each such discontinuity. By combining the first two of these, one finds

$$\rho \dot{s}^2 = \frac{\hat{\sigma}(\overset{+}{\gamma}) - \hat{\sigma}(\bar{\gamma})}{\overset{+}{\gamma} - \bar{\gamma}} \geq 0. \tag{5.1}$$

From (5.1) and (3.1) one sees that shock waves in both the high-strain and low-strain phases have the same constant propagation speed $c = (\mu/\rho)^{\frac{1}{2}}$. In addition, the linearity

of the stress response function $\hat{\sigma}(\gamma)$ between the two strains $\bar{\gamma}$ and $\overset{+}{\gamma}$ allows one to infer that the driving force f on a shock wave vanishes. Thus, in particular, the entropy inequality $f\dot{s} \geq 0$ is automatically satisfied at all shock waves in this trilinear material.

Turning next to phase boundaries, it is readily seen from Figure 2 and (5.1) that the propagation speed $|\dot{s}|$ of a phase boundary cannot exceed the shock wave speed c; thus phase boundaries in this material are necessarily *subsonic*. (This would *not* be true if the two phases had different elastic moduli, say μ_H and μ_L with $\mu_H \neq \mu_L$; such a case is discussed in [3], demonstrating that subsonic and supersonic phase boundaries behave quite differently.) The kinetic relation is to hold at all phase boundaries. The results that follow hold for *any* kinetic relation of the form $\dot{s} = V(f)$ with $V(f)f \geq 0$ and V monotonically increasing (and are *not* restricted to the particular form (3.5), (3.4), which in any case was constructed by assuming the process to be quasi-static).

In order to elucidate the distinct roles played by the entropy inequality, the kinetic relation and the nucleation criterion, we now discuss a particular initial-value problem. We consider an infinite bar and seek solutions of the fields equations $\hat{\sigma}'(\gamma)\gamma_x - \rho v_t = 0$ and $\gamma_t - v_x = 0$, subject to the corresponding jump conditions, the entropy inequality, the kinetic relation, the nucleation criterion and the (Riemann) initial conditions

$$\gamma(x,0+) = \begin{cases} \gamma_L, & -\infty < x < 0, \\ \gamma_R, & 0 < x < \infty, \end{cases} \qquad v(x,0+) = \begin{cases} v_L, & -\infty < x < 0, \\ v_R, & 0 < x < \infty. \end{cases} \tag{5.2}$$

The initial strains and velocities γ_L, γ_R, v_L and v_R are given constants and we restrict attention to the case where γ_L and γ_R are both in the *low-strain phase*. Finally, it will be convenient to introduce a parameter h, having the dimension of stress, that depends only on the given data:

$$h = (\mu/2)\left[\gamma_L + \gamma_R + (v_R - v_L)/c\right]. \tag{5.3}$$

We seek solutions that are piecewise smooth and invariant under the scale change $x \to kx$, $t \to kt$, where k is a constant. Such solutions must have the piecewise constant form shown in Figure 8 where γ_j and v_j are the constant values of the strain $\gamma(x,t)$ and the particle velocity $v(x,t)$ in the j*th*-wedge of the (x,t)-plane. This solution form involves N strain discontinuities on lines $x = \dot{s}_j t$ which may be shock waves or phase boundaries. For the trilinear material, fans cannot occur. Since the strain and velocity fields corresponding to Figure 8 are piecewise constant, the governing field equations are satisfied automatically. The one-dimensional versions of the jump conditions (2.1), (2.2) and the entropy inequality $f\dot{s} \geq 0$ must be enforced at each of the N discontinuities. We momentarily *postpone* consideration of the kinetic relation and the nucleation criterion.

Entropy inequality: The jump conditions and entropy inequality can be invoked to show that any solution of the form shown in Figure 8 must necessarily take one of the two simpler forms shown in Figure 9: when the initial data are such that $h > \sigma_0$, where

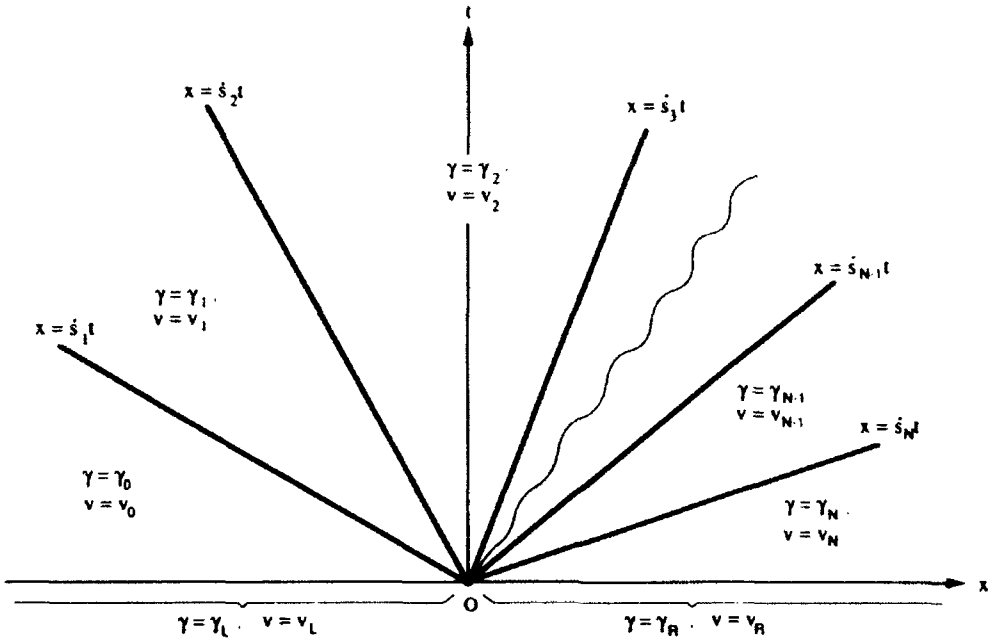
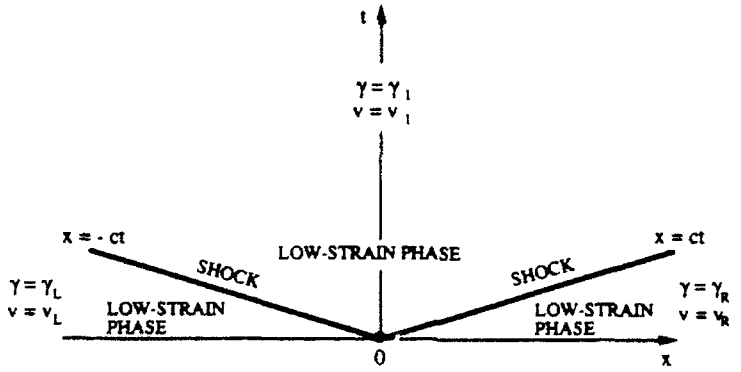
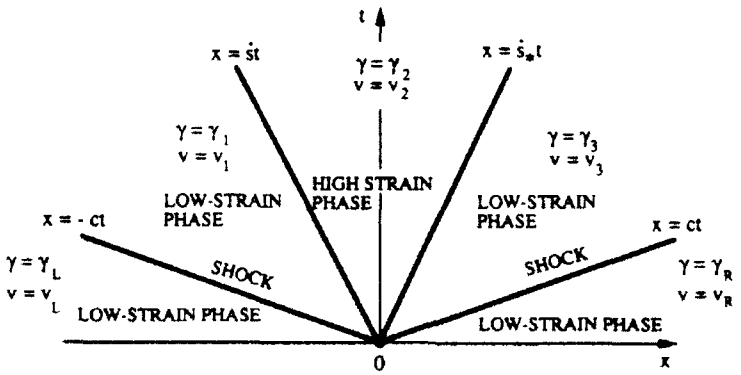


Figure 8. Assumed form of solutions to the Riemann problem.



(a) NO PHASE BOUNDARIES



(b) TWO PHASE BOUNDARIES

Figure 9. Form of solutions to Riemann problem permitted by entropy inequality.

σ_o is the Maxwell stress given by (3.3), there is a *two-parameter family* of solutions of the form shown in Figure 9(b); the two parameters are the velocities $\dot{s}$ and $\dot{s}_*$ of the two phase boundaries. On the other hand, when $-\mu < h \leq \sigma_M$, there is a solution of the form shown in Figure 9(a) involving *no* phase boundaries. For initial data in the common interval $\sigma_o < h < \sigma_M$, *both* types of solutions exist.

This reveals two distinct types of non-uniqueness. First, there is the lack of uniqueness arising because of the undetermined velocities of the moving phase boundaries; this is exemplified in the first case. Second, for initial data such that $\sigma_o < h \leq \sigma_M$, *two* types of solutions can occur, one in which the bar ultimately changes phase, the other in which it does not.

Kinetics, nucleation and uniqueness: We now indicate how the nucleation criterion and the kinetic relation resolve this lack of uniqueness. Within the class of solutions for which the bar changes phase (Figure 9(b)), it is possible to show that the kinetic relation $\dot{s} = V(f)$, with V monotonically increasing, serves to determine the velocities of the two phase boundaries, so that the Riemann problem, when augmented by the kinetic relation, has a unique solution. The only remaining issue arises for values of h in the intermediate range $(\sigma_o, \sigma_M]$; in this case, it is required to choose between precisely two solutions, in one of which the bar changes phase (Figure 9(b)), while in the other it does not (Figure 9(a)). In order to make this choice, one appeals to the nucleation criterion, according to which a phase transition from the low-strain phase to the high-strain phase is initiated at a critical level $f = f_*$ of driving force. It is possible to show by invoking this criterion that the change-of-phase solution is selected for initial data in the range $\sigma_{L\to H} \leq h < \sigma_M$, while for $\sigma_o < h < \sigma_{L\to H}$ the criterion picks the solution with no phase change; here $\sigma_{L\to H}$ is the nucleation stress-level defined in (3.6).

6. EFFECT OF VISCOSITY AND STRAIN-GRADIENT

By modeling the material as being thermoelastic, we neglect, in particular, the effects of viscosity and strain-gradients. Such effects could have important local consequences in the vicinity of a phase boundary, even if it is appropriate to neglect them when describing the bulk behavior of the material elsewhere in the body. In this section we describe the influence of such effects in a one-dimensional isothermal setting.

Suppose that the effect of viscosity and strain-gradient on the constitutive relation is to replace the elastic stress-strain law $\sigma = \hat{\sigma}(\gamma)$ by

$$\sigma = \hat{\sigma}(\gamma) + \rho\nu\gamma_t - \rho\lambda\gamma_{xx}, \tag{6.1}$$

where the constants $\nu \geq 0$ and $\lambda \geq 0$ are the viscosity and the strain-gradient coefficient respectively, and $\hat{\sigma}$ is the trilinear function shown in Figure 2. A propagating phase boundary in the *elastic* bar, associated with the states $(\bar{\gamma}, \bar{v})$ and $(\overset{+}{\gamma}, \overset{+}{v})$, corresponds, in this augmented theory, to a travelling wave that smoothly connects these two states:

$$\gamma = \gamma(\xi), \quad v = v(\xi), \quad \xi = x - \dot{s}t, \tag{6.2}$$

$$\gamma(-\infty) = \bar{\gamma}, \quad v(-\infty) = \bar{v}, \qquad \gamma(+\infty) = \overset{+}{\gamma}, \quad v(+\infty) = \overset{+}{v}, \tag{6.3}$$

where $\dot{s}$ is a constant. The functions $\gamma(\xi)$, $v(\xi)$ are to be twice continuously differentiable on $(-\infty, \infty)$ and must satisfy the field equations $\sigma_x - \rho v_t = 0$, $\gamma_t - v_x = 0$ and (6.1), and the boundary conditions (6.3). The trilinearity of $\hat{\sigma}$ allows one to determine $\gamma(\xi)$ and $v(\xi)$ explicitly, though we shall not display the result here.

Such travelling waves $\gamma(\xi)$, $v(\xi)$ have a limit as the viscosity ν and the strain-gradient coefficient λ tend to zero, with the dimensionless parameter $\omega = 2\lambda^{\frac{1}{2}}/\nu$ and the data $\bar{\gamma}, \bar{v}, \overset{+}{\gamma}, \overset{+}{v}$, held fixed; indeed, $\gamma(\xi)$, $v(\xi)$ tend to $\bar{\gamma}, \bar{v}$ for $\xi < 0$, and to $\overset{+}{\gamma}, \overset{+}{v}$ for $\xi > 0$ in this limit. The limiting functions $\gamma(\xi)$, $v(\xi)$ satisfy all of the differential equations and jump conditions of the elastic theory; *in addition*, one finds that the limiting discontinuity satisfies a *particular kinetic relation* $\dot{s} = V(f; \omega)$. This kinetic relation characterizes the influence of viscosity and strain-gradient on a phase boundary even when these effects are negligible in the bulk material.

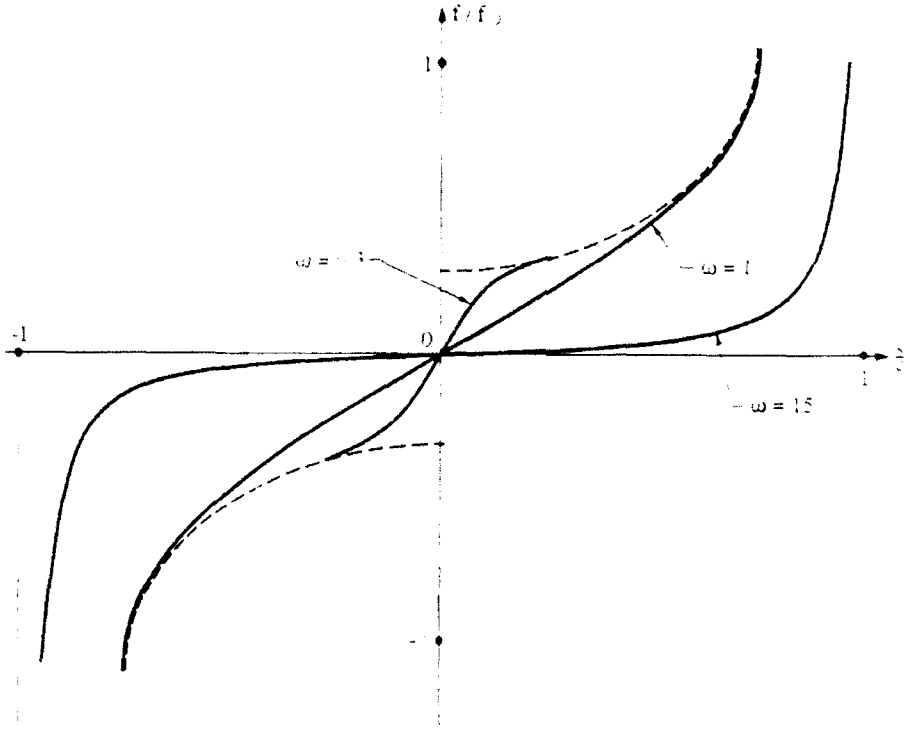


Figure 10. The kinetic relation induced by viscosity and strain-gradient effects.

The kinetic relation inherited in this way may be fully and explicitly determined, though its representation is too complicated to display here. Figure 10 describes it by plotting f/f_o vs. $\dot{s}/c$ where $f_o = \frac{1}{2}\mu\gamma_T(\gamma_m + \gamma_M)$. One finds that f/f_o depends only on the three material parameters $\sigma_M/\sigma_m, \gamma_M/\gamma_m$ and ω; the figure is plotted for fixed values of the first two of these and for various values of ω. The kinetic response function V increases monotonically with f for the values of the parameters considered. It can be

shown that V satisfies the requirement $V(f;\omega)f \geq 0$ imposed on every kinetic response function by the entropy inequality.

The effect of viscosity and strain-gradients (or capillarity) on shock waves in gas dynamics is well-known. Here too, the discontinuities are smoothed-out and replaced by narrow zones in which the fields vary continuously but rapidly. In the limit of vanishing viscosity and capillarity, one recovers *the* particular shock wave that is associated with the states $(\bar{\gamma},\bar{v})$, $(\overset{+}{\gamma},\overset{+}{v})$ and all traces of viscosity and capillarity disappear in this limit. *In contrast* for phase boundaries, the limit of vanishing viscosity and strain-gradient effects leads to *a particular* discontinuous solution from among the many that can be associated with the states $(\bar{\gamma},\bar{v})$, $(\overset{+}{\gamma},\overset{+}{v})$, the particular phase boundary selected by this limiting procedure being the one that obeys the aforementioned kinetic relation. Thus the effects of viscosity and strain-gradient *persist* (in the form of a kinetic relation $\dot{s} = V(f;\omega)$) even in the limit $\nu, \lambda \to 0$, through the presence of the parameter ω.

Acknowledgements

We have had the benefit of questions and comments from Kaushik Bhattacharya, Richard James Kazuhiro Otsuka, Phoebus Rosakis, James Rice and Lev Truskinovsky, to whom we express our thanks.The results described here were obtained in the course of investigations supported by the U.S. Office of Naval Research and the U.S. Army Research Office.

7. REFERENCES

1. R. Abeyaratne and J. K. Knowles, On the dissipative response due to discontinuous strains in bars of unstable elastic material, *Int. J. Solids Struct.* **24**(1988), 1021-1044.
2. R. Abeyaratne and J. K. Knowles, On the driving traction acting on a surface of strain discontinuity in a continuum, *J. Mech. Phys. Solids*, **38**(1990), 345-360.
3. R. Abeyaratne and J. K. Knowles, Kinetic relations and the propagation of phase boundaries in solids, *Archs ration. Mech. Analysis*,**114**(1991),119-154.
4. R. Abeyaratne and J. K. Knowles, Implications of viscosity and strain gradient effects for the kinetics of propagating phase boundaries in solids, *SIAM J. Appl. Math.* **51**(1991),1205-1221.
5. R. Abeyaratne and J. K. Knowles, On the propagation of maximally dissipative phase boundaries in solids, *Quart. Appl. Math.* **50**(1992) , 149-172.
6. R. Abeyaratne and J. K. Knowles, Nucleation, kinetics and admissibility criteria for propagating phase boundaries, in *Shock Induced Transitions and Phase Structures in General Media*, (Editors - R. Fosdick, E. Dunn and M. Slemrod), IMA Volumes in Mathematics and Its Applications, Vol. 52, Springer-Verlag, 1992.
7. R. Abeyaratne and J. K. Knowles, A continuum model of a thermoelastic solid capable of undergoing phase transitions, to appear in the *Journal of the Mechanics and Physics of Solids.*
8. J. M. Ball and R. D. James, Fine phase mixtures as minimizers of energy,*Archs ration. Mech. Analysis*.**100**(1987),15-52.

9. K. Bhattacharya, Wedge-like microstructure in martensite, *Acta Metal. Materiala*, **39**(1991), 2431-2444. Self-accommodation in martensite, (1992), to appear in the *Archs ration. Mech. Analysis.*
10. K. Bhattacharya, Linear and nonlinear thermoelasticity theory for crystalline solids, in *Recent Advances in Elasticity*, AMD Volume 124, Proceedings of the ASME Winter Annual Meeting, Atlanta, (1991b).
11. M. W. Burkart and T. A. Read, Diffusionless phase change in the indium-thallium system, *Trans. AIME: J. Metals*, **197**(1953), 1516-1524.
12. H. B. Callen, *Thermodynamics and an Introduction to Thermostatistics*,1985, Wiley, New York.
13. J. W. Christian,*The Theory of Transformations in Metals and Alloys*, Part I,1975, Oxford: Pergamon Press.
14. C. Collins and M. Luskin, The computation of the austenitic-martensitic phase transition, in *Partial Differential Equations and Continuum Models of Phase Transitions*, (Editors - M. Rascle, D. Serre, and M. Slemrod), Lecture Notes in Physics **344**,1989, Springer-Verlag, 34-50.
15. C. M. Dafermos, Hyperbolic systems of conservation laws, in *Systems of nonlinear partial differential equations*, (Editor - J. M. Ball), 1983, 25-70, D. Riedel, Dordrecht.
16. J. L. Ericksen, Equilibrium of bars, *J. Elasticity*, **5**(1975), 191-201.
17. J. L. Ericksen, Some phase transitions in crystals, *Archs ration. Mech. Analysis*, **73**(1980), 99-124.
18. J. L. Ericksen, Constitutive theory for some constrained elastic crystals, *Int. J. Solids Struct.* **22**(1986), 951-964.
19. J. D. Eshelby, Continuum theory of lattice defects, in *Solid State Physics*, (Editors - F. Seitz and D. Turnbull), 1956, Academic Press, New York. **3**, 79-144.
20. F. Falk, Model free energy, mechanics and thermodynamics of shape memory alloys, *Acta Metal.*, **28**(1980), 1773-1780.
21. M.E. Gurtin and A. Struthers, Multi-phase thermomechanics and interfacial structure 3. Evolving phase boundaries in the presence of bulk deformation, *Archs ration. Mech. Analysis*, **112**(1990), 97-160.
22. M. Grujicic, G.B. Olson and W.S. Owen, Mobility of the $\beta_1 - \gamma_1'$ martensitic interface in Cu-Al-Ni: Part I. Experimental measurements.*Met. Trans A*, **16A**(1985), 1723-1734.
23. H. Hattori, The Riemann problem for a van der Waals fluid with entropy rate admissibility criterion. Isothermal case. *Archs ration. Mech. Analysis*, **92**(1986), 247-263.
24. W. Heidug and F.K. Lehner, Thermodynamics of coherent phase transformations in nonhydrostatically stressed solids, *Pure Appl. Geophys.* **123**(1985), 91-98.
25. R. D. James, Displacive phase transformations in solids, *J. Mech. Phys. Solids*, **34**(1986), 359-394.
26. Q. Jiang, A continuum model for phase transformations in thermoelastic solids, to appear in *J. Elasticity.*

27. A.G. Khachaturyan and G.A. Shatalov, Theory of macroscopic periodicity for a phase transition in the solid state, *Soviet Physics JETP*, **29**(1969), 557-561.
28. J. K. Knowles, On the dissipation associated with equilibrium shocks in finite elasticity, *J. Elasticity*, **9**(1979), 131-158.
29. R. V. Kohn and S. Müller, Branching of twins near an austenite-twinned martensite interface, *Phil. Mag. (Part A)*, 1991, to appear.
30. R. V. Krishnan and L. C. Brown, Pseudo-elasticity and the strain-memory effect in an Ag-45 at. pct. Cd alloy, *Met. Trans.*, **4**(1973), 423-429.
31. P. Le Floch, Propagating phase boundaries: formulation of the problem and existence via Glimm's scheme, *Archs ration. Mech. Analysis*,1992,to appear.
32. I. Müller and K. Wilmansky, Memory alloys - phenomenology and ersatzmodel, in *Continuum Models of Discrete Systems*, (Editors - O. Brulin and R.K.T. Hsieh), 1981 North Holland, Amsterdam, pp.495-509.
33. I. Müller and H. Xu, On the pseudo-elastic hysteresis, *Acta Met. Materialia*, **39**(1991), 263-271.
34. Z. Nishiyama,*Martensitic Transformation*,1978, Academic Press, New York.
35. K. Otsuka, C. M. Wayman, K. Nakai, H. Sakamoto, and K. Shimizu, Superelasticity effects and stress-induced martensitic transformations in Cu-Al-Ni alloys, *Acta Met.* **24**(1976), 207-226.
36. K. Otsuka and K. Shimizu, Pseudoelasticity and shape memory effects in alloys, *Int. Metals Reviews*, **31**(1986), 93-114.
37. R. Pego, Phase transitions in one-dimensional nonlinear viscoelasticity: admissibility and stability, *Archs ration. Mech. Analysis*, **97**(1986), 353-394.
38. H. Pops, Stress-induced pseudo-elasticity in ternary Cu-Zn-based beta prime phase alloys, *Met. Trans.*, **1**(1970), 251-258.
39. M. Pitteri,Reconciliation of the local and global symmetries of crystals, *J. Elasticity*, **14**(1984), 175-190.
40. J. R. Rice, Continuum mechanics and thermodynamics of plasticity in relation to microscale deformation mechanisms, in *Constitutive Equations in Plasticity* (Editor - A. S. Argon), 1975, MIT Press, 1975, pp. 23-79.
41. A. L. Roitburd, Martensitic transformation as a typical phase transformation in solids, *Solid State Phys.* **33**(1978), 317-390.
42. M. Shearer, The Riemann problem for a class of conservation laws of mixed type, *J. Diff. Eqns.* **46**(1982), 426-443.
43. M. Shearer, Nonuniqueness of admissible solutions of Riemann initial value problems for a system of conservation laws of mixed type, *Archs ration. Mech. Analysis*, **93**(1986), 45-59.
44. S. A. Silling, Numerical studies of loss of ellipticity near singularities in an elastic material, *J. Elasticity*, **19**(1988), 213-239.
45. S. A. Silling, Phase changes induced by deformation in isothermal elastic crystals, *J. Mech. Phys. Solids*, **37**(1989), 293-316.
46. M. Slemrod, Admissibility criteria for propagating phase boundaries in a van der Waals fluid, *Archs ration. Mech. Analysis*, **81**(1983), 301-315.

47. C. Truesdell, *Rational Thermodynamics*, 1969, Springer-Verlag, New York.
48. C. Truesdell and W. Noll, *The nonlinear field theories of mechanics*, Handbuch der Physik, III/3, (Editor-S.Flügge), 1965, Springer, Berlin.
49. L. Truskinovsky, Structure of an isothermal phase discontinuity, *Soviet Phys. Dok.* **30**(1985), 945-948.
50. Z. Z. Yu and P. C. Clapp, Growth dynamics study of the martensitic transformation in Fe-30alloys: Part II. Computer simulation of martensitic growth, *Met. Trans. A*, **20**(1989), 1617-1629.

Theoretical and Applied Mechanics 1992
S.R. Bodner I Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.

1993 IUTAM.

Propagating instabilities in structures

Stelios Kyriakides

Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, Austin, Texas 78712, USA

SUMMARY

Instability is one of the factors which limit the extent to which structures can be loaded or deformed and plays a pivotal role in the design of structures loaded in compression but also in tension. The major issue of practical concern is the establishment of the critical buckling load and deformation, that is, the lowest load or smallest deformation at which alternate equilibria become possible. This is usually established through linearized bifurcation analyses. In practice, the critical condition can be affected by imperfections, material nonlinearities and other factors. Incorporating these in the problem usually requires solution of the more complete nonlinear problems.

This classical approach, that is, concern about the onset of instability, is sufficient for the design of most structures. In the past fifteen years, a class of structural instabilities have been identified in which the classical approach is not sufficient. In general, these affect structures of larger size in which, following the onset of instability, collapse is confined to a relatively small part of the structure, that is, it localizes. However, under prevailing conditions, the initially local collapse can propagate, or spread, often in a dynamic fashion, over the rest of the structure. The load required to propagate such instabilities is often substantially lower than that required to initiate them in the original structure.

In practice, propagating instabilities are initiated from local imperfections which locally weaken the structure and cause local changes in its geometry (collapse). Very much like a set of dominoes once the geometric integrity of such structures is compromised locally, the instability has the potential of spreading over the whole structure.

The aim of this lecture is to demonstrate the breadth of relevance of the phenomenon by pulling together results from studies of four structural examples which exhibit this behavior. Although the practical relevance of the examples varies from insignificant to very significant, each demonstrates different aspects of this behavior.

The first problem is that of finite inflation of an elastic tube. This problem demonstrates the complexity of localization-type instabilities, which in this case takes the form of a bulge, and the subsequent propagation or spreading of the bulge. The problem, once understood, can be solved quite adequately with relative ease using finite-deformation membrane theory. Thus, it provides a model problem for introducing some of the features of this type of structural behavior.

The second problem is that of propagating buckles in offshore pipelines used for transporting oil and gas. Under some circumstances, a local disturbance can initiate a collapse which can propagate dynamically along the length of the line and totally destroy it. In this case the problem is governed by interaction between geometric nonlinearities with inelastic characteristics of the pipe material. The understanding of many aspects of this problem has over the last fifteen years reached a certain degree of maturity and the progress achieved to date will be reviewed.

The third problem is that of localized collapse and the subsequent spreading of this collapse in shells lining stiffer cylinders. This problem has the additional nonlinearity of contact. It can occur for linearly elastic as well as elastic-plastic shells. The fourth problem involves a long panel with an arch cross section, loaded by uniform pressure. It will be demonstrated that following the classical onset of instability the panel can develop a localized buckle which then can propagate at a relatively low pressure and collapse the whole structure.

It will be shown that the underlying common characteristic of structures exhibiting this class of instabilities is a nonlinear, local load-deformation response with two branches with positive slope joined by an intermediate one with negative slope, as shown in Fig. 1. Thus, for some range of loads, such structures have three possible equilibria for each value of load. The one on the first ascending branch will represent prebuckling deformations. The one on the second ascending branch will represent buckled configurations. Under some conditions, the two states can co-exist in the same structure.

Details of each problem including experimental and analytical/numerical simulations as well as a comprehensive literature review can be found in the review article in the reference.

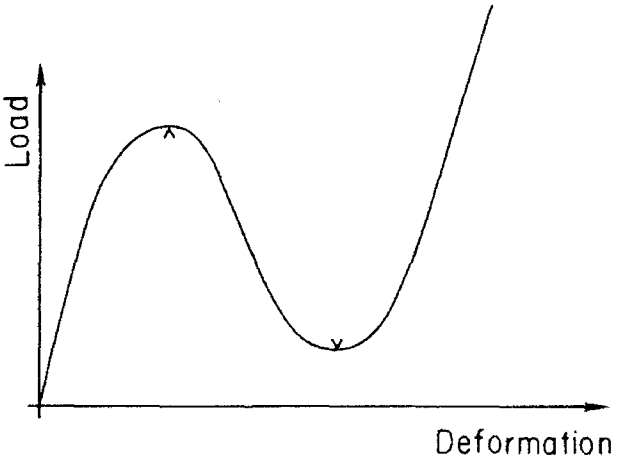


Figure 1. Local "load-deformation" response characteristic of this class of problems

REFERENCE

S. Kyriakides, Propagating Instabilities in Structures. In, Advances of Applied Mechanics, T.Y. Wu and J.W. Hutchinson (eds.), Vol. 30, Academic Press, Boston, Massachusetts, 1993.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
1993 IUTAM.

COMPUTATIONAL APPROACHES TO PLASTIC INSTABILITY IN SOLID MECHANICS

Yoshihiro TOMITA

**Department of Mechanical Engineering, Faculty of Engineering,
Kobe University, Nada, Kobe 657, Japan**

Abstract

The paper presents computational approaches to predict the onset and growth of such plastic instability phenomena as buckling, necking and flow localization. The regularization schemes to remedy the problems associated with spurious mesh sensitivity and incorrect convergence in finite element prediction of flow localization are discussed. The instability of thick circular tubes deformed under pressure and combined loading, neck and bulge propagation in polymeric materials, wrinkling of plates and shells under forming processes, flow localization of thermo-elasto-viscoplastic materials under various deformation rates, and flow localization of mono- and polycrystalline solids are briefly reviewed.

1. INTRODUCTION

Instability phenomena in plastic solids under deformation present themselves in various ways. The decrease in the stiffness due to the geometrical change and/or material softening caused by the deformation is responsible for the occurrence of such instability phenomena as buckling, necking and shear banding. Because the formation of these instabilities is an important precursor to collapse or failure, the computational prediction of the onset and growth of these instabilities is indispensable to understanding the final strength of the structures and materials.

The purpose of the present paper is to provide an introduction to computational predictions relating to plastic instability in solid mechanics. The basic equations, including the virtual work principle and the weak form of the energy balance equation, and constitutive equations covering various aspects of material response are first outlined. To clarify the instabilities associated with the bifurcation which governs the load-carrying capability, and the forming limit of the thin-walled bodies, the specializations and extensions of Hill's theory of uniqueness and bifurcation [1] to cover these specific problems will be discussed. The computational strategies for the prediction of deformation behavior deep in the postbifurcation range are given.

With regard to shear-band-type localization, since the deformation process manifests itself as autocatalytic, and the problem likely becomes an ill-posed one, the computational results exhibit an inherent mesh dependence. Several adaptive remeshing schemes to resolve the sharp strain gradients, and regularization procedures proposed to remedy the mesh-dependent features of the computational results are briefly reviewed.

As illustrative examples, instability behaviors associated with the thick circular tubes under pressure and combined loading, instability propagation in polymeric materials, and wrinkling behavior of plates and shells under different forming processes are briefly reviewed. Subsequently, the flow localization behavior of thermo-elasto-viscoplastic material under various deformation rates will be presented. We close with direct physical approaches to flow localization of mono- and polycrystalline solids which strongly depend on the evolution of microstructure, and their interaction.

2. BASIC EQUATIONS

2.1 Governing Equations

Consider an equilibrium state for a body with volume V and surface S subjected to a velocity constraint on S_v and traction on S_t. An updated Lagrangian formulation is employed where each material particle is labeled by a set of curvilinear coordinates x^i with metric tensor G_{ij}, which are embedded in the body in the current state. The weak form of the equation governing the rate of stress and traction yields the virtual work principle [1,2],

$$\int_V (\dot{S}^{ij} + \sigma^{ij} V^i{}_{;l})\, \delta V_{i;j} dV = \int_{S_t} \dot{P}^i\, \delta V_i dS, \tag{2.1}$$

where δV_i is the virtual velocity satisfying the homogeneous boundary condition over surface S_v. Here, S^{ij} is the Kirchhoff stress, which is identical to Cauchy stress σ^{ij} in the current configuration. An overdot denotes a material time derivative and $(\,)_{;l}$ denotes the covariant derivative with respect to current coordinates. Generally the configuration-dependent part of traction rate $\dot{P}_i$ is a homogeneous function of degree one in the velocity V_i and velocity gradient $V_{i;j}$ [3] which has a potential for a specific case. We can generalize this equation to a dynamic case and a case with residual non equilibrium force at the previous step.

Meanwhile, the weak form of the energy balance equation for the same body subjected to heat flux $q=-n_i q^i=Q$ on S_q with the surface unit normal n_i and temperature constraint on S_T is

$$\int_V \delta T \rho c \dot{T} dV + \int_V \delta T_{,i}\, \kappa^{ij} T_{,j} dV = \int_V \delta T \alpha w^p dV + \int_{S_q} \delta T Q dS, \tag{2.2}$$

where ρ, c and κ^{ij} are the mass density, specific heat and thermal conductivity tensor, respectively. Here, δT is the variation of the temperature which satisfies the homogeneous boundary condition on S_T. The fraction of irreversible work, w^p, which is converted to heat is α, where α is in the range of 0.85 to 0.95 for many metals. The specific form of heat flux Q depends on the respective boundary conditions.

2.2 Constitutive Equations

Since instability predictions strongly depend on the assumed material response, here constitutive equations which have been employed in instability analysis are briefly summarized. In elastic-plastic material response, the plastic part, D^p_{kl}, of the strain rate, $D_{kl}=(V_{k;l}+V_{l;k})/2$, is specified through various kinds of constitutive equations. Elastic material response takes the form of a linear relationship between the elastic strain rate, $D^e_{ij}=D_{ij}-D^p_{ij}$, and a suitable objective stress rate, $\mathring{S}^{kl}$. Then the elastic-plastic constitutive equation can be expressed by

$$\mathring{S}^{ij} = L^{ijkl} D_{kl}, \tag{2.3}$$

where L^{ijkl} is the constitutive tensor and depends on the current stress and the deformation histories.

Within the rate-independent response, for the rather wide class of materials which obey the flow rule, D^p_{ij} can be expressed by

$$D^p_{ij} = n_{mn} \mathring{S}^{mn} m_{ij}/h, \tag{2.4}$$

where n_{mn}, m_{ij} and h are the normal to the yield surface, direction of plastic strain rate and hardening modulus, respectively. The constitutive equations due to J_2 flow [1,4,5] and kinematic hardening [6] theories and to many anisotropic theories (see refs.7,8) fall into the special case of Eq.(2.4) with $n_{ij}=m_{ij}$. In Rudnicki and Rice [9], for material where $n_{ij} \neq m_{ij}$ and $\mathring{S}^{mn}$ is replaced by $\mathring{\sigma}^{mn}$, the specific form of n_{ij} and m_{ij}, for example, can be determined by the

Gurson-type yield function [10,11]. Several extended versions accounting for the kinematic hardening [12], the change in elasticity modulus due to void volume fraction [13] and the interaction of voids [14] were also suggested.

The deformation-type constitutive equation by Budiansky [15] was focused on as the simplest model which permits the development of a corner on the yield surface. It has been generalized to account for the finite strain [16-18] and anisotropy [19], and extensively employed for buckling- and shear-band-type instability analyses. The plastic strain rate is

$$D^p_{ij}=n_{nm}\mathring{S}^{mn}n_{ij}/h+(\mathring{S}'_{ij}-n_{nm}\mathring{S}^{nm}n_{ij})/h_s, \tag{2.5}$$

where h_s is a new hardening modulus. This constitutive equation is validated for the deformation satisfying the total loading condition [15]. For the strongly nonproportional stress histories, Christoffersen and Hutchinson [20] proposed the corner theory in which angular measure, θ, of the stress rate direction with respect to corner direction of the yield surface is defined. The plastic strain rate can be expressed by

$$D^p_{ij}=\frac{\partial^2 E^p_c}{\partial\mathring{S}^{ij}\,\partial\mathring{S}^{kl}}\mathring{S}^{kl},\quad E^p_c=\frac{1}{2}f(\theta)C_{ijkl}\mathring{S}^{ij}\mathring{S}^{kl}, \tag{2.6}$$

where C_{ijkl} is a plastic compliance tensor of Eq.(2.5). The transition function $f(\theta)$ is unity throughout the total loading range $0\le\theta\le\theta_o$ and zero for $\theta_c\le\theta\le\pi$. Here, $f(\theta)$ decreases from unity to zero as θ increases from θ_o to θ_c. The Bauschinger effect [21] and anisotropy [22] have been introduced in Eq.(2.6).

Constitutive descriptions of single-crystal plasticity in terms of crystallographic shearing on specific slip systems have been given by Asaro and Rice [23], and Asaro [24]. In the current configuration, the plastic strain rate is described by

$$D^p_{ij}=\Sigma\ P^{(a)}_{ij}\ \dot{\gamma}^{(a)},\quad 2P^{(a)}_{ij}=s^{e(a)}_i m^{e(a)}_j+m^{e(a)}_i s^{e(a)}_j, \tag{2.7}$$

where (a) designates the crystal-structure-dependent (a)th slip system with the slip direction $s^{e(a)}_i$ and normal $m^{e(a)}_i$ in the current state. Here, $\dot{\gamma}^{(a)}$ is the shear rate on slip system (a), and the summation in Eq.(2.7) extends over the active slip systems. The constitutive equation for the single crystal can be established by Eq.(2.7) and the evolutional equation for the flow stress on slip system (a). In the description of the hardening moduli, self-hardening and latent hardening are introduced to duplicate the experimental observations [24]. Recently, more elaborate descriptions of latent hardening have been proposed [25].

The materials generally possess strain rate and temperature sensitivities to some extent, and they have an important effect on instability behavior. Generalized versions of the above-mentioned constitutive equations accounting for the thermo-elasto-viscoplastic response of the materials have been established [26,27] and employed in thermocoupled instability analyses.

In the meantime, in order to avoid numerical instability and maintain the required accuracy, suitable integration schemes of constitutive equations are employed [28-30]. Iterative methods, however, which can achieve numerically stable computational processes with large increment, may have the drawback of creating problems associated with the exact account of the deformation history.

3. BIFURCATION AND POSTBIFURCATION ANALYSIS

3.1 Bifurcation Analysis

The bifurcation criterion employed is Hill's theory of uniqueness and bifurcation [1] for elastic-plastic solids. This theory states that the solution is not unique when a nontrivial solution can be found for the eigenvalue problem given by the following variational equation:

$$\delta I=0,\ I=\int_V (L^{ijkl}D^*_{ij}D^*_{kl}+\sigma^{lj}V^{*i}_{;l})V^*_{i,j}dV-\int_{S_t}\dot{P}^{*i}V^*_i dS. \tag{3.1}$$

Here, $(\ )^*$ denotes the difference between the fundamental solution and the second one. The surface integral of Eq.(3.1) arises from the configuration dependence of loading and L^{ijkl} is assumed to be a constitutive tensor for a linear comparison solid. When plastic loading continues throughout the plastic zone in the fundamental solution, the bifurcation point for underlying materials is also given by Eq.(3.1).

Now, when the bifurcation functional in Eq.(3.1) is approximated by the finite element method, we can arrive at the following homogeneous algebraic equation:

$$[K]\{\hat{\delta}^*\}=0, \tag{3.2}$$

where $\{\hat{\delta}^*\}$ denotes the values of V^*_i at the nodal points. When this homogeneous equation has a nontrivial solution, bifurcation may occur. The bifurcation condition becomes

$$\det[K]=0, \tag{3.3}$$

and the corresponding mode is obtained as the eigenmode of Eq.(3.2). At every computational step of fundamental analysis, the change of the sign of det[K] is checked and then an iterative method is used to determine the vanishing point of det[K].

In the course of the deformation analysis of the block and bar, Kitagawa et al. [31,32] found that the sign of the determinant of the stiffness matrix for the finite element model, which has the capability of expressing bifurcation mode, changes at the bifurcation points on the fundamental path, and the associated eigenmode is the corresponding bifurcation mode. This has been successfully applied to various kinds of bifurcation problems. A similar and extended discussion can be seen in refs. 33-35.

Next, we consider an axisymmetric body subjected to axisymmetric deformation. Here, x^1 and x^2 are curvilinear coordinates in a meridian plane and x^3 is circumferential. Nonaxisymmetric admissible velocities V^*_i are considered in the forms

$$V^*_{m1}=\hat{V}^*_{m1}\cos mx^3,\ V^*_{m2}=\hat{V}^*_{m2}\sin mx^3,\ V^*_{m3}=\hat{V}^*_{m3}\cos mx^3, \tag{3.4}$$

where m is the mode number, and $\hat{V}^*_{mi}$ is the function of x^1 and x^2 and independent of x^3. Substituting these into Eq.(3.1), we arrive at the bifurcation functional for mode m by

$$I_m=\int_A (L^{ijkl}\hat{D}^*_{ij}\hat{D}^*_{kl}+\sigma^{lj}\hat{V}^{*i}_{;l}\hat{V}^*_{i;j})rdA-\int_{S_t}\dot{P}^{*i}\hat{V}^*_{mi}rdS, \tag{3.5}$$

where $\hat{D}^*_{ij}$ and $\hat{V}^*_{i;j}$ can be described by $\hat{V}^*_{mi}$ and its derivatives with respect to x^1 and x^2, and mode number m, and so on. Here, r is the radial distance of a generic point in the axisymmetric body. A and S are the cross-sectional area of the body and its boundary, respectively.

The finite element approximation is then employed to describe $\hat{V}^*_{mi}$ over an element, and the corresponding bifurcation condition in Eq.(3.1) becomes

$$\det[K_m]=0\ (m=1,2,3...). \tag{3.6}$$

The first bifurcation point is referred to as the critical bifurcation point, and the corresponding mode is the critical bifurcation mode.

When the materials follow the nonassociative flow law, the bifurcation condition in Eq.(3.1) becomes invalid because of the nonsymmetry of tensor L^{ijkl} in Eq.(2.3). Raniecki and Brunhs

[36] introduced two comparison solids, through which the search for the genuine bifurcation state is replaced by a search for upper and lower bounds. Investigation concerning the improvement of the lower bounds can be seen in [37]. Furthermore, for materials obeying the nonlinear constitutive equation, Tvergaard [38] and Triantafyllidis [39] proposed methods to determine the upper and lower bounds, respectively, for the first bifurcation. See also the recent paper [40].

3.2 Wrinkling Analysis of Thin-Walled Bodies

Consider a bifurcation into a more complicated membrane and bending type of deformation from membrane deformation. Any general point on a thin-walled body can be identified by $(y^I,y^3)(I=1,2)$ where (y^1,y^2) are curvilinear coordinates on the middle plane, and the y^3 axis is perpendicular to the plane. Now, let v_i and V_i be the velocities of the body at $(y^I,0)$ and (y^I,y^3), respectively. According to Naghdi's thin-shell theory [41], velocity V_i can be expressed as

$$V^I=v^I+y^3\,\delta^I \quad , V^3=v^3+y^3\,\delta^3, \tag{3.7}$$

where $y^3\delta_i$ describes the director or relative velocity at a point (y^I,y^3) relative to the velocity at point $(y^I,0)$. Specifically, the Kirchhoff-Love hypothesis for thin shells provides the following relationship:

$$\delta_I=-(v_{3,I}+b^L{}_I v_L),\ \ \delta_3=0, \tag{3.8}$$

where b_{IJ} is the curvature tensor of the middle plane and vanishes for flat plates. When Eq.(3.7) and the corresponding velocity gradient and strain rate are substituted into Eq.(3.1) with $P^{*i}=0$, the bifurcation functional for this problem becomes

$$\begin{aligned} I=\int_A [t\{&\bar{L}^{IJMN}d^*_{IJ}d^*_{MN}+4\bar{L}^{I3N3}d^*_{I3}d^*_{N3}+\sigma^{MJ}(v^{*L}{}_M v^*_{LJ}+v^*_{3M}v^*_{3J})\}\\ &+t^3/12\{\bar{L}^{IJMN}\kappa^*_{IJ}\kappa^*_{MN}+4\bar{L}^{I3N3}\kappa^*_{I3}\kappa^*_{N3}+\sigma^{MJ}(\omega^{*L}{}_M\omega^*_{LJ}+\omega^*_{3M}\omega^*_{3J})\}]dA. \end{aligned} \tag{3.9}$$

The details of the notations are explained in Tomita and Shindo [42]. Tensor $\bar{L}^{ijkl}$ is the plane stress ($\sigma^{33}=0$) constitutive tensor, where

$$\bar{L}^{ijkl}=L^{ijkl}-L^{ij33}L^{kl33}/L^{3333}, \tag{3.10}$$

which is uniform throughout the thickness because of the assumption of a membrane state prior to bifurcation. Adopting the Kirchhoff-Love hypothesis in Eq.(3.8) for the bifurcation functionals in Eq.(3.9), we can get the simplified functional of

$$\begin{aligned} I_k=\int_A [t\{&\bar{L}^{IJMN}d^*_{IJ}d^*_{MN}+\sigma^{MJ}(v^{*L}{}_M v^*_{LJ}+v^*_{M3}v^*_{J3})\}\\ &+t^3/12\{\bar{L}^{IJMN}\kappa^*_{IJ}\kappa^*_{MN}+\sigma^{MJ}(\omega^{*L}{}_M\omega^*_{LJ}+\omega^*_{3M}\omega^*_{3J})\}]dA. \end{aligned} \tag{3.11}$$

Since the in- and out-of-plane deformation modes are coupled with each other, the necking type of mode is affected by the bending type of mode and vice versa. On the other hand, for a plate, we should recall that $b_{IJ}=0$, and we can arrive at the decoupled bifurcation functionals [42].

Due to the complexity of constructing a finite element which satisfies the C^1 continuity requirement for a general shell, the functional in Eq.(3.9) is recommended for use in such problems. One exceptional case is the problem of an axisymmetric shell which has been

discussed [42]. The search for the bifurcation point is the same as that explained before.

3.3 Postbifurcation Analysis

The solution of the boundary value problem at the bifurcation point can be expressed by the sum of the fundamental solution at the bifurcation point and a suitably normalized bifurcation mode. As discussed by Hutchinson [43], the specific amplitude of the bifurcation mode is determined such that loading occurs everywhere in the current plastic zone, except at one point where neutral loading takes place. The highly nonlinear nature of the postbifurcation behavior can be clarified by employing the virtual work principle with the finite element approximation. When the bifurcation point obtained from Eq.(3.1) is not the real bifurcation point, the postbifurcation behavior must be traced by the specially developed method [40] or calculated approximately using bodies with initial imperfections through, again, the virtual work principle. The proper magnitude of imperfection, which depends on the problems and the number of significant figures of the employed computational facility, and of the incremental step must be introduced to suitably predict bifurcation and postbifurcation behavior [31,32,44].

4. FLOW LOCALIZATION ANALYSIS

4.1 Problems Associated with Flow Localization Analysis

Once the flow localization into narrow bands of intense straining starts, problems associated with the incompressibility of the materials, and an excessive distortion of the element due to high deformation gradients appear. Furthermore, due to the ill-posedness of the problem, the numerical solution of localization problems for a rate-independent solid exhibits an inherent mesh dependence [45].

The problem associated with the incompressibility of the materials was comprehensively discussed by Nagtegaarl et al. [46], who proposed the well-known crossed triangular element. Subsequently, many elements with reduced or selective reduced integrations [47] and with hourglass control [48] have been developed. The aspect ratio of the quadrilaterals (crossed triangular elements) is chosen so that they deform by alignment with the most critical shear band direction [49]. By contrast, in the conventional quadrilateral elements, suitable hourglass control must be introduced to capture a sharp shear band [29].

Regardless of the elements employed, the elements in the localized zone are excessively distorted and the performance of the approximation of the velocity distribution becomes poor. For this problem, Ortiz and Quigley [50] proposed the adaptive h-method in which the refinement of the mesh is performed in specific regions such that the variation of the solution is within a prescribed tolerance over the element in the mesh. An adaptive remeshing algorithm to produce constant error finite elements was also proposed [51]. In the adaptive method, a proper transfer operator must be defined to correctly redistribute the variables for the new mesh [50]. However, the problem associated with mesh dependence still remains an open question. This will be discussed next.

4.2 Localization Analysis with Regularization

The problem associated with the ill-posedness can be resolved by introducing a suitable length scale. The viscoplastic regularization is aimed at introducing the length scale by the strain rate sensitivity of the material to make the original problem into a well-posed one [45,52,53]. Without regularization, the widths of the computed shear bands are set by the mesh spacings, as observed in many localization analyses. However, for mesh sizes which appropriately resolve the localization with suitable regularization, convergence occurs with the refinements [53]. Thus viscosity can be a regularization parameter from a computational point of view or a micromechanical parameter to be determined from observed shear-band width from a physical point of view [53]. The possibility of the existence of a stable stationary shear band of a finite thickness has been shown [35]. Nevertheless, when a flow localization developes sufficiently, the localization zones again exhibit mesh dependency [35,54,55].

The length scale can be naturally introduced by employing the nonlocal continuum. When the magnitude of inelastic strain rate $\mathbf{D}^p$ at location $\mathbf{x}$ is defined by spatial averaging while the elastic behavior is treated as local [54], the elastic-plastic constitutive equation can be expressed by Eq.(2.3). Then the same finite element model can be used [56]. For damaging materials, Bazant and Cabot [56] defined the nonlocal damage parameter expressed by spatial averaging as in the plastic strain rate $\mathbf{D}^p$ and succeeded in the substantial reduction of the mesh sensitivity of the flow localizations.

Aifantis [57] introduced higher strain gradient terms in the flow stress to represent the microstructural dependent nature of the material response. This provides a natural length scale to the problem and makes it possible to investigate the structure of shear bands [58,59]. Using strain-gradient-dependent flow stress for the thermal softening material

$$\bar{\sigma} = \bar{\bar{\sigma}}(\bar{\varepsilon}, \dot{\bar{\varepsilon}}, T) - c\nabla^2\bar{\varepsilon}, \tag{4.1}$$

Tomita and Nakao [60,61] showed that the finite element predictions of the evolution of shear band thickness and strain distribution can duplicate the analytical results [59] without mesh sensitivity. Figure 4.1 shows the finite element mesh-size-independent prediction of shear localization and the specimen-size-independent nature of shear banding [60,61]. It has been proven numerically that the incorporation of the length scale in a different manner equally eliminates the mesh sensitivity of the flow localization process when the length scales are suitably estimated [62].

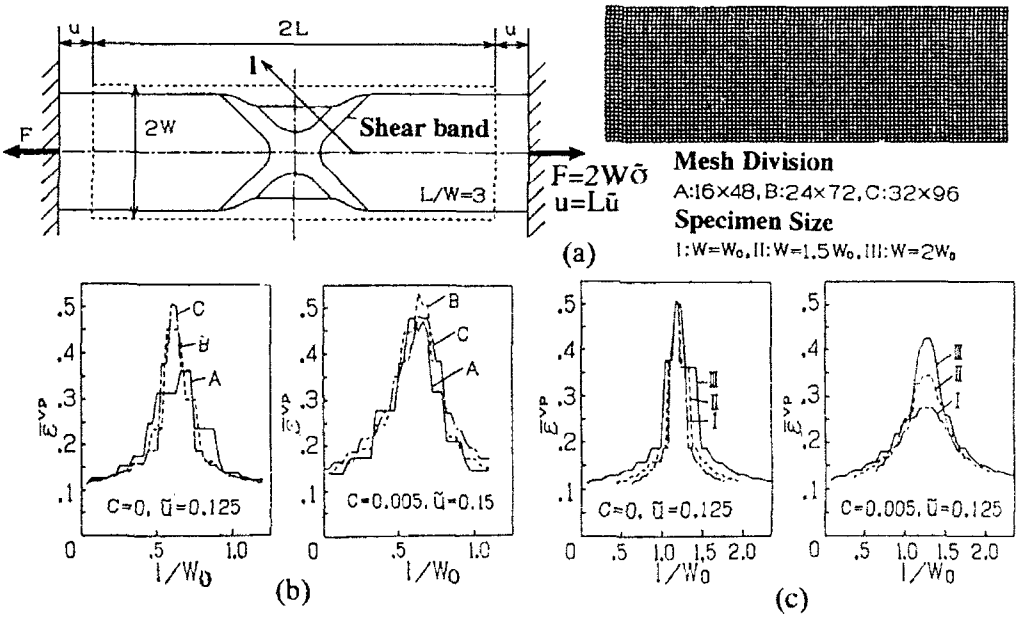


Figure 4.1 Representative strain distribution in the shear band. (a) Computational model, (b) Effect of mesh division with specimen size I, (c) Effect of specimen size with mesh division C, c: positive parameter in Eq.(4.1). (from [60]).

Numerical simulations using a Cosserat continuum [63] and a constitutive equation of dipolar materials [64] clarified that the strain profile over the localization zone and the magnitude of energy required for deformation converge to certain unique values for static as well as for transient loading.

It is noted that a more rigorous investigation concerning the introduction of the higher-order gradient term is necessary [65]. Furthermore, the problem associated with the identification of length scales is still open. It would be natural to relate them to the size of the real localization zone.

To resolve the sharp strain gradients, it has been proposed to add a higher-order deformation mode to the conventional finite element [66,67]. In the element, velocity fields containing a shear band/ plane can be expressed by

$$v_i = \Sigma N_a v_{ia} + \Sigma M_{ib} \dot{\gamma}_b, \tag{4.2}$$

where the first term is a C^o finite element interpolation and the second stands for the incompatible term expressed by the localized mode amplitude $\dot{\gamma}_b$ and the shape function M_{ib}. The direction of the shear band/plane is estimated at the Gauss integration points by the shear band analysis [68] and the linear perturbation analysis [69] for strain-rate-independent and -dependent materials, respectively.

When the thickness of the shear band is smaller than the size of the element, an embedded formulation where the bandwidth or the profile of the strain fields is assumed to be known a priori [70] restores the well-posedness of the problem for strain-rate-independent materials. For strain rate-dependent materials, since the problem is wellposed, the enriched formulation [71] with a spectral interpolation is used to approximate the localized strain fields. Then the mesh dependence is substantially reduced in the embedded element and when used for a coarse mesh, it predicts the effective strain much better than the standard element. Recently, the ability of different finite element enhancements to capture localization was discussed [72].

5. PLASTIC INSTABILITIES IN SPECIFIC PROBLEMS

5.1 Instability Behavior of Circular Tubes

The problem of instabilities in elastic-plastic tubes has been investigated by a number of authors. The instability at the maximum pressure point is overcome by volume loading [73]. The bifurcation points for internal pressure [74,75] and external pressure[76] under axially plane strain conditions, and for combined loading of internal pressure and axial force [44,77] have been obtained by employing Eq.(3.5).

The initial-to-medium postbifurcation behavior [75,76] and localization of the deformation accompanied by shear bands [78] have been clarified under axially plane strain conditions. Tomita et al. [22,44] investigated the loading-path-dependent bifurcation and postbifurcation behavior of tubes subjected to axial tension and internal/external pressure. The influence of axial loads, the Bauschinger effect and corner structure on the yield surface on the onset and growth of surface instability and shears band were investigated. In these analyses, to avoid instability in the numerical analysis in the vicinity of the maximum pressure point, computations were performed under displacement control [79]. Quite recently, Tvergaard [80] predicted the nonaxisymmetric modes leading to localized necking and the subsequent shear fracture on one side of the tube [78] in connection with secondary bifurcation points after the development of an axisymmetric localized bulge.

When an elastic-plastic circular tube is drawn axisymmetrically through a frictionless die, bifurcation with short circumferential wavelength over the traction-free inner surface, a surface-type bifurcation [81,82], or long wavelength mode, becomes possible [83]. As the tube deforms further beyond the bifurcation point, the surface wave will grow into a considerable surface undulation and ultimately may lead to the development of a shear band connecting the highly strained regions beneath the traction-free surface in a zigzag fashion [84]. In the composite tube [85], besides the surface-type bifurcation, the boundary-type bifurcation with large mode amplitude at the boundary plane of two materials appears. Figure 5.1 clarifies that the boundary-type bifurcation may cause strong strain localization which starts near the plane and extends on either side of the plane.

5.2 Instability Propagations in Polymeric Material

The mechanical aspects of instability propagation in polymeric material have recently received much attention. Hutchinson and Neale [86] and Chater and Hutchinson [87] clarified the characteristic feature of the propagations of necking in tension blocks, bulging in balloons

and buckling in tubes under lateral pressure in terms of simplified analyses. In further studies, full finite element analyses for solid circular bars [88], plane strain blocks [89] and tubes [90] have been conducted. Figure 5.2 shows quasi-static neck and bulge propagations with respect to circumferential and axial directions, respectively, in cylinders under plane strain and axisymmetric deformation conditions, respectively, with internal pressure. It is noted that the thickness nonuniformity induced by neck propagation along the circumferential direction diminishes when the propagation of axisymmetric necking and bulging begins. For further discussion concerning buckle propagation, see Kyriakides [91].

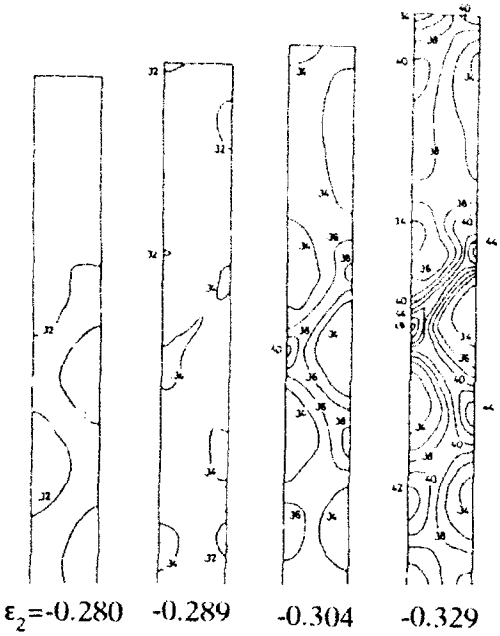


Figure 5.1 Representative strain distribution for different stages of deformation.
ε_2: average strain (from[85]).

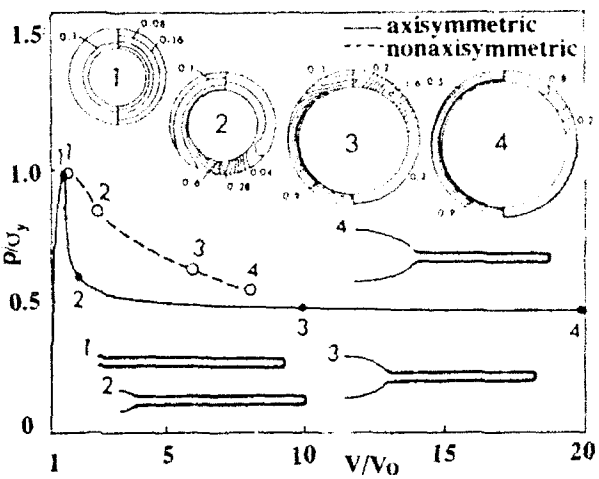
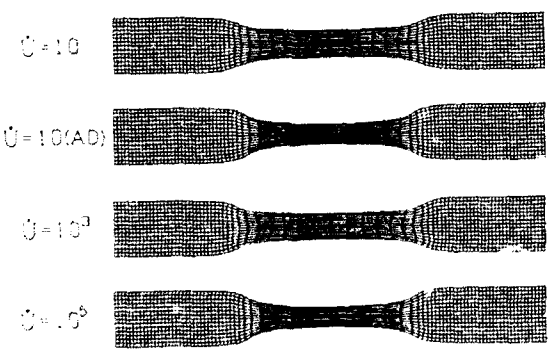
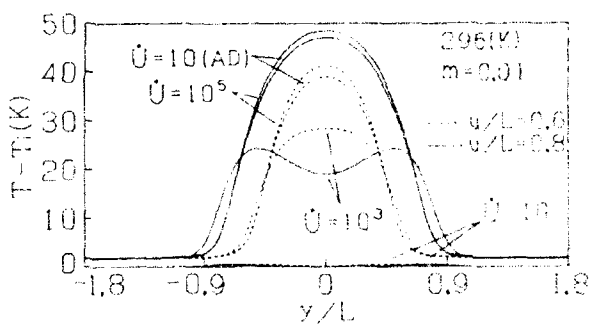


Figure 5.2 Neck and bulge propagations of circular tubes under internal pressure p. Plastic strain (left) and its increment (right) distributions (from [90]).



(a)



(b)

Figure 5.3 Thermo-elasto-viscoplastic neck propagation. (a) Neck profile, (b) Temperature distribution along tension axis y. u: end displacement, 2L: initial length, T· initial temperature, Average strain rate: $2*10^{-5}$Ů, AD: Locally adiabatic deformation (from [94]).

In a subsequent study, the effects of strain rate sensitivity [92] and the temperature dependence [93,94] on the neck propagation behavior were investigated. Figure 5.3 shows the thermo-elasto-viscoplastic neck propagation behavior. At a low rate of deformation, stabilization effects by upturns in the uniaxial stress-strain relation overcome the destabilization

effect due to thermal softening, and neck propagation with a heat source can be observed. The deformation-induced heating and its conduction strongly affect the neck propagation behavior for a relatively low rate of deformation; predictions with adiabatic assumptions do not provide proper results.

Due to the nonuniform and three-dimensional nature of the deformation behavior of polymeric material, identification of the constitutive response from the load elongation curve is quite difficult [95] and requires both precise experiments and computational simulations [94]. Furthermore, anisotropy caused by the microscopic mechanisms of the molecular chains and to the distribution of their orientations due to the plastic deformation is quite important and can be characterized by the molecular orientation developing with plastic straining [96].

5.3 Wrinkling Formation During Forming Processes

In connection with deep drawing of circular cylinders, the nonaxisymmetric wrinkling behavior of a flat flange was first investigated by employing the simplified bifurcation functional (3.11) with $b_{ij}=0$ in conjunction with admissible velocities v^*_i (3.4) [97]. Subsequently, it extended to the wrinkling behavior of general axisymmetric shells [98]. The same or similar procedures have been employed to clarify the effect of material characteristics and the restraining effect of the blank-holding force [97,99,100], die angles [97], normal anisotropy and details of die geometry [98] on the onset of wrinkling during the deep drawing processes. Figure 5.4 shows nosing ratios χ_w and χ_{wc} at the wrinkling and critical wrinkling points, respectively, versus the initial thickness to mean radius ratio (t_o/R_o), and the number of wrinkles m for different die semi-angles α. The typical wrinkle modes are illustrated in the figure.

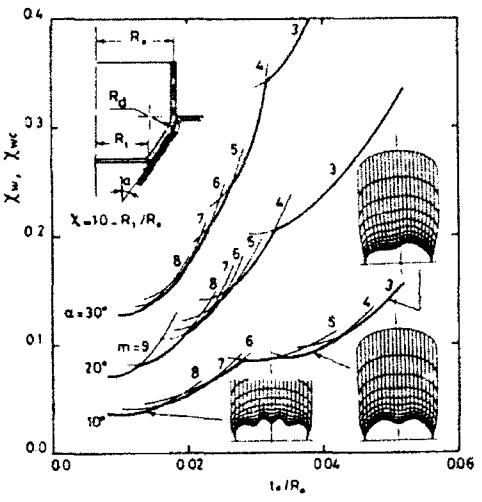
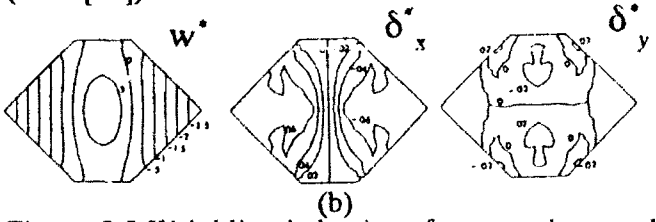
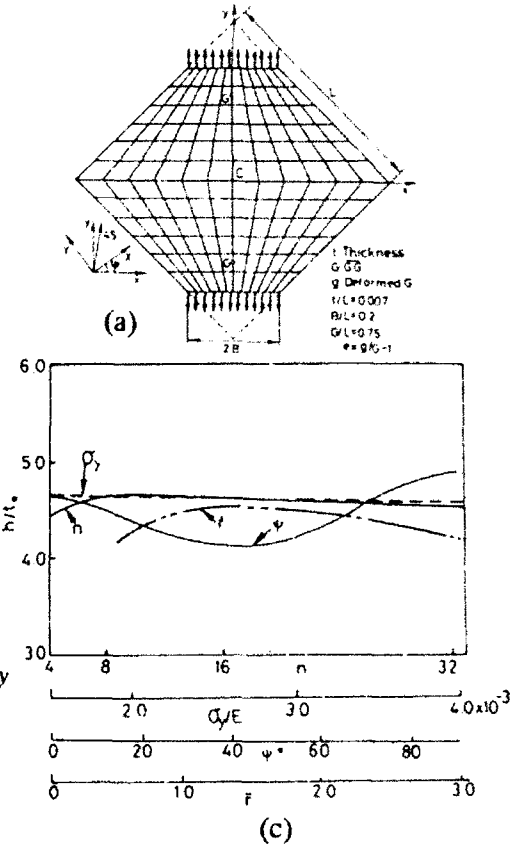


Figure 5.4 Nonaxisymmetric wrinkling of circular tubes subjected to nosing operation.
χ_w, χ_{wc}: wrinkling and critical wrinkling points (from [98]).



(b)



(c)

Figure 5.5 Wrinkling behavior of square sheet under diagonal tension. (a) Computational model, X, Y: orthotropic axes, (b) Wrinkling modes, w^*, δ^*_x, δ^*_y: deflectional velocity, rotational velocities around y and x axes, respectively, (c) Effect of work hardening exponent n, yield stress σ_y, normal anisotropy $\bar{r}$ and tensile direction Ψ in (a) on wrinkle height h at e=0.02 (from [21]).

In stretch forming with a draw bead, the fundamental deformation path includes the strong deviation from the proportional loading path or unloading, and the bifurcation analysis mentioned in section 3.1 with corner theory [20] is performed [39]. The strong deviations from proportional loading have an important stiffening effect in the material, and thus the onset of wrinkling is considerably delayed as compared with that predicted by the deformation theory [39]. Furthermore, the results clarified that the draw bead restrains puckering and promotes the onset of localized necking, which shows good agreement with those of experiments [101].

Stretching a square sheet along one of its diagonals (see Fig.5.5), which has been developed to simulate the wrinkling behavior in press forming of auto body panels, has been investigated [19,21] by employing the simplified bifurcation functional in Eq.(3.9) with $b_{ij}=0$ in conjunction with corner theory [20] with Hill's orthotropic yield function [102]. A substantial delay of wrinkling is observed when the plate is stretched to the minimum r value direction. The effect of the orthotropy of a material on wrinkling behavior is suppressed as compared with that predicted with the flow theory type of constitutive equation [19].

To investigate the growth rate of wrinkles and the residual height of wrinkles, three-dimensional shell analysis with a slight initial imperfection which has a similar shape to that of the wrinkling mode shown in Fig.5.5(b) has been performed. Figure (c) shows dependency of wrinkle height h on material parameters and tensile direction. The same tendency is observed for residual wrinkle height. The same problems have been recently traced [103].

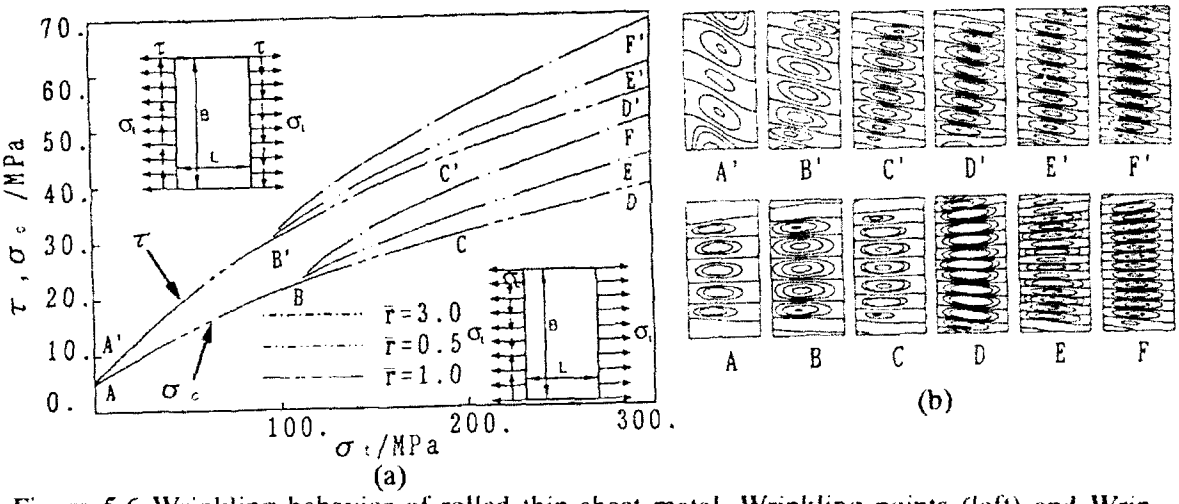


Figure 5.6 Wrinkling behavior of rolled thin sheet metal. Wrinkling points (left) and Wrinkling modes (right) (from [104]).

The same procedure may be applied to the wrinkling behavior in sheet metal rolling in Fig.5.6. Wrinkles at 45 degrees and 0 degree with respect to the rolling direction are called cross and vertical buckles, respectively. Figure 5.6 shows that the compressive force due to friction or shear force induced by the slight inclination of the work roll acting at both ends of the sheet is responsible for the formation of these wrinkles [104]. A different type of wrinkling is caused by nonuniform tension due to rolling with a crown roll [104].

As discussed in section 3.2, since the interaction between the wrinkling- and necking-type bifurcations may be automatically introduced for sheets with curvature, the effect of sheet geometry and material properties on the critical stress states for wrinkling and necking can be clarified. Furthermore, when the wrinkling is confined to a local region of the sheet metal and such a region can be isolated from the remainder of the sheet metal, the curvature and the thickness-dependent forming limit diagram (FLD) [105] and wrinkling limit diagram (WLD) [106,107] may be established. WLD, FLD and information from the sheet forming codes based on the membrane theory will provide information concerning the onset of wrinkling and necking.

Most of the problems associated with sheet metal forming with a tool fall into the category of unilateral contact [108,109], which is not discussed here.

5.4 Flow Localization of Thermo-Elasto-Viscoplastic Solids

Real strain-rate-dependent flow localization manifests itself as different features depending on the rate of deformation, whereas it is essentially the same under the isothermal condition [110,111], meaning that thermocoupled analysis is inevitable. Plane strain tension [29,30,55,60,61,112], compression [26], and plane strain compression near a stress-free surface [27] have been analyzed without the inertial effect by assuming an adiabatic process which represents the upper bound of the temperature [29,112], and by accounting for heat conduction [26,27,55,60,61]. Plane strain quadrilateral elements with hourglass control [29,113] and crossed triangular elements [26,27,55,60,61] are employed. An intensification of shear localization has been observed for larger specimens, and the adiabatic assumption may provide suitable information only in the specific order of the strain rate which increases as the specimen size decreases. Localization of the deformation is delayed by the strain rate effect and strain gradient dependence of the yield stress, as seen in Eq.(4.1), while increase in the curvature of the flow potential surface, heat generation and noncoaxiality of the strain rate to deviatoric stress have the opposite effect. The results also illustrate that the effect of the interaction of material properties and thermal softening and growth of voids are two competing and interacting softening mechanisms in porous material.

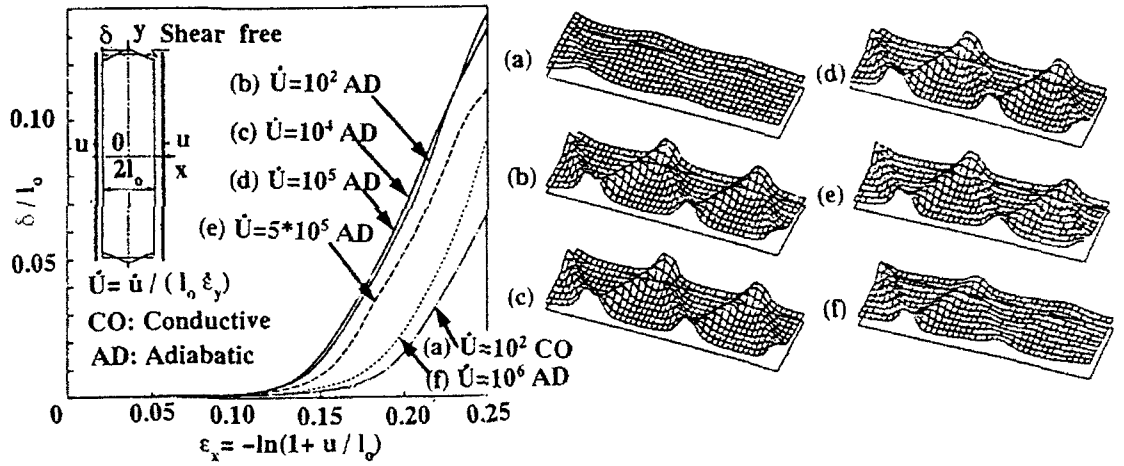


Figure 5.7 Growth of surface undulation (left) and representative viscoplastic strain distribution (right) for dynamic compression of blocks near stress free surface (from [116]).

Needleman [114] has performed dynamic flow localization analyses for plane strain compression with the strain softening constitutive equation as a simple model for a thermally softening solid. This equation phenomenologically describes the various microscopic softening mechanisms including damage due to voids and microcrack development, and texture softening. As a result, the problem can be treated from a purely mechanical point of view. Except for a significant delay of shear band development due to the inertial effect, the main features of shear band development are the same as under quasi-static loading conditions. Batra and Liu [115] investigated a similar problem by introducing a temperature bump at the center of the block showing thermally softening viscoplastic response.

Figure 5.7 shows the compression of the block near the stress-free surface [116]. Deformation process with heat conduction (CO) and locally an adiabatic condition (AD) have been assumed. A competing effect of thermal softening and inertia is observed in the evolution of undulation and representative strain. According to the experimental observation, flow stress increases with abrupt increase in strain rate [117,118]. The effect of this material strain rate history dependence on flow localization behavior has been investigated [116]. Figure 5.8

depicts the results of dynamic compression of plane strain blocks. Differences observed in representative strain distribution (c) for II and III are attributed to the strain rate history dependence of the material, which certainly stabilizes the deformation. Furthermore, the comparison between two results for I and II clarifies that the dynamic deformation subsequently applied to the quasi-static deformation stabilizes the deformation. Meanwhile, the effect of strain rate sensitivity exponent m on the flow localization behavior is quite substantial [29,116]. In Nemat-Nasser et al. [29], m was switched to unity at a certain strain rate so that the bands were broader and tended to disappear at very high strain rate. For dipolar materials, the localization of deformation is significantly delayed, and localized deformation zones are wider as compared with those for nonpolar materials [119].

The three-dimensional aspects of localized deformation without inertial effect [120] and with inertial effect [69,113] have been investigated. Zbib and Jubran [113] clarified the smooth transition of plane stress to plane strain deformation by employing very thin to thick specimens. The orientations of the shear band are 35.25 and 45 degrees, and they are consistent with theoretical predictions. Again, a softening mechanism and an initial imperfection are the main causes of shear banding. The multiaxial effect stabilizes the deformation and causes delay of the localization [113,120].

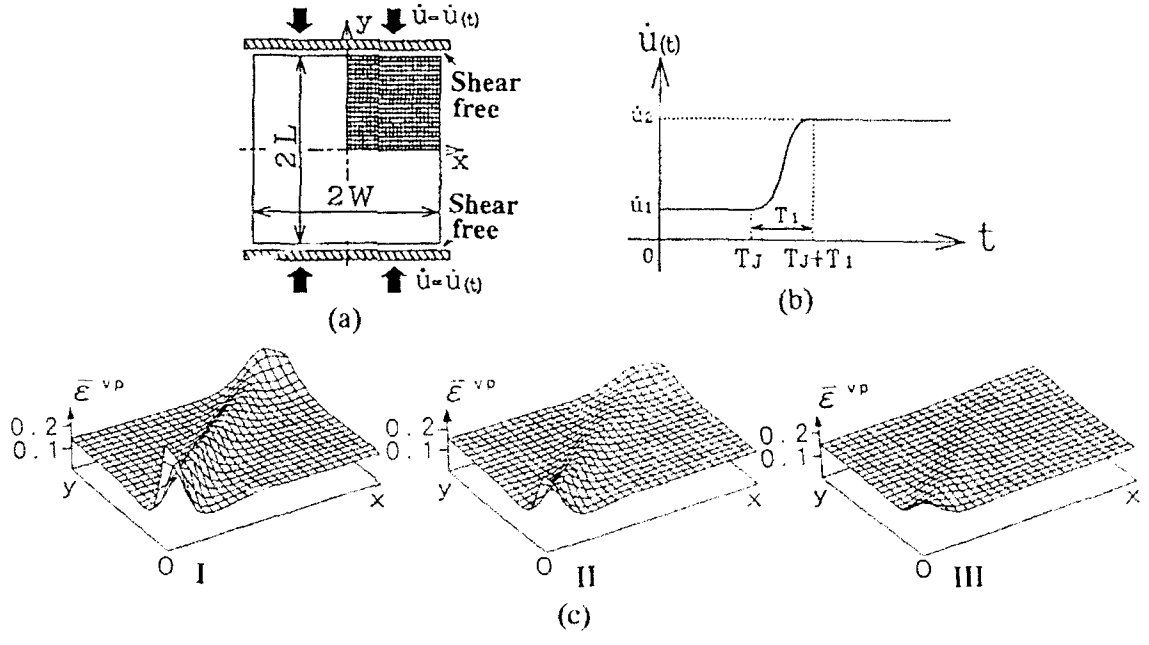


Figure 5.8 Dynamic compression of blocks under plane strain condition. (a) Computational model, W/L=1, (b) Boundary condition applied at both ends, $\dot{u}_1/L=0.002/s$, $\dot{u}_2/L=10^4/s$, $T_1=10^{-7}s$, I: constant displacement rate ($\dot{u}=\dot{u}_2$), II: displacement rate jump ($\dot{u}=\dot{u}_1$ to $\dot{u}=\dot{u}_2$ at $t=T_J$, corresponding end displacement u/L=0.05) without strain rate history dependence, III: II with strain rate history dependence, (c) Representative strain distribution at u/L=0.125 (from [116]).

5.5 Flow Localization of Mono- and Polycrystalline Solids

Physically observed features of flow localizations induced by the evolution of microstructure and their interaction are quite complicated, and hence the limitations of the phenomenological approaches may be undeniable. In such a situation, an essential procedure is the direct physical approach.

In the crystalline model, constitutive equations originated by Asaro and Rice [23] and Asaro [24] have been extensively employed. Peirce et al. [121] carried out a finite element analysis of macroscopically localized deformation in single crystals deformed by tension by employing the two-dimensional double-slip model of Asaro [24]. The results clarified the geometrical

softening due to lattice rotation, and the effect of self- and latent hardening on shear localization, which are in close agreement with experiments. When the single crystals deform under a high strain rate, thermal softening coupling with geometrical softening causes substantial acceleration in shear band initiation [122]. Kitagawa and Matsushita [123,124] clarified that when the number of slipdirections increases, the shear banding behavior becomes close to that for J_2 deformation theory.

Figure 5.9(a) shows mode I crack tip displacement in a silicon iron plate deformed under plane strain conditions observed experimentally [125], and (b),(c) depict the displacement vector and contour of representative strain calculated using a single-crystalline model, respectively [126]. Although the material properties, crack shape and loading conditions are not always the same, good correspondence can be found concerning the formation of the sharp corner at the crack tip and the manner of spread of the deformation field, which is limited to a narrow band emanating from the tip in a radial direction.

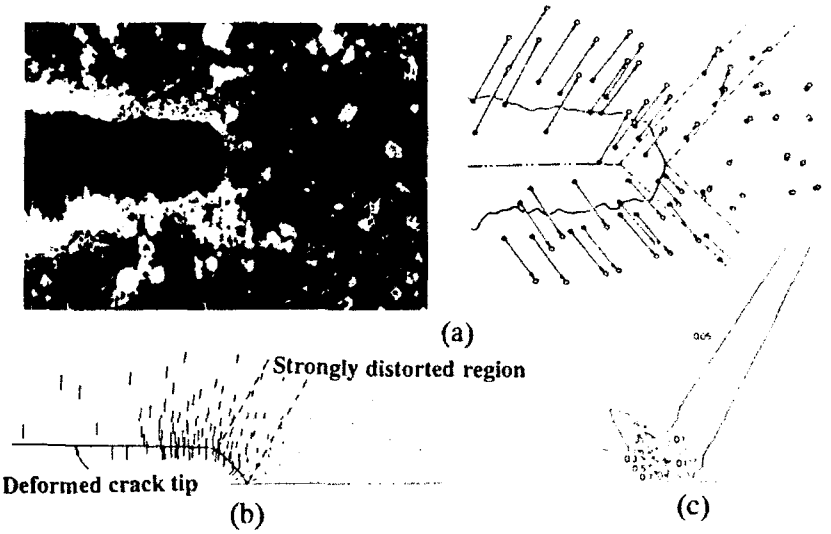


Figure 5.9 Deformation behavior of silicon iron cracked plate. (a) Crack tip deformation and displacement observed experimentally (from [125]), (b) Displacement vector, (c) Representative strain distribution (from [126]).

In order to understand the role of grain boundaries on the localization, Lemonds et al. [127] investigated bicrystals which have the same geometry but have different levels of latent hardening and grain boundaries across which crystallographic misorientation exists. It was found that the boundaries cause a nonuniform deformation from the onset of plastic flow, and the formation of shear bands is strongly affected by the initial misorientation across the boundary and by the difference in latent hardening.

The primitive polycrystalline model has been constructed as a heterogeneous aggregation of the single crystal with different slip systems [123,124]. To account for the inhomogeneous deformation within the single-crystal grain, the grain is divided into several finite elements whose constitutive response is expressed by a single-crystal model such as Asaro's model [24] with the same slip direction. The grain boundaries are the same in bicrystals. The compatibility of the deformation and equilibrium are satisfied. The model provides insight as to how the flow localization depends on the grain size, grain profile and the orientation of the slip system. The evolution of representative plastic strain in Fig.5.10 captures the characteristic feature of the flow localization in tension blocks. That is, microscopically, a heterogeneous structure of the material microscopically produces an inhomogeneous deformation field, and then surface roughness increases steadily with increase of macroscopic strain. Irregularity of deformation develops more rapidly near the free surface and the central part of the body. These deformation fields cause the macroscopic shear bands seen in the

figure. As the number of grains along the thickness decreases, the deformation behavior becomes more dependent on such initial grain characteristics as the shape, size and orientation of the slip plane [128].

Harren and Asaro [129] investigated nonuniform deformation behaviors in polycrystals using similar but more elaborate models than these mentioned above. Two-dimensional polycrystals are the aggregate of hexagonal single crystals with three slip directions. A patchy nonuniform deformation mode associated with the substantial jumps in current lattice orientation, i.e., the subgrain boundary, occurred. It was also clarified, for example, that the grain boundaries act as an impedance to the slip activity, while the triple points act as initiators of nonuniform deformation modes. How shear bands cross the grain boundary was elucidated, and the validity of Taylor's model was also precisely examined. However, again, proper modeling of the grain boundary and high resolution is required for the prediction of initial grain level shear banding [129].

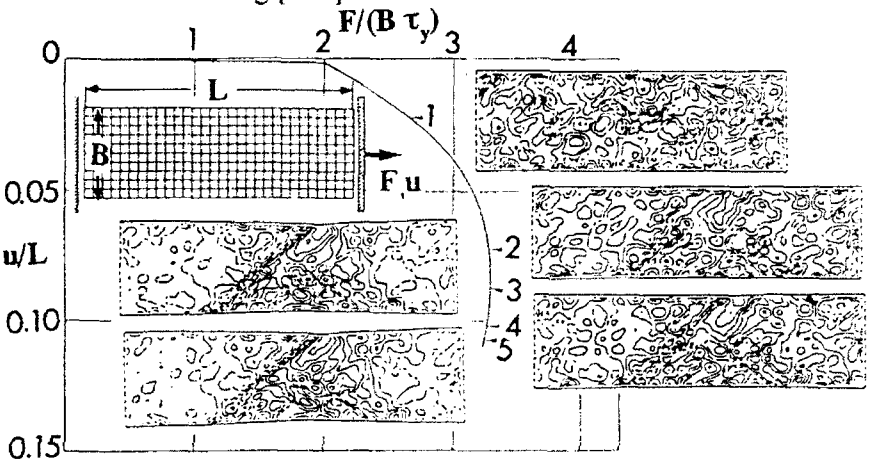


Figure 5.10 Load vs elongation and evolution of representative plastic strain in polycrystalline block under tension (from [123, 124]).

Solution of these problems can be expected with the recent development of mechanical analysis employing molecular dynamic (MD) simulation, which is the most elemental method and has potential for treatment of the dynamics of the crystal defect, such as vacancy, dislocation and twinning. Although the large discrepancy in time and space in the continuum analysis and MD analysis is inevitable, further development is expected, and indeed, quite intensive studies are now under way to bridge the micro- and macroscopic deformation behaviors [130].

6. CONCLUDING REMARKS

An introduction has been given to the computational prediction of plastic instabilities. The computational strategies to predict the onset of bifurcation for general solids and thin-walled bodies, and to mesh independent predictions of flow localization behavior were discussed. Several topics, mainly selected from the contributions during the last decade, including instability behaviors of thick circular tubes under various kinds of loading, instability propagation in thermo-elasto-viscoplastic polymeric materials, wrinkling of thin walled bodies under different type of metal forming processes, dynamic flow localization of thermo-elasto-viscoplastic blocks and flow localization of mono- and polycrystalline solids were reviewed.

Thus, computational studies can play an important role in understanding and modeling the wide range of instability behavior. The problem, however, associated with flow localization still to be solved is the precise prediction of the chain of events observed in deforming

material, that is, the initial stage of deformation which includes the microscopically inhomogeneous deformation leading to instability to the final stage of macroscopic failure. To clarify the continuous transition from micro- to macroscopic deformation behavior, the development of direct physical approaches for each class and bridging of the substantial gap in size and time between the different classes are essential. The suggestive report related to this is given in ref.131.

ACKNOWLEDGMENTS

Partial support for my work reviewed here from the Ministry of Education, Science and Culture of Japan (MESCJ), from Kobe University through KURNS, and from Amada Foundation for Metal Work Technology, is gratefully acknowledged. I also appreciate the support provided by MESCJ for presentation of this paper.

REFERENCES

1 R.Hill, J. Mech. Phys. Solids, **8** (1958) 236.
2 Y.Seguchi, H.Kitagawa, Y.Tomita and A.Shindo, Memo. Fac. Engng. Kobe Univ. **17** (1971) 51.
3 M.J.Sewell, Arch. Ratl. Mech. Anal., **23** (1967) 327.
4 H.Kitagawa and Y.Tomita, Proc. 21st Japan Natl. Congr Appl. Mech., (1971) in: Theoretical and Applied Mechanics, **21** (1973) 243, University of Tokyo Press.
5 J.W.Hutchinson, in: R.F.Hartung (ed.), Numerical Solution of Nonlinear Structural Problems, ASME (1973) 17.
6 V.Tvergaard, Int. J. Mech. Sci., **20** (1978) 651.
7 K.W.Neale, SM Arch., **6** (1980) 79.
8 Y.Tomita, Computational Elasto-Plasticity (1990) Yokendo.
9 J.W.Rudnicki and J.R.Rice, J. Mech. Phys. Solids, **23** (1975) 371.
10 A.L.Gurson, Trans ASME, J. Eng. Mater. Tech., **99** (1977) 2.
11 V.Tvergaard, Int. J. Fract., **17** (1981) 389.
12 M.E.Mear and J.W.Hutchinson, Mech. Mater., **4** (1985) 395.
13 Y.Tomita, A.Shindo and K.Tohne, Trans JSME., **54A** (1988) 578.
14 N.A.Fleck, T.Kuhn and R.M.McMeeking, J. Mech. Phys. Solids, **40** (1992) 1139.
15 B.Budiansky, Trans ASME., J. Appl. Mech., **26** (1959) 259.
16 S.Storen and J.R.Rice, J. Mech. Phys. Solids, **23** (1975) 421.
17 A.Needleman and V.Tvergaard, ditto., **25** (1977) 159.
18 J.W.Hutchinson and K.W.Neale, in: D.P.Koistinen and N-M.Wang (ed.), Mechanics of Sheet Metal Forming, (1978), 127, Plenum Press.
19 Y.Tomita and A.Shindo, in: J.Salencon (ed.), Plastic Instability, (1985) 203, Press Pont et Chaussees.
20 J.Christoffersen and J.W.Hutchinson, J. Mech. Phys. Solids, **27** (1979) 465.
21 Y.Tomita and A.Shindo, Int. J. Mech. Sci., **30** (1988) 921.
22 Y.Tomita, A.Shindo, Y.S.Kim and K.Michiura, ditto., **28** (1986) 263.
23 R.J.Asaro and J.R.Rice, J. Mech. Phys. Solids, **25** (1977) 309.
24 R.J.Asaro, Acta Metall., **27** (1979) 445.
25 Y.Zhou, K.W.Neale and L.S.Toth, in: M.Jono and T.Inoue (eds.), Mechanical Behaviour of Materials, (1991) 135, Pergamon Press.
26 J.Lemonds and A.Needleman, Mech. Materials, **5** (1986) 339.
27 Y.Tomita, S.Asada and A.Shindo., Trans JSME., **55A** (1989) 1865.
28 D.Peirce, C.F.Shih and A.Needleman, Computers & Struct., **18** (1984) 875.
29 S.Nemat-Nasser, D.-T. Chung and L.M.Taylor, Mech. Materials, **7** (1989) 319.
30 A.Zavaliangos and L. Anand, Int. J. Num. Meth. Engng., **33** (1992) 595.
31 H.Kitagawa, J.Tanabe, K.Shykuri and M.Hamada, Trans JSME., **46A** (1980) 486.
32 H.Kitagawa, J.Tanabe, M.Hamada and Y.Tomita, ditto., **48A** (1982) 141.
33 R.de Borst, Ing. Arch., **59** (1989) 160.

34 P.Bardet, Computers & Struct., **36** (1990) 993.
35 Y.M.Leroy and O.Chapuis, Comp. Meth. Appl. Mech. Engng., **90** (1991) 969.
36 B.Raniecki and O.Brunhs, J. Mech. Phys. Solids, **29** (1981) 153.
37 Y.Tomita, A.Shindo and A.Fatnassi, Int. J. Plasticity, **4** (1988) 251.
38 V.Tvergaard, D.C.A.A.M. Report **4**, No.250 (1982)
39 N.Triantafyllidis, J. Mech. Phys. Solids, **33** (1985) 117.
40 H.Petryk and K.Thermann, Int. J. Solids Struct., **29** (1992) 745.
41 P.M.Naghdi, Handbuch der Physik, **VI**, (1972) Springer-Verlag Berlin.
42 Y.Tomita and A.Shindo, in: T.Inoue et al. (eds.), Computational Plasticity. (1990) 165, Elsevier.
43 J.W.Hutchinson, J. Mech. Phys. Solids, **21** (1973) 163.
44 Y.Tomita , A.Shindo and M.Nagai, Int. J. Mech. Sci., **26** (1984) 437.
45 A.Needleman, Comp. Meth. Appl. Mech. Engng., **67** (1988) 69.
46 J.C.Nagtegaarl, D.M.Parks and J.R.Rice, Comp. Meth. Appl. Mech. Engng., **4** (1974) 177.
47 O.C.Zienkiewicz, The Finite Element Method (3rd edn). (1977) McGraw-Hill.
48 W.K.Liu, H.Hang, S.Chen and T.Belyscheko, Comp. Meth. Appl. Mech. Engng., **53** (1985) 13.
49 V.Tvergaard, A.Needlemen and K.K.Lo, J. Mech. Phys. Solids, **29** (1981) 115.
50 M.Ortiz and J.J.Qurgley, Comp. Meth. Appl. Mech. Engng., **90** (1991) 781.
51 M.Pastor, J. Peraire and O.C.Zienkiewicz, Arch. Mech., **61** (1991) 30.
52 B.Loret and J.H.Prevost, Comp. Meth. Appl. Mech. Engng., **83** (1990) 247.
53 J.H.Prevost and B.Loret, ditto., **83** (1990) 275.
54 Z.P.Bazant and F.P.Lin, Int. J. Num. Meth. Engng., **26** (1988) 1805.
55 Y.Tomita, A.Shindo and T.Sasayama, Int. J. Mech. Sci., **32** (1990) 613.
56 Z.P.Bazant and G.P.Cabot, Trans ASME., J. Appl. Mech., **55** (1988) 287.
57 E.C.Aifantis, Int. J. Plasticity, **3** (1987) 211.
58 H.M.Zbib and E.C.Aifantis, Res. Mechanica, **23** (1988) 261.
59 H.M.Zbib and E.C.Aifantis, ditto., **23** (1988) 279.
60 Y.Tomita and T.Nakao, in: M.Jono and T.Inoue (eds.), Mechanical Behaviour of Materiais, (1991) 197, Pergamon Press.
61 Y.Tomita and N.Nakao, in: D.Besdo and E.Stein (eds.), Finite Inelastic Deformations, Theory and Applications, (1992) in press.
62 Y.Tomita and Maruhashi, Proc JSME., **920** (1992) in press.
63 R.de Borst and L.J.Sluys, Comp. Meth. Appl. Mech. Engng., **90** (1991) 805.
64 R.C.Batra, Int. J. Plasticity, **3** (1987) 75.
65 H.Muhlhaus and E.C.Aifantis, MME. Report No.17 (1989).
66 M.Ortiz, Y.Leroy and A.Needleman, Comp. Meth. Appl. Mech. Engng., **61** (1987) 189.
67 A.Nacar, A.Needleman and M.Ortiz, ditto., **73** (1989) 235.
68 J.R.Rice, W.Koiter(ed.), Proc. 14th IUTAM Congress, Delft (1976) 207, North-Holland.
69 Y.Leroy and M.Ortiz, Int. J. Num. Anal. Meth. Geomech., **14** (1990) 93.
70 J.Fish and T.Belytschko, Computers & Struct., **35** (1990) 309.
71 J.Fish and T.Belytschko, Comp. Meth. Appl. Mech. Engng., **78** (1990) 181.
72 P.Steinmann and K.Willam, Arch Mech., **61** (1991) 259.
73 A.Needleman, J. Mech. Phys. Solids, **23** (1975) 357.
74 C.C.Chu, Trans ASME., J. Appl. Mech., **46** (1979) 889.
75 Y.Tomita , A.Shindo and H.Kitagawa, Int. J. Mech. Sci., **23** (1981) 723.
76 Y.Tomita and A.Shindo, Comp. Meth. Appl. Mech. Engng, **35** (1982) 207.
77 H.Kitagawa, R.Yokoyama and M.Hamada, Trans JSME., **50A** (1984) 209
78 M.Larrson, A.Needleman, V.Tvergaard and B.Storakers, J.Mech. Phys. Solids, **30** (1982) 121.
79 V.Tvergaard, ditto., **24** (1976) 291.
80 V.Tvergaard, Eur. J. Mech., A/Solids, **9** (1990) 21.
81 R.Hill and J.W.Hutchinson, J. Mech. Phys. Solids, **23** (1975) 236.
82 J.W.Hutchinson and V.Tvergaard, Int. J. Mech. Sci., **26** (1980) 339.
83 Y.Tomita, A.Shindo and Y.S.Kim, Trans JSME., **51A** (1985) 1872.
84 Y.S.Kim, Y.Tomita and A.Shindo, Computational Mechanics, **4** (1989) 211.
85 Y.Tomita, A.Shindo and T.Umakoshi., Trans JSME., **55A** (1989) 765.

86 J.W.Hutchinson and K.W.Neale, J. Mech. Phys. Solids, **31** (1983) 405.
87 E.Chater and J.W.Hutchinson,Trans ASME., J. Appl. Mech., **51** (1984) 269.
88 K.W.Neale and P. Tugcu, J. Mech. Phys. Solids, **33** (1985) 323.
89 L.O.Fager and J.L.Bassani, Int. J. Solids. Struct., **22** (1986) 1243.
90 Y.Tomita, T.Takahashi and A.Shindo, Int. J. Mech. Sci., **32** (1990) 335.
91 S.Kyriakides (1992) in this volume.
92 P.Tugcu and K.W.Neale, Int. J. Solids Struct., **23** (1987) 1063.
93 P.Tugcu and K.W.Neale, Int. J. Mech. Sci., **32** (1990) 405.
94 Y Tomita and K.Hayashi, in: J.P.Boehler and A.S.Khan (eds.), Anisotropy and Localization of Plastic Defomation, (1991) 524, Elsevier.
95 C.G'Sell and J.J.Jonas, J. Mater. Sci., **14** (1979) 583.
96 M.C.Boyce, D.M.Parks and A.S.Argon, Mech. Mater., **7** (1988) 15.
97 Y.Tomita, A.Shindo and K.Itoh, Proc. 32nd Japan Natl. Congr. Appl.Mech., (1982)in:Theoretical and Applied Mechanics, **32** (1984) 367, Univ. of Tokyo Press.
98 A.Fatnassi, Y.Tomita and A.Shindo, Int. J. Mech. Sci., **27** (1985) 643.
99 K.Naruse and H.Takeyama, Trans JSTP., **20** (1979) 386.
100 N.Triantafyllidis and A.Needleman, Trans ASME., J.Engng. Mater.Tech., **102** (1980) 241.
101 T.Yoshida, K.Itoh, and T.Sagawa, Trans JSTP., **32** (1991) 1286.
102 R.Hill, Mathematical Theory of Plasticity, (1950) Oxford.
103 X.Wang and L.H.N.Lee, in: J.P.Boehler and A.S.Khan (eds.), Anisotropy and Localization of Plastic Defomation, (1991) 676, Elsevier.
104 Y.Tomita and H.Shao, Proc. AEPA' 92, (1922) in press.
105 N.Triantafyllidis, in: J.Salencon (ed.), Plastic Instability, (1985) 115, Press Pont et Chaussees.
106 J.W.Hutchinson and K.W.Neale, ditto., (1985) 71.
107 K.W.Neale and Tugcu, Int. J. Num. Meth. Engng., **30** (1990) 1595.
108 K.Itoh, H.Ohki and H.Takeyama, Trans JSTP., **24** (1983) 920.
109 X.Chateau and Q.S.Nguyen, Eur. J. Mech., A/Solids, **10** (1991) 71.
110 K.Chung and R.Wagoner, Metall. Trans., **A17** (1986) 1632.
111 Y.Tomita, A.Shindo, S.Asada and H.Gotoh, Trans JSME., **54** (1988) 1124.
112 H.M.Zbib, MME Report No.91-01 (1991).
113 H.M.Zbib and J.S.Jubran, ditto., No.91-01 (1991)
114 A.Needleman, Trans ASME., J. Appl. Mech., **56** (1989) 1.
115 R.C.Batra and De-Shin Liu, ditto., **56** (1989) 527.
116 Y.Tomita and K.Mimura, Proc. AEPA' 92, (1992) in press.
117 J.D.Campbell,A.M.Eleiche and M.C.C.Tsao, Fundamental Aspect of Structural Alloy Design, (1977) 545, Plenum Publ.
118 K.Mimura and Y.Tomita, Journal de Physique IV, DYMAT 91, Les editions de Physique, France (1991) 813.
119 R.C.Batra and C.H.Kim, Int. J. Plastaicity, **6** (1990) 127.
120 M.Chu, Y.Tomita and A.Shindo, Trans JSME., **55A** (1989) 1872.
121 D.Peirce , R.J.Asaro and A.Needleman, Acta. Metall., **30** (1982) 1087.
122 M.A.Zikry and S.Nemat-Nasser, Mech. Mater., **10** (1990) 215.
123 H.Kitagawa and H.Matsushita, Trans JSME., **51A** (1984) 2120.
124 H.Kitagawa and H.Matsushita, Int. J. Solids Struct., **23** (1987) 351.
125 M.Kikukawa, M.Jono, M.Iwahashi, M.Ichikawa and N.Uesugi, J. Soc. Mater. Sci. Japan, **31** (1982) 669.
126 H.Kitagawa, in: T.Inoue et al. (eds.), Computational Plasticity, (1990) 165, Elsevier.
127 J.Lemonds, R.J.Asaro and A.Needleman, Mech. Materials, **4** (1985) 417.
128 Y.Tomita, K.Mimura and T.Sasayama, Trans JSME., **57A** (1991) 3036.
129 S.V.Harren and R.J.Asaro, J. Mech. Phys. Solids, **37** (1989) 3036.
130 H.Kitagawa and Nakatani, in: M.Jono and T.Inoue (eds.), Mechanical Behaviour of Materials, (1991) 111, Pergamon Press.
131 L.Anand, O.Dillon, T.A.Place and B.F.von Turkovich, Int. J. Plasticity, **6** (1990) I.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

The Role of Wave Breaking in Air-Sea Interaction

W. K. Melville[a]

[a]Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA 92093-0213.

Abstract

Some of the roles of breaking in air-sea interaction are reviewed. At present we have no rational way of theoretically predicting the evolution of an air-water interface beyond breaking, so progress depends to a great extent on experimental measurements in both the laboratory and the field. Even the task of identifying breaking in a random wave field is not simple and much effort has been expended in determining rational breaking criteria and their measurement. As a consequence of breaking, much of the local momentum flux transmitted from the atmosphere to the ocean via the wave field is rapidly transferred to currents. Breaking is the primary dissipative mechanism for the surface wave field, making energy available for mixing the surface layers. Local equilibrium wave models, in which wind input, nonlinear transfers, and dissipation due to breaking are in balance, lead to predictions of the wave spectrum and to estimates of the dissipation. It is shown that recent field measurements of enhanced dissipation in the surface layers are consistent with equilibrium modelling and laboratory measurements of dissipation and mixing. Enhanced levels of surface turbulence and bubble mediated gas transfer can lead to significant increases in gas transfer across the surface with the onset of breaking. Breaking is one of the major sources of ambient sound in the ocean. Recent results on the use of ambient sound to identify and quantify breaking are presented.

1. Introduction

The study of surface waves has a long tradition in theoretical and applied mechanics, but even before the formal investigation of the subject it must have been of interest for as long as man has been involved in maritime exploration and trade. The dangers presented to early explorers by wind and waves are readily apparent when it is observed that many were putting to sea in ships which were barely longer than many of today's cruising yachts. Even today, fishing vessels capsize and are lost in steep and breaking seas. While steep and breaking seas provide the extreme environments that are encountered, it is still surprisingly the case that most descriptions of the ocean surface, whether for engineering or oceanographic purposes, are based on linear theory. The foundations of linear wave theory were laid by many of the founders of theoretical hydrodynamics in the nineteenth century. In 1847 Stokes published his investigation of weak nonlinear effects on surface waves, but it took more than a century for the subject to extend to an investigation of nonlinear interactions between surface waves (Phillips, 1960) and subsequently the stability of weakly nonlinear waves (Benjamin & Feir, 1967).

Convenient measures of the nonlinearity are the wave steepness ak, where a is the wave amplitude and k is the wave number, or equivalently, u/c where u is a measure of the horizontal fluid velocity at the surface and c is the phase speed. In breaking waves u/c may exceed unity while ak is $O(1)$, and so breaking waves are strongly nonlinear. Rigorous theoretical investigations of breaking have been confined to describing local features of the flow (Longuet-Higgins, 1988). Numerical codes are available to describe the evolution of two-dimensional surface waves up to breaking, but they are unable to continue beyond the point at which the surface impacts on itself (Dommermuth et al., 1987). Thus the study of wave breaking is based in large part on experiments in both the laboratory and the field.

The most common expression of breaking at the sea surface is a whitecap due to the entrained air.[1] Whitecaps become obvious at scales of $O(1m)$ extending up to $O(100m)$, but breaking may occur without significant air entrainment down to scales of centimeters. This lack of a universal visual feature identifying breaking and the difficulty of measuring variables in a two-phase free-surface turbulent flow have made progress difficult, especially in the field. However, this difficulty is justified by the importance of breaking to the processes occurring at the air-sea interface. Breaking plays a number of related roles in air-sea interaction. These include:

a. Limiting the height (slope) of surface waves.

b. Being a source of vorticity and turbulence.

c. Dissipating surface wave energy, some of which is available for turbulent mixing.

d. Generating ocean currents by transferring momentum from the wave field.

e. Enhancing gas transfer due to surface turbulence and bubble entrainment.

f. Generating sound at the ocean surface which may be used as a diagnostic tool for air-sea interaction studies (Acoustical Oceanography).

g. Providing passive and active microwave signatures which may prove useful in remote sensing of the ocean.

It may seem surprising that a process which is intermittent in space and time, and which, at least at the larger scales, is only occurring over a small fraction of the ocean surface at any time, may be significant. We shall see below that the direct effects of breaking are typically felt to a depth of the order of the wave height. In relatively benign conditions this may be to depths of the order of metres, but in stormy seas this may be depths of $O(10m)$. The sensitivity of the atmosphere-ocean system to the conditions in the first ten metres of the ocean are graphically presented by Gill(1982):

a. Of the solar radiation incident on the atmosphere 40% is absorbed in the first $10m$ of the ocean; more in coastal waters.

[1] For the proceedings of a recent conference on the subject of whitecaps, see Monahan & McNiocaill, 1986.

b. The first $10m$ of the water column has the same weight as all the atmosphere above.(This has important implications for gas transfer by bubbles.)

c. The first $2.5m$ of the water column has the same heat capacity as all the atmosphere above.

Now breaking is observed to begin at a wind speed of approximately $3m/s$. This, together with the fact that the momentum and mechanical energy fluxes across the air-sea interface typically scale as the square and cube of the wind speed, respectively, implies that even relatively short periods of high winds and waves may contribute significant fractions of the total fluxes across the air-sea interface, with breaking absent only under the most benign conditions.

In this paper we shall review some of the recent contributions to our knowledge of breaking and its role in air-sea interaction. This is not meant to be an exhaustive review of the subject, a task which would expand well beyond the format of these proceedings. However, it is hoped that the reader will achieve an appreciation for the wide range of techniques, both theoretical and experimental, that are being used to investigate this important phenomenon.

2. Breaking Criteria and Detection

Historically, the first "breaking" flow to be identified was Stokes' 120^0 corner flow at the crest of a periodic wave of limiting slope ($ak = 0.443$). In a frame of reference travelling at the phase speed the fluid at the crest is at rest, and so in the laboratory frame it is moving at the phase speed of the wave. This then satisfies the kinematic criterion that the fluid velocity matches or exceeds the phase velocity. However, it is most unlikely that such a flow can ever be realized, since uniform wave fields of much smaller slope are subject to intrinsic instabilities which may lead to breaking. Melville (1982) has shown experimentally that for wave slopes below approximately 0.3 two-dimensional waves are unstable to two-dimensional Benjamin-Feir instabilities which lead to breaking, while at larger slopes three-dimensional instabilities dominate (See also Su et al., 1982; McLean et al., 1981). Numerical solutions have shown that uniform wave trains with slopes as small as 0.1 may evolve to breaking following instability (Dold & Peregrine, 1986).

Stokes' limiting form gives an expression for the wave height,H, in terms of the wave period,T, and the gravitational acceleration, g:

$$H = 0.027gT^2. \tag{1}$$

Laboratory experiments by a number of authors (Figure 1) have shown that breaking occurs for average values of the numerical factor in the range 0.016-0.022, or at slopes based on the wave height and length of approximately two-thirds that of the limiting slope (Coincidentally, this is close to the slope of 0.3 at which fast three-dimensional instabilities begin.) However, as Figure 1 shows, the scatter in the data is considerable.

Field observations of breaking have been based on a so-called "jump meter" which uses the time derivative of the surface displacement measured by a wave gauge along with the characteristic phase speed of the wave to determine spatial gradients in the surface displacement (Longuet-Higgins & Smith, 1983; Thorpe & Humphries, 1980). Gradients

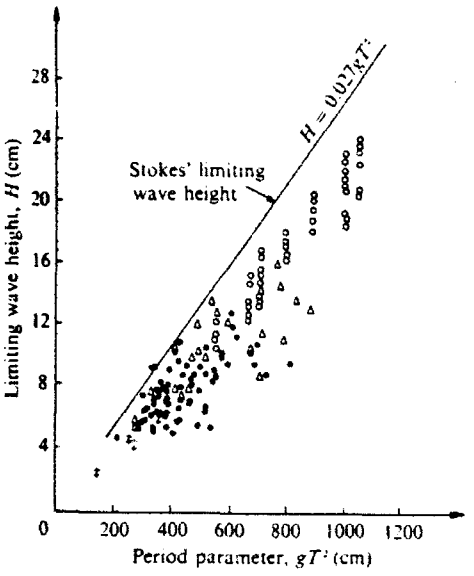


Figure 1. Laboratory measurements of wave heights and periods at breaking compared with Stokes limiting form (Bonmarin, 1989).

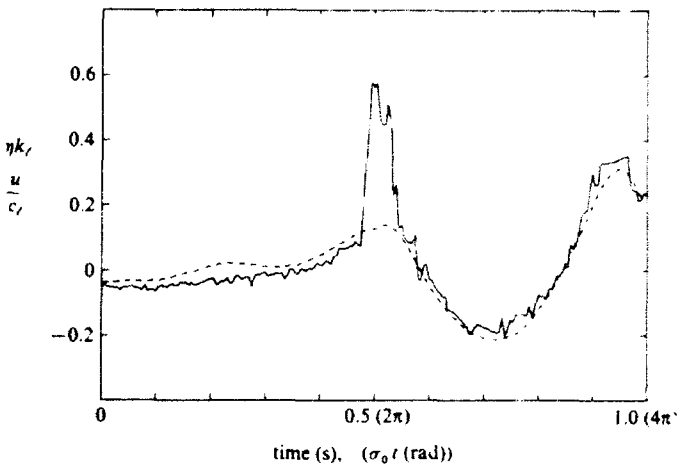


Figure 2. Laboratory measurements of normalised surface displacement and horizontal fluid velocity in breaking and unbroken waves (Melville & Rapp, 1988).

which exceed an arbitrary threshold are considered to be breaking. Weissman et al.(1984) have used high-passed wave gauge records to detect breaking in lake experiments. Since large temporal gradients are resolved by high frequency Fourier components, the jump-meter and high-frequency approaches are related. Holthuijsen & Herbers(1986) visually observed whitecaps at the location of a wave measuring buoy and analysed their data in terms of the joint height-period probability distributions of both the broken and unbroken waves passing the buoy. These distributions showed considerable overlap and no clear threshold for breaking based on the combined wave height and period. This contrasts with the laboratory measurements described above. However, the difficulty with point measurements, whether in the laboratory or field, is that in a random breaking wave field they may catch the wave at any stage in the breaking process, and since the wave parameters may change during that time they are not necessarily the same as those at the instant of the onset of breaking.

This point was made very clear by a laboratory study by Melville & Rapp(1988) who used a laser anemometer to measure the horizontal velocity of the fluid at the surface to detect breaking. Figure 3 shows an example of a small section of coincident surface displacement and velocity time series. Of the two waves in the frame the smaller less steep wave is breaking while the larger steeper wave is not. The wave began breaking near the peak of the wave envelope but decreased in amplitude and slope as it continued to break. This example displays the danger of using observations of breaking at any stage in its evolution to infer breaking criteria. It also points to the difficulties inherent in using the

surface displacement or its first derivative in time as an indicator of breaking.

In recent years other non-hydrodynamic methods of detecting breaking have been investigated and exploited. These include acoustic methods (about which we will have more to say below), and microwave methods. The use of active microwave methods to detect and quantify breaking has been investigated by a number of authors. Keller et al. (1986) presented an illustrative summary of the use of coherent microwave systems in measuring breaking waves. They pointed out that both specular, Bragg and volume scattering may contribute to the scattered components. Banner & Fooks(1985) associated enhanced scattering by a quasi-steady spilling wave with the Bragg components in the smaller-scale waves and surface disturbances that accompany the turbulence. The author and his co-workers (Melville et al., 1988; Jessup et al, 1990, 1991a,b; Loewen & Melville, 1991) have attempted to quantify the scattering by breaking waves and correlate it with the dynamics of the breaking waves. In laboratory studies we have found that the scattering correlates with the dissipation due to breaking, although the detailed reason for this result remains to be investigated. In the field we have found that the lower order moments of the microwave Doppler spectrum taken together provide a good indicator of breaking and that the contribution of the breaking waves to the radar cross section at moderate incidence angles is significant and increases approximately as the cube of the friction velocity of the wind. Notwithstanding these successes, the fact remains that steep but unbroken waves may have similar microwave signatures to breaking waves, and resolving the differences may prove difficult.

In summary, the only reliable methods for identifying breaking at all scales include measurements of both the surface geometry and the surface velocity field. Such methods are confined to the laboratory at present. The problem of detection and quantification of the parameters describing breaking waves remains an impediment to field studies of breaking.

3. Momentum Flux across the Air-Sea Interface

The wind generates waves! The wind generates currents! Even to the lay observer these statements are obviously true. Now waves can transport momentum with a mean momentum flux that is proportional to the square of the wave height. What is not so obvious is the partitioning of the fluxes across the air-sea interface between waves and currents. If the waves were to continuously receive momentum from the wind they would continue to grow and the momentum flux they carry would continue to increase. The growth of the waves is arrested by breaking and in breaking the wave field gives up part of its momentum flux to currents. Thus breaking is an important process in the generation of ocean currents.

These issues were addressed by Mitsuyasu (1985). The essentials may be discussed by considering a locally uniform two-dimensional wave field which has spatial gradients on a scale much larger than the wave length. The wave momentum flux, S_{11}, is given by

$$S_{11} = \frac{1}{2}E, \tag{2}$$

where $E = \frac{1}{2}\rho_w g a^2$ is the wave energy density. The momentum flux from the air to the water τ_0 is given by $\tau_0 = \tau_w + \tau_t$, where τ_w is the flux directly into the waves, and τ_t is

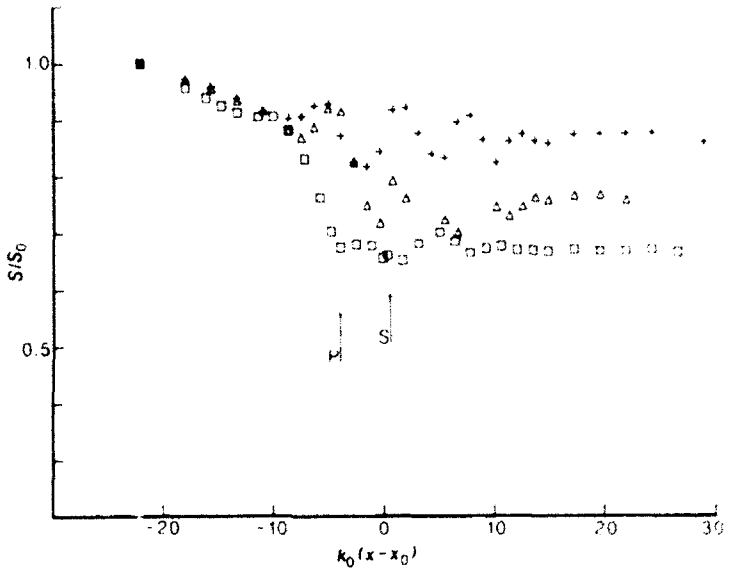


Figure 3. Laboratory measurements of normalized variance of the free surface displacement of a wave packet as a function of downstream position. Note the decrease across the breaking region. Symbols refer to values of a slope parameter $a_0 k_0$: $+, 0.26; \triangle, 0.30; \square, 0.39$. (Melville & Rapp, 1986).

the remaining turbulent flux which is not coherent with the wave field. Now τ_w can be related to the gradient in the momentum flux,

$$\tau_w = \frac{dS_{11}}{dx} = \frac{\beta}{c} E \tag{3}$$

where $\beta = \frac{c_g}{E}\frac{dE}{dx}$ is the energy density growth rate, and c and c_g are the phase and group velocities, respectively. From laboratory experiments in the absence of breaking Mitsuyasu found that $\tau_w = 22(ak)^2\tau_0$, giving significant values of τ_w for representative values of the wave slope ak. This estimate of the momentum flux directly into the wave field was supported by laboratory and field measurements giving ratios of $\frac{\tau_w}{\tau_0}$ in the range $0.4 - 0.6$ (Hsu et al., 1981, 1982; Snyder et al., 1981). However, field measurements of the secular wave growth showed that $\frac{dS_{11}}{dx} = 0.05\tau_0$; that is, only 5% of the momentum flux from the atmosphere is carried by waves propagating out of the generation region. This is only 10% of the measured momentum flux from the wind to the waves. Mitsuyasu conjectured that the remainder of τ_w is lost from the wave field by breaking in the generation region.

This argument was supported by laboratory experiments on unsteady breaking by Melville & Rapp(1986). Using the dispersive properties of deep-water waves to focus a wave packet in a laboratory channel, and by measuring the wave field upstream and downstream of the break, we were able to measure the excess momentum flux lost from the wave field (Fig. 3). Using field measurements of the incidence of breaking by Thorpe & Humphries (1980), we were able to show that Mitsuyasu's conjecture was supported by the combined laboratory and field data. The laboratory studies were considerably extended by Rapp & Melville(1990).

We can conclude from this work that while a significant fraction of the momentum flux from the atmosphere (50% or so) may pass through the wave field; most of it is lost to currents by breaking in the generation region.

4. Wind-Wave Modelling

4.1. The radiative transfer equation

While a rigorous theoretical foundation for prediction of wind-wave generation evolution and decay does not yet exist, a sufficient number of the important pieces of the puzzle can be assembled to formulate the problem in terms of the radiative transfer equation (Phillips, 1977)

$$\frac{dN}{dt} = \frac{\partial N}{\partial t} + (\vec{c}_g + \vec{U}).\nabla N = -\nabla_k.\vec{T}(\vec{k}) + S_w - D, \tag{4}$$

where $N(\vec{k}) = \frac{g}{\sigma}F(\vec{k}) = (g/k)^{1/2}F(\vec{k})$ is the wave action spectral density; $\vec{T}(\vec{k})$ is the action spectral flux due to wave-wave interactions; S_w is the wind input, and D represents the dissipation which is thought to be due primarily to breaking. $I \equiv -\nabla.\vec{T}(\vec{k})$ is the "collision integral" representing the nonlinear wave-wave interactions as formulated by Hasselmann (1962). In operational wave models I is computed numerically. The wind input term is based on Miles theory (1957, 1962), and empirical input (Snyder et al., 1981; Plant, 1982) leading to an expression of the form

$$S_w = m\cos^{2p}\theta\sigma(u_{*a}/c)^2 N(\vec{k}) \tag{5}$$

(Phillips, 1985), where m and p are empirical constants and u_{*a} is the friction velocity in the air. The dissipation term is the least well understood. Hasselmann (1974) has made heuristic arguments subsequently modified by Komen et al.(1984) leading to an expression of the form

$$D \propto \bar{\sigma}(\sigma/\bar{\sigma})^2(\bar{\alpha}/\alpha_{\bar{PM}})^2 N(\vec{k}) \tag{6}$$

where $\bar{\alpha}$ is a slope parameter and the subscript 'PM' refers to a value representative of the Pierson-Moskowitz spectrum . Figure 4 shows an example of the spectrum and the contributions from the various terms as computed by Komen et al (1984). According to this model the wind input and dissipation predominate at frequencies greater than the spectral peak, while the effect of the nonlinear interactions is to add energy at the lower frequencies while subtracting it from the higher frequencies. Clearly, if this class of models is representative of the physics, then the dissipation due to breaking is comparable to the wind input and must have a profound effect on the details of the spectral evolution.

4.2. Equilibrium Models and Spectral Slopes

With the example of the success of Kolmogorov's universal equilibrium and inertial subrange in describing turbulence spectra , it is not surprising that similar approaches have been made in an attempt to explain the essential characteristics of the surface wave energy spectrum. The motivation for this work has expanded in recent years with the need to understand active microwave remote sensing systems which are directly affected by the small scale waves having wavelengths of O(1m) or less.

Phillips (1958) speculated that at high frequencies and wavenumbers saturation due to breaking would occur, and that the form of the spectrum could only depend on the frequency σ and gravity g. This led to a frequency spectrum of the form $F(\sigma) = \alpha g^2 \sigma^{-5}$, with α becoming known as 'Phillips constant'. The deep-water dispersion relationship, $\sigma^2 = gk$, then gives a wavenumber spectrum $F(k) \propto k^{-4}$. During the 1970's, as more measurements of wave spectra became available, it became apparent that α was not constant. On the basis of laboratory measurements and dimensional analysis Toba (1973) proposed that F should also depend on the ratio u_{*a}/c (where u_{*a} is the friction velocity in the air) giving

$$F(\sigma) \propto (u_{*a}/c) g^2 \sigma^{-5} = u_{*a} g \sigma^{-4}. \tag{7}$$

This result was supported by subsequent laboratory and field data (Kawai et al., 1977; Forristall, 1981; Donelan et al., 1985) and the parameter c/u_{*a} has become known as the "wave age". Kitaigorodskii (1983) followed Kolmogorov's inertial subrange idea more explicitly considering an equilibrium region in which $I = S_w = D = 0$. As a consequence of this hypothesis he found that

$$F(k) \approx \epsilon_0^{1/3} g^{-1/2} k^{-7/2}, \tag{8}$$

$$F(\sigma) \approx \epsilon_0^{1/3} g \sigma^{-4}; \tag{9}$$

this frequency spectrum agreeing with the more recent observations. Kitaigorodskii's model was quickly followed by an equilibrium model from Phillips (1985). The basis of Phillips' modelling was that the right hand side of Equation 4 was zero through an equilibrium in which all terms were significant:

$$I + S_w + D = 0 \tag{10}$$

$$I \propto S_w \propto D. \tag{11}$$

This work provides the most readable and rational account of this class of wave modelling. Phillips found that the energy spectrum was given by

$$F(\vec{k}) = \beta (\cos\theta)^p u_* g^{-1/2} k^{-7/2} \tag{12}$$

and the spectral rate of energy loss was given by

$$\epsilon(\vec{k}) = \gamma \beta^3 (cos\theta)^{3p} u_* k^{-2}. \tag{13}$$

It may at first seem surprising that the very different hypotheses of Kitaigorodskii and Phillips should give the same spectral slope of $-7/2$; however, as pointed out by Phillips, the form of I, which is the same in both cases, the assumption of proportionality between the three terms, and dimensional constraints, determine the slope. (This work also provides a general more reminder that the prediction of a spectral slope may be a necessary but not sufficient test of a theory.)

Most recently, Banner (1990) has synthesized a spectral model based on observations. Writing the wavenumber directional spectrum in the form

$$F(k, \theta) = F(k, \theta_{max}) D(\theta; k) \tag{14}$$

where $F(k, \theta_{max}) \propto k^{-4}$ and$D(\theta; k)$ are based on observations; he calculates a frequency spectrum proportional to $u_{*a}\sigma^{-4}$.

We summarize by noting that our inability to directly measure the spectral distribution of dissipation due to wavebreaking is the greatest impediment to the further development of the current generation of wind-wave models. In the absence of these measurements, the local equilibrium hypothesis provides a rational way of estimating the dissipation based on our better knowledge of wind input and weak nonlinear transfers. It is also consistent with a multiple-scale approach in which on short length and time scales the spectrum is in equilibrium, while on longer scales secular changes can occur.

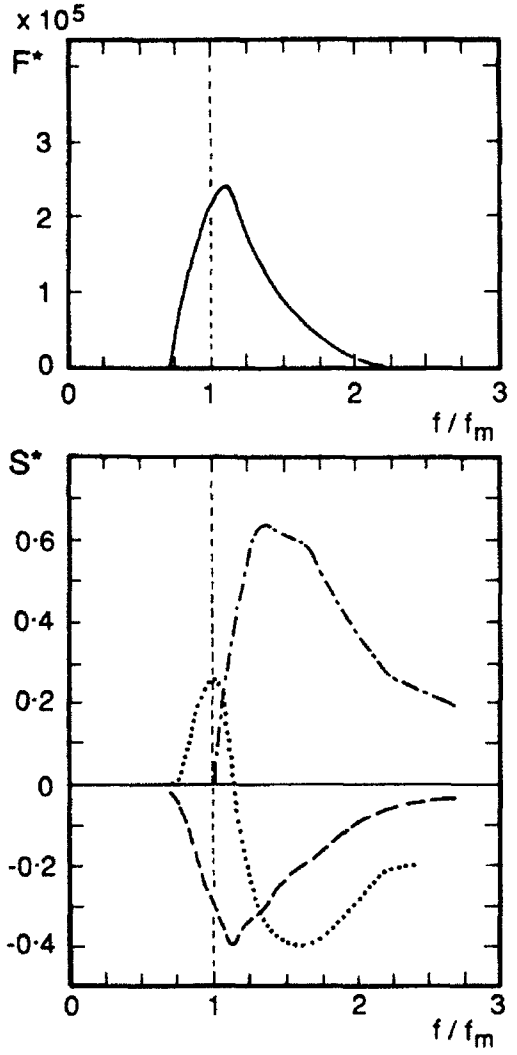


Figure 4. Model wave spectrum and contributions of the wind input, $-\cdot-$, nonlinear transfer, $\cdots$, and dissipation,$- - -$, in the wind direction (Komen et al., 1984).

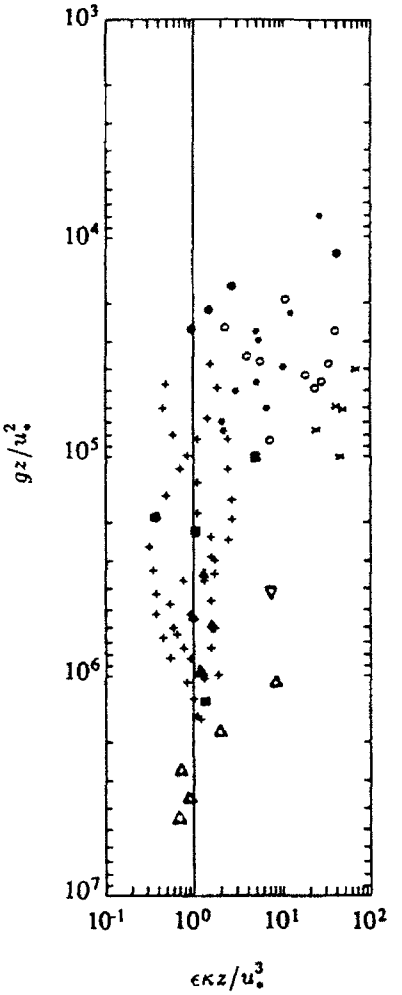


Figure 5. Field measurements of enhanced dissipation in the surface layer (Agrawal et al., 1992).

5. Dissipation due to breaking

In the absence of a direct method for testing the available models of spectral dissipation, we can at least determine whether they are consistent with the available observations. We consider Phillips'(1985) equilibrium model as the basis for such a comparison. Phillips' integrated spectral dissipation rate (over the equilibrium range) , $\epsilon(\vec{k})$, gives a total rate of energy dissipation, ϵ_0, where

$$\epsilon_0 \approx (2\gamma\beta^3 I(3p)\frac{\rho_w}{\rho_a})\rho_a u_{*a}^3 \ln(rC_D^{-1}). \tag{15}$$

There are a number of empirically based parameters included in this expression which Phillips estimated to be given by the following approximate values: $\alpha \approx 6 - 11.10^{-2}$; $\gamma\beta^3 I(3p) > 4.10^{-4}$, and $r \approx 0.4 - 0.7$. The drag coefficient, C_D is taken to be equal to 1.2×10^{-3}. With these estimates we find that

$$\epsilon_0 > 5.\rho_a u_{*a}^3. \tag{16}$$

Now, laboratory experiments on unsteady breaking by Rapp & Melville(1990) have shown that more than 90% of the energy lost from the wave field is dissipated in the water column within four wave periods of the onset of breaking; the active breaking itself lasting for a time comparable to the wave period. This very rapid rate of dissipation was difficult to reconcile with preconceptions until our recent work (Lamarre & Melville, 1991) showed that up to 50% of the energy lost from the wave field is initially expended in overcoming the buoyancy forces acting on the entrained air. (See below). As the air returns to the surface it acts as a buoyant source of kinetic energy, but also enhances dissipation due to the shear in the bubble boundary layers. The experiments of Rapp & Melville (1990) also showed that the turbulent region generated by the breaking wave advected and mixed down to a dimensionless depth $kD \approx 0.5 - 1$, where k is the characteristic wavenumber of the breaking waves. This suggests that the bulk of the dissipation in the water column due to breaking will take place in a layer which is much less than the wave length , and if breaking occurs for $ak \approx 0.3$ (c.f. Figure 1), then the thickness of this highly dissipative layer would be in the range $1 - 1\frac{1}{2}$ wave heights.

We can be a little more specific with these ideas by assuming that the total energy is dissipated over a depth D at a volumetric rate ϵ_w,

$$\epsilon_0 = \int_0^D \rho_w \epsilon_w dz \tag{17}$$

$$= \rho_w \bar{\epsilon_w} D \tag{18}$$

which defines the mean dissipation $\bar{\epsilon_w}$ over the depth D.

In a logarithmic boundary layer the dissipation rate is typically scaled with $u_*^3/\kappa z$ where κ is von Karman's constant (0.4). This scaling gives a dimensionless dissipation rate of unity. It is of interest to determine whether the estimated dissipation rate in the water column under breaking waves based on the modelling of Phillips (1985) and the measurements of Rapp & Melville (1990) gives values significantly different from those in the classical wall layer. If we use the mean values of ϵ_w and D to scale z, we have that

$$\frac{\epsilon_w \kappa z}{u^3{}_w} = O(\frac{\epsilon_0 \kappa}{\rho_w u^3{}_w}) > \frac{5\kappa\rho_a u_{*a}^3}{\rho_w u^3{}_w} \tag{19}$$

It remains to determine the friction velocity in the air, u_{*a}, in terms of the friction velocity in the water, u_{*w}. We can investigate this relationship by considering the disposition of the momentum flux lost from the wave field by breaking. This momentum flux is transferred to currents. Given the intense mixing due to breaking we anticipate that there will a distinct surface current within a depth of the order of one wave height from the surface. Since we are assuming that the wave field is in approximate local equilibrium, to be consistent we must also assume that the horizontal gradients in the current are not large, and in consequence are changing on the same slow spatial scales as the wave variables. This implies that if such a surface current exists then the momentum flux through its upper boundary (the surface) is almost matched by that passing through its lower surface at a depth comparable to the wave height. Thus we expect that $\rho_a u_{*a}^2 \approx \rho_w u_{*w}^2$, whence, $u_{*a}/u_{*w} \approx (\frac{\rho_w}{\rho_a})^{1/2}$, and $\epsilon_w \kappa z u_{*w}^{-3} > 60 = O(10 - 10^2), say.$

These estimates were motivated by recent measurements of Agrawal et al.(1992). Working from a platform in Lake Ontario they used optical, acoustical and electro-mechanical instruments to measure the dissipation near the surface. Their principal results are presented in Figure 5, which shows the measured dissipation scaled in wall variables, as a function of depth scaled by u_{*w}^2/g. The figure may be summarized by the observation that for $\tilde{z} = gz/u_{*w}^2 > 10^5$ the law of the wall scaling appears to hold, whereas for $\tilde{z} < 10^5$ the dissipation rate is up to two orders of magnitude greater. How does a value of $\tilde{z} = 10^5$ compare with the wave height? Agrawal et al.(1992) did not give complete information on the supporting data for their measurements; however, they did indicate that the fetch limited waves were of the order of 30cm significant wave height (H_w), and the winds were greater than $8ms^{-1}$. Taking these values and assuming a value of the drag coefficient in the air of $C_D = 1.2 \times 10^{-3}$, the scaled significant wave height, $\hat{H}_w = 0.4 \times 10^{-5}$. Thus the characteristics of the measured enhanced dissipation layer are in agreement with the estimates based on Phillips (1985) modelling and our laboratory measurements of unsteady breaking (Rapp & Melville, 1990). While Agrawal et al.(1992) speculated that the source of the high dissipation events was breaking waves, they stated that independent (unpublished) estimates of energy fluxes to steep waves at frequencies greater than the peak of the spectrum were consistent with the enhanced kinetic energy dissipation they had measured. As shown above, enhanced dissipation also requires that the energy be dissipated over a sufficiently small depth. The estimates presented here, which are based on both measurements and modelling, provide support for the conclusion that the source of the enhanced dissipation is breaking.

6. Gas Transfer

Public debate has heightened awareness of the importance of the oceans in modulating secular changes in climate. Perhaps nothing has attracted as much attention as the influence of the increase in greenhouse gases on global warming. Quantitative estimates of the role of the oceans in taking up increases in CO_2 are based in part on global budgets which involve considerable uncertainty. As Baggeroer & Munk (1992) have recently observed in their status report on acoustic monitoring of ocean warming, "...all we know for sure is that the oceans are an important sink of heat, and CO_2, and of ignorance."

For moderately soluble gases such as CO_2 and O_2 the transfer across the air-sea in-

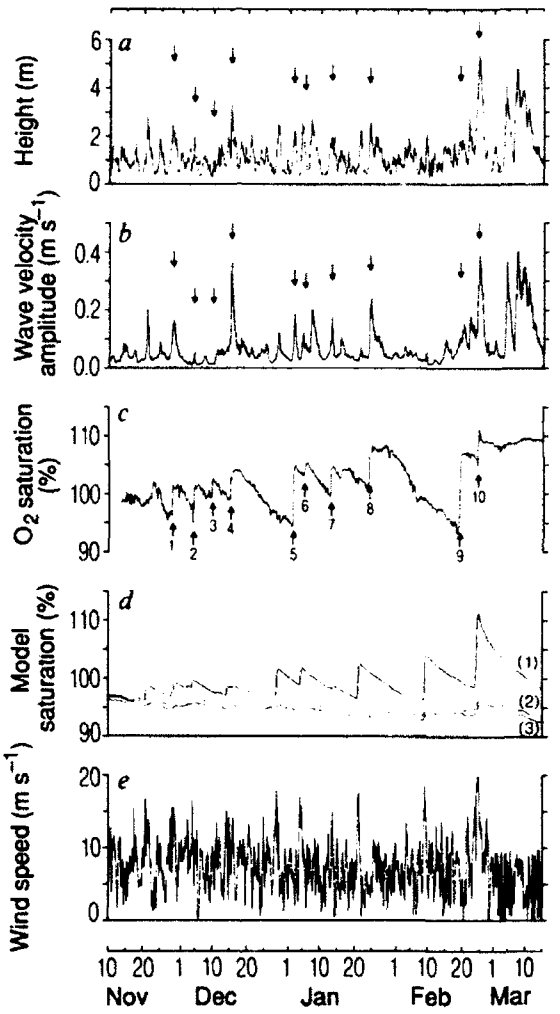


Figure 6. Field measurements of dissolved oxygen and other environmental variables, including wave height, wave orbital velocity and wind speed. Note the rapid increase in the dissolved gas with storm events indicated in the wave orbital velocity time series, and the slower relaxation back to lower levels (Wallace & Wirick, 1991).

terface is believed to be controlled by the aqueous boundary layer (Jaehne, 1990) and is parameterised in terms of a gas transfer velocity k where the molar flux of gas across the interface, $j = k\Delta c$, and Δc is the difference between the saturation and bulk concentration of the dissiolved gas. For low wind speeds, wind-wave tunnel studies have shown that

$$k \approx 0.15 Sc^{-0.5} u_{*w} \tag{20}$$

(Jaehne, 1990; Khoo & Sonin, 1992), where $Sc = \nu/\kappa$ is the Schmidt number, the ratio of the kinematic viscosity of the fluid to the diffusion coefficient of the dissolved gas. At higher wind speeds, above $8-10ms^{-1}$, there is a break point in the linear rise, with higher wind speeds giving a correspondingly higher transfer velocity. This increase in the transfer velocity has been attributed to the onset of breaking and bubble entrainment, but recent experiments by Khoo & Sonin (1992) in which the turbulence is induced by agitation from below with no breaking nor air entrainment show a similar break and enhancement of the gas transfer. Notwithstanding this uncertainty regarding the mechanism leading to enhanced gas transfer at higher wind speeds, Jaehne (1990) has demonstrated that the addition of a mechanically generated breaking wave to a wind-generated wave field in the laboratory can enhance the gas transfer by a factor of two.

The opportunites for field studies of gas transfer have improved recently with the availability of relatively rapid- response dissolved gas sensors which can be deployed on long term moorings. In a recent field experiment in the Middle Atlantic Bight, Wallace & Wirick (1992) deployed a pair of dissolved oxygen sensors at two depths (19m and 34m) for a period of four months. They found that the time series of dissolved oxygen were distinguished by sudden large increases associated with surface wave activity, followed by longer periods of degassing between storms (Figure 6). They compared their measurements with semi-empirical models due to Thorpe (1984) and Spitzer & Jenkins(1989) (each including bubble transport) and a thin-film model (excluding bubble transport). [2] They concluded that Thorpe's model gave the best agreement with the measurements, but by being forced by wind rather than waves it led to some discrepancies. One of the primary causes of the enhanced gas transfer due to bubbles is the increase in diffusion of gas across the bubble surface due to the increase in hydrostatic pressure with depth. As noted in the introduction, the hydrostatic pressure at 10m depth is twice that at the surface. Significant bubble populations are commonly observed at 10m depth, especially in storms. Impediments to improved modelling of bubble transport include uncertainties about initial distibutions of bubble sizes and subsequent breakup, turbulent transport and surface chemistry and contamination. A more detailed review of this important and fascinating subject is beyond the scope of this review, but the reader is referred to Thorpe (1992) for a current assessment.

It is believed that enhancement of gas transfer by breaking is due to both the local increase in turbulent intensity and the entrainment and dissolution of bubbles. Kitaigorodskii (1984) has modelled the influence of patches of enhanced turbulence due to breaking. He finds that the transfer velocity k is proportional to $Sc^{-1/2}[\nu\epsilon_\nu(0)]^{1/4}$, where $\epsilon_\nu(0)$ is the dissipation of turbulent energy near the surface. [3] While noting the similarity of the

[2]Note that these three model results are denoted by 1,2, 3, respectively, in Figure 6.

[3]Note that Kitaigorodskii used the Prandtl number to denote the ratio of the diffusion of momentum to

scaling of the velocity to Kolmogorov scaling, Kitaigorodskii asserted that the arguments leading to this result are independent. It is of interest to determine the effects of the enhanced dissipation discussed in the previous section in the light of Kitaigorodskii's result. From our estimates of Section 5 we would expect an enhancement of the dissipation by a factor of at least 60. According to Kitaigorodskii this would lead to an increase in the transfer velocity by at least a factor of three.

7. The Acoustics of Breaking Waves

Until recently, ambient sound in the audio range ($20Hz - 20kHz$) in the ocean was broadly classified as being due to ocean turbulence ($< 100Hz$), shipping traffic ($< 1kHz$), and "wind" noise ($100Hz - 20kHz$), and was represented by smooth empirical spectra (Wenz, 1962). While it was recognized that much of the wind noise was in fact due to breaking, it is only in the last ten years that attention has focussed more directly, and in a quantitative fashion, on the relationship between ambient noise at the ocean surface and surface wave breaking. The beginnings of much of this work are summarized in the proceedings of a wonderful conference which brought together acousticians and hydrodynamicists in Lerici in 1987 (Kerman, 1988). By that time field measurements had shown that discrete acoustic events could be identified with breaking (Farmer & Vagle, 1988) and there was considerable speculation about the source of the sound. Sources of sound associated with the entrainment of air are almost certainly the dominant source of sea-surface sound. After entrainment the volume of air is broken up into smaller and smaller bubbles. At the time of its generation a bubble is not in its equilibrium shape and as it relaxes to this spherical shape it oscillates in its axisymmetric ("breathing") mode. This volume mode of oscillation is the most efficient for the radiation of sound and has a characteristic frequency

$$\omega^2 = 3\gamma P/\rho r^2, \tag{21}$$

where γ is the ratio of the specific heats of the gas, P is the equilibrium pressure, ρ is the density of the liquid and r is the equilibrium bubble radius. For a bubble of radius 1mm in water at atmospheric pressure, the resonant frequency is approximately 3.26 kHz. Banner & Cato(1988) presented a video recording of the sound generated in this fashion by bubbles pinching off in a quasisteady spilling breaker. In a series of laboratory experiments, Medwin and his colleagues (Medwin & Beaky, 1989; Medwin & Daniel, 1990) have shown that this is the dominant source of sound at frequencies greater than approximately 400-500Hz.

We were concerned to determine whether the sound generated by breaking could be used to do more than just identify breaking events. Preliminary experiments were undertaken which showed that for frequencies greater than 500Hz the sound radiated by breaking waves in the laboratory correlated with the energy dissipated (Melville et al., 1988). Now if the causes of this correlation could be understood, and if the result also applied to the field, then this would offer an efficient way of monitoring surface wave dissipation, a difficult measurement to make directly (See above.) These preliminary results were confirmed in more extensive measurements (Loewen & Melville, 1991a) which showed that of the

mass. This is usually reserved for heat transfer.

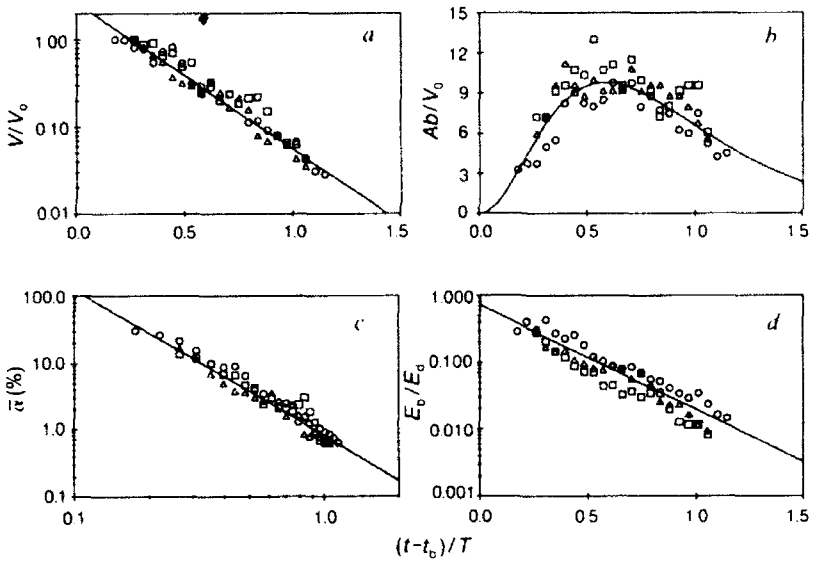


Figure 7. Moments of the void fraction field measured under unsteady breaking waves in the laboratory. V/V_0: normalized total volume of air; Ab/V_0: normalized area of the bubble cloud, where b is the breadth of the channel; $\bar{\alpha}$: mean void fraction by percent; E_b/E_d: ratio of the potential energy of the bubble cloud to the total energy dissipated in breaking. (Lamarre & Melville, 1991a).

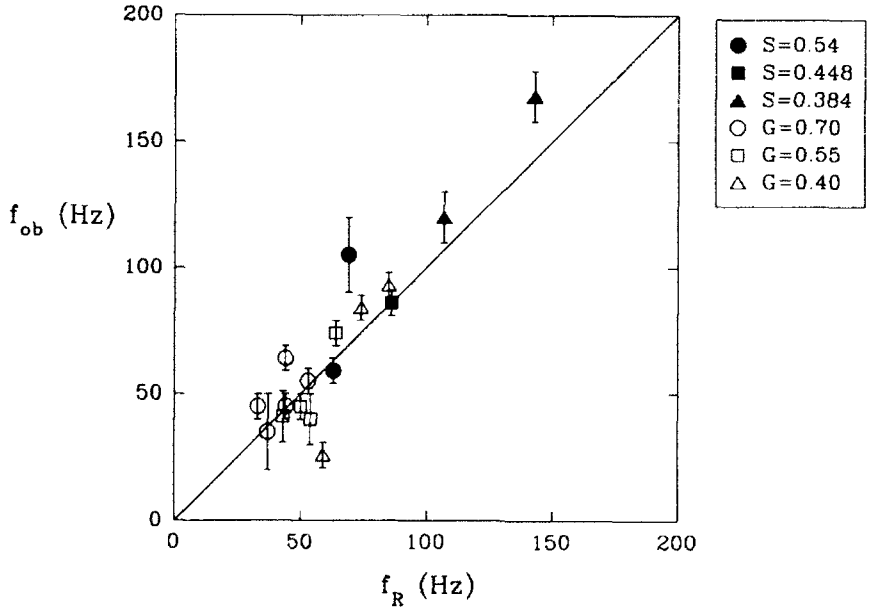


Figure 8. Comparison of measured frequency of collective oscillations under unsteady breaking waves and computed frequency based on void fraction measurements (Loewen& Melville, 1992).

order of 10^{-8} of the mechanical energy dissipated was radiated as sound. Simple semi-empirical modelling (Loewen & Melville, 1991b) showed that for gently spilling waves the sound radiated above approximately 400Hz (Medwin & Beaky, 1989; Medwin & Daniel, 1990) could be reproduced with a simple dipole model of bubbles entrained at the surface. Inplicit in this result is the fact that the sound radiated is proportional to the volume of air entrained.

While these results were encouraging for gently spilling waves they did not account for the air entrainment and acoustics of more energetically breaking waves in which large volumes of air are entrained. In these cases it more useful to consider the whitecap as a continuum described by its geometry and void fraction field, rather than a collection of discrete bubbles. By using the electrical conductivity of the air-water mixture in the whitecap we were able to measure the void fraction field and hence the evolution of the bubble cloud in unsteady breaking waves (Lammare & Melville, 1991a) . It was found that all of the lower order moments of the void-fraction field evolved as simple exponential or power law functions in time (Fig. 7). But the most important result of this work was to show that the energy expended in entraining the air against buoyancy forces could account for up to 50% of the total mechanical wave energy dissipated. This result serves to support the earlier empirical correlation between energy dissipation and sound radiation. Subsequent field work using similar conductivity probes showed that the large void fractions measured in the laboratory ($O(10\%)$) were also to be found in the field (Lamarre & Melville, 1991b).

Bubble clouds may also oscillate as a whole in what are called "collective oscillations". The possibility that such oscillations could occur in bubble clouds entrained by breaking waves had been considered by Prosperetti (1988) and by Carey and Browning (1988) as a source of low frequency sound ($< 500Hz$) at the ocean surface. The lowest mode of oscillation of a spherical bubble cloud has a frequency

$$\Omega = \frac{1}{R}\left(\frac{3\gamma P_0}{\rho_w \bar{\alpha}}\right)^{1/2}, \tag{22}$$

where R is the radius of the cloud, γ is the ratio of the specific heats, P_0 is the equilibrium pressure inside the individual bubbles comprising the cloud, and $\bar{\alpha}$ is the mean void fraction. If isothermal conditions apply $\gamma = 1$. For example, if $R_0 = 0.3m$, $\bar{\alpha} = 0.1$, and $P_0 = 101kPa$ then $\Omega = 30 \times 2\pi rads^{-1}$. Confirmation of the essential characteristics of collective oscillations in laboratory models have been obtained by Yoon et al. (1991) who generated a vertical bubble column above a bubble generator, and by Kolaini et al.,(1991) who dropped volumes of water onto still water surfaces. However, Hollet (1989) and Farmer & Vagle (1988) had earlier shown in the field that low frequency sound accompanies wave breaking. Evidence that collective oscillations may be the source of low frequency sound in breaking waves was provided in two complementary sets of laboratory measurements. In both two- and three-dimensional breaking waves Lamarre & Melville (1992) measured the void fraction. Following Lu et al. (1990) these data were then used to predict the eigenfrequencies of the collective oscillations of the bubble clouds and compared with low frequency sound measurements by Loewen & Melville(1992). It was found that for spilling breaking the sound radiated below approximately 500Hz did not vary with the strength of the break; however, on the transition to plunging breaking the

low frequency sound was clearly a function of the strength of the break, and that spectral peaks could be identified. The results of this comparison between measured and predicted collective oscillation are shown in Figure 8.

The success of these relatively simple acoustic models in predicting parameters describing the sound radiated by breaking waves is reason for confidence that ambient sound may be a useful tool for learning more about surface fluxes. For example, acoustic characterization of collective oscillations may lead to estimates of the volume of air entrained by breaking waves. This is important initial data for models of gas transfer.

8. ACKNOWLEDGEMENTS

Many of the results reviewed here were obtained in joint work with my students at Massachusetts Institute of Technology and Scripps Institution of Oceanography. My work in the area of surface waves has been generously supported by grants from the National Science Foundation, the National Aeronautics and Space Administration, and the Office of Naval Research.

9. REFERENCES

Agrawal, Y.C., Terray, E.A., Donelan, M.A., Hwang, P.A., Williams, A.J. III, Drennan, W.M., Kahma, K.K. & Kitaigorodskii, S.A., Nature, in press (1992).
Baggeroer, A. & Munk, W., Physics Today, 45 (1992) 22.
Banner, M.L., J. Phys. Ocean., 20 (1990) 966.
Banner, M.L. & Cato, D.H., in *Sea Surface Sound - Natural Mechanisms of Surface generated Noise in the Ocean*, Ed. B.R. Kerman, 429 Kluwer, Dordrecht, 1988.
Banner, M.L. & Fooks, E.H., Proc. Roy. Soc. Lond. A 399 (1985) 93.
Benjamin, T.B. & Feir, J.E., J. Fluid Mech., 27 (1967) 417.
Bonmarin, P. J. Fluid Mech., 209 (1989) 405.
Carey, W.M., & Browning, D.G., in *Sea Surface Sound - Natural Mechanisms of Surface generated Noise in the Ocean*, Ed. B.R. Kerman, Kluwer, Dordrecht, 1988.
Coles, K.A. *Heavy Weather Sailing.* John de Graff Inc., Clinton Corners, N.Y., 1978.
Dold, J.W. & Peregrine, D.H., *Proc. 20th Conf. Coastal Eng.*, Taipei, A.S.C.E., 1 (1986) 163.
Dommermuth, D.G., Yue, D.K.P., Rapp, R.J., Chan, E.-S. & Melville, W.K., J. Fluid Mech., 89 (1987) 432.
Donelan, M.A., Hamilton, J. & Hui, W.H., Phil. Trans. Roy. Soc. Lond. A 315 (1985) 509.
Farmer, D.M. & Vagle, S., J. Geophys. Res., 93 (1988) 351 91.
Forristall, G.Z., J. Geophys. Res., 86 (1981 8075.
Gill, A.E., *Atmosphere-Ocean Dynamics*, Academic, London, 1982.
Hasselmann, K., J. Fluid Mech., 12 (1962) 481.
Hasselmann, K., Boundary- Layer Met., 6 (1974) 107.
Hollet, R., J. Acous. Soc. Am., 85(S1) (1989) S145.
Holthuijsen, L.H. & Herbers, T.H.C., J. Phys. Ocean., 16 (1986) 290.
Hsu, C.T., Hsu, E.Y. & Street, R.L., J. Fluid Mech., 105 (1981) 87.

Hsu, C.T., Wu. H.W., Hsu, E.Y. & street, R.L., J. Phys. Ocean., 12 (1982) 929.
Jaehne, B., In *Proc. Air-Water Mass Transfer: Second Int. Symp.*, Minneapolis, ASCE, 1990
Jessup, A.T., Keller, W.C. & Melville, W.K., J. Geophys. Res., 95 (1990) 9679.
Jessup, A.T., Melville, W.K. & Keller, W.C., J. Geophys. Res., 96 (1991a) 20547.
Jessup, A.T., Melville, W.K. & Keller, W.C., J. Geophys. Res., 96 (1991b) 20561.
Kawai, S, Okada, K. & Toba, Y., J. Ocean. Soc. Japan, 33 (1977) 137.
Keller, W.C., Plant, W.J. & Valenzuela, G.R., In *Wave Dynamics and Radio Probing of the Sea Surface*, Eds O.M. Phillips & K. Hasselmann. 295 Plenum, New York, 1986.
Kerman, B.R., Ed., *Sea Surface Sound - Natural mechanisms of Surface Generated Noise in the Ocean*, Kluwer, Dordrecht, 1988.
Khoo, B.C. & Sonin, A.A., J. Geophys. Res., 97 (1992) 14413.
Kitaigorodskii, S.A., J. Phys. Ocean., 13 (1983) 816.
Kitaigorodskii, S.A., J. Phys. Ocean., 14 (1984) 960.
Kolaini, A., Roy, R.A., Crum, L.A., J. Acous. Soc. Am., 89 (1991) 2452.
Komen, G.J., Hasselmann, S. & Hasselmann, K., J. Phys. Ocean., 14 (1984) 1271.
Lamarre, E. & Melville, W.K., Nature 351 (1991a) 469.
Lamarre, E. & Melville, W.K., IEEE J. Ocean. Eng., 17 (1991b) 204.
Lamarre, E. & Melville, W.K., J. Acous. Soc. Am., sub judice (1992).
Loewen, M.R. & Melville, W.K., J. Fluid Mech., 224 (1991a) 601.
Loewen, M.R. & Melville, W.K., J. Acous. Soc. Am., 90 (1991b) 2075.
Loewen, M.R. & Melville, W.K., J. Acous. Soc. Am., sub judice (1992).
Longuet-Higgins, M.S., in *Sea Surface Sound - Natural Mechanisms of Surface Generated Noise in the Ocean* B.R. Kerman, Ed., Kluwer, Dordrecht, 1988.
Longuet-Higgins, M.S. & Smith, N.D., J. Geophys. Res., 88 (1983) 9823.
Lu, N.Q., Prosperetti, A. & Yoon, S.W., IEEE J. Ocean. Eng., 15 (1991) 275.
Mc Lean, J.W., J. Fluid Mech., 114 (1982) 315.
Medwin, H. & Beaky, M.M., J. Acous. Soc. Am., 86 (1989) 1124.
Medwin, H. & Daniel, A.C., J. Acous. Soc. Am., 88 (1990) 408.
Melville, W.K., J. Fluid Mech., 115 (1982) 163.
Melville, W.K., J. Fluid Mech., 128 (1983) 489.
Melville, W.K. & Rapp, R.J., Nature 317 (1986) 514.
Melville, W.K. & Rapp, R.J., J. Fluid Mech., 189 (1988) 1.
Melville, W.K., Loewen, M.R., Felizardo, F.C., Jessup, A.T. & Buckingham, M.J., Nature 336 (1998) 54.
Miles, J.W., J. Fluid Mech., 3 (1957) 185.
Miles, J.W., J. Fluid Mech., 13 (1962) 433.
Mitsuyasu, H., J. Geophys. Res., 90 (1985) 3343.
Monahan, E.C. & Mc Niocaill, G., Eds, *Ocean Whitecaps and their Role in Air-Sea Exchange Processes.* Reidel, Dordrecht, 1986.
Phillips, O.M., J. Fluid Mech., 4 (1958) 426.
Phillips, O.M., J. Fluid Mech., 9 (1960) 193.
Phillips, O.M., *The Dynamics of the Upper Ocean.* 2nd ed., Cambridge, 1977.
Phillips, O.M., J. Fluid Mech., 156 (1985) 505.
Prosperetti, A., J. Acous. Soc. Am., 84 (1988) 1042.

Plant, W.J., J. Geophys. Res., 87 (1982) 1961.
Rapp, R.J. & Melville, W.K., Phil. Trans Roy. Soc. Lond. A 331 (1990) 735.
Snyder, R.L., Dobson, F.W., Elliott, J.A. & Long, R.B., J. Fluid Mech., 102 (1981) 1.
Spitzer, W.S. & Jenkins, W.J., J. Marine Res., 47 (1989) 169.
Su, M.-Y., Bergin, M., Marlev, P. & Myrick, R., J. Fluid Mech., 124 (1982) 45.
Thorpe, S.A., Ann. Geophys., 2 (1984) 53.
Thorpe, S.A., Q. J. Roy. Met. Soc., 118 (1992) 1.
Thorpe, S.A. & Humphries, P.N., Nature 283 (1980) 463.
Weissman, M.A., Katsaros, K.B., Atakturk, S.S., J. Phys. Ocean., 14 (1984) 1608.
Wallace, D.W.R. & Wirick, C.D., Nature 356 (1992) 694.
Yoon, S.W., Crum, L.A., Prosperetti, A. & Lu, N.Q., J. Acous. Soc. Am., 89 (1991) 700.
Wenz, G.M., J. Acous. Soc. Am., 34 (1962) 1936.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Extreme Waves and Breaking Wavelets

O.M. Phillips

Dept. of Earth & Planetary Sciences
The Johns Hopkins University
Baltimore, MD 21218, USA

Abstract

This review summarises some recent findings about ocean surface waves at the two extremes of scale - the expected configurations of extreme wave groups (giant waves) on the one hand and on the other, the processes whereby short wind-generated gravity waves are suppressed by their interaction with longer waves. A simple approximate theory expresses the space-time configuration of extreme waves in terms of the autocorrelation function of the wave field as a whole, and this is tested by buoy and Surface Contour Radar measurements made in the field during the SWADE experiment. The interactions between (not so extreme) long wave groups and shorter wind-generated gravity waves have been explored in the laboratory in the NASA-Wallops Wind Wave facility. The short waves are suppressed by breaking during the passage of the group, their significant slope at the long wave crests remaining constant as the amplitude and wavelength both decrease. The gradual regeneration of these waves by the wind is interrupted in the tank by a wave energy front originating from the upwind end.

Introduction

People who look at waves on the surface of the sea must have noticed both their randomness and their quasi-regularity. Their detailed configuration changes constantly, yet the dominant wave groups usually have a comforting regularity. Occasionally, an extremely high crest or deep trough 'appears from nowhere' (as eyewitnesses have claimed) sometimes causing damage or loss of life. In a moderate wind, the surface may be covered by distinct wave groups or trains a meter or two in length, but under high winds with much longer big waves, the surface, streaked with spume and rippled with very short capillary waves, often seems to have largely lost these intermediate scale gravity waves. In a hard driven wind sea when the longest dominant waves are breaking, the breaking zones erase shorter slower waves leaving a smoother surface behind. When the longest waves themselves are not breaking, any shorter waves present do tend to break, intermittently and sporadically, near the crests of the largest waves. The dynamics of the breaking process have been elucidated greatly by the fine experiments of Duncan (1981), Melville and Rapp (1985) and others.

There is clearly a melange of processes occurring, and I would like to concentrate on just two of these. The work that I review is partly theoretical; the measurements have been made both in the laboratory and in the field, and a number of people have contributed substantially. Daifang Gu was involved with the theory and did the data analysis of extreme waves; Mark Donelan and Edward Walsh generously provided, and allowed our use of, hard-won data from the Surface Wave Dynamics Experiment (SWADE) in the western Atlantic Ocean off Virginia; Jacob Chu and Steven R. Long conducted the NASA-Wallops experiments and the extensive data analysis. To them all, my sincere thanks.

Extreme Wave Form

Extreme wave events, giant waves or rogue waves, have been responsible for many marine accidents, some involving loss of life. They occur under storm conditions when the waves are already high, perhaps amplified further by refraction in currents such as the Agulhas current off eastern South Africa or the Gulf Stream. The very highest individual waves, the rogue waves, seem to appear with little warning and may be regarded as the statistical extremes in an already rough random sea, occurring sporadically in space and time. Although the *occurrence* of these events may be random, it is of interest to inquire whether in the vicinity of extreme wave crests, there is any predictable, expected configuration of the sea surface, any 'organized structure' of the surface in space and time with which a mariner and his ship must cope.

Any such 'organized structure' in a random sea (or random function in general) is not of course deterministic in the usual sense, but the selection of regions surrounding extreme maxima does extract a regularity from the randomness of the field as a whole. Suppose, for example, we have an extensive record $\zeta(\underline{x},t)$ of surface displacement in a random wind-generated wave field as a function of position and time. From this record, let us pick out the high wave crests - extract instances of wave maxima lying between ζ_m and $\zeta_m + d\zeta_m$, where ζ_m is a large multiple of the r.m.s. wave height, and consider the nature of the surface displacements surrounding these maxima. *At the maxima, the surface displacements are all in essence the same,* ζ_m. Close to the maxima, the surface displacements $\zeta < \zeta_m$. The expected or mean surface displacement is somewhat less than ζ_m while still being very large, and the variance about this mean is small. As the distance in space or time from the maxima increases, the expected value of ζ decreases and the variance among different realizations increases until ultimately, far from the extreme events, the order is lost, the expected position of the free surface is simply the mean water level, $\bar{\zeta} = 0$ and the variance is that of the overall wave field. This paper is concerned with the question: Given the existence of an extreme wave crest, what is the *expected* surface configuration surrounding this crest or, equivalently, the *mean over many realizations of extreme events*, and what is the distribution of variance about this mean? This question seems to have been asked first by Boccotti (1981-1989) of the Università di Reggio Calabria in Italy. In a random sea, one cannot predict where or when extreme crests or troughs will appear, but one can predict what their expected configuration will be when they do occur.

The direct calculation of the expected surface displacement $\bar{\zeta}(\underline{x},t)$, given that at the origin, say, ζ has an instantaneous maximum whose value ζ_m is a substantial multiple of the r.m.s. value $\sigma = (\overline{\zeta^2})^{1/2}$, is extremely cumbersome and has not been done. The defi-

nition of a maximum requires that the surface slope $\nabla\zeta$ be zero, the curvature be negative, ζ be zero and the vertical acceleration negative and conditional probabilities with all these restrictions make for impossibly tedious algebra. Let us ask the question in a different way.

Rather than seeking the precise points where realizations of $\zeta(\underline{x},t)$ attain maxima, let us consider those regions where $\zeta \geq \gamma(\overline{\zeta^2})^{1/2} = \gamma\sigma$, where γ is (formally) a number large compared with unity. For a given γ, at any instant these regions consist of isolated islands each containing at least one maximum and as γ increases, the islands shrink, converging towards the maxima and then disappearing. Our interest is in large values of γ, where there are rare, small, isolated islands in which $\zeta \geq \gamma\sigma$ and let us pose the question thus: Given that at $\underline{x},t$, say, $\zeta > \gamma\sigma$ where γ is large, what is the expected distribution of ζ in the vicinity and what is the standard deviation about this expected value? The expected distribution in space and time describes the configuration and evolution of the extreme wave events and the standard deviation is the random uncertainty. This formulation allows a very much simpler calculation.

Accordingly, let $\zeta_1 = \zeta(\underline{x},t)$, $\zeta_2 = \zeta(\underline{x}+\underline{r},t+\tau)$, $\overline{\zeta^2} = \sigma^2$ and $\rho(\underline{r}\cdot\tau) = \overline{\zeta_1\zeta_2}/\sigma^2$, the autocorrelation function for the surface displacement. From the theorem of conditional probability, the distribution of ζ_2 given that $\zeta_1 \geq \gamma\sigma$ is, in the usual notation,

$$p(\zeta_2|\zeta_1 \geq \gamma\sigma) = p(\zeta_2, \zeta_1 \geq \gamma\sigma)/p(\zeta_1 \geq \gamma\sigma).$$

and, for a Gaussian process, the righthand side is easy to calculate (Phillips, Gu and Donelan, 1992). After some calculation, it is found that the expected value of ζ_2 given that ζ_1 is $> \gamma\sigma$ is then

$$\overline{\zeta_2} = \int_{-\infty}^{\infty} \zeta_2 p(\zeta_2|\zeta_1 \geq \gamma\sigma)d\zeta_2 \ ,$$

$$= \gamma\sigma\rho[f(\gamma)]^{-1} \ , \qquad (2.6)$$

where $f(\gamma)$ is an integral expression whose asymptotic form is

$$f(\gamma) \sim 1 - \gamma^{-2} + 3\gamma^{-4} - \ldots \quad \text{for} \quad \gamma \gg 1. \qquad (2.4)$$

In particular, when $r = 0, \tau = 0$, the autocorrelation function $\rho = 1$ so that the average height of those waves higher than $\gamma\sigma$ is $\gamma\sigma[f(\gamma)]^{-1}$ so that

$$\frac{\overline{\zeta(\underline{x}+\underline{r},t+\tau)}}{\overline{\zeta(\underline{x},t)_{\zeta>\gamma\sigma}}} = \rho(\underline{r},\tau).$$

The simplicity of this result is a surprise. It is an approximation in two senses - in the use of a Gaussian distribution and in averaging over points in islands surrounding maxima rather than averaging relative to the maxima themselves. Let us see how well this simple approximation works.

When $r = 0$, we are concerned with waves measured in time at a single point by a wave staff or buoy. In this simple case, the direct calculation of the expected history of surface displacement surrounding a high maximum is manageable with a narrow spectrum approximation (Phillips, Gu and Donelan, 1992). Though more cumbersome than the calculation leading to the result above, it does provide a theoretical check on the latter. Time series data obtained in SWADE by a buoy at 38.37°N, 73.65°W during a storm on 26 October, 1990 appeared to be closely Gaussian, and the autocorrelation functions $\rho(\tau)$ were calculated for each of four data segments. The records were then searched for maxima larger than a chosen multiple of σ. The time of occurrence of a high wave is denoted by t_m and the average surface displacement in the vicinity, $\bar{\zeta}(t_m + \tau)$ was found as a fraction of $\zeta(t_m)$ by averaging over the ensemble of maxima. A sample of this is given in Figure 1, the solid line showing the expected configuration using the direct calculation and the narrow spectrum

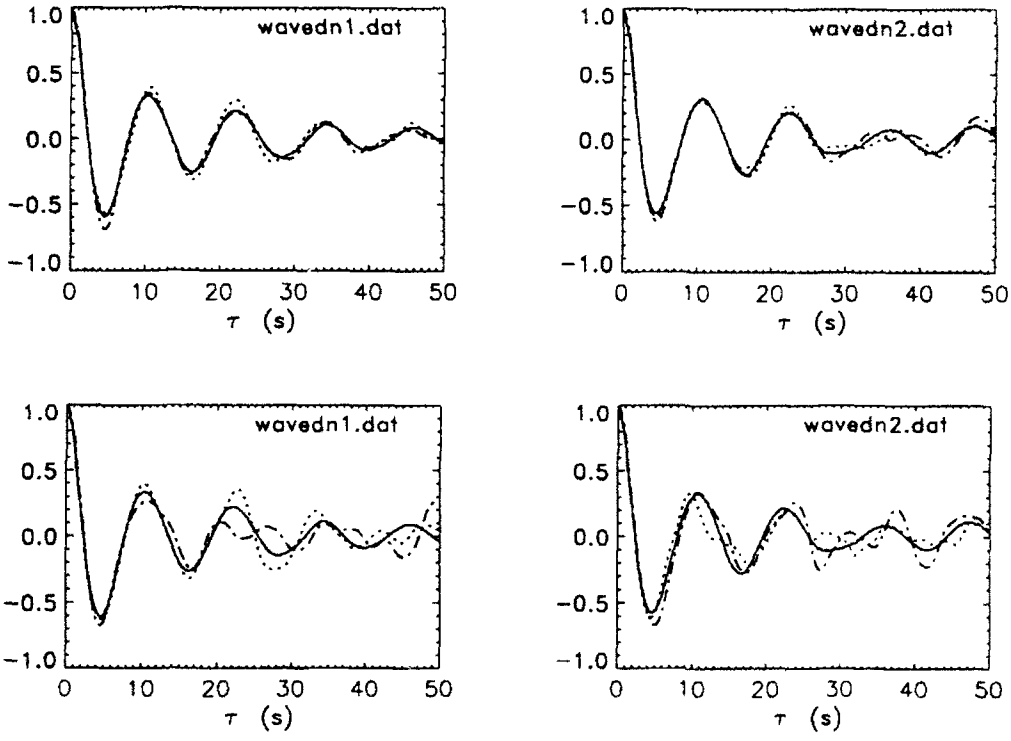


Figure 1. Averaged time histories of $\zeta(t)$ in buoy data surrounding high maxima with $\gamma = 2$ (top) and 2.5 (bottom) compared with the expected configuration (solid line) calculated for a narrow spectrum. For $\gamma = 2.5$ the sample size is of course smaller.

approximation, while the dotted and dot-dash lines show the means of the displacements following and preceding the maximum. Figure 2 shows a comparison between the approximate theory (solid line) and the measured averaged time histories surrounding maxima. The direct calculation and the approximate theory are virtually indistinguishable in this case (which is typical of others examined); near the maximum the measured average displacements are very close to the theory but develop more random variations as the time interval

increases and correlation is lost. Far from the maxima of course, the expected surface displacement is zero with a standard deviation of σ, that of the wave field as a whole.

A much more interesting case, and a much more demanding test of the result above, concerns the expected *spatial* configuration of the surface surrounding high maxima. No direct calculation is available for this, but according to the approximate theory, the expected instantaneous configuration surrounding high crests should be given by the spatial autocorrelation function, with $\tau = 0$. This function is of course the Fourier Transform of the two-dimensional wave spectrum and its form will depend on the directional characteristics of this spectrum as well as its scalar wave-number dependence.

Extensive data on the spatial configuration of the sea surface under various wave conditions have been obtained by Dr. Edward Walsh's group during SWADE using their surface contour radar.

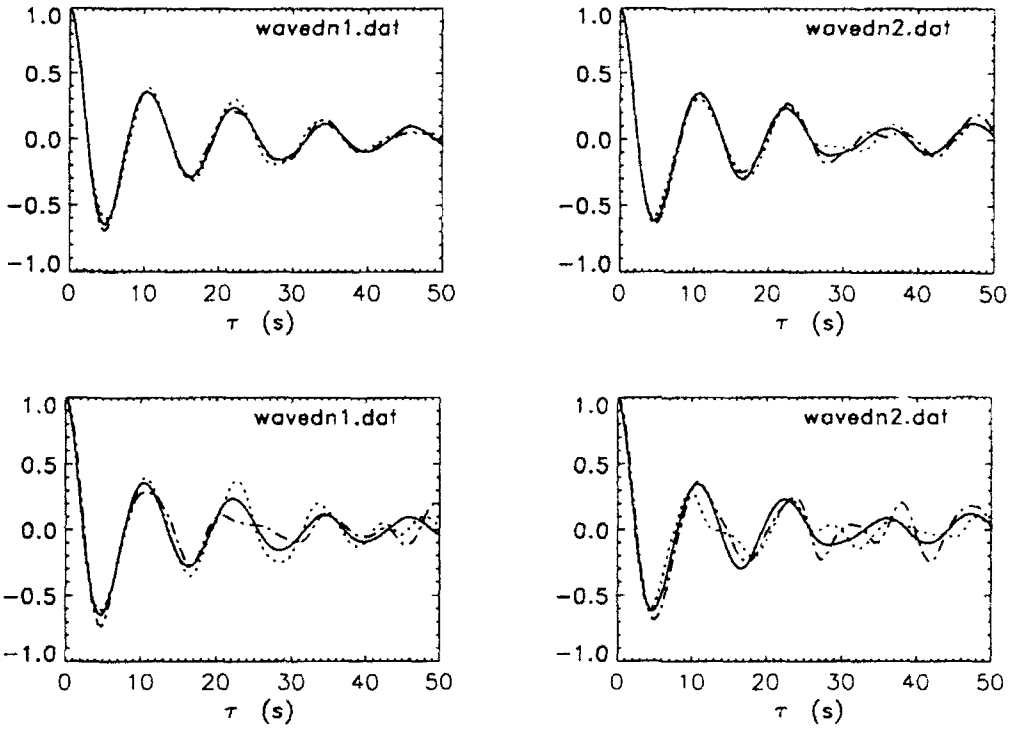


Figure 2. As in figure 1, but now compared with the temporal autocorrelation function $\rho(\tau)$.

Briefly stated, the surface contour radar is an aircraft-mounted radar altimeter whose beam sweeps to and fro transverse to the flight path; the distance from aircraft to the water surface being measured by the return time of the individual radar pulses. The transverse sweeps mark out a swath along the flight track whose width is proportional to the aircraft altitude and the measurements provide a nearly instantaneous 'map' of the wave pattern in the swath below. From the digitized data, the spatial autocorrelation

function $\rho(\underline{r})$ can readily be found; as before, the data are then searched for high maxima (or low minima) and the ensemble of surface configurations around these points is averaged to define $\bar{\zeta}(\underline{r})/\bar{\zeta}_m$. This is to be compared with $\rho(\underline{r})$.

As examples, consider two contrasting wave conditions encountered on 4 and 5 March 1991. For a couple of days prior to the 4th, the wind was fairly steady from the south, and the wave field at the measurement region (south and east of the Gulf Stream) was uni-modal, with dominant waves moving north, having a narrow directional spread. The significant wave height was about 5 m. The spatial autocorrection function of waves in the entire swath, shown in Fig-

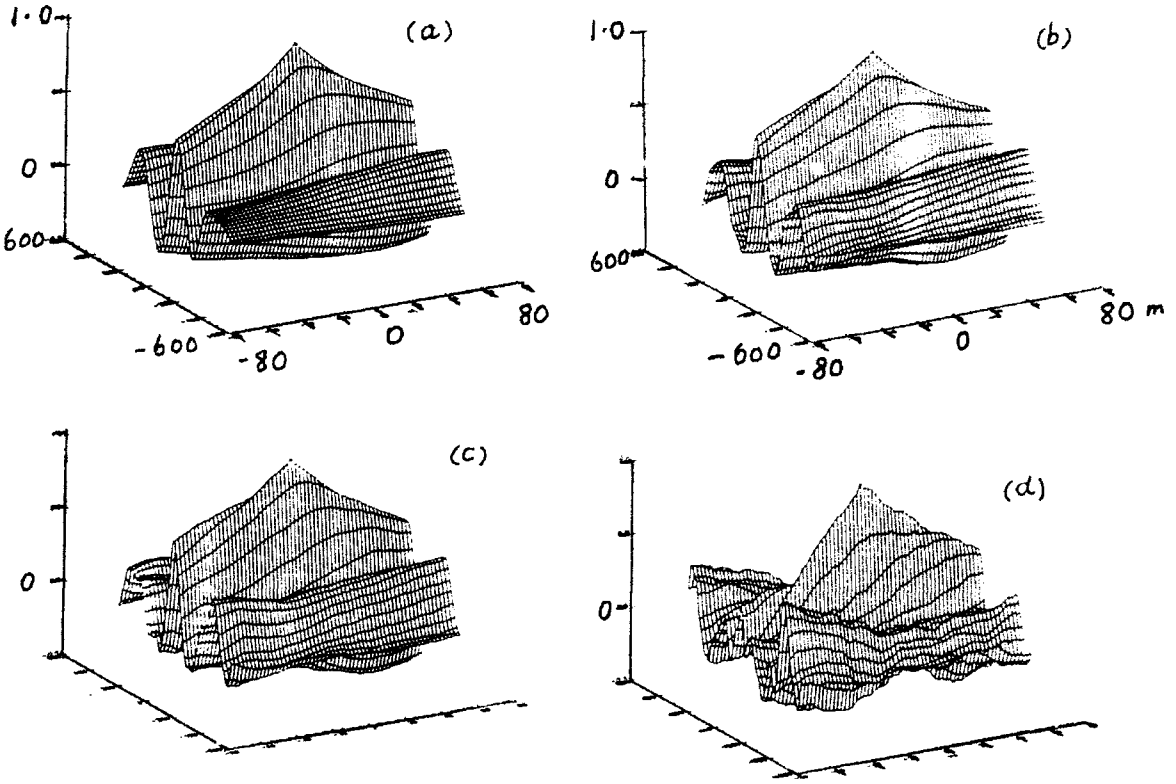


Figure 3. (a) Spatial autocorrelation function $\rho(\underline{r})$ in a uni-modal wind sea. (b, c, d), the mean surface configurations of the sea surface surrounding high maxima with $\gamma = 2, 2.5$ and 3.5 respectively.

ure 3 a, is consistent with a field of rather long-crested waves, with a dominant wavelength of about 250 m travelling slightly to the right of the flight path. Maxima for $\zeta > \gamma\sigma$ were then identified for values of $\gamma = 2$, 2.5 and 3.5 and the ensemble averaged surface configurations in the vicinities of the maxima calculated. The results are shown in Figures (3b, c, d). For $\gamma = 2$ the average configuration

is remarkably close to $\rho(\underline{r})$; as γ increases, the theory should get better but the statistics deteriorate (fewer realizations) so that the 'noise' increases. The dominant characteristics, however, persist.

During the following night there was a cold air outbreak, with strong winds from the North-West and on 5 March, further measurements were taken a little to the south of the previous site. The combination of wind sea, originally travelling to the south east, but refracted and scattered by passage through the Gulf Stream, a residual 10-second swell travelling north produced a sea state with a very wide directional spread in its dominant waves. The autocor-

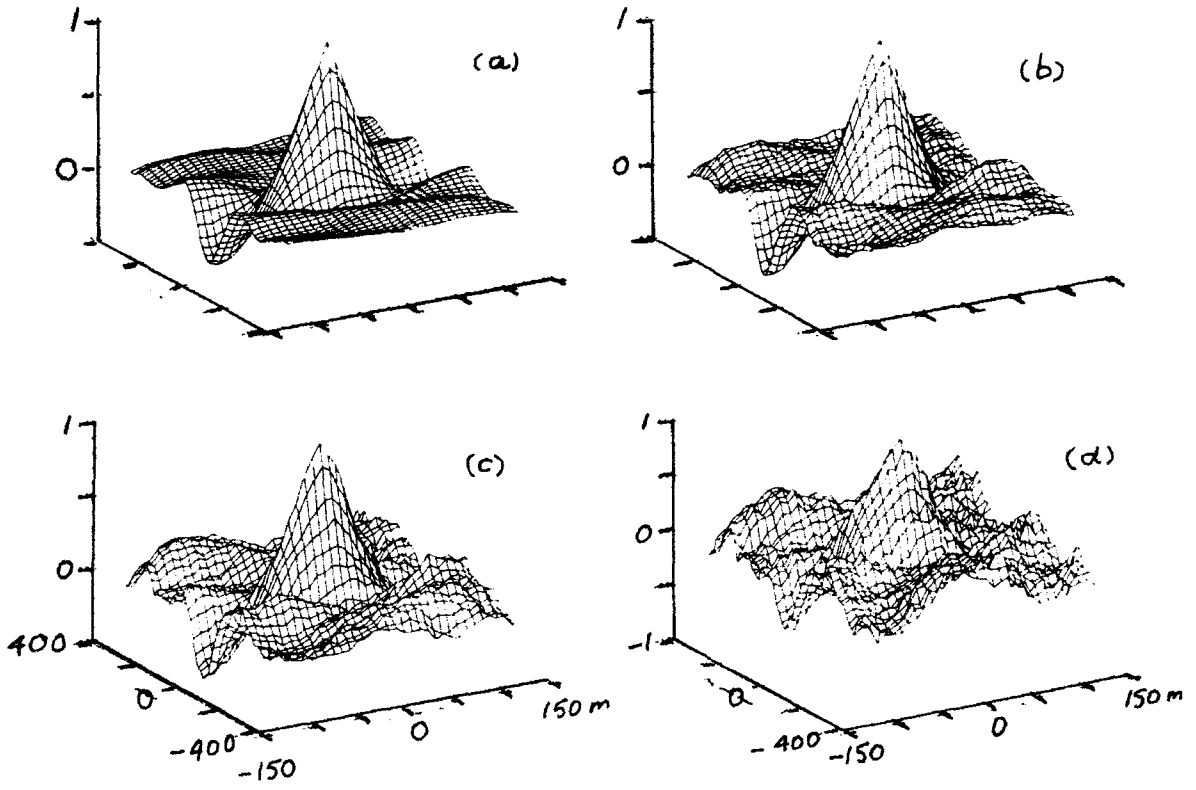


Figure 4. (a) Spatial autocorrelation function $\rho(\underline{r})$ in a wave field with a wide directional spread of the dominant frequencies. (b, c, d) mean instantaneous surface configurations surrounding high maxima with $\gamma = 2, 2.5$ and 3 respectively.

relation function $\rho(\underline{r})$ had the rather remarkable form shown in Figure (4a) and the mean instantaneous configurations surrounding high maxima are illustrated in Figures (4b, c, d) for increasing values of γ (remember that the vertical scale *is* exaggerated). The waves are evidently very short crested as is consistent with their wide directional spread, having more the nature of a transient standing

wave formed from wave groups travelling in almost opposite directions. The average configuration surrounding minima with $\gamma \approx 2$ is shown in figure 5 - to a helmsman, the trough would look like a 'black hole'. Eyewitness reports of giant waves frequently described them as "coming out of nowhere". It is interesting to note that this location is near "the north wall of the Gulf Stream", a region infamous for erratic and extreme waves.

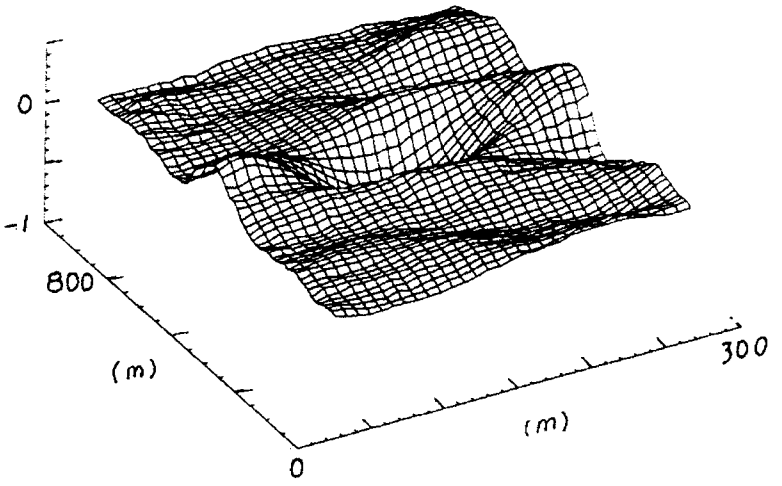


Figure 5. Mean configuration surrounding extreme troughs in the same data set as for figure 4, $\gamma = 2$. The bottom of the trough (out of sight) extends to -1.

Short Wave Suppression

At the other end of the spectrum, we have the short gravity and capillary waves. Mitsuyasu (1966) discovered that in the presence of swell, short gravity wave components of a wind-wave field were considerably smaller than they were without the swell, all other conditions being the same. The effect was also observed by Phillips and Banner (1974) and Donelan (1987) in wind-generated waves superimposed on continuous trains of swell. Although the phenomenon is well known, there is no consensus about the mechanisms involved. Phillips and Banner ascribe it to enhanced breaking of the wind waves at the swell crests, while Donelan pointed out that the oscillatory straining of the short waves by the swell can disrupt the energy transfer to them from longer waves by resonant wave interactions. It occurred to us that the nature of the mechanism might be clarified if the wind waves were subjected to the passage of a group of longer waves rather than a continuous train of them. Enhanced wave breaking provides a rapid energy loss and if this is the process responsible, the reduction in energy density should also be found after a long wave group had passed. On the other hand, wave-wave intereactions have time scales very much longer than the wave period so the disruption of them as the group passes should have little net effect.

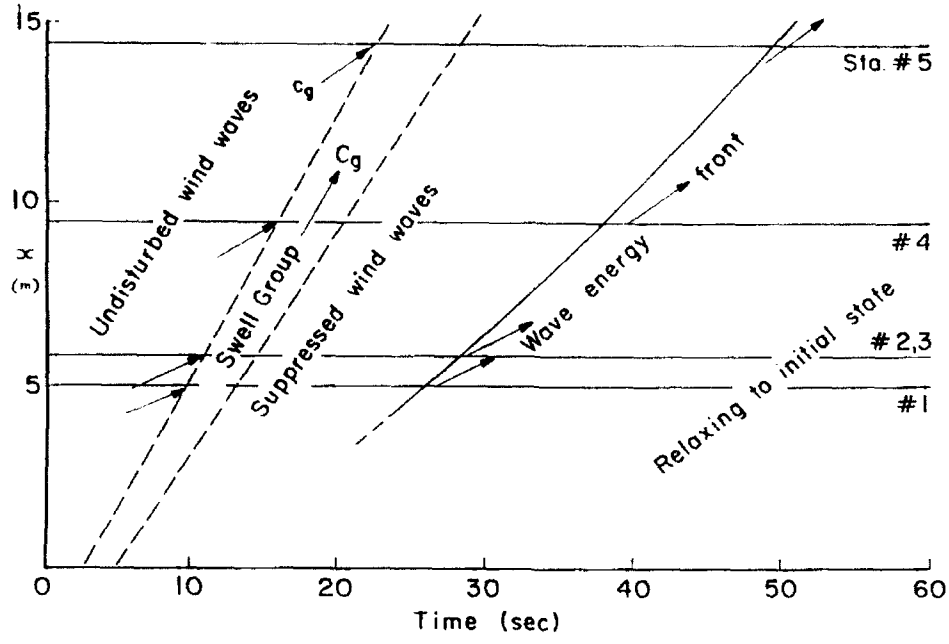


Figure 6. An $x-t$ diagram of the wind-wave, swell group interaction experiment.

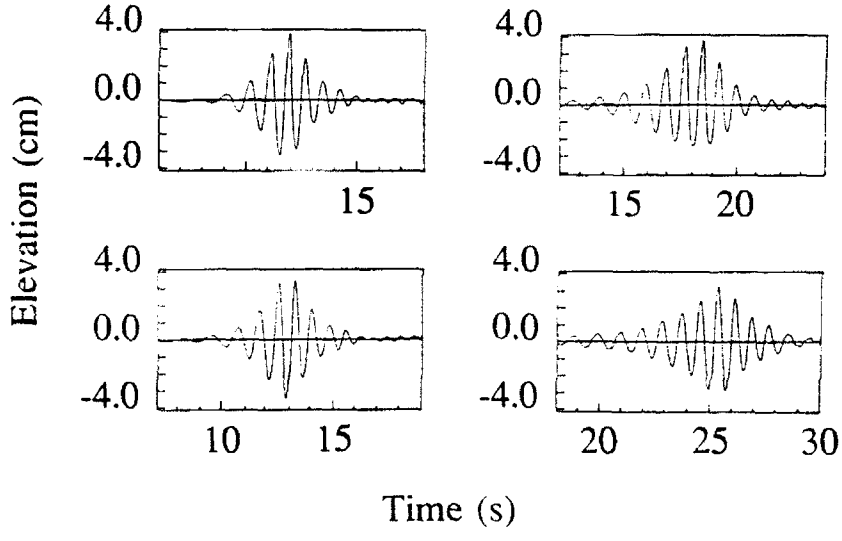


Figure 7. Ensemble means of the surface displacement specify the propagation of wave groups past the measurement stations. Left column, fetches of 5.2, 5.9m; right, 9.5, 14.4m.

The experiments that I will describe now, then, involve a pre-existing statistically steady wind wave field in a laboratory wind wave facility into which, at the upstream end, groups of longer, mechanically generated waves are injected, the time histories of the wave fields being measured at various points down the tank before, during and after the passage of the groups. Because of the randomness of the wind-generated waves and the absence of stationarity, it

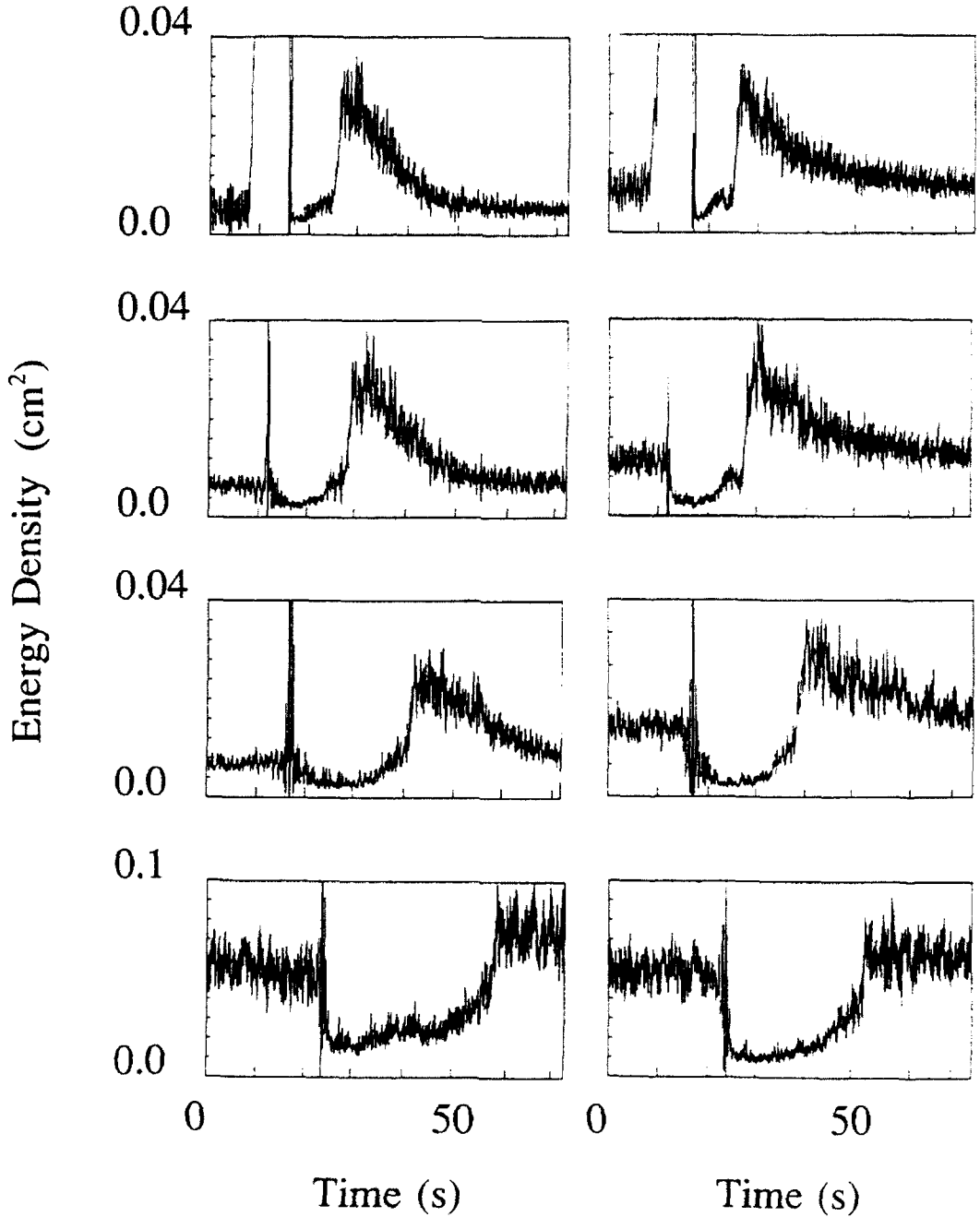


Figure 8. Time histories of $\overline{\zeta^2}$ for the wind waves at $u_* = 15 cm/s$ (left column) and 27 cm/s (right) at fetches 5.1, (top row), 5.9m, 9.5m and 14.4m (bottom row).

was necessary to repeat each experiment a large number of times (100), recording the data digitally in order to obtain ensemble averages. Although initially motivated by the questions described above, the experiment revealed a number of different phenomena and several surprises, and to put them into context, consider the x-t diagram of Figure 6 which provides a 'road-map', of the experiment, with fetch down the tank plotted vertically and elapsed time horizontally. Initially, there is a statistically steady field of wind waves growing with fetch; the stationary observation points are indicated by the horizontal lines. A group of longer waves is then generated at zero fetch and passes down the tank, moving with its group velocity, catching up and overtaking the wind waves whose group velocity is less. For each set of conditions, the experiment is repeated 100 times with identical input pulses of the wavemaker. The ensemble average of the surface displacement is just the (deterministic) long wave group (Figure 7); when this is subtracted from each realization, the residual is the distribution of short wind-generated waves (as a function of time at each measurement station) as modified by interaction with the long wave groups. Care had to be exercised in extracting short wave information particularly during the interaction, and details are given by Chu, Long and Phillips (1992).

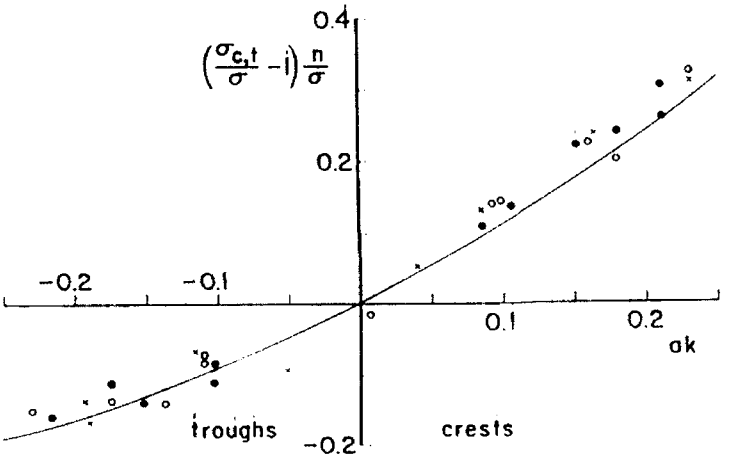


Figure 9. Apparent frequencies of wind waves at crests and troughs of the swell, with slope (ak) and frequency n. The solid line is calculated from the kinematic conservation equation.

Figure 8 represents the time histories at fixed points of the mean square surface displacement of the short waves in a representative sample of the experiments, each being averaged over the 100 realizations. It is clear that in each case, the wind waves are suppressed immediately during the passage of the long wave group, the event being marked by large spikes on the left of the figures resulting from 'phase jitter' of the long waves among the different realizations. This suggests that in these experiments, enhanced breaking of the wind-generated waves is indeed the process responsible for suppression, since it provides a rapid energy loss mechanism. This sugggestion is reinforced by measurements of the short waves during the passage of the group. A short wave energy packet is stretched and compressed by the orbital velocities of the longer waves, the magnitude increasing in the first half of the overtaking

long wave group as successive wave crests are higher and higher. Its apparent frequency thus oscillates with increasing amplitude and Figure 9 shows the (scaled) measured frequency variation as a function of long wave slope. The line shows what one would expect from wave kinematics; the observed frequency is a little higher than the line for high crests, possibly because of amplified wind drift which is neglected in the calculation.

The mean square surface displacement of the wind waves at the swell crests, relative to that before the arrival of the swell group, is shown in Figure 10. In the absence of wave breaking, this would *increase* with ak because of the convergence of flow at the crests of the long waves combined with the decrease in local apparent gravity. Clearly, in the experiments $\overline{\zeta'^2}$ decreases as successive wave crests become higher. Interestingly, though, the 'significant slope' $(\overline{\zeta'^2})^{1/2}/\lambda_d$, where λ_d is the dominant wavelength, remains sensibly constant as Figure 11 shows - the wavelength at the crest of the swell varies as (1 - ak) - indicating that as they are deformed by the swell, the short waves retain a level of geometrical similarity, as one might expect of intermittently breaking waves.

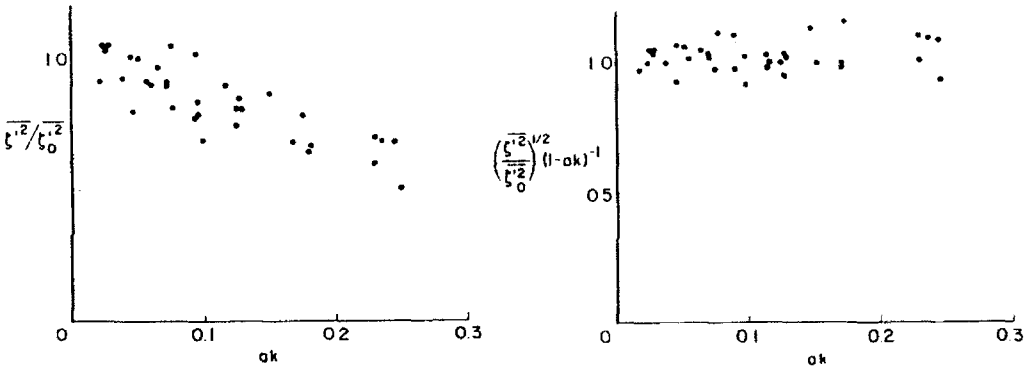


Figure 10. (left) Mean square surface displacement of the wind waves at the swell crests (referred to undisturbed values) as a function of swell slope ak.

Figure 11. (right) The corresponding significant slopes.

After the highest swell crest has swept through the short waves, breaking appears largely to cease on the successively smaller crests that follow. When the group has gone by, the water surface appears much smoother than before even to the eye. This is evident also in measurements of the mean square slope shown in Figure 12. At the higher slopes of the swell, the wind wave slope is reduced enormously; the smaller scales in the wind wave field which contribute significantly to the slope seem to have in effect disappeared.

Following the suppression, the wind waves gradually regenerate, though at a rate μ apparently slower than that indicated by Plant's formula $\mu = 0.04(u_*/c)^2$. The best-fit coefficient from these experi-

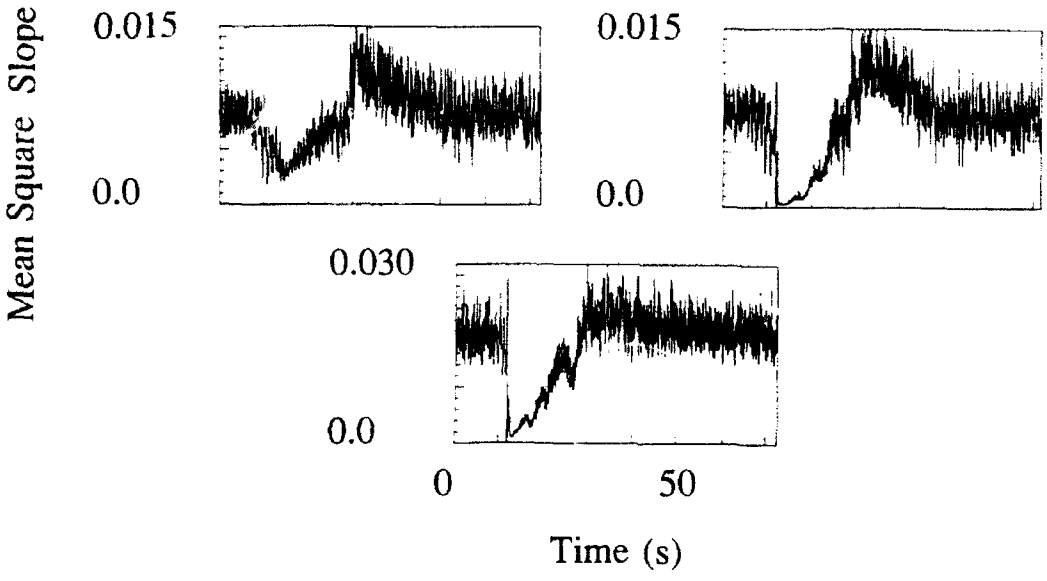


Figure 12. Mean square slope at a fetch of 5.9m as a function of time, measured using two probes 1 cm apart. left panel: $u_* = 15 cms^{-1}$, ak = 0.1; right: $u_* = 15 cms^{-1}$, ak = 0.26; bottom: $u_* = 27 cms^{-1}$, ak = 0.26.

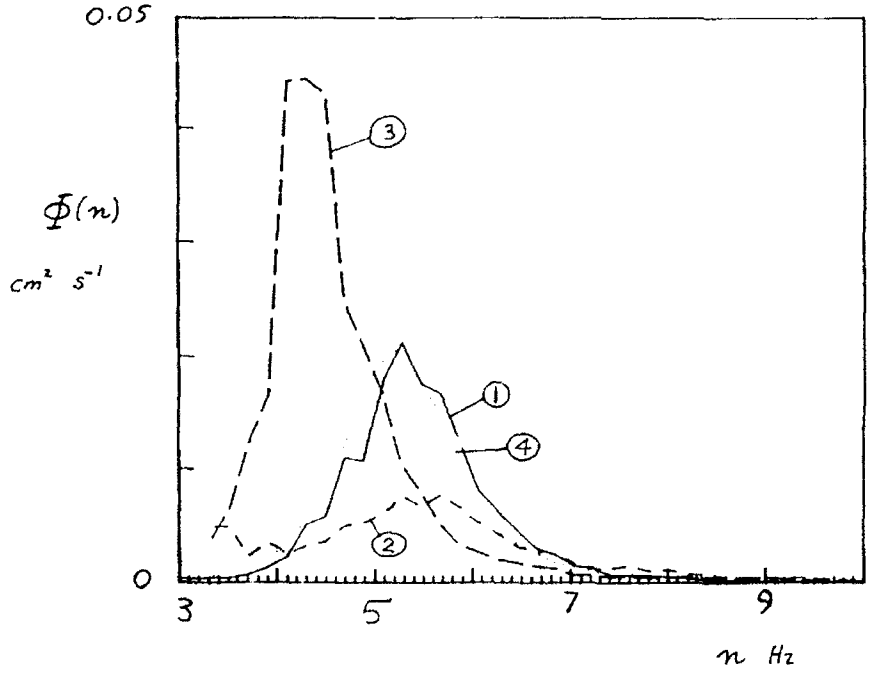


Figure 13. Local spectra at different phases of the interaction. (1) initially undisturbed, (2) after passage of the long wave group, (3) in the wave energy front and (4) after the recovery phase.

ments is 5.2×10^{-3}, being smaller possibly because of the local smoothness of the water surface under these conditions.

This recovery is interrupted, particularly at small fetches, by the sudden arrival of a front of waves longer and higher than the original undisturbed wind waves. As Figure 7 shows, this is a consistent and reproducible occurrence. Figure 13 shows local spectra at various points in the interaction - one is the initial field, two the waves after suppression (note that the dominant frequency is the same but the spectral density is reduced over the whole range), three is the spectrum in the wave energy front with a higher energy density and lower frequency than the initial field, four is the spectrum at the end of the window, when the waves have apparently returned to their initial state. When we calculate the significant slope of the waves in the energy front, we find that this again is the same as in the undisturbed field - they are higher because they are longer. We do not altogether understand yet the detailed dynamics of this wave energy front. They are clearly very different from the sharp front in the frequency spectrum of wind waves. The wave frequency is found to gradually decrease as the front propagates down the tank, and I think its evolution involves leakage of low frequency components ahead of the front by dispersion, which are then amplified by the wind until they reach saturation.

Now that I have come to things that I do not understand, I think I should stop. I am happy to acknowledge the support of the U.S. Office of Naval Research for these projects and the grants N00014-90-J-1623 and 1482.

References

1. Boccotti, P. 1981 On the highest waves in a stationary Gaussian process. *Att Acc. Lig.*, **38**.

2. Boccotti, P. 1989 On mechanics of irregular gravity waves. *Atti Acc. Naz. Lincei, Mem* **19**, 109-170.

3. Chu, J.S., Long, S.R. and Phillips, O.M. 1992 Measurements on the interaction of wave groups with shorter wind-generated waves. *J. Fluid Mech.*, (in press).

4. Donelan, M.A. 1987 The effect of swell on the growth of wind waves. Johns Hopkins APL: *Technical Digest* **8**, 18-23.

5. Duncan, J.H. 1981 An experimental investigation of breaking waves produced by a towed hydrofoil. *Proc. R. Soc. London A* **377**, 331-348.

6. Melville, W.K. and Rapp, R.J. 1985 Momentum flux in breaking waves. *Nature* **336**, 54-59.

7. Mitsuyasu, H. 1966 Interaction between water waves and wind I *Rep. Res. Inst. Appl. Mech.*, Kyushu Univ. **14**, 57-88.

8. Phillips, O.M. and Banner, M.L. 1974 Wave breaking in the presence of wind drift and swell. *J. Fluid Mech.*, **66**, 625-640.

9. Phillips, O.M., Gu, D. and Donelan, M. 1992 On the expected structure of extreme waves in a Gaussian sea. I. Theory and SWADE buoy measurements. *J. Phys. Oceanogr.* (in press).

10. Phillips, O.M., Gu, D. and Walsh, E. 1992 On the expected structure of extreme waves in a Gaussian sea. II. SWADE surface contour radar measurements. *J. Phys. Oceanogr.* (submitted).

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Effect of wind and water shear on wave instabilities

P. G. Saffman

Caltech, Pasadena, CA 91125

Abstract

The effect of vorticity or shear on the inviscid evolution and instability of surface waves is discussed. The instability of a flat surface is considered in detail. For the case of shear in the water, the instability is examined by means of Howard's semi-circle theorem, which is shown to elucidate the qualitative nature of the instability domain. This can then be determined numerically without recourse to analysis of the critical layer. The dependence of growth rate on the air profile for the case of wind blowing over water at rest is similarly considered. For bounded laminar profiles, the results are found to be sensitive to the profile shape, and Miles's approximate formula is found to be inaccurate. On the other hand, for unbounded logarithmic profiles, the growth rates are relatively insensitive to the profile shape. However, there is still disagreement between numerical calculations and Miles's formula, which overestimates the growth rate by a factor of two. A simple model is presented for the case of shear in both water and air.

1. INTRODUCTION

There are two problem areas in which wind and water shear, or equivalently vorticity in the air and the water, affect the instabilities of surface waves at an air-water interface. First, there is the question of the instability of existing waves, generated mechanically for example, or by earlier winds. In this situation, the dynamics of the air is unimportant and the hydrodynamics of the water motion is the question. There are then three categories of wave instability, as follows.

(i) Two- and three-dimensional long wave envelope modulations or side band instabilities:
These go generally under the name of the Benjamin and Feir [1] instability. Instabilities of this kind are also described by Lighthill [2] (who used Whitham's variational principle), Zakharov [3], [4] and others. These instabilities have non-water wave interest, because they exhibit the phenomenon of recurrence, and also because the envelope modulation while weakly nonlinear is described by the cubic Schrödinger equation.

(ii) Resonant interactions, Phillips [5]:
This describes the transfer of energy between wave trains containing waves of wavenumbers k_i and frequencies ω_i when $\sum k_i = 0$ and $\sum \omega_i = 0$.

(iii) Two- and three-dimensional, sub- and super-harmonic instabilities of finite amplitude waves of permanent form:
The two-dimension case was studied by Longuet-Higgins [6], [7]. Saffman [8] and MacKay and Saffman [9] applied Hamiltonian methods and the concept of signature to explain qualitatively some of the results, such as the change of stability when the wave energy as a function of wave height is stationary or modes collide. The three-dimensional case was studied by McLean, Ma, Martin, Saffman and Yuen [10], see also McLean [11], We shall refer to this as the MMMSY instability.

The phenomena in (i) and (iii) can, of course, be interpreted in terms of resonant interactions, but it is useful to categorize them separately since small disturbances grow at exponential rates rather than the linear rate associated with the resonances of (ii).

The resonances and envelope modulations are basically slowly varying, weakly nonlinear phenomena, and the modifications due to shear will arise primarily through changes in the dispersion equations and the speed of weakly nonlinear, uniform wave trains (Stokes waves). For example, Li, Hui and Donelan [12] have calculated the effect of constant water shear on the side-band instability for the case of infinite depth, and found that small shear tends to enhance instability whereas large shear tends to suppress it.

The effect of shear on the MMMSY, basically short wave instabilities, is a more difficult problem and will require massive computations in general, as potential theory cannot be used to relate perturbations of the flow to perturbations of the surface. Also, there is no scalar stream-function for three-dimensional long wave disturbances. However, Benney and Chow [13] have used perturbation methods for weakly nonlinear waves and found that weak shear generates relatively strong three-dimensional long wave instabilities. The properties of the instability are significantly different from those of the Benjamin and Feir type. For finite ampliftude waves and zero surface tension, Okamura and Oikawa [14] have calculated the three-dimensional instability of waves of permanent form which were calculated by Teles da Silva and Peregrine [15]. The disturbance $\eta'(x, z, t)$ of the free surface has the form

$$\eta' \propto e^{2\pi i(px+qz)}\, e^{\sigma t} \sum_{-\infty}^{\infty} \eta_n\, e^{2\pi i n x}. \tag{1}$$

They recover the finite depth MMMSY results (McLean [16]) for class I (which includes the Benjamin and Feir instability as the subharmonic wave number p tends to zero) and class II (principally period doubling) disturbances, as the shear tends to zero. They also see the Benney-Chow long oblique waves, which are absent when there is no shear.

The second problem area is concerned with wave generation by wind or current. In particular, we can consider the generation of waves as a problem of the hydrodynamic instability of a uniform state. For this phenomena, there are three main categories.

(i) Kelvin-Helmholtz (KH) instability:
This is the classic instability of a vortex sheet, modified by a density difference between the fluids on the two sides of the sheet.

(ii) Miles [17] instability:
This is the instability of a shear flow in the air over water initially at rest which can input energy into waves whose phase speed lies between the minimum and maximum of the wind velocity.

(iii) Drift layer instability:
This is the hydrodynamic instability of a current with shear in the water, which could be produced, for example, by the drag of the air over the water or mechanically by running a stream over the surface, as for example in an estuary. It was investigated by Esch [18] for gravity waves on piecewise linear profiles, by Stern and Adam [19] for capillary-gravity waves on piecewise linear profiles and by Morland, Saffman and Yuen [20] for capillary-gravity waves on smooth profiles.

This review will concentrate on these three mechanisms of wave generation. Note that the discussion is limited to inviscid instability mechanisms, and the direct effect of viscosity on the wave evolution is neglected. (Viscosity may, of course, play an important role in the generation of the shear profiles, but this is assumed to be on a much longer time scale than that of the instability.) Viscosity is sometimes employed to circumvent analytical difficulties associated with a critical layer; this introduces the Orr-Sommerfeld equation (e.g. Valenzuela [21] and subsequent studies of coupled air-water shear flows). But a consistent purely inviscid treatment is possible and the critical layer can be ignored except that it has an indirect effect in requiring fine resolution in the numerical integrations when the growth rate is small.

The further mechanism (Phillips [22]), in which waves are generated by random pressure fluctuations due to a turbulent wind and the main effect of shear is the modification of the dispersion relation, will not be considered here.

2. GENERATION OF WAVES BY WATER SHEAR

We consider the instability of an incompressible, infinitely deep, steady uniform flow $u(x,y,t) = U(y)$, $v(x,y,t) = 0$. Here, x is the horizontal axis in the direction of the flow, y is the vertical coordinate (increasing upwards), and u and v are the velocity components. The undisturbed free surface is taken as $y = 0$. The air motion is neglected. It is sufficient to consider two-dimensional motion, as the inviscid form of Squire's theorem holds. That is, for every unstable three-dimensional disturbance there is a corresponding two-dimensional disturbance with larger growth rate (Yih [23]). Introducing a stream function, we consider disturbances of the form

$$u'(x,y,t) = i\phi_y(y)/k\, e^{ik(x-ct)}, \quad v'(x,y,t) = \phi(y)\, e^{ik(x-ct)}. \tag{2}$$

It follows from the Euler equations, and the boundary conditions that the pressure is constant on the free surface and the disturbance vanishes at infinite depth, that ϕ

satisfies

$$\phi_{yy} - k^2\phi - \frac{U_{yy}\phi}{U-c} = 0, \tag{3}$$

with boundary conditions

$$(U-c)^2\phi_y - (U_y(U-c) + g + Tk^2)\,\phi = 0 \qquad \text{on } y = 0, \tag{4}$$

and

$$\phi \to 0 \quad \text{as } y \to -\infty. \tag{5}$$

Here, g is gravity, T is surface tension and k is the wavenumber of the disturbance. This is an eigenvalue problem for the (possibly complex) disturbance wave speed c. If $c_i = \Im c \neq 0$, then the disturbance grows at an exponential rate and the flow is unstable. (The roots come in complex conjugate pairs so it is sufficient to find a non-zero c_i.) For a given shear profile, with $U_d = U(0)$ giving a characteristic speed, and

$$\Delta = \frac{2}{U_d}\int_{-\infty}^{0} U(y)\,dy \tag{6}$$

a characteristic thickness of the drift layer, it follows that

$$c = \text{fn}\ (U_d, \Delta, g, T, k). \tag{7}$$

The mathematical task is to determine the values of the parameters which make c complex.

We know that there are two kinds of eigenfunction and eigenvalue. There are the proper ones, which are associated with a discrete spectrum. The eigenvalues are isolated, and the eigenfunctions are non-singular. In the pesent case, the proper or discrete spectrum contains values of c which are either complex, or real with $c < 0$ or $c > U_d$. The improper eigenvalues are values of c for which the Rayleigh equation is singular. These have $c_i = 0$ and $0 < c < U_d$. The spectrum is continuous and and the eigenfunctions are singular. The continuous spectrum is needed for the initial value problem or analysis of the Rayleigh equation, regarded as the small viscosity limit of the Orr-Sommerfeld equation, when the proper spectrum is not complete. This is the case here, where there are only two proper eigenvalues. But the continuous spectrum usually contributes a disturbance which decays algebraically, and is not needed for considerations of stability. (An exception is a vortex flow studied by Smith and Rosenbluth [24] for which the continuous spectrum grows algebraically.) We therefore restrict attention to the discrete spectrum.

Rayleigh's criterion (see, e.g. Drazin and Reid [25] asserts that a necessary condition for instability is the existence of a point of inflection in the velocity profile, i.e. there exists y_c in $-\infty < y_c < 0$ such that $U_{yy}(y_c) = 0$. However, on repeating the details of the argument, it is found that Rayleigh's criterion does not apply to this problem owing to the different boundary condition at the free surface.

A more powerful result for qualitative analysis of the proper eigenvalues is Howard's [26] semi-circle theorem, which was shown by Yih [23] to be still valid for the present problem. According to this result, the unstable proper roots of the eigenvalue problem lie in a semi-circle in the $c_r - c_i$ plane whose diameter lies on the real axis from 0 to U_d. We can now describe the behaviour of the eigenvalues as $\lambda = 2\pi/k$ varies from ∞ to 0.

When $U_d = 0$ and there is no shear in the water, the dispersion relation is $c = \pm c_o$ where

$$c_o = c_m \sqrt{\frac{\lambda}{2\lambda_m} + \frac{\lambda_m}{2\lambda}}. \tag{8}$$

Here $\lambda_m = 2\pi(T/g)^{1/2}$ and $c_m = (4gT)^{1/4}$ are the wavelength and phase speed of the slowest capillary-gravity wave.

Now introduce shear and vary λ keeping all other parameters constant. When λ is very large, the effect of shear is negligible, being confined to a relatively thin layer, and the speeds are given by (8). There are thus two modes, one co-flowing (i.e. with velocity in the direction of the shear current), and the other counter-flowing. As the wavelength decreases, the magnitudes of the speeds decrease.

Consider first the co-flowing wave. Clearly its speed c will be greater than U_d, as shear convects the wave. Also, as $\lambda \to 0$, the speed goes to infinity as the effect of capillarity becomes dominant. Therefore there must be a value of λ such that the speed is a minimum, and this minimum speed is greater than U_d. Hence, the co-flowing wave never enters the Howard semi-circle, and this mode is stable. As the Rayleigh equation never becomes singular, the mode is unique and there are no other members of the discrete spectrum with $c > U_d$.

The situation for the counter-flowing wave is fundamentally different. The entrainment effect of the shear now increases the wave speed, measured algebraically. But when λ is small, the wave speed goes back to $-\infty$, and hence there is a value of λ, λ_c say, for which the wave speed is a maximum, denoted by V, say. There are now two possibilities. First, the entrainment of the shear is smaller than c_m. Then $V < 0$ and c is always real and negative, and the mode is stable. Second, the entrainment is sufficiently large that $V > 0$. Then the semi-circle is entered and complex roots appear. However, as λ continues to decrease, the capillary effect eventually dominates the entrainment and the semi-circle is exited, the mode changing back from unstable to stable. Thus, there will be a range of wavelengths which are unstable, the limiting values being neutrally stable as the semi-circle is entered and left at $c = 0$. The deciding factors are the values of the ratios U_d/c_m and Δ/λ_m.

We now give arguments to show that U_d/c_m is a necessary condition for instability, i.e. the waves are always stable when this condition is violated. Further, when this condition is not satisfied, unstable modes exist only when Δ exceeds a critical value which depends on U_D/c_m.

The transition wave numbers for which the counter-flowing wave speed enters or leaves the semi-circle are the eigenvalues k of the equation

$$\phi_{yy} + \frac{U_{yy}}{U}\phi = k^2\phi, \qquad -\infty < y < 0, \tag{9}$$

with boundary conditions

$$U_d^2\phi_y - (U_d U_{d,y} + g + k^2 T)\,\phi = 0 \qquad \text{on } y = 0 \tag{10}$$

and (5). These are obtained by putting $c = 0$ in the Rayleigh equation (3) and its boundary condition (4). By inspection, (9) and (10) can be scaled for a given profile in terms of c_o/U_d and λ/Δ. Then in a $\lambda/\Delta - c_o/U_d$ plane, the curve on which $c = 0$ separates the plane into an upper stable region, in which $c_r < 0$ and $c_i = 0$, and a lower unstable region in which $0 < c_r < U_d$ and $c_i \neq 0$. The separating curve passes through the point $c_o/U_d = 1$ when $\lambda/\Delta = 0$. Now c_o and λ are not independent, being coupled by the dispersion equation (8), which is fixed when c_m/U_d and λ_m/Δ are known and can be written

$$\frac{c_o}{U_d} = \frac{c_m}{U_d}\sqrt{\frac{\lambda}{\Delta}\frac{\Delta}{2\lambda_m} + \frac{\Delta}{\lambda}\frac{\lambda_m}{2\Delta}}. \tag{11}$$

If $c_m/U_d < 1$, the dispersion relation intersects the stability boundary when Δ is sufficiently large, because the minimum of the scaled dispersion relation (11) moves to the left as Δ increases.

For the case of an exponential profile, $U(y) = U_d\, e^{2y/\Delta}$, the $c = 0$ neutral curve can be determined analytically, and is

$$\frac{\lambda}{\Delta} = \frac{1}{2}\pi\left(\left(\frac{U_d}{c_o}\right)^2 - \left(\frac{c_o}{U_d}\right)^2\right). \tag{12}.$$

To obtain growth rates, numerical methods must be resorted to in general. However, this is fairly straightforward and care is required only when c_i is very small. For instance, the Rayleigh equation (3) can be regarded as an equation for complex k as a function of c. The value of c_r is fixed and c_i is varied until the solution of a centred finite difference approximation to the differential equation and boundary conditions gives k real. Continuation methods can be employed to find further solutions. The critical layer is unimportant in the calculation, except for the need to employ a fine mesh near the value of y for which $U(y) = c_r$ when c_i is particularly small.

When $T = 0$, the locus in the semi-circle is qualitatively different. In this case, the co-flowing waves are again always stable, but their speed tends to U_d as $\lambda \to 0$.The counter-flowing waves are stable until their wavelength decreases to a value λ_c at which they are stationary. For $\lambda < \lambda_c$, the entrainment by the shear makes the speed positive, and the waves are unstable until $\lambda = 0$, when c_r has reached the value U_d. For the exponential profile,

$$\frac{\lambda_c^2}{\Delta^2} = \frac{\pi^2 U_d^2}{\Delta g}\left(1 + \frac{g\Delta}{4U_d^2}\right)^{-1}. \tag{13}$$

The mechanism for the generation of waves by drift layer instability is qualitatively attractive. It provides a delay mechanism for the wave generation after a wind starts blowing, because the instability mechanism will not work until the layer thickness Δ

reaches a critical value. But quantitative predictions are very poor. The necessary condition $U_d > c_m$ is not satisfied by the wind drift layer, and the predicted phase speeds are of order $\frac{1}{2} c_m$ which is much too fast. On the other hand, the theory may be applicable to estuary discharge (J. Jimenez, private communication) or tidal flow (D.W. Moore, private communication) in the presence of wind. An onshore wind, for example, would tend to amplify counter flowing waves on an estuary current or ebbing tide which would be unstable and produce choppiness, whereas waves would not be supported with an onshore wind and a rising tide and the surface would remain relatively smooth.

An interesting question, currently under study, is the behaviour of the instability in the presence of a stable long wave. This is the MMMSY problem, for the case in which a superharmonic disturbance is unstable. For capillary gravity waves, it is possible that there might be enhanced instability because water shear causes extra corrugation of the interfaces (Milinazzo and Saffman [27]).

3. EFFECT OF WIND SHEAR ON WAVE GENERATION

We now consider the case of shear in the air and not in the water. We neglect surface tension, and again suppose the flow is laminar and inviscid. In this case, the eigenfunction $\phi(y)$ for two-dimensional waves of wavenumber k satisfies the Rayleigh equation (3) for $0 < y < \infty$, where $U(y)$ now denotes the shear profile in the air. In the water, $-\infty < y < 0$, the Rayleigh equation reduces to

$$\phi_{yy} - k^2 \phi = 0. \tag{14}$$

Further, $\phi \to 0$ as $y \to \pm\infty$, and at the interface $y = 0$,

$$[\rho\{(U-c)^2 - (U_y\,(U-c) + g)\,\phi\}]^{air}_{water} = 0. \tag{15}$$

These equations again constitute an eigenvalue problem for the wave speed c.

Miles [17] and Lighthill [28] demonstrate the existence of growth by a quasi-steady analysis of the critical layer for waves whose wavenumber is such that $0 < c_w < U_\infty$, where $c_w = \sqrt{g/k}$. they give the formulae

$$c_i = \frac{r\beta}{2c_w}, \qquad \beta = -\pi \frac{w_c''}{w_c'^3} \left(\int_{\xi_c}^{\infty} e^{-\xi}\, w^2\, d\xi \right)^2, \tag{16}$$

where $r = \rho_a/\rho_w$, $\xi = ky$, $w = U(y) - c_w$, and a prime denotes differentiation with respect to ξ. A subscript c indicates evaluation at the critical layer $\xi_c = ky_c$, which is defined by $U(y_c) = c_w$. If c_w does not lie between the maximum and minimum values of $U(y)$, then β is defined to be zero. We shall refer to (16) as the Miles formula. In the case of exponential profiles in the air,

$$U = U_\infty\,(1 - e^{-2y/\Delta}), \tag{17}$$

the expression for β can be evaluated in closed form to yield

$$kc_i = \frac{16\,\pi r k^2 \Delta U_\infty^2}{c_w (2+k\Delta)^2 (4+k\Delta)^2} \left(1 - \frac{c_w}{U_\infty}\right)^{2+k\Delta}. \tag{18}$$

It can be shown that Squire's theorem still holds (see Morland and Saffman [29]), and it is therefore sufficient to consider two-dimensional disturbances.

As for the case of shear in the water, Howard's semi-circle theorem still holds and combined with straightforward numerical methods eliminates the need for a critical layer analysis. For a bounded flow (i.e. $U_\infty < \infty$) such as the exponential profile (17), or an error function profile $U = U_\infty \operatorname{erf}(2y/\sqrt{\pi}\Delta)$, there are two modes for infinite wavelength. As λ decreases, the speed of the co-flowing wave decreases until at a critical wavelength λ_c, the wave enters the semi-circle at $c_r = U_\infty$. As λ now tends to zero, the locus is inside the semi-circle and the co-flowing wave is unstable for $0 < \lambda < \lambda_c$. On the other hand, the speed of the counter-flowing wave is always negative and approaches zero as $\lambda \to 0$. These waves are therefore always stable. Outside the semi-circle, the phase speed is given approximately by $c_r = \pm c_w(\lambda_c) \approx U_\infty$; which is to be expected as the air is just a small perturbation of the water motion.

The details of the complex wave speeds in the semi-circle depend on the values of r ($\approx 10^{-3}$ for the air-water interface) and $\Delta g/U_\infty^2$, and on the profile. If $\Delta g/U_\infty^2 > r$, then $c_r \approx c_w(\lambda)$ inside the semi-circle. For $\Delta g/U_\infty^2 = 0.1, kc_i \sim 5 \times 10^{-3} g/U_\infty$. For $\Delta g/U_\infty^2 = 0.01$, $kc_i \sim 0.2 g/U_\infty$. The growth rates for these cases given by the Miles formula (16) are different by a factor about 2 or 1/2 depending on the profile.

If $\Delta = 0$, the shear in the air condenses to a vortex sheet, and the instability mechanism is the classic KH instability (see Lamb [30]). The condition for the instability of the smooth profile to be of this class is $\Delta g/U_\infty^2 \ll 2\pi r$. The complex wave speed for the KH instability is

$$c = rU_\infty \pm \sqrt{\lambda g/2\pi - rU_\infty^2}\,. \tag{19}$$

These values are retrieved by the numerical computations when $\Delta g/U_\infty^2 \to 0$ with $\lambda \gg \Delta$. In this limit, the approximation $c_r \sim c_w$ becomes invalid for the unstable modes since from (19) their phase velocity is approximately constant with value rU_∞. The KH instability becomes significant for smooth profiles when Δ is small enough that waves long compared to Δ are KH unstable.

Unbounded profiles, e.g. lin-log profiles in which $U \propto y$ for $y < y_*$ and $U \propto \log y$ for $y > y_*$, where y_* is the width of the viscous sub-layer, are of interest as models of turbulent flow. In this case, the semi-circle occupies the whole positive c_r axis, and all wavelengths are unstable. However, the numerics is still straightforward. It is found that the numerical results agree qualitatively with those calculated from the Miles formula (16) provided $\lambda \gg u_*^2/g$, where u_* is the friction velocity. This condition ensures that the critical layer lies outside the viscous sublayer so that the Miles formula is applicable. For further details, see Morland and Saffman [28]. However, there is a significant quantitative difference, with the Miles formula (16) overestimating the growth rate obtained from numerical solution by a factor of two.

When there is shear in both wind and water, the problem becomes more difficult. A preliminary analysis using a piecewise linear profile indicates that new phenomena may be present. The advantage of such a profile is that the dispersion relation is obtained from solving a quartic, and the dependence on the various parameters is more easily examined than is the case for smooth profiles. But there is still the difficulty of making sense of the results when there are many parameters in the problem. The profile used

is

$$U(y) = \begin{cases} 0 & \text{if } y < -\Delta_w; \\ V_w + V_w y/\Delta_w & \text{if } -\Delta_w < y < 0; \\ V_w + V_a y/\Delta_a & \text{if } 0 < y < \Delta_a; \\ V_w + V_a & \text{if } y > \Delta_a. \end{cases} \tag{20}$$

The main conclusion from an initial study is that water shear tends to stabilize the Miles type instability, and extends the boundary of the region where KH instability is seen. Also, there is a range of windspeeds for which waves with a wavelength in a range do not grow, i.e. different wave components behave differently under a given wind in the presence of water shear. For further details, see Caponi *et al* [31].

4. REFERENCES

1 T.B. Benjamin and J.E. Feir, J. Fluid Mech., 27 (1970) 417.
2 M.J. Lighthill, J. Inst. Math. Appl., 1 (1965) 269.
3 V.E. Zakharov, Sov. Phys. J.E.T.P., 51 (1966) 1107.
4 V.E. Zakharov, J. Appl. Mech. Tech. Phys., 2 (1968) 190.
5 O.M. Phillips, J. Fluid Mech., 9 (1960) 193.
6 M.S. Longuet-Higgins, Proc. Roy. Soc., A360 (1978) 471.
7 M.S. Longuet-Higgins, Proc. Roy. Soc., A360 (1978) 489.
8 P.G. Saffman, J. Fluid Mech., 159 (1985) 169.
9 R.S. MacKay and P.G. Saffman, Proc. Roy. Soc., A406 (1986) 115.
10 J.W. McLean, Y.C. Ma, D.U. Martin, P.G. Saffman and H.C. Yuen, 1981 Phys. Rev. Lett., 46 (1981) 817.
11 J.W. McLean, J. Fluid Mech., 114 (1982) 315.
12 J.C. Li, W.H. Hui and M.A. Donelan, M.A., Nonlinear water waves (Eds K. Horikawa and H. Maruo) Springer-Verlag, 1988.
13 D.J. Benney and K. Chow, Stud. App. Math., 74 (1986) 227.
14 M. Okamura and M. Oikawa, J. Phys. Soc. Japan, 58 (1989) 2386.
15 A.F. Teles da Silva and D.H. Peregrine, J. Fluid Mech., 195 (1988) 381.
16 J.W. McLean, J. Fluid Mech., 114 (1982) 331.
17 J.W. Miles, J. Fluid Mech., 3 (1957) 185.
18 R.E. Esch, J. Fluid Mech., 12 (1962) 192.
19 M.E. Stern and Y.A. Adam, Mem. Soc. Roy. Sci. Liege. Ser 6, 6 (1973) 179.
20 L.C. Morland, P.G. Saffman and H.C. Yuen, Proc. Roy. Soc., A433 (1990) 441-450.
21 G.R. Valenzuela, J. Fluid Mech., 76 (1976) 229.
22 O.M. Phillips, J. Fluid Mech., 2 (1957) 417.
23 C.S. Yih, J. Fluid Mech., 51 (1972) 209.
24 R.A. Smith and M.N. Rosenbluth, Phys. Rev. Lett., 64 (1990) 649.
25 P.G. Drazin and W.H. Reid, Hydrodynamic Stability, Cambridge Univ. Press, 1981.
26 L.N. Howard, J. Fluid. Mech., 10 (1961) 509.
27 F.A. Milinazzo and P.G. Saffman, J. Fluid Mech. 216 (1990) 93.
28 M.J. Lighthill, J. Fluid Mech., 14 (1962) 385.
29 L.C. Morland and P.G. Saffman, submitted J. Fluid Mech.

30 H. Lamb, Hydrodynamics (6th ed.), Cambridge Univ. Press.
31 E.A. Caponi, M.Z. Caponi, P. G. Saffman and H.C. Yuen, Proc. Roy. Soc. A438 (1992) 95.

Acknowledgment

I wish to acknowledge support by the Office of Naval Research, Grant No. N00014-92-J-1435. My work was carried out in collaboration with Dr. E. Caponi, Dr. M. Caponi, Dr. L. Morland and Dr. H. Yuen.

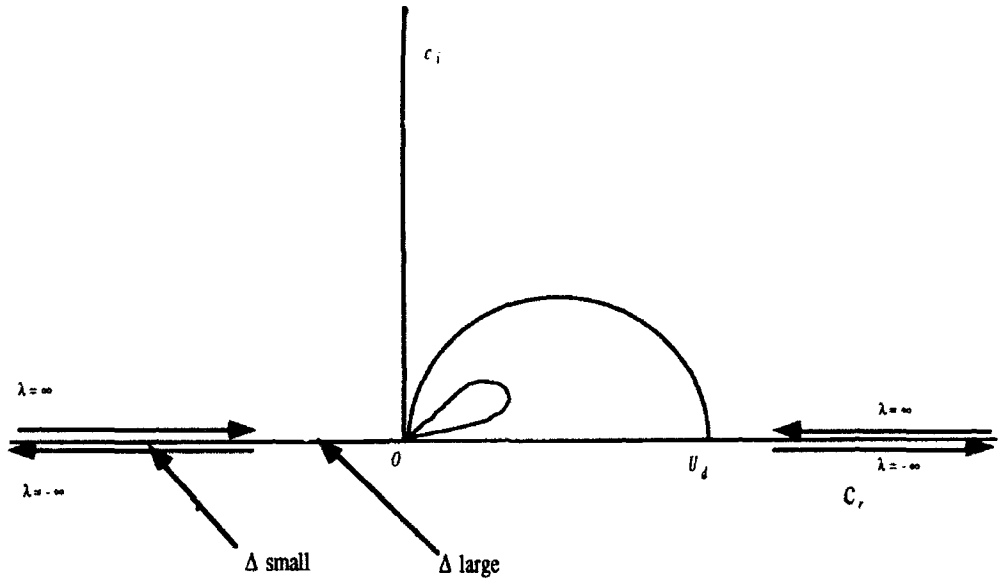


Figure 1. Howard semi-circle for drift wave instability

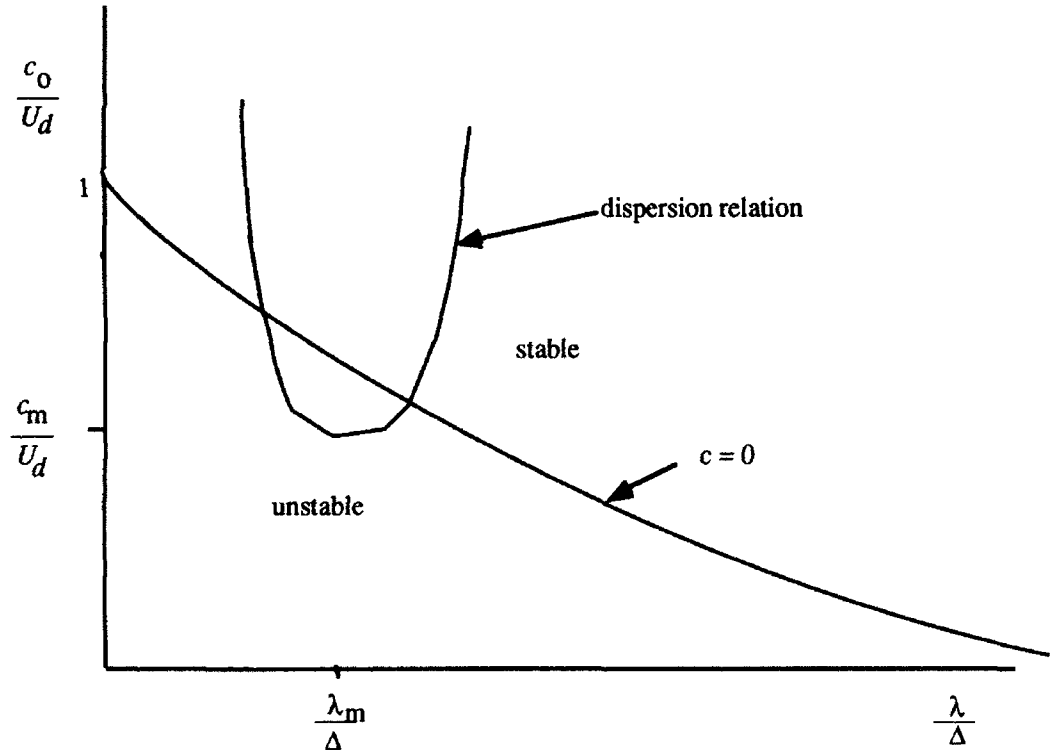


Figure 2. Stability boundary for drift wave instability

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.

1993 IUTAM.

Nature's Structural Engineering of Bone on a Daily Basis

Stephen C. Cowin

Department of Mechanical Engineering, The School of Engineering of The City College and The Graduate School of The City University of New York, New York, N. Y. 10031

Abstract

Living tissue adapts to its environmental mechanical loading on a daily basis. The adaptation begins within seconds of a significant loading change and, while the change can take months in bone, it can happen in weeks in soft tissues. Wolff's law is one of the observations of functional adaptation of living tissue to mechanical loading. It is the observation that the axes of the individual bony struts in the cancellate (spongy or porous) bone coincide with the principal axes of stress induced by the external loads applied to the whole bone. After a brief historical introduction to the subject, two topics of current research concerning Wolff's law are described. The first topic is a phenomenological continuum model whose objective is to describe the temporal evolution of both the density changes and the reorientation of the anisotropic microarchitecture when the stress state applied to a bone is changed. The second topic is the modeling of a possible cellular mechanosensory transduction mechanism by which communicating bone cells (osteocytes) sense the very small in vivo strains in the bone and signal other bone cells to build bone (osteoblasts) or to remove bone (osteoclasts).

1. HISTORICAL BACKGROUND

The history of development of what is today called Wolff's law begins at a meeting of a Natural Science Society in Zurich held exactly 126 years ago, in August of 1866. In attendance were Professor H. Meyer, an anatomist, and Professor K. Culmann who was Professor of Engineering Science at the newly created Swiss Federal Technical Institute (ETH) in Zurich. Professor Culmann was a leading structural engineer and his book [1] on graphical statics, the then-current standard procedure for structural stress analysis, appeared that year. He is credited with the creation of what we now call the two-dimensional Mohr's circle (i.e., Mohr extended the Culmann circle to three dimensions). At the meeting Meyer made a presentation in which he displayed preparations and drawings of the trabecular architecture seen in a frontal section of the pelvic end of the human thigh bone as shown in Figure 1. Note that a strut of the cancellate (spongy or porous) bone is called a trabecula, from the Latin for "little beam." Culmann observed that the form of the trabeculae in the pelvic end of the thigh bone closely resembled the form of the stress flow lines (i.e., stress trajectories) in an object of the same shape as the bone and loaded in the same way. He made that observation based on his experience with a crane design that was similar in shape to the frontal section of a human

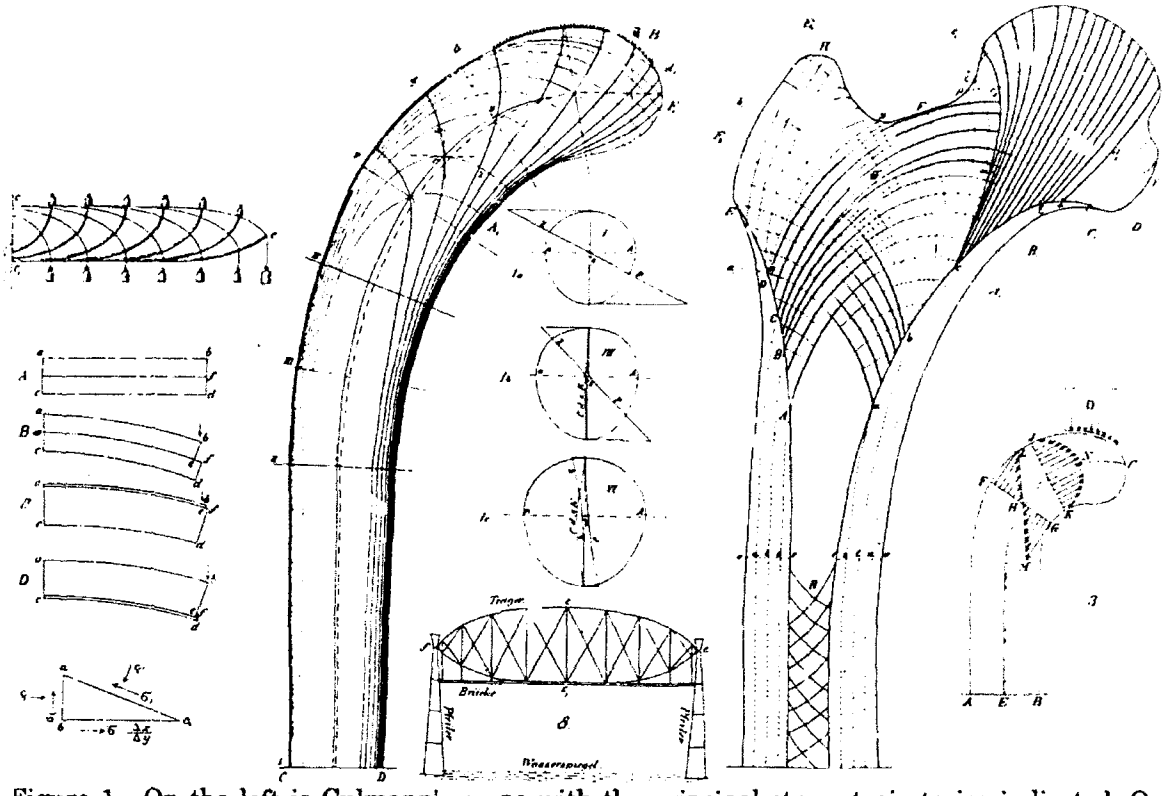


Figure 1. On the left is Culmann's crane with the principal stress trajectories indicated. On the right is von Meyer's sketch of the trabecular architecture in a section through the proximal end of the human femur. Both the femur and Culmann's crane are loaded transversely at their cantilevered ends as illustrated on the little insert at the lower far right. Taken from Wolff [2].

thigh. Figure 1, taken from an 1870 paper [2] of Wolff's, illustrates the relationship between trabecular architecture and the stress flow lines of Culmann's crane. The coincidence of trabecular directions with the directions of the stress flow lines is called Wolff's law. The interaction between Meyer and Culmann is considered by many to be the origin of what is perceived to be Wolff's law. The name of Julius Wolff, an anatomist and a contemporary of Meyer and Culmann, is associated with the observation because Wolff wrote voluminously about it, Wolff [3, 4]. The meeting of Meyer, the anatomist, and Culmann, the engineer, was an important historical event, the apparent genesis of the important biological concept of structural adaptation. Culmann's involvement with the ideas of structural adaptation in bone was a minor incident in a highly successful professional engineering career. Culmann did not publish any material on this topic, although he published a great deal on graphical statics and a very influential 1851 article on the latest advances in bridge, railway and river-boat construction in England and the United States, Charlton [5].

With the observation of Culmann as his basis, Wolff [2] in 1870 stated that, when the environmental loads on a bone are changed by trauma, pathology or change in life pattern, functional remodeling reorients the trabeculae so that they align with the new principal stress trajectories. Wolff never attempted to prove this assertion. In 1885, Roux [6], who broadly generalized the concept of functional adaptation, produced an analysis of the

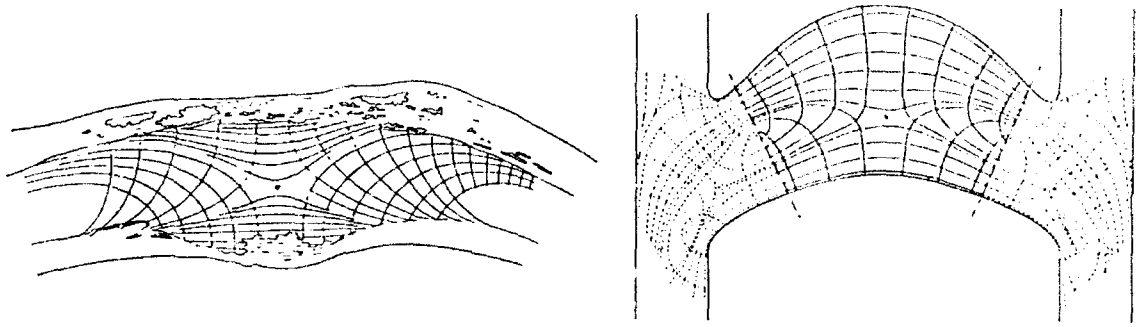


Figure 2. Left: Pauwels' [12] sketch of the cancellous architecture in the ankylosed knee he considered. Right: Pauwels' sketch of the principal stress trajectories in the photoelastic model he constructed. The two transverse dotted lines in the photoelastic model are the boundaries of the anatomically correct model of the knee.

structure of the cancellous tissue of an ankylosed knee as a proof of Wolff's law. The development of the theory of the trajectorial architecture of cancellous bone appears to have slowed after the 1885 book of Roux because the theory had developed as far as it could with the then current techniques for stress analysis.

In 1954 Pauwels [7] examined another ankylosed knee using two-dimensional photoelastic models to estimate the principal stress trajectories in the knee. Figure 2(a) shows Pauwels' sketch of the cancellous architecture of the ankylosed knee and Figure 2(b) shows the stress trajectories in his photoelastic model of the situation. Pauwels indicates that this analysis is a proof of Wolff's hypothesis.

An increased research interest in the trajectorial theory of bone architecture has occurred in recent years. This increased research interest is driven by two synergistic factors. First, there is the clinical need to understand the mechanical and remodeling behavior of bone tissue because the use of implanted bone prostheses has increased very greatly. Prostheses transfer joint forces to the bone cortex via the trabecular network. Second, scientific and engineering experimental techniques developed in the physical and biological sciences and in engineering in the last thirty years now supply the technology that blocked further development in the 1880's. For example, Hayes and Snyder [8] employed quantitative stereology techniques and a two-dimensional finite element mesh to study the trabecular architecture of the patella and concluded that the trabecular structure aligns itself with principal stress trajectories. Little et al. [9] reported a three-dimensional finite element analysis of the upper tibia and made drawings of the stress trajectory patterns. They concluded that these patterns are very similar to the orientations of the trabeculae in the section views of actual tibiae.

The underlying idea of the studies of Little et al. [9] and Hayes and Snyder [8] is the same as that that underlay the analysis of Roux and of Pauwels, namely, the verification of Wolff's trajectorial architecture hypothesis by comparison of the stress trajectories predicted by elastic stress analysis with the actual trabecular trajectories. Roux used rubber models in the shape of the ankylosed knee and noted the deformation patterns of inscribed lines on the surface of the rubber when it is loaded as a knee is loaded. Pauwels used the

photoelastic stress analysis. Unlike the experimental models of Roux and Pauwels, the finite element technique can be applied to the stress analysis of objects of any shape and of any degree of anisotropy and inhomogeneity, thus permitting cortical bone to be distinguished from cancellous bone, and cancellous bone tissue of different densities and different elastic constants to be distinguished in the model.

2. THE PHENOMENOLOGICAL CONTINUUM AND MECHANOSENSORY TRANSDUCTION MECHANISM MODELS

A phenomenological continuum model for the temporal evolution of both the density changes and the anisotropic microstructure of cancellous bone is described in the next three sections. In this model a quantitative stereological measure of the microstructural arrangement of trabeculae and pores of cancellous bone, called the fabric tensor, is related to the elastic constants of the tissue. This formulation of the elastic constitutive equation in terms of the fabric tensor leads to a mathematical statement of Wolff's law at remodeling equilibrium. The fabric tensor and its experimental determination are discussed in the following section. The relationship between the elastic constants and the fabric tensor in a local region of cancellous bone is described in a subsequent section. A mathematical formulation of Wolff's law of trabecular architecture at remodeling equilibrium is sketched in the fifth section.

The remainder of this paper concerns a cellular level mechanosensory transduction mechanism for Wolff's law. A new hypothesis is advanced for the cellular level feedback mechanism associated with Wolff's law. The hypothesis concerns the mechanism by which the osteocytes housed in the lacunae of mechanically loaded bone sense the mechanical load applied to the bone by the detection of cyclic dynamic strains. The model incorporating this hypothesis employs a hierarchical model of bone tissue structure as the medium for the conversion of cyclic mechanical loading applied to the whole bone to the fluid shear stress at the surface of the osteocytic cell process. The proposed mechanosensory model is described in section six and the hierarchical anatomical and permeability model levels are detailed in the next two sections. In the final section the predictions of the mechanosensory model and their significance are described.

3. FABRIC

It is recognized that porosity or solid volume fraction is the primary measure of local material structure in a porous material. Porosity does not, however, reflect any directionality of the specimen's microstructure. The fabric tensor is constructed to represent the orientation of the local material microstructure in many porous materials, [10-16]. The term fabric tensor is used here to indicate any positive definite second-rank tensor in three dimensions that characterizes the local anisotropy of the material's microstructure. The fabric tensor is a point property (even though its measurement requires a finite test volume) and is therefore considered to be a continuous function of position in the material model. Methods exist to measure fabric ellipsoids in cellular materials, cancellous bone, rocks and granular materials and are described in [10-16].

One rather general way to construct a fabric tensor for a material is from a set of N measurements of material microstructural features, each measurement characterized by a scalar m_n and a unit vector $\mathbf{c}^n$ ($\mathbf{c}^n \cdot \mathbf{c}^n = 1$), where n = 1, 2,...N. The normalized fabric tensor **H** is defined in terms of the N observations as,

$$\mathbf{H} \sum_{n=1}^{N} m_n = \sum_{n=1}^{N} m_n \mathbf{c}^n \otimes \mathbf{c}^n, \tag{1}$$

where the encircled cruciform indicates the open or tensor product of $\mathbf{c}^n$ with itself. Due to the normalization, tr**H** = 1, since $\mathbf{c}^n \cdot \mathbf{c}^n = 1$ for all N.

The experimental procedure for the measurement of cancellous bone fabric is described by Whitehouse [15], [16], Harrigan and Mann [12] and Turner *et al.* [13]. It consists of determining the mean distance, m_n, between the change of phases, bone to void and vice versa, along each test direction $\mathbf{c}^n$. The work of these authors has shown that the fabric tensor is a good measure of the structural anisotropy in cancellous bone tissue. The methodology of making measurements is easily adapted to an automated computational system as shown by Harrigan and Mann [12], Turner *et al.* [13] and Hodgskinson and Currey [14]. The determination of **H** for the cancellous bone from a horse is illustrated in Figure 3, which is taken from [14].

4. WOLFF'S LAW AT REMODELING EQUILIBRIUM

A continuum formulation of *Wolff's law at remodeling equilibrium (RE)* was suggested by Cowin [17]. This formulation is based on an elastic constitutive relation for cancellous bone in which the elastic coefficients depend on the solid volume fraction of bone as well as the fabric tensor. The constitutive equation for this continuum formulation is

$$\mathbf{T} = \mathbf{C}(\mathbf{H}, \nu)[\mathbf{E}] \quad , \quad \textbf{or} \; T_{ij} = C_{ijkm}(\mathbf{H}, \nu)E_{km} \quad , \tag{2}$$

where **T**, **H**, **E** are the stress, fabric and strain tensors whose Cartesian components are T_{ij}, H_{ij} and E_{ij}, respectively, $\mathbf{C}(\mathbf{H}, \nu)$ is the fourth rank elasticity tensor whose components are C_{ijkm}, and ν is the solid volume fraction of the trabecular tissue. The quantity $\mathbf{C}[\mathbf{E}]$ is the second rank tensor with components $C_{ijkm}E_{km}$. It is assumed that all the structural anisotropy of the material is characterized by the fabric tensor **H**; this assumption can be shown to require that the material have orthotropic symmetry, [19]. The absence of bone remodeling is remodeling equilibrium. Specifically, *remodeling equilibrium* (RE) is the set of conditions under which there is no realignment of trabecular architecture and no net deposition or resorption of cancellous bone tissue. RE is thus characterized by a particular equilibrium architecture, denoted by ν^* and $\mathbf{H}^*$, and a particular stress and strain state, denoted by $\mathbf{T}^*$ and $\mathbf{E}^*$, respectively. The RE stress $\mathbf{T}^*$ and strain $\mathbf{E}^*$ may actually be ranges of stress and strain, or some special measures of stress and strain history. This matter is discussed by Cowin [20]. It is reasonable to think of $\mathbf{T}^*$ and $\mathbf{E}^*$ as long time averages of the environmental stress and strain, although they may be other functions of the stress and strain history of the bone

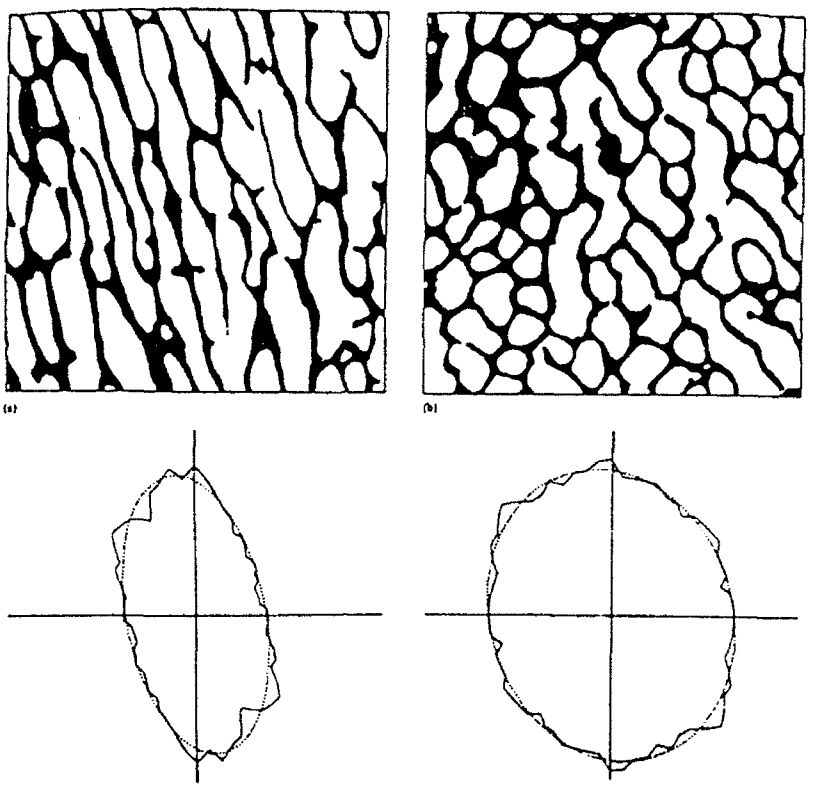


Figure 3. The fabric ellipses for cancellous bone; faces of two cubes of cancellous horse bone.

tissue.

Wolff's law of trabecular architecture is the observation that the principal stress axes coincide with the principal trabecular directions in cancellous bone in RE. This means that, in RE, the principal axes of stress $\mathbf{T}^*$ must coincide with the principal axes of fabric $\mathbf{H}^*$. This coincidence of principal axes is assured if the matrix multiplication of $\mathbf{T}^*$ and $\mathbf{H}^*$ is commutative, i.e., if

$$\mathbf{T}^*\mathbf{H}^* = \mathbf{H}^*\mathbf{T}^*. \tag{3}$$

This equation, which is a simple statement of the commutability of the matrix multiplication $\mathbf{T}^*$ and $\mathbf{H}^*$, can be viewed as an algebraic statement of the phenomenological Wolff's law. If the principal axes of $\mathbf{T}^*$ and $\mathbf{H}^*$ are coincident, it is shown in [17] that the principal axes of $\mathbf{E}^*$ are also coincident with those of $\mathbf{T}^*$ and $\mathbf{H}^*$, thus

$$\mathbf{T}^*\mathbf{H}^* = \mathbf{H}^*\mathbf{T}^*, \ \mathbf{E}^*\mathbf{H}^* = \mathbf{H}^*\mathbf{E}^*, \ \mathbf{T}^*\mathbf{E}^* = \mathbf{E}^*\mathbf{T}^*, \tag{4}$$

the principal axes of $\mathbf{T}$, $\mathbf{E}$ and $\mathbf{H}$ are all coincident or they are all distinct. Since the eigenvectors of $\mathbf{H}$ are the axes of material symmetry for orthotropy, this coincidence of the principal axes of $\mathbf{T}$, $\mathbf{E}$ and $\mathbf{H}$ is equivalent to the observation that the principal axes of $\mathbf{T}$ and $\mathbf{E}$ coincide in an orthotropic material only when the axes of $\mathbf{T}$ and $\mathbf{E}$ are coincident with the axes of material symmetry.

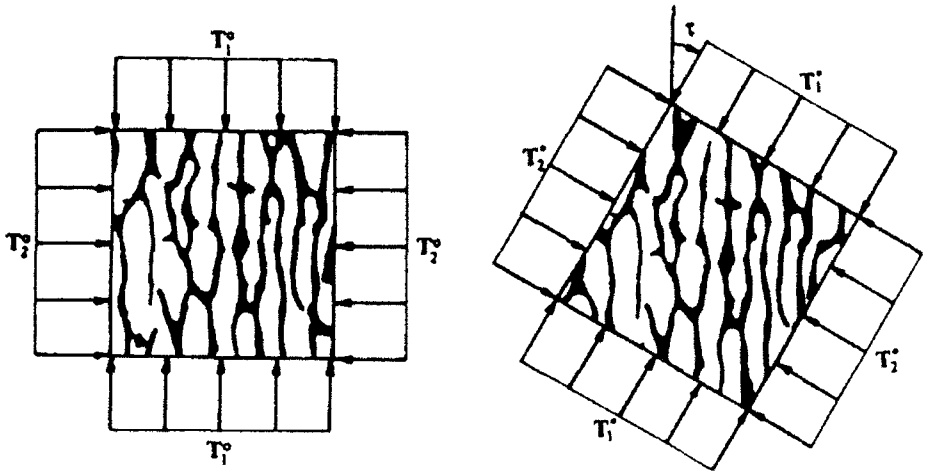


Figure 4. An illustration of the formulation of the evolutionary fabric problem considered. a.) The remodeling equilibrium situation associated with a stress state $\mathbf{T}^\circ$ that exists for $t < 0$. b.) The situation at $t = 0^+$ when a new stress state $\mathbf{T}^*$ is applied and held constant for all subsequent time.

5. AN EVOLUTIONARY WOLFF'S LAW

A constitutive hypothesis to describe the temporal evolution of trabecular architecture, an extension of the formulation of Wolff's law at remodeling equilibrium described above, was put forward by Cowin *et al.* [21]. That constitutive hypothesis will be described here, without equations. This is accomplished by describing what the theory will predict in a particular, spatially homogeneous evolutionary fabric problem.

The evolutionary fabric problem considered is a situation in which a spatially homogeneous trabecular architecture, characterized by the fabric $\mathbf{H}^\circ$ and a solid volume fraction v_0, is in RE under the stress $\mathbf{T}^\circ$ and the strain $\mathbf{E}^\circ$ for a long time ($-\infty < t < 0$). At time $t = 0^+$ the stress is changed to a new stress $\mathbf{T}^*$ and held at the new stress for all time ($0 < t < \infty$). As time becomes very large a new RE state denoted by $\mathbf{T}^*$, $\mathbf{H}^*$, $\mathbf{E}^*$ and v^* is achieved. The two RE states of the problem are thus characterized as follows: the one that exists before $t = 0$ is characterized by $\mathbf{T}^\circ$, $\mathbf{E}^\circ$, $\mathbf{H}^\circ$ and v_0; and the one that exists as time becomes very large, is characterized by $\mathbf{T}^*$, $\mathbf{E}^*$, $\mathbf{H}^*$ and v^*. These two stress states are illustrated in Figure 4.

The fabric and strain tensors as they evolve in time between these two RE states are denoted by $\mathbf{H}(t)$ and $\mathbf{E}(t)$. The various coordinate systems employed are illustrated in Figure 5. Figure 5 is a sequence of four illustrations showing the temporal evolution of the trabecular architecture from an RE situation characterized by states of stress, strain and fabric denoted as $\mathbf{T}^\circ$, $\mathbf{E}^\circ$ and $\mathbf{H}^\circ$, respectively, to an RE situation characterized by states of stress, strain and fabric denoted as $\mathbf{T}^*$, $\mathbf{E}^*$ and $\mathbf{H}^*$, respectively. Figure 5a illustrates the remodeling equilibrium situation that exists for $t < 0$. The principal axes of $\mathbf{T}^\circ$, $\mathbf{E}^\circ$ and $\mathbf{H}^\circ$ are coincident. Figure 5b illustrates the situation when remodeling is initiated at $t = 0^+$. The applied stress is now represented by the temporally

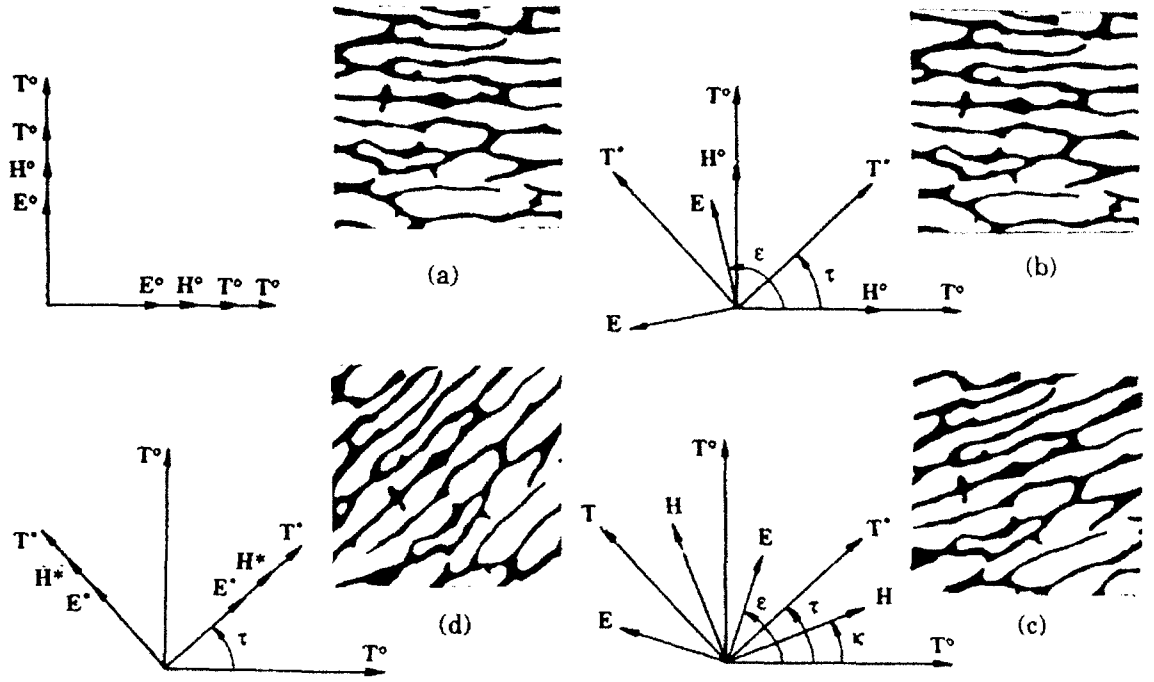


Figure 5. This sequence of four illustrations shows the evolution of the trabecular architecture from an RE situation characterized by states of stress, strain and fabric denoted as $\mathbf{T}^\circ$, $\mathbf{E}^\circ$ and $\mathbf{H}^\circ$, respectively, to an RE situation characterized by states of stress, strain and fabric denoted as $\mathbf{T}^*$, $\mathbf{E}^*$ and $\mathbf{H}^*$, respectively. The clockwise sequence starting from the upper left hand corner, (a), (b), (c) and (d) is described in the text.

constant stress tensor $\mathbf{T}^*$, whose principal axes make an angle τ with the principal axes of the previously applied stress tensor $\mathbf{T}^\circ$. Furthermore, the temporally varying strain tensor $\mathbf{E}(t)$ has instantaneously assumed a new value while the fabric tensor $\mathbf{H}^\circ$ remains at its $t < 0$ value. Figure 5c illustrates the situation at a typical time t as remodeling is proceeding. The angles that the principal axes of strain and fabric make with the horizontal axes are denoted by $\varepsilon(t)$ and $\kappa(t)$, respectively. As time increases, the principal axes of strain and fabric move into coincidence with the principal axes of the applied stress. Figure 5d illustrates the new remodeling equilibrium situation which is achieved as t tends to infinity. A new RE situation has been established; it is characterized by states of stress, strain and fabric with coincident principal axes denoted as $\mathbf{T}^*$, $\mathbf{E}^*$ and $\mathbf{H}^*$, respectively, and by a solid volume fraction ν^*.

The model just described is necessarily non-linear. This non-linearity is fundamental in that it stems from the fact that, during remodeling, the relationship between stress and strain is changing as the stress and strain variables themselves are changing. In order to preserve the remodeling property of the model, terms that are of the order strain times the changes in density and/or microstructural properties must be retained. If these terms were dropped, there would be no feedback mechanism for architectural adaptation and no adaptation of the trabecular architecture. There is, therefore, no linearized version of this model of the temporal evolution of

trabecular architecture. A significant limitation of the proposed continuum model is the length scale below which it cannot be applied, about 5mm, because of the inhomogeneity of the porous structure. This precludes application of the model at a bone-implant interface.

6. MECHANOSENSORY MECHANISM

The continuum model considered above is described as "phenomenological" because it has as its objective a description of the observed phenomena, without addressing the mechanism that controls the phenomena. The remainder of this paper addresses this shortcoming by describing a model for a cellular level mechanosensory transduction mechanism for Wolff's law. This is a new hypothesis for the cellular level feedback mechanism, recently advanced by Weinbaum *et al.* [22]. The hypothesis is that the mechanical loads applied to a whole bone cause the movement of bone water which, in turn, causes shear stresses on the bone cells encased in the whole bone; the bone cells sense these stresses and communicate with other bone cells to deposit or resorb bone tissue. The bone cells that sense the fluid movement are the osteocytes housed in the lacunae. The strains that are detected are cyclic dynamic strains. The model incorporating this hypothesis employs a hierarchical model of bone tissue structure as the medium for the conversion of cyclic mechanical loading applied to the whole bone to the fluid shear stress at the surface of the osteocytic cell process.

There are three levels of scale in this model. The smallest scale level is the spacing Δ (approximately 7nm) between the proteoglycan fibers on the membrane surface of the osteocytic process. The middle scale level is the gap (approximately 0.1 μm) between the wall of a canaliculus and the surface of the cell process in the canaliculus. The largest scale level is a unit cell (approximately a cubic volume 40 μm on each edge) containing a lacuna and half the length of all the canaliculi associated with the lacuna. Using viscous fluid theory, the Brinkman approximation for fluid flow through porous media and the Biot theory - the theory of stress-induced flow of interstitial fluid in porous solids - we relate the fluid shear stress at the osteocytic surface to the mechanical load applied to the bone.

7. HIERARCHICAL MICROANATOMICAL MODEL LEVELS

The purpose of this section is to describe the three hierarchical anatomical levels of the model. An anatomical illustration of a lacuna housed osteocyte and its cytoplasmic processes, taken from page 225 of Krstic [23], is shown in Figure 6. A schematic model of the lacunar-canalicular porosity associated with the fluid space surrounding the osteocytes in a trabecula is illustrated in Figure 7. An idealized model of the cross section of a canaliculus is shown in Figure 8. This figure illustrates the details of the fluid annulus of the canaliculus with its cell process and fiber matrix structure. While Figure 7 is on the length scale L (approximately 40μm) of the lacunar spacing, the field shown in Figure 7 is 5L by 5L (200μm by 200μm); both panels of Figure 8 are on a scale of about 1/500 that of Figure 7, approximately 0.4μm by 0.4μm. The top of Figure 8 shows the annular shape of the transverse cross section and the bottom shows a longitudinal cross section. The annular region exists between

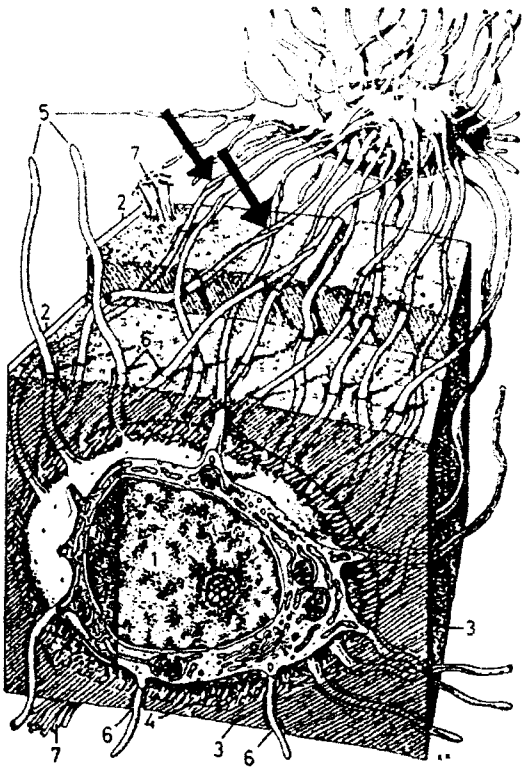


Figure 6. Diagram of two osteocytes (1) in the lamellar bone of calcified bone matrix (3). Two neighboring lamellae (2) with different collagen orientations (7) are visible. The osteocytic cell bodies are located in lacunae and are surrounded by a thin layer of uncalcified matrix (4). Their cell processes (5) are housed in canaliculi (6). Some of the gap junctions between the cell processes are indicated (arrows).

the surface of the osteocytic cell membrane and the surface of the canalicular wall. Figure 8 shows an idealization of the cross-bridging fibers and the open gap Δ between the fibers. The cross-bridging fibers, which traverse the annular region, represent the cross linking proteoglycans of the surface glycocalyx structure. It is important to note that the canaliculus shown in Figure 8 is one of the many canaliculi indicated on the much larger scale idealization shown in Figure 7. Figure 9 gives an enlarged view of the cross linking proteoglycan net in a small central section of a typical canaliculus, such as the one labeled [1] in Figure 7. Figure 9 is another idealized model of the cross linking proteoglycan net in the annular region between the osteocytic cell membrane and the surface of the canalicular wall. Figure 9 is on a scale of about 1/4 the scale of Figure 8 and 1/2000 the scale of Figure 7. Although the model of the cross linking proteoglycan net used in the analysis of [22], shown in Figure 8, does not mimic the detail of the representation of the cross linking proteoglycan net shown in Figure 9, the work of Weinbaum *et al.*[25] shows that the effective fluid resistance is roughly the same for both periodic and random matrix structures with the same average fiber spacing.

The four fundamental lengths that characterize the local fluid behavior are shown in these figures. One is the center to center distance between two lacunae L illustrated in Figure 7 and a second is the radius of curvature of the annulus b as shown in Figure 8. The other two fundamental lengths are shown in Figures 7 and 8: the gap height of the annulus, (b - a), and the matrix fiber spacing Δ.

8. HIERARCHICAL PERMEABILITY MODEL LEVELS

The purpose of this section is to describe the three permeability constants associated with the three hierarchical length scales Δ , (b - a) and L. These

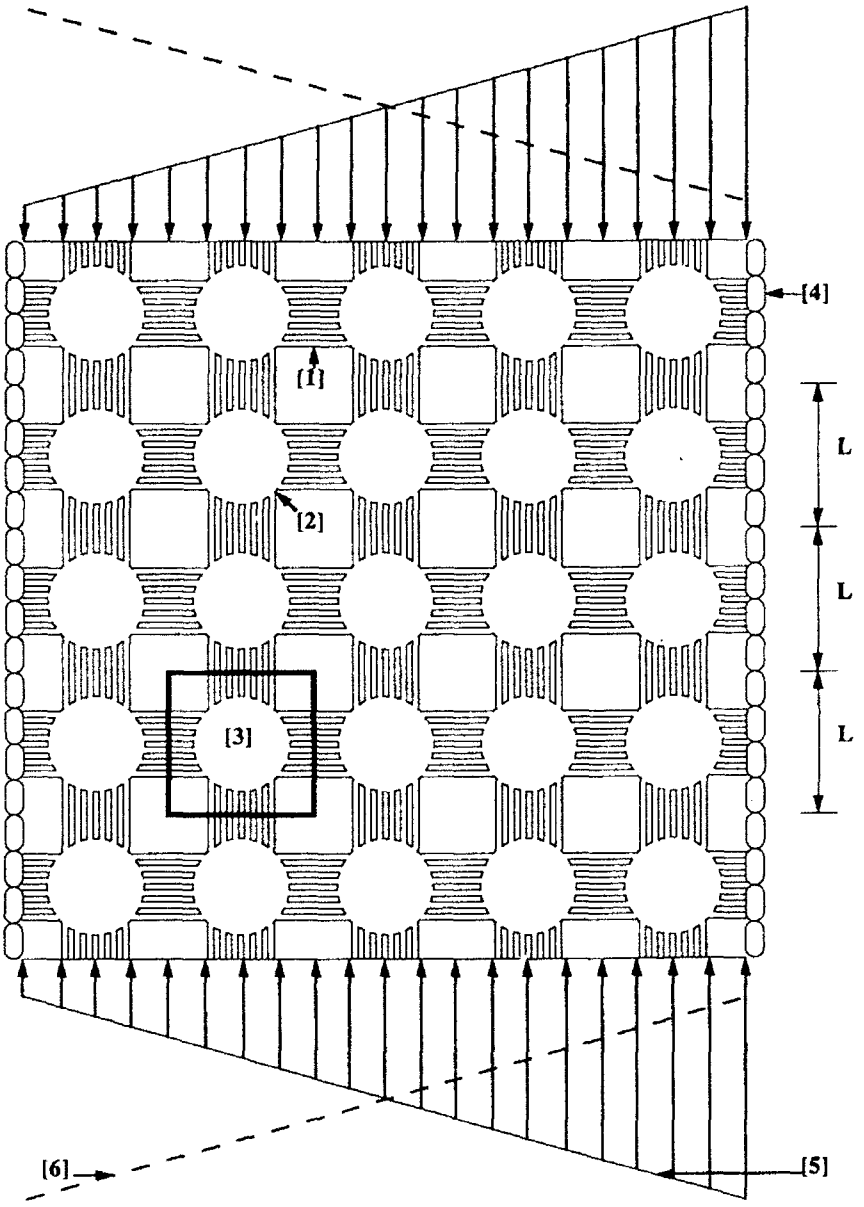


Figure 7. An idealized model of a trabecular cross section illustrating the lacunar-canalicular porosity associated with the fluid space surrounding the osteocytes in a trabecula. The number [1] indicates one of the numerous canaliculi that connect lacunae [2]. The number [1] also indicates the region of this Figure that is enlarged many times in Figure 8. Each lacuna [2] houses an osteocyte (not shown), and osteocytic processes (not shown) run from each osteocyte about half way down each canaliculus to connect, with a gap junction (not shown), to the osteocytic processes from the osteocyte housed in the adjacent lacuna. The number [3] indicates the unit cell used in our calculation of the large scale permeability constant k, [4] indicates bone cells on the surface of the trabecula. The linear stress distribution that the cross section is subjected to at one extreme of the cyclic loading is indicated by [5] and at the other extreme by [6]. This linear stress distribution represents a periodic, combined axial and bending loading. The length L is the center to center distance between two lacunae.

levels correspond to the three hierarchical levels of the model described in the previous section. The three permeability constants are denoted by k_p, $k_{p,eff}$ and k, respectively, and each has the units of area. The local or small scale permeability constant k_p represents the permeability of the bone fluid through the cell membrane glycocalyx of cross-bridging fibers of the glycosaminoglycans (GAGS) of the proteoglycans, without consideration of the boundaries of flow, namely the surface of the cell process and the wall of the canaliculus. In order to account for the boundary layer effects at these boundaries of the pore space, the effective permeability $k_{p,eff}$ constant is introduced. It represents an averaging of the small scale permeability k_p over

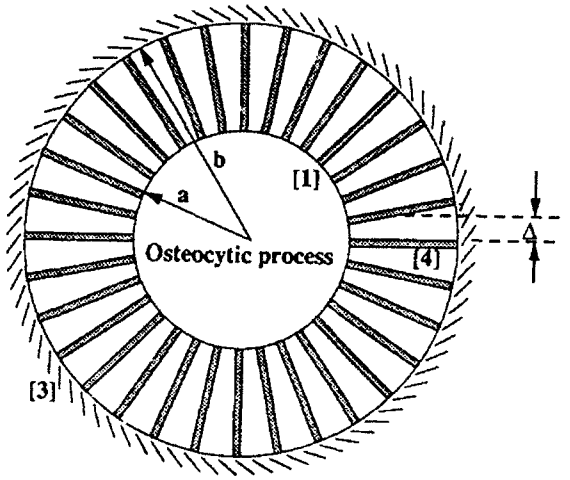
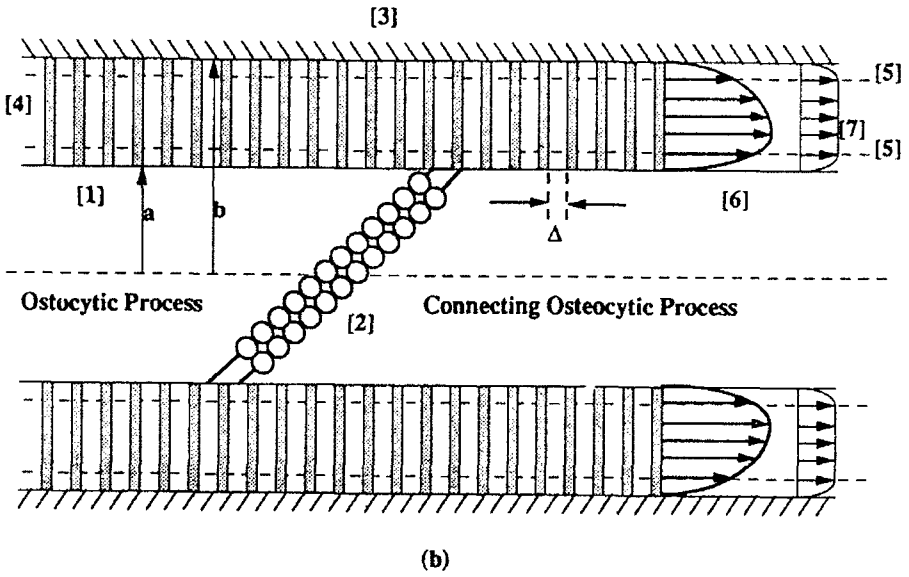


Figure 8. An idealized model of the cross section of a canaliculus: (top) transverse cross section showing the annular shape of the region, (bottom) the longitudinal cross section. This Figure is on a scale of about 1/500 times the scale of Figure 7. This Figure gives an enlarged view of a small central section of a typical canaliculus, such as the one labeled [1] in Figure 7. The number [1] of this Figure indicates the surface of the osteocytic cell membrane, [2] the gap junction between two connecting osteocytic cell processes, [3] the surface of the canalicular wall, [4] a typical fiber traversing the annular region and representing the cross linking proteoglycans. The dotted lines [5] represent the outer edge of the viscous boundary layer.The velocity distribution varies from a radially distorted parabola [6], to a plug flow [7] as the effective viscous boundary layer diminishes due to an increase in the fiber density. Three of the four fundamental lengths that characterize the local fluid behavior are shown: the radius of curvature of the annulus b, the gap height of the annulus, (b - a), and the open gap Δ between the fibers. The fourth fundamental length is the center to center distance length L between two lacunae shown in Figure 7.



the annular pore space of the canaliculus. The large scale permeability constant k represents the permeability of the entire canalicular system of a lacuna, a system consisting of all the canaliculi associated with one lacuna. The following items are discussed below in the order indicated: the largest scale constant k; then the smallest scale constant k_p; formulas for $k_{p,eff}$ and k; and the relationship of k_p to D.

The large scale permeability constant k is introduced as part of Darcy's law. If **u** represents the fluid velocity vector and ∇p the gradient of the pore fluid pressure p, then Darcy's law is

$$\mathbf{u} = -\kappa \nabla p, \tag{5}$$

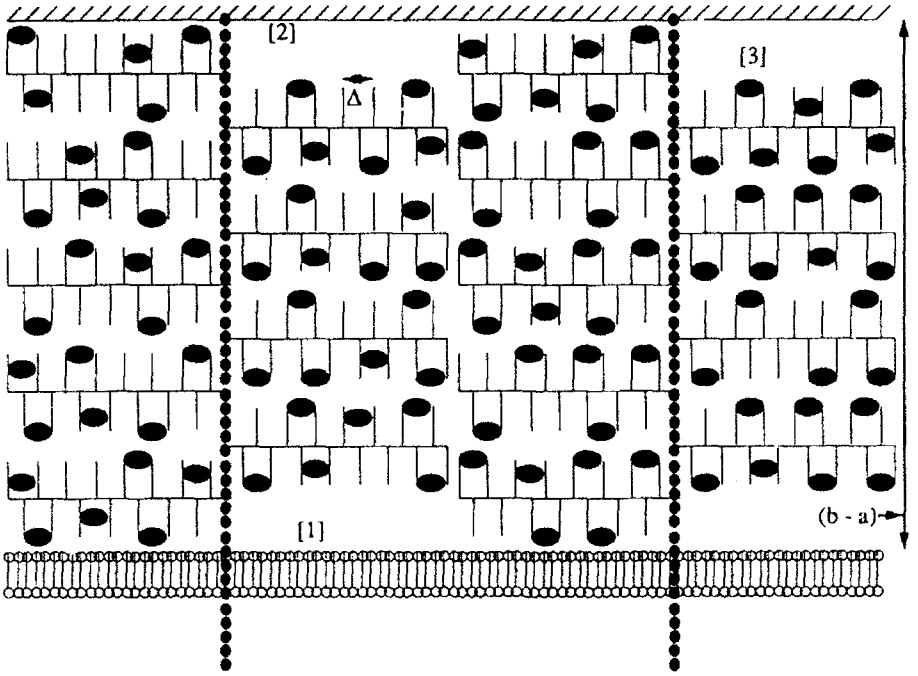


Figure 9. An idealized model of the cross linking proteoglycan net in the annular region between the osteocytic cell membrane and the surface of the canalicular wall. This Figure is on a scale of about 1/4 the scale of Figure 8 and 1/2000 the scale of Figure 7. This Figure gives an enlarged view of the cross linking proteoglycan net in a small central section of a typical canaliculus, such as the one labeled [1] in Figure 7. The number [1] of this Figure indicates the osteocytic cell membrane, [2] the surface of the canalicular wall, and [3] the darkened ellipses represent albumin molecules.

where the coefficient of permeability, $\kappa = k/\mu$, is a ratio of the permeability constant to the fluid viscosity μ.

The local or small scale permeability constant k_p for the fluid filled matrix is introduced as part of Brinkman's [26] equation. The theory in Tsay and Weinbaum [27] has shown that Brinkman's equation provides a highly accurate approximation for bounded flow in a channel with cross-bridging fibers provided the aspect ratio of the fibers is greater than five. This criterion for the aspect ratio is easily satisfied for long slender proteoglycan fibers. We thus describe the fluid annulus surrounding the osteocytic process by a Brinkman equation for a fiber filled medium,

$$\nabla p = -\frac{\mu}{k_p}\,\mathbf{u} + \mu\nabla^2\mathbf{u}\,. \tag{6}$$

Brinkman's equation (6) reduces to Darcy's law (5) in the dense fiber limit when k_p is small and to the standard Stokes equation (*i.e.*, $\nabla p = \mu\nabla^2\mathbf{u}$) in the limit where the fibers are dilute and k_p is large. The Brinkman equation (2) shows that the thickness of the fiber induced boundary layer is of the order $\sqrt{k_p}$.

The average velocity of flow through a canalicular pore annulus containing the fiber matrix is obtained by dividing the volume flow rate Q

through the canaliculus by the annular area $\pi(b^2 - a^2)$. The effective permeability $k_{p,eff}$ is determined from the average velocity $Q/\pi(b^2 - a^2)$, where Q is calculated from the velocity profiles in the annulus obtained from the Brinkman equation (6), the fluid viscosity μ, and the pressure gradient in the axial direction of a canaliculus, $|\nabla p|$, thus

$$k_{p,eff} = \frac{Q}{\pi(b^2 - a^2)} \mu |\nabla p|^{-1} \quad . \tag{7}$$

The large scale permeability constant k can then be determined by first calculating the large scale volume flow rate by multiplying the volume flow rate Q through one canaliculus by n/L^2- the number of canaliculi n per unit area L^2 crossing the face of a unit cell. The quantity nQ/L^2 is equal to the average velocity u in the axial direction of a canaliculus. Darcy's law (1) can then be inverted to obtain an expression for the large scale permeability constant k,

$$k = \frac{\mu n Q}{L^2} |\nabla p|^{-1} \quad . \tag{8}$$

This is the large scale permeability constant k for the unit cell illustrated in Figure 7. The cell is assumed to have n canaliculi per unit area in each of its three principal directions.

The value of the small scale permeability constant k_p for an infinite medium can be related to the fiber spacing parameter Δ. A relationship of this type can be obtained from the rigorous solution for the flow perpendicular to a two-dimensional square array of fibers. This solution, however, is a complicated function of the fiber volume fraction for this geometry and Tsay and Weinbaum [27] show that the following simple approximate formula for k_p as a function of Δ is accurate to within ten percent for all volume fractions less than 0.7:

$$k_p = 0.0572 a_o^2 \left(\frac{\Delta}{a_o}\right)^{2.377} , \tag{9}$$

where a_o is the fiber radius. A rough approximation to equation (9) is that k_p is approximately equal to the square of Δ; thus, since the Brinkman equation shows that the thickness of the fiber-induced boundary layer is of the order $\sqrt{k_p}$, it follows that the boundary layer thickness is approximately equal to the fiber spacing parameter Δ.

9. MODEL PREDICTIONS

The model described above, in conjunction with the theory of stress-induced flow of interstitial fluid in porous solids (Biot theory), is used to relate the externally applied cyclic mechanical loading applied to the whole bone to the fluid shear stress experienced by the osteocytic cell process in a bone canaliculus. The first step is to employ the Brinkman equation (6) to find the

fluid velocity in a canaliculus through the fiber net representing the cell membrane glycocalyx of cross-bridging proteoglycan fibers. This velocity can then be used to relate the fluid shear stress on the cell process to the pressure gradient in the canaliculus. The pressure gradient in the canaliculus is related to the applied mechanical loading by the Biot theory [28]. Biot theory has two basic assumptions: (i) Darcy's law (5) governs the flow of the interstitial fluid; (ii), the matrix material of the porous medium follows Hooke's law. The permeability k is employed in the Biot theory. It is calculated, as is $k_{p,eff}$, from the fluid velocity in a canaliculus determined above. The velocity is integrated over the cross section to obtain the total flux Q which, in turn, is used in the formulas (7) and (8) for $k_{p,eff}$ and k, respectively.

Biot theory is then used to calculate the pore fluid pressure relaxation time from the expression for the large scale permeability k and known data on the elastic constants of bone. This relaxation time has been measured experimentally and shown to be approximately 2 sec. However, for the typical values of a and b of the inner and outer radii of the canaliculus, 0.087 and 0.20 μm, the model predicts a fluid pressure relaxation time of the order 10^{-2} sec., much less than the experimentally measured 2 sec., if the proteoglycan matrix were not present in the fluid annulus. However, when the annulus is filled with a matrix whose principal component is the GAG side chains of the proteoglycans, fibers with a typical radius of typical radius of a_o = 0.6nm, one finds for the same values of b and a that a pore relaxation time of 2 sec. can be achieved when the fiber spacing Δ in (9) is 7 nm. This predicted spacing corresponds nearly exactly with the size selective matrix structures that have been observed at the membrane surface of endothelial cells in many tissues where this structure serves as a capillary filter for plasma proteins, Michel [29]. It is of particular significance that bone pore fluid is a plasma derivative that is rich in albumin. This molecule has been hypothesized to order the side chains of the proteoglycans, as shown in Figure 9, and thus determine the characteristic dimension of the molecular sieve, for virtually all capillary tissues.

The second important prediction of the model is the shear stress acting on the membranes of the osteocytic processes. It is important to note that the theory predicts that no fluid movement will occur, and thus no shear stress, when the loads applied to the bone are steady. The results we report are for the cyclic loading conditions at 1 Hz. For a matrix with the above fiber spacing one finds that for whole bone subject to physiological strains, a peak shear stress of between 10 and 20 dyne/cm^2 will be obtained, the maximum value depending on the ratio of the applied bending to axial load. These shear stresses are precisely in the same range as the fluid shear stresses acting on endothelial cells where an adaptive remodeling of the vasculature has been observed. From geometric considerations the portions of the membranes of the osteocytes that lie within the lacunae would experience considerably lower shear stress levels. Although the strain is mechanically sensed by the osteocytes, it can be communicated to the osteoblasts and bone lining cells at the surface of the trabeculae by changes in ion currents through cell to cell membrane ion channels in the gap junctions at the apical ends of the osteocytic process, as shown in Figure 8b.

These theoretical predictions have important implications for the possible nature of the excitation process that control stress-related resorption and deposition of bone tissue. The results suggest that the same fiber matrix

microstructure characteristic of other filtration tissues exists in the fluid sheath surrounding the osteocytic processes. Theoretical predictions for the fluid shear stress on the membranes of the osteocytic processes correspond closely to the experimentally measured range of stresses where Ca^{++} has been observed to be released from intracellular stores in shear stress-activated endothelial cells. The presence of Ca^{++} is known to control the opening and closing of the ion channels in the gap junctions, thereby changing the potential difference between cells. Weinbaum *et al.*[22] propose that the resulting change in potential through the intracellular pathway is the electrical stimulus for the activation of the osteoblasts.

ACKNOWLEDGMENT

This investigation was supported by NSF Grant No. BSC-9103236 and by grant number 662319 from the PSC-CUNY Research Award Program of the City University of New York. The author thanks T. J. Pedley and S. Weinbaum for constructive comments on an earlier draft.

REFERENCES

1. K. Culmann, *Die Graphische Statik.* (Meyer & Zeller, Zurich, 1866).
2. J. Wolff, *Arch. path. Anat. Physio. Med., Virchovs Arch.* 50 (1870) 389.
3. J. Wolff, *Das Gesetz der Transformation der Knochen.* (Hirschwald, Berlin, 1892).
4. J. Wolff, *The Law of Bone Remodelling,* (Springer, Berlin, 1986).
5. T. M. Charlton, *A History of the Theory of Structures in the Nineteenth Century,* (Cambridge University Press, Cambridge, 1982).
6. W. Roux, *Gesammelte Abhandlungen uber die Entwicklungsmechanik der Organismen,* (W. Englemann, Leipzig, 1885).
7. F. Pauwels, *Biomechanics of the Locomotor Apparatus,* (Springer, Berlin, 1980).
8. W. C. Hayes and B. Synder, in *Mechanical Properties of Bone,* edited by S. C. Cowin (American Society of Mechanical Engineers, New York, 1981).
9. R. B. Little, H. W. Wevers, D. Siu, and T. D. V. Cooke, *J. Biomech. Eng.* 108 (1986) 111.
10. M. Oda, *Soils Found.* 12 (1972) 17.
11. M. Oda, J. Konishi, S. Nemat-Nasser, *Geotechnique* 30 (1980) 497.
12. T. P. Harrigan, R. W. Mann, *J. Mater. Sci.* 19 (1984) 761.
13. C. H. Turner, S. C. Cowin, J. Y. Rho, R. B. Ashman, J. C Rice, *J. Biomech.* 23 (1990) 549.
14. R. Hodgskinson and J. D. Currey, *Proc. Instn. Mech. Engrs.* 204 (1990) 101.
15. W. J. Whitehouse, *J. Microscopy* 101 (1974) 153.
16. W. J. Whitehouse, E. D. Dyson, *J. Anatomy* 118 (1974) 417.
17. S. C. Cowin, *J. Biomech. Eng.* 108 (1986) 83.
18. S. C. Cowin, *Bull. Hosp. Joint Dis. Orthop. Inst.* 49 (1989) 206.
19. S. C. Cowin, *Mech. Mater.* 4 (1985) 137.
20. S. C. Cowin, *Cal. Tissue Int.* 36 (1984) S99.
21. S. C. Cowin, A. M. Sadegh, and G. M. Luo, *J. Biomech. Eng.* 114 (1992) 129.
22. S. Weinbaum, S. C. Cowin and Yu Zeng, submitted to *J. Biomech.*
23. Krstic, R.V. . *Die Gewebe des Menschen und der Saugetiere.* (Springer Verlag, Berlin, 1978).
24. S. C. Cowin, L. Moss-Salentijn and M.L. Moss, *J. Biomech. Eng.* 113(1991) 191.
25. S. Weinbaum, R. Tsay and F. E. Curry, Microvasc. Res. 44(1992)85.
26. H. C. Brinkman, *Physica.* 13(1947) 447.
27. R.-Y. Tsay and S. Weinbaum, *J. Fluid Mech.* 226(1991)125.
28. M.A. Biot, *J. Appl. Phys.* 12(1941)155.
29. C. C. Michel, *J. Physiol.* 404(1988) 1.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.

© 1993 IUTAM. All rights reserved.

LIQUID LAYER DYNAMICS IN PULMONARY AIRWAYS

R.D. Kamm

Massachusetts Institute of Technology, Department of Mechanical Engineering, Fluid Mechanics Laboratory, Cambridge, Massachusetts 02139, USA

Abstract

The pulmonary airways are lined with a surfactant-rich liquid layer which thickens markedly with respiratory disease. The dynamics of this liquid layer have been investigated, primarily through numerical simulation. Two specific cases were examined: (i) When the layer is sufficiently thick, a fluid dynamic instability may occur, causing the liquid layer to form a meniscus that obstructs the airway lumen. The numerical results indicate that this occurs during deflation at approximately 25% maximal lung volume for a liquid layer initially 10 μm thick and that the presence of surfactant can delay or even prevent this process. (ii) When a bolus containing exogenous surfactant is instilled into the lung, it spreads due to Marangoni convection. The influence on spreading rate of the characteristics of the instilled bolus and of the resident liquid layer have been examined.

1. INTRODUCTION

When you exhale as far as possible, a volume of gas remains in the lung -- usually about 25% of the Total Lung Capacity (TLC) -- that cannot be eliminated no matter how much effort is expended. While several factors contribute to this observation, the reason in many subjects, especially those with various forms of respiratory disease, is that there is a thin liquid layer that lines the airways of the lung from the gas exchange region in the periphery (the "alveolar zone") all the way to the trachea. The liquid layer, if sufficiently thick relative to the airway radius, can undergo a fluid dynamic instability akin to the Rayleigh instability of a liquid jet; a small perturbation in film thickness grows for wavelengths greater than $2\pi a$ where a is the undisturbed inner radius of the film [1]. Moreover, if the liquid layer is thin relative to the airway radius, this instability ultimately leads to the formation of an unduloid or portions of an unduloid separated by dry regions [2]. For thicker films, as might be expected in respiratory disease or at lower lung volumes, the instability leads instead to the formation of a meniscus that blocks the lumen of the airway producing what is termed "airway closure" [3]. This phenomenon is thought to be responsible for the observation that lungs trap gas at low volume and may also be the cause of an increase in minimal lung volume in diseases which are associated with excess airway liquid. In such diseases, airway closure, if it persists, can lead to gas absorption peripheral to the meniscus ("atelectasis"), with a concomitant impairment in gas exchange capability.

The thickness of the pulmonary liquid layer varies from about 0.1 μm in the smallest airspaces, the alveoli, to about 10 μm in the trachea in a normal, healthy lung [4], and can increase many-fold in respiratory disease. What regulates its thickness remains largely unknown although the epithelial cells, especially those in the alveolar zone, have the capability of pumping sodium ion out of the airway liquid, thus setting up an osmotic gradient that draws water out of the airway lumen and into the surrounding tissue space, the interstitium [5]. In disease, water originating in the pulmonary capillaries enters the interstitial space causing pulmonary edema, and often leaks through the alveolar epithelium into the alveolar airspace.

The liquid layer has many functions, among which are the transport of airborne impurities toward the mouth, the protection of the thin layer of epithelial cells that line the airways, and the humidification of the air that you breathe in. Although largely comprised of water, the properties of this fluid are complicated by the presence of a potent surfactant produced by epithelial cells in the alveolar zone, of mucus, a highly viscous, non-Newtonian substance, introduced by goblet cells contained in the lining of the airway walls, and of proteins, some of which are needed for normal surfactant function and others, such as those originating from the blood plasma, which act to impair normal function of surfactant. Far from static, this layer is in constant motion due in part to the action of ciliated cells along the airway that continually drive the liquid toward the trachea. Additional factors that can affect the motion and distribution of the airway liquid include coughing, transport across the airway epithelium, gravity, changes in lung volume, and interfacial shear due to normal breathing.

One of the more fascinating yet little-studied characteristics of this liquid layer is the potential for flows generated by surface-tension gradients (the Marangoni effect). In the normal lung, there exists a standing gradient with surface tension low in the alveolar zone and high in the central airways [6,7]. This gradient, in combination with the existence of a contiguous liquid layer, must tend to generate a steady flow of liquid toward the mouth totally apart from flows generated by the cilia. Such naturally-occurring Marangoni flows are an aspect of liquid transport in the lung that has never been subjected to systematic, quantitative analysis.

The role of pulmonary surfactant in disease is widely recognized. For example, in neonatal respiratory distress syndrome (RDS) the primary pathology is a lack of normal surfactant production. The currently accepted and generally successful treatment for infants suffering from RDS is direct instillation of an exogenous surfactant, known as surfactant replacement therapy (SRT) [8]. In a small but significant subgroup, however, SRT fails for reasons that are not yet fully understood, but may be due to factors that influence the transport of exogenous surfactant to the lung periphery. These issues have led to the recent interest in transient Marangoni flows in thin films [9,10,11]. The observation that surfactant is efficiently and rapidly distributed throughout the lung has also led to the notion that surfactants might be used as vehicles for drug delivery [12], either by direct instillation of a bolus or via an aerosol.

In this paper, we discuss two topics related to the dynamics of the pulmonary liquid layer. The first is airway closure which we simulate for the smallest non-alveolated airways (the terminal bronchioles) found in the lung, including the effects on closure of endogenous surfactant. The second is the transient distribution of a bolus of exogenous surfactant, placed in an axi-symmetric distribution on the walls of a tube of constant diameter. In addition, some qualitative experimental results are presented for airway closure in a large-scale model.

2. METHOD OF ANALYSIS

The purpose of this modeling study is (i) to elucidate the time-scales for airway closure and the lung volumes at which closure occurs for a given liquid layer thickness, and (ii) to identify the rate of spread of an instilled surfactant monolayer. For the analysis of these problems in a thin liquid layer lining the inside of a tube of cylindrical cross-section (Fig. 1),we employ a numerical method solving the coupled continuity, surfactant conservation and axial momentum equations. We assume that (i) the surfactant monolayer is insoluble in the aqueous sub-phase, (ii) that, for cases simulating expiration, the model airway retains a fixed geometry as its dimensions decrease, (iii) the aqueous sub-phase behaves as a Newtonian fluid, (iv) surface viscosity can be neglected, and (v) the lining is thin and disturbances are of long wavelength. For a more detailed derivation of the following equations see [13].

Consistent with these assumptions, the continuity equation can be expressed as:

$$\frac{\partial A}{\partial t} + \frac{\partial}{\partial z}(\overline{w}A) = 0 \qquad (1)$$

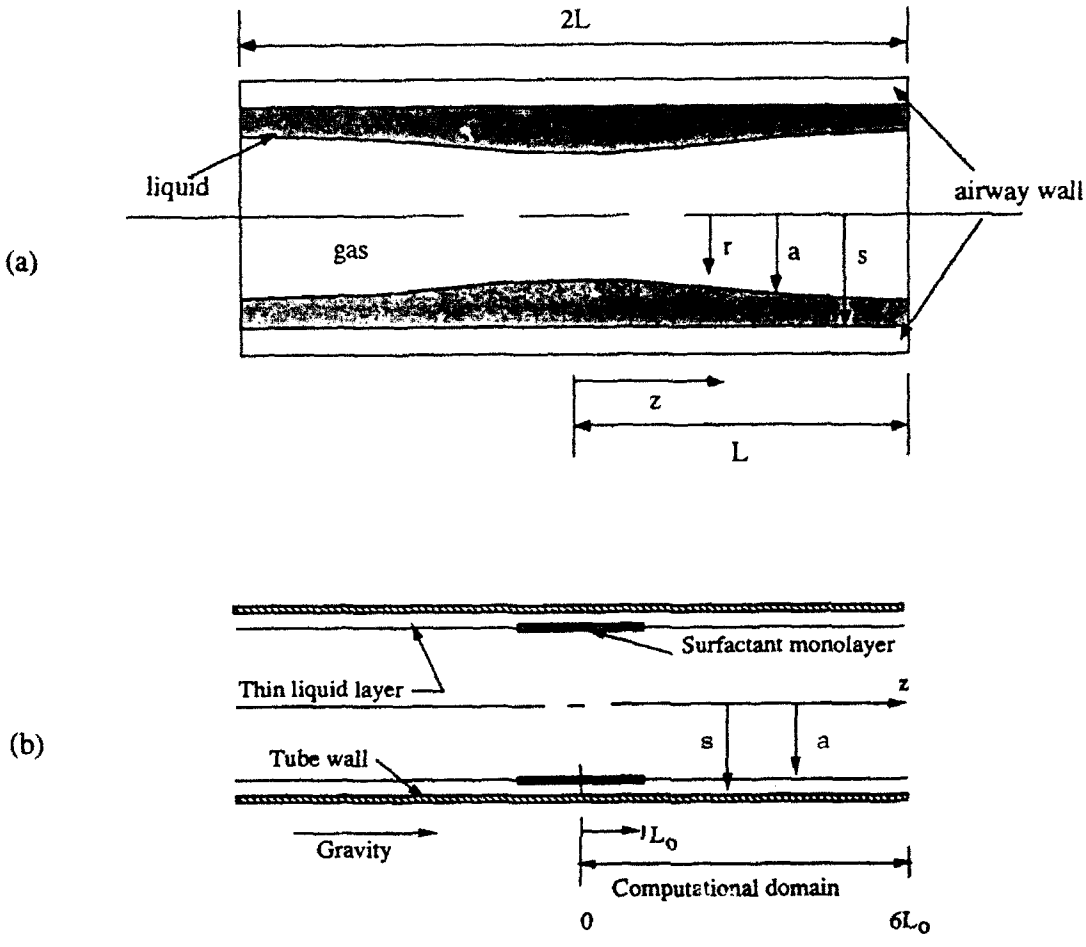


Fig. 1. Geometry used in the numerical simulations. (a) Airway closure model. (b) Surfactant spreading model.

where $A = (s^2-a^2)/2$ (all terms are defined in the nomenclature). In order to deal with problems in which the airway dimensions change with time, we introduce the following transformation:

$$x = \frac{z}{L(t)} \tag{2}$$

which, when introduced to eqn. (1), yields:

$$\frac{\partial A}{\partial t} + \frac{1}{L}\frac{\partial (A\bar{V})}{\partial x} + \frac{A\dot{L}}{L} = 0 \tag{3}$$

where $\bar{V}$ is the mean fluid velocity relative to the moving wall, $\bar{V} = \bar{w} - x\dot{L}$.

Applying a similar transformation to the surfactant conservation equation produces the expression:

$$\frac{\partial}{\partial t}\left(\frac{a\Gamma}{\gamma}\right)+\frac{1}{L}\frac{\partial}{\partial x}\left(\frac{a\Gamma V_a}{\gamma}\right)+\frac{a\Gamma\dot{L}}{L\gamma}-\frac{1}{L^2}\frac{\partial}{\partial x}\left(aD_s\gamma\frac{\partial\Gamma}{\partial x}\right)=0 \tag{4}$$

where

$$\gamma=\left[1+\left(\frac{1}{L}\frac{\partial a}{\partial x}\right)^2\right]^{-1/2} \tag{5}$$

The axial momentum equation, with gravity oriented parallel to the tube axis is written

$$\rho\left(\frac{\partial w}{\partial t}+u\frac{\partial w}{\partial r}+w\frac{\partial w}{\partial z}\right)=-\frac{\partial p}{\partial z}+\frac{\mu}{r}\frac{\partial}{\partial r}\left(r\frac{\partial w}{\partial r}\right)+\rho g \tag{6}$$

the solutions of which must satisfy boundary conditions on the radial velocity:

$$u_s=\frac{\partial s}{\partial t}+w_s\frac{\partial s}{\partial z} \qquad\qquad u_a=\frac{\partial a}{\partial t}+w_a\frac{\partial a}{\partial z} \tag{7}$$

axial velocity at the wall

$$w_s=x\dot{L} \tag{8}$$

and the velocity gradient at the interface

$$\mu\left(\frac{\partial w}{\partial r}\right)_a=-\frac{\partial\sigma}{\partial z} \tag{9}$$

which is what gives rise to the Marangoni effect. For all cases discussed below we assume symmetry about z = 0 and impose the condition of zero axial velocity at z = L.

The radial momentum equation is not needed since, in the long-wavelength approximation, the radial pressure gradient can be neglected compared with its longitudinal variation. Therefore, pressure in the liquid film is determined by Laplace's law applied across the gas-liquid interface, used in the form:

$$p(z,t)=\sigma\left(-\frac{\gamma}{a}+\frac{\partial}{\partial z}\left(\gamma\frac{\partial a}{\partial z}\right)\right)+p_{gas} \tag{10}$$

where gas pressure, p_{gas}, is assumed constant. Combining eqns. (6)-(9) and integrating the momentum equation across the liquid layer in the manner of Johnson et al.[3], we obtain

$$\left(\frac{\partial \bar{V}}{\partial t}\right)+x\ddot{L}+\frac{\bar{V}}{L}\left(\frac{\partial \bar{V}}{\partial x}\right)+\frac{\bar{V}\dot{L}}{L}=-\frac{1}{\rho L}\left(\frac{\partial p}{\partial x}\right)-\frac{4\mu}{\rho}\left[\frac{\bar{V}+\dfrac{a}{2\mu L}\dfrac{d\sigma}{dx}\left(\dfrac{a^2\ln\dfrac{s}{a}}{A}-1\right)}{A-a^2\left(1-\dfrac{a^2}{A}\ln\dfrac{s}{a}\right)}\right]+g \tag{11}$$

This system of equations is solved numerically using a third-order Adams-Bashforth technique for time-derivatives and central differences for spatial derivatives. In some of the following calculations, the effects of axial inertia were negligible, resulting in a somewhat simpler form of the equations. In addition, although surface diffusion is included in some of the following results, it was found to be negligible in all cases when realistic numerical values are used.

For simulation of the lung during expiration, all linear dimensions are taken to vary as the cube root of lung volume [14], which is assumed to decrease linearly with time:

$$L(t)=L_{TLC}\left(1-\frac{t}{T_{exp}}\right)^{1/3} \quad s(t)=\frac{L(t)}{3} \tag{12}$$

3. EXPERIMENTS ON AIRWAY CLOSURE

To provide a qualitative illustration of airway closure, experiments were conducted in a large-scale model of a segment of the airway tree, manufactured by a computer-aided milling process by Hammersly and Olson [15]. Two immiscible liquids were used, silicone oil (μ = 9.7 dyne·s/cm^2, ρ = 0.97 g/cm^3) and a water-ethanol mixture (88.4% water, 11.6% ethanol by weight) that was density-matched to minimize buoyancy effects. The water-alcohol mixture was used to simulate the relatively inviscid air and was dyed to distinguish it from the clear silicone oil that simulated the liquid lining in the lung.

The system was prepared for an experiment by first filling the entire network with silicone oil, then slowly adding the water-alcohol mixture, leaving a thin, nearly uniform layer of silicone oil on the walls of the network. The initial thickness of the liquid layer was approximately proportional to the airway diameter for each generation (h_0/s = 0.30±0.02) except for regions of near zero thickness near the flow divider (carina) of each bifurcation. When the entire network had been purged, the flow was stopped and a sequence of photographs were taken during the time it took for all of the branches to become occluded by a meniscus. A selected set of these photographs is shown in Fig. 2a-d.

Following the stoppage of flow, the uniform annular film in each airway initially grew toward an unduloid-like shape and then finally to complete closure by a meniscus. The process began slowly. then proceeded more and more rapidly as the amplitude of the disturbance grew, taking a total of from about 15 to 25 minutes. The final stages in the joining of the opposing interfaces occurred on a time scale of seconds whereas the process leading up to this event took many minutes. By 25 minutes, all branches were occluded, producing a collection of "dog-bone-shaped" dark regions.

Despite the tendency for the menisci to be slightly nearer the larger neighboring branch (parent airway) in each instance, there was little evidence to suggest that a significant quantity of liquid was exchanged between branches on the time scale of film collapse; each meniscus occupied roughly the same fractional volume of the branch at the end of the experiment. This may have been at least partly a result of the tendency for wall liquid to be drawn from the bifur-

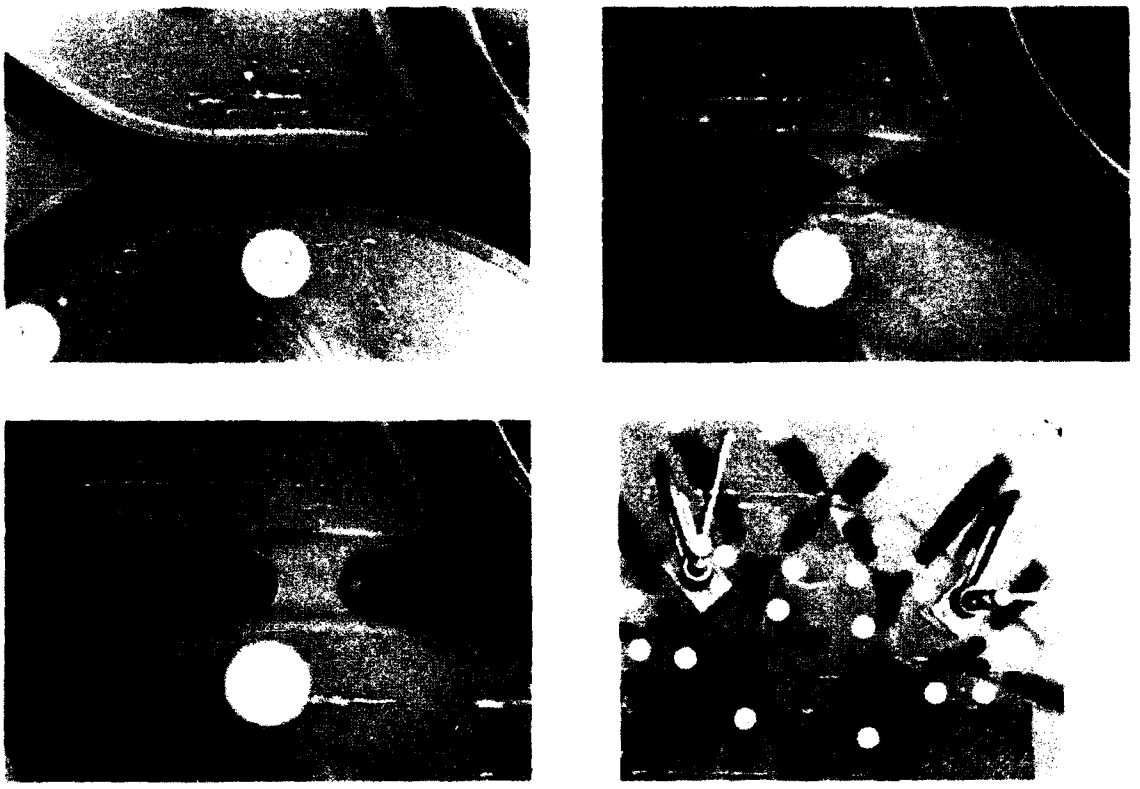


Fig. 2. Sequence of photographs showing the progression of a viscous liquid film lining the inside walls of a plexiglas model from a layer with h_0/s nearly uniform at 0.23 to a final state with all airways occluded by a meniscus. Photos correspond to the following times, referenced to the time that the airway shown in (a)-(c) is first purged: (a) t = 3:45 min. (b) t = 22:35 min. (c) t = 23:05 min. (d) t = 30:45 min..

cation zones toward the center of the branch as the instability grew, leaving behind a narrowed layer where exchange would have to take place. In every instance, the wavelength of the instability that ultimately led to closure was comparable to the length of the branch itself.

4. PREDICTIONS OF AIRWAY CLOSURE

Airway closure was modeled for a lung during expiration using the full form of eqns. (3), (4), (8) - (12). A single airway was considered, having the idealized geometry depicted in Fig. 1a. The airway size and liquid layer thickness at maximal lung volume were selected to represent typical values for a terminal bronchiole (see caption for Fig. 4). It is assumed that the airway remains geometrically similar and that no liquid escapes from the ends at x = 1,-1. In addition, the equation of state relating surfactant concentration to surface tension is given in Fig. 3, taken from the measurements of Schurch [16].

A typical result is shown in Fig. 4. In Fig. 4a, the thickness of the liquid layer is shown as a function of time, demonstrating the transition from an initially uniform layer (perturbation

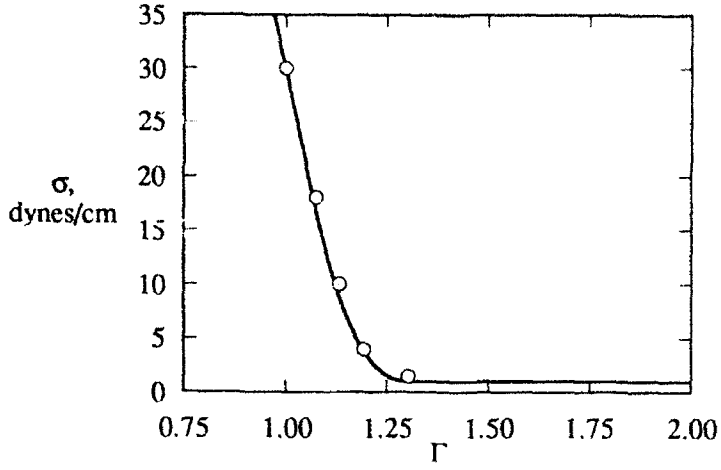


Fig. 3 Surfactant equation of state for lung surfactant [16] used in the calculations of airway closure. Minimum surface tension is taken to be 1 dyne/cm. Surfactant concentration is shown here normalized to the value at $\sigma = 30$ dyne/cm which corresponds to 2.3×10^{18} molecules/m^2.

amplitude $0.001a_0$) at t = 0, through the growth of the instability and ultimate closure. For times less than about 400 msec the layer remains fairly uniform, but thickens due to the reduction in airway size. For t > 400 msec, the instability rapidly grows and liquid is drawn from the ends of the tube toward the center until at t = 436 msec, the protrusion at x = 0 reaches the centerline and closure occurs. It should be noted that, in all these simulations, the approximations made in the analysis break down in the final stage of closure, but this stage passes so quickly that it has relatively little influence on the predicted closure times [3].

For the case shown, the liquid film is at first stable to small perturbations and the initial disturbance begins to *relax* toward a cylindrical surface. By the time lung volume falls to 0.89TLC, however, the film has thickened to the point that the mean inner film radius is less than L/π and the Rayleigh instability takes over, reversing the initial tendency toward a uniform film and generating a flow from the ends toward a growing "bulge" near the center. Given sufficient time at a given lung volume, this instability would proceed to a new equilibrium configuration (an unduloid) *without* airway closure. However, as lung volume continues to fall, the ratio of liquid volume to tube volume increases above a critical value of 0.297 (at a lung volume of 0.26TLC) and, rather than forming an unduloid, a closed meniscus is the stable, minimum surface energy state. For different geometries and/or different liquid volumes, this sequence can be altered and the critical volume condition might be met before the Rayleigh instability causes the film to become unstable; in that case it is never possible to obtain a stable unduloid. These events are illustrated in the stability map of Fig. 5 where a reduction in lung volume can be viewed as moving downward along a line of fixed $L/\pi s$. For case 1, the film begins as a stable, uniform-thickness layer, passes through a region in which the stable configuration is an unduloid, and ends up forming a liquid bridge as a/s falls below a value of about 0.8. Case 2 is one in which the only instability ultimately leads to a liquid bridge, by-passing the unduloid state entirely.

The pressure distributions shown in Fig. 4b correspond to the same times and lung volumes as shown in 4a. The interesting feature here is that the minimum pressure at the wall is observed at *maximal* lung volume when the radius of curvature is *largest*. This counter-intuitive result is due to the presence of surfactant at concentrations typical of those found in the lung.

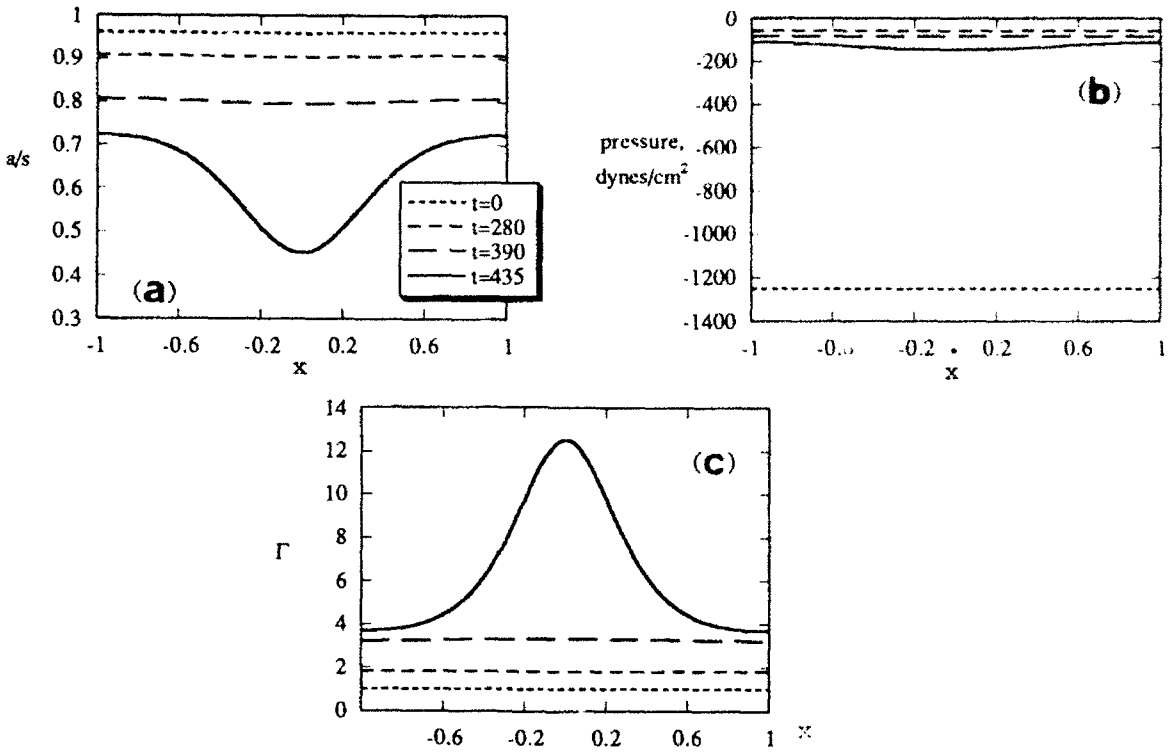


Fig. 4. Airway closure in a contracting airway with variable surface tension. At TLC, L = 0.075 cm, s = 0.025 cm, a = 0.024 cm. Fluid properties: μ = 0.01 dyne-sec/cm^2, ρ = 1g/cm^2. Expiration rate = 200% TLC/sec. (a) Film thickness at four times (in msec), (b) pressure in liquid layer at same times, (c) surfactant concentration at same times. Closure occurs at 435 msec.

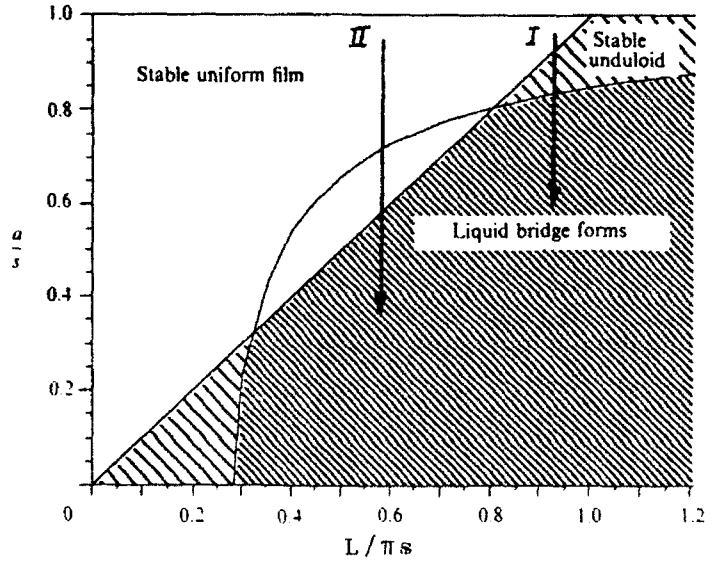


Fig. 5 Stability map for the formation of an unduloid and liquid bridge.

As the interface contracts during expiration, the surfactant concentration rises (see Fig. 4c) and the surface tension consequently falls. With sufficient surfactant present, it falls so rapidly that σ/a actually *decreases* as the airway contracts. The magnitude of the liquid pressure falls from an initial value of about 1250 dynes/cm^2 to about 100 dynes/cm^2 and remains low for the duration of expiration. The gradients in surfactant concentration remain relatively small until the final stages of closure, but are, in fact, large enough to effectively "freeze" the interface due to the Marangoni effect. Thus, closure is delayed by pulmonary surfactant as a result of two mechanisms. First, the reduction in surface tension as the gas-liquid interface becomes increasingly compressed during expiration brings about a consequent reduction in the pressure gradient driving the instability for a given displacement amplitude. This driving force scales with the surface tension and its reduction due to airway contraction is, in most cases, the dominant influence. Second, the presence of a surface tension gradient produces a shear stress at the interface (see eqn. (9)) that effectively reduces the interfacial velocity to near zero values. By itself, this slows the instability by a factor of about four -- for a given pressure gradient and film thickness, the flow rate between two stationary surfaces is lower by a factor of four than that between one stationary and one free surface. As long as the surface tension remains non-zero, neither mechanism causes an unstable condition to become stable; they only affect the *rate* at which the instability proceeds and, therefore, can influence the lung volume at which closure occurs for a given rate of expiration.

Fig. 6 shows the results from many calculations of the type just described, both with and without surfactant and with different values of surface tension. The two limiting cases (high and low values of the expiration rate) can be understood as follows. In the quasi-static limit (far to the left in the figure) closure occurs as soon as both conditions ($L > \pi a$ and $V_{liq}/s^3 > 5.6$) are first satisfied and is independent of the rate of expiration. In the limit at the right, the airway contracts so rapidly that the instability has no time to develop and closure occurs essentially when the volume of liquid equals the airway volume. In between these limits, a competition exists between the growth of the instability and the reduction in airway volume. Interestingly, the case which is thought to be most representative of the real lung (labeled case IV for expiration rates of about 20%TLC/s) fall into this transition range, suggesting that surfactant does indeed function to reduce the closing volume of the lung under normal conditions. Note

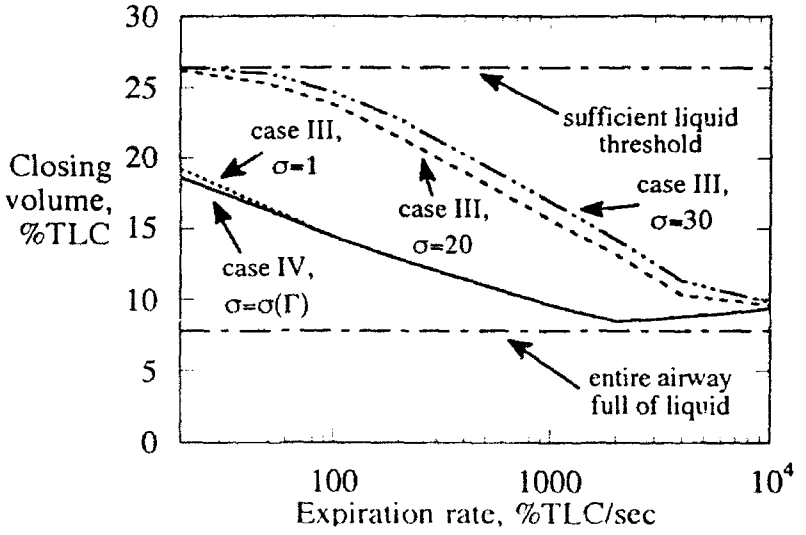


Fig. 6 Closing volume for different rates of expiration. For the three curves labled Case III, the surface tension is held constant at 1, 20 or 30 dynes/cm. Horizonatal lines show limits for slow (upper) and fast (lower) expirations.

that closing volume in a normal lung usually occurs at a lung volume of about 30%TLC, roughly consistent with these predictions. Also consistent with these predictions, lungs with insufficient surfactant (as in respiratory distress syndrome) exhibit gas trapping and atelectasis at higher than normal lung volumes and often require mechanical ventilation with high airway pressures to maintain patent airways.

Other factors, not accounted for in this model, may also be important. The airways are compliant vessels, the stiffness of which is largely determined by the tethering forces exerted on their external surface by the surrounding lung tissue (parenchyma). This tethering force scales with the inflating pressure of the lung and approaches zero for lung volumes in the range of 15%TLC [17]. Although the precise influence of airway compliance in the real lung is not known, the effects of compliance have been incorporated into a model of airway closure [18] and demonstrate that compliance hastens the closure process and would thus cause it to occur at somewhat greater lung volumes.

5. SPREADING OF AN EXOGENOUS SURFACTANT

The geometry used to study the axial dispersion of an instilled bolus of exogenous surfactant is shown in Fig. 1b. In this model, a linear equation of state of the form, $\sigma = \sigma_{max} - \alpha\Gamma$, is employed, where σ_{max} is the maximum value of surface tension. The surfactant is deposited in an axi-symmetric distribution between x = 0 and 1 at time t = 0. This introduces a gradient in surface tension which draws the interface to the right, dragging with it fluid in the sub-phase. As shown in Fig. 7, liquid in the vicinity of the deposited surfactant moves outward in the form of a propagating, shock-like disturbance of nearly constant amplitude. Over most of the domain, flow is generated by the (nearly linear) surface tension gradient (Marangoni convection) with capillarity associated with surface curvature playing an insignificant role. At the leading edge of the advancing front, however, the steep gradient in film thickness gives rise to large interfacial curvatures and capillarity becomes the dominant factor leading to pressure gradients in the film that generate the observed flow. Surface curvature and capillarity also play a role in the trailing region near the site of surfactant deposition, although to a somewhat lesser extent. A simple scaling analysis, confirmed by more extensive numerical simulations [11], demonstrates that the height of the surface film will approach twice the initial thickness of the liquid layer and that its leading edge will propagate at a speed that scales as $(\alpha M h_0 \mu^{-1} t^{-2})^{1/3}$.

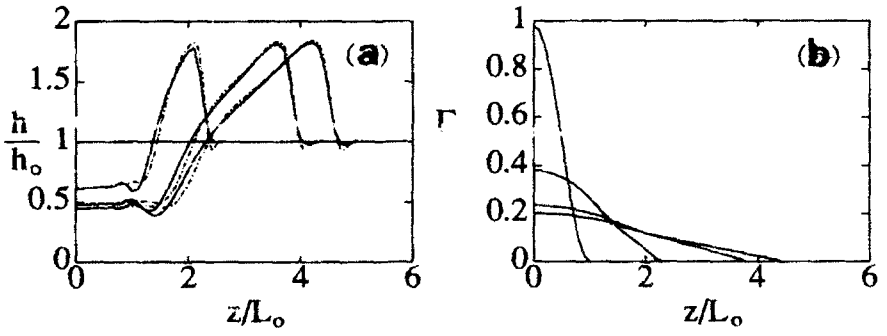


Fig. 7 Dimensionless liquid layer thickness (a) and surfactant concentration (normalized by concentration at $z/L_0 = 0$) (b) for the spreading of a surfactant from an initial deposition between $z/L_0 = 0$ and 1. Parameter values: $h_0/L_0 = 0.01$, $h_0/s = 0.01$, $H_{bol} = 0.0$, $Bo = 0.0$. Curves are for dimensionless times $t\sigma_{max}h_0/(\mu L_0^2) = 0$, 2, 10, and 16.

A refined discretization from 60 to 240 nodes dampens the waves at the trailing edge of the propagating disturbance, but has little effect on those at the leading edge. Further refinement, to 480 nodes, produces no further change leading us to conclude that while the wave structure at the back is a numerical artifact, the precursor wave structure is real and is presumably associated with the dispersive character of the equation set. Subsequent results were obtained using the coarser grid since this resolution seems adequate to capture the primary spreading behavior.

The parameters of this problem include the ratio of initial film height to airway radius (h_0/s), the ratio of initial film height to the axial extent of the region of deposited surfactant (h_0/L_0), the initial surface tension difference, the Bond number representing the ratio of gravitational to surface tension effects, and the amount of bulk liquid deposited on the film in a bolus along with the surfactant, characterized by the parameter H_{bol} as defined by the following expression for the initial film thickness distribution:

$$\frac{h}{h_0} = 1 + H_{bol}\cos^2\left(\frac{\pi}{2}\frac{z}{L_0}\right) \tag{13}$$

for $-1 < z/L_0 < 1$ and $h/h_0 = 1$ elsewhere. Each of these was varied independently in order to ascertain their effect on spreading behavior.

Simulations in which h_0/s was varied between 0.001 and 0.04 (the range of values anticipated in the lung) showed no significant differences from the behavior depicted in Fig. 7. Airway curvature therefore plays essentially no role in surfactant spreading. Consequently, models in which the effects of airway curvature are neglected would seem perfectly appropriate for most cases.

Variations in h_0/L_0 proved to be somewhat more significant as shown in Fig. 8. When expressed in this non-dimensional format, the main influence of h_0/L_0 is on the shape of the propagating disturbance rather than its speed of propagation with larger values giving rise to a generally more rounded appearance and a broader, less well-defined shock zone. In dimensional terms, since time is normalized by $\mu L_0^2/(\sigma_{max} h_0)$ the rate at which the wave propagates is proportional to h_0/L_0^2; thinner films or larger regions of surfactant coverage tend to reduce the rate of spread.

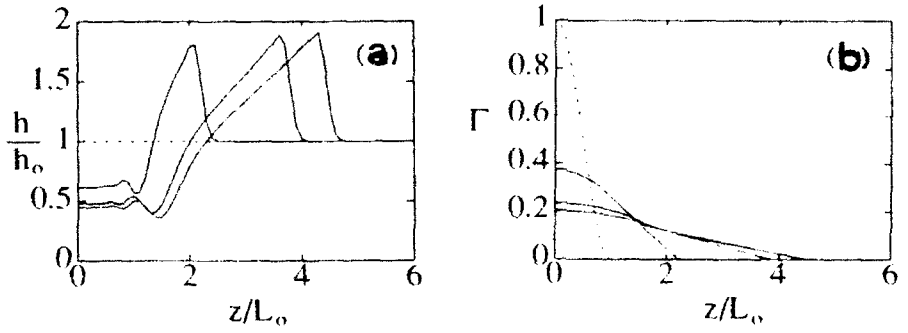


Fig. 8 Dimensionless liquid layer thickness (a) and surfactant concentration (b) showing the effect of changing h_0/L_0. Parameter values: are the same as in Fig. 7 except $h_0/L_0 = 0.01$. Curves are for dimensionless times $t\sigma_{max}h_0/(\mu L_0^2) = 0$, 2, 10, and 16.

Given this dependence of propagation speed on film thickness, it seemed reasonable to attempt to augment the rate of transport by introducing the surfactant in a liquid volume sufficient to significantly increase the underlying film thickness. However, when bolus volumes were increased (Fig. 9), we found that the rate of propagation was virtually unaffected.

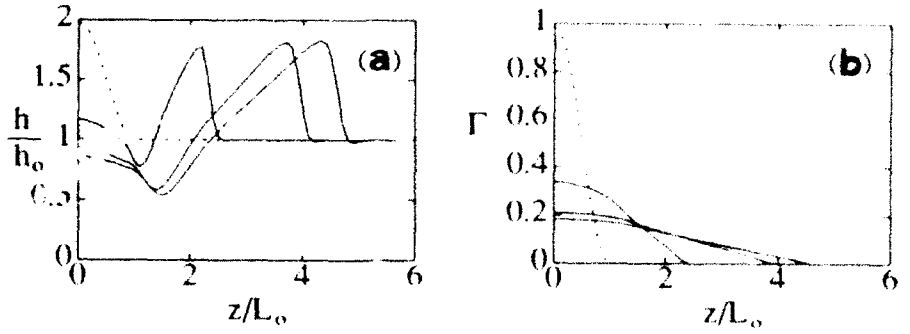


Fig.9 Dimensionless liquid layer thickness (a) and surfactant concentration (b) showing the effects of changing bolus volume. Parameter values are the same as in Fig. 7 except H_{bol} = 1.0. Curves are for dimensionless times $t\sigma_{max}h_0/(\mu L_0^2)$ = 0, 2, 10, and 16.

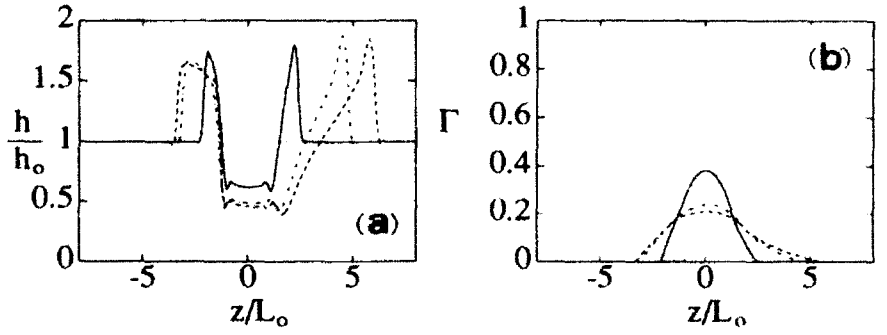


Fig. 10 Dimensionless liquid layer thickness (a) and surfactant concentration (b) showing the effects of changing bolus volume. Parameter values are the same as in Fig. 7 except for Bo = 0.1. Curves are for dimensionless times $t\sigma_{max}h_0/(\mu L_0^2)$ = 0, 2, 10, and 16.

Moreover, the additional instilled volume appears to have remained in a central pool at the site of deposition which was proportionately of greater depth than the original case (compare Figs. 7 and 9).

Gravity was also considered by choosing values of the Bond number (Bo = 0.001 to 0.1) corresponding to a vertically oriented airway ranging in size from a terminal bronchiole to the trachea. Only in the tracheal simulation (Bo = 0.1) did gravity exert a significant influence, causing the spread to be noticeably asymmetric as shown in Fig. 10. Even in this extreme case, the rate of spread of the surfactant was only slightly increased, suggesting that, at least for relatively small volumes of deposited surfactant, gravity plays a minor role. This is clearly not the case, however, in experiments that have been reported in which the volume of the instilled bolus is comparable to that of the entire central airway system [19] in which case other factors, not represented in these simulations, take precedence.

Lastly, we investigated the perhaps more realistic situation of placing an exogenous surfactant on a thin film containing an endogenous surfactant layer (Fig. 11). In this case, the

propagating wave compresses the interface ahead of it, producing a secondary gradient in surface tension due to increased concentration of the *endogenous* surfactant. This generates flow away from the deposition site and toward more remote locations. By this mechanism, the effect of the instilled surfactant spreads much more rapidly (compare Figs. 7 and 11). However, because the height of the disturbance is correspondingly reduced, the amount of wall liquid transported away from the deposition site is actually less than when there is no endogenous surfactant. This finding is also true in cases when the volume of the bolus is increased with endogenous surfactant present (results not shown).

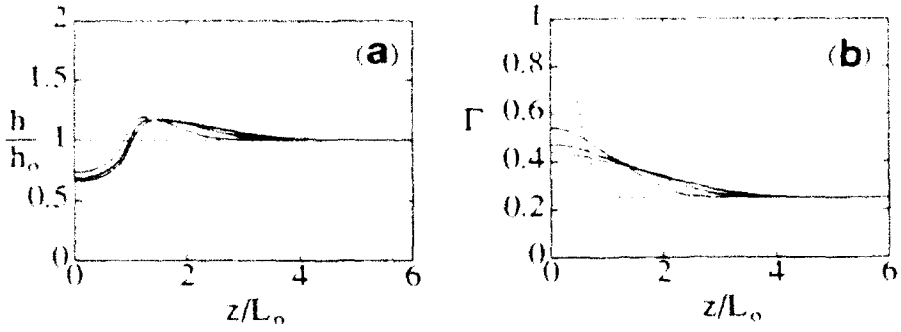


Fig. 11. Dimensionless liquid layer thickness (a) and surfactant concentration (b) showing the effects of changing bolus volume. Parameter values are the same as in Fig. 7 except for the presence of surfactant on the initial interface. Curves are for dimensionless times $t\sigma_{max}h_0/(\mu L_0^2) = 0$, 1, 2 and 3.

These results can be used to predict the rate of spread of a bolus of surfactant deposited in the trachea of a newborn infant. In the absence of endogenous surfactant, and using the parameter values given in the caption, surfactant would traverse the entire length of the trachea (1.5 cm) in just 0.2 seconds. If we neglect the influence of bifurcations, the action of other mechanisms such as the beating of the cilia in propelling liquid toward the mouth, and the rapid increase in surface area per unit length of airway with distance toward the periphery, we calculate that this front would reach the alveolar zone in about 7 seconds. Such transport rates help to explain why surfactants instilled in the trachea of infants produce so favorable an effect, and also argues for the use of surfactants in the delivery of other medications that need to be drawn into the lung.

6. CONCLUSION

In this paper we have examined two problems pertaining to the stability and mobility of the airway liquid lining as a means of raising a variety of intriguing and significant fluid dynamic questions. As the lung and airway network expand and contract during normal breathing the thickness of the liquid layer varies and can, if sufficiently thick, become unstable at low lung volumes by a mechanism analogous to the Rayleigh instability of a liquid jet and form a meniscus that blocks the airway. If this blockage persists, gas absorption peripheral to the meniscus produces regional atelectasis and a loss of gas exchange capability. Since excess airway liquid and atelectasis are common manifestations of respiratory disease, this phenomenon may be critical. If atelectasis is to be avoided, the airways, if closed, must reopen on inspiration. This raises questions concerning the movement and eventual elimination of the liquid meniscus.

A related set of fluid dynamic issues is raised by the presence of pulmonary surfactant which is produced in the gas exchange region of the lung but found in varying amounts throughout the airway tree. In lungs with normal surfactant production, a surface tension gradient exists that has the capability of drawing liquid from the periphery toward the mouth by the action of surface tension gradients. The importance of pulmonary surfactant is underscored by the prevalence of pulmonary diseases in which surfactant production is impaired such as neonatal or adult respiratory distress syndrome (RDS). Instillation of exogenous surfactant is now the accepted treatment for neonatal RDS with generally excellent results. For a small but significant subgroup, however, this therapy is ineffective. Since the efficacy of this method may depend on how effectively surfactant is transported to the lung periphery, studies are now underway to determine how an exogenous surfactant spreads in the lung. The spreading is effected by Marangoni flows, raising the exciting prospect of using surfactants as a vehicle for transporting other medications to the periphery where they can be readily absorbed.

The solutions presented here address but a few of the intriguing fluid dynamic problems relating to the dynamics of the pulmonary liquid lining layer. Studies relating to the formation, movement, and reopening of liquid menisci are essential to further our understanding of these phenomena in healthy and diseased lungs. Important questions remain concerning the factors that regulate the distribution of liquid between the vascular, interstitial and airway compartments. The movement of airway liquid through a multiple-generation network due to surface tension gradients, either due to endogenous or exogenous surfactant needs further study. In particular, the potential role of instilled surfactant as a vehicle for drug delivery warrants investigation.

Acknowledgements

Many individuals have made significant contributions to the work presented in this paper. They include M. Johnson, J. Haid, F. Espinosa, T.J. Pedley, J.J. Fredberg, A.H. Shapiro, D.R. Otis. *This work was supported by grants from the National Heart, Lung, and Blood Institute, HL33009 and HL45679.*

NOMENCLATURE

a	=	radius to free surface, cm
A	=	$(s^2 - a^2)/2$, cm^3
Bo	=	Bond number, $\rho g h_0 L_0/\sigma$
D_s	=	surfactant surface diffusion coefficient, cm^2/s
g	=	acceleration due to gravity, cm/s^2
h	=	liquid layer thickness, (s-a), cm
L_0	=	initial monolayer half length, cm
L	=	airway half-length, cm
M	=	mass of surfactant per unit circumference, g/cm
p	=	pressure in liquid film, $dyne/cm^2$
p_{gas}	=	gas pressure in airway, $dyne/cm^2$
s	=	radius to airway wall, cm
t	=	time, s
u	=	radial velocity of liquid layer, cm/s
TLC	=	total lung capacity
T_{exp}	=	time in which the lung is emptied from TLC, s
V_a	=	$w_a - x\dot{L}$ = surface axial flow velocity relative to airway wall, cm/s
$\bar{V}$	=	$\bar{w} - x\dot{L}$ = mean axial flow velocity relative to airway wall, cm/s
w	=	axial velocity of airway film, cm/s
$\bar{w}$	=	mean axial velocity of airway film, cm/s
x	=	z/L

z = axial coordinate in airway, cm

α = slope of the linear equation-of-state for σ vs. Γ, cm^2/s^2
γ = $[1 + ((1/L)\partial a/\partial x)^2]^{-1/2}$
Γ = dimensionless surfactant concentration
μ = viscosity of airway liquid, dyne·s/cm^2
ρ = density of airway liquid, g/cm^3
σ = surface tension, dynes/cm
σ_{max} = maximum value for σ

Subscripts

a ~ (r=a)
s ~ (r=s)

REFERENCES

1 L. Rayleigh, London Math Soc. Proc. 10 (1879) 4.
2 P.S. Hammond, J. Fluid Mech., 137 (1983) 363.
3 M. A. Johnson, R. D. Kamm, L. W. Ho, A. H. Shapiro and T. J. Pedley, J. Fluid Mech., 233 (1991) 141.
4 W. C. Hulbert, B. B. Forster, W. Laird, C. E. Phil and D. C. Walker, Lab. Invest., 47 (1982) 354.
5 J. A. Nadel, J. H. Widdicombe and A. C. Peatfield, Vol. 1 Handbook of Physiology, Respiratory System, Chapt. 13, Waverly Press, Inc., (1986) 419.
6 S. Schurch, Resp. Phys., 48 (1982) 339.
7 P. Gehr, S. Schurch, Y. Berthiaume, V. Im Hof and M. Gerser, J. Aero. Med., 3(1), (1991) 27.
8 A. Jobe and M. Ikegami, Am. Rev. Respir. Dis., 136 (1987) 1256.
9 D. P. Gaver and J. Grotberg, J. Fluid Mech., 213 (1990) 127.
10 D. P. Gaver and J. Grotberg, J. Fluid Mech., 235 (1992) 399.
11 F. F. Espinosa, A. H. Shapiro, J. J. Fredberg and R. D. Kamm, J. Appl. Physiol. Submitted for publication.
12 V. S. Kharasch, T. D. Sweeney, J. J. Fredberg, J. Lehr, A. I. Damokosh, M. E. Avery and J. D. Brain, Am. Rev. Respir. Dis., 144 (1991) 909.
13 D. R. Otis, M. Johnson, T. J. Pedley and R. D. Kamm, J. Appl. Physiol. Submitted for publication.
14 J. M. B. Hughes, F. G. Hoppin and J. Mead, J. Appl. Physiol., 32 (1972) 25.
15 J. R. Hammersley, B. Snyder and D. E. Olson, Ad. Bioeng: ASME (1986) 66.
16 S. Schurch, H. Bachofen, J. Goerke and F. Possmayer, J. Appl. Physiol., 67(6) (1989) 2389.
17 F. G. Hoppin, J. C. Stothert, Jr., I. A. Greaves, Y.-L. Lai and J. Hildebrandt, Vol. 3 Handbook of Physiology. The Respiratory System, Chapt. 13 (1986) 195.
18 D. Halpern, J. B. Grotberg, J. Fluid Mech., (1992). In press.
19 N. Gilliard, P. M. Richman, T. A. Merritt and R. G. Spragg, Am. Rev. Respir. Dis., 141 (1990) 743.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.

© 1993 IUTAM. All rights reserved.

ENERGY-SAVING MECHANISMS IN ANIMAL MOVEMENT

R. McN. Alexander

Department of Pure and Applied Biology, University of Leeds, Leeds LS2 9JT, U.K.

Abstract

Animals use legs and flapping wings instead of the wheels and rotors of man-made vehicles. Some largely avoid the energetic penalties of using reciprocating instead of rotary motion by means of springs which store the kinetic energy that is lost, as the body or a limb decelerates, and return it in an elastic recoil. Pendulum mechanisms involving exchange of kinetic energy with gravitational potential energy provide alternative means of saving energy at low speeds.

1. INTRODUCTION

Energy saving is important to animals, in many ways. An animal may be unable to obtain more than a limited amount of food (and so energy) each day. Alternatively, more food may be available only at the cost of taking risks, or of having less time available for other activities such as reproduction. An animal that moves economically may be able to run faster or further for the same energy cost, so may be better at catching prey or escaping from enemies. The structures and patterns of movement on which economy of energy depends are largely inherited, so can be modified by evolution - other things being equal, animals that are more economical of energy are more likely to survive and have offspring. This paper shows some of the many energy-saving devices that have evolved in animals.

Unfortunately, we have to start with an awkward complication. It is often relatively easy to predict from engineering mechanics how a change in structure or pattern of movement will affect the mechanical power needed for locomotion. However, the costs that drive evolution are not work, but fuel (food) consumption. We want to know the effects of changes not on mechanical power but on metabolic power, the rate at which food energy (measured as heat of combustion) is used.

We might hope to be able to calculate metabolic from mechanical power by dividing by some constant efficiency, but that approach would lead, in some cases, to serious errors. Metabolic energy must of course be used when muscles do work (conservation of energy requires it) but it is also used when a muscle exerts tension while maintaining constant length and so doing no work. Muscle consists of an array of inter-

digitating protein filaments that can slide past each other, allowing it to lengthen and shorten like a telescope. It is made to shorten by crossbridges between the filaments which attach, pull and detach like a team of men pulling in a rope hand-over-hand. These cycles of attachment and detachment continue as long as tension is being exerted, whether the muscle is shortening or not, and each cycle costs energy.

The relationship between mechanical performance and metabolic energy cost is imperfectly understood (Ma & Zahalak, 1991; Alexander, 1991), but to a first approximation we can think of metabolic cost as the sum of two components:
(i) the cost of work, proportional to the work performed; and
(ii) the cost of force, proportional to the number of active crossbridges multiplied by the number of cycles made by each of them. The number of crossbridges is proportional to the volume of muscle that has to be activated to exert the force, and the rate at which they cycle is higher in muscles that can shorten fast than in muscles that can contract only slowly. Thus evolution can be expected to favour changes that reduce the work that muscles have to do, that reduce the volume of muscle involved, or that enable the animal to use slower, more economical muscles.

2. RUNNING

The energy cost of running is particularly interesting for the paradoxical reason that very little net work is needed, to run at constant speed over level ground. Work is of course required to overcome aerodynamic drag, and friction in the joints, but the quantities are only tiny fractions of the metabolic energy requirement (Alexander, 1977). Additional energy costs arise because animals travel on legs rather than wheels. It is difficult to imagine how wheels could have evolved but La Barbera (1983) has argued that legs are in any case preferable largely because of their capability of travel over irregular terrain.

To understand the additional costs of travel on legs we must take account not only of the positive work that muscles do when they shorten while exerting force, but also of the negative work done when they are forcibly lengthened, degrading mechanical energy to heat. One muscle in a leg may do positive work while another does negative work. Alternatively, muscles may do positive work at one stage of the stride and negative work at another. In either case, the sum of all the quantities of positive work performed in a journey may greatly exceed the net work requirement. Similarly, the fuel consumption of your car is increased if you drive with the brakes on.

The most obvious additional cost of travel on legs arises because wheels rotate but legs reciprocate. The kinetic energy of an animal (or of any other system of masses) can be thought of as the sum of external kinetic energy associated with the

movement of the centre of mass and internal kinetic energy associated with movement of parts relative to the centre of mass (see, for example, Alexander, 1988). The internal kinetic energy of a wheeled vehicle may be constant, when it travels at constant speed, but that of a legged animal rises and falls as its legs swing backward and forward. Positive work must be done on each leg twice in each stride, and negative work must be done twice, to keep it swinging. The quantities involved can be reduced by reducing the moment of inertia of the limb about its proximal end. Specialised running animals such as horses and antelopes have large muscles in the upper parts of their legs, where their mass contributes little to the moment of inertia, while their feet are slender and correspondingly light.

A pendulum can be kept swinging with very little energy input, because potential energy changes compensate for the fluctuations of kinetic energy. Legs may swing forward passively in walking (Mochon & McMahon, 1980; McGeer, 1992) but at higher speeds they have to be driven well above their pendulum frequency. We will examine later the possibility that legs may be spring-mounted, so as to vibrate at the appropriate frequency, but it will be convenient to leave that until after discussing other, more firmly established, elastic mechanisms.

We have been discussing energy costs that arise because legs have mass and make reciprocating movements. We turn now to other energy costs. Fig. 1 shows some simple wolking devices which will be assumed in the discussion that follows, to have legs of zero mass.

In Fig. 1(a) each leg has two joints, powered by appropriate actuators : a hinge joint at the proximal end, which enables the leg to be pointed in any required direction within a plane, and a sliding joint which allows changes of leg length. It might have seemed more realistic to have drawn a leg with two hinges (a "hip" and "knee") but the arrangement shown allows the same number of degrees of freedom of movement. Suppose that this device walks over level ground keeping the forces on its feet vertical and equal to body weight. Suppose also that the moment of inertia of its body is so high that the amplitude of pitching movements is negligible. In that case, kinetic and potential energy will remain constant throughout the stride and no net work will be required at any stage, except to replace frictional losses. However, at the stage shown in Fig. 1(a)(i) the hinge actuator is doing positive work and the sliding joint actuator negative work; while at the stage shown in Fig. 1(a)(ii) the reverse is the case. Similarly, if a leg with hip and knee joints walked in this way, the actuator of one joint would have to do work against the actuator of the other, during much of the stride.

This wastage of energy could be avoided by having legs with two sliding joints as shown in Fig. 1(b), or by more complicated pantograph mechanisms such as have been used in some legged robots (Waldon et al., 1984), but I know no cases of animals saving energy in this way.

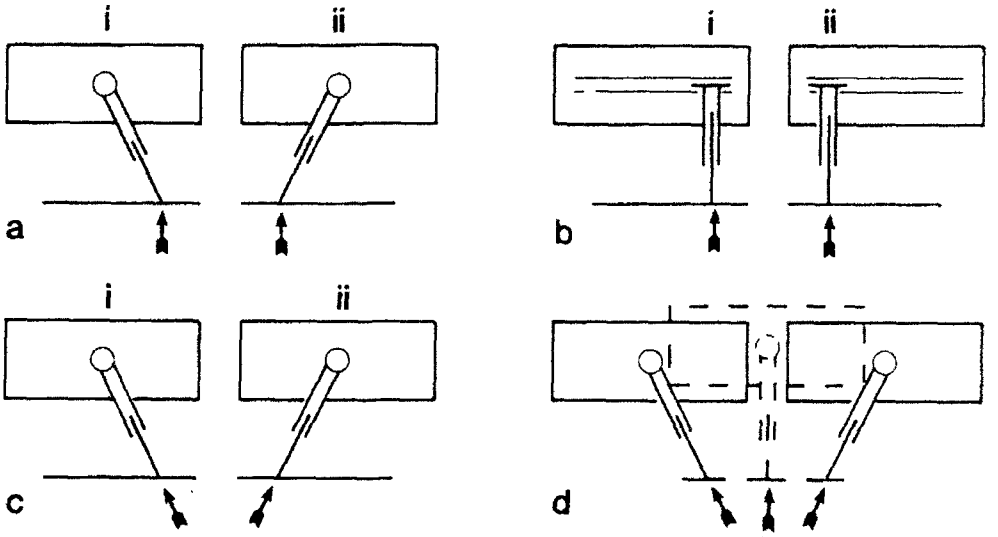


Fig. 1 Diagrams of simple walking devices

Another walking style that avoids having actuators working against each other is shown in Fig. 1(c). The torque actuator at the proximal end of the limb remains inactive, so the ground force is aligned with the leg. The sliding joint actuator does negative work at the stage of the stride shown in Fig. 1(c)(i), reducing the kinetic energy of the body; and positive work at the stage shown in Fig. 1(c)(ii), restoring the kinetic energy. Thus positive work done at one stage is counteracted by negative work done at another, but it is easily shown that the quantities of work are less than in the style of walking shown in Fig. 1(a), for all reasonable step lengths. Alexander (19>1) shows similarly for a leg with two hinge joints, that positive and negative work requirements can be minimized by aligning the ground force with the proximal joint. The forces exerted on the ground in walking and running by dogs, sheep and other vertebrates tend to be aligned with the proximal limb joints, or with points above them (Jayes & Alexander, 1978).

Even this reduced work requirement can be eliminated (leaving only the work needed to replace frictional losses) if the actuator of the sliding joint is replaced by a spring : the theory is presented by McMahon & Cheng (1990). A vehicle bouncing along on sprung legs must rise and fall, accelerate and decelerate, but the fluctuations of kinetic energy plus gravitational potential energy are balanced by fluctuations of elastic strain energy.

Some of the muscles in the legs of mammals have long tendons in series with them. These serve as springs, and have

excellent elastic properties : Bennett et al., (1986) show that when tendons are stretched to physiologically relevant stresses, and allowed to recoil, percentage energy dissipation is no more than about 7%. The forces on leg tendons and the extent to which they stretch in running, have been calculated for many mammals (Alexander, 1988) but the most thorough studies are of kangaroos, most recently by Griffiths (1989). He measured the force exerted by the gastrocnemius (calf) muscle and its changes in length, as the animal hopped. He showed that it stretched by about 12 mm as the force increased and shortened as the force fell, as if it were a passive spring. He also measured the elastic compliance of its tendon and showed that 41% of the negative and positive work was attributable to passive stretching and recoil of the tendon, leaving only 59% of the work to be done by the muscle fibres.

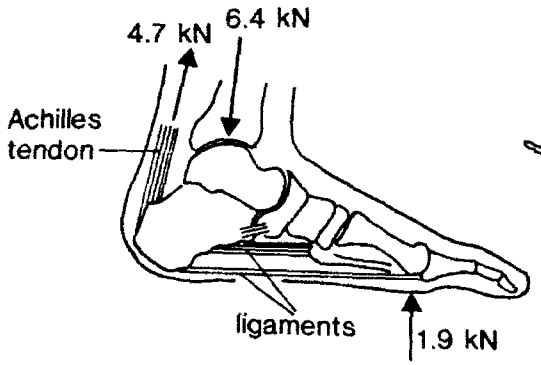


Fig. 2 The force on the foot of a human runner, calculated from force plate records.

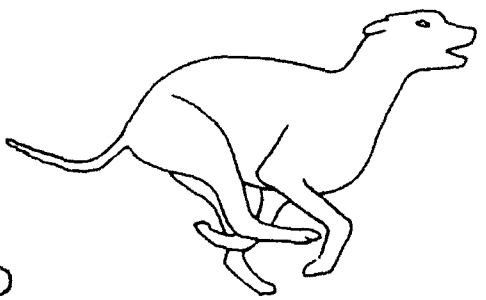


Fig. 3 At this stage of a galloping stride, all four legs change their direction of movement and energy can be saved by elastic extension and recoil of a sheet of tendon in the back.

The ligaments in the arch of the human foot form an additional spring. Fig. 2 shows a foot at the stage of a running stride at which the forces are greatest. The ground force and its point of application (on the ball of the foot) are known from force plate records. The force in the Achilles tendon and the reaction at the ankle joint have been calculated. These three forces flatten the arch, storing strain energy in the ligaments. Experiments with amputated feet (Ker et al., 1987) show that the arch is a reasonably good spring, with about 20% energy dissipation, and enable us to calculate the strain energy stored in it. It seems that of the (kinetic plus potential) energy lost and regained in each step, about 35% is

stored as elastic strain energy in the Achilles tendon and about 17% in the arch of the foot. Thus half of the negative and positive work requirement is accommodated by deformation and recoil of springs, leaving only half to be supplied (at metabolic energy cost) by muscles. The muscles that attach to the human Achilles tendon have fibres about 50 mm long, but muscles associated with tendon springs in more specialized running mammals have much shorter fibres. For example one of the digital flexor muscles in the fore leg of the horse has muscle fibres only about 3 mm long attached to a 700 mm tendon. Analysis of films of galloping horses shows that this muscle-plus-tendon lengthens and shortens by about 50 mm in each stride (Dimery, Alexander & Ker, 1986). The vestigial muscle fibres can contribute very little to this length change, which must be almost entirely due to passive stretching and recoil of the tendon. The advantage of the reduced size of the muscle fibres is presumably that very little metabolic energy is needed to develop tension in them.

An alternative means of saving energy is to balance kinetic energy fluctuations against gravitational potential energy fluctuations. In Fig. 1(d) the length of the leg is kept constant and the ground force is kept aligned with the leg. At the illustrated stages of the stride, kinetic and potential energy changes are balanced and no work is required of either actuator. However, work is required when one foot is lifted and the other set down (McGeer, 1992). Alexander (1993) has shown that gaits involving exchange of kinetic with gravitational potential energy can be economical at low speeds, and that human walking is close to making the best possible use of this principle.

I return now to the kinetic energy changes associated with forward and backward swinging of the legs. When quadrupeds trot, each fore leg moves in synchrony with the diagonally opposite hind leg. At higher speeds mammals gallop, swinging the fore legs back while the hind legs swing forward and vice versa. Fig. 3 shows a stage of the strike at which all four legs change their directions of movement and internal kinetic energy passes through a minimum. Alexander, Dimery & Ker (1985) have shown that much of the lost kinetic energy may be stored as elastic strain energy in a sheet of tendon in the back, saving metabolic energy. There is another stage of the stride at which the legs are at the opposite extremes of their ranges of movement, at which it would again be useful to store kinetic energy taken from them, as elastic strain energy. Bennett (1989) has identified a sheet of tendon in the thigh as a possible store for some of this energy.

3. FLIGHT

When insects, small birds and bats hover they beat their wings back and forward in a near-horizontal plane, driving air downwards in the manner of a helicopter rotor. When they fly

fast they beat their wings up and down, adjusting their angle of attack in such a way as to supply not only the upward force needed to counteract the body's weight but also the thrust needed to overcome drag. In either case, we have a reciprocating motion : kinetic energy must be supplied to the wings twice in each wing beat cycle, and removed twice. This may add little or nothing to the work required of the muscles, if the inertial work required to accelerate the wings is small compared to the aerodynamic work required for flight. Work done to accelerate the wing at the beginning of a stroke may be recovered as useful work done against the aerodynamic forces that decelerate it at the end of the stroke. However, if the inertial work requirement is large, some of it may have to be dissipated by muscles doing negative work at the end of the stroke.

The inertial work is relatively large in slow and hovering flight. Ellington (1984) calculated for hovering by various insects that the ratio of inertial to aerodynamic work lay between 0.7 and 3. In such cases, worthwhile energy savings can be made by elastic mechanisms. This could be done by placing springs of appropriate stiffness either in parallel or in series with the muscles. Springs in parallel would reduce the forces required of the muscles without affecting their length changes while springs in series would do the reverse. Bennett, Ker & Alexander (1987) show that the two cases require springs of different stiffness. Weis-Fogh (1960) described parallel springs in the wing hinges of locusts and a series spring on a small muscle in dragonflies. However, in advanced insects (bees, flies, bugs, beetles etc.) the muscles themselves serve as springs. This was demonstrated by Machin & Pringle (1959), who showed that muscles from such insects, mounted in apparatus that simulated the loads that would act on them in flight, could be made to behave like springs with negative hysteresis : instead of dissipating energy in each cycle of extension and shortening, they did work in each cycle.

The cross-bridges seem to be the springs in these insect muscles. In Machin & Pringle's (1959) experiments the muscle fibres were made to lengthen and shorten by about 5% (peak to peak). Such small length changes might be accommodated by stretching of cross-bridges, without detachment : when a cross-bridge detaches any strain energy stored in it is lost. The muscles are believed to work over similarly small length ranges in the intact flying insect. Because the range is small, a large volume of muscle has to be activated to supply the work for each wing beat, so it may seem doubtful whether metabolic energy is saved in comparison with possible alternative arrangements that would not exploit muscle elasticity.

The wing muscles of birds are arranged differently from those of the insects we have been discussing, with much large velocity ratios. Consequently, they have to make length changes of about 20% peak to peak (Cutts, 1986) and there is little scope for exploiting the elastic properties of cross bridges. Pennycuick & Lock (1976) suggested that useful energy savings

might be made by storage of strain energy in bent feathers, but Alexander (1988) argued that the mechanism could not work when it would be most useful, in hovering. Jenkins, Dial & Goslow (1988) took X-ray cine film of a starling flying in a wind tunnel and showed that the width between the shoulders increases and decreases in each wing beat cycle. They remarked that the furcula (wishbone) was behaving like a spring, but it is easy to show from data in their paper that the strain energy stored in it is only about 5% of the inertial work required to accelerate the wings. So far, no spring has been identified that could have an important energy-saving role in bird flight.

4. SWIMMING

Fishes swim by beating their tails from side to side, and whales beat their tail flukes up and down. Inertial work must be done at the beginning of each stroke, and there may be scope for energy saving by elastic storage. Bennett, Ker & Alexander (1987) thought that the long tendons in the tails of dolphins might serve as in-series springs, but their (admittedly very rough) calculations indicated that they were too compliant to save energy. Wainwright, Vosburgh & Hebrank (1978) suggested that the mesh of helically-wound collagen fibres in fish skin might be a useful spring, but its geometry makes it difficult to estimate how much strain energy is stored (Alexander, 1987) and the possibility remains open.

REFERENCES

Alexander, R.McN. (1977) Terrestrial locomotion. In R.McN. Alexander & G. Goldspink (edit.) Mechanics and Energetics of Terrestrial Locomotion. 168-203.

Alexander, R.McN. (1987) Bending of cylindrical animals with helical fibres in their skin or cuticle. Journal of Theoretical Biology. 124, 97-110.

Alexander, R.McN. (1988) Elastic Mechanisms in Animal Movement. Cambridge : Cambridge University Press.

Alexander, R.McN. (1991) Energy-saving mechanisms in walking and running. Journal of Experimental Biology. 160, 55-69.

Alexander, R.McN. (1992) A model of bipedal locomotion on compliant legs. Philosophical Transactions of the Royal Society B (in press).

Alexander, R.McN., Dimery, N.J. & Ker, R.F. (1985) Elastic structures in the back and their role in galloping in some mammals. Journal of Zoology. A 207, 467-482.

Bennett, M.B. (1989) A possible energy-saving role for the major fascia of the thigh in running quadrupedal mammals. Journal of Zoology. 219, 221-230.

Bennett, M.B., Ker, R.F. & Alexander, R.McN. (1987) Elastic properties of structures in the tails of cetaceans (Phocaena and Lagenorhynchus) and their effect on the energy cost of swimming. Journal of Zoology 211, 177-192.

Bennett, M.B., Ker, R.F., Dimery, N.J. & Alexander, R.McN. (1986) Mechanical properties of various mammalian tendons. Journal of Zoology A 209, 537-548.

Cutts, A. (1986) Sarcomere length changes in the wing muscles during the wing beat cycle of two bird species. Journal of Zoology A 209, 183-185.

Dimery, N.J., Alexander, R.McN. & Ker, R.F. (1986) Elastic extension of leg tendons in the locomotion of horses (Equus caballus). Journal of Zoology A 210, 415-425.

Ellington, C.P. (1984) The aerodynamics of hovering insect flight. VI. Lift and power requirement. Philosophical Transactions of the Royal Society B 305, 145-181.

Griffiths, R.I. (1989) The mechanics of the medial gastrocnemius muscle in the freely hopping wallaby (Thylogale billardierii). Journal of Experimental Biology. 147, 439-456.

Jayes, A.S. & Alexander, R.McN. (1978) Mechanics of locomotion of dogs (Canis familiaris) and sheep (Ovis aries). Journal of Zoology. 185, 289-308.

Jenkins, F.A., Dial, K.P. & Goslow, G.E. (1988) A cineradiagraphic analysis of bird flight : the wishbone in starlings is a spring. Science 241, 1495-1498.

Ker, R.F., Bennet, M.B., Bibby, S.R., Kester, R.C. & Alexander, R.McN. (1987) The spring in the arch of the human foot. Nature. 325, 147-149.

La Barbera, M. (1983) Why the wheels won't go. American Naturalist. 121, 395-408.

Ma, S. & Zahalak, G.I. (1991) A distribution-moment model of energetics in skeletal muscle. Journal of Biomechanics. 24, 21-35.

Machin, K.E. & Pringle, J.W.S. (1959) The physiology of insect fibrillar muscle. II. Mechanical properties of a beetle flight muscle. Proceedings of the Royal Society. B 151, 204-225.

McGeer, T. (1992) Principles of walking and running. In RMcN. Alexander (editor) Mechanics of Animal Locomotion. 113-139, Berlin, Springer.

McMahon, T.A. & Cheng, G.C. (1990) The mechanics of running : how does stiffness couple with speed? Journal of Biomechanics. 23, suppl. 1, 65-78.

Mochon, S. & McMahon, T.A. (1980) Ballistic walking : an improved model. Mathematical Biosciences. 52, 241-260.

Pennycuick, C.J. & Lock, A. (1976) Elastic energy storage in primary feather shafts. Journal of Experimental Biology. 64, 677-689.

Wainwright, S.A., Vosburgh, F. & Hebrank, J.H. (1978) Shark skin : function in locomotion. Science. 202, 747-749.

Waldron, K.V., Vohnout, V.J., Pery, A. & McGhee, R.B. (1984) Configuration design of the Adaptive Suspension Vehicle. International Journal of Robotics Research. 3, 37-48.

Weis-Fogh, T. (1960) A rubber-like protein in insect cuticle. Journal of Experimental Biology. 37, 889-906.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

CONTROLLING CHAOTIC CONVECTION

Haim H. Bau

Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA 19104-6315, USA
E-mail: bau@eniac.seas.upenn.edu

Abstract

Through the use of a feedback control strategy, one can profoundly alter the bifurcation structure of convective flows. For instance, with the use of a feedback controller, one can maintain the no-motion state in fluid layers heated from below and cooled from above at significantly higher temperature differences than the critical one corresponding to the onset of convection in an uncontrolled system. Likewise, one can maintain steady, time-independent flow under conditions in which the flow would otherwise be chaotic. Similarly, one can stabilize periodic non-stable orbits which exist in the chaotic regime of the uncontrolled system. Finally, the controller also can be used to destabilize otherwise stable flows and thus induce chaos in laminar, non-chaotic flow. In this paper, I will discuss the progress that has been made in achieving the foregoing in the convective loop and the saturated porous layer paradigms.

1. INTRODUCTION

In recent years, active control strategies have been successfully used for, among other things, noise reduction and vibration suppression. The area of active control of fluid flow is no less important from a technological point of view; for in many situations, it may be desirable to maintain flow conditions other than the naturally occurring ones. Below, I mention a few examples.

In some situations, convection suppression may be desirable. For instance, when a crystal is pulled from a liquid melt heated from below, it may be desirable to operate at Rayleigh numbers higher than the critical one at which convection normally begins and yet have no convection so as to minimize transport of contaminants from the crucible walls to the crystal.

In other situations, suppression (laminarization) of chaotic or turbulent motion and maintenance of a steady, time-independent flow may be desirable in order to minimize flow unpredictability, remove temperature oscillations which may exceed safe operational conditions, and/or reduce drag. For instance, stabilizing the Blasius profile and delaying the laminar-turbulent transition in the boundary layer flow may provide a means for reducing aircraft drag and thus obtaining significant savings in fuel cost. Likewise, suppressing oscillatory convection in crystal growth processes may lead to an improvement in crystal quality.

In yet other situations, it may be desirable to stabilize one of the time-periodic orbits embedded in the turbulent attractor in order to achieve a desired

flow behavior. Indeed, from the purely scientific point of view, the ability to stabilize and observe various non-stable, periodic orbits embedded within the chaotic attractor may help enhance our understanding of the chaotic dynamics and assist us in describing various important properties of the chaotic attractor.

In still other situations, it may be advantageous to induce chaos, under conditions in which it would not normally occur, so as to enhance mixing, heat transport or chemical reactions in combustion chambers or chemical reactors.

Surprisingly given all the foregoing, the idea of altering the bifurcation structure of flows through the use of active control has received relatively little attention. Most studies to date have focused on passive control methods.

Passive control strategies do not require a direct investment of energy and they are usually accomplished by changing the geometry of the solid boundaries. For example, mixing can be enhanced through the use of pipes with sinusoidally modulated walls rather than straight pipes or by introducing spoilers in the flow. Boundary layer transition to turbulence can be delayed by using perforated solid walls to facilitate suction (Bushnell and McGinley, 1989); and thermal convection can be suppressed by introducing buffers in cavities. Passive devices may have indirect energy effects as their presence may affect the amount of energy needed to propel the fluid or to propel an object submerged in the fluid.

In contrast, active control strategies require a direct investment of energy for actuation purposes. Examples of active control include inducement of periodic motion of the solid boundaries, suction and/or blowing, induction of acoustic pressure waves, and heating and cooling of the fluid to affect its viscosity. See, for example, Liepman et. al. (1982) and Liepmann and Nosenchuck (1982) who advocated the use of thermal actuation to suppress or enhance Tollmien-Schlichting waves in laminar boundary layer flow. Active control strategies can be further sub-classified into predetermined and feedback.

Predetermined control consists of pre-programmed actuation designed to accomplish a desired outcome. The actuation, however, is not sensitive to the fluid's behavior. For example, one may periodically modulate the motion of a cavity's boundary to induce chaotic advection and enhance mixing (Ottino, 1991). Another instance of predetermined control is the utilization of magnetic fields or time-periodic modulation of wall temperature to suppress convective processes.

In contrast, a feedback controller reacts to events occurring in the fluid so as to suppress or enhance naturally occurring instabilities. To make the concept of feedback control more concrete, let's consider the Rayleigh-Benard problem of a horizontal fluid layer heated from below and cooled from above (Fig. 1). In the classical problem, as the temperature difference between bottom and top (expressed in non-dimensional form as the Rayleigh number) increases, the state of the system changes from no-motion to time-independent motion to time-dependent, possibly chaotic, motion. These transitions occur because once the Rayleigh number exceeds certain critical values, conditions are ripe for instabilities to manifest themselves. The objective of the feedback controller is to enhance the dissipating mechanisms so as to suppress the naturally occurring instabilities. For example, Fig. 1 describes the control strategy (dealt with more fully later in the paper) we used to delay the transition from the no-motion to the time-independent motion state. In this example, a sensor measures the fluid's temperature at mid-height and the bottom temperature is slightly modified in proportion to the deviation of the sensed temperature from its desired (no-motion) value. At locations where

flow behavior. Indeed, from the purely scientific point of view, the ability to stabilize and observe various non-stable, periodic orbits embedded within the chaotic attractor may help enhance our understanding of the chaotic dynamics and assist us in describing various important properties of the chaotic attractor.

In still other situations, it may be advantageous to induce chaos, under conditions in which it would not normally occur, so as to enhance mixing, heat transport or chemical reactions in combustion chambers or chemical reactors.

Surprisingly given all the foregoing, the idea of altering the bifurcation structure of flows through the use of active control has received relatively little attention. Most studies to date have focused on passive control methods.

Passive control strategies do not require a direct investment of energy and they are usually accomplished by changing the geometry of the solid boundaries. For example, mixing can be enhanced through the use of pipes with sinusoidally modulated walls rather than straight pipes or by introducing spoilers in the flow. Boundary layer transition to turbulence can be delayed by using perforated solid walls to facilitate suction (Bushnell and McGinley, 1989); and thermal convection can be suppressed by introducing buffers in cavities. Passive devices may have indirect energy effects as their presence may affect the amount of energy needed to propel the fluid or to propel an object submerged in the fluid.

In contrast, active control strategies require a direct investment of energy for actuation purposes. Examples of active control include inducement of periodic motion of the solid boundaries, suction and/or blowing, induction of acoustic pressure waves, and heating and cooling of the fluid to affect its viscosity. See, for example, Liepman et. al. (1982) and Liepmann and Nosenchuck (1982) who advocated the use of thermal actuation to suppress or enhance Tollmien-Schlichting waves in laminar boundary layer flow. Active control strategies can be further sub-classified into predetermined and feedback.

Predetermined control consists of pre-programmed actuation designed to accomplish a desired outcome. The actuation, however, is not sensitive to the fluid's behavior. For example, one may periodically modulate the motion of a cavity's boundary to induce chaotic advection and enhance mixing (Ottino, 1991). Another instance of predetermined control is the utilization of magnetic fields or time-periodic modulation of wall temperature to suppress convective processes.

In contrast, a feedback controller reacts to events occurring in the fluid so as to suppress or enhance naturally occurring instabilities. To make the concept of feedback control more concrete, let's consider the Rayleigh-Benard problem of a horizontal fluid layer heated from below and cooled from above (Fig. 1). In the classical problem, as the temperature difference between bottom and top (expressed in non-dimensional form as the Rayleigh number) increases, the state of the system changes from no-motion to time-independent motion to time-dependent , possibly chaotic, motion. These transitions occur because once the Rayleigh number exceeds certain critical values, conditions are ripe for instabilities to manifest themselves. The objective of the feedback controller is to enhance the dissipating mechanisms so as to suppress the naturally occurring instabilities. For example, Fig. 1 describes the control strategy (dealt with more fully later in the paper) we used to delay the transition from the no-motion to the time-independent motion state. In this example, a sensor measures the fluid's temperature at mid-height and the bottom temperature is slightly modified in proportion to the deviation of the sensed temperature from its desired (no-motion) value. At locations where

the flow is hotter than usual and the fluid tends to ascend, the boundary temperature is reduced to assist in dissipating the excess heat. The reverse occurs at locations where the fluid descends. This strategy can be used to significantly increase the magnitude of the critical Rayleigh number for the onset of convection.

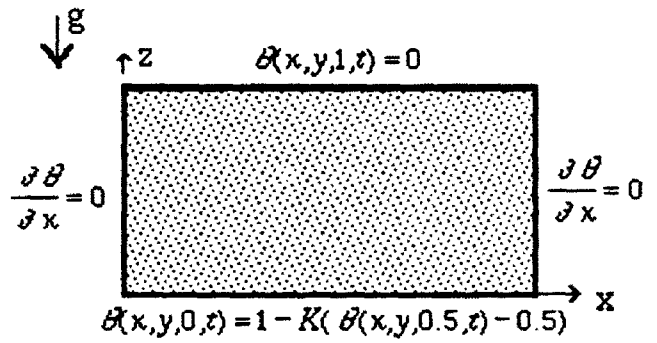


Fig. 1 A schematic description of the use of feedback control to stabilize the no-motion state in the Rayleigh-Benard problem. θ is the non-dimensional temperature; x and y are coordinates in the horizontal platform; z is a vertical coordinate; t is time; and K is the controller's gain.

For the remainder of this paper, I will focus on the use of active feedback control strategies to modify the nature of convective flows.

2. GENERAL CONCEPTS

In this section, I will present the general concepts underlying feedback control utilizing mathematical notation. First, let's consider the dynamic system (i.e., the Navier-Stokes equations),

$$\dot{\Xi} = F(\Xi;R), \tag{1}$$

where Ξ is a vector of dependent variables (i.e., velocity components, temperature, and density); superscript dot denotes the time (t) derivative; R is the parameters' vector (i.e., the Reynolds and/or Rayleigh number); and F is a non-linear (possibly differential) operator with the appropriate boundary conditions. Let $\Xi=\Xi_0$ denote an equilibrium, non-transient (i.e., time-independent, time-periodic or quasi-periodic) solution of (1). For example, Ξ_0 may be one of the non-stable periodic orbits embedded in the chaotic attractor. As R is varied, equilibrium states may lose stability at bifurcation points and give rise to new, non-transient, stable solutions. Often several successive bifurcations will eventually result in chaotic behavior. Ξ_0 may be either a stable solution for a limited range of the parameters' R (say, for $R_1<R<R_2$) or a non-stable solution for all R values. Let us assume that for a particular process, a selected Ξ_0 has better properties than the normally occurring flow structure. For example, in Rayleigh-Benard convection, Ξ_0 may represent the no-motion state which we wish to maintain under conditions in which convection would normally occur. In turbulent boundary layer flow, Ξ_0 may

represent either the laminar, time-independent state or one of the time-periodic orbits embedded in the turbulent attractor. Our objective is to render Ξ_0 stable in regions of the parameter space (R) in which it would normally be non-stable.

2.1 DIRECT APPLICATION OF FEEDBACK CONTROL

Feedback control is implemented by augmenting the system (1) with the controller C(·):

$$\dot{\Xi} = G(\Xi;R,K) = F(\Xi;R) + C(\Xi - \Xi_0;R,K), \tag{2}$$

where C(·), the controller's output, is either a linear or a non-linear operator with C(0; R, K)=0 for all R and K values; K is the controller's gain vector; (Ξ-Ξ_0) is the deviation of the measured variables from their desired values; and Ξ_0 is an equilibrium solution which may be stable for some range of the parameters R or non-stable for all R values. In the former case, the controller may be used to move bifurcation points in parameter space (i.e., delay the bifurcation to higher values of R), thus delaying the appearance of chaos. In the latter case, the controller can be used to stabilize states which otherwise would be non-stable for all values of the parameter R such as non-stable periodic orbits embedded in a chaotic attractor.

Note that $\Xi=\Xi_0$ is a solution of both the original system (1) and the augmented system (2). That is, the controller preserves the system's equilibria. In other words, the controller attempts to preserve the stability of Ξ_0 in regions of parameter space in which Ξ_0 would normally be non-stable. The proposed method of control is not the only possible one. Alternatively, one might attempt to achieve the desired flow behavior by generating a new flow structure (Ξ_N) which is an equilibrium state of (2) but not of (1), i.e., $C(\Xi-\Xi_N; R,K)\neq 0$. Such a control strategy might conceivably be necessary if none of the equilibrium states (Ξ_0) of (1) has the desired behavior. However, an attempt to induce a flow structure (Ξ_N) which is not an equilibrium solution of (1) may require a fair amount of power investment. In contrast, the stabilization of Ξ_0 in a noise-free environment requires an infinitesimal amount of power. Here, we discuss only the stabilization of equilibrium solutions.

In practice, the controller C(·) can be implemented by manipulating the system's boundary conditions. For example, one can equip some of the system's surfaces with sensors and actuators.

In relatively simple cases, the analysis of the controlled system (2) can be carried out using tools of stability theory to establish the effects of controller gains on the bifurcation structure of the system. In more difficult cases, one might utilize a "black box" approach and implement the control without knowing either the bifurcation structure of the system (1) or the mathematical model F(·). One does need, however, to have an estimate for Ξ_0.

2.2 TRAJECTORY DIRECTING

A somewhat different, but related, control strategy based on analyzing the behavior of a system's trajectories in phase space has been advanced by Ott et. al. (1990a and b). The major advantage of this technique is that it provides a procedure to implement control when the mathematical model for the system is not known or is too complex to analyze. The method has been used thus far only in low-dimensional systems. The procedure requires the identification of the non-stable periodic orbit (P), which one wishes to stabilize and which is embedded in the chaotic attractor, and its linearly stable ($E^S P$) and unstable ($E^U P$) manifolds. This can be done by using embedding to reconstruct the phase space of the attractor from a time series of one of the system's variables obtained by numerical or physical experiments. Subsequently, one can construct a return map to a selected (Poincare) cross-section obtained by stroboscopically sampling the data at the period of the orbit one wishes to stabilize. This periodic orbit is a fixed point on the map. The controller is applied in such a way as to force the system's trajectories to approach the stable manifold ($E^S P$) of the periodic orbit (P). Ott et. al. (1990) have successfully stabilized some periodic orbits of the Henon attractor by applying the controller once within each period. The same technique has been implemented experimentally by Ditto et. al. (1990) to stabilize two otherwise non-stable, periodic orbits embedded in the chaotic vibrations of a magnetostrictive ribbon.

2.3 OPTIMAL CONTROL

Optimal control involves the minimization of an appropriately defined cost function (Lions, 1971). For example, when one is applying heating and cooling for drag reduction in boundary layer flow, the cost function may be defined as:

$$O(q) = \frac{1}{T}\int_0^T dt\left(\iint_S w\, q(x,y,t)^2 dxdy + D\right), \tag{3}$$

where q(x,y) is the distribution of heating/cooling fluxes along the surface; the integration is carried out over the entire surface over time interval T; D is the drag; and w is a weight which represents the control's cost. In the case of convection suppression in the Rayleigh-Benard problem, D may represent the absolute value of the vertical velocity component integrated along a horizontal cross-section or the Nusselt number. The objective is to obtain the distribution q which minimizes the cost function O(q) subject to various constraints. Thus, the optimal control problem becomes a problem of the calculus of variations.

Depending on the system's constraints, the resulting flow structure of the controlled system may or may not be an equilibrium solution (Ξ_0) of the system (1). The problem can be formulated, however, as a feedback control problem in which $q=q(\Xi-\Xi_0)$ with $q(0)=0$.

The advantage of the "optimal control" technique is that it provides a systematic algorithm to obtain a desired state of the system and to optimize the control strategy. Unlike the method described in section 2.1, one need not select a pre-specified functional form for the operator $C(\cdot)$. The disadvantage is that the

method is computationally intensive and the minimization of the cost function requires that the system (1) be solved repetitively. This may be beyond present computational capabilities in situations in which the system (1) consists of the Navier-Stokes equations. To alleviate this problem, Choi et. al. (1992) have designed a simplified procedure which they have entitled "suboptimal control" and have applied it to the stochastic Burgers equation.

3. EXAMPLES OF THE IMPLEMENTATION OF FEEDBACK CONTROL STRATEGIES

In this section, I describe some of the work that has been done at the University of Pennsylvania on feedback control of convective flows in a thermal convection loop (Singer et. al. 1991, Singer and Bau, 1991, Wang et. al., 1991) and in a saturated porous layer heated from below and cooled from above (Tang and Bau, 1992a and b). To date, we have been able to demonstrate both theoretically and experimentally in the case of the thermal convection loop that a feedback controller, such as the one described in section 2.1, can successfully and profoundly alter the bifurcation structure of convective motions by effecting small perturbations in the boundary conditions. Currently, we are verifying that similar results can be achieved in a fluid layer heated from below and cooled from above.

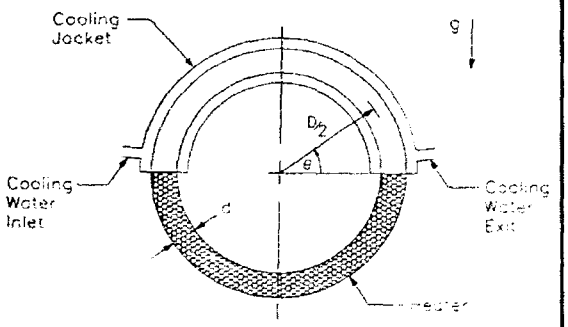


Fig. 2: Schematic description of the experimental apparatus.

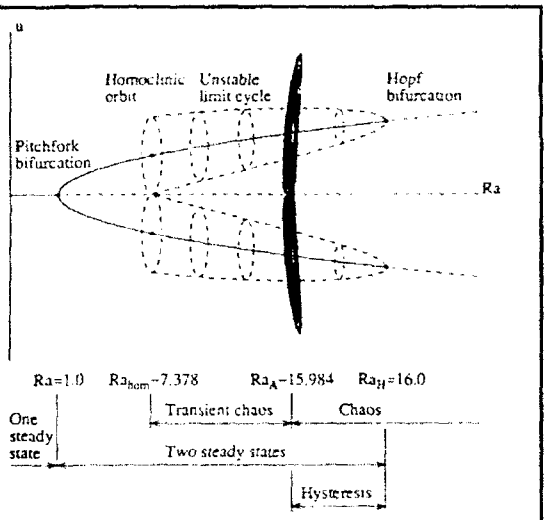


Fig. 3: An unscaled bifurcation diagram depicting various possible solutions as Ra is increased (P=4). Stable/non-stable solutions are denoted by solid/dashed lines. The dark region represents the appearance of the strange attractor which exists for $Ra > Ra_A$.

3.1 THE THERMAL CONVECTION LOOP

Consider a thermal convection loop constructed from a pipe bent into a torus and standing in the vertical plane as depicted in Fig. 2. The diameter of the pipe is d; and the diameter of the torus is D. q is the angular location of a point on the torus. The time-independent wall temperature of the pipe $T_W(q)$, which may vary with the angular location q, is symmetric with respect to the torus axis that is parallel to the gravity vector. Variations in the wall temperature may cause a spatial temperature distribution inside the fluid which, under appropriate conditions, may induce fluid motion in the loop. The convection in such loops has been investigated by, among others, Creveling et. al. (1975) and Gorman et. al. (1984, 1986).

The motion in the loop can be described within the framework of Boussinesq's approximation by using a one-dimensional model consisting of mass, momentum and energy balances (Bau and Wang, 1991):

$$u = u(t), \tag{4}$$

$$\dot{u} = \frac{1}{\pi} Ra\, P \oint T\cos(\theta)d\theta - Pu, \qquad \textit{and} \tag{5}$$

$$\dot{T} = -u\frac{\partial T}{\partial \theta} + B\frac{\partial^2 T}{\partial \theta^2} + \left[T_\infty(\theta) - T\right]. \tag{6}$$

The fluid is assumed to be incompressible and Newtonian. In the above, all quantities are non-dimensional. The length scale is the torus' radius D/2;

$Ra = \frac{g\beta\Delta T\tau^2}{DP}$ is the loop's Rayleigh number; β is the thermal expansion coefficient; g is the gravitational acceleration; and ΔT is the averaged wall temperature difference between the loop's bottom and top. The time scale is $\tau = \rho_0 C_p d/(4h)$, where ρ_0 is the fluid's average density, C_p is the thermal capacity, and h (which we assume to be constant) is the heat transfer coefficient between the fluid and the pipe's wall. $P = 32\nu\,\tau/d^2 = 8\,Pr/Nu$ is the loop's Prandtl number, where ν is the kinematic viscosity. $Pr = \nu/\alpha$ and $Nu = hd/k$ are the conventional Prandtl and Nusselt numbers, respectively. α and k are the fluid's thermal diffusivity and conductivity. $B = (d/D)^2/Nu$ is the Biot number.

The system (4-6) exhibits multiple, non-transient states $\{u_0(t), T_0(\theta,t)\}$ which may exist only for certain Ra values and whose stability characteristics may vary in accordance with the magnitude of the Rayleigh number. Our objective is to stabilize and destabilize some of these equilibrium states through the use of feedback control. Our control strategy is to measure the deviation of the actual temperature from its desired value at various locations around the loop and modify the wall temperature accordingly. For example, a convenient input signal to the controller is $\Delta T_{3\text{-}9}(t) - \Delta T_{3\text{-}9,0}(t)$, where $\Delta T_{3\text{-}9}(t)$ and $\Delta T_{3\text{-}9,0}(t)$ denote, respectively, the actual and the desired temperature differences between positions

3 and 9 o'clock around the loop. The wall temperature is modified in proportion to the above deviation, i.e.,

$$T(\theta,t)=T_{nominal}(\theta)+C[\Delta T_{3\text{-}9}(t)-\Delta T_{3\text{-}9,0}(t)], \tag{7}$$

where the controller C(·) is typically chosen to be a linear function of its argument.

To facilitate the analysis, the temperature distribution in the fluid and the wall is expanded into a Fourier series in terms of θ (Malkus, 1977).

$$T_w(\theta,t) = W_0(t) + \sum_{n=1}^{\infty} W_n(t)\,\sin(n\theta), \quad \text{and} \tag{8}$$

$$T(\theta,t) = \sum_{n=0}^{\infty} S_n(t)\,\sin(n\theta) + C_n(t)\,\cos(n\theta). \tag{9}$$

Expansions (8-9) are substituted into equations (4-6). By requiring that the equations be satisfied in the sense of weighted residuals, the infinite dimensional system (4-6) is transformed into an infinite set of ordinary differential equations. Three of these equations, similar in nature to the Lorenz (1967) equations, decouple from the rest of the set and can be solved independently with exact closure. In other words, the entire dynamics of the system can be captured by three ordinary differential equations:

$$\frac{\dot{u}}{P} = c - u, \tag{10}$$

$$\dot{c} = -us - c \tag{11}$$

$$\dot{s} = uc - s + RaW_1 \tag{12}$$

In the above, the dependence on the Biot number, B, was removed via the simple, algebraic transformation {u, c, s, *Ra*, P, t} <- 1/(1+B) {u, *Ra* C_1, *Ra* S_1, *Ra*/(1+B), P, t(1+B)^2}. The quantities c and s, respectively, represent the temperature differences in the fluid between positions 3 and 9 and 6 and 12 o'clock around the loop. W_1 represents the difference in the wall temperature between positions 6 and 12 o'clock. For the uncontrolled case, without loss of generality, we set W_1 =-1.

The three ODEs (10-12) possess a number of equilibrium, non-transient solutions, some of which are listed below:

(A) no motion state (u=0);

(B) time-independent motion, either in the clockwise (denoted B_-) or counterclockwise (denoted B_+) direction ($u=\pm\sqrt{Ra-1}$);

(C) chaotic motion; and

(D) periodic motions of various periodicities (i.e., for certain values of the Rayleigh number, one observes stable periodic solutions. Also, non-stable periodic orbits are embedded in the chaotic attractor).

Some of the above solutions and their stability characteristics are depicted schematically in Fig. 3 for a loop Prandtl number P=4 which we estimate to approximate the loop Prandtl number of our experimental apparatus (Singer, 1991). In Fig. 3, we denote stable and non-stable solutions by solid and dashed lines, respectively.

Briefly, if one were to follow the chain of events in the uncontrolled system as the Rayleigh number (Ra) increases, one would observe no net motion in the loop for $Ra<1$. At $Ra=1$, the no-motion solution loses its stability through a supercritical pitchfork bifurcation and is replaced by time-independent motion. Depending on random disturbances, this motion will be either in the clockwise (B_-) or counterclockwise (B_+) direction. The motion solution is stable for $1<Ra<16$. At Ra=16, the steady solution loses stability through a subcritical Hopf bifurcation. The resulting limit cycle is non-stable and its period increases to infinity as the Rayleigh number decreases to Ra_{hom}~7.378, where the periodic orbit becomes an homoclinic orbit and passes through the no-motion state (A). At $Ra=Ra_{hom}$, there is a bifurcation (the homoclinic explosion) which results in an assortment of non-stable periodic and non-periodic orbits known collectively as the non-wandering set which is initially non-attracting. As the Rayleigh number is further increased beyond Ra_A~15.984$\leq Ra_H$, the non-wandering set becomes a strange (the Lorenz) attractor. The chaotic regime exists for $Ra>Ra_A$ with occasional windows of periodic behavior. Non-stable periodic orbits of various periodicities are embedded in the chaotic attractor. In the chaotic regime, the motion in the loop consists of irregular oscillations with occasional reversals in the direction of the flow as shown, for example, in Fig. 3 for $Ra=3Ra_H(4)=48$. In Fig. 4, the experimentally observed temperature difference between positions 3 and 9 o'clock are depicted as functions of time. Qualitatively similar behavior is exhibited by the variable c when the equations (10-12) are integrated numerically. The positive and negative values of c in Fig. 4 correspond to flow in the counterclockwise and clockwise directions, respectively.

For example, the control (7) for suppressing chaotic motion is incorporated into the mathematical model by setting,

$$W_1 = -1 - \frac{K}{Ra}(c(t) - c_0), \tag{13}$$

where for demonstration purposes, proportional control is employed. In the above, c(t) and c_0 denote, respectively, the actual and the desired temperature differences between positions 3 and 9 o'clock.

With the use of similar control strategies, we can alter the bifurcation structure depicted in Fig. 3. More specifically, we can:

(i) maintain the no-motion state at significantly higher temperature differences between the heated and the cooled portions of the loop than the critical one corresponding to the onset of convection in the uncontrolled system (verified thus far by theoretical calculations only, Singer and Bau, 1991);

(ii) maintain steady, time-independent flow under conditions in which the flow would otherwise be chaotic (verified by experiments and theory, Singer et. al., 1991, and Wang et. al., 1992).

(iii) stabilize periodic non-stable orbits which exist in the chaotic regime of the uncontrolled system (verified thus far by theoretical calculations only, Singer and Bau, 1991); and
(iv) induce chaos in otherwise laminar (fully predictable), non-chaotic flow (verified by theory and experiments, Wang et. al., 1992).

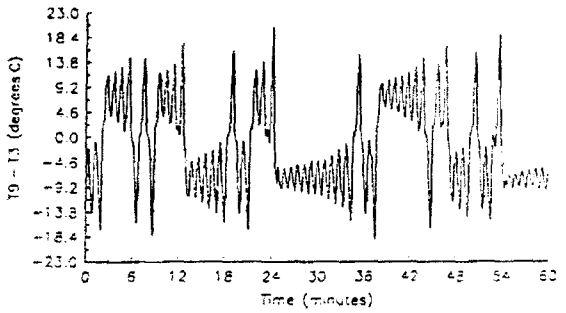


Fig. 4: The experimentally measured temperature differences $\Delta T_{3\text{-}9}$ depicted as functions of time.

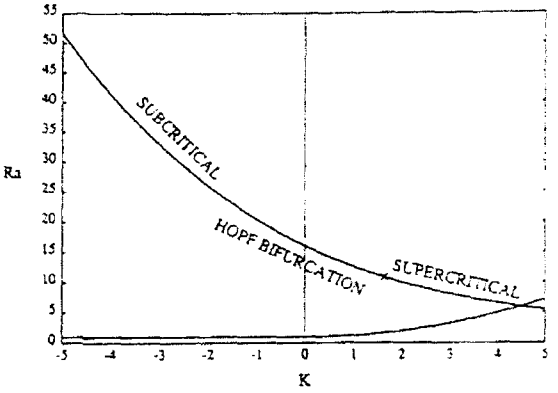


Fig. 5: Stability characteristics of the counterclockwise motion solution under proportional control. The critical Rayleigh number is depicted as a function of the controller's gain (K).

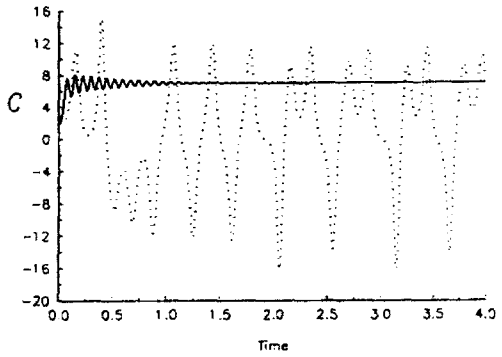


Fig. 6: The temperature difference between positions 3 and 9 o'clock is depicted for Ra=50 and P=4 as a function of time for the controlled (K=-35, solid line) and the uncontrolled (K=0, dashed line) systems.

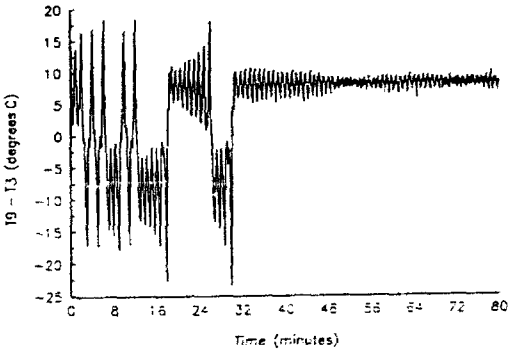


Fig. 7: The experimentally measured temperature difference $\Delta T_{3\text{-}9}$ is depicted as a function of time for similar conditions to those in Fig. 6. The controller was activated 33 minutes into the run.

For example, we stabilized the time-independent motion (B_+), by employing the control strategy described by equation (13) with $c_0(t) = \sqrt{Ra - 1}$. Fig. 5 depicts our theoretical results pertaining to the effect of the controller's gain on the location of the Hopf's bifurcation point Ra_H. Clearly, negative (positive) values of K serve to

stabilize (destabilize) the counterclockwise motion solution. As the magnitude of the controller's gain decreases, the magnitude of the Rayleigh number at the Hopf bifurcation point increases and the onset of chaos is delayed.

Another interesting phenomenon exhibited in Fig. 5 is that for sufficiently high values of the controller's gain (i.e., K>1.63), the Hopf bifurcation becomes supercritical. For example, in an experiment (Fig. 5) in which we set K=1.7 and gradually increased the Rayleigh number, we observed bifurcation from time-independent to stable, time-periodic flow.

Figures 6 and 7 illustrate the impact of the feedback control on the system. In Fig. 6, the theoretically obtained temperature difference between positions 3 and 9 o'clock is depicted as a function of time for the controlled (solid line) and uncontrolled (dashed line) systems for the same nominal operating conditions ($Ra=50\sim3Ra_H$). The uncontrolled system exhibits chaotic oscillations while the controlled system exhibits laminar behavior. Similar behavior is observed in experiments. Fig. 7 depicts the experimentally obtained temperature difference between positions 3 and 9 o'clock as a function of time for the same nominal operating conditions as in Fig. 6. To illuminate the impact of the feedback control, the behavior of the uncontrolled system is depicted for about 33 minutes prior to the application of the controller. Witness that once the controller has been engaged, the seemingly random, violent oscillations disappear and the flow is "laminarized". The relatively low amplitude oscillations seen in the system after the application of the controller are attributable to the fact that our experimental system is relatively noisy and sensitive to changes in the environment.

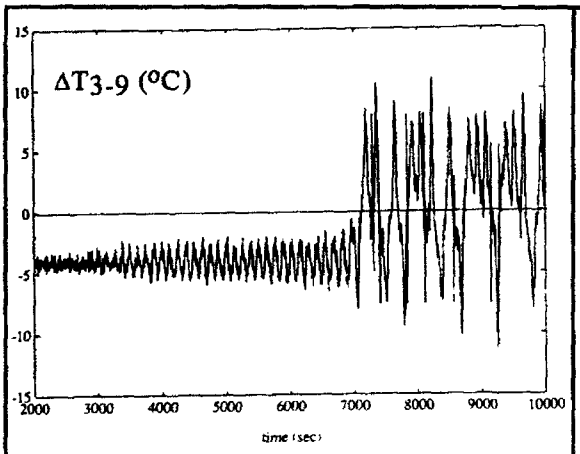


Fig. 8: The controller is being used to induce chaos at $Ra=0.6Ra_H$ with K=35W/C. The figure depicts the temperature difference between positions 3 and 9 o'clock as a function of time.

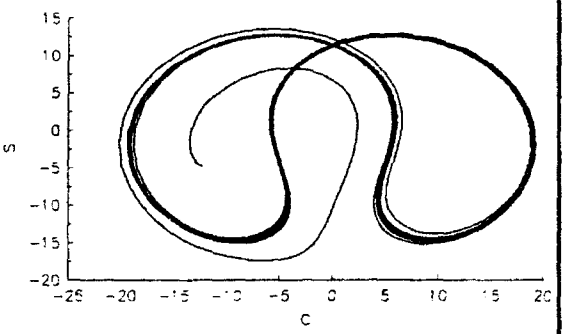


Fig. 9: Stabilization of an otherwise non-stable periodic orbit embedded in the chaotic attractor for Ra=50.

We can also induce chaos under circumstances in which chaos will not normally occur by using positive feedback (K>0) in (13). In this case, though, we will need to destabilize simultaneously both the counterclockwise (B_+) and the clockwise (B_-) time-independent motion states. This we accomplish by employing

the control strategy $W_1 = -1 - \frac{K}{Ra}[Sgn(c)c(t) - c_0)$. The experimental results of this strategy are depicted in Fig. 8, where we show c(t) as a function of time for subcritical Rayleigh numbers Ra~0.6 Ra$_H$=10. To illustrate the difference between the uncontrolled and the controlled systems, we initially depict the uncontrolled behavior for t<1800s. Witness the relatively time-independent, counterclockwise motion. The controller was activated at t=1800s. It caused flow instabilities to amplify, as is evident in Fig. 8, until eventually chaotic flow developed.

Finally, we can use the controller to stabilize one of the otherwise non-stable, periodic orbits embedded in the chaotic attractor. We do this by employing a control strategy similar to the one described by equation (13), with $c_0(t)$ now representing the periodic orbit we desire to stabilize. The periodic orbit was identified by examining a return map. Alternatively, one can employ one of the techniques described in Mindlin and Gilmore (1992). The controller can be employed either continuously in time or at discrete time intervals such as once every period (i.e., $K=K_0\delta(t-n\mathfrak{T})$, where $\delta(\cdot)$ is the Dirac delta function, $\mathfrak{T}$ is the period of the orbit we wish to stabilize, and n is an integer) or when the orbit passes in close vicinity to a point on the periodic orbit we wish to stabilize. Fig. 9 depicts a representative result of the "cs"-plane projection of a stabilized periodic orbit under conditions in which, in the absence of the controller, the flow would be chaotic (Ra=50).

As we have seen, the thermal convection loop paradigm exhibits complicated temporal behavior. Next, I describe a strategy for controlling convection in a system which is not only temporally but also spatially complex.

3.2 THE LAPWOOD PROBLEM

Let us consider thermal convection in a saturated porous layer confined in an upright box, heated from below and cooled from above. The side walls of the box are insulated. As in the case of the thermal convection loop paradigm, as the temperature difference between the bottom and top (or the Rayleigh number) increases, the equilibrium state of the system undergoes a sequence of bifurcations from no-motion to time-independent, cellular motion to oscillatory, chaotic flow. Our objective is to shift the bifurcation points with the aid of a feedback controller. I describe below only the stabilization of the no-motion state (Tang and Bau, 1992 a and b).

I use, respectively, the height of the box (H), the thermal diffusion time H/α (where α is the equivalent thermal diffusivity of the porous medium), and the nominal temperature difference (ΔT^*) between the bottom and top of the box as the length, time and temperature scales. In the uncontrolled system, the non-dimensional bottom and top temperatures are assigned the values one and zero, respectively. The non-dimensional, horizontal dimensions of the box are L_x and L_y. The motion of the fluid is described by the Darcy-Oberbeck-Boussinesq (**DOB**) equations (i.e., Joseph, 1976):

$$\dot{u} = B^{-1}(-u - \nabla p + R\theta\,\hat{e}_z)$$
$$\dot{\theta} = \chi^{-1}(u_z - u\cdot\nabla\theta + \nabla^2\theta)' \qquad (14)$$

where $u^T=\{u_x,u_y,u_z\}$ is the divergent free ($\nabla\cdot u=0$), non-dimensional velocity vector which satisfies the impermeability boundary conditions on all solid surfaces; x and y are horizontal, Cartesian coordinates; z is the vertical coordinate; $\theta=T-(1-z)$ is the deviation of the non-dimensional temperature from the linear (conduction) profile;

$R=\frac{g\lambda\beta H\Delta T^*}{\upsilon\,\alpha}$ is the Darcy-Rayleigh number· g is the gravitational acceleration; λ is the medium's permeability; β and v are, respectively, the thermal expansion coefficient and the kinematic viscosity of the saturating fluid; p is the pressure deviation from its hydrostatic value; $B=\frac{\lambda}{\phi H^2}\frac{\alpha}{\upsilon}$ (usually, B<<1); ϕ is the porosity; and $\chi=\frac{(\rho_0 c_p)_{eq}}{(\rho_0 c_p)_f}$ is the ratio between the equivalent thermal capacity of the medium and that of the saturating fluid.

The **DOB** equations admit the no-motion solution {u, θ}=0, which for R<Rc is globally stable. The magnitude of Rc depends on the aspect ratios L_x and L_y. For example, for an unconfined fluid layer with rigid bottom and top boundaries, $Rc=4\pi^2$ (i.e., Lapwood, 1948). At R=Rc, the no-motion state loses stability through a supercritical bifurcation into time-independent, cellular motion. The linearized problem of stability is self-adjoint, thus the growth rate at the onset of instability is real. In other words, the bifurcation occurs through a real eigenvalue, and the "principle of exchange of stability" is valid.

To extend the range of stability of the no-motion solution to higher values of the Rayleigh number, we modify the bottom temperature in proportion to the deviation of some interior temperatures from their (no-motion) conduction value. We will assume that the temperature distribution is measured at some horizontal cross-section, i.e., at z=0.5, as a function of the horizontal co-ordinates (x, y). The controller will sense the deviation $\theta_{0.5}=\theta(x,y,0.5,t)$ of the actual temperature from its desired value (θ=0) in the plane z=0.5. (For R>Rc, in the absence of the controller, $\theta_{0.5}\neq 0$). It will, then, modify slightly the spatial distribution of the base's temperature in proportion to the measured deviation and its time derivative, i.e.,

$$\theta(0,x,y)=C(\theta_{0.5})=-K_P\theta_{0.5}-K_D\frac{d\theta_{0.5}}{dt}, \qquad (15)$$

where K_P and K_D represent, respectively, the gains of the proportional and the differential controllers. Note that boundary condition (15) replaces the boundary

condition $\theta(x,y,0,t)=0$ in the uncontrolled problem. This can be accomplished in practice, for example, by embedding a network of computer-controlled heaters in the base plate and varying their heat input in an appropriate manner.

Next, we investigate the linear stability of the controlled system in the presence of the controller. The analysis is somewhat more complicated than in the uncontrolled case since the system (14-15) is no longer self-adjoint and one cannot ignore the possibility of the no-motion state undergoing a Hopf bifurcation resulting in time-periodic motion. As in the classical case, the disturbances can be decomposed into even and odd modes with respect to the mid plane (z=0.5). The odd modes, which lose stability at $R=16\pi^2$ (r=4), have a node at z=0.5. Thus, the control strategy employed here (equation 15) will affect only the stability characteristics of the even modes. The results of the analysis are summarized in Fig. 10, where we depict the ratio between the Rayleigh numbers of the controlled and uncontrolled systems at marginal stability ($r=R/4\pi^2$) as a function of the proportional controller's gain (K_P) for various gains of the differential controller (K_D=0., 0.02, and 0.06) for a box with aspect ratios $L_x=L_y=1$. The various symbols in Fig. 10 represent the "most dangerous" modes (n,m). The first number (n) represents the number of cells in one direction and the second number (m) represents the number of cells in the other direction. For example, the modes (1,0) and (1,1) correspond, respectively, to unicellular two and three-dimensional convection. The solid and dashed lines represent, respectively, bifurcations through real and complex eigenvalues. The figure demonstrates that through the action of the controller, the critical Rayleigh number at the onset of convection can be increased by as much as a factor of 4. Through optimization, a further increase in the critical Rayleigh number may be attainable.

We have also successfully applied similar techniques to stabilize the no-motion state in the Rayleigh-Benard problem (Tang and Bau, 1992c).

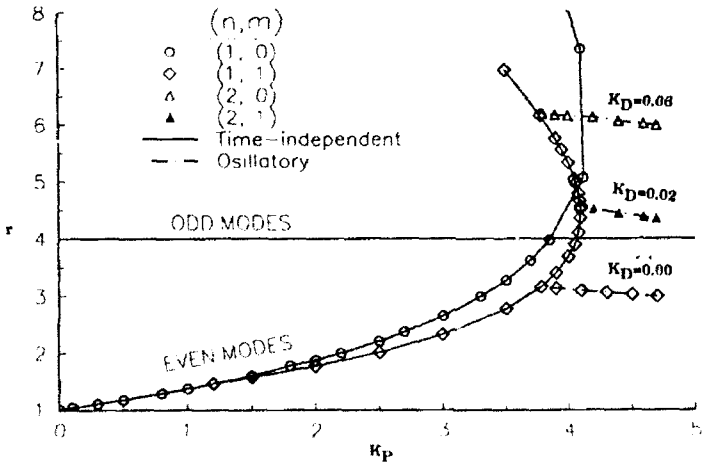


Fig. 10: The critical normalized Rayleigh number (r) at the onset of convection is depicted as a function of the controller gain K_P for K_D=0.0, 0.02, and 0.06. The symbols represent the various modes (n,m). The solid and dashed-dot lines correspond, respectively, to bifurcations into time-independent and time-periodic convection.

4. CONCLUSION

As the simple examples described here indicate, the bifurcation structure of convective flows can be modified in a significant way through the use of feedback control. In the future, similar techniques may profoundly impact diverse technological processes, enabling us, for instance, to substantially reduce drag in turbulent flows and significantly modify convection in crystal growth and material processing industries.

ACKNOWLEDGMENT

This work was supported, in part, by the Department of Energy, Office of Basic Energy Sciences, through Grant DE-FG02-92ER14271. A travel grant from the US National Committee on Theoretical & Applied Mechanics is also gratefully acknowledged.

REFERENCES

Bau H.H. and Wang Y-Z, 1991, Chaos: a Heat Transfer Perspective, to be published in Annual Reviews in Heat Transfer, IV, 1-50, (C.L. Tien, editor), Hemisphere.

Bushnell, D.M., and McGinley, C.B., 1989, Turbulence Control in Wall Flows, in Annual Reviews Fluid Mechanics, 21, 1-20.

Choi, H., Temam, R., Moin, P., and Kim, J., 1992, Feedback Control for Unsteady Flow and its Application to Burgers Equation, Center For Turbulence Research, CTR MS 131.

Creveling H.F., De Paz J.F., Baladi J.Y., and Schoenhals R.J., 1975, Stability characteristics of a single phase thermal convection loop, J. Fluid Mechanics, 67, 65-84.

Ditto W.L., Rauseo S.N., and Spano M.L., 1991, Experimental Control of Chaos, Physical Review Letters, 65, 3211-3214.

Gorman M., Widmann P.J., and Robins K.A., 1984, Chaotic Flow Regimes in a Convection Loop, Physical review Letters, 52, 2241-2244.

Gorman M., Widmann P.J., and Robins K.A., 1986, Nonlinear Dynamics of a Convection Loop: a quantitative comparison of experiment with theory, Physica 19D, 255-267.

Joseph D.D., 1976, Stability of Fluid Motions II, Springer-Verlag.

Lapwood, E.R., 1948, Convection of Fluid in a Porous Medium, Proceeding Cambridge Phil. Soc., 44, 508-521.

Liepmann, H.W., Brown, G.L., and Nosenchuck, D.M., 1982, Control of Laminar-instability waves using a new technique, J. Fluid Mechanics 118, 187-200.

Liepmann, H.W., and Nosenchuck, D.M., 1982, Active Control of Laminar-Turbulent Transition, J. Fluid Mechanics 118, 201-204.

Lions, J.L., 1971, Optimal control of Systems Governed by Partial Differential Equations, Springer-Verlag.

Lorenz E.N., 1963, Deterministic Nonperiodic Flow, J. of the Atmospheric Sciences, 20, 130-141.

Malkus W.V.R, 1972, Non-periodic convection at high and low Prandtl number, Mem. Societe Royale des Sciences de Liege, 6^{e}serie, tome IV, 125-128.

Mindlin, G.B., and Gilmore R., 1992, Topological Analysis and Synthesis of Chaotic Time Series, to appear in Physica D.

Ott E., Grebogi C, and Yorke J.A., 1990a, "Controlling Chaos" ,Physical Review Letters, 64, 1196-1199.

Ott E., Grebogi C, and Yorke J.A., 1990b, "Controlling Chaotic Dynamical Systems", in CHAOS: Soviet-American Perspectives on Non-linear Science, (D.K. Campbell, editor), American Inst. of Physics, New York, pp. 153-172.

Singer, J., Wang, Y-Z, and Bau, Haim H., Controlling a Chaotic System, Physical Review Letters, 66, March 4, 1991, 1123-1125.

Singer, J., and Bau, Haim H., 1991, Active Control of Convection, Physics of Fluids A, 3, 2859-2865.

Tang, J., and Bau, Haim H., 1992a, Stabilization of the No-motion State of a Horizontal, Saturated, Porous Layer Heated From Below, to appear in ASME Proceedings of the 1992 Winter Annual Meeting.

Tang, J., and Bau, Haim H., 1992b, Feedback Control Stabilization of the No-Motion State of a Fluid Confined in a Horizontal, Porous Layer Heated from Below, submitted for consideration for publication.

Tang, J., and Bau, Haim H., 1992c, Stabilization of the No-motion State of the Rayleigh-Benard Problem through Feedback Control, submitted for consideration for publication.

Wang, Y-Z, Singer J., and Bau, Haim H., 1992, Controlling Chaos in a Thermal convection Loop, J. Fluid Mechanics 237, 479-498.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Application of Structural Mechanics to Biological Systems

C.R. Calladine

Department of Engineering, University of Cambridge, Trumpington Street, Cambridge CB2 1PZ, United Kingdom

Abstract

I describe two examples of the application of mechanics to biological rod-like structures at the molecular level.

The first is the filament of a bacterial flagellum. This is a slender but rigid corkscrew-like structure by which a bacterium propels itself through the water. The filament, of diameter 20 nm, is built from identical subunits in a definite cylindrical lattice. The fact that the filament is curved rather than straight poses a paradox, which can be resolved by mechanics.

The second example is DNA, the material of heredity. This molecule is a rod of diameter 2 nm, with a double-helical structure. Mechanical properties, such as intrinsic curvature and flexibility, are a function of the sequence of bases along the molecule; and they are used in the recognition of particular sequences of DNA by protein molecules. Again, mechanics enables us to understand the behaviour.

I also make some general remarks about some hazards which are likely to emerge when we try to apply mechanics to problems in molecular biology.

1. INTRODUCTION

In this talk I intend to discuss two specific examples of the application of mechanics – as we understand that term here in IUTAM – to the solution of structural problems in molecular biology. The first example is a slender, rigid, corkscrew-like filament, the *bacterial flagellum*, which is part of the machinery for propelling bacteria through water: it is a filament of diameter 20 nm, and is made from a single type of "building block", which is a molecule of the protein *flagellin*. The problem is to understand the principles of construction of the filament. The second example is the double-helical DNA molecule, a thread of diameter 2 nm, which carries the complete genetic information in every cell of a living organism. Here the problem is to understand the mechanical properties of the thread – such as its intrinsic curvature and flexibility – in terms of its base sequence, and how these play a role in the ways by which DNA works within the cell.

I thought it would be more interesting to present here two examples of this genre rather than one, in spite of the obvious difficulty inherent in such a plan over the time available for necessary detail. My main aim will be to try to convince you that mechanics has some interesting applications in the field of molecular biology – even if you are quite unfamiliar with biology as a discipline.

Let me begin by making some general remarks about collaboration between mechanics people and biologists. Both groups work, of course, within the same general framework of "scientific method"; but in practice there is usually a high barrier between the two sets of people. One element of this barrier is *jargon*: each side coins words and expressions which will be convenient for the communication of ideas and thoughts. However, as most of us are

aware, jargon within any field can actually *inhibit* clear, productive thought. Thus, someone from outside a given field often has a double task in penetrating the jargon to the core ideas of the subject. I shall try, of course, to keep this talk relatively free from jargon.

In trying to overcome barriers between ourselves and biologists, it is necessary for us to *talk* with biologists; and even more important for us to *listen* to them. For a tremendous wealth of experimental observation is available in biology; and a mechanician must hope that somewhere among the data a suitable problem may be found. We are definitely not in the business of having a *solution* and looking for a suitable *problem* to which it may be applied. Rather, we are looking for tractable problems among a veritable haystack of empirical data. This is an area of work where, if one can but *state* the problem, one has already come a long way towards its solution.

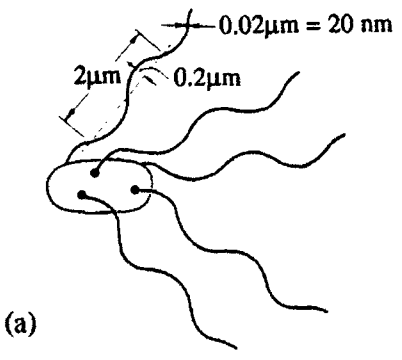
A special difficulty which mechanics people will experience in this field is that biologists talk and think a great deal about "structure" at many different levels. Most biologists have "hands-on" experience with three-dimensional chemical model structures of the ball-and-spoke variety; and many are also competent crystallographers. When they discuss possible alternative structural conformations for a molecule – whether large or small – they find it natural to think in terms of "minimum energy" configurations. Very few biologists have ever encountered what we call the *concept of stress*; which is, of course, absolutely central in all of our applied-mechanics approach to engineering structures. This difference in outlook can be a big stumbling-block to cooperation.

As I have intimated, this lecture will be rather scanty on details. Those who seek a more complete description of the work will get help from the bibliography.

2. CONSTRUCTION OF THE BACTERIAL FLAGELLUM [1–4]

I was fortunate to have been given this problem by my friend Aaron Klug. His main interest is the structural organisation in biological systems; and he detected in the bacterial flagellum a paradox which seemed to be beyond the range of crystallography or chemistry or thermodynamics; but which he thought might be susceptible to an attack from the standpoint of structural mechanics [5].

The main dimensions of a bacterial flagellar filament are as shown in the sketch of Figure 1. The filament is of length a few microns (μm) and diameter 20 nm. It has the form



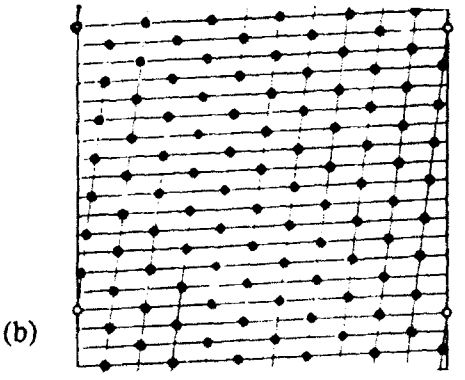


Figure 1 (*left*). Sketch of *Salmonella* bacterium "at rest", showing the cell ("body"), rotary motors and "normal" flagellar filaments.

Figure 2 (*right*). The surface lattice for a "straight" flagellar filament, to a much larger scale: this is an "unrolled" sheet which was formerly wrapped around a short piece of filament. The position of each subunit is marked by a spot; and the two sides join at the open spots.

of a "rigid" and uniform helix, typically with radius 0.2 μm (= 200 nm) and pitch 2 μm. The filament is built from a single kind of protein building block or subunit, and the simple cylindrical surface lattice of the filament contains near-longitudinal 11-start helices, as shown in Figure 2. This and other information about the packing-pattern of the flagellin molecules comes from electron microscopy by image-reconstruction techniques [6]; while data on the overall helical shape comes from both electron- and light-microscopes [7,8].

The paradox perceived by Klug is this. If all the building blocks in the cylindrical lattice were in a strictly *equivalent* relationship with their neighbours, then the rod would be *straight*. But the rod is actually *curved* into a helical shape. How, then, is the construction arranged?

Our early thinking about this problem suggested that the solution might lie in the idea of some sort of "bistable switch" between two states [9], either within the subunit or at an interface or connection-point between neighbouring subunits. (In terms of the mechanics of the complete assembly, there can hardly be a difference between these two situations.) At the molecular level, a bistable feature of this kind had already been found within the four parts of the haemoglobin molecule [10].

On the whole, progress was slow until we took into account data which increasingly became available on a *range of different helical forms* which a flagellar filament can adopt in a variety of different circumstances: the filament is *polymorphic*. Briefly, there are (i) different natural helical shapes in different mutants of *salmonella* and *E. coli*. bacteria [7]; (ii) different shapes under different environments – such as different acidity and salt content of the surrounding water [11]; (iii) different forms when flagellin monomers from different strains are mixed and co-polymerised into artificial helical filaments [7]; and (iv) changes in helical shape when a filament is subjected to torsional stress [12], and particularly when its rotary motor reverses direction, as sketched in Figure 3 [13,14]. Furthermore, it turned out that the "family" of distinct helical waveforms which can be built under all of these different circumstances had a limited number of members, as shown in Figure 4: the same helical forms appeared over and over again in the different assays.

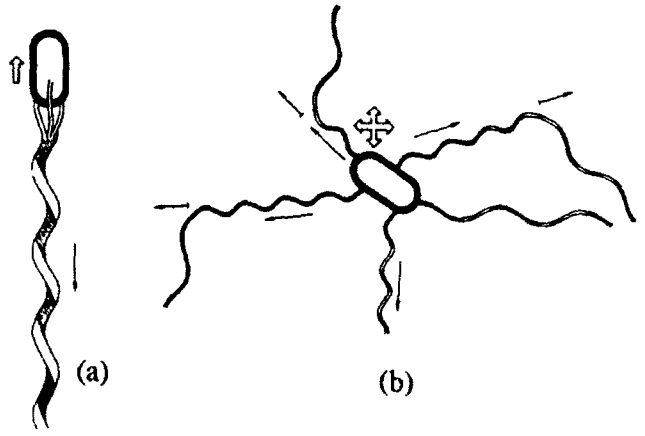


Figure 3 (*left*). (a) In smooth swimming, the various filaments (cf. Figure 1) associate to form a single bundle. (b) Occasionally the rotary motors go into reverse, and a new waveform is seen at the distal end. These sketches were taken from light-microscope assays [14].

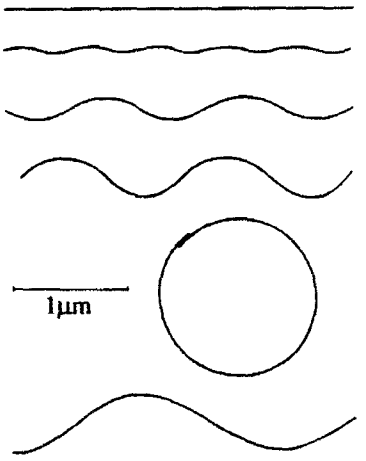


Figure 4 (*right*). Sketch of the family of discrete helical waveforms of flagellar filaments which are made by copolymerisation of protein monomer subunits from two strains of bacteria [1], and which appear also in other assays.

Now to a structural engineer investigating the building of a helical rod in a curious way, it seemed obvious that one should map the available data on radius and pitch of the various polymorphic helices into a space of (uniform) *curvature* κ and *twist* τ. When we did this and took account of the helical handedness of the discrete forms (right-handed (+) or left-handed (–)) [8] we found that the points corresponding to the family of discrete forms lay on a single arc [2–4]; and indeed that they lay close to an array of 12 points equally spaced on a sinusoidal curve, as shown in Figure 5.

At this stage it became not difficult to envisage a situation in which the subunits – which are represented, conventionally, by featureless spots in Figure 2 – were rectangular in shape, and connected at their corners in a chess-board fashion, with one type of corner-connection having a bi-stable feature, as shown in Figure 6(a) [2,3]. The different helical forms could then correspond to a number (from 0 to 11) of entire longitudinal strands of connections being one configuration rather than the other, as shown in Figure 6(b). If we then arranged that the chain of "1" (or B) connections was slightly *longer* than the corresponding chain of "0" (or A) connections, then the different numbers of "1" connections would produce different amounts of curvature in the filament; and indeed, provided that the strands of "1" connections were all bunched together in one portion of the circumference of the rod, as shown in the examples of Figure 6(b), the points in the κ, τ plot all lay on a single, sinusoidal arc as shown in Figure 6(c). The points are equally spaced in the τ direction because each additional strand of "1" connections introduces an equal extra twist to the filament overall, as explained in the caption of Figure 6(b).

The way in which the strands of "1" connections produce curvature may be seen from a simplified "lion's cage" model [2], shown in Figure 7, in which two rigid discs are separated by 11 parallel elastic bars, some of which (in the "1" state) are slightly longer than the others ("0" state) when they are relaxed. It is an elementary matter to show that the angle of tilt between the discs – which corresponds, of course, to curvature in the filament – is proportional to $\sin(n\pi/11)$, where n is the number of bars in the longer, "1" configuration. The complete range of 12 possible states includes two straight forms ($n = 0$, $n = 11$): one of these has left-handed twist and the other right-handed twist.

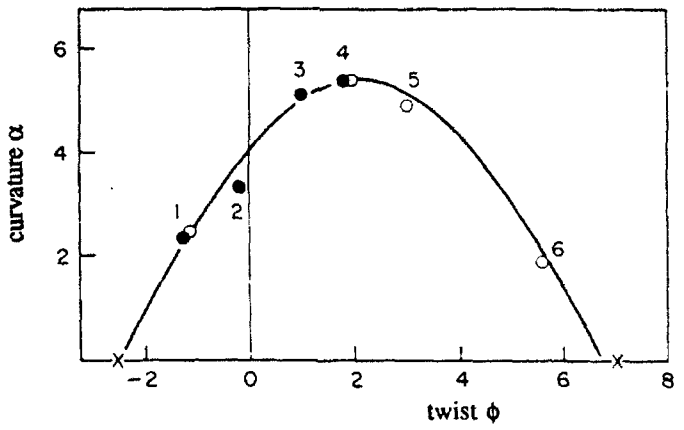


Figure 5. Twist-curvature plot of helical forms observed in homopolymers from different strains or mutants of bacteria (●) and copolymers of subunits from two different strains (cf. Figure 4) (○). Arbitrary units for curvature and twist. Point X on the right corresponds to a mutant with straight flagellar filaments having right-handed twist, as in Figure 2 [6]; and point X on the left corresponds to another "straight" mutant, discovered later, with filaments having left-handed twist.

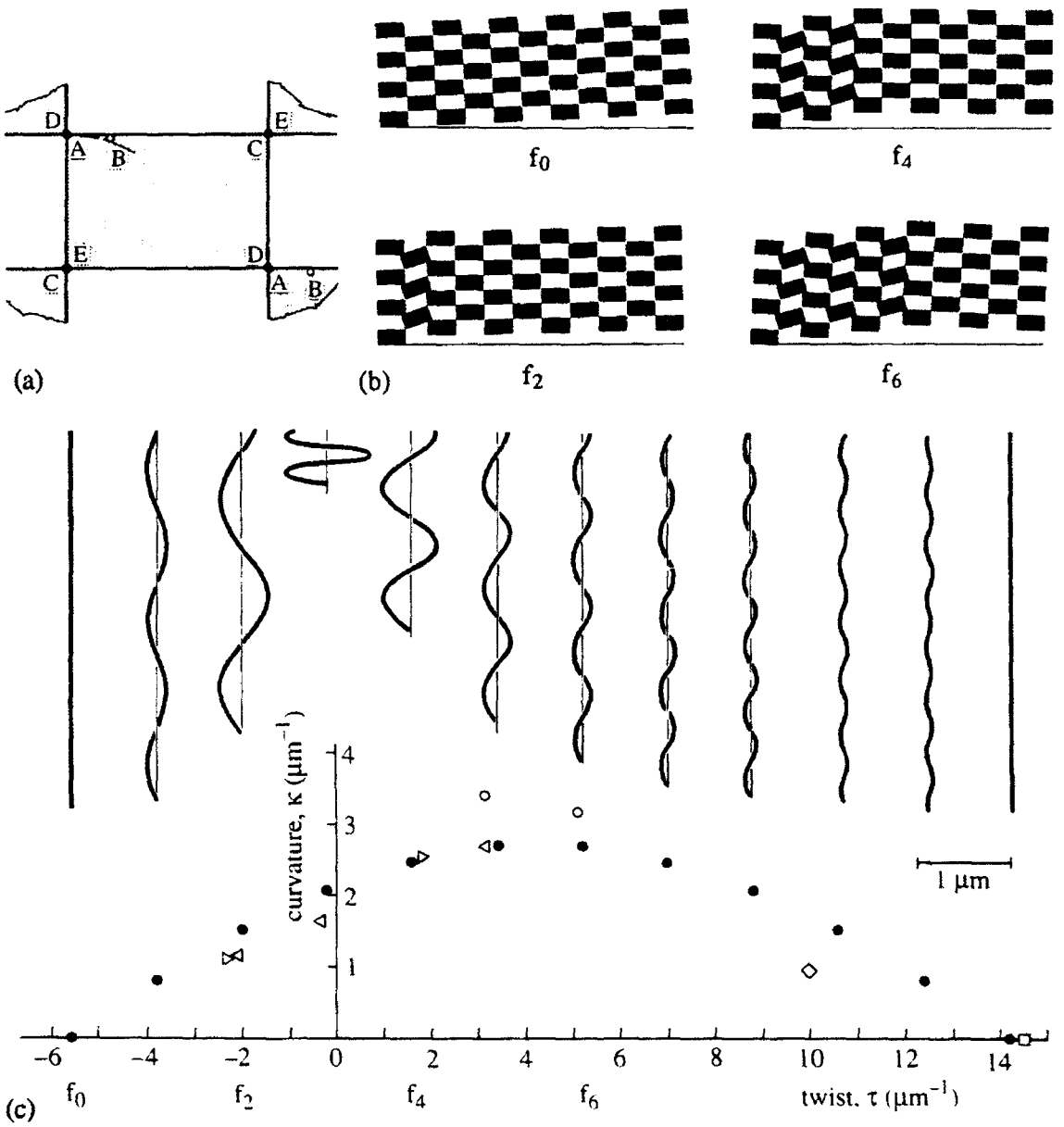


Figure 6. (a) Proposed near-rectangular subunit, showing inter-connections with identical neighbours. The rule of connection is that C joins to E, and D joins *either* to A *or* the neighbouring point B. (b) Four possible packing arrangements in the surface lattice (cf. Figure 2), in which the subscript indicates the number of near-longitudinal (vertical) strands of connections in the longer "1" configuration, corresponding to connections B in (a). The horizontal lines are circumferential. Note how the inclination of the near-longitudinal columns of subunits with respect to the circumference changes as the number of "1" strands increases, as a direct consequence of the geometry of packing; so that the overall *twist* of the filament changes in proportion to the number of "1" strands. (c) Curvature (κ), twist (τ) plot for the twelve theoretical packing states (of which four are shown in (b)) together with the helical shape (for a given contour length) of the corresponding waveforms. The geometrical parameters of the model have been chosen to fit the experimental observations from Figure 5 and from other sources, which are marked by open symbols.

In principle, of course, there could be several different configurations with, say, n = 9 sets of longitudinal strands in the "1" configuration and two sets in the "0" configuration. Why, then, have we chosen only one, as shown in Figures 6(b) and 7, in which these two are bunched together on one side of the filament? It is clear that such a "bunching" feature will make the range of distinct states lie on the single arc as in Figure 6: but can we adduce any physical reason for this restriction?

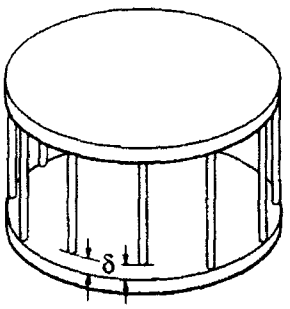
A plausible answer is that if the two slightly shorter rods in the model of Figure 7 were *not* adjacent to each other in a single group, then the elastic strain energy would be greater. It is straightforward to obtain this result; and indeed a similar result holds for all values of n [2].

Let us now look more closely at the "chess-board" model of the arrangement of subunits. The essential feature of the chess-board scheme is that it provides a way of disposing subunits on the known surface lattice while providing for longitudinal strands of connection which can give the short/long "0/1" feature that is needed for curvature. It is crucial, of course, to the model that the "0/1" pattern should *propagate* unchanged along the entire length of a filament; and again an elastic calculation – though admittedly on a plane grid of elastic units – shows that the "0/1" switch will not propagate in any direction of the surface lattice (Figure 2) other than the near-longitudinal 11-start direction [3].

Some features of the structural model are critical, while others are not. Thus, the general layout, with a single "0/1" bistable feature for each subunit, and an otherwise simple elasticity, are all crucial. But it is probably not essential to have the subunits *planar*, for example. We have also assumed tacitly that inter-unit contacts are of the "frictionless spherical joint" variety. This is not unreasonable as a picture of connections between adjacent protein molecules; but, again, it is not an essential feature.

Some may object, of course, that our flat subunits do not look at all like protein molecules. The key point here is that our unit is, strictly, a model for the proper spatial connection between the four points at which the molecule connects to its neighbours.

We were led to use the scheme of *four* connecting points for each sub-unit by the following reasoning; which turned out to be incorrect, as we shall see. (There must, of course, be an *even* number of connecting points per subunit if *two* units connect at each joint.)



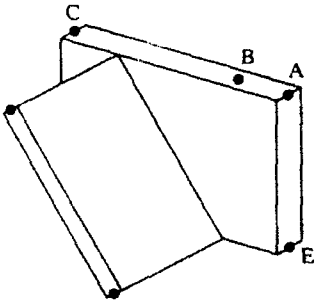


Figure 7 (*left*). "Lion's cage" model to explain how curvature of the filament depends on the number of longitudinal strands of connections in the "1" (long) and "0" (short) configurations. Here the strands are represented by rods which are bistable in their relaxed form – as shown – but are otherwise linear-elastic. When the two short rods are connected up to the discs, the discs tilt relative to each other, corresponding to curvature in the filament.

Figure 8 (*right*). Six-connected version of the model subunit, in which the four-connected "surface" unit of Figure 6(a) is equipped with an inward-facing extension which, *inter alia*, ensures that the subunits pack in a unique surface lattice.

Thus, if one rigid block is fixed, and another rigid block is then connected to it by means of a spherical joint, the act of coupling removes *three* degrees of freedom from the second block: for, whereas it originally had 6 independent degrees of freedom in 3-space, it now has only 3, in the form of 3 components of rotation. Hence we may say that a spherical joint removes 1.5 degrees of freedom per connecting block. Therefore a block with *four* connecting points loses a total of 6 degrees of freedom when it is connected up; and so it becomes fixed in space, and part of a rigid structure.

This argument is almost entirely satisfactory *provided the blocks are built into the correct surface lattice.* But in fact the four-connected subunits of the type which we are discussing can clearly be built into an *arbitrary* cylindrical surface lattice, and not only the specific lattice of Figure 2 in which we are interested. So what determines a specific surface lattice? The only acceptable answer to this question is *the subunits themselves.* The details of the surface lattice must be a consequence, somehow, of the geometrical design of the subunit: and the subunits can then *self-assemble* into the correct packing pattern.

The extra geometrical constraints that produce the proper global packing pattern can be provided by the addition of two further points of connection with neighbouring units. A possible arrangement is shown in Figure 8 [4]. It is clear that the total number of units around the circumference of the cylinder will be determined by the geometrical layout of the inward-projection; and indeed the overall *twist* of the assembly to ensure correct closure is also fixed by the additional geometrical constraints.

The scheme which I have outlined here enables us to see, in principle, how a single kind of subunit can build only into one particular surface lattice with, indeed, a specific number of longitudinal strands of connections in the "1" configuration: the assembly simply will not fit together in any other way. And a very small change in the geometry of the subunit would make the units assemble only into another, specific member of the discrete family of waveforms.

There is not much point in pursuing a detailed analysis of the small changes in geometry of the subunit which would effect changes between particular helical forms within the discrete range of available states. The crucial point is that the changes are small; and moreover they are of the kind which would be produced in the actual protein by a variety of means, such as a change of hydrogen activity, or a mechanical torque. And it is not difficult to see that a copolymerisation of two different monomers, each of which by itself would construct a particular member of the family of distinct helical waveforms, could construct, between them, an intermediate member of that family.

In summary, we have here a "mechanical" model which produces all of the main aspects of the known behaviour of the bacterial flagellar filament. The essential features of the design are clear, but much of the detailed geometry of the subunit is still uncertain. The data now available on the detailed shape of a subunit [15–17] are much better than they were in the mid 1970s, when the work which I have described was done; but they are still insufficient for the construction of a physically accurate model. It seems likely that new experimental techniques will enable some progress to be made on this problem within the next few years. A particularly interesting recent result is that the part of the flagellin molecule which is responsible for the special scheme of construction of the filament occupies only a small fraction of the molecule, at the two ends of the protein chain [18].

3. STRUCTURAL MECHANICS OF DNA [19–23]

It has been well known for nearly 40 years that DNA, the molecule within every cell of our bodies, that stores our individual genetic "design blueprint", has a double-helical structure, as shown schematically in Figure 9. Attached to each of the helical "backbones" or strands are chemical *bases* of four different kinds – A, T, G, C – which carry the genetic information in three-letter code-words. The bases form *pairs* in a cross-chain sense; and since things are so arranged that A always pairs with T and G with C, as in Figure 9, the second

strand carries the *complement* of the code-message of the first. The helical strands are made from sugar-phosphate units, and their degree of intertwining is such that there are about 10 base-pairs in one double-helical turn of DNA.

When a cell is about to divide into two cells – this being the basic scheme for growth or construction in all living organisms – the DNA has to be duplicated, so that a complete and accurate copy of it can be incorporated in each daughter-cell. This is achieved in the cell by first separating the two strands from each other and then assembling onto each, from available "spare parts", the missing complementary strand.

All of this is well-known, high-school science. What is perhaps not so well-known is that the DNA thread in any "chromosome" or packaging unit in the cell is extremely long in comparison with its width: its aspect ratio is typically about 2×10^7. Thus, if we were to enlarge a piece of DNA by a linear factor of 10^6, it would be like a piece of string 2 mm in diameter and about 40 km long. On the same scale, the diameter of a single cell would be about 10 m. We have to conclude that there is a colossal packaging problem, to get the DNA thread bundled up into the available space, and without tangling.

Nature uses a hierarchy of packing procedures to achieve this massive degree of compaction. The details are rather complicated; but for present purposes it suffices to say that the DNA thread in the cell is wrapped round a series of "bobbins" made from special proteins called histones. The DNA wraps twice round each "histone spool" in a left-handed superhelical sense, as shown schematically in Figure 10. There are about 80 base-pairs in each turn of DNA around the spool. The picture of Figure 10 is not drawn to scale: in fact, the total volume of DNA in the arrangement is approximately equal to the total volume of histones.

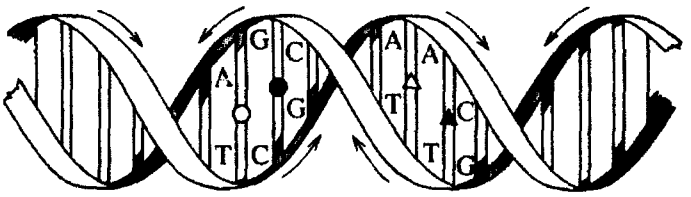


Figure 9. Double-helical structure of DNA: a schematic view showing the "anti-parallel" directions of the two chains, and bases, base-pairs and base-pair steps.

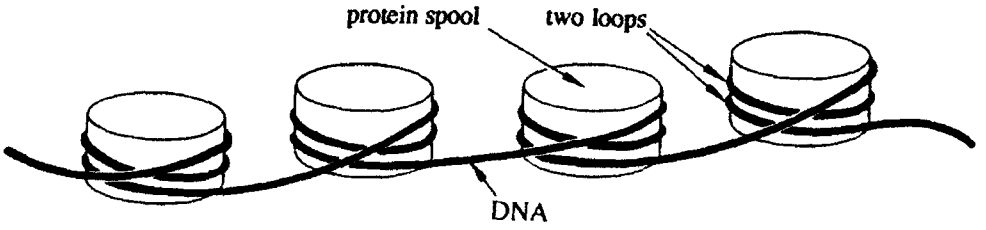


Figure 10. DNA wrapped around histone spools.

I have used the word "superhelical" here to describe this wrapping of DNA around a spool to distinguish the *path* of the DNA thread from the right-handed helical twisting of the two strands about each other, on a much smaller scale, within the DNA thread itself.

Before we can proceed further, I must explain some simple labelling conventions. Each of the sugar-phosphate chains has a distinct sense of *direction*, which can be seen directly from the atoms in "all-atom" pictures (cf. Figure 12, below) but which has to be shown by an arrow in a schematic drawing like that of Figure 9. The *sequence* of DNA is described by giving a letter-by-letter description of the bases along one strand, in the conventional direction. Later on we shall discuss double-helical DNA as an assembly of base-pair *steps* ("dinucleotide steps"). For example, base-pair steps TC (o), CG (●), AA (Δ) and AC (▲) may be seen on the strand which runs from left to right in Figure 9. These same four steps would be read – in the opposite order – as GT (▲), TT (Δ), CG (●) and GA (o) with reference to the other strand. Thus the step TC (o) may be described alternately as TC or GA, while the step Δ may be described alternatively as and TT or AA; but step CG (●) reads the same with respect to either strand. Altogether, there are 10 different kinds of base-pair step, according to sequence.

Two further points need to be made here. The first is that A and G bases ("purines") are *larger* than T and C bases ("pyrimidines"). This difference will play a part in the structural mechanics of DNA, as we shall see. The second is that modern analytical technology makes it a relatively straightforward matter to determine the precise *sequence* of a given piece of DNA a few hundred base-pairs long.

3.1. Some questions about DNA

I propose now to discuss some questions about DNA for which applied mechanics can help to provide some answers. I shall necessarily be brief; but I hope to be able to give sufficient detail for the discussion to be comprehensible. Full details may be found in the book which my collaborator Horace Drew and I have written recently [23].

3.2. Why is DNA double-helical?

If DNA were stretched out to make a plane "ladder", there would be *gaps* between successive base-pairs, because the bases are 3.3 Å thick (1 Å = 10^{-10} m = 0.1 nm) whereas the standard length of the segments of connecting chain is about 6 Å. One step of such an artificially straight ladder is shown schematically in Figure 11(a).

The DNA molecule is surrounded by water in the cell. Now the bases are *hydrophobic*; and thus they hate to have water sandwiched between them. The twisting of the plane ladder into a double-helix – shown for one step in Figure 11(b) – thus allows the bases to stack directly onto each other, while the 6 Å chain-lengths adopt a spiral shape so that they climb by only 3.3 Å per step. A simple geometrical calculation, based on the dimensions shown in Figure 11(a), indicates an angle of *helical twist* in (b) of about 34°, which is typical of actual

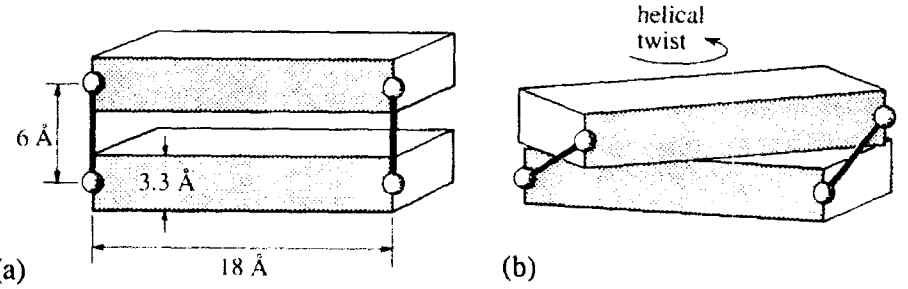


Figure 11. (a) A base-pair step of an artificially straightened "ladder" of DNA. (b) The same, but with the base-pairs arranged with helical twist, so that they can stack directly onto each other.

DNA. (Such a calculation is, however, too crude to predict whether the double helix will be left- or right-handed.)

This picture of the double-helical structure of DNA leads to an important point. The sugar-phosphate chains are often described as "backbones" in textbooks, as if they were providing some rigid support to the base-pairs – as if to the treads of a spiral staircase. In fact, the structure is controlled by the *base-stacking* arrangements; and the flexible chains have to adjust their positions accordingly. These chain-links provide some constraint, of course, to the relative positions which can be adopted by the stacking bases; but they are not in themselves the primary structural element of the double-helix.

3.3. Why is DNA wrapped around the histone spools in a left-handed fashion?

When the coded message of DNA is being read by the machinery within the cell, it is necessary for the two sugar-phosphate strands to come apart, so that the special "reading-enzymes" can gain access to the individual bases on one particular strand of the DNA.

Left-handed wrapping of the DNA around the spools is helpful in this respect. Thus, if we remove a spool and pull the now naked segment of DNA out straight, the DNA becomes strongly *untwisted* if the wrapping around the spool was in a *left-handed* sense. You can easily demonstrate this effect yourself, by wrapping some electric power cord around a beer can, removing the beer can, and pulling the cord out straight [24,25,23].

Thus Nature provides a neat and simple way of *untwisting* DNA when this is needed, as the first stage in the process of separating the two strands. In general, of course, it is not an easy matter to untwist a piece of string: there is a difficult problem in gripping the filament strongly enough. The arrangement in which the DNA goes in a left-handed sense round the

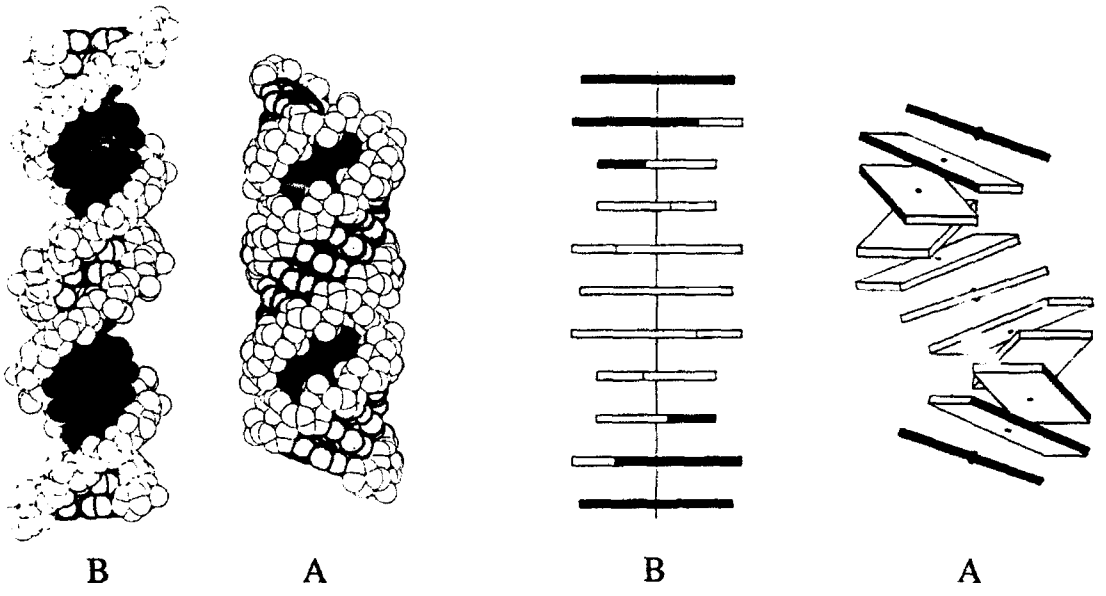


Figure 12 (*left*). All-atom representations of DNA in the classical "B" and "A" forms, showing two double-helical turns.

Figure 13 (*right*). Rigid-block representations of the base-pairs of DNA in the classical "B" and "A" forms. The sugar-phosphate chains are not shown, in order to display the geometry of the base-pair steps; and only one helical turn is shown. (The scale is different from that of Figure 12.)

removable histone spools provides a way of gripping and untwisting the DNA – in addition, of course, to achieving an overall compaction in length of the DNA.

3.4. How does DNA change from the "B" to the "A" form?

For this third question we must go back almost forty years, to the time when Rosalind Franklin obtained the first x-ray diffraction pictures, from samples of random-sequence DNA which had been prepared by pulling a tangled mass of DNA threads out into fibres: this procedure aligned the individual pieces of DNA with respect to each other, and allowed the x-ray method to reveal the mean or average configuration of the DNA molecule [26].

This exercise produced two distinct diffraction patterns: "A" when the fibre was dry, and "B" when it was wet. When these diffraction patterns had been analysed properly, so that the detailed structure could be discerned, the two corresponding kinds of DNA were both right-handed double helices: but there were some clear differences between them, as may be seen from Figure 12. Thus, relative to the "B"-form, a piece of "A"-form DNA having the same number of base pairs: (i) is *shorter*, (ii) is fatter, (iii) has its bases *tilted* with respect to the overall axis, and (iv) has about 10% less helical twist per base pair [27]. The first three of these features can be seen clearly in Figure 12.

For many years it was not at all clear what mediated the "B"-to-"A" transition when the water-activity was reduced. Usually the "backbone" was adduced as being somehow responsible – even though it could hardly explain the well-established fact that DNA having specific base sequences switched more readily to the "B" form with increasing water activity than DNA having other sequences.

As soon as we accept the idea that the double-helical structure of DNA is driven by the arrangements for stacking the bases on top of each other, we can begin to see how the transition between the two forms comes about. What we need is an interpretation of the two pictures in Figure 12 from the point of view of the *stacking arrangements* within a typical base

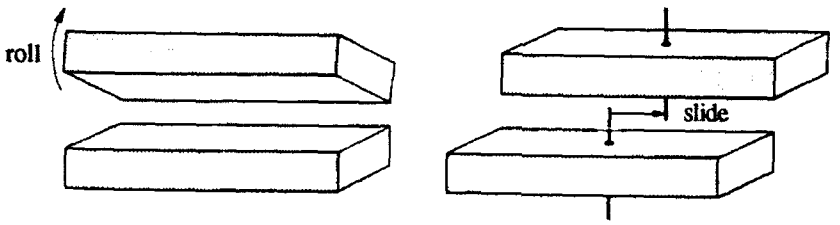


Figure 14. Definition of the "roll" and "slide" relative motions at a base-pairs step: "helical twist" is shown in Figure 11(b).

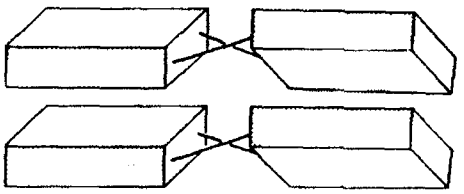


Figure 15. Propeller-twist in two base-pairs. The bases are shown here as if they were all of the same size.

step. This is shown in Figure 13, where the base pairs have been drawn as rigid blocks, and the sugar-phosphate chains have been omitted entirely [19]. At this point we need to introduce a technical definition. The so-called "minor-groove side" of the DNA has been marked in Figure 13 by marking the edges of the blocks in black. (See also the shading in Figure 11, which serves the same purpose.) This feature can also be seen in Figure 12: the "groove" is the space between the sugar-phosphate chains.

To go from the characteristic "B" step to the characteristic "A" step we need simply to apply both the "roll" and "slide" internal motions, as defined in Figure 14. These are in fact two of the six degrees of freedom which one rigid block has, in general, with respect to its neighbour; and of the six these two are the least constrained by the combined effects of the stacking arrangements and the sugar-phosphate links. A third degree of freedom, *helical twist*, is what we have already seen in Figure 11(b). The transition from the "B" to the "A" form requires, on average, a change in roll angle of 12°, and a change in slide of about 3 Å.

In recent years it has become possible to examine the stacking arrangements within the base-pair steps of DNA to atomic resolution by means of x-ray analysis of *single crystals* that have been made from very pure samples of short pieces of DNA having identical sequences. Such structures are almost always rather *irregular*, with strikingly different geometries in individual base-pair steps, although the double-helical structure appears to be uniform to first-order [28,29]. An almost-universal feature which emerges from such studies is that the base-pairs are not actually planar, but have "propeller twist" [28,30]. This is sketched in Figure 15. It can be seen from Figure 16 that if two propeller-twisted base-pairs are stacked onto each other, then geometry requires that the "slide" motion will be directly linked to the "roll" motion. And this is precisely the arrangement which we have already described as being characteristic of the "B"-to-"A" transition.

What, then, "drives" the transition? The explanation is, broadly, that in "wet" conditions, the hydrophobic forces make the base-pairs "cover" each other as well as possible, thereby giving low "slide", as in Figure 16(a). But in "dry" conditions the hydrophobic effect is weaker, and other stacking-forces push the base-pairs into an "offset" or high-slide position, as in Figure 16(b). (Unfortunately, these pictures are too crude to show this effect convincingly.)

It is not obvious from what I have said so far that the "A"-to-"B" transition will take place as a clear "switch" – as it does in practice, when the water activity is steadily reduced – rather than in a "smooth" fashion. The key to the situation is that in double-helical structures which have been studied by single-crystal x-ray methods, certain kinds of base-pair step, particularly of the kind Pyr-Pur (Pyr = pyrimidine, i.e. A or G; Pur = purine, i.e. T or C)

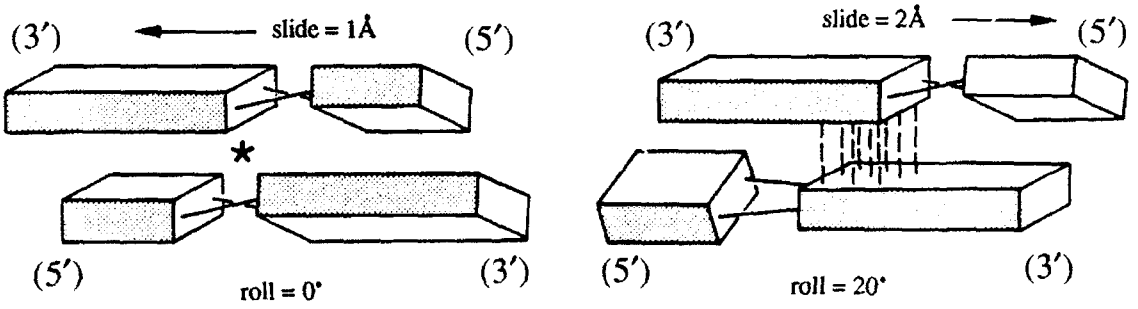


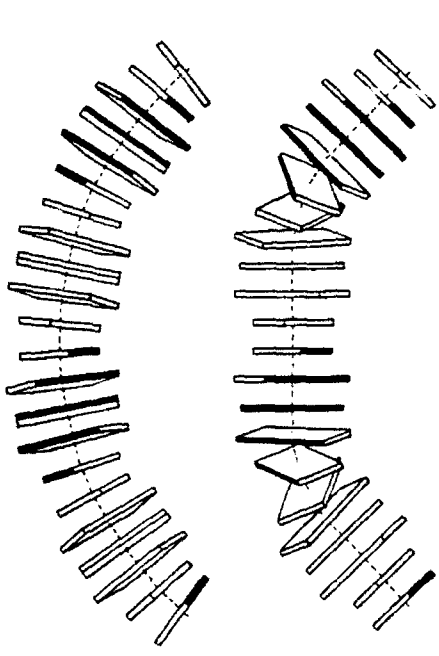
Figure 16. Two distinct stacking-arrangements of a base-pair step, showing the coupling between roll and slide on account of propeller-twist. The sugar-phosphate chains run in the "5′ → 3′" direction (cf. Figure 9), so the smaller (Pyrimidine) base precedes the larger (Purine) base on each chain.

adopt *either* a high-slide, high-roll *or* a low-slide, low-roll configuration; but not an intermediate form. This is the situation shown in Figure 16. It seems that it is base-pair steps of this kind which produce the "switch" feature of the transition in a tract of DNA. Other types of step prefer a fixed configuration: for example, AA steps always have a low-roll, low-slide shape; and thus they are reluctant to move over into the "A"-form.

3.5. How does DNA curve around a histone spool?

The bending of DNA is more complicated than the bending of an ordinary elastic rod for two main reasons. First, it is clear that the *helical structure* of the DNA will play an important part in bending. Second, the relative rotations at base-pair step level which are necessary for changes in curvature in DNA can occur most easily by means of "roll" within the individual base-pair steps.

Figure 17(a) shows schematically a piece of DNA which has been bent into a curve of about 4.5° per step, on average [20]. (This is precisely the degree of curvature required for bending DNA around a histone spool: 360° in 80 steps.) The curvature is produced by positive roll in some base-pair steps, but negative roll in others which are half-a-turn of helical twist away. Simple geometry shows that curvature requires a *periodic variation of roll*, at a period equal to the helical repeat of the DNA. In Figure 17(a) the roll angles vary sinusoidally with position along the molecule. A different way of achieving the same overall curvature is to have alternating patches of "B"-form and "A"-form DNA (as in Figure 13), with the pattern



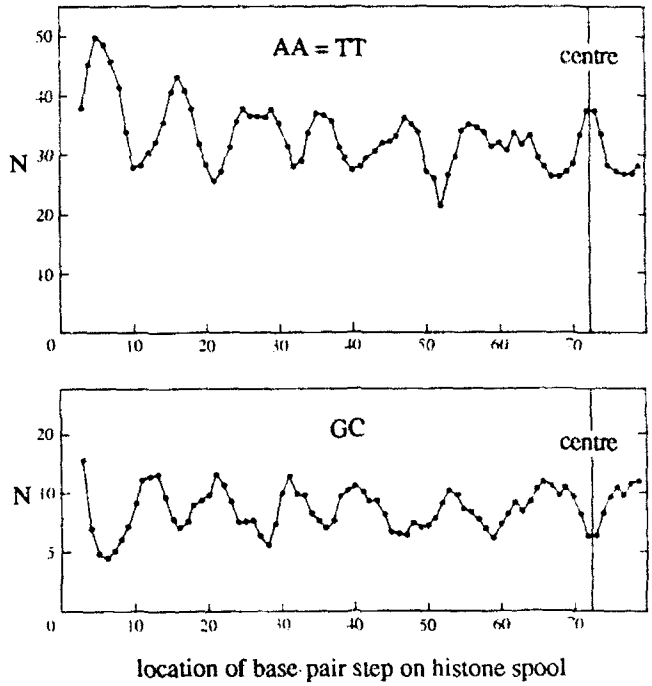


Figure 17 (*left*). Two double-helical turns bent through a 90° curve (a) by a smooth variation of roll angles and (b) by step-wise variation of roll.

Figure 18 (*right*). Total number N of (a) AA = TT and (b) GC base-pair steps in (many pieces of) curved chicken DNA, plotted against location on the histone spool (cf. Figure 10). The overall length of the DNA is 145 base-pairs, but the data were symmetric, and so could be plotted, as shown, from end to centre.

repeating every 10 base pairs, as in Figure 17(b). Plainly, there is no unique way of imparting a given curvature to DNA.

Now if a piece of ordinarily-straight DNA is to be curved tightly around a histone spool, it is necessary for *changes* of roll to occur at some of the base-pair steps along the molecule. It should be clear from previous remarks that some specific types of base-pair steps prefer a low-roll position, while others prefer a high-roll position; and that others readily adopt either a low-roll or a high-roll configuration.

Figure 18 shows the results of sequencing pieces of DNA from chicken blood cells which are known to have been wrapped twice round histone cores [31]. More than one hundred distinct 145-base-pair lengths of DNA of this kind were sequenced, and statistical analyses were performed on the positional preferences for various kinds of base-pair step. Without going into much detail, we can see immediately that the frequency of occurrence of steps AA and steps GC at different positions around the circumference of the spool varies periodically with position. In both cases the periodicity is about 10 base pairs; i.e. it corresponds to the helical repeat of the DNA. But closer examination shows that AA steps prefer to occupy "low-roll" positions – where the minor-groove side of the DNA faces inwards – while GC steps prefer "high-roll" positions.

Time does not permit me to go through the details of this substantial assay; but I may remark that the positional preferences of this kind that are shown by particular dinucleotide steps are in general accord with configurations that are found in single-crystal assays on the same types of step.

3.6. How do proteins recognise particular sequences of DNA?

There are many biological processes in which a particular protein has to search, and find, and attach itself to a specific sequence – such as ACAATATATATTGT – in a very long piece of DNA. Such a sequence, for example, might mark the start of a gene.

Recognition processes, whether between protein and protein or protein and DNA, have a strong geometrical component. Thus, for strong binding or attachment of one to the other there must be not only a good geometric fit but also an arrangement of atoms within the mating surfaces which allow the formation of a sufficient number of specific hydrogen bonds. In this way, particular proteins can find and bind to specific sequences of DNA by what is usually called a "direct reading" process.

But there is also a second, different element in the recognition process, which has only recently been elucidated. The binding of "434 repressor" to the 14 base-pair DNA sequence ACAATXXXXATTGT provides a good example [32]: here X denotes a variable base.

For binding to take place, the two 5-letter end-sequences of the DNA must be exactly as given above. But it turns out that a rather wide variation of sequence is tolerable within the

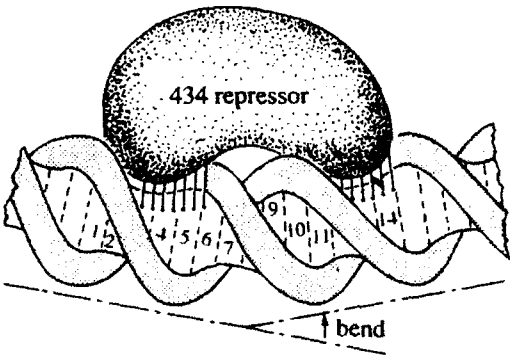


Figure 19. Schematic view of the binding of 434 repressor protein to DNA, showing a central "bend" of about four base-pairs which are not in direct contact with the protein.

central four-base region. And indeed the *strength of binding* depends significantly on the DNA sequence of this somewhat variable portion.

Figure 19 shows a sketch of the binding arrangement. The two end regions of the DNA are tightly bound to the protein by "direct reading" recognition. Notice that the shape of the protein is such that the central portion of DNA has to *bend* through about 20° in order to fit. Now we have seen above that different dinucleotide steps have different positional preferences in DNA that is curved; and it turns out that there is an excellent correlation between the strength of binding on the one hand and the position preferences – as determined from the chicken blood cell DNA wrapped around histone spools (cf. Figure 18) – corresponding to the sequence of the central region in the 434 recognition sequence.

4. CONCLUSION

In this lecture it has been appropriate to present the various thoughts and ideas by means of words and pictures. In particular, I have not employed any mathematical symbolism, although in a full account (e.g. [23]), there would be plenty of mathematics. (For example, a full treatment of base-pair stacking in a piece of DNA clearly requires careful computation of finite angular rotations.)

In fact, the main task in each of the two problems which I have described has been to set up a suitable conceptual reference-frame for thinking about the structural mechanics of the situation. As I have tried to demonstrate, the resulting models enable us to explain, within the framework of structural mechanics, a rather wide range of physical phenomena.

The thought-processes which lead to these appropriate models have involved structural mechanics at a most elementary level. Thus, for example, in each of the two problems we have used, in different ways, the primitive observation that a rigid body has six degrees of freedom in three-dimensional Euclidean space. It is perhaps rare nowadays to find problems where the key to the solution lies in such a simple and elementary idea. But we have, of course, also employed other considerations than those of elementary kinematics. For the most part these have involved, explicitly or implicitly, some *linearisation* of the various equations and an invocation of elementary elasticity, coupled with some simple ideas of mechanical bistability. The fact that the various models produce satisfactory results in spite of ample linearisations suggests that the underlying ideas are both robust and appropriate to the problems under investigation. And indeed, we may perhaps put both of our problems into the category of those that are solved in the first instance by a judicious amount of *oversimplification*. It remains to be seen, of course, whether there are many more problems in molecular biology which will yield to the same mode of attack.

5. REFERENCES

1. Calladine, C.R. (1975) *Nature* **255**, 121.
2. Calladine, C.R. (1976) *J. theor. Biol.* **57**, 469.
3. Calladine, C.R. (1978) *J. molec. Biol.* **118**, 457.
4. Calladine, C.R. (1983) Sci. Prog., Oxf., 365.
5. Klug, A. (1967) *Symposium of the International Society for Cell Biology*, Vol. 6 (ed. by J.F. Danielli) pp. 1–18. Academic Press, New York.
6. O'Brien, E.J. & Bennett, P.M. (1972) *J. molec. Biol.* **70**, 133.
7. Asakura, S. & Iino, T. (1972) *J. molec. Biol.* **64**, 251.
8. Shimada, K., Kamiya, R. & Asakura, S. (1975) *Nature* **254**, 332.
9. Wakabayashi, K. & Mitsui, T. (1972) *J. Phys. Soc. Japan* **33**, 175.
10. Perutz, M.F. (1970) *Nature* **228**, 726.
11. Kamiya, R. & Asakura, S. (1977) *J. molec. Biol.* **108**, 513.
12. Hotani, H. (1982) *J. molec. Biol.* **156**, 791.

13. Macnab, R.M. & Koshland, D.E. (1974) *J. molec. Biol.* **84**, 399.
14. Macnab, R.M. & Ornston, M.K. (1977) *J. molec. Biol.* **112**, 1.
15. Shirakihara, Y & Wakabayashi, T. (1979) *J. molec. Biol.* **131**, 485.
16. Namba, K., Yamashita, I. & Vonderviszt, F. (1989) *Nature* **342**, 698.
17. Trachtenberg, S. & DeRosier, D.J. (1991) *J. molec. Biol.* **220**, 67.
18. Vonderviszt, F., Aizawa, S.-I. & Namba, K. (1991) *J. molec. Biol.* **221**, 1461.
19. Calladine, C.R. & Drew, H.R. (1984) *J. molec. Biol.* **178**, 773.
20. Calladine, C.R. & Drew, H.R. (1986) *J. molec. Biol.* **192**, 907.
21. Calladine, C.R., Drew, H.R., & McCall, M.J. (1988) *J. molec. Biol.* **201**, 127.
22. Drew, H.R., McCall, M.J. & Calladine, C.R. (1990) *DNA topology and its biological effects* (ed. by N.R. Cozzarelli & J.C. Wang) Chapter 1, pp. 1–56. Cold Spring Harbour Laboratory Press.
23. Calladine, C.R. & Drew, H.R. (1992) *Understanding DNA: the molecule and how it works.* Academic Press, London.
24. Cozzarelli, N.R., Boles, T.C. & White, J.H. (1990) *DNA topology and its biological effects* (ed. by N.R. Cozzarelli & J.C. Wang) Chapter 4, pp. 139–184. Cold Spring Harbour Laboratory Press.
25. Fuller, F.B. (1971) *Proc. Nat. Acad. Sci. USA* **68**, 815.
26. Judson, H.F. (1979) *The eighth day of creation.* Simon & Schuster, New York.
27. Alden, C.J. & Kim, S.-H. (1979) *J. molec. Biol.* **152**, 723.
28. Wing, R.M., Drew, H.R., Takano, T., Broka, C., Tanaka, S., Itakura, K. & Dickerson, R.E. (1980) *Nature* **287**, 755.
29. Kennard, O. & Hunter, W.N. (1989) *Quart. Rev. Biophys.* **22**, 327.
30. Nelson, H.C.M., Finch, J.T., Luisi, B.F. & Klug, A. (1987) *Nature* **330**, 221.
31. Satchwell, S.C., Drew, H.R. & Travers, A.A. (1986) *J. molec. Biol.* **191**, 659.
32. Koudelka, G.B., Harrison, S.C. & Ptashne, M. (1987) *Nature* **326**, 886.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Viscous fingering as a pattern forming system

Y. Couder

Laboratoire de Physique Statistique de l'Ecole Normale Supérieure
24 rue Lhomond, 75231 Paris Cedex 05. France

The conference I gave in Haifa corresponded very closely to a review article that I had written earlier[1]. In order to avoid repetition I will only give here a brief summary. The interested reader can find more complete information in the above-mentionned article and in the references therein.

Saffman Taylor viscous fingering is the process which occurs in a Hele Shaw cell when a fluid of small viscosity forces a fluid of large viscosity to recede. In such a situation the pressure field in the whole cell is Laplacian and the interface, which moves with a velocity proportional to the local pressure gradient, is unstable. As the interface is stabilizedat small scales by surface tension, the instability has a characteristic length-scale called the capillary length l_c which can be found by the linear stability analysis of a plane front. Experimentally the viscous fingering is an unbounded instability and creates typical patterns, the problem being to understand their formation.

Amongst the large variety of morphogenetic instabilities, two are of a particular interest for comparison with Saffman Taylor viscous fingering: the dendritic crystal growth and the diffusion-limited aggregation (DLA).

Dendrites are formed during the growth of a monocrystal in a melt or in a solution in conditions far from equilibrium. Here, due to the growth of the solid its surface moves with a velocity which is proportionnal to the gradient of either the temperature or the impurity concentration. Both these quantities have a diffusive field, comparable to a Laplacian field in the limit where the diffusion length is very large. The front is stabilized by surface tension through the Gibbs-Thomson effect so that there is also a typical length scale l_s to this instability. The dendrites form needles with a parabolic tip, the direction of their growth corresponding to one of the main cristallographic directions of the cristal.

Diffusion limited aggregation is a model system introduced by Witten and Sander in which random walkers moving on a lattice are emitted far away from a central seed. If a walker visits any of the sites neighbouring this seed it remains there. The process is repeated, each walker stopping whenever it visits a site neighbouring an already occupied point. It was shown that the probability of visit of a given site by the walker obeys a Laplacian law and that a region of the aggregate grows with a velocity proportionnal to the local gradient of this probability. The typical scale l_u of this system is simply the mesh size of the computing grid.

In very general terms the growth morphologies obtained in Saffman Taylor fingering and their comparison with those of the two other systems show that three

factors are determinant.
- The geometry: for Laplacian systems the geometry of the fixed boundaries define families of possible analytic solutions for the possible shapes of the interface as well as the selection amongst them.
- The isotropy or anisotropy of the system
- The length scales involved in the growth. These are usually the instability length scale l_c and the total width of the moving front. If these length scales are of comparable order of magnitude stable curved fronts can be obtained. If they are very far apart, fractal growth will occur in the intermediate range separating them.

The geometry.

The simplest problem in viscous fingering is the growth in a linear channel. If the width is of the order of a few l_c a single finger moving steadily along the cell is obtained occupying half of the cell's width. This is one of the archetypes of stable curved fronts and its investigation has a long history. A family of solutions was found in the initial work of Saffman and Taylor if surface tension was neglected. However it is only much later that the selection of the actually observed fingers was understood as resulting from the effect of isotropic surface tension. The only other geometries in which the understanding of the stable fronts has reached the same level is that of the recently introduced sector-shaped cells. In these geometries the observed fingers also occupy a fraction of the cells' angular width. This fraction is an increasing function of the cell's angle.

The isotropy or anisotropy of the system.

If either a local disturbance of the tip or anisotropy is imposed to a Saffman-Taylor finger growing in a channel, the finger becomes narrower and more stable. In the limit of a wide channel it becomes obvious that the finger's tip is parabolic, is followed by lateral side branching and has the general aspect of a cristalline dendrite. The introduction of these disturbances thus lifts the specific selection due to isotropic surface tension. The new finger is no longer selected by its width but by the radius of curvature at its tip which is proportionnal to l_c. The striking aspect of this result is that this is precisely, due to their anisotropy, the type of selection of the dendrites. The isotropy or anisotropy of a system is thus an essential ingredient in its growth morphology because it changes the selection mechanism. In general terms the isotropic fronts undergo a selection related to the large scale of the system while the anisotropic ones are selected by the small scale of the system.

The length scales.

Within the limit of infinite viscosity contrast and very rigid glass plates, it is possible to obtain fractal growth in viscous fingering. This occurs whenever the length scale of the front width is very large compared to l_c and was mostly observed in radial geometries. The resulting fractal dimension is similar to that of diffusion limited aggregation.

Recent experiments on very unstable Saffman fingering were done in those geometries in which the stable curved fronts are well known, i.e. the linear channel and the sector shape cells. For each of these experiments a large number of identical runs were performed so as to obtain a statistical analysis of the occupancy of the cell by the patterns. The result is that the region of the cell with a mean occupancy larger

than average has the shape and width of the stable curved front in this geometry. It thus appears that the selection mechanism has survived the finger instability. Diffusion limited aggregation numerical experiments were done in the same geometries. They form patterns that are fractal in a much larger range because the scales are further apart from each other. However their averaging produces the same results. We can note that this averaging method works both for isotropic or anisotropic fractal growth, the average isotropic fractal pattern being selected by its large scale and the average anisotropic patterns on its small scale. A theoretical interpretation of these results in terms of a modified mean field theory has been given recently.

(1) Y. Couder, Growth Patterns: from stable fronts to fractal structures in *Chaos, Order and Patterns* , Edited by R. Artuso, P. Cvitanovic and G. Cassati, Plenum New York 1991

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Mechanics in sport

G. Grimvall

Department of Theoretical Physics, Royal Institute of Technology,
S–100 44 Stockholm, Sweden

Abstract

Sports offer many examples that illustrate both simple and complex aspects of mechanics. This paper considers examples that require only an introductory undergraduate course in mechanics, but often show surprising or thoughtprovoking results. Several of them will deal with official rules for track and field events, and the implication of those rules on the recorded result. Other examples deal with the precision by which some movements must be executed, and the surprising ability of the human brain to cope with this. One aim of the paper is to provide the unversity teacher with material that can be used to illustrate some principles of mechanics.

1. INTRODUCTION

There is a rich but scattered literature on various aspects of the mechanics of sport. Frohlich [1] has written a "resource letter" on the physics of sports, with references to general books and also to 90 specialised articles on particular sports, ranging from archery and athletics to cycling and karate. There is also a selection of reprints on the physics of sports [2]. In the general field covered by the present article, two books stand out as exceptionally good. One is "The physics of ball games" by Daish [3]. It is divided into two parts The first part contains a general treatment of various ball games, and the second part gives a mathematical treatment at the level of undergraduate courses. The other book is "Dyson's mechanics of athletics", originally written by Dyson and with its latest edition revised by Woods and Travers [4]. It gives a detailed description of the mechanics of track and field events, with a large number of illustrative drawings, but with a minimum of algebra.

Much of what is published on the mechanics and physics of sports deals with sports that are strong in the US, e.g., baseball and golf. In addition to references given by Frohlich [1], one may mention books by Watts and Bahill [5], Adair [6], Schrier and Allman [7], Brody [8], Cochran and Stobbs [9], Brancazio [10] and de Mestre [11], and recent papers on the aerodynamics of sports balls (baseball, golf, cricket) [12] and on features of a baseball bat [13]. Below we shall discuss how an athlete's result may depend on parameters that are restricted, by the rules of the sport, to have certain values. In major league baseball, the coefficient of restitution of the ball is required to be 0.546 ± 0.032. Kagan [14] has considered the mechanical implications of such a variation.

The classic book "Bicycling science" by Whitt and Wilson is out in a new and revised edition [15]. It contains many data on the energy required in transport, with emphasis on bicycles.

Biomechanics is not treated in the present paper. This broad field covers subjects ranging from how runners save energy by bouncing on their tendons and ligaments and on their running shoes [16] to how a videofilm of athlete's motions can be digitalised to guide improvements in the performance in sports [17]. The book by Hay [18] is a standard reference on biomechanics and sport.

2. EFFECT OF INCLINED TRACKS AND FIELDS

The international rules for the athletic throwing events (shot, discus, hammer, javelin) say that "the overall downward inclination of the landing sector, in the throwing direction, shall not exceed 1:1000". (In this paper, exact quotes from the International Amateur Athletic Federation Handbook 1992–93 [19] are within quotation marks.) Similarly, the overall inclination of the tracks in the running direction shall not exceed 1:1000. What is the advantage of a maximum allowed inclination in throwing and in 100 m sprint?

2.1. Throwing events

Let the implement strike the field at a distance L from the athlete. With a maximum allowed inclination, that point lies L/1000 below the true horizontal plane. If the trajectory of the landing implement makes an angle φ to the horizontal plane, it will move an additional distance $L/(1000 \tan\varphi)$ before it strikes the field. 1. In hammer throwing the trajectory follows fairly well an ideal parabola, with $\varphi \approx 45°$. Then $\tan\varphi \approx 1$ and we get the simple rule that the recorded length is as much longer as the landing point lies below the ideally horizontal field, see Figure 1. The world record is

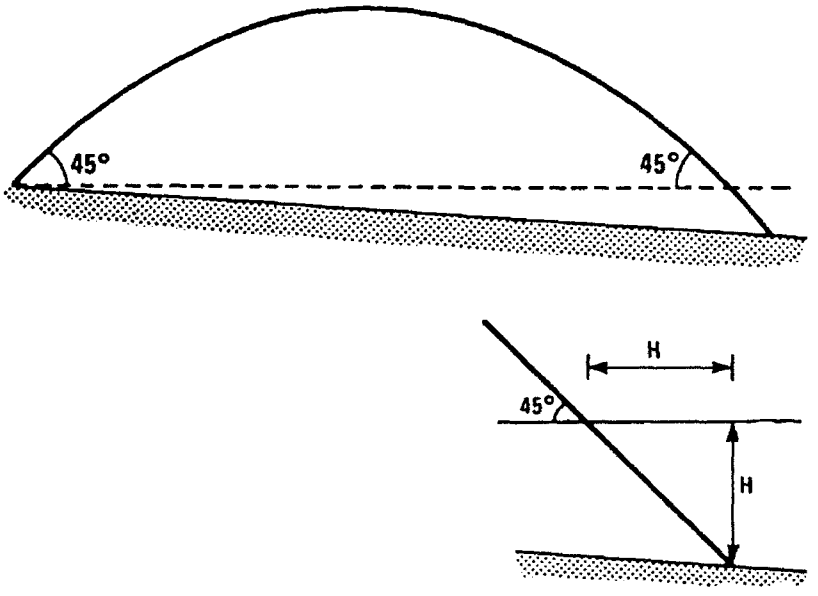


Figure 1. A throw following the parabola of maximum range will come as much farther as the landing point lies below the ideally horizontal plane.

presently 86.74 m, so the advantage of a maximum allowed inclination is about 9 cm. We remark that in hammer throwing (and also in discus and javelin) the recorded length is given to the nearest even centimeter below the measured length, or the measured length if that is obtained as a whole even centimeter. In shot putting, the recorded length is given in whole centimeters.

In javelin throwing the measurement is made at the point where the metal head first strikes the ground, even if the javelin afterwards falls flat. The javelin may often seem to land at a small angle φ to the horizontal plane, but this is largely an illusion. Even if the javelin axis is almost horizontal, the trajectory of the center of gravity (and hence also of the metal head) approaches the ground more steeply. For an estimate we can take $\tan\varphi \approx 1$, as for hammer throwing. The present world record is 94.74 m, so also in javelin throwing does the maximum allowed inclination of the field give an advantage of about 9 cm.

In shot putting the world record is 22.47 m, and hence the advantage of an inclined field is at most about 2 cm. Shot puts that under ideal measurements differ by 0.9 cm may sometimes be recorded as being of equal length, according to the rule cited above.

2.2. 100 m sprint

The throwing events allowed a satisfactory treatment based on elementary mechanics. The sprint events are considerably more difficult to discuss. For illustrative purposes we shall treat them under some elementary, although possibly dubious, assumptions.

As a rule of thumb, a reasonably well trained person walking in mountain areas can climb 300 m in a hour. The famous Baedeker tourist guide suggests 320 m/h for vertical ascent, while the instruction of the Austrian army says 300 m/h when carrying a load. These numbers correspond to a power of about 80 W, for the increase in potential energy alone. When the climb is of more limited duration, man can produce a much higher power. For instance, the 1987 record for running up the stairs in the Empire State Building in New York (320 m vertically) is 10 min 59 s, corresponding to a power of somewhat more than 300 W for the change in potential energy. All the previous excercises require the burning of oxygen. In sprint, the athlete makes direct use of chemical energy stored in the muscles. To get a crude estimate we assume that he can develop 1.5 kW.

Now consider 100 m sprint, and let there be an *upward* inclination of 1:1000, i.e. a total of 0.1 m. Suppose that the athlete needs an extra time t', so that the energy Pt' corresponds to the increase in potential energy due to the slope. With a power $P = 1.5$ kW, t' will be about 0.05 s. For all races up to and including 10 000 m, times are recorded to within 0.01 s.

If there is a downward slope along the running direction, it is not obvious how the gain in potential energy can be used to increase the speed. However, between two consecutive steps the body follows a trajectory that is similar to that described by the implements in throwing events, although with a much smaller angle φ. If the character of the leg motion is not changed due to the inclined track, each step in free flight will be longer by a factor $[1 + 1/(1000 \tan\varphi)]$. Therefore, with an inclined track the steps need not be taken so often and the muscles may provide more energy per step, leading to an increased running speed. With $\tan\varphi = 1$ and a power consideration as above we again get about 0.05 s advantage. With a realistic and much smaller $\tan\varphi$, the advantage would be quite significant. However, it is not clear that the power can be used as we have assumed, and it is not clear that the inclined track will not slightly but detrimentally affect the leg motion.

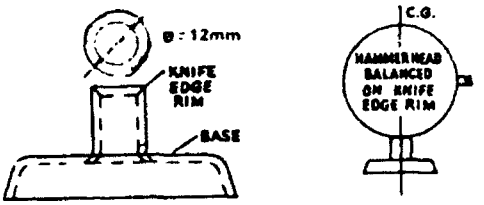
As a final illustrating calculation we again assume that the athlete can directly benefit from the gain in potential energy. McFarland [20] considered the effect of the reduced air pressure in Mexico City and assumed that the reduced energy waste on air

resistance, compared to the sea level, could be used directly to increase speed. From his data we find that a gain in energy by about 1 kJ during the 100 m race results in a time that is shorter by 0.09 s. In our case an athlete of mass 75 kg gains 73 J in potential energy on a track with maximum allowed inclination, corresponding to an advantage of 0.007 s, i.e., a negligible amount.

3. RULES ON THE EQUIPMENT

3.1. The hammer

A hammer consists of a metal head, a wire and a grip. "The head shall be of solid iron or other metal not softer than brass or a shell of such metal, filled with lead or other solid material with a minimum diameter of 110 mm. It must be completely spherical in shape." Further, its "centre of gravity shall be not more than 6 mm from the centre of the sphere". The centre of gravity is normally not at the centre of the sphere, in spite of the requirement that the head shall be "completely spherical in shape", because the head must have some kind of attachment to the string. The rules suggest an ingenious way to test whether the head satisfies the requirement on the centre of gravity. "It must be possible to balance the head, less handle and grip, on a horizontal sharpedged circular orifice 12 mm in diameter", cf. Figure 2. Of course it is then understood that the head must balance for all of its orientations.



Suggested apparatus for testing centre of gravity of Hammer head

Figure 2. Illustration in the International Amateur Athletic Federation Handbook [19].

3.2. The hammer cage

"All hammer throws shall be made from an enclosure or cage to ensure the safety of spectators, officials and competitors. The cage should be designed, manufactured and maintained so as to be capable of stopping a 7.26 kg hammer head moving at a speed of up to 29 metres per second." One may wonder, why 29 m/s?

The well known expression for the length s of a throw over a horisontal plane, with an elevation angle α to that plane and with the neglect of air resistance, is

$$s = \frac{v^2}{g} \sin(2\alpha) \qquad (1)$$

The maximum range is v^2/g, for $\alpha = 45^o$. The present world record in hammer throwing is 86.74 m. Therefore a lower limit to the velocity v is 29.2 m/s, neglecting the small correction related to the fact that the launch is not exactly from where the distance is measured. Since air resistance is not entirely negligible (cf. Section 4.1) the actual speed at which it might hit the cage in a world record throw must be somewhat larger than 29 m/s, although not as much as 30 m/s. Obviously the rules do not allow for a safety factor.

When throwing the discus one may either use a (slightly modified) cage for hammer throwing or a specially designed cage. The latter should be "capable of stopping a 2 kg discus moving at a speed of up to 25 metres per second". The present world record is 74.08 m. Arguing as above for the hammer, the discus record corresponds to a release speed of at least 26.9 m/s, i.e., significantly higher than the design value of 25 m/s for the cage. The apparent inconsistency is resolved if we note that a discus "rides" on the air and therefore may travel farther in a normal atmosphere than in vacuum. Actually, a head wind will often increase the length of a discus throw [4,21].

3.3. Geometrical specifications on the implements

The hammer is specified geometrically by the diameter of the head (110 – 130 mm), by the entire length of the hammer measured from inside the grip (1175 – 1215 mm) and by the position of the centre of gravity (±6 mm). All three parameters affect the distance D from the grip to the centre of gravity, which thus may vary from $1175 - 65 - 6 = 1104$ mm to $1215 - 55 + 6 = 1166$ mm. To this we should add a distance of almost 1 m, corresponding to the length of the arms. However, it is difficult to analyse how D (with arm length added) affects the length of throw. The release velocity depends on both D and on the angular velocity ω of the string at the moment of release. An increase in D is likely to decrease ω. In fact, if ω had been independent of D, the largest and the smallest D allowed by the rules would result in throws that differ in length by 12 %, i.e., about 10 m for a world class result. Obviously an argument based on a constant ω is useless. In this context it can also be remarked that the rotational motion of the hammer does not have the athlete as its centre. If that were the case, the horizontal centripetal force would be up to 3 times the weight of the athlete and that can not be achieved with any reasonable frictional force between the athlete and the ground. The motion is more like that of the Earth and the Moon, rotating about a common centre of gravity.

In discus and javelin throwing, the air affects the trajectory of the implement. This is reflected in the precise specifications on the implement geometry. The discus geometry is specified by 5 lengths, and the javelin by 5 lengths along its axis and 10 values referring to various diameters. In 1984 new rules were introduced for the javelin, moving the centre of gravity to the rear by 4 cm and making the rear end somewhat thinner. That reduced the world record by about 15 m. In hammer throwing the effect of the air is difficult to estimate more precisely, because of the string and grip attached to the head. In Section 4.1 we crudely estimate it to reduce the length of a throw by about 2 m. In major events the hammers are provided by the organisors. Since the diameter of the head, according to the rules, should lie between 110 mm and 130 mm, two acceptable hammers may differ in the area of the head by 40 %, while still having the same mass. Air resistance is proportional to the cross sectional area, so the two hammers may be expected to give recorded lengths that differ by more than a meter (cf. Section 4.1).

Continuing on an analogous theme, we note that in discus throwing the range of supply of equipment for competition must be in the range 2.005 kg to 2.025 kg, according to the rules. It is difficult to calculate what a difference by 0.020 kg in mass does to the length of a throw because the ability to exert muscular force on the implement is not independent of its opposing inertial force. However, we are somewhat

helped by the fact that the discus for women has a mass of 1 kg, and the world record (76.80 m) is longer than that for men (74.08 m). In view of known differences between men and women in other sports requiring strength, it seems likely that the mass difference of 1 kg affects the length of a throw by, say, 20 m. A mass difference of 0.020 kg would then correspond to almost half a meter in the recorded result, if we can extrapolate linearly from the 1 kg through the 2 kg discus.

4. AIR RESISTANCE AND WIND VELOCITY

Air resistance affects the results not only in the throwing events but in many other sports as well. One may think of the world record by Beamon in the long jump at the Olympic games in 1968 in Mexcico City, that is often given as an example although we shall see below that the effect was quite small. In golf there is not only a retarding air resistance. The Magnus effect on a spinning ball may help to carry the ball typically 75 m farther than in an air and 50 m farther than in vacuum [9]. In this Section we shall compare a dimensionless "figure–of–merit" for the ordinary air resistance on various balls and then discuss wind speed versus air pressure for some sports.

4.1. A figure–of–merit for air resistance

The standard expression for the air resistance F acting on a body with a smooth surface moving with a speed v in air of density ρ is

$$F = \frac{1}{2} C_d A \rho v^2 \tag{2}$$

Here A is the cross sectional area and $C_d(Re)$ is a drag coefficient that varies with the Reynolds number Re. For a sphere with radius r,

$$Re = \frac{2rv}{\nu} \tag{3}$$

where ν is the kinematic viscosity of air. For a sphere, in air at normal pressure and temperature, $Re = 12.8 \cdot 10^4 rv$, if v is expressed in m/s and r in m. For $10^3 < Re < Re_c \approx 2 \cdot 10^5$, and for a sphere, we may take $C_d = 0.5$. When $Re > Re_c$, $C_d(Re)$ decreases significantly, i.e., the air resistance is lowered.

In a uniformly retarded linear motion, the distances covered in a time t, with initial velocity v_o and retardation a, is

$$s = v_o t - \frac{1}{2} a t^2 \tag{4}$$

If a is small it is useful to introduce a dimensionless quantity ξ that measures how the retardation affects s. We define

$$\xi = \frac{at^2}{2v_o t} = \frac{at}{2v_o} \tag{5}$$

which yields

$$s = v_o t(1 - \xi) \tag{6}$$

Similarly, the velocity v after a time t is

$$v = v_o - at = v_o(1 - 2\xi) \tag{7}$$

Let m be the mass of the moving body, and take a = F/m, with F from Eq. (2). Further, let $t = t_o$ be the time of interest (e.g., the time to complete the trajectory of a ball). Since t_o is usually not directly available, we replace it by ℓ_o/v_o where ℓ_o is a characteristic length, e.g., the length of a throw. Then

$$\xi = \frac{1}{2} C_d A \rho v_o^2 \left(\frac{\ell_o}{v_o}\right) \left(\frac{1}{2 v_o m}\right) = \frac{1}{4} \left(\frac{C_d A \rho \ell_o}{m}\right) \tag{8}$$

In Table 1 are summarised characteristic parameters and the resulting ξ for some sports. For all cases we have taken $C_d = 0.5$. When the trajectory of the object is not straight, the relation between t_o, v_o and ℓ_o is complex. We have taken v_o to be the speed of release etc., without resolving it into vector components. The distance ℓ_o is chosen to be the straight distance between the end points of the trajectory of interest. Variations in ξ by 50 % are not significant.

Table 1
Parameters governing the air resistance

Sport	Area	Mass	Speed	Length	ξ	Re
	A (cm^2)	m (kg)	v_o (m/s)	ℓ_o (m)		
Shot	95	7.26	15	22	0.005	100 000
Hammer	95	7.26	29	90	0.02	200 000
Golf	14	0.046	70	200	1	190 000
Tennis	32	0.058	60	20	0.2	240 000
Table tennis	11	0.0025	25	7	0.5	60 000
Soccer	380	0.43	30	30	0.5	420 000
Long jump	5000	70	10	8	0.01	500 000

The values in Table 1 for the ball games do not include any effect of a spinning ball. The density of air is taken to be that at sea level. The velocities are those at the start of the trajectory, for world class performances.

We see that for soccer $Re > Re_c$ so C_d is significantly smaller than the value 0.5 assumed in Table 1. This reduces ξ to about 0.3. A kick that would carry the soccer ball about 40 m if the standard value $C_d = 0.5$ is used, actually travels about 60 m [3]. In golf $Re < Re_c$ according to Table 1, but our estimate of Re_c refers to a smooth surface of the ball. Golf balls have dimpled surfaces which considerably reduces the air resistance. For long jump, Re estimated from a sphere may be less relevant.

The approach in the beginning of this Section assumed that $\xi << 1$. It is obvious that the ball games cannot be treated by considering air resistance as a small perturbation.

4.2. Wind speed versus air pressure

The expression (2) for the air resistance contains a velocity v which of course is to be taken relative to the air. A tail wind with a speed v' would reduce the retarding

force F. This is insignificant when $\xi << 1$. In fact, ξ does not contain $v - v'$ because the tail wind gives a second–order effect when $\xi << 1$.

In the track and field events long jump, triple jump and running up to and including 200 m, the international rules state that "if the wind velocity measured in the direction of jumping" (etc.) "averages more than 2 metres per second, the record will not be accepted". The wind speed is measured and averaged automatically, during a certain time that varies with the event.

The air resistance depends not only on the speed through the air but also on the air density, i.e., indirectly on the altitude. The air pressure at sea level is about 1010 hPa. In Mexico City, located at an altitude of 2200 m, the pressure is about 780 hPa. From the gas law, $pV = RT$, we get that the density ρ is proportional to the pressure p, at equal temperatures T. Hence ρ is 23 % lower in Mexico City than in, e.g., Rome. Now compare an event in still air in Mexico City with an event with a tail wind in Rome. The air resistance in the two cases is the same if there is a tail wind v' in Rome so that $(v–v')^2/v^2 = 780/1010$. When $v \approx 11$ m/s, as is approximately the case in 100 m sprint, $v' = 1.3$ m/s. McFarland [20] has considered how much the lower air resistance in Mexico City affects the time for 100 m sprint. He allows for the varying speed during the run, and assumes that the kinetic energy (i.e., the speed) increase corresponds to the lower energy loss due to air resistance at low pressure. This model calculation suggests that running in Mexico City rather than at sea level, and in both cases in still air, gives an advantage of about 0.09 s. Based on this result we would expect that a maximum tail wind of 2 m/s gives an advantage of about 0.13 s.

The combined effect of the low air density in Mexico City and a maximum allowed tail wind of 2.0 m/s there, would be equivalent to competing at a place where the air pressure is reduced to 52 % of that at sea level. This combined effect may lower the time by 0.2 s in 100 sprint.

We finally note that a change in temperature affects the air density. An increase by 20 °C, which is not unreasonable if we compare, e.g., Stockholm and Rome, corresponds to a difference in ρ by 6 % when the pressure is held constant. That means an advantage by 0.02 s in 100 m sprint, but only about 1 cm in long jump. In hammer throwing its effect is larger than about 10 cm, i.e., larger than the effect of a maximum allowed inclination of the field.

5. RECORDS AND GRAVITATION

The acceleration of gravity, g, varies from place to place. One might first think that high altitudes would reduce g significantly, but that is a small effect compared to the variation of g with the latitude φ. It is a good approximation to take

$$g = 9.832 - 0.052 \cos^2\varphi \ \ (m/s^2) \qquad (9)$$

The dependence on φ has two causes, one being a "direct" centrifugal effect due to the rotation of the earth and the other coming indirectly from the fact that the rotation also changes the shape of the earth to a somewhat flattened sphere. The actual acceleration of gravity may differ slightly from the result of Eq. (9) due to local variations in the mass distribution on the Earth, but for our purposes Eq. (9) is sufficient. From it we get the results of Table 2.

Table 2
The acceleration of gravity, g (m/s^2), in some cities.

Stockholm	9.818	Rio de Janeiro	9.788
Berlin	9.813	São Paulo	9.789
Mexico City	9.780	Melbourne	9.800

Now consider a standard parabolic trajectory, for which $s = v^2\sin(2\alpha)/g$. With a given launch speed v and elevation angle α, a change in g changes the distance s by Δs. From Eq. (9) we can write

$$\Delta s/s = |\Delta g/g| = 0.0053 \cos^2\varphi \qquad (10)$$

One of the classic world records is that of the Japanese Tajima, whose 16.00 m in triple jump at the Olympic games in Berlin in 1936 was not matched until 1950, when Ferreira da Silva from Brazil also reached 16.00 m, in Sao Paulo. The following year he jumped 16.01 m in Rio de Janeiro. However, we see from Table 2 that the lower acceleration of gravity in São Paulo and Rio de Janeiro, compared to Berlin, corresponds to an advantage of about 4 cm for Ferreira da Silva. This fact started a public discussion at the time, whether not Tajima's result was still the best. The same argument, when applied to hammer throwing, gives an advantage to competitors in Mexico City, over those in say Stockholm, by 34 cm. This is from the difference in gravity alone. About an equally large advantage may come from the lower air pressure, and hence reduced air resistance on the hammer, in Mexico City, cf. Section 4.2.

McFarland [20] has simulated numerically a long jump, a triple jump and a high jump in Mexico City and in Moscow, allowing for variations both in gravity and in air resistance. He concludes that the famous world record by Beamon got an extra length by about 2 cm from the decreased gravity, 3 cm from the reduced air pressure and another 3 cm from the fact that there was a maximum allowed tail wind of 2 m/s. These results of his detailed calculations are in all essentials in agreement with the results we arrive at using the simplified arguments in this paper.

A word of caution is appropriate when one discusses the effect of reduced gravity on sports records. In some textbooks one is asked to estimate the world record in high jump, if competitions could be held on the Moon where the gravity is only 1/6 of the value on the Earth. Obviously the record would not be 6 times higher on the Moon than on the Earth, because what matters is how much the centre of gravity is raised in the jump. For a world class male athlete on the Earth, it is raised from about 1.1 m to about 2.4 m when the bar is passed. About 8 or 9 meters would then be a reasonable answer to the Moon problem. However the altered gravity certainly also affects the athlete's approach and take-off. That will be less forceful on the Moon.

We now return to throwing events. The inertial forces are much more important than gravity forces during the athlete's throw. Consider, as an illustration, shot putting. For our purpose a crude model is sufficient. We assume that during the final stage of the put, the launch velocity v is reached in a uniformly accelerated motion over a distance h. That acceleration requires a force $F_i = mv^2/(2h)$, or $F_i = 400$ N if m = 7 kg, h = 2 m and v = 15 m/s. During the throw the athlete also raises the shot and so has to overcome the gravity force. The difference in g between Stockholm and Mexico City corresponds to a change in the gravity force on the shot by 0.2 N, or 1/2000 of F_i. The arm and other parts of the athlete's body must also be accelerated and raised in the gravity field, but that will lead to approximately the same ratio between altered gravity force and inertial forces. If the total force applied by the athlete is constant (i.e., independent of the small gravity changes) the launch velocity

would increase by 1/4000 and the length of the throw by 1/2000 in our example. That means 1 cm in a 20 m shot put, to be compared with about 6 cm from the effect of the reduced gravity force during the free flight of the implement.

6. THROWING AT A TARGET – THE PRECISION OF RELEASE

In some sports an implement is released towards a mark or another implement at some distance, and the aim is to come as close as possible to the target. Here we shall first consider sports akin to boule, and then curling. Finally, references are given to work on tennis, golf and basketball.

6.1. *Varpa*, boule and related sports

A vertical stick is placed 20 m from the competitors, who are throwing flat stones (real stones, or cast metal discs) trying to land as near the stick as possible. This particular version of a universally spread game is called *varpa* and has its roots on the island of Gotland, in the Baltic Sea. Boule is a similar game, with metal balls thrown at another ball at a distance of about 15 m. In one version of the *varpa* one measures the distance between the stick and the nearest part of the stone. A champion has an accumulated distance to the stick of about 400 cm after a completed competition of 36 throws, i.e., about 10 cm per throw.

In order to come close to the target, the direction sideways has to be correct. We shall neglect that and focus on the problem of landing the stone at the correct distance from the launch point. Suppose that a pointlike mass is thrown so as to land within the distances d ± b. We neglect air resistance and use the result for the parabolic trajectory, i.e., we require that

$$d - b < \frac{v^2}{g} \sin(2\alpha) < d + b \tag{11}$$

When d = 20 m and b = 10 cm, the inequality is fulfilled for all elevation angles α and launch speeds v that lie within the shaded area in Figure 3. The allowed interval in launch speed, for an angle $\alpha = 45^\circ$, is (14.00±0.04) m/s. It is remarkable how well the human brain is able to achieve such a fantastic precision. But it is not necessary to exemplify with a champion. Even an unexperienced person can come within 1 m of the stick 20 m away, after a short practice. That requires a launch velocity (14.00±0.35) m/s, i.e., a variation by less than 2.5 %. For completeness it should be noted that the distance to the stick is measured from the closest part of the stone. Since the stone has a certain size the requirements on v and α are somewhat less stringent than obtained in our model of a pointlike particle.

6.2. Curling

In curling a stone slides over the ice towards a target about 36 m away. We analyse this in an idealised model where a mass m moves over a horisontal surface with a constant coefficent of friction f. The speed of delivery is v. After a time t the mass stops, at a distance s. Thus

$$v - at = 0 \tag{12}$$

$$s = vt - at^2/2 \tag{13}$$

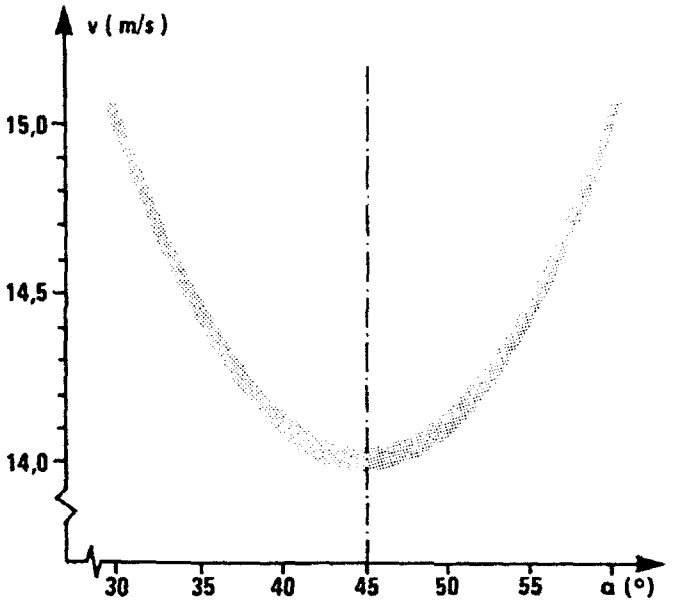


Figure 3. The "window of acceptance" for the launch velocity v and the elevation angle α in the sport *varpa*, where a flat stone is thrown towards a stick at a distance of 20 m and the center of the stone lands within 10 cm of the stick.

The retardation is $a = gf$, where g is the acceleration of gravity. Hence

$$f = 2s/(gt^2) \tag{14}$$

The typical time for the stone to slide to the target is $t = 22$ s. With $s = 36$ m, Eq. (14) gives $f = 0.015$. The coefficient of friction between the curling stone and the ice is very low!

In analogy to the discussion above for *varpa* we now ask what is the allowed interval for the velocity of delivery if the stone is to stop within a certain distance interval. Eliminating t in the equations above gives

$$v = \sqrt{2gfs} \tag{15}$$

With, say, $s = (36 \pm 0.5)$ m and $f = 0.015$ one has $v = (3.25 \pm 0.03)$ m/s. Several factors complicate the situation in the real world of curling. The ice is not flat but has bubbles from being sprinkled with water. The underside of the stone is not flat, and it is in contact with the ice only along its rim. Further, the stone is given a slight rotation, that leads to a curl in the path and to complicated frictional forces. During a competition the coefficient of friction may change, and different rinks can have frictions that vary by 20 % or more. On the other hand the rules allow the sweeping of the ice with brooms, in front of the stone while it slides forward. That may affect the sliding distance by several meters. However, we should note that an increase in s by 3 m, in our model example only requires that v increases from 3.25 m/s to 3.39 m/s, so v still has to be controlled within a few per cent.

6.3. Other sports – tennis, golf and basketball

We end by giving references to some works on the required precision in other sports. With computer techniques one may realistically calculate the trajectory of a tennis ball or a golf ball, even when it is spinning in air. In tennis one can define an "acceptance window" in speed and velocity of a returned ball, so that it clears the net and hits inside the court. An advanced popular book has been written by Brody [8]. He has also written some more technical articles on tennis [22,23]. Similar work, but relating to golf, has been published in a popular and lavishly illustrated book by Cochran and Stobbs [9]. Holmes [24] has modelled golf putting.

Shots in basketball are particularly complex when performed while the athlete is in the air, since the launch velocity is affected by the translational and rotational motion of the body, as discussed by Brancazio [10,25]. Townend [26] has considered the free (ie., penalty) shot in basketball, with particular emphasis on the launch angle, the speed and the height of the launch point.

7. CONCLUDING REMARKS

In the first part of this paper we considered various factors that the athlete has no control over, but which will affect the recorded result. Table 3 summarises some of the findings. The columns refer to (i) the efffect of a maximum slope of 1:1000 of the field or track; (ii) variations in the acceleration of gravity between a place close to the equator (e.g., Mexico City, São Paulo) and a place further to the poles (e.g., Stockholm); (iii) reduced air pressure at high altitudes (e.g., Mexico City) compared to sea level; and (iv) maximum allowed tail wind of 2 m/s. The numbers given are of course only approximate and of varying uncertainty, as discussed above. Each entry refers to the size of the possible improvement due to the factor under consideration. The units in Table 3 are meter (hammer, long jump) and second (100 m sprint). It is seen that the combined effect of the most favourable inclination, gravity, air pressure and tail wind implies a total advantage of about 1 % of the recorded result in hammer throwing and long jump and about 2 % in 100 m sprint.

Table 3
Summary of advantages related to various factors, given in meters (hammer, long jump) and in seconds (100 m sprint)

Sport	Slope	Gravity	Air	Wind
Hammer	0.1	0.3	0.5	0.2
Long jump	0.01	0.02	0.03	0.03
100 m sprint	<0.02?	<0.01	0.09	0.13

In the last part of this paper we considered the precision by which an athlete has to release an implement in order to succesfully reach a certain target. Often the speed given to the implement must be controlled to within 1 %, which is an impressive indication of how well the human brain can coordinate movements.

8. REFERENCES

1 C. Frohlich, Am. J. Phys., 54 (1985) 590.
2 C Frohlich (ed.), Physics of sports; Selected reprints, American Association of Physics Teachers, College Park, MD, 1986.
3 C.B. Daish, The physics of ball games, Hodder and Stoughton, London, 1972.
4 B.D. Woods and P.E. Travers, Dyson's mechanics of athletics, Hodder and Stoughton, London, 1986.
5 R.G. Watts and A.T. Bahill, Keep your eye on the ball: The science and folklore of baseball, Freeman, New York, 1990.
6 R.K. Adair, The physics of baseball, Harper and Row, New York, 1990.
7 E.W. Schrier and W.F. Allman (eds.), Newton at the bat: The science in sports, Scribner's, New York, 1987.
8 H. Brody, Tennis science for tennis players, Univ. Pennsylvania Press, Philadelphia, PA, 1987.
9 A. Cochran and J. Stobbs, The search for the perfect swing, Lippincott, Philadelphia, 1976.
10 P.J. Brancazio, Physical laws and optimum performance, Simon and Shuster, New York, 1984.
11 N. de Mestre, The mathematics of projectiles in sport, Cambridge Univ. Press, Cambridge, 1990.
12 R.D. Mehta, Ann. Rev. Fluid Mech., 17 (1985) 151.
13 T. Bahill and W. Karnavas, New Scientist, 130 (1991) No. 1763, p. 26.
14 D.T. Kagan, Am. J. Phys., 58 (1990) 151.
15 F.R. Whitt and D.G. Wilson, Bicycling Science, MIT, Cambridge MA, (1982).
16 R. McNeill Alexander, New Scientist, 114 (1987) No. 1538, p. 42.
17 A. Turnbull, New Scientist, 135 (1992) No. 1831, p. 23.
18 J.G. Hay, The biomechanics of sports techniques, Prentice–Hall, Englewood Cliffs, NJ, 1985.
19 International Amateur Athletic Federation Handbook 1992–1993, London, 1992.
20 E. McFarland, Am. J. Phys., 54 (1986) 513.
21 C. Frohlich, Am. J. Phys., 49 (1981) 1125.
22 H. Brody, Am. J. Phys., 47 (1979) 482.
23 H. Brody, Am. J. Phys., 49 (1981) 816.
24 B.W. Holmes, Am. J. Phys., 59 (1991) 129.
25 P.J. Brancazio, Am. J. Phys. 49 (1981) 356, Erratum: 49 (1981) 356.
26 S. Townend, in Mathematical modelling: A source book of case studies, edited by I.D. Huntley and D.J.G. James, Oxford Univ. Press, Oxford, p. 97, 1990.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Aerodynamic sound associated with vortex motions: observation and computation

T. Kambe

Department of Physics, University of Tokyo
Hongo, Bunkyo-ku, Tokyo 113, Japan

Abstract

Theory of vortex sound in the aeroacoustics can describe characteristic features of the waves radiated by localized vortex motions at low Mach numbers. Some evidences and illustrative examples are presented from our studies to show remarkable agreement between observed and computed wave profiles emitted by the vortex motions with or without presence of solid body. Among the cases to be discussed here are (i) *head-on* collision or *oblique* collision of two vortex rings, (ii) *vortex-cylinder* interaction, (iii) *vortex-edge* interaction. Corresponding to each of the three cases, the radiation patterns are characterized by the quadrupole, dipole and cardioid type, respectively. Monopolar (isotropic) component detected in the experiment is likely due to the viscous dissipation of kinetic energy. In the oblique collision of two vortex rings, reconnection of vortex-lines occurs, and causes a violent motion of the vorticity, resulting in excitation of a characteristic type of acoustic wave.

1. INTRODUCTION

In the theory of vortex sound at low Mach number and high Reynolds number [1-8], the source flow is characterized by a localized vortex motion scaled on a length l which is regarded as the vortex size too. This flow field is surrounded by *outer* wave field scaled on the length $\lambda \approx c\tau = O(l/M)$, where $\tau = l/u$, u being a typical velocity of the flow field. The sound speed and density in the undisturbed uniform medium are denoted by c and ρ_0, respectively. Typical Mach number M is assumed to be much less than unity: $M = u/c \ll 1$. This enables separate analysis of the two fields: *inner* flow and *outer* wave regions, because of the *compactness* of the source flow, $l \ll \lambda$. The theory of vortex sound is one of the most successful parts in the theory of aerodynamic sound [9-12].

In our investigation of the vortex sound in the past decade, evidences showing the validity of the theory are accumulating. Our experimental and theoretical studies include three typical cases. The first one is the sound generated by two interacting vortices in free space in the absence of external body: *i.e.* (*i*) head-on collision and (*ii*) oblique collision of two vortex rings. In the first stage of the collision, the vortex interaction is almost inviscid. This is followed by the second viscous stage : cancellation of opposite signs of vorticity at the time of collision, reconnection of vortex-lines, or turbulent disintegration. Acoustic waves generated by these motions were detected experimentally and investigated in detail. The second case is the sound wave generated by a vortex

interacting with a solid body of length scale l. Presence of an external body in the vicinity of an eddy causes emission of waves of the dipole type [13], whereas the waves in the former case are characterized as the quadrupole. The third case considered below is the wave generation by a vortex ring moving near an edge plate whose size is larger than the scale λ. Remarkable feature of the wave field are the effects of the edge and non-compactness of the plate, giving rise to the scattering of waves by the edge plate. Here the temporal wave profile depends on the time history of the vortex motion [29].

In the present paper, asymptotic representation of the velocity potential of the inner flow is sought first, and then the solution of the outer wave equation is determined so as to be matched to the inner solution. This method is especially successful to the first case of the vortex sound in free space, and also applicable to the second case of compact body. The third case must be treated differently because of the effect of scattering. Main interest here is comparison of the wave profiles thus determined computationally with the corresponding profiles detemined experimentally. The vortex motions considered are three dimensional. Although two dimensional problems of sound generation by rectilinear vortices were studied in the early times of development ([2], see [10] for reviews), experimental observations and validation are rare.

2. MATHEMATICAL FORMULATION

Basic equation of the aerodynamic sound is given by the Lighthill's equation [9]:

$$\rho_{tt} - c^2\nabla^2\rho = \frac{\partial^2}{\partial x_i \partial x_j} T_{ij} \tag{1}$$

where $$T_{ij} = \rho v_i v_j + (p - c^2\rho)\delta_{ij} - \tau_{ij} \ ,$$

p is the pressure, ρ the density, v_i the i-th component of velocity and τ_{ij} the viscous stress tensor. Summation convention is assumed on the right hand side (rhs) of (1) and henceforth.

In the inner region scaled by the length l and time τ, the ratio of the two terms on the left hand side is estimated as $|\rho_{tt}|/|c^2\nabla^2\rho| = O(l^2/c^2\tau^2) = O(M^2)$. In the first approximation neglecting terms of $O(M^2)$ and hence the term ρ_{tt}, the equation (1) reduces to the Navier-Stokes equation for an incompressible fluid, governing solenoidal vortex motion treated as a source flow.

The solenoidal velocity field $\boldsymbol{v}(\boldsymbol{x},t)$ can be represented in terms of a vector potential $\boldsymbol{A}(\boldsymbol{x},t)$: $\boldsymbol{v} = \nabla \times \boldsymbol{A}$. Given the vorticity $\boldsymbol{\omega}(\boldsymbol{x},t)$, the vector potential is expressed as

$$\boldsymbol{A}(\boldsymbol{x},t) = \frac{1}{4\pi}\int \frac{\boldsymbol{\omega}(\boldsymbol{y},t)}{|\,\boldsymbol{x}-\boldsymbol{y}\,|}\mathrm{d}^3\boldsymbol{y} \tag{2}$$

(*e.g.* see [14]). At a large distance $r = |\boldsymbol{x}|$ away from the localized vorticity region having either a compact support of $\boldsymbol{\omega}(\boldsymbol{y},t)$ or such a property that $|\boldsymbol{\omega}|$ decays expoentially as $r \to \infty$, the factor $1/|\boldsymbol{x}-\boldsymbol{y}|$ can be expanded in a Taylor series which is convergent for $|\boldsymbol{y}|/|\boldsymbol{x}| < 1$, and the velocity field $\boldsymbol{v}(\boldsymbol{x},t)$ tends to be represented asymptotically by a velocity potential Φ. Asymptotic expansion of the potential $\Phi(\boldsymbol{x},t)$ which is obtained after some nontrivial calculation for $r/l \to \infty$ is given in the form [15],

$$\Phi(\boldsymbol{x},t) = \Phi_0(\boldsymbol{x},t) + Q_i\,\partial_i\frac{1}{r} + Q_{ij}\,\partial_i\partial_j\frac{1}{r} + Q_{ijk}\,\partial_i\partial_j\partial_k\frac{1}{r} + O(r^{-5}) \ , \tag{3}$$

where $\partial_i = \partial/\partial x_i$, and

$$Q_i(t) = \frac{1}{8\pi}\int (\boldsymbol{y}\times\boldsymbol{\omega})_i \,\mathrm{d}^3\mathbf{y} \quad , \tag{4}$$

$$Q_{ij}(t) = -\frac{1}{12\pi}\int (\boldsymbol{y}\times\boldsymbol{\omega})_i\, y_j \,\mathrm{d}^3\mathbf{y} \quad , \tag{5}$$

$$Q_{ijk}(t) = \frac{1}{32\pi}\int (\boldsymbol{y}\times\boldsymbol{\omega})_i\, y_j\, y_k \,\mathrm{d}^3\mathbf{y} \quad , \tag{6}$$

$\cdots \qquad \cdots\cdots$

The vector $\boldsymbol{x} = (x_i)$ is used for the point of observation and $\boldsymbol{y} = (y_i)$ for the point of integration (the souce point). It is almost trivial to show the following properties,

$$Q_{ii}(t) = 0 \ , \quad Q_{iik}(t) = 0 \ , \quad Q_{iji}(t) = 0 \tag{7}$$

by the orthogonality of the two vectors $\boldsymbol{y}$ and $\boldsymbol{y}\times\boldsymbol{\omega}$. The resultant impulse of the vortex system is defined by $4\pi Q_i$, which is an invariant of motion for unbounded fluid wihtout external body.

The first term $\Phi_0(\boldsymbol{x},t)$ represents the effects which cannot be represented by the expression (2) (*e.g.* fluid compressibility or presence of a solid body in the inner region). The compressibility of the source flow ($|$ div $\boldsymbol{v}|$ being $\mathrm{O}(M^2)$) is assumed to be localized, apart from the generated acoustic wave which is of larger scale $\lambda \gg l$ and treated in the next stage. From the theory of irrotational motion with localized rate of expansion $\Delta = \mathrm{div}\,\boldsymbol{v}$, the first term Φ_0 is given by the following asymptotic expansion [14],

$$\Phi_0 = R_0(t)\frac{1}{r} + R_i(t)\,\partial_i\frac{1}{r} + R_{ij}(t)\,\partial_i\partial_j\frac{1}{r} + R_{ijk}(t)\,\partial_i\partial_j\partial_k\frac{1}{r} + O(r^{-5}) \tag{8}$$

where $R_0, R_i, R_{ij}, \cdots$ are all functions of time t and determined by the distribution of $\Delta(\boldsymbol{x},t)$. Presence of a solid body in the vicinity of the vortex motion can be represented by an additional velocity potential whose asymptotic expansion is again of the form (8) [14]. In the case of solid body, the first monopole term must vanish by the condition of no net outflow over the body surface. Thus both effects of compressibility and presence of solid body are represented by the expansion (8).

Corresponding pressure field at large distances from the vortex motion is given by the linear approximation, $p^{(i)}(\boldsymbol{x},t) = -\rho_0\Phi_t$, namely

$$p^{(i)}(\boldsymbol{x},t) = -\rho_0\dot{P}_0\frac{1}{r} - \rho_0\dot{P}_i\,\partial_i\frac{1}{r} - \rho_0\dot{P}_{ij}\,\partial_i\partial_j\frac{1}{r} - \rho_0\dot{P}_{ijk}\,\partial_i\partial_j\partial_k\frac{1}{r} + O(r^{-5}) \tag{9}$$

where the over-dot denotes a time differentiation, and $P_0(t) = R_0$, $P_i(t) = Q_i + R_i$, $P_{ij}(t) = Q_{ij} + R_{ij}$, $P_{ijk}(t) = Q_{ijk} + R_{ijk}$, $\cdots$.

In the outer region, the rhs of (1) becomes vanishingly small and the pressure is governed by the wave equation : $p_{tt} - c^2\nabla^2 p = 0$, because of the (assumed) adiabatic relation $\mathrm{d}p = c^2\mathrm{d}\rho$. The *wave* pressure $p^{(\mathrm{w})}$ matching to the inner solution (9) asymptotically as $|\,\boldsymbol{x}\,|/\lambda \to 0$ is given by the multi-pole expansion,

$$p^{(\mathrm{w})}(\boldsymbol{x},t) = -\rho_0\frac{\dot{P}_0(t_r)}{r} - \rho_0\,\partial_i\left[\frac{\dot{P}_i(t_r)}{r}\right] - \rho_0\,\partial_i\partial_j\left[\frac{\dot{P}_{ij}(t_r)}{r}\right] - \rho_0\,\partial_i\partial_j\partial_k\left[\frac{\dot{P}_{ijk}(t_r)}{r}\right] + \cdots \tag{10}$$

where $t_r = t - r/c$ is the retarded time. This is verified as follows.

Obviously, each term of (10) satisfies the wave equation. Matching of the two solutions $p^{(\mathrm{i})}$ and $p^{(\mathrm{w})}$ is carried out in an intermediate region. In terms of the inner variable defined by the normalization $\bar{x}_i \equiv x_i/l$ (and $\bar{r} = r/l$), the outer variable $\hat{x}_i \equiv x_i/\lambda$ and intermediate variable ξ_i are given by

$$\hat{x}_i = M\bar{x}_i = M^{1-\alpha}\xi_i\,, \qquad \xi_i = M^{\alpha}\bar{x}_i = (M^{\alpha}/l)\,x_i \tag{11}$$

where α is a parameter in the range $0 < \alpha < 1$. Normalized retarded time is $\bar{t}_r \equiv t_r/\tau = \bar{t} - \hat{r} = \bar{t} - M^{1-\alpha}\xi$, where $\bar{t} = t/\tau$, $\hat{r} = r/\lambda$, $\xi = M^{\alpha-1}\hat{r} = M^{\alpha}\bar{r}$. Using ξ, we have $\bar{r} = M^{-\alpha}\xi$ and $\hat{r} = M^{1-\alpha}\xi$. Thus in the limit $M \to 0$ with keeping ξ and ξ_i fixed (the intermediate limit), we obtain $\bar{r} \to \infty$ and $\hat{r} \to 0$. As an example we consider matching of the second order term. The third term of the *inner* pressure (9) is written as

$$-\rho_0 \dot{P}^{(\mathrm{i})}_{ij}\,\partial_i\partial_j \frac{1}{r} = -\rho_0 \frac{M^{3\alpha}}{l^3}\dot{P}^{(\mathrm{i})}_{ij}(\bar{t})\,\frac{\partial^2}{\partial\xi_i\partial\xi_j}\frac{1}{\xi}$$

whereas the third term of the outer *wave* pressure (10) is

$$-\rho_0\frac{M^{3\alpha}}{l^3}\frac{\partial^2}{\partial\xi_i\partial\xi_j}\left[\frac{\dot{P}^{(\mathrm{w})}_{ij}(\bar{t}-M^{1-\alpha}\xi)}{\xi}\right] = -\rho_0\frac{M^{3\alpha}}{l^3}\frac{\partial^2}{\partial\xi_i\partial\xi_j}\left[\frac{\dot{P}^{(\mathrm{w})}_{ij}(\bar{t})}{\xi} + \frac{1}{2}M^{2(1-\alpha)}\xi\,\ddot{P}^{(\mathrm{w})}_{ij}(\bar{t}) + \cdots\right], \tag{12}$$

where the superscripts (i) and (w) are used to distinguish the functions in the two regions. It is found that both expressions have the same order of magnitude $M^{3\alpha}$ and we have $\dot{P}^{(\mathrm{w})}_{ij} = \dot{P}^{(\mathrm{i})}_{ij}$ in the intermediate limit as $M \to 0$. Crow [4] shows that the second term of the order $\mathrm{O}(M^{\alpha+2})$ in the parenthesis [ ] on the rhs of (12) can be matched to the term in the next approximation (effect of compressibility, *e.g.* the term ρ_{tt}) of the inner expansion. Precisely speaking, an arbitrary term $C(\bar{t})\delta_{ij}$ can be added to $\dot{P}^{(\mathrm{w})}_{ij}(\bar{t})$ in the leading order matching. If so, in the next order $\mathrm{O}(M^{3\alpha}M^{2(1-\alpha)}) = O(M^{\alpha+2})$, the second term in [ ] of rhs of (12) will include the additional term of the form $\ddot{C}\nabla^2(\xi/2) = \ddot{C}/\xi$. This arbitrariness drops out when we introduce a monopole term of the form $\dot{P}^{(\mathrm{w})}_0 = -(M^2/l^2)\ddot{C}(\bar{t})$ by the consistency argument. Then the combined outer solution $-\ddot{C}(\bar{t}-\hat{r})/\hat{r} + \nabla^2[C(\bar{t}-\hat{r})/\hat{r}]$ vanishes identically [4]. Matchings of the other terms are verified similarly. In particular, in the matching of the fourth terms, another arbitrariness in the dipole term comes in, but can be dropped out by the same reasoning.

The origin of the first monopole term of (10) is considered by Kambe [16] (see also [17, 18]). It is shown from the dynamical equation of motion (Navier-Stokes equation) that

$$P_0(t) = -\frac{5-3\gamma}{12\pi}\frac{1}{c^2}\dot{K}(t)\,, \quad K(t) = \frac{1}{2}\int v^2(\mathbf{y},t)\,\mathrm{d}^3\mathbf{y} \tag{13}$$

where K is the total kinetic energy and γ the ratio of specific heats ($\gamma = 7/5$ for the air). We assume that significant effect of the compressibility of the vortex motion appears only in this isotropic term due to the assumption of the compact source flow and $M \ll 1$, mentioned in the beginning. The rest factors P_i, P_{ij}, P_{ijk}, $\cdots$ of the multipole components of $p^{(\mathrm{w})}$ are given by Q_i, Q_{ij}, Q_{ijk}, $\cdots$ associated with the vorticity ω and R_i, R_{ij}, R_{ijk}, $\cdots$ representing the influence of a solid body. The functions Q_i, Q_{ij}, $\cdots$ are integrals of the moments of the vorticity $\omega(\boldsymbol{y},t)$ as given in (4)–(6). Thus the multipole components of the generated wave $p^{(\mathrm{w})}$ are expressed in terms of the vorticity of the source flow. This is called as ***vortex sound***. The the second term of (10) represents dipole emission due to change of the total impulse $4\pi Q_i$ and/or presence of a solid body.

3. COLLISION OF TWO VORTEX RINGS

We first consider the vortex sound in an unbounded fluid with no solid body, generated by collisions of two vortex rings. In this situation, the impulse $4\pi Q_i$ is conserved, and hence the dipole term disappears. In fact, the dipole emission (considered in the next section) is related to the rate of change of the resultant external force $F_i(t)$ exerted on the fluid [13] which is absent by the assumption. Therefore P_{ij} and P_{ijk} are given only by Q_{ij} and Q_{ijk} respectively. The pressure in the acoustic far-field ($\hat{r} \to \infty$) is much simplified in its form because the space derivatives applied to r^{-1} in (10) become higher order of smallness than those applied to the functions of t_r as ($\hat{r} \to \infty$). Thus the pressure observed at a point $\boldsymbol{x} = (x_1, x_2, x_3)$ in the acoustic far-field is given as

$$p^{(f)}(\boldsymbol{x}, t) = -\rho_0 \dot{P}_0^{(1)}(t_r)\frac{1}{r} - \frac{\rho_0}{c^2} Q_{ij}^{(3)}(t_r)\frac{x_i x_j}{r^3} + \frac{\rho_0}{c^3} Q_{ijk}^{(4)}(t_r)\frac{x_i x_j x_k}{r^4} + \cdots \qquad (14)$$

where superscript (n) denotes the n-th time derivative. The second quadrupole term (Möhring's quadrupole [6]) can be shown to derive from the non-isotropic part of the Reynolds stress $\rho_0 v_i v_j$ [16]. The conservation of the resultant moment of impulse (angular impulse) leads to the symmetry property $\dot{Q}_{ij} = \dot{Q}_{ji}$ (and $Q_{ii}^{(3)} = 0, etc.$ by (7)). The first isotropic (monopole) term arises when the total kinetic energy K of the system changes, but vanishes identically in an inviscid fluid. Here we have written the formula up to the third order terms. An experimental observation (described in § **3.2.**) is showing existence of this order.

Using the length scale l, the vorticity scale $\omega = u/l$ and the time scale $\tau = l/u$, the scaling law of the wave pressure of the quadrupole sound is deduced as follows. The tensor Q_{ij} is normalized by $l^4 u$ and hence $Q_{ij}^{(3)}$ by $l^4 u/\tau^3 = l u^4$ (the first monopole term gives the same by (13)). Thus we find the scaling law for the quadrupole sound as

$$p_Q \sim \frac{\rho_0}{c^2} l u^4 \frac{1}{r} = \frac{\rho_0 u^4}{c^2}\frac{l}{r} .$$

The sound intensity I_Q is given by $p_Q^2/\rho_0 c$. Hence we obtain the well-known intensity law [9]: $I_Q \sim (\rho_0 u^8/c^5)(l/r)^2 \propto u^8$.

3.1. Head-on collision

Axisymmetric collision of two vortex rings [17] is a particularly simple example of the vortex sound. In an inviscid fluid, the first term of the formula (14) vanishes and the second quadrupole term reduces to

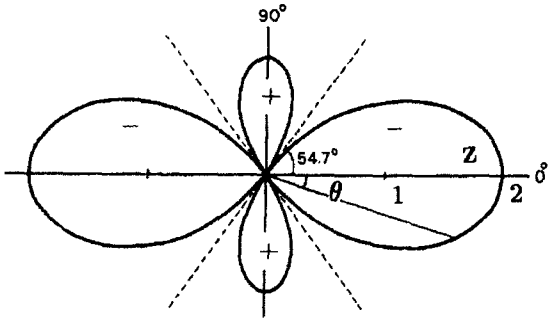


Figure 1. Directivity of the acoustic pressure: $1 - 3\cos^2\theta$.

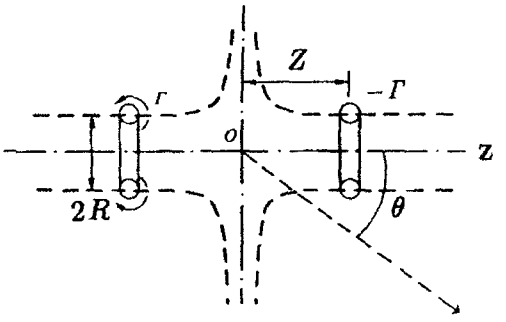


Figure 2. Head-on collision of two vortex rings (definition sketch).

$$p = -\frac{\rho_0}{12c^2} Q^{(3)}(t_r)\ \frac{1}{r}(1 - 3\cos^2\theta) \tag{15}$$

where the vorticity is assumed to have only azimuthal ϕ-component $\omega(\sigma, z)$ in the cylindrical coordinate system (z, σ, ϕ), leading to

$$Q(t) = \int \omega(\sigma, z)\ \sigma^2 z\ \mathrm{d}\sigma\,\mathrm{d}z\ , \tag{16}$$

with $z = x_3$ and $\sigma = \sqrt{x_1^2 + x_2^2}$ and $\theta = \arccos(z/r)$. In this case the temporal behavior of the pressure is described by the single scalar function of 3rd-order derivative $Q^{(3)}(t)$, and spatial distribution is given by the axisymmetric four-lobe directivity (Figure 1).

(a) Computation

Suppose that we have two vortex rings having a common symmetry axis z with one vortex being a mirror image of the other with respect to the plane $z = 0$ (Figure 2), and that they approach each other according to the equation of motion. When the vortex is characterized by its strength $-\Gamma$, ring radius $R(t)$ (core radius δ) and the distance $Z(t)$ from the symmetry plane $z = 0$, the interacting motion of two vortex rings of very thin cores is described by a system of first-order ordinary differential equations [19] originally derived by Dyson [20]. Then the profile function $Q(t)$ is expressed by $-2\Gamma R^2(t)Z(t)$. The wave pressure form is given by $Q^{(3)}(t)$, which is calculated numerically by solving this system of differential equations and shown by the curve I in Figure 3, together with the observed one E (see *(b)*). When the two vortices come close to each other, forward speed $|\dot{Z}(t)|$ decreases, but the ring radius $R(t)$ grows rapidly, and the cores of the two vortices get closer to come into contact [19]. The first peak of the profile corresponds to the initial inviscid stage of the colliding motion.

Effect of finite core-size and core deformation on the wave profile is studied in detail by the contour dynamics [21]. This analysis suggests that the dip observed in the curve $-Q^{(3)}(t)$ is due to core deformation, although the core size used in the numerical simulation seems to be larger than in the experiment and the fluid is inviscid.

A numerical simulation of axisymmetric viscous vortex collision was carried out to estimate the profile function $Q^{(3)}(t)$, shown by the curve V in Figure 3, at the Reynolds number $Re \approx 1.3 \times 10^3$, based on the initial translation velocity $U = |\dot{Z}(0)|$ and ring diameter $2R_0 = 2R(0)$ [31].

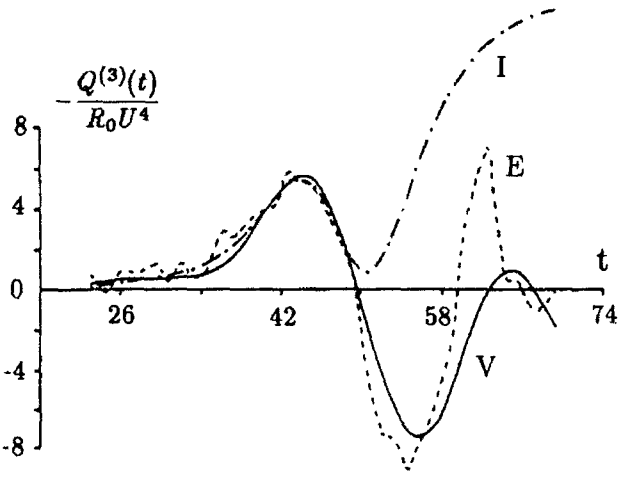


Figure 3. Temporal profiles $-Q^{(3)}(t)$ (normalized): I, inviscid; E, experimental; V, viscous simulation.

(b) Observation

Experimental observations of the corresponding acoustic waves due to the vortex collision are reported in [17] and [22]. In the latter study, the waves were observed in all directions at $\theta = 10°$ to $350°$ with $10°$ interval in a meridional plane including

the symmetry axis x_3.

The observed pressures at 35 stations are represented in the truncated Fourier series: $p(\theta,t) = a_0(t) + a_1(t)\cos\theta + b_1(t)\sin\theta + a_2(t)\cos 2\theta + b_2(t)\sin 2\theta$. The two terms $a_0(t)$ and $a_2(t)\cos 2\theta$ are found to be dominant and rewritten as

$$p_{\rm mq}(\theta,t) = p_{\rm m}(t) + p_{\rm q}(t)(1 - 3\cos^2\theta) \ ,$$

where $p_{\rm m}$ and $p_{\rm q}$ represent the monopolar and quadrupolar components respectively. Significant amplitudes of $p_{\rm m}(t)$ and $p_{\rm q}(t)$ are detected (Figure 4). This experimental study suggests that there exists monopole component associated with dissipation of the kinetic energy (13). The Reynolds number Re is of the order 10^4 or larger in this experiment. A general feature of the observed curve $p_{\rm m}$ implies, helped by (13) and (10), that the kinetic energy K decreases like a step function (or rather like the function $-\tanh$) at the time around $t \approx 2000(\mu{\rm s})$.

An experimental visualization in color was made recently in water at lower Reynolds number of the order 10^3 [23]. This clearly shows, when the two vortex rings are close to one another, reconnection of vortex lines and formation of smaller rings by instability. At higher Reynolds number, turbulent disintegration is observed.

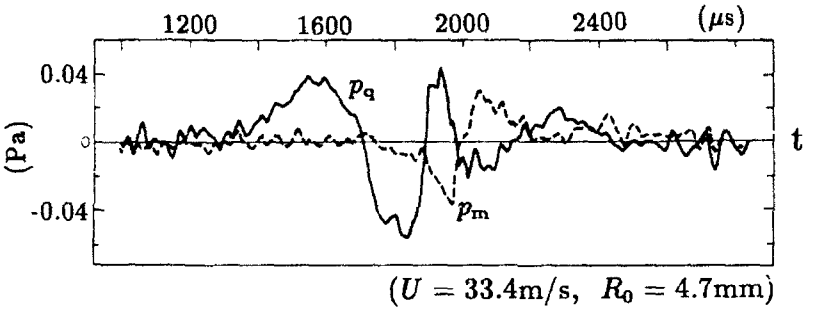


Figure 4. Observed temporal profiles of $p_{\rm m}(t)$ and $p_{\rm q}(t)$.

3.2. Oblique collision

Oblique collision of two vortex rings at right angles is studied experimentally and computationally [15]. Evidently this oblique collision (Figure 5) has no axisymmetry like that in the previous case of head-on collision. This requires an increased amount of data for the analysis. At the oblique collision, opposite senses of vortex lines are forced to come in contact at the inner part. This event is followed by violent motion of vorticity and excitation of acoustic waves. Analysis of the wave data provides some information of the complex vortex motion associated with vortexline reconnection, sometimes turbulent.

Consider a problem that the initial state is given in such a way that two vortex rings are set to move along the paths intersecting at right angles at the origin and collide with one another. The bisecting straight line between the two paths of the vortex center is taken as the polar axis $\theta = 0$ (along the x_3 axis) of the spherical coordinate system. The plane perpendicular to the x_3 axis is the (x_1, x_2) plane on which $\theta = \pi/2$. There are two symmetry planes including the x_3 axis: one includes the trajectories of the vortex centers which is defined as (x_2, x_3) plane and the plane (x_1, x_3) perpendicular to it is also a symmetry plane which bisects the two trajectories. The plane $\phi = 0$ is taken along the positive x_1 axis. Thus the two vortex rings (centers) move toward the origin along the direction of the angles $(\theta, \phi) = (\pi/4, \pi/2)$ and $(\pi/4, 3\pi/2)$ before colliding

0.1ms 0.9ms 1.2ms 1.4ms

Figure 5. Schlieren photographs of the oblique collision of two vortex tings (made by shock impulse from the two nozzles seen above).

interaction. From the geometrical arrangement just mentioned, the acoustic pressure $p(\theta, \phi, t)$ is characterized by the symmetry:

$$p(\theta, -\phi, t) = p(\theta, \phi, t) \ , \qquad p(\theta, \phi + \pi, t) = p(\theta, \phi, t) \ . \tag{17}$$

The formula (14) is rewritten by using the spherical polar coordinates (r, θ, ϕ): $x_1 = r\sin\theta\cos\phi$, $x_2 = r\sin\theta\sin\phi$, $x_3 = r\cos\theta$. Then the n-th order form, like the expression $F_2 = C_{ij}x_ix_j/r^2$ for $n = 2$ or $F_3 = C_{ijk}x_ix_jx_k/r^3$ for $n = 3$, can be represented in terms of the n-th order (or lower order) spherical harmonics,

$$P_n^0(\zeta), \quad P_n^1(\zeta)\,(\cos\phi,\ \sin\phi), \quad \cdots, \quad P_n^n(\zeta)\,(\cos n\phi,\ \sin n\phi), \tag{18}$$

where $\zeta = \cos\theta$, $P_n^0(\zeta)$ and $P_n^k(\zeta)$, $(k = 1, \cdots, n)$ are the Legendre polynomials. In fact, we have $F_2 = (1/3)(-C_{11} - C_{22} + 2C_{33})P_2^0 + (1/6)(C_{11} - C_{22})P_2^2\cos 2\phi$ + (linear combination of $P_2^1\cos\phi, P_2^1\sin\phi$ and $P_2^2\sin 2\phi$). Similarly, regarding the third-order form F_3, we give an explicit expression to the terms to be used later:

$$F_3 = \frac{1}{5}(2Q_{333}^{(4)} - \tilde{Q}_{113} - \tilde{Q}_{223})P_3^0 + \frac{1}{30}(\tilde{Q}_{113} - \tilde{Q}_{223})P_3^2\cos 2\phi \ + \ \cdots \tag{19}$$

where $Q_{ijk}^{(4)}$ are used in place of C_{ijk}, and the tilde symbol denotes $\tilde{Q}_{113} = Q_{113}^{(4)} + Q_{131}^{(4)} + Q_{311}^{(4)}$, *etc.*. The coefficient of P_3^0 of (19) reduces to $Q_{333}^{(4)} - (1/5)Q_{3kk}^{(4)}$ by the relations (7).

In view of the symmetry (17), the pressure is represented as

$$\begin{aligned} p(\theta, \phi, t) \ = \ & A_0(t) + A_1(t)P_2^0(\cos\theta) + A_2(t)P_2^2(\cos\theta)\cos 2\phi \\ & + B_1(t)P_3^0(\cos\theta) + B_2(t)P_3^2(\cos\theta)\cos 2\phi \ , \end{aligned} \tag{20}$$

where higher order terms are omitted since observed amplitudes are not significant. Thus it is found that the far-field acoustic pressure (14) is represented in terms of the five normal modes with five coefficient functions of time $[A_0(t), A_1(t), A_2(t), B_1(t), B_2(t)]$. Here the Legendre functions are $P_2^0 = (1/2)(3\cos^2\theta - 1)$, $P_2^2 = 3\sin^2\theta$, $\cdots$, and $P_3^0 = (1/2)(5\cos^3\theta - 3\cos\theta)$, $P_3^2 = 15(\cos\theta - \cos^3\theta)$, $\cdots$.

(a) Observation

The acousttic waves emitted by the 90° collision were detected at 102 different angular positions on the three great circles of radius $r = 620mm$ on the three orthogonal pianes: (1) $\phi = \pi/2, 3\pi/2$; (2) $\phi = 0, \pi$; (3) $\theta = \pi/2$. The trajectories of the vortex cores in the (x_2, x_3) plane observed by a photosensor are shown in Figure 6, where the ring radius $R_0 = R(0)$ of the single (unperturbed) vortex is 4.7mm. ($U = 27$m/s)

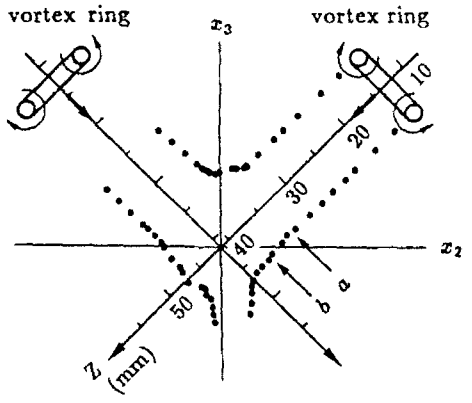


Figure 6. Observed core trajectories of two colliding vortices in the (x_2, x_3) plane including the vortex center. The marked positions a and b correspond to the times of Figure 7 (broken lines).

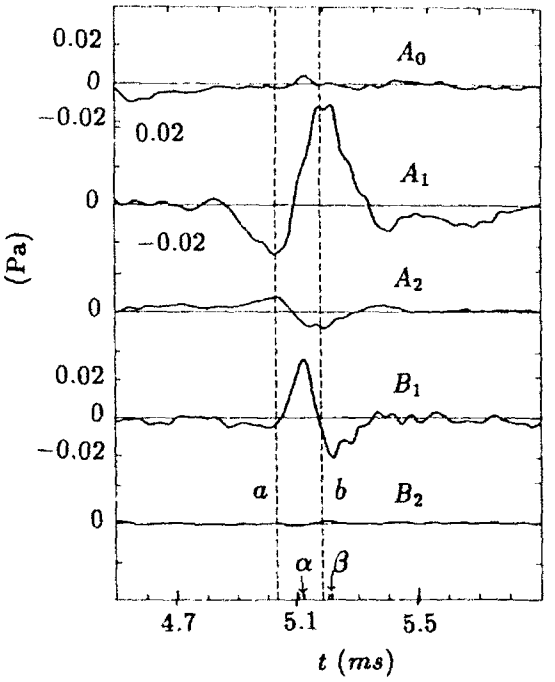


Figure 7. Main mode amplitudes of observed waves.

The symmetry relation (17) is found to be consistent with the observed data. In fact, the acoustic pressure detected in the plane $\theta = \pi/2$ can be expanded into Fourier series with respect to the angle ϕ, and it is found that the Fourier coefficients of $\sin m\phi$ $(m = 1 \sim 4)$ and $\cos m'\phi$ $(m' = 1, 3, 4, 5)$ are negligible.

Three sets of profile functions of $[A_0, A_1, A_2, B_1, B_2]$ can be determined from the observed data. It is found that they coincide almost with each other. These profiles are shown in Figure 7.

(b) Computation

Numerical simulation of the vortex collision at right angles, which corresponds to the experiment, was carried out by M. Takaoka in order to estimate the acoustic emission, using the method of [24]. The incompressible Navier-Stokes equation is solved numerically, together with the continuity equation, by the spectral method on 64^3 grid points. Imposed boundary conditions are 2π periodicity in the three coordinate directions.

The isotropic component is proportional to the second time derivative of the kinetic energy, and the quadrupole components and higher modes are related to the change of moments of vorticity distribution. Using the data from the simulation, we can calculate the tensors $Q_{ij}(t)$ and $Q_{ijk}(t)$ of (5) and (6). Thus we readily obtain the main mode

coefficients $[\tilde{A}_0, \tilde{A}_1, \tilde{A}_2, \tilde{B}_1, \tilde{B}_2]$. The main mode amplitudes of the computation and the observation are qualitatively similar, but quantitatively different, probably because of the difference of the values of the Reynolds number. Unfortunately, the Reynolds number of the computer simulation is smaller by two orders of magnitude than that of the experiment.

(c) *Implication of dominant wave components*

Observed amplitudes A_1 and A_2 of the two quadrupoles P_2^0 and $P_2^2 \cos 2\phi$ in (20) are significantly large. The amplitudes $A_1(t)$ and $A_2(t)$ are considered as consisting of three longitudinal quadrupoles Q_{11}, Q_{22}, Q_{33} (to the directions x_1, x_2, x_3), as understood from the form of the coefficients of F_2. It is probable that there exists non-negligible amplitude in the monopole.

Out of the two third-order components, the amplitude B_2 of $P_3^2 \cos 2\phi$ is negligibly small, but the appearance of the mode P_3^0 is substantial. To see the significance of this mode P_3^0, we examine its coefficient B_1 which is given by $Q_{333}^{(4)}(t) - \frac{1}{5} Q_{3kk}^{(4)}(t)$, where

$$Q_{333} = \frac{1}{32\pi} \int (\boldsymbol{y} \times \boldsymbol{\omega})_3 \, y_3^2 \, \mathrm{d}^3 \mathbf{y}, \quad Q_{3kk} = \frac{1}{32\pi} \int (\boldsymbol{y} \times \boldsymbol{\omega})_3 \, y^2 \, \mathrm{d}^3 \mathbf{y} \ ,$$

$$p_3 \equiv (\boldsymbol{y} \times \boldsymbol{\omega})_3 = y_1 \omega_2 - y_2 \omega_1 = \boldsymbol{y}_* \times \boldsymbol{\omega}_* \ ,$$

the vectors $\boldsymbol{y}_* = (y_1, y_2)$ and $\boldsymbol{\omega}_* = (\omega_1, \omega_2)$ being projection of $\boldsymbol{y}$ and $\boldsymbol{\omega}$ to the plane (x_1, x_2). Restricting our attention to the location of vortexline reconnection at the inner part of the collision and reminding the geometry of the vortexlines, it is expected that the variable p_3 (near $\boldsymbol{y}_* = 0$) changes its sign before and after the reconnection, and that its rapid change will give substantial contribution to the terms $Q_{333}^{(4)}(t)$ and $Q_{3kk}^{(4)}(t)$. Thus it is suggested that the significant amplitude $B_1(t)$ observed in the experiment (see Figure 7) is representing an acoustic signal of a local violent motion at the inner part of the collision. The observed acoustic pressure shows a characteristic, asymmetric directivity at the times when the B_1 mode is substantial (see Figure 8). As a test observation, we took a shadowgraph (Figure 9) of the waves generated by the core collision with the vortex speed much higher than that of the acoustic measurement. It is supposed that the wave generation corresponds to the phase α of Figure 8.

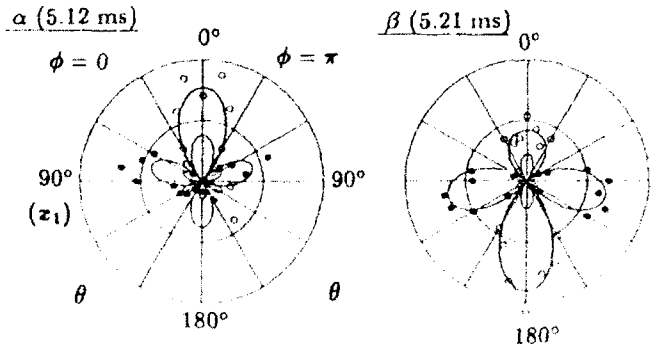


Figure 8. The asymmetric directivity of the acoustic pressure at the times α and β marked in Figure 7 : positive (○), negative(●).

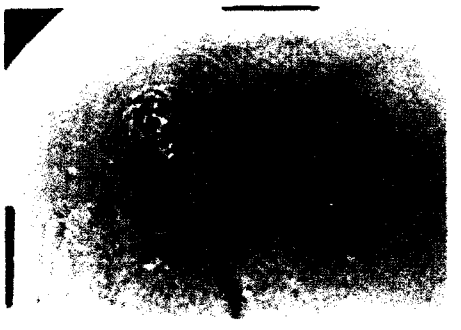


Figure 9. Waves generated by the core collision at a higher vortex speed (taken by T. Minota).

4. INTERACTION OF A VORTEX RING WITH A BODY

Presence of a solid body in a flow causes a considerable increase in the intensity of generated sound. The degree of increase depends on the boundary conditions. If the body is finite with its dimension being small compared with a typical wavelength, then the sound wave generated by nearby eddies is characterized by a dipole emission whose intensity is proportional to the sixth power of the typical velocity u. On the other hand, the intensity of sound generated by eddies in the vicinity of a sharp edge of a flat semi-infinite plate (considered in the next section) is proportional to the fifth power of u and the wave field takes a cardioid directivity pattern. [see [7] for general formulation; [25] for a circular cylinder; [26] for a sphere ; [27] for a wedge.]

Solenoidal vortex motion in the neiborhood of a solid body at rest is dominated by a dipole field. This is interpreted as follows. The first and second terms of (8) disappear by the conditions of vanishing net outflux across the body surface and no translational motion of the body [14], respectively. However, the impulse $4\pi Q_i$ (the coefficient of the second term of (3)) changes according to the momentum exchange between the body and the surrounding fluid. The rate of change of $4\pi Q_i(t)$ is given by the resultant force $F_i(t)$ exerted on the fluid by the relation, $4\pi \dot{Q}_i = F_i$ (see [19]).

It is shown [7] that the temporal wave profile is determined by the following integral which is a function of time t:

$$\Pi_i(t) = \int \omega(y,t) \cdot \Psi_i(y)\, \mathrm{d}^3 y \tag{21}$$

where ω is the vorticity of the source flow and the function $\Psi_i(y)$ $(i = 1,2,3)$ is a vector potential (with div $\Psi_i = 0$) for the velocity of an imaginary potential flow (around the body) of a *unit velocity at infinity in the i-th direction*. The function $\Psi_i(y)$ represents the presence of the body in the vicinity of the vortex motion.

4.1 General formula

If the eddy and the body without sharp edge are both compact relative to the acoustic wavelength, the acoustic pressure obtained by Curle [13] is expressed as

$$p(\boldsymbol{x},t) = \frac{1}{4\pi c}\frac{x_i}{r^2}\,\dot{F}_i(t-\frac{r}{c}) \tag{22}$$

in the far-field where $-F_i(t)$ is the force exerted on the body by the surrounding fluid, represented by an integral of the stress over the body surface. The expression (22) is clearly a dipole wave with its axis in the direction of $\dot{F}_i(t)$

A different expression of the acoustic pressure is obtained by Obermeier [28] and Kambe [7] for the same problem. The pressure generated by the interaction of a solid body and a closed-loop vortex (σ) as

$$p(\boldsymbol{x},t) = \frac{\rho_0}{4\pi c}\frac{x_i}{r^2}\,\ddot{\Pi}_i(t-\frac{r}{c}) \tag{23}$$

where the function $\Pi_i(t)$ of (21) reduces to

$$\Pi_i(t) = \Gamma \oint_\sigma \Psi_i \cdot \mathrm{d}\mathbf{s} = \Gamma \int_S (\nabla \times \Psi_i) \cdot \mathbf{n}\, \mathrm{d}S = \Gamma J_i(t)\ . \tag{24}$$

The constant Γ is the strength of the vortex, the infinitesimal variables ds and dS are a line element of the closed loop σ and an elemnt of the open surface S bounded by

σ, respectively, and n is a unit normal to the surface S. The function $J_i(t)$ denotes the volume flux of the potential flow Ψ_i passing through the loop σ.

To obtain the scaling law of the dipole emission, we note that Π_i is normalized by ul^3 since $\Gamma \sim ul$ and $|\nabla \times \Psi_i|$ being of order unity away from the body. Thus the scaling law is

$$p_D \sim \frac{\rho_0}{c}\frac{ul^3}{\tau^2}\frac{1}{r} = \frac{\rho_0 u^3}{c}\frac{l}{r}.$$

The sound intensity is $I_D \sim (\rho_0 u^6/c^3)(l/r)^2 \propto u^6$.

4.2 A vortex passing nearby a circular cylinder

(*a*) *Computation*

Consider a circular vortex ring of radius R (core radius δ) passing by the side of an infinite straight circular cylinder of radius a (Figure 10). The x_3 axis is taken along the cylinder axis. (In this case, Ψ_3 denotes a uniform flow to the cylinder axis.) If the vortex is sufficiently distant from the cylinder, the vortex path can be regarded as rectilinear. It is assumed that the vortex ring keeps its circular form (approximately) and the ring center moves within the plane (x_1, x_2), which means that the normal to the ring plane lies within the (x_1, x_2) plane. The above expression (23) reduces to

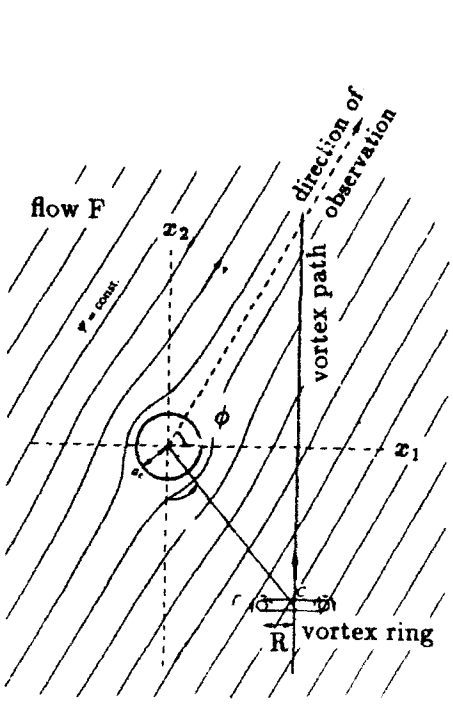


Figure 10. Schematic diagram of the problem and the flow F to the observation direction.

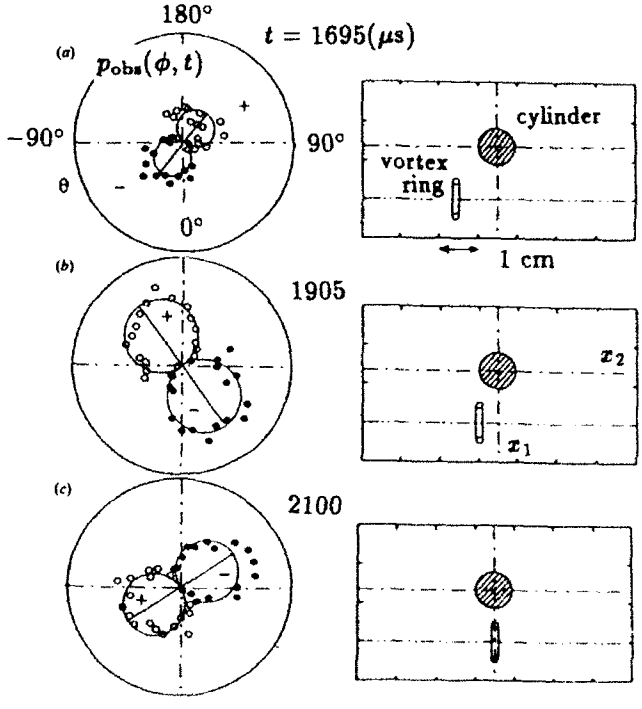


Figure 11. Dipole character of the far-field acoustic pressure $p(\phi, t)$ and corresponding vortex position relative to the circular cylinder at three times, where U = 27m/s, R = 4.7mm and a = 4.5mm .

$$p(\boldsymbol{x}, t + \frac{r}{c}) = \frac{\rho_0}{4\pi c} \Gamma \frac{\sin\theta}{r} [\ddot{J}_1(t) \cos\phi + \ddot{J}_2(t) \sin\phi] \quad (25)$$

$$= \frac{\rho_0}{4\pi c} \Gamma \frac{\sin\theta}{r} \ddot{J}(t) \cos(\phi - \Theta(t)) \quad (26)$$

where $\ddot{J}^2 = \ddot{J}_1^2 + \ddot{J}_2^2$ and $\tan\Theta = \ddot{J}_2/\ddot{J}_1$, since $\ddot{J}_3$ is taken to be zero and Ψ_i $(i = 1, 2)$ has only the x_3 component: $(0, 0, \Psi_i)$. The angle ϕ denotes the azimuthal angle of the projection of the position vector $\boldsymbol{x}$ on the (x_1, x_2) plane, measured from the x_1 axis. The angle θ is the polar angle of the direction $\boldsymbol{x}$ from the x_3 axis. Evidently the directivity of the acoustic emission (26) is of dipole character with its axis in the direction $\Theta(t)$ determined by $\ddot{J}_1$ and $\ddot{J}_2$. The coefficient $\ddot{J}(t)$ is the second time derivative of the volume flux $J(t)$ passing through the vortex ring of the flow F around the cylinder to the direction of observation ϕ ([25], Figure 10). Note that the wave pressure (26) vanishes toward the cylinder axis $\theta = 0$ and π. Thus the non-compactness of the body to the direction of the cylinder axis does not influence the characteristic feature of the wave field.

(*b*) *Observation*

Corresponding acoustic waves were detected at various angles ϕ in the plane $\theta = 90°$ and various angles θ in the plane $\phi = 0°$ and 180°. Average wave observed at every 10° position (in the plane $\theta = 90°$) is expanded into Fourier series with respect to ϕ. It is found that the main component is given in the form,

$$p_{\text{obs}}(\phi, t) = -a(t) \sin\phi + b(t) \cos\phi \ ,$$

in accordance with the expression (25), and the dipole character is clearly seen in Figure 11. The solid curves a, b in figure 12 are the observed profiles. The broken curves show computed (normalized) amplitudes $a = -\ddot{J}_2$ and $b = \ddot{J}_1$ for the vortex ring moving along a straight line parallel to the x_2 axis but with observed variable speed (the velocity decays gradually). Agreement in absolute values between the observed and predicted profiles are fairly good.

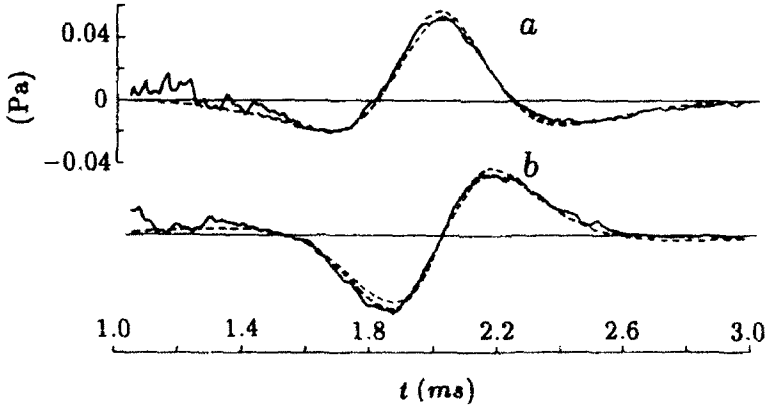


Figure 12. Comparison of the coefficients $a(t)$ and $b(t)$: observation (solid) and computation with δ/R =0.15 (thick broken) and 0.1 (thin broken).

5. INTERACTION OF A VORTEX RING WITH AN EDGE OF A FLAT PLATE

Presence of a flat plate with a sharp edge in the vicinity of a vortex alters the wave field significantly, as well as the inner field bounded by the plate. When the edge plate is non-compact, the temporal wave profile depends on the history of the vortex motion, which is the scattering effect of the edge. This effect is taken into account by using the low-frequency Green's function [5], and the formulation is made differently from the previous ones ([29], [30], [10]).

In the present case of a vortex ring moving near the edge of the plate (Figure 13, the edge being taken as the x_3-axis, the polar axis), the plate is placed at $y_2 = 0$ ($y_1 < 0$) or $\phi = \pm 180°$. The wave pressure generated by this interaction is obtained [7, 29] as

$$p(\boldsymbol{x},t) = \frac{\rho_0}{(2\pi^3 c)^{1/2}} \frac{\sin\frac{1}{2}\phi\,(\sin\theta)^{1/2}}{r} f(t-\frac{r}{c}) \tag{27}$$

where $$f(t) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{t} \frac{\ddot{\Pi}(s)}{(t-s)^{1/2}}\,\mathrm{d}s\,, \quad \Pi(t) = \Gamma \int_S (\nabla \times \Psi)\cdot \mathbf{n}\,\mathrm{d}S, \tag{28}$$

the latter expression is obtained from (21) like (24) with $\Psi_i = \Psi = (0,0,\psi(y))$. The streamfunction $\psi(y)$ is defined by $-Y^{1/2}\cos\frac{1}{2}\phi$, representiong a hypothetical potential flow around the edge from below to above the plate, where $Y = (y_1^2 + y_2^2)^{1/2}$. Note that the velocity $\nabla \times \Psi$ of the hypothetical flow is scaled as $L^{-1/2}$ (L being the nearest distance to the edge), and that the time scale for the flux change should be given by $\tau_L = L/u$. Using the argument which is similar to the case of dipole, we obtain the scaling law of the wave pressure for the *edge* sound as

$$p_E \sim \frac{\rho_0}{c^{1/2}} \frac{ulL^{-1/2}l^2}{\tau_L^{3/2}} \frac{1}{r} = \frac{\rho_0 u^{5/2}}{c^{1/2}} \left(\frac{l}{L}\right)^2 \frac{l}{r}\,.$$

The acoustic intensity I_E defined by $p_E^2/\rho_0 c$ is now proportional to $u^5 L^{-4}$. The angular dependence of the wave pressure (27) is given by the function $F(\theta,\phi) = \sin\frac{1}{2}\phi\,(\sin\theta)^{\frac{1}{2}}$

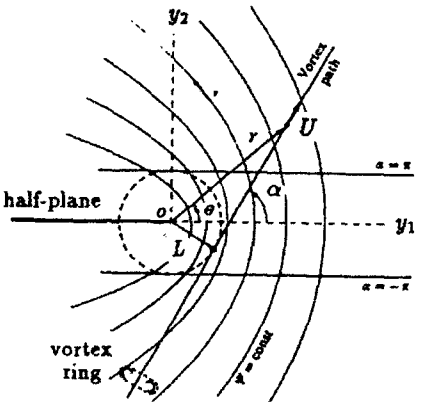


Figure 13. Trajectory of the vortex ring moving near the edge and the streamlines given by ψ =const.

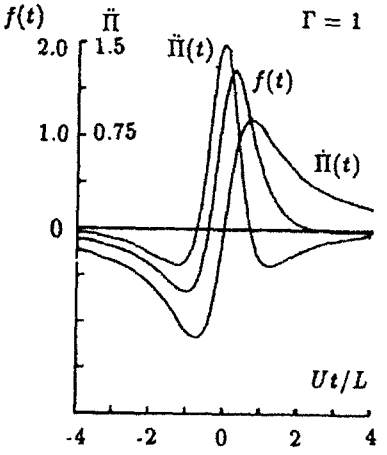


Figure 14. Three functions, $f(t)$, $\dot{\Pi}(t)$, $\ddot{\Pi}(t)$ for the angle of trajectory $\alpha = -\frac{1}{2}\pi$.

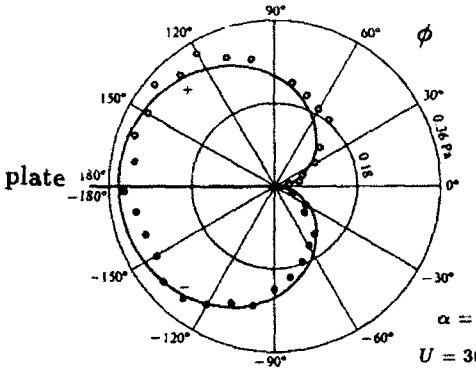


Figure 15. Polar plot of the acoustic pressure observed at a fixed time t_0 with $\theta = \pi/2$. The solid curve is $f_{\exp}(t_0)\sin\frac{1}{2}\phi$.

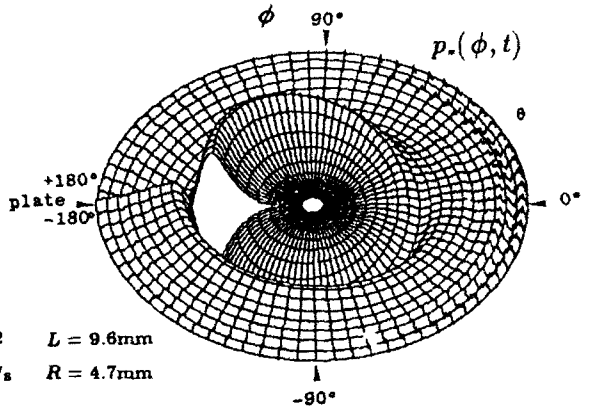


Figure 16. Perspective profile of the observed wave: $p_*(\phi,t) = f_{exp}(t)\sin\frac{1}{2}\phi$.

(Figure 15). In the plane perpendicular to the x_3 axis ($\theta = \pi/2$), the directivity of the acoustic intensity is the *cardioid*, $I_E \sim \sin^2\frac{1}{2}\phi$ on a fixed circle r =const and θ =const.

(*a*) *Computation*

The temporal wave profile of the pressure is represented by $f(t)$ of (28) :

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty}(-i\omega)^{\frac{1}{2}}\dot{\Pi}_F(\omega)\exp[-i\omega t]\,d\omega = \left(\frac{d}{dt}\right)^{\frac{1}{2}}\dot{\Pi}(t) \tag{29}$$

[26], where $\Pi(t)/\Gamma$ is the volume flux of the flow ψ passing through the vortex ring of radius R (core radius δ), and $\dot{\Pi}_F(\omega)$ is the Fourier transform of $\dot{\Pi}(t)$. The vortex trajectory is a straight line with the angle α to the y_1 axis. The three functions $f(t)$, $\dot{\Pi}(t)$, $\ddot{\Pi}(t)$ are shown in Figure 14 for $\alpha = -\pi/2$. It is observed that the curves $f(t)$ are intermediate between $\dot{\Pi}(t)$ and $\ddot{\Pi}(t)$. This intermediate behavior exhibited by the half-integer differential operation $(d/dt)^{\frac{1}{2}}$ is related to the scattering of the wave by the non-compact edge plate. It is interesting to compare this with $\ddot{\Pi}(t)$ of (23) for the compact case.

(*b*) *Observation*

The sound emission from a vortex ring travelling near the edge of a flat plate along a nearly straight path was investigated in a laboratory experiment [29]. Both of the spatial profile $F(\theta,\phi)$ and temporal profile $f(t)$ were found to agree with the predicted ones described above.

The pressure $p(\phi,t)$ observed in the plane $\theta = 90°$ is expressed by a truncated Fourier series. It is found that the main component is of the form $p_*(\phi,t) = f_{\exp}(t)\sin\frac{1}{2}\phi$ in accordance with (27). Experimental study was done mainly for the vortex path of $\alpha = -\pi/2$. Figure 15 is a polar plot of the pressure $p_*(\phi,t)$ at a fixed time t_0: $p_*(\phi,t_0)$. A perspective plot of the pressure profile $p_*(\phi,t)$ is illustrated in figure 16. The half-plane lies at $\theta = \pm 180°$, on both sides the pressure takes opposite signs. The vortex moved from the positive to negative y_2-axis with the translation velocity $U = 30.4$m/s, $R = 4.7$mm and $L = 9.6$mm.

Figure 17 shows comparison between the observed curve and computed curves on

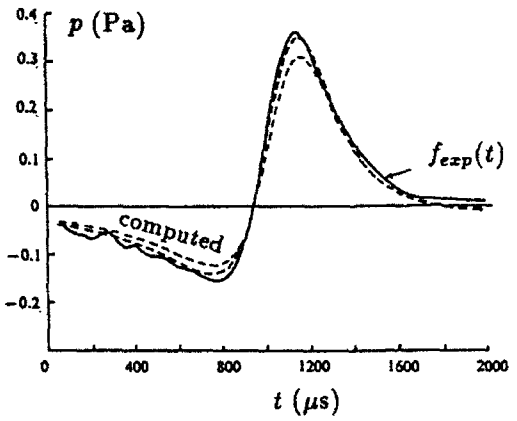


Figure 17. Comparison of $f_{\rm exp}(t)$ and corresponding computed curves.

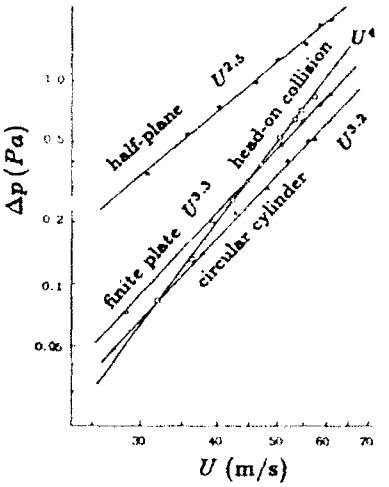


Figure 18. Observed power laws of Δp (Pa) vs. U (m/s).

the absolute scales. The solid curve is the experimentally determined profile $f_{\rm exp}(t)$ and the broken curves are the computed ones of $p/F(\phi, \theta = 90°)$ obtained by using (27, 28) for the relative core size $\delta/R = 0.2$ (thin broken) and 0.3 (thick broken). Agreement of the two curves is remarkable.

6. CONCLUSION

Concerning the problem of the vortex sound, a new formulation is developed for a compact source flow of localized vorticity distribution, based on asymptotic expansion of the velocity potential of the inner flow. This is successfully applied not only to the sound generation by two interacting vortices in free space, but also to the case of wave generation by a vortex interacting with a compact external body. The third case of the wave generated by a vortex moving near a non-compact edge plate must be treated differently, taking account of the scattering effect. In each of the three cases, the computed wave profiles show remarkable agreement with the observed ones, implying validity of the theory of vortex sound. Figure 18 is a summary of our experiment, illustrating observed power laws (in the log-log plot) of the acoustic pressure (peak) amplitude Δp versus the translation velocity U of a single vortex (in isolated state).

In the section 3.1 of head-on collision of two vortex rings, a consistent account is given by relating the phases of the vortex collision with those of the wave profiles. A new feature is noticed concerning asymmetric wave emission in the oblique collision in the section 3.2.

Finally, some remarks are to be made about other successful comparisons between observation and computation, which are out of scope of the present article. Acoustic noise of dipole-type is investigated recently, based on the computer simulation of a uniform flow around an aerofoil, and compared with corresponding experiment at $M_\infty = 0.23$ and $Re = 3 \times 10^5$ (with reference to the chord and the freestream velocity) [32]. Fair agreement is found between the two noise spectra obtained from the computer simulation and the experiment. The computed data shows structure of the surface dipole over the wing surface. Another study [33] of a shock wave interacting with a vortex filament also revealed reasonable agreement between the observation and the calculation.

Acknowledgement

The author wishes to thank Dr. Toyoko Minota for collaborative work on the present subject and Dr. Masanori Takaoka on the computational aspect of the oblique vortex collision.

REFERENCES

1 A. Powell, *J. acoust. Soc. Am.*, **36** (1964) 177.
2 F. Obermeier, *Acustica*, **18** (1967) 238.
3 V.R. Lauvstad, *J. Sound Vib.*, **7** (1968) 90.
4 S.C. Crow, *Stud. appl. Math.*, **49** (1970) 21.
5 M.S. Howe, *J. Fluid Mech.* **71** (1975) 625.
6 W. Möhring, *J. Fluid Mech.* **85** (1978) 685.
7 T. Kambe, *J. Fluid Mech.* **173** (1986) 643.
8 E.-A. Müller and F. Obermeier, *Fluid Dyn. Res.* **3** (1988) 43.
9 M.J. Lighthill, *Proc. R. Soc. Lond.*, A**211** (1952) 564.
10 D.G. Crighton, *Prog. Aerospace Sci.* **16** (1975) 31.
11 M.E. Goldstein, *Aeroacoustics*, McGraw-Hill, 1976.
12 A.P. Dowling and J.E. Ffowcs Williams, *Sound and Sources of Sound*, Ellis Horwood Ltd., 1983.
13 N. Curle, *Proc. R. Soc. Lond.* A**231** (1955) 505.
14 G.K. Batchelor, *An Introduction to Fluid Dynamics*, Chap. 2 & 6, Cambridge University Press, 1967.
15 T. Kambe, T. Minota, T. Murakami and M. Takaoka, in *Topological Fluid Mechanics* (eds.: H.K. Moffatt & A. Tsinober), 515-521 (1990).
16 T. Kambe, *J. Sound Vib.* **95** (1984) 351.
17 T. Kambe and T. Minota, *Proc. R. Soc. Lond.*, A**386** (1983) 277.
18 F. Obermeier, *J. Sound Vib.* **99** (1985) 111.
19 T. Kambe and T. Minota, *J. Sound Vib.* **74** (1981) 61.
20 F.W. Dyson, *Phil. Trans. R. Soc. Lond.* A**184** (1893) 1041.
21 K. Shariff, A. Leonard and J.H. Ferziger, NASA Tech. Mem. 102257 (1989).
22 T. Minota and T. Kambe, *J. Sound Vib.* **111** (1986) 51.
23 T.T. Lim and T.B. Nickels, *Nature* **357** (1992, May 21).
24 S. Kida, M. Takaoka and F. Hussain, *Phys. Fluids* A**1** (1989) 630; and *J. Fluid Mech.* **230** (1991) 583.
25 T. Minota and T. Kambe, *J. Sound Vib.* **119** (1987) 509.
26 T. Minota, T. Kambe and T. Murakami, *Fluid Dyn. Res.* **3** (1988) 357.
27 T. Minota, T. Murakami and T. Kambe, *Fluid Dyn. Res.* **4** (1988) 57.
28 F. Obermeier, *J. Sound Vib.* **72** (1980) 39.
29 T. Kambe, T. Minota and Y. Ikushima, *J. Fluid Mech.* **155** (1985) 77.
30 J.E. Ffowcs Williams and L.H. Hall, *J. Fluid Mech.* **40** (1970) 657.
31 T. Kambe and U Mya Oo, *J. Phys. Soc. Jpn.* **53** (1984) 2263.
32 S. Adachi, K. Ishii, S. Akishita and K. Kuwahara, DGLR/AIAA-92-02-157 (DGLR/AIAA-14th Aeroacoustics Conference, Aachen, May 1992).
33 T.M. Weeks and D.S. Dosanjh, *AIAA Journal*, **5**(4) (1967) 660.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
1993 IUTAM.

Nonlinear Membrane Theory

A. Libai

Faculty of Aerospace Engineering, Technion, Haifa 32000, Israel

Abstract

Nonlinear membrane theory can be regarded as an approximation to the nonlinear behavior of shells of finite thickness under suitable conditions, or else studied as a mathematical model, complete in itself, for shells which are formally assumed to be incapable of carrying moments. Selected topics of interest in the statics of nonlinear elastic membranes will be discussed. These include: the basic equations, the behavior of a membrane adjacent to a stressed initial state, the undeformed membrane, boundary conditions, membrane edge effect, lines of rotational discontinuity, axisymmetric membranes, cylindrical membranes, wrinkle fields and others. The intrinsic form of the shell equations will be adopted in most of the discussions. Cases with small-strain geometric nonlinearities as well as some finite strain problems will be considered. Emphasis will be put more on general and formulative issues rather than on specific solutions, but the analysis of a specific case will be outlined.

1. INTRODUCTION

Membrane theory is commonly regarded as an approximation to the shell problem, obtained by formally suppressing the stress couples and transverse shears in the shell equations, compared with the tangential force resultants ("membrane forces"). The suppression takes place in the force equilibrium equations, force boundary conditions, energy, and moment equilibrium around the normal to the shell. There is ample experimental and theoretical evidence to the effect that membrane theory is a valid "interior" approximation to the shell equations, which dominates the behavior far from boundaries and discontinuities (either in the geometry or the loading), for deformation processes with a typical length measure 'r' of the reference surface, and provided that there exists adequate support at the shell boundaries. Processes with length measure $\sqrt{rt}$, where t is the shell thickness (such as buckling, short-wavelength vibrations, edge effects), and also those involving substantial inextensional deformations, are to be treated by the more general bending theory. Also, bending corrections to the membrane solution which are based on its deformation patterns can be calculated. These can be used to estimate the accuracy of the membrane approximation and to correct the stress levels (which are of a lower order in t/r). Chapters on membrane theory can be found in many standard books on shell theory [1-13]. Some consider nonlinear behavior specifically [7-14].

Another approach to membrane theory regards membranes as two-dimensional deformable sheets, plane or curved, which are *formally incapable* of carrying any stress couples or transverse shears, and can sustain tangential force resultants only. These are the "ideal" or "true" membranes (the term "ideal" membranes will be used to distinguish them from "shell membranes" which *can* carry stress couples, as previously discussed). The theory of ideal membranes can be formally modelled as a complete entity in itself.

Several important characteristics distinguish the behavior of ideal membranes from shell membranes. To name a few:

(1) Ideal membranes cannot sustain compressive stresses, whereas shell membranes can carry compressive stresses until buckling takes place.

(2) Ideal membranes exhibit "wrinkle fields" or "tension fields" to compensate for the nonexistence of compressive stresses. In shell membranes, postbuckling behavior is out of the scope of the model, but the approximation of an "incomplete tension field" can be sometimes used to investigate the transition from postbuckling to wrinkle-fields [15].

(3) Ideal membranes can accept lines of angular discontinuity in the deformation pattern on the boundary or in the interior. In shell-membranes, angular continuity is preserved.

(4) In the same vein: existing lines of angular discontinuity in an ideal membrane can be smoothed out under suitable conditions. In shell-membranes, the theory breaks down near such lines.

(5) Large inextensional deformations can be sustained in an ideal membrane, but not in a shell membrane.

(6) A large variety of kinematic and force boundary conditions can be imposed on an ideal membrane within the scope of the model (with obvious exceptions - such as imposed stress couples or transverse shears). Exact edge conditions can rarely be specified for a shell-membrane. Bending or 3D theories are used for edge effect analysis, with component 3D, bending and nonlinear membrane effects.

(7) The "undeformed shape" of an ideal membrane needs only a metric specified in a 2D (curved) space $u^{\alpha}(\alpha=1,2)$ plus a specified closed line in this space. It can be "packed in a box". A shell membrane starts from a 3D thin walled body which constitutes its undeformed shape. Not all the geometrical properties of this shape appear in the equations, but they are necessary for the analysis of its limitations.

The question as to whether ideal membranes do actually exist as physical entities need not be answered: the analyses of mechanics problems involve some degree of idealization and modelling of the behavior of the physical world, and the success of a model should be measured by its ability to predict this behavior in some approximate sense. With this in mind, several classes of shell problems can be modelled as ideal membranes. Among those, the most important is the asymptotic behavior of thin shells as (t/r) approaches zero. In fact, ideal membrane behavior can be arrived at solely by this limiting process. Other classes of problems which can be modelled as ideal membranes include shells made from special materials (biological membranes, cloth and other thin films, inflatables), problems involving large loads of the pressure variety, many large strain problems, etc.

In what follows, no direct distinction between the two types of membranes will be made. The typing, if needed, should be based on the problem at hand. For example, compressive stresses will be included in the analysis, but problems related to wrinkle fields will also be discussed. Boundary conditions for an ideal membrane will be looked at, with the understanding that in a shell membrane this may be just a component of a more complex edge effect problem.

A "real" shell is, of course, a thin 3D continuum. Shell theory is an approximation to 3D behavior. Membrane Theory is an approximation to shell theory and, putting aside special cases, the behavior of ideal membranes is a subset of Membrane Theory. However, if the parameters and objectives are right, the latter can give very good predictions of the behavior of this 3D body under conditions involving strong nonlinearities which, otherwise, would have been extremely difficult to treat.

2. NONLINEARITIES IN MEMBRANE THEORY

These follow the usual classification into geometrical and material nonlinearities. *Small strain theory* with nonlinearities due to the inextensional component of the deformation and to local strain gradients is the more common case. Tension field theory can be included within this class since the wrinkling process involves mostly geometrical changes. A special subclass of the above is "rigid-inextensional" theory which assumes that the strains are zero and that deformation can take place through inextensional and/or wrinkle fields only.

The relative simplicity of the application of two-dimensional hyperelastic constitutive relations to ideal membranes has led to the development of the materially nonlinear large strain membrane theory which, in addition to many practical uses, can be applied to problems involving the loss of stability in tension due to combined material and geometrical effects.

The main feature that nonlinearity adds to membrane theory is shape-adaptability: both geometry and membrane forces participate simultaneously in taking on the applied loads. This is facilitated by considering equilibrium with respect to the deformed configuration. The process gives rise to a sixth order partial differential system which can accommodate, in most case, three boundary conditions. This should be contrasted with linear membrane theory which is taken with respect to an undeformed geometry and can accommodate, in principle, only two boundary conditions. In some cases of well supported membranes (such that the boundary conditions prevent inextensional deformations), *linear* membrane theory yields good predictions away from the boundary. In such cases, the corresponding small strain nonlinear theory predicts the formation of boundary layers of thickness $r\sqrt{\varepsilon}$ at the edges, where ε is a typical strain measure of the membrane.

A well studied group of problems which belong to this class is the axisymmetric deformation of membranes of revolution. However, as the strain increase, the boundary layers are obliterated and boundary effects extend over a large part of the membrane. See section 5 for more details.

At the other extreme lie cases where almost all of the deformation is inextensional and the contribution of the strains is minor. For a trivial but illuminating example, consider a rectangular ideal membrane sheet in the xy plane, clamped along its edge x=0 and loaded along its edge x=a by a uniform line load P_z in the z direction. In its deformed shape, the sheet rotates through 90° and a line of angular discontinuity forms along x=0. The addition of normal pressure causes the deformed shape to be curved but the essential feature of almost inextensional deformation is preserved.

The "spherical barrel" example of section 7 serves to demonstrate some of these cases. This is a truncated ideal spherical membrane attached to rigid rings which are pulled apart by forces P. The membrane collapses in to form a cylindrical wrinkle field. For small P, "rigid-inextensional" theory holds with the wrinkles extending clear to the rings. For larger P, a narrow unwrinkled axisymmetric boundary layer is formed adjacent to the rings. For larger P, the boundary layer no longer exists. The unwrinkled stretched membrane which covers a large part of the surface is analyzed by axisymmetric large strain theory.

3. A REVIEW OF THE MEMBRANE EQUATIONS

(a) Notations

Tensorial component notation is adopted in the general discussion. Physical components are used in examples. Vector notation is used as needed, referred to a fixed Cartesian coordinate system x,y,z. The membrane is to be analyzed in three possible states: *Current state* (denoted with an upper bar), *undeformed state* (without bars), and

reference state (denoted with a subscript or superscript (o)). The latter is a known deformed state with respect to which incremental solutions are sought.

In the sequel, a comma, a double vertical bar, and a single vertical bar, denote partial and covariant differentiations with respect to the current and reference states, respectively. Upper case letters denote full quantities and lower case letters denote incremental quantities (from the reference state). A membrane is identified either by the position vector $\bar{\mathbf{R}}(u^\alpha)$ of material points on its deforming reference surface, or by its current metric and curvature tensors $\bar{A}_{\alpha\beta}$ and $\bar{B}_{\alpha\beta}$, respectively, which satisfy the three continuity equations of Codazzi-Mainardi and Gauss [16,17]. Associated quantities are $\bar{A}=\det(\bar{A}_{\alpha\beta})$, permutation tensor $\bar{\varepsilon}_{\alpha\beta}$, unit normal $\bar{\mathbf{N}}$ and Christoffel symbols $\bar{\Gamma}^\alpha_{\beta\gamma}$ given by $\bar{\Gamma}^\alpha_{\beta\gamma}=\frac{1}{2}\bar{A}^{\alpha\lambda}(\bar{A}_{\lambda\gamma,\beta}+\bar{A}_{\lambda\beta,\gamma}-\bar{A}_{\beta\gamma,\lambda})$. The corresponding quantities in the undeformed and reference states are $\mathbf{R}$, $A_{\alpha\beta}$ etc. and $\mathbf{R}^{(o)}$, $A^{(o)}_{\alpha\beta}$ etc., respectively.

The boundary of a membrane is a curve Γ with arclength parameter $\bar{s}$:Γ: $u^\alpha = u^\alpha(\bar{s})$, where $d\bar{s}^2 = \bar{A}_{\alpha\beta}du^\alpha du^\beta$.

Unit vectors along the curve are $\bar{\lambda}$, $\bar{\nu}$, $\bar{\mathbf{N}}$ where $\bar{\lambda}^\alpha=u^\alpha_{,\bar{s}}$ and $\bar{\nu}_\alpha=\bar{\varepsilon}_{\beta\alpha}\bar{\lambda}^\beta$.

The normal ($\bar{K}_n$) and geodesic ($\bar{K}_g$) curvatures along the curve are:

$$\bar{K}_n=\frac{1}{\bar{r}_n}=\bar{B}_{\alpha\beta}\bar{\lambda}^\alpha\bar{\lambda}^\beta \qquad \bar{K}_g=\bar{\lambda}^\gamma\bar{\nu}_\beta\bar{\lambda}^\beta{}_{\|\gamma}=\bar{\lambda}^\gamma\bar{\nu}_\beta(\bar{\lambda}^\beta{}_{,\gamma}+\bar{\Gamma}^\beta_{\theta\gamma}\bar{\lambda}^\theta)$$

The symmetric force resultant tensor $\bar{N}^{\alpha\beta}$ acts on elements of the *current* membrane, in *current* directions and is measured in terms of the *current* metric. Kirchhoff ("engineering") resultants act in the same directions but use the undeformed metric. They are given by $\sqrt{A}\,N^{\alpha\beta}=\sqrt{\bar{A}}\,\bar{N}^{\alpha\beta}$.

The loading vector per unit deformed/undeformed area is $\bar{\mathbf{p}}=\bar{p}^\alpha\bar{\mathbf{R}}_{,\alpha}+\bar{p}\bar{\mathbf{N}}/p=p^\alpha\bar{\mathbf{R}}_{,\alpha}+p\bar{\mathbf{N}}$. Components with respect to undeformed directions (f^α,f) or to cartesian directions ($f^i\mathbf{e}_i$) are occasionally used.

Two-point tensors and vectors associated with the geometry of the deformation are the Green Strain Tensor $E_{\alpha\beta}$, change of curvature tensor $K_{\alpha\beta}$ and displacement vector $\mathbf{V}$. These are given by $2E_{\alpha\beta}=\bar{A}_{\alpha\beta}-A_{\alpha\beta}$; $K_{\alpha\beta}=\bar{B}_{\alpha\beta}-B_{\alpha\beta}$; $\mathbf{V}=\bar{\mathbf{R}}-\mathbf{R}=v^\alpha\mathbf{R}_{,\alpha}+w\mathbf{N}=v^i\mathbf{e}_i$.

(b) A Review of the Basic Equations

The important issues are the choice of variables and choice of directions. The *intrinsic form* utilizes $\bar{A}_{\alpha\beta}$ and $\bar{B}_{\alpha\beta}$ as geometric field variables and relates the equations to the current directions. The *displacement form* utilizes the displacements (v^α,w) as field variables and relates the equations to the undeformed directions. Other variants are possible too. The equations are occasionally related to extrinsic directions such as the axial and radial directions in membranes of revolution or Cartesian directions. The intrinsic form is preferred in the sequel.

(b.1) Equilibrium

The fundamental form is the vector equation of equilibrium $(\sqrt{A}\,\bar{T}^{\alpha})_{,\alpha}+\sqrt{A}\,\bar{p}=0$ where $\bar{T}^{\alpha}=\bar{N}^{\alpha\beta}\bar{R}_{,\beta}$ is the stress resultant vector.

Alternate component forms with respect to the current directions are: $\bar{N}^{\alpha\beta}{}_{\|\alpha}+\bar{p}^{\alpha}=0$; $\bar{N}^{\alpha\beta}\bar{B}_{\alpha\beta}+\bar{p}=0$ or $\frac{1}{\sqrt{A}}\left(\sqrt{A}\,N^{\alpha\beta}\right)_{,\alpha}+\bar{\Gamma}^{\beta}_{\alpha\gamma}N^{\alpha\gamma}+p^{\beta}=0$; $N^{\alpha\beta}\bar{B}_{\alpha\beta}+p=0$.

Component forms with respect to the undeformed directions utilize the "displacement gradients" and "rotations" defined by $S_{\alpha\beta}=v_{\alpha}|_{\beta}-B_{\alpha\beta}w$; $\varphi_{\alpha}=-(w_{,\alpha}+B_{\alpha\beta}v^{\beta})$. The equations are [18]:

$$\left[\left(A_{\gamma\beta}+S_{\gamma\beta}\right)N^{\alpha\beta}\right]\Big|_{\alpha}+B_{\gamma\alpha}\varphi_{\beta}N^{\alpha\beta}+f_{\gamma}=0$$

$$\left(A_{\gamma\beta}+S_{\gamma\beta}\right)B^{\gamma}_{\alpha}N^{\alpha\beta}-\left(\varphi_{\beta}N^{\alpha\beta}\right)\Big|_{\alpha}+f=0$$

(b.2) Kinematics and Compatibility

The $\bar{A}_{\alpha\beta}$ and $\bar{B}_{\alpha\beta}$ must satisfy the continuity (compatibility) equations of surface theory. These are [16]:

$$\varepsilon^{\beta\gamma}\left(\bar{B}_{\alpha\beta,\gamma}-\bar{\Gamma}^{\lambda}_{\alpha\gamma}\bar{B}_{\lambda\beta}\right)=0 \qquad \text{Codazzi equations}$$

$$\varepsilon^{\alpha\gamma}\varepsilon^{\beta\delta}\left(\bar{B}_{\alpha\beta}\bar{B}_{\gamma\delta}+A_{\alpha\beta,\gamma\delta}+\bar{A}_{\lambda\mu}\bar{\Gamma}^{\lambda}_{\alpha\beta}\bar{\Gamma}^{\mu}_{\gamma\delta}\right)=0 \quad \text{Gauss equation.}$$

In the case of small strains, simplified equations are available, which are linear in the strains and their first order derivatives [19,20,14]. See section 4(a).

For the displacement formulation, the strain-displacement equations are [18]:

$$E_{\alpha\beta}=\frac{1}{2}\left(S_{\alpha\beta}+S_{\beta\alpha}+S^{\gamma}_{\alpha}S_{\gamma\beta}+\varphi_{\alpha}\varphi_{\beta}\right).$$

(b.3) Constitutive Relations:

Only hyperelastic materials are considered. Let $\Phi(E_{\alpha\beta})$ be the strain energy density per unit undeformed area. Then $N^{\alpha\beta}=\frac{1}{2}\left(\frac{\partial\Phi}{\partial E_{\alpha\beta}}+\frac{\partial\Phi}{\partial E_{\alpha\beta}}\right)$. Methods for obtaining Φ have been discussed elsewhere. See, for example Green and Adkins [10], Wu [21], Libai and Simmonds [13]. Approximate strain energy functions for many biological membranes have been obtained, although these apply under restricted conditions only [22,23].

"Rigid-inextensional" behavior where all the $E_{\alpha\beta}$ are set to zero (except for wrinkling) is also a constitutive relation.

While the assumption of dependence on the $E_{\alpha\beta}$ only appears to be valid for ideal membranes, this is not necessarily true in "shell membranes", where, in the most general

case, dependence on the $K_{\alpha\beta}$ might be expected too. However, to conform with common usage, the possible dependence on the $K_{\alpha\beta}$ will be ignored in the sequel.

(b.4) Boundary Conditions

These come out of the virtual work identity for the *deformed* membrane

$$\int_{\bar{\Gamma}} \bar{T}^* \cdot \delta\bar{V} d\bar{s} + \int\int_{\bar{a}} \bar{p} \cdot \delta\bar{V} d\bar{a} = \int\int_{\bar{a}} \bar{T}^{\alpha} \cdot \delta\bar{V}_{,\alpha} d\bar{a}$$

where $\bar{T}^*$ is the applied force per unit length and $\delta\bar{V}$ is the virtual displacement. Integration by parts yields the work form of the equilibrium equations and also the boundary conditions $(\bar{T}^* - \bar{T}^{\alpha}\bar{\nu}_{\alpha}) \cdot \delta\bar{V} = 0$.

The natural conditions in component form are $\bar{T}^*_s = \bar{T}^* \cdot \bar{\lambda} = \bar{N}^{\alpha\beta}\bar{\nu}_{\alpha}\bar{\lambda}_{\beta}$; $\bar{T}^*_{\nu} = \bar{T}^* \cdot \bar{\nu} = \bar{N}^{\alpha\beta}\bar{\nu}_{\alpha}\bar{\nu}_{\beta}$ and $\bar{T}^* \cdot \bar{N} = |\bar{T}^*| \cos(\bar{T}^*, \bar{N}) = 0$. The first two conditions equate the applied shearing and normal forces in the tangent plane to the corresponding stress resultants. The third is a condition on the field variable $\bar{N}$ itself (since the direction of $\bar{T}^*$ is known). It is an equilibrium condition, but it sets up a *kinematic* requirement on the *direction*. At a free edge, $\bar{T}^* = 0$ and only two conditions can be applied. This agrees with the nature of the membrane equations, as will be seen later. The vector form of the boundary conditions can also be related to undeformed directions, leading to natural conditions in the displacement formulation [18]:

$$\left[A_{\gamma\beta} + S_{\gamma\beta}\right] N^{\alpha\beta}\nu_{\alpha} = \mathbf{T}^* \cdot \mathbf{R}_{,\gamma}$$

$$-\varphi_{\beta} N^{\alpha\beta}\nu_{\alpha} = \mathbf{T}^* \cdot \mathbf{N} .$$

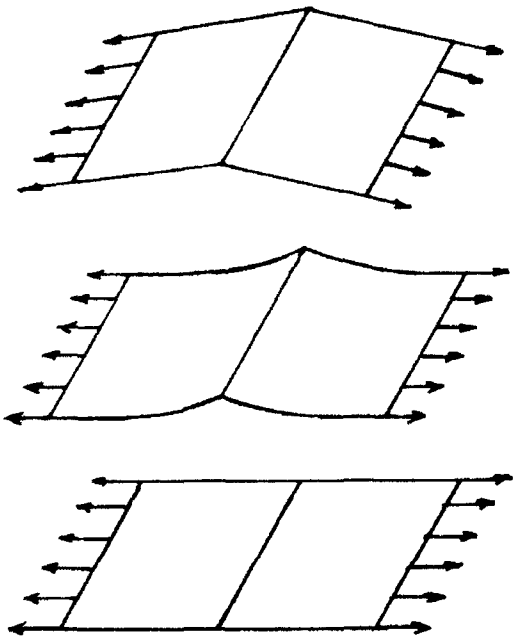
Again, if $\mathbf{T}^* = 0$, then, from the first equation $N^{\alpha\beta}\nu_{\alpha} = 0$ and the second equation is identically satisfied, so that no condition on φ_{β} can be set.

The kinematic conditions emerge from $\delta\bar{V} = 0$. In the displacement formulation they require the specification of the displacements v^{α}, w from the undeformed configuration. Taken with respect to the current state, the conditions $\delta\bar{v}^{\alpha} = 0$, $\delta\bar{w} = 0$ may be used in problems with *evolving* kinematic conditions. More commonly, they are translated into conditions on the *intrinsic* quantities on the boundary: the extensional strain ε_s, the normal curvature $\bar{K}_n$ and the geodesic curvature $\bar{K}_g$. For example, if the boundary does not deform, then $\varepsilon_s = 0$, $\bar{K}_g = K_g$, $\bar{K}_n = K_n$. In many engineering applications, the curvatures and strain are the obvious data.

A more general approach relates the natural and kinematic conditions to the behavior of a surrounding "boundary beam", where the ordinary differential equations of the beam become the boundary conditions of the membrane [24]. It is also possible to utilize variational formulations for a direct derivation of the boundary conditions in intrinsic form. As a word of caution, it must be emphasized that the boundary conditions strictly apply to ideal membranes only. In shell-membranes the bending component of the boundary conditions is always present to some degree. Also, compressive boundary loads may cause wrinkling in ideal membranes.

(c) Angular Discontinuities

Conditions of kinematic and static continuities exist across internal lines in the membrane: The deformed position and stress resultant vectors must be continuous, and, *unless the normal force across the line vanishes*, the unit normals $\bar{N}$ *must* coincide. The latter can also be obtained from the elementary observation that membrane *equilibrium* cannot be locally maintained in the presence of an angular discontinuity with a non-zero tensile force (see Fig. 1). This leads to two conclusions: (a) preexisting lines of



A. Not in equilibrum

B. Shell Membrane

C. Ideal Membrane

Figure 1. Angular discontinuity with tensile force.

angular discontinuity are smoothed out by a nonzero tensile force. (b) If the force vanishes, then a line of angular discontinuity is permitted (even if it did not pre-exist). As before, the above is strictly valid for ideal membranes under tension only. In shell membranes the membrane state breaks down near the line and bending effects intermix. Hence, the immediate region near the fold should be taken out of the analysis. Yet, the procedure of disregarding very local bending zones is common in the analysis of sufficiently thin highly loaded shell membranes. They can be thus analyzed as ideal membranes.

4. THE NATURE OF THE MEMBRANE PROBLEM

(a) Behavior of the Equations

The membrane field equations form a sixth order nonlinear partial differential system. Compared with linear theory, the nonlinear terms raise the order from four to six. They occur in the *highest derivatives* and involve the current state of stress of the membrane. Consequently, the membrane may behave in radically different ways, depending on the nature of the state of stress and deformation in the neighborhood of different solutions. A helpful and convenient method for the study of the system in a specific region of its variable-space is to expand it around a solution point ("reference state") in the region, and consider the perturbation series from this point.

Properties near this state can be gleaned from the linear terms in the expansion in the sense of "small deformations superposed on large" which has many applications in Nonlinear Mechanics [9,25]. In membranes it has been used to investigate stability and vibrations [18,26-29] but has additional practical applications in cases where a solution is sought which is "not far" in some sense from a known state [18,24,14,30]. Let (o) denote the reference state. Let: $e_{\alpha\beta}=E_{\alpha\beta}-E^{(o)}_{\alpha\beta}$, $k_{\alpha\beta}=\bar{B}_{\alpha\beta}-B^{(o)}_{\alpha\beta}$, $\gamma^{\alpha}_{\beta\gamma}=\bar{\Gamma}^{\alpha}_{\beta\gamma}-\Gamma^{\alpha(o)}_{\beta\gamma}$, $n^{\alpha\beta}=N^{\alpha\beta}-N^{\alpha\beta}_{(o)}$, Δp^{α}, Δp be the increments from the reference state. Restrictions are imposed that the $e_{\alpha\beta}$ and $\gamma^{\alpha}_{\beta\gamma}$ be sufficiently small such that *linearization* can be performed in these variables, and that no wrinkling takes place. Substitution into the field equations yields the following set of *incremental equations*:

$$n^{\alpha\beta}\big|_{\alpha} + N^{\alpha\beta}_{(o)}\gamma^{\beta}_{\alpha\gamma} + \Delta p^{\beta} = 0$$

$$n^{\alpha\beta}B^{(o)}_{\alpha\beta} + N^{\alpha\beta}_{(o)}k_{\alpha\beta} + \underline{n^{\alpha\beta}k_{\alpha\beta}} + \Delta p = 0$$

$$\varepsilon^{\beta\gamma}(k_{\alpha\beta}\big|_{\gamma} - B^{(o)}_{\lambda\beta}\gamma^{\lambda}_{\alpha\gamma}) = 0$$

$$\hat{B}^{\alpha\beta}_{(o)}k_{\alpha\beta} + \hat{e}^{\alpha\beta}\big|_{\alpha\beta} + \frac{1}{2}\underline{\hat{k}^{\alpha\beta}k_{\alpha\beta}} - K_o e^{\alpha}_{\alpha} = 0$$

$$n^{\alpha\beta} = C^{\alpha\beta\gamma\delta}e_{\gamma\delta} = \left[\Phi,_{E_{\alpha\beta},E_{\gamma\delta}}\right]^{(o)}e_{\gamma\delta}$$

where K_o is the Gaussian curvature of the reference state (o), and all tensorial operations are performed in this state. For quadratic strain energy densities, the $C^{\alpha\beta\gamma\delta}$ is the linear elastic tensor. Since the $k_{\alpha\beta}$ are not assumed to be small, nonlinear (underlined) terms are retained (in fact, were the (o) system chosen as the *undeformed* state, then a small-strain-finite-rotation membrane theory would have resulted!). In the above, the cofactor tensor is defined by $\hat{f}^{\alpha\beta}=\varepsilon^{\alpha\beta}\varepsilon^{\beta\delta}f_{\gamma\delta}$ and the incremental Christoffel tensor is given by its linearized form:

$$\gamma^{\alpha}_{\beta\gamma} = e^{\lambda}_{\alpha}\big|_{\beta} + e^{\lambda}_{\beta}\big|_{\alpha} - e_{\alpha\beta}\big|^{\lambda}$$

For a complete linearization, the underlined terms should be dropped. Henceforth, $\Delta p^{\alpha}=0$ is assumed for simplicity. To facilitate further study, a D.M.V.-type approximation is made by replacing *tangential* equilibrium and compatibility with a stress function F and curvature function G:

$$n^{\alpha\beta} = \hat{F}\big|^{\alpha\beta} + R_n \qquad k_{\alpha\beta} = G\big|_{\alpha\beta} + R_k$$

where the R_i are lower order remainder terms. This approximation has a wide range of applicability in shell theory. For the present study it suffices to note that it is also useful for a qualitative evaluation since it preserves the highest derivatives of the system (however, more exact substitutions are available, as will be seen later). Furthermore, the incremental constitutive relations are assumed to be isotropic with effective moduli Et and ν (this is exact for quadratic strain energies). Substitution

and rearrangement yield:

$$L_b(F) + L_n(G) + \Delta p = R_N \qquad \text{(normal equilibrium)}$$

$$\frac{1}{Et} \nabla^4(F) + L_b(G) = R_B \qquad \text{(compatibility)}$$

Here, R_i are remainder terms involving lower order derivatives and the differential operators L_b, L_n and ∇^4 are

$$L_b(f) = \hat{B}^{\alpha\beta}_{(o)} f\Big|_{\alpha\beta} = \text{"curvature operator"}$$

$$L_n(f) = N^{\alpha\beta}_{(o)} f\Big|_{\alpha\beta} = \text{"stress operator"}$$

$$\nabla^4(f) = f|^{\alpha\beta}_{\ \ \ \

(2) In cases with sufficiently rigid boundaries and small strains, the $N^{\alpha\beta}_{(o)}$ of the linear membrane solution plus the $B^{(o)}_{\alpha\beta}$ of the undeformed state can be used for approximate analysis. In shell-membranes this has the advantage of easy merging with bending effects where appropriate.

(3) Obviously, the chosen state should be sufficiently "close" to the final solution in order to ensure meaningful results (easier said then done).

(b) Dependence on the initial shape

A study of the intrinsic form of the equations shows that the only initial data which is formally needed is the metric $A_{\alpha\beta}$ (which appears in the constitutive relations) and the definition of the boundary $\Gamma: u^{\alpha}=u^{\alpha}(S)$. All other properties of **R**, including the undeformed curvatures $B_{\alpha\beta}$ do not appear in the equations. The boundary data and loading are usually related to the final form (e.g. pressure) or to extrinsic directions (e.g. axial loads in a shell of revolution), but rarely to the initial shape. It follows that all membranes having the same $A_{\alpha\beta}$ and Γ should have the same final state. This includes the class of surfaces which are isometric to the initial surface, *including those with any number of angular discontinuities (folds)*. For all practical purposes, the membrane can be, initially, even *"packed in a box"*.

The displacement formulation requires an initial surface to measure the displacements from. There is some arbitrariness in the choice since any surface from the isometric set can be chosen. If the *initial* metric plus kinematic boundary data ensure the existence of a unique $B_{\alpha\beta}$ in the class of smooth functions, then the resulting surface is commonly chosen for **R**. It should be, however, noted that this is an artifice, since the boundary data is provided for the final state only, and a deformation of the boundary between the initial and final states is possible. To cite an example: the shapes of the ends of a closed pressurized cylindrical membrane are assigned to be plane ellipses. Yet, the initial shapes can be circular, with specified edge displacements as kinematic boundary conditions. This choice of initial shape might be easier to solve than the choice of an initial elliptical shape.

A shell-membrane *does have* a unique undeformed body to start from. The need for it arises whenever an interaction with bending theory is required. However, it is not a part of the formal membrane data, unless it is chosen as the initial state.

(c) Wrinkling and Tension Fields

Very thin shells and other shells having low bending rigidities can support little compressive stress before they buckle. If such shells are pulled in one direction and compressed in the other (such as when subjected to shear), a large number of high-aspect-ratio buckles forms with creases oriented in the tension direction. As the bending rigidity goes down, so do the buckling stress and the distance between crease lines. In ideal membranes, the instability stress approaches zero and the number of buckles becomes very large, thus forming a typical "tension field", which is not an uncommon sight in very thin, flexible membranes. The assumed inability to carry compressive stresses is common to all ideal membranes. It has been initially used by Wagner [31] to examine the postbuckling behavior of thin plates in shear, and was shown by Steigmann [32] to be exact. Several investigators undertook the study of tension fields in *plane* membranes [33-38]. Wrinkled biological membranes were studied by Danielson et al. [39]. Curved wrinkled membranes, which are of more interest in this review, were studied by Kondo [40,41], Wu [42,43], Zak [44] and others. A more extensive study of the foundations of plane and curved tension field theory was initiated by Pipkin [45], Steigmann [46] and in collaboration [47,48]. A numerical approach was

developed by Roddeman et al. [49]. The instability due to wrinkling of bent pressurized cylindrical membranes was investigated by Comer and Levy [50], Koga [27] and Łukasiewicz et al. [51]. The ponding instability of partly wrinkled spherical membranes was analyzed by Glockner and his associates [52-56], Croll [57] has studied the formation of wrinkles in transversely loaded plates and Libai analyzed boundary layers adjacent to wrinkles [58]. The list is not exhaustive and will, in all probability expand in the near future as partial wrinkle fields will be routinely incorporated into nonlinear membrane analysis. Also, more effort should be put into "incomplete tension fields" which describe the transition from shell to wrinkled membrane, and into the wrinkling of biological membranes.

The underlying idea of tension field theory is to replace the complex wrinkled surface (with its random form of detailed wrinkles) with a *smoothed out* "pseudo surface" [43] or "wrinkle surface" which could be imagined as a continuous distribution of wrinkles or fibers, and constitutes a regular surface in space. This is equivalent to the concept of a generalized curve or surface in the calculus of variations [59]. The pseudo-surface should satisfy the equilibrium equations, with one of its principal stress resultants being positive and the other set to zero. For convenience, a non-material orthogonal coordinate system which coincides with the principal stress directions is employed. Nontensorial notation is used, where $\bar{N}_1$ is the nonvanishing principal resultant, $\bar{\mathbf{e}}_\alpha$ are unit vectors in the principal directions and $\bar{A}_\alpha=\sqrt{\bar{A}_{\alpha\alpha}}$. Then:

$$\bar{\mathbf{T}}^1 = \bar{N}_1\bar{\mathbf{e}}_1, \qquad \bar{\mathbf{T}}^2=0, \qquad \bar{\mathbf{e}}_{1,1}= -\frac{1}{\bar{A}_2}\bar{A}_{1,2}\bar{\mathbf{e}}_2 + \frac{1}{\bar{r}_1}\bar{A}_1\bar{\mathbf{N}}, \qquad \bar{\mathbf{p}}=\bar{p}_{(\alpha)}\bar{\mathbf{e}}_\alpha + \bar{p}\bar{\mathbf{N}}.$$

The vector and component equations of equilibrium become, respectively:

$$(\bar{N}_1\bar{A}_2\bar{\mathbf{e}}_1)_{,1} + \bar{A}_1\bar{A}_2\bar{\mathbf{p}} = 0$$

$$(\bar{A}_2\bar{N}_1)_{,1} + \bar{A}_1\bar{A}_2\bar{p}_{(1)} = 0$$

$$-\bar{A}_{1,2}\bar{N}_1 + \bar{A}_1\bar{A}_2\bar{p}_{(2)} = 0$$

$$\frac{1}{\bar{r}_1}\bar{N}_1 + \bar{p} = 0$$

where both $\bar{N}_1$ and the geometry $(\bar{A}_1,\bar{A}_2,\bar{r}_1)$ are unknowns. The nature of the loading determines the properties of the stress trajectories:

(1) If $\bar{p}_{(2)}=0$, then $\bar{\mathbf{e}}_{1,1}$ is parallel to $\bar{\mathbf{N}}$ and the trajectories are *geodesics*.

(2) If $\bar{p}=0$, then $\frac{1}{\bar{r}_1}=0$ and the lines are *asymptotic* (with $\bar{K}\leq 0$).

(3) If both are zero, then $\bar{\mathbf{e}}_{1,1}=0$. The lines are straight, and the surface is *ruled*.

(4) If $\bar{p}_{(1)}=0$, then the force-per-element $(\bar{N}_1\bar{A}_2)$ is constant. This is analogous to *flow lines* in fluid flow.

The geometric variables $(\bar{A}_1,\bar{A}_2,\bar{r}_1,\bar{r}_2,\bar{r}_{12})$ are connected through the continuity

equations of Codazzi and Gauss.

To tie the solution to the rest of the membrane and to the initial metric, constitutive relations are needed. Only isotropic membranes will be considered at this stage (non-isotropic membranes were studied by Mansfield [60] and Roddeman [49]). Here, the principal directions of stress and strain coincide. The condition $N_2=0$ can be used to eliminate the true principal strain E_{22} from the constitutive relations, which leaves $N_1=f(E_{11})$. The difference between the *apparent* strain $\bar{E}_{22}$ of the pseudo-surface and the true strain is the "pseudo strain" $E^*_{22}=\bar{E}_{22}-E_{22}$. It was introduced by Wu [43] and is used as an extra variable in lieu of the extra equation $N_2=0$ which was imposed on the field equations. The constitutive relation does not add an additional variable, but should be used to relate the tension field to the material system. If there are no additional symmetries, the problem may be complex [46].

Other methods for constitutive modification where employed by Reissner [33] who used a special orthotropic material law and Stein et al. [34] who introduced a variable Poisson ratio. More recently, Pipkin [45] introduced a "relaxed" energy density to account automatically for the modified constitutive relations.

The boundary conditions with the unwrinkled portions of the membrane are obtained from continuity requirements:

(1) The resultant $N_1 \geqq 0$ is continuous across the boundary.

(2) $N_2=0$ as the boundary is approached from the unwrinkled side.

(3) If $N_1 \neq 0$, then angular continuity is preserved.

(4) Positional continuity is maintained, thus implying $E^*_{22}=0$.

Examples: See section 5 for the axisymmetric case. See section 7 for the edge zone near a wrinkle field. Further investigations and details may be found in the literature cited above.

5. AXIMEMBRANES

These are membranes whose undeformed shapes are surfaces of revolution and are axisymmetrically deformed. Because of their relative "simplicity" they have been studied extensively in literature and solutions to many nonlinear problems exist. Hence, their discussion here will be limited in scope. More information and details can be found in the literature and some have been gathered in chapters in books. Examples are Green and Adkins [10] who discuss constitutive relations and some finite strain problems, and Libai and Simmonds [13, chapter 5] who made a more extensive survey with examples.

The reference surface of the undeformed membrane is defined in a fixed cylindrical coordinate system (r,θ,z) by its meridional position vector $\mathbf{R}=r(s)\mathbf{e}_r+z(s)\mathbf{e}_z$ where s is arclength. The angle α between $\mathbf{R}_{,s}$ and $\mathbf{e}_r$ satisfies $r_{,s}=\cos\alpha$; $z_{,s}=\sin\alpha$. The corresponding quantities in the deformed meridian are $\bar{r}$, $\bar{z}$, $\bar{\alpha}$. A nontensorial approach is adopted. Meridional and circumferential stretches are defined by $\lambda_s=\frac{d\bar{s}}{ds}$ and $\lambda_\theta=\frac{\bar{r}}{r}$, respectively. Stress resultants are $\bar{N}_s$ and $\bar{N}_\theta$ (or N_s and N_θ per unit undeformed length). The loading vector is $\bar{\mathbf{p}}=\bar{p}_H\mathbf{e}_r+\bar{p}_v\mathbf{e}_z$ and may depend on the field variables. The equilibrium

equations are

$$(rN_s\sin\bar{\alpha})_{,s} + rp_v = 0 \qquad (rN_s\cos\bar{\alpha})_{,s} = N_\theta - rp_H$$

In the case of uniform pressure $\bar{p}$ one may also use the partly integrated form:

$$rN_s\sin\bar{\alpha} = \frac{1}{2}\bar{p}\,\bar{r}^2 - c_1 \qquad (rN_s)_{,s} = N_\theta\cos\bar{\alpha}$$

To these are added the equations of geometry and constitution:

$$\bar{r}_{,s} = (r\lambda_\theta)_{,s} = \lambda_s\cos\bar{\alpha} \qquad \bar{z}_{,s} = \lambda_s\sin\bar{\alpha}$$

$$N_s = \Phi_{,\lambda_s} \qquad N_\theta = \Phi_{,\lambda_\theta}$$

where $\Phi(\lambda_s,\lambda_\theta)$ is the strain energy density. For boundary conditions, the position or stress resultant vectors should be specified, subject to overall equilibrium. The system is 4th order (but can be usually reduced to second order plus two quadratures). It is a two-point b.v.p. which requires appropriate numerical procedures (e.g. Runge Kutta plus shooting) or approximate techniques for solution.

Pressurized wrinkled aximembranes with meridional tension fields: Here N_θ=0. Further integration yields $rN_s=\bar{r}\,\bar{N}_s=c_2$. The meridian of the pseudo-surface can be expressed explicitly in terms of the Jacobi elliptic function dn with modulus k:

$$\bar{r} = \frac{2}{k}\sqrt{c_2/\bar{p}}\;\mathrm{dn}\left[\frac{1}{k}\sqrt{\bar{p}/c_2}\;(\bar{s}-\bar{s}_o)\right] \qquad \text{where} \qquad k^2 = 2(1 + c_1/c_2)^{-1}$$

To relate the result to the undeformed surface, the equations N_θ=0 and $N_s r(s)=c_2$ are used in the constitutive relations to eliminate λ_θ and express λ_s in terms of s. This establishes $\bar{s}(s)=\int\lambda_s ds$. If stretching is neglected (λ_2=1) then $\bar{s}$= s. The constants c_i are obtained from the edge conditions of the wrinkled zone.

Aximembranes cannot deform inextensionally: any radial displacement and any nonuniform axial displacement produce straining (an exception is near the apex where a local semi-inextensional deformation is possible). However, if wrinkles are permitted, then inextensional deformation become possible and even common [13,42,52-56]. Barring wrinkles, it follows from small strain moderate rotation theory that *linear* theory is valid interior solution for many aximembranes, with narrow boundary layers of size $r\sqrt{\epsilon}$ added to account for support conditions. Many studies of this phenomenon exist [61-65]. The use of nonlinear correction for internal continuity finds its use in toroidal membranes [66,67]. In problems involving large strains, boundary layers usually do not exist, and problems associated with inflation and stability are the more important topics. Constitutive relations assume a major role and combine with the evolving geometry to produce limit point [10,68] or bifurcation type [69-72] instabilities.

Aximembranes can be classified by their undeformed shapes. Within each class the small-strain problem is studied separately. Some cases will be briefly reviewed:

Plane Circular Membranes: The small strain moderate relation theory is due to Föppl [73]. A less restrictive theory is due to Reissner [74]. Many solutions for pressure and/or end loaded circular and annular membranes are found in the literature. Some were summarized by Libai and Simmonds [13]. More have recently appeared [75-79]. In the case of large strains, inflation [10,80-83] and deformation under concentrated loads [84-86]

are some of the studies of interest. The inverse problem of deformation into a predetermined inflated shape has also been considered [81,82,87,88].

Spherical Membranes: Small strain problems lead to the study of edge effects. Shallow spherical membranes were treated by Dickey [76]. The common large strain problem is that of the inflation of spherical balloons. Of special interest is the possible existence of a maximum pressure. In volume-controlled problems, a further *increase*; in volume would lead to a *decrease* in the pressure. this trend may reverse itself at higher volumes, and is strongly material dependent [10,69]. The bifurcation of a spherical balloon into a nonspherical shape has also been studied [69-72].

Circular Cylindrical Membranes: In the small strain range the main issue is the edge effect due to kinematic constraints or nonaxial end loads. Also if the end load is radial, a small part of the cylinder deforms into an annulus, while the rest remains cylindrical. At the line of angular discontinuity, $N_s=0$ [13,62]. The large strain problem has the first integral [89] $\Phi-\lambda_s\Phi_{,\lambda_s}=C$, where C is a constant of integration. This has been used to reduce the large strain problem of pressurized cylindrical membranes to quadratures [13,90,91]. The stretching of a cylindrical membrane attached to rigid end rings has also been studied [82,90]. Finally, as in spherical membranes, the possibility of a maximum pressure exists here too [92,68].

Toroidal Membranes: The important issue in the small strain theory for *circular* toroidal membranes is adjustment of the apparent discontinuity at the apex. This was accomplished by Jordan [66] and by Sanders and Liepin [67]. Problems involving large strains were treated by Kydoniefs and Spencer [93]. The major difficulty in *nonsymmetric* pressurized toroidal ideal membranes (where $r(\alpha=0)\neq r(\alpha=\pi)$) lies in the fact that a *necessary condition* for equilibrium is $\bar{r}(\bar{\alpha}=0)=\bar{r}(\bar{\alpha}=\pi)$. This must hold regardless of the magnitude of the pressure. Slight asymmetries can be taken care of by local nonlinear rotations which, as pointed out by Kuznetsov [94], are nearly inextensional at $\alpha=0,\pi$. However, if the asymmetry is substantial, then large strains are required to take place even if the pressure is very small. This "toroidal paradox" was used by Goldenveizer [2] in discussion of the existence of membrane solutions, however, if wrinkling is accepted as a part of the deformation process, then a solution might be possible, and should be tried. To account for axial equilibrium, the tension field must be meridional ($N_\theta=0$) wherever it occurs. In addition, wherever $N_s=0$, a circumferential line of angular discontinuity is permitted. The membrane is unstable at its undeformed form, so that large scale wrinkling takes place even at low pressures.

For the shape of the membrane in unwrinkled regions, the general equations should be used, but if the pressure is low, the extensional strains can be neglected and the forces calculated by linear membrane theory. In the wrinkled regions, the theory for wrinkled axisymmetric membranes [43,47] applies, and the shape can be calculated without difficulty, as was seen earlier in this section.

Large Strain Asymptotic Behavior: In the case of very large strains a simplified asymptotic analysis can be made. Of some interest is the possible tendency of closed inflated aximembranes to assume a *spherical* shape as the pressure and stretches become very large. This was shown for a class of materials of the Mooney type by Isaacson [95], Wu [21] and Sagiv [96]. Antman and Calderer [97] investigated the restrictions that must be put on the energy density of a spheroidal shell is to approach a spherical shape. The behavior of Neo-Hookean membranes with very large meridional strains was studied by Foster [98].

6. CYLINDRICAL MEMBRANES

In this section, some problems related to pressurized and stretched noncircular cylindrical membranes are discussed. Included are cases where the boundary has a non-circular (oval) form, or where processes cause an initially circular membrane to become non-circular in its deformed state. The general procedure is that of an incremental analysis, with the pressurized and stretched unbounded circular cylindrical membrane taken as the "reference state" (section 4a). This is a point in the solution domain for all cylindrical membranes. It has a simple unvarying geometry to which all homogeneous pressurized oval membranes tend, far enough from their ends. The main restriction is that the final deformed state should be sufficiently "close" to the reference state so as to make an incremental analysis (including linearization in the metric and its first order derivatives) valid in some approximate sense.

Let x and s denote longitudinal (generator) and circumferential (hoop) arclength coordinates *in the reference state*. The membrane material is assumed to be nonlinearly homogeneous and orthotropic, such that a circular cylindrical membrane of length L and radius r is transformed by pressure $\bar{p}$ and total tension T into a circular cylinder of length L_o and radius r_o. Let $\bar{N}^{xx}_{(o)}=P_x=\frac{T}{2\pi r_o}$ and $\bar{N}^{ss}_{(o)}=P_s=\bar{p}r_o$ be the physical Cauchy stress resultants in this reference state. Let n_x, n_s, n_t, ε_x, ε_s, ε_t, κ_x, κ_s, κ_t be the incremental stress resultants, strains and curvatures, all measured with respect to the Cartesian metric (x,s). Then, using (4a), the following field equations result:

$$(n_x - P_s\varepsilon_s + P_x\varepsilon_x)_{,x} + (n_t + 2P_s\varepsilon_t)_{,s} = 0$$

$$(n_s - P_x\varepsilon_x + P_s\varepsilon_s)_{,s} + (n_t + 2P_x\varepsilon_t)_{,x} = 0$$

$$P_x\kappa_x + P_s\kappa_s + \frac{1}{r_o} n_s + \underline{(n_x\kappa_x + n_s\kappa_s + 2n_t\kappa_t)} = 0$$

$$(\kappa_x - \frac{1}{r_o}\varepsilon_x)_{,s} - (\kappa_t - \frac{2}{r_o}\varepsilon_t)_{,x} = 0$$

$$(\kappa_s - \frac{1}{r_o}\varepsilon_s)_{,x} - \kappa_{t,s} = 0$$

$$\frac{1}{r_o}\kappa_x + \varepsilon_{x,ss} + \varepsilon_{s,xx} - 2\varepsilon_{t,xs} + \underline{\kappa_x\kappa_s - \kappa_t^2}\ 0$$

$$\varepsilon_x = f_x n_x - f_{xs} n_s$$

$$\varepsilon_s = f_s n_s - f_{xs} n_x$$

$$2\varepsilon_t = f_t n_t$$

where the f_i are constitutive coefficients obtained by inverting the incremental constitutive relations of the reference state in terms of Cauchy resultants. For quadratic energy densities, these are the elastic constants, in which case the equations should be similar to those of Libai [30] and Danielson and Simmonds [99]).

Following section (4a), but without making the approximations therein, stress and curvature functions are defined by putting $n_t=-F_{,xs}$ and $K_t=G_{,xs}$. Substitution yields:

$$n_x = (1 + \alpha_x)F_{,ss} + \alpha_{xs}F_{,xx} \quad ; \qquad n_s = (1 + \alpha_s)F_{,xx} + \alpha_{sx}F_{,ss}$$

$$\varepsilon_x = f_x(1 + \beta_x)F_{,ss} - f_{xs}(1 + \beta_{xs})F_{,xx}$$

$$\varepsilon_s = f_s(1 + \beta_s)F_{,xx} - f_{xs}(1 + \beta_{sx})F_{,ss} \quad ; \qquad 2\varepsilon_t = -f_tF_{,xs}$$

$$\kappa_s = G_{,ss} + \frac{1}{r_o}\varepsilon_s \quad ; \qquad \kappa_x = G_{,xx} + \frac{1}{r_o}(\varepsilon_x - 2f_tF_{,xx})$$

where the α_i and β_i are constants having the order of magnitude of the strains at the reference state, and are easily calculated [100]. Introduction into the remaining two equations leads to one 4th order and one 2nd order equation in F and G. These are non-linear due to the (FG) and (GG) terms - as is common in similar formulations.

In many practical applications all the nonlinear terms in the increments may be suppressed (at least as a first approximation). If, in addition, the strains at the reference state are small, the equations simplify to

$$\frac{1}{r_o}G_{,xx} + \nabla_1^4F + \frac{1}{r_o^2}[f_xF_{,ss} + (2f_t - f_{xs})F_{,xx}] = 0$$

$$P_xG_{,xx} + P_sG_{,ss} + \frac{1}{r_o}F_{,xx} - \frac{2}{r_o}P_sf_{xs}F_{,ss} = 0$$

where the modified biharmonic operator ∇_1^4 is defined by

$$\nabla_1^4F = f_sF_{,xxxx} + f_xF_{,ssss} + 2(f_t - f_{xs})F_{,xxss}$$

Further elimination yields the 6th order equation

$$r_o^2\nabla_1^4(P_xF_{,xx} + P_sF_{,ss}) - F_{,xxxx} = (P_xf_x + P_sf_{xs} - 2P_sf_t)F_{,xxss} - P_sf_xF_{,ssss}$$

In the DMV approximation, the right hand side of the equation is set to zero. For quadratic energy densities of *isotropic* membranes, set: $f_x=f_s=\frac{1}{Et}$, $f_{xs}=\frac{\nu}{Et}$, $f_t=\frac{1+\nu}{Et}$, $\nabla_1^4F=\frac{1}{Et}\nabla^4F$. As a practical application, consider the problem of an oval cylindrical membrane subjected to uniform pressure $\bar{p}$ and an equally distributed tensile force $P_x=T/S$ on its boundaries x=0,L. Diaphragm-type supports maintain the oval shape r. Small strain theory is assumed. The reference surface is chosen to be a circular cylinder with radius r_o and length L_o which are obtained by applying the loads to the *isoperimetric* cylinder with radius $r_a=S/2\pi$ and length L. The approach consists of applying to the incremental problem the boundary conditions $\varepsilon_s=\frac{r_a-r_o}{r_a}=\text{const.}$; $\kappa_s(s)=\frac{1}{r(s)} - \frac{1}{r_o}$; $n_x = 0$.

Of the three conditions, that of ε_s leads to an edge-effect correction, which is strongly affected by P_x, while that of κ_s affects large parts of the membrane ("slow

decay"), which is of more interest. A similar situation exists in the semi-membrane analysis of tubes [101] where the effects of ε_s tend to be suppressed.

Solutions to the case of lateral pressure (P_x=0) were studied by Libai [24] who also suppressed the boundary conditions on ε_s, and used a Fourier expansion in the s direction. The solutions possess important features which compare favorably with those of the more common linear membrane analysis of an oval. The latter has some serious deficiencies which render it unacceptable in many applications: (a) It requires the continuity of r(s) and its derivatives (up to the 4th !). (b) It does not tend to the circular form far from the edges. In fact, it *diverges* as L→∞. (c) It has no decay mechanism for boundary disturbances. (d) It cannot accommodate boundary conditions along its generators (important for cylindrical panels).

The behavior of the incremental solution relates to the parameter $\gamma = \frac{\bar{p}L^4}{Etr_a^3} = \varepsilon_s\left(\frac{L}{r_a}\right)^4$. (a) It yields smooth solutions even if the slope or curvature of the initial form have discontinuities. (b) It approaches the circular shape as γ increases. (c) The typical decay length θ_x for end disturbances with harmonic index n is $\theta_x=r_a n^{-3/2}\varepsilon_s^{-1/4}=Ln^{-3/2}\gamma^{-1/4}$. (d) It can accept "flexible" boundary conditions along the generators [30]. (e) It approaches the linear membrane solution for γ<<1, and the original shape as $\bar{p}\to 0$. (f) For large n, it approaches the solutions of the biharmonic equation of *plane elasticity*. The method is being applied to other problems [100] including those of cracks in cylindrical membranes.

An akin problem is that of nonlinear effects in beam-bending of tubes (Brazier Problem). Here, circumferential shell bending is commonly assumed to resist the ovalization which accompanies the beam bending of circular cylindrical tubes. However, if the tube is thin and not very long, membrane resistance to ovalization which comes out of the supports provides the dominant mechanism. Pressurization enhances the effect.

7. EXAMPLE: EDGE EFFECT IN A PARTLY WRINKLED AXIMEMBRANE

Consider a truncated ideal spherical membrane of radius R, connected to rigid rings of radius r_c along its boundaries $\alpha=\alpha_c$, $\alpha=\pi-\alpha_c$ (Fig. 2). If the rings are pulled apart by forces P, the spherical membrane collapses inward, and a meridional tension field in the form of a pseudocylinder is created. It extends *almost* to the supports, but a *very narrow* wrinkled region forms near the rings [47]. The equations in the unwrinkled region are those of aximembranes (section 5). Axial equilibrium yields $2\pi r N_r \sin\bar{\alpha}=P$. The other equations can be reduced to a second order equation in $y=\cot\bar{\alpha}$ as function of $\alpha=s/R$. The boundary conditions at α_c is $\bar{r}_c=r_c$(or $\lambda_\theta=1$). At the boundary of the wrinkled region $\alpha=\alpha_B$, the conditions are those of continuity: $\bar{\alpha}_B=\pi/2$ and $N_\theta=0$. These can be converted into $y=y_{,\alpha}=0$. Note that α_B *is also unknown*, so that three conditions are needed for the differential equation and α_B.

The availability of *two conditions* at α_B facilitates the solution of the equation as

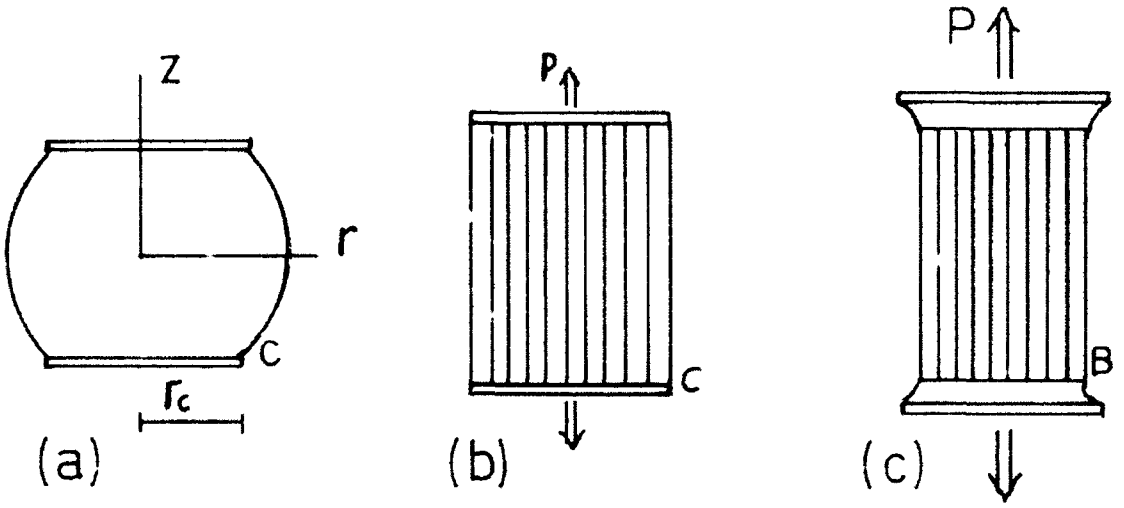


Figure 2. Partly wrinkled truncated spherical membrane.
(a) Original (b) Wrinkled (c) Wrinkled with edge zone

an *initial value problem* starting from α_B by either a power series or numerical integration [58]. The condition $\lambda_\theta=1$ determines α_c and consequently α_B. In small strain theory, the equations are:

$$(\sin\alpha y_{,\alpha})_{,\alpha} - (\csc\alpha)y = \varepsilon^{-1}\left[\frac{y}{\sqrt{1+y^2}} - \cos\alpha\right]$$

$$(\sin\alpha)y_{,\alpha} = \nu\sqrt{1+y^2} \quad \text{at} \quad \alpha = \alpha_c \;\; ; \;\; \varepsilon = \frac{P}{2\pi REt}$$

Construction of a formal power series solution yields the 1st order solution:

$$\alpha_B - \alpha_c = \frac{\nu\varepsilon}{\cos\alpha_c} + 0(\varepsilon^2) \qquad (\alpha_c \neq \pi/2).$$

Near $\alpha_c=\pi/2$ this solution breaks down and an additional term is needed, yielding

$$\alpha_B - \alpha_c = \sqrt{2\nu\varepsilon} + 0(\varepsilon^{3/2}) \qquad (\alpha_c \sim \pi/2)$$

Large-strain results for a Neo-Hookean material were obtained by numerical integration. These indicate the breakdown of the edge effect for large strains, as expected. All the other relevant parameters of the problem were also calculated [58].

Boundary layers $0(\varepsilon)$ are unusual in shell theory. In small strain theory they could be virtually ignored if not for the large rotations within the layer which indicate the possibility of strong interaction with bending effects in thin shell-membranes. They are not easily "captured" by pure numerical methods, unless looked for.

8. REFERENCES

1 S.P. Timoshenko and S. Woinowsky-Krieger, Theory of Plates and Shells, 2nd ed., McGraw Hill, 1959.
2 A.L. Goldenveizer, Theory of Elastic Thin Shells, Pergamon Press, 1961.
3 H. Kraus, Thin Elastic Shells, Wiley, 1967.
4 V.V. Novozhilov, Thin Shell Theory, 2nd ed., Wolters-Nordhoff, The Netherlands, 1970.
5 W. Flügge, Stresses in Shells, 2nd Ed., Springer Verlag, 1973.
6 C.R. Calladine, Theory of Shell Structures, Cambridge University Press, 1983.
7 P.L. Gould, Analysis of Shells and Plates, Springer Verlag, 1987.
8 K.M. Mushtari and K.Z. Galimov, Nonlinear Theory of Thin Elastic Shells, Israel Program for Scientific Translations, Jerusalem, 1961 (NASA-TT-F62).
9 A.E. Green and W. Zerna, Theoretical Elasticity, 2nd Ed., Clarendon Press, Oxford, 1968.
10 A.E. Green and J.E. Adkins, Large Elastic Deformations, 2nd Ed., Clarendon Press, Oxford, 1970.
11 F.I. Niordson, Shell Theory, North Holland, 1985.
12 J.W. Leonard, Tension Structures, McGraw Hill, 1988.
13 A. Libai and J.G. Simmonds, The Nonlinear Theory of Elastic Shells, Academic Press, San Diego, 1988.
14 A. Libai and J.G. Simmonds, Nonlinear Elastic Shell Theory, in Advances in Applied Mechanics (Hutchinson and Wu, eds.), 23, 1983, 271-371.
15 P. Kuhn, Stresses in Aircraft and Shell Structures, McGraw Hill, New York, 1956.
16 L.P. Eisenhart, An Introduction to Differential Geometry, with Use of the Tensor Calculus, 1947.
17 D.J. Struik, Lectures on Classical Differential Geometry, Addison-Wesley, 1950.
18 B. Budiansky, Notes on Nonlinear Shell Theory, ASME J. Appl. Mech., 35 (1968) 393.
19 W.T. Koiter, On the Nonlinear Theory of Thin Elastic Shells, I-III, Proc. Kon. Ned. Ak. Wet., B69, 1966, 1-54.
20 D.A. Danielson, Simplified Intrinsic Equations for Arbitrary Elastic Shells, Int. J. Eng. Sci., 8 (1970) 251.
21 C.H. Wu, Large Finite Strain Membrane Problems, Q. Appl. Math., 36 (1979) 347.
22 Y.C. Fung, On Pseudoelasticity of Living Tissue, Mechanics Today, 5 (1980) 49.
23 A.E. Evans and R. Skalak, Mechanics and Thermodynamics of Biomembranes, CRC Press, Boca Raton, Florida, 1980.
24 A. Libai, The Nonlinear Membrane Shell with Application to Noncircular Cylinders, Int. J. Solids Struct., 8 (1972) 923.
25 A.E. Green, R.S. Rivlin and R.T. Shield, General Theory of Small Elastic Deformations Superposed on Finite Elastic Deformations, Proc. R. Soc. London, A211 (1952) 123.
26 A.H. Corneliussen and R.T. Shield, Finite Deformation of Elastic Membranes with Application to the Stability of an Inflated and Extended Tube, Arch Rat. Mech. Anal., 7 (1961) 273.
27 T. Koga, Bending Rigidity of an Inflated Circular Cylindrical Membrane of Rubbery Material, AIAA J., 10 (1972) 1485.
28 R.T. Shield, On the Stability of Finitely Deformed Elastic Membranes, ZAMP, 23 (1972) 16.
29 A. Needleman, Necking of Pressurized Spherical Membranes, J. Mech. Phys. Solids, 24 (1976) 339.
30 A. Libai, Pressurized Cylindrical Membranes with Flexible Supports, IASS Int. Symp. on Pneumatic Structures, Delft, 1972.

31 H. Wagner, Ebene blechwanträger mit sehr dünnem stegblech, Z. Flugtech. Motorluftschiffart, 20 (1929) 200 (translated as NASA TM 604-606).
32 D.J. Steigmann, Proof of a Conjecture in Elastic Membrane Theory, J. Appl. Mech. 53 (1986) 955.
33 E. Reissner, On Tension Field Theory, Proc. 5th Int. Cong. Appl. Mech., 1938, 359-361.
34 M. Stein and J.M. Hedgepeth, Analysis of Partly Wrinkled Membranes, NASA TN D-813, 1961.
35 G.P. Cherpanov, On the Buckling Under Tension of a Membrane Containing Holes, PMM, 27 (1963) 275.
36 E.H. Mansfield, Tension Field Theory, Proc. 12th Inst. Cong. Appl. Mech., Springer Verlag, Berlin, 1969, 305-320.
37 E.H. Mansfield, Load Transfer via a Wrinkled Membrane, Proc. R. Soc. London, A316 (1970) 269.
38 C.H. Wu and T.R. Canfield, Wrinling in Finite Plane-Stress Theory, Q. Appl. Math. 39 (1981) 179.
39 D.A. Danielson and S. Natarajan, Tension Field Theory and the Stress in Stretched Skin, J. Biomechanics, 8 (1975) 135.
40 K. Kondo, On the General Theory of the Curved Tension Field, J. Soc. Aero. Sci. Nippon, 10 (1943).
41 K. Kondo, I. Takeshi, S. Moriguti and T. Murasaki, Tension Field Theory, Memoirs of the Unifying Study of the Basic Problems in Engineering Science by Means of Geometry, Vol. 1 Division C, 1955, 417-441.
42 C.H. Wu, The Wrinkled Axisymmetric Air Bags Made of Inextensible Membranes, J. Appl. Mech., 41 (1974) 963.
43 C.H. Wu, Nonlinear Wrinkling of Nonlinear Membranes of Revolution, J. Appl. Mech., 45 (1978) 533.
44 M. Zak, Statics of Wrinkling Films, J. Elast. 12 (1982) 51.
45 A.C. Pipkin, The Relaxed Energy Density for Isotropic Elastic Membranes, IMA J. of Appl. Math., 36 (1986) 85.
46 D.J. Steigmann, Tension Field Theory, Proc. R. Soc. London, A429 (1990) 141.
47 D.J. Steigmann and A.C. Pipkin, Axisymmetric Tension Fields, Z. Angew. Math. Phys., 40 (1989) 526.
48 D.J. Steigmann and A.C. Pipkin, Wrinkling of Pressurized Membranes, ASME J. Appl. Mech., 56 (1989) 624.
49 D.J. Roddeman, J. Drukker, C.W.J. Oomens and J.D. Janssen, The Wrinkling of Thin Membranes, Part I - Theory, Part II - Numerical Analysis, J. Appl. Mech., 54 (1987) 884.
50 R.L. Comer and S. Levy, Deflection of an Inflated Circular Cylinder Cantilever Beam, AIAA J., 1 (1963) 1652.
51 S.A. Lukasiewicz and P.G. Glockner, Stability of Lofty Air Supported Cylindrical Membranes, J. Struct. Mech. 12 (1985) 543.
52 D.J. Malcolm and P.G. Glockner, Collapse by Ponding of Air-Supported Spherical Caps, Proc. ASCE, 107 (1981) 1731.
53 G. Ahmadi and P.G. Glockner, Collapse by Ponding of Pneumatic Spherical Caps Under Distributed Loads, Can. J. Civ. Engr., 40 (1983) 740.
54 S.A. Lukasiewicz and P.G. Glockner, Collapse by Ponding of Shells, Int. J. Solids Struct., 19 (1983) 251.
55 W. Szyskowski and P.G. Glockner, Finite Deformation and Stability Behaviour of Spherical Inflatables Under Axisymmetric Concentrated Loads, Int. J. Non-Lin. Mech., 19 (1984) 489.
56 W. Szyskowski and P.G. Glockner, Finite Deformation and Stability Behavior of Spherical Inflatables Subjected to Axisymmetric Hydrostatic Loading, Int. J. Solids Struct., 20 (1984) 1021.

57 J.G.A. Croll, A Tension Field Solution for Nonlinear Circular Plates, in Aspects of the Analysis of Plate Structures, Dav[illegible] al., Eds., Clarendon Press, Oxford, 1985, 309-323.
58 A. Libai, The Transition Zone Near Wrinkles in Pulled Spherical Membranes, Int. J. Solids Struct. 26 (1990) 927.
59 L.C. Young, Lectures on the Calculus of Variations and Optimal Control Theory, W.B. Saunders, Philadelphia, 1969.
60 E.H. Mansfield, Analysis of Wrinkled Membranes with Anisotropic and Nonlinear Elastic Properties, Proc. R. Soc. London, A.353 (1977) 475.
61 E. Bromberg and J.J. Stoker, Non-Linear Theory of Curved Elastic Sheets, Quat. Appl. Math., 3 (1945) 246.
62 W. Flügge and P.M. Riplog, A Large Deformation Theory of Shell Membranes, Tech. Rep. No. 102, Div. Eng. Mech. Stanford Univ., 1956.
63 E. Reissner, The Edge Effect in Symmetric Bending of Shallow Shells of Revolution, Comm. Pure Appl. Math., 12 (1959) 385.
64 A.E. Clark and O.S. Narayanaswamy, Nonlinear Membrane Problems for Elastic Shells of Revolution, Proc. Symp. on the Theory of Shells, D. Muster, Ed., 1967, University of Houston Press, 83.
65 F.Y.M. Wan and H.J. Weinitschke, Boundary Layer Solutions for Some Nonlinear Elastic Membrane Problems, ZAMP, 38 (1987) 79.
66 P.E. Jordan, Stresses and Deformations of the Thin Walled Pressurized Torus, J. Aerospace Sci., 29 (1962) 213.
67 J.L. Sanders, Jr. and A.A. Liepins, Toroidal Membrane Under Internal Pressure, AIAA J., 1 (1963) 2105.
68 R. Benedict, A. Wineman and W.H. Wang, The Determination of Limiting Pressure in Simultaneous Elongation and Inflation of Nonlinear Elastic Tubes, Int. J. solids Struct., 15 (1979) 241.
69 A. Needlman, Necking of Pressurized Spherical Membranes, J. Mech. Phys. Solids, 24 (1976) 339.
70 V.I. Feodosev, On Equilibrium Modes of a Rubber Spherical Shell Under Internal Pressure, PMM, 32 (1968) 335.
71 A. Needleman, Inflation of Spherical Rubber Balloons, Int. J. Solids Struct., 13 (1977) 409.
72 R.T. Shield, On the Stability of Finitely Deformed Elastic Membranes, ZAMP, 23 (1972) 16.
73 A. Föppl, Vorlesungen Über Technische Mchanik, Vol. III, Teubner, Leipzig, 1907.
74 E. Reissner, On Finite Deflections of Circular Plates, Proc. Symp. Appl. Math. 1 AMS Providencee (1949) 213.
75 R.W. Dickey, The Nonlinear Circular Membrane Under a Vertical Force, Q. Appl. Math., 41, (1983) 331.
76 R.W. Dickey, Membrane Caps, Q. Appl. Math., 45 (1987) 697.
77 H.J. Weinitschke, On Finite Displacements of Circular Elastic Membranes, Math. Meth. in the Appl. Sci., 9 (1987) 76.
78 H.J. Weinitschke, Stable and Unstable Axisymmetric Solutions for Membranes of Revolution, Appl. Mech. Rev. 42, part 2, (1989) 289.
79 S.P. Joshi and L.M. Morphy, Large Axisymmetric Deformation of a Laminated Composite Membrane, ASME J. Appl. Mech., 57 (1990) 150.
80 W.W. Klingbeil and R.T. Shield, Some Numerical Investigations on Empirical Strain Energy Functions in the Large Axisymmetric Extensions of Rubber Membranes, ZAMP, 15 (1964) 609.
81 L.J. Hart-Smith and J.D.C. Crisp, Large Elastic Deformations of Thin Rubber Membranes, Int. J. Eng. Sci., 5 (1967) 1.
82 W.H. Yang and W.W. Fang, On Axisymmetrical Deformations of Nonlinear Membranes, ASME J. Appl. Mech., 37 (1970) 1002.

83 P. Pujara and T.J. Lardner, Deformations of Elastic Membranes: Effects of Different Constitutive Relations, ZAMP, 29 (1978) 315.
84 E. Schwerin, Über Spannungen und Formänderungen Kreisringförminger Membrananen, Zeit. f. Tech. Physik, 12 (1929) 651.
85 J.P. Fulton and J.G. Simmonds, Large Deformations Under Vertical Edge Loads of Annular Membranes with Various Strain Energy Densities, Int. J. Nonlinear Mech., 21 (1986) 257.
86 J.G. Simmonds and M.A. Horn, Asymptotic Analysis of the Nonlinear Equations for an Infinite Rubber-Like Slab Under an Equilibrated Vertical Line Load, J. Elasticity, 24 (1990) 105.
87 C.H. Wu, Tube to Annulus - an Exact Nonlinear Membrane Solution, Q. Appl. Math., 27 (1970) 489.
88 H. Vaughan, Pressurizing a Prestressed Membrane to Form a Paraboloid, Int. J. Eng. Sci., 18 (1980) 99.
89 A.D. Pipkin, Integration of an Equation in Membrane Theory, ZAMP, 19 (1968) 818.
90 C.H. Wu, On Certain Integrable Nonlinear Membrane Solutions, Q. Appl. Math., 28 (1970) 81.
91 W.L. Yin, Nonuniform Inflation of a Cylindrical Elastic Membrane and Direct Determination of the Strain Energy Function, J. Elasticity, 7 (1977) 265.
92 A.D. Kydoniefs and A.J.M. Spencer, Finite Axisymmetric Deformation of an Initially Cylindrical Elastic Membrane, Q. J. Mech. Appl. Math., 22 (1969) 87.
93 A.D. Kydoniefs and A.J.M. Spencer, The Finite Inflation of an Elastic Toroidal Membrane of Circular Cross Section, Int. J. Engng. Sci., 5 (1967) 367.
94 E.N. Kuznetsov, The Membrane Shell as an Underconstrained Structural System, ASME J. Appl. Mech., 56 (1989) 387.
95 E. Isaacson, The Shape of a Balloon, Comm. Pure Appl. Math., 18 (1965) 163.
96 A. Sagiv, Inflation of an Axisymmetric Membrane: Stress Analysis, ASME J. Appl. Mech., 57 (1990) 682.
97 S.S. Antman and M.C. Calderer, Asymptotic Shapes of Inflated Spheroidal Nonlinearly Elastic Shells, Math. Proc. Camb. Phil. Soc., 97 (1985) 541.
98 H.O. Foster, Very Large Deformations of Axially Symmetrical Membranes Made of Neo-Hookean Materials, Int. J. Eng. Sci., 5 (1987) 95.
99 D.A. Danielson and J.G. Simmonds, Accurate Buckling Equations for Arbitrary and Cylindrical Elastic Shells, Int. J. Eng. Sci., 7 (1969) 459.
100 A. Libai, The State of Stress Near a Longitudinal Slit in a Membrne Tube-Formulation of the Problem, Report to the Fund for the Promotion of Research at the Technion.
101 E.L. Axelrad, Theory of Flexible Shells, North Holland, 1987.

APPENDIX: SOME TOPICS IN CYLINDRICAL SHELL-MEMBRANES

(a) Nonlinear Bending of Finite Length Membrane Tubes

A circular cylindrical tube subjected to a *beam bending moment* M(x) deforms nonlinearly due to the "ovalization" of its cross sections, until it collapses or buckles locally. In very long tubes, ovalization is resisted by circumferential *shell stress-couples*, and is known as the *Brazier effect*. However, in finite-length tubes, *membranes resultants* which stem from the supports participate in this resistance and even dominate it for appropriate geometries.

The membrane tube problem can be stated in terms of a simplified mixed variational equation $\delta U^*(F,G)=0$, where:

$$U^* = \int_A\int\left[F_{,xxs}G - N_x^2/(2E_xt) - N_{xs}^2/(2E_{xs}t)\right]dA - \int_0^L M^2/(2E_x\bar{I})dx$$

with: $\theta = s/r$; $q = \theta + G_{,s}$; $\bar{y} = \int_0^s \cos q\,ds$

$$N_x = -\left[(F_{,qq} + F)q_{,s}\right]_{,s} \qquad N_{xs} = \left[(F_{,qq} + F)q_{,s}\right]_{,x}$$

$I = \pi r^3 t$; $\bar{I} = \oint t\bar{y}^2 ds$: cross-sectional moments of inertia.

Solutions can be obtained by expanding F and G as truncated Fourier series in θ, starting with m=2. The strongest nonlinearity is due to $\bar{I}$, and in most cases q may be replaced by θ in N_x and N_{xs}. It has been demonstrated before that the dominant Fourier component of the ovalization is m=2. Putting

$$F = r^2 f(x)\sin(2\theta) ; \qquad G = -rg(x)\cos(2\theta)$$

into the reduced functional and performing $\delta U^* = 0$, differential equations and boundary conditions are obtained:

$$(r^2/E_xt)f_{,xx} + (rM/2E_xI)^2(I/\bar{I})_{,g} = 0$$

$$f/E_xt - r^2(f/4E_{xs}t - g/18)_{,xx} = 0$$

$$(2g - 9f/E_{xs}t)_{,x}\delta f = 0 ; \qquad g\delta f_{,x} = 0$$

$$\bar{I} = I(1 - 2g - 8/9\ g^2 + 28/9\ g^3 + 32/75\ g^4.....)$$

Traction boundary conditions are the specification of f and $f_{,x}$(N_x and N_{xs}). The kinematic conditions are the vanishing of their coefficients: the geodesic (warping) and normal curvature changes.

Equilibrium paths (up to collapse) and local buckling curves for finite length tubes subjected to a constant bending moment and with diaphragm and clamped support conditions were presented by Libai and Bert in ICTAM '92. Good correlation with previous numerical work on local buckling was observed for membranes with $(L/r)(t/r)^{1/2} \leq 1.2$. Beyond this, circumferential shell bending effects became significant.

(b) The Decay of Edge Disturbances in Loaded Cylindrical Shells and Membranes

The addition of a *bending* term $-D\nabla^8 F$ to the DMV form of the *cylindrical membrane* equation converts it to a DMV *cylindrical shell* equation which is suitable for the qualitative analysis of edge effects and their decay into the shell. To study decay *from the curved edge*, a solution $F=Ce^{(-x/d)}\cos(ms/r)$ is put into the equation. Here, d is a characteristic decay length and m is the harmonic index. This yields a characteristic equation in d:

$$\left(\frac{1}{d^2}-\frac{m^2}{r^2}\right)^2\left(\frac{\varepsilon_x}{d^2}-\frac{m^2\varepsilon_s}{r^2}\right)-\frac{D}{Et}\left(\frac{1}{d^2}-\frac{m^2}{r^2}\right)^4-\frac{1}{d^4r^2}=0$$

Here, the first term is the *nonlinear membrane* effect (with *initial strains* $\varepsilon_x=P_x/Et$, $\varepsilon_s=P_s/Et$), the second is the *bending* effect, and the third is the *linear membrane* effect. Decay lengths are given below for some important special cases:

(1) Axisymmetric linear bending (m=0, $\varepsilon_x=0$): $d/r=0[(t/r)^{1/2}]$.

(2) Axisymmetric nonlinear membrane (m=0, D=0): $d/r=0(\varepsilon_x^{1/2})$.

(3) Slow decay, linear bending ($\varepsilon_x=\varepsilon_s=0$): $d/r=0[(r/t)^{1/2}m^{-2}]$.

(4) Slow decay, nonlinear membrane (D=0): $d/r=0(\varepsilon_s^{-1/4}m^{-3/2})$.

(5) Slow decay, nonlinear membrane (D=0, $\varepsilon_s=0$): $d/r=0(\varepsilon_x^{-1/2}m^{-1})$.

(6) Plane sheet (1/r=0, m=0): $d/t=0(\varepsilon_x^{-1/2})$.

Cases (1)-(2) typify rapidly decaying edge disturbances due, for example, to radial and rotational continuity requirement. Cases (3)-(5) represent the slow decay of circumferentially varying edge input. As is well known, linear membrane theory cannot handle the diffusion or decay of nonuniform input. For example, nonuniform loads N_x applied to the edges, retain the same nonuniformity through the membrane, even for very long shells, contrary to *St. Venant's principle*. Linear bending (case 3) or nonlinear membrane theories (case 4) provide weak mechanisms with long decay lengths. This justifies the approximation $F_{,s}>>F_{,x}$ made in these cases. Note that for thin, highly stressed shells, the membrane decay length may be considerably shorter than the linear bending length and be the dominant effect.

Case (6) represents the decay of a bending-membrane edge disturbance from a long stretched sheet. As d→0 the *ideal membrane* case is approached. This occurs as ($t^{3/2}$, E, $P^{-1/2}$)→0, and demonstrates the three main reasons for *asymptotic ideal membrane behavior*: thickness→0, rigidity→0, loads→∞.

The characteristic equation is also useful for the evaluation of the relative importance of nonlinear membrane effects (ε_x, ε_x) vs bending effects ($t^2/12r^2$) in shells. In many practical cases, the first may be large compared with the second. For example, if $\varepsilon_x=0.01$ and r/t=100, then the nonlinear membrane term would be a *thousand times* larger than the bending term. This simple example should explain why nonlinear membrane theory is important.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

On the role of wave propagation and wave breaking in atmosphere–ocean dynamics

M. E. McIntyre

Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge CB3 9EW, UK (*mem2@phx.cam.ac.uk*)

Abstract

Wave propagation, wave breaking, and the concomitant wave-induced momentum transport are ubiquitous processes in the Earth's atmosphere and oceans, a classic example being the surface gravity waves from storms in the Southern Ocean that drive mean longshore currents on California beaches. Such processes are not only interesting in themselves, but are also fundamental to making sense of the various 'wave-turbulence jigsaw puzzles' with which the atmosphere and ocean, separately and in combination, confront us. For instance, what used to be regarded as an enigmatic 'negative viscosity' of the subtropical atmosphere is now straightforwardly comprehensible in terms of Rossby wave propagation and breaking. Other examples include understanding (*a*) why the mean east–west winds in the equatorial lower stratosphere reverse every 14–18 months, throughout a belt encircling the globe, (*b*) why the lowest temperatures on earth (as low as 110 K, or −163°C) are found not in the winter hemisphere but near the summer pole, at altitudes between 80 and 90 km, (*c*) why the e-folding atmospheric lifetimes of certain man-made chlorofluorocarbons are of the order of a century, and (*d*) why the greatest concentrations of stratospheric ozone are found where photochemical ozone production rates are least.

This lecture will discuss some of the theoretical-mechanical concepts relevant to understanding these phenomena, including the concepts of wave 'momentum' and wave 'breaking'. There emerge, somewhat unexpectedly, what might prove to be some useful new ideas about the problem of water-wave generation by wind. The main point is that the water waves can be systematically amplified by certain irreversible, ratchet-like, non-superposable effects that depend on spatio-temporal inhomogeneities, such as wind gustiness and wave 'groupiness'. These include the effects of what might be called 'Rossby lee waves' in the airflow downstream of water-wave groups. The resulting wave drag can amplify non-breaking water waves and might, for instance, help to explain the growth of the 'energy front' reported by O. M. Phillips in this Proceedings.

1. INTRODUCTION

I want to widen the context of this Minisymposium and talk about some phenomena and concepts that appear fundamental to a whole range of problems in atmosphere ocean dynamics. I also want to say something about that old but still problematic topic, the generation of water waves by wind. I cannot claim to be an expert on that problem, let alone on air-sea interaction in general, but it is possible that a fresh look from another perspective might help to advance our understanding.

The atmosphere and oceans used to be thought of simply as 'turbulent' fluids on a vast scale, and early attempts to understand them often involved Reynolds averaging together with the hope that the resulting eddy-flux terms might be able to be characterized, at least roughly, in terms of the notion of 'eddy diffusivity'. It was therefore a surprise when even the signs of these fluxes sometimes turned out contrary to expectation. This was found, for example, from global-scale atmospheric 'general circulation statistics' (e.g. Lorenz 1967); and for a time the phenomenon was thought of as a kind of mysterious 'negative eddy viscosity' (Starr 1968). This meant of course that no-one understood what was going on. Eastward momentum was seen to be transported poleward by large-scale eddies, against the local mean horizontal shear in subtropical latitudes. The eddy viscosity became infinite a little further poleward. Many other such examples are now known, one of the most conspicuous being the celebrated 'quasi-biennial oscillation' (QBO) of the equatorial lower stratosphere, to be described in §4.

The mystery was gradually solved as it was recognized how important for this purpose, as well as ubiquitous, are the various wave propagation mechanisms that operate in the atmosphere and oceans, such as internal gravity wave and Rossby wave* propagation. Among the important pioneering contributions were those of Charney and Drazin (1961), Eady (unpublished, but see Green 1970), Booker and Bretherton (1967), Wallace and Holton (1968), Lindzen and Holton (1968), Dickinson (1969), and Rhines (1975). Today we have a relatively clear view of these problems, both through data from clever terrestrial and space-based observing techniques, and through a better understanding of the basic theoretical principles and of how to apply both numerical and idealized theoretical-mechanical modelling. Such modelling is used not only in hypothesis-testing thought-experiments, but also as pointing toward better ways to make observational data tell us, in a dynamically intelligible way, more about what is going on in the real atmosphere and oceans (e.g. Thorncroft *et al.* 1993, & refs.). Better understanding includes seeing what is robust about an idealized model, hence which aspects of it are likely to carry over to more realistic situations. This lecture will mention a few such models and their contribution to our present-day understanding. Also touched on will be the intimate relation between wave propagation mechanisms and shear instability mechanisms, and the concept of 'wave breaking' and its frequent relevance — when appropriately defined — to phenomena involving wave-induced momentum transport.

It is interesting to view the wind–wave problem from the conceptual vantage point thus arrived at. There emerge what may turn out to be some new ideas about wave generation mechanisms, to be discussed briefly at the end of the lecture. Besides being of general significance for our understanding of ocean waves and air–sea interaction, these ideas might help to explain, for instance, the growth of the 'short wave energy front' seen in the experiments described in Professor Phillips' Minisymposium Lecture in this Proceedings and in Chu *et al.* (1992). The key is to recognize all the wave propagation and wave breaking mechanisms that come into play, in the air as well as in the water.

2. THE MIDDLE ATMOSPHERE: SOME OBSERVED FACTS

The phenomena to be discussed, including the phenomenon of 'negative viscosity', are seen very clearly in what is now usually called the 'middle' atmosphere (but still, occasionally, the 'upper' atmosphere). It comprises the stratosphere, extending to about 50 km altitude, and the mesosphere above it, extending to somewhere between 80 and

*Rossby waves (historically, 'Kelvin–Kirchhoff–Rayleigh–Rossby waves') may also be called 'vorticity waves' or, more generally, 'potential-vorticity waves' (§6 below). I am following today's established usage.

100 km. Figure 1 includes the bulk of the middle atmosphere; it shows a typical mean January temperature distribution as a function of altitude and latitude. The units on the right-hand scale are e-folding pressure scale heights, roughly 7 km, so that the altitude range is from sea level to roughly 85 km. (Somewhere above this altitude, temperatures rise steeply into what is called the thermosphere, where simple thermodynamics and continuum mechanics break down and plasma physics becomes important.) This section and §§4.8 sketch briefly some of what is known and understood about the middle atmosphere; a more extensive discussion and bibliography can be found in a recent review of mine (1992) written for a non-specialist audience of physicists, and in other, more specialized review material cited therein; see for instance the book by Andrews, Holton and Leovy (1987).

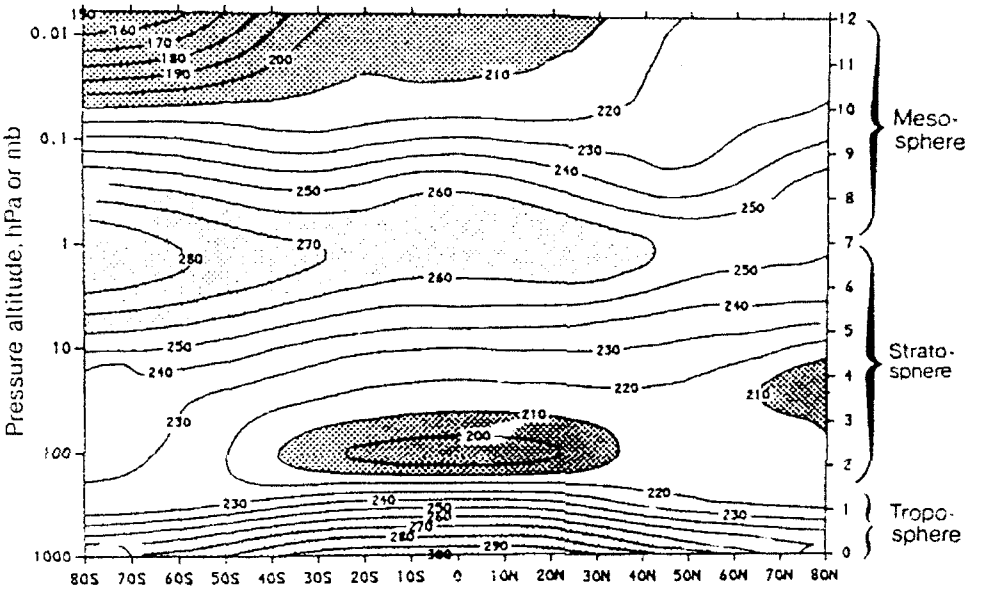


Figure 1 Temperatures T between sea level and about 85 km, for typical January conditions (degrees Kelvin, longitude and time averaged). Northward is toward the right, and the right-hand altitude scale is in e-folding pressure scale heights, roughly 7 km. From satellite and other observations analyzed by Barnett and Corney (1985).

The light shading in Figure 1 shows the warmest regions, and the dark shading the coldest, in January. Some of these features can be understood largely from considerations of radiative heat transport. For instance the high temperatures $T \gtrsim 260$ K near 7 scale heights or 50 km, defining the 'stratopause', are largely due to heating by absorption of solar ultraviolet radiation by ozone, balanced by infrared cooling to space. T increases southward at the stratopause mainly because the diurnally averaged normal solar irradiance has an absolute maximum at the south pole at the December solstice, and throughout a substantial part of December and January. The south pole is then, so to speak, the sunniest place on earth.*

A number of other features are not so simply explained. One example is the very cold region at the top left of the picture, where the lowest mean temperature plotted is $T = 150$ K. In fact individual rocket soundings have shown temperatures as low as 110 K

*For solar declination α relative to a spherical, rotating earth, the fractional length of day $\lambda(\phi)$ at latitude ϕ is $\pi^{-1} \arccos[\max\{-1, \min(1, -\tan\alpha \tan\phi)\}]$, and the diurnally averaged vertical component of solar irradiance is the full solar irradiance multiplied by $\lambda(\phi) \sin\alpha \sin\phi + \pi^{-1} \sin\{\pi\lambda(\phi)\} \cos\alpha \cos\phi$. This has an absolute maximum at the south pole when α is within 2.8° of its minimum value $-23.6°$.

on occasion. At these altitudes, the sunniest place on earth is also the coldest place on earth. Until about a decade ago, this was regarded as one of the great enigmas of atmospheric science, the enigma of the 'cold summer mesopause'. There is no doubt that the observed temperatures are far below what they should be on the basis of radiative heat transport and photochemistry alone. Temperatures substantially higher than their radiative values also occur. They occur, for instance, throughout most of the depth of the north-polar region on the right of Figure 1.

Let us note one other observed fact, at first sight having little to do with the observed pattern of temperature anomalies, but actually closely connected. This is the now-notorious fact (with its potentially serious implications for the stratospheric ozone layer) that the man-made chlorofluorocarbons known as CFC-11 and CFC-12 have very long atmospheric lifetimes, of the order of a century. These are e-folding times. Even if all leakage of these CFCs into the atmosphere could be stopped tomorrow, it would take several centuries for their concentrations to decrease to, say, 1% of present values.

How are these facts connected? The connecting link is indicated in Figure 2, which shows an estimate of the mean circulation of the stratosphere and mesosphere. It is the mean in a sense very roughly equivalent to a Lagrangian mean with suitable re-initialization of particle ensembles (see the caveats in my 1992 review and its bibliography). This mean circulation gives us a roughly correct explanation both of the temperature anomalies and of the observed lifetimes of CFCs, and the lifetimes of certain other long-lived chemical tracers. The temperature anomalies are accounted for by adiabatic expansion in the rising branches of the circulation, and adiabatic compression in the descending branches, pulling temperatures away from the radiative values toward which they would otherwise tend to relax. The rising branches carry chemical tracers upward, for example CFCs and other tropospheric tracers through the tropical stratosphere. The rate at which they are carried upward governs the rate at which they are exposed to sufficiently energetic solar ultraviolet radiation, and hence destroyed photochemically. This is the main way in which CFCs are destroyed, removal rates at sea level being relatively small. The typical strength of the mean circulation required to account for the observed temperature anomalies also gives a CFC destruction rate consistent with the observed lifetimes.

Again, the rising branch of the circulation at higher altitudes, shown schematically by the heavy dashed curve at the top left of Figure 2, explains the extraordinarily low summer mesopause temperatures. The circulation also carries small amounts of water vapour upward to the mesopause. A phenomenon observed at these altitudes, the sporadic formation of 'noctilucent' and 'polar mesospheric' clouds, requires a supply of water vapour as well as exceptionally low temperatures.* Water vapour is photochemically destroyed near the mesopause, where ultraviolet photons are even more energetic than in the stratosphere. Considerations like these, and other observational evidence, give confidence in the picture suggested by Figure 2.

The mean circulation is also part of what controls the distribution of ozone, and the rate of replenishment of the ozone layer. The photochemistry of ozone is complicated and can interact more subtly with the fluid dynamics than, for instance, CFC destruction, which depends mainly on total exposure to ultraviolet photons. Ozone can be expected to be more sensitive to processes hidden by the averaging in Figure 2; we return briefly to this point in §8.

---

*See for instance Thomas *et al.* (1989, & refs.). These incidentally are the world's highest clouds, by far, occurring at altitudes around 83–85 km. Anyone who lives between about 50° and 60° latitude can observe them as an electric blue glow above the poleward horizon, sometimes intricately patterned, on a few clear nights after midnight in the two months following midsummer.

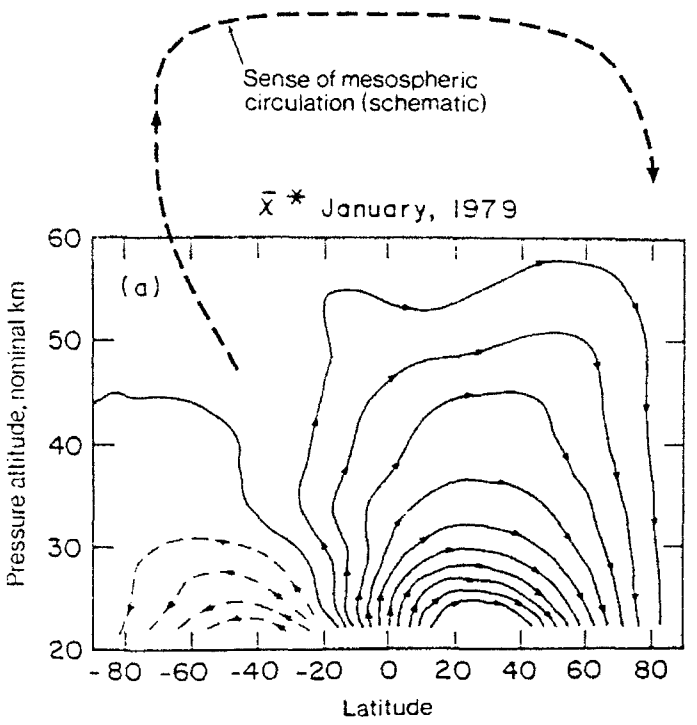


Figure 2. *Light curves*: mass transport streamlines of the longitude and time averaged global-scale circulation in an altitude range between about 3 and 9 pressure scale heights (cf. right-hand scale in Figure 1), estimated from satellite data for January 1979 by Solomon *et al.* (1986). The pressure altitude is in nominal kilometres, defined as 1/7 of a scale height. The circulation is defined in a quasi-Lagrangian sense giving a simplified, but roughly correct, indication of the vertical advective transport of chemical tracers (and is broadly consistent with direct observations of such tracers). The time for a notional fluid element to rise from the tropical tropopause to, say, 40 km is typically of the order of two years. *Heavy dashed curve*: qualitative indication of the extension of the circulation into the upper mesosphere and lower thermosphere.

How is the mean circulation driven? This again forces us to consider processes hidden by the averaging. The key point is that the Earth is a rapidly rotating planet. It is rapidly rotating in the sense that the distribution of angular momentum M per unit mass in the atmosphere is dominated by the Earth's rotation, and only weakly affected by air motion relative to the Earth. In the extratropics there is a strong latitudinal gradient of M. So a circulation like that shown in Figure 2 cannot persist unless something exerts a persistent torque on the extratropical atmosphere, in an appropriate sense (in fact, against the Earth's rotation everywhere except in the summer mesosphere, where it must be with the rotation). In a thought-experiment in which this torque is switched off, the circulation begins to die down, and temperatures to relax toward their radiative values (e.g. Haynes *et al.* 1991, & refs.).

It is now believed on good evidence that in the real middle atmosphere this torque arises from wave-induced angular momentum transport, in fact mostly from Rossby waves and internal gravity waves generated in the troposphere. The Rossby waves account for most of the lower part of circulation shown in Figure 2, and the gravity waves account for the upper mesospheric branch. The sense of the angular momentum transport is related to the phase speeds of the waves and to the different places where they are generated and dissipated, and not locally to such things as the sign of the mean shear. It is no wonder, then, that attempts to apply ideas like 'eddy viscosity' can produce incongruous results, such as negative or infinite eddy viscosities.

3. WAVE-INDUCED MOMENTUM TRANSPORT

The fact that wave propagation and diffraction are generally accompanied by a systematic flux or transport of momentum is a well-known rule in theoretical physics, applying to waves in fluids as well as to the more obvious case of waves in a vacuum, such as photons. The key phenomenon is that, if progressive waves are generated in one place A and dissipated in another B, this is accompanied by an irreversible, cumulative transport of momentum from A to B. The sense and rate of this transport is usually given, to some useful approximation, by what might be called the photon analogy, or 'quasimomentum rule' (§5).

The physical reality of such wave-induced momentum transport is easy to demonstrate in the laboratory. For instance, it manifests itself in the phenomenon known as the 'sonic wind' or 'quartz wind' (e.g. Lighthill 1978a,b), in which a piezoelectric transducer emits a beam of ultrasound that transports momentum from the transducer to locations where the waves dissipate. The resulting mean force on the fluid often generates a turbulent jet. One can demonstrate what is fundamentally the same thing without any special apparatus at all, using capillary-gravity waves at frequencies of order 5 Hz. This was done during the lecture using a small cylindrical wavemaker (Figure 3a) and a glass dish containing water with chalk dust on the surface. Making the wavemaker oscillate vertically shows that the observed mean flow (arrowed curves) is predominantly wave-induced, and not Rayleigh–Schlichting boundary-layer streaming from the surface of the wavemaker. The latter has the opposite sense (e.g. Lighthill 1978a, p.348; Van Dyke 1982, Fig. 31). Carefully stopping the wavemaker and observing the persistence of the mean flow shows that it is more than a 'Stokes drift': irreversible, cumulative momentum transport has indeed taken place. With a larger area of water, one can use a curved wavemaker to focus the waves on a more distant spot (Figure 3b), illustrating the fact that the wave generation and dissipation sites can be well separated spatially.

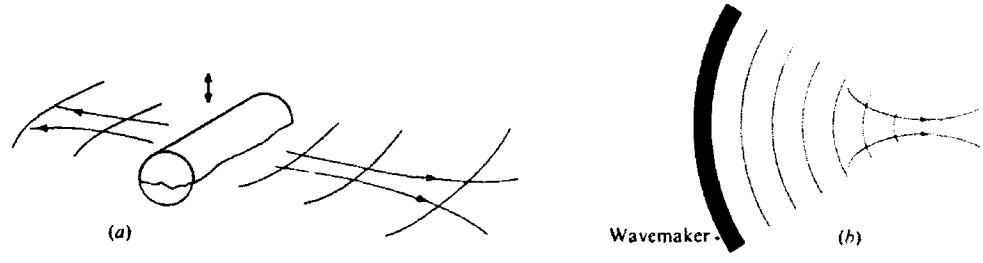


Figure 3. Simple demonstrations of momentum transport by 5 Hz surface capillary-gravity waves (see text). Configuration (a) works well with a wavemaker whose diameter ~ 3 to 4 cm, and configuration (b) with a wavemaker whose radius of curvature $\gtrsim 50$ cm. From McIntyre and Norton (1990).

This last point is underlined by the classic work of Munk *et al.* (1963) in which, following earlier work by Barber and Ursell, they demonstrated that surface gravity waves generated by storms in the Southern Ocean can, and often do, propagate across the Pacific all the way to beaches in California, where they break and generate longshore mean currents (see also Snodgrass *et al.* 1966). This is a clear case of irreversible wave-induced momentum transport over many thousands of kilometres! The internal gravity waves that sustain the mesospheric circulation of Figure 2 also dissipate mostly by breaking. They have sources mostly in the denser layers of the atmosphere far below, so that horizontal momentum is transported over vertical distances not of thousands, but certainly many tens, of kilometres (e.g. Fritts 1984, 1993).

Essentially the same things happen with Rossby waves, except that Rossby-wave dynamics has a peculiar 'one-signedness' that can constrain the sense of the momentum

transport in a ratchet-like way. As well as being interesting for this reason, the case of Rossby waves is arguably the simplest to understand from first principles. This will be seen in §7. First, however, I want to mention an example involving internal gravity and other kinds of waves, in which the wave-induced momentum transport interacts with wave refraction to produce an interesting feedback oscillation of the *mean flow* that has been observed both in the laboratory and in the real atmosphere — the celebrated 'quasi-biennial oscillation' or 'QBO' — and then I want to touch briefly on the photon analogy and on what is sometimes called wave 'momentum'.

4. THE QBO AND ITS LABORATORY ANALOGUE

The 'laboratory QBO' was demonstrated in a famous experiment by Plumb and McEwan (1978). The system used was a salt-stratified fluid contained in a large laboratory annulus, of depth 50 cm and gap width 12 cm. Internal gravity waves are excited by making a flexible lower boundary oscillate in a standing wave, equivalent to equal amplitudes of clockwise and anticlockwise progressive waves. The response of the fluid breaks this symmetry (the annulus is at rest relative to the laboratory, and the Earth's rotation can be neglected); and a wave-induced mean flow arises, horizontally around the annulus. Soon a regime is established displaying a characteristic spacetime signature, in which the mean flow reverses periodically at a given altitude, and does so earlier at higher altitudes. This was illustrated by a movie of the original experiment shown in the lecture.

The two mechanisms involved are first the wave-induced angular momentum transport, cumulatively changing the mean velocity profile as the waves dissipate (viscously in this case), and second the effect of mean shear in Doppler shifting and refracting the waves (somewhat like the selective surface-wave refraction that can make a bathtub vortex appear, at first glance, to be rotating the wrong way). In the Plumb McEwan experiment the wave dissipation rate is least, and the vertical group velocity greatest, when the waves propagate against the mean flow. Such waves therefore penetrate highest. They transport angular momentum in the same sense as their intrinsic angular phase speed and can therefore reverse the mean motion where they dissipate, first at higher and then at lower and lower altitudes in the annulus. Mathematical models that express these ideas tend to behave non-chaotically, and to produce the spacetime signature very robustly (e.g. Yoden and Holton 1988; Haynes *et al.* 1993), as long as waves of sufficient amplitude are excited in both senses, clockwise and anticlockwise.

The same spacetime signature (albeit not exactly periodic) is conspicuously present in the mean east–west winds of the tropical lower stratosphere, throughout a belt encircling the globe, at altitudes between about 16 and 35 km and latitudes between about $\pm 15°$. Whereas in the laboratory it typically takes about half an hour for the mean flow to reverse, in the real stratosphere it takes about 14 ± 3 months. The wind reversals are clearly seen in the tropical radiosonde data that became routinely available from the early 1950s onwards; and there is indirect evidence for their existence many decades before then (Hamilton and Garcia 1984; Teitelbaum *et al.* 1993). Although our detailed understanding is incomplete, it seems overwhelmingly likely that the resemblance to the laboratory experiment and to mathematical models of it is not accidental, and that the mean flow changes are wave-driven in much the same way. Indeed, before wave driving was thought of at all in this context — the observed QBO used to be another great enigma of atmospheric science — a strong case had already been made for the existence of some strange kind of eddy or fluctuation-induced angular momentum transport, again entailing 'negative eddy viscosity'. Without some such eddy transport, it seemed impossible to make dynamical sense of the observed mean flow changes, on

the basis of careful physical arguments and numerical experiments (Wallace and Holton 1968). It was only after this that the realization came (Lindzen and Holton 1968) that wave-induced angular momentum transport could behave in just the required manner.

We do not yet have a convincing *quantitative* model of the real QBO including, for instance, the effect of the mean upwelling illustrated in Figure 2. This advects angular momentum upward on a comparable timescale, and we lack sufficient quantitative knowledge of its strength and seasonal variation. There is also uncertainty over exactly which wave types are significant in the tropical stratosphere. It used to be assumed that the principal such wave types are the equatorially trapped Kelvin and Rossby–gravity waves (e.g. Andrews *et al.* 1987, Gill 1982). There is observational support for this in the case of the Kelvin wave, whose observed amplitude is not far from having an appropriate order of magnitude. But it has seemed more and more likely, for instance, that planetary-scale Rossby waves originating in the extratropical troposphere are more significant than equatorial Rossby–gravity waves — meaning locally significant in the tropics (e.g. Dickinson 1968; Lindzen and Tsay 1975; Andrews and McIntyre 1976; Dunkerton 1983; Takahashi and Boville 1992; O'Sullivan and Hitchman 1992, see §8 below) as well as indirectly significant through the strength of the mean upwelling.

We have even less knowledge, either observational or theoretical, of possible wave generation mechanisms. The real QBO seems to involve highly complicated, chaotic, nonlinear wave generation processes (mainly in the troposphere, both tropical and extratropical) to which there is a robustly non-chaotic response in the tropical stratosphere — with, it might be added, potentially important consequences for long-range weather forecasting, such as a feedback on the depth of cumulonimbus convection and hence on tropical cyclone intensities and El Niño timings (Gray *et al.* 1992). These aspects remain a challenge and opportunity for the future.

5. LONG-DISTANCE TRANSPORT AND THE PHOTON ANALOGY

The 'photon analogy' or 'quasimomentum rule' says that the rate at which momentum is transported from location A to location B, when a wave packet is generated at A and dissipated at B, is the same as if

(*a*) the fluid were absent, and
(*b*) the wave packet had a certain amount of momentum **q** that it carries around with it, like a photon in a vacuum.

The quantity **q** is not actually a momentum. It is a property of the wave packet that may more aptly be called its *quasimomentum* or *pseudomomentum*, in order to distinguish it from momentum. Momentum, in the presence of a material medium, is a fundamentally different quantity.* Hence the alternative term 'quasimomentum rule'. Indeed a better, more precise statement is that "the rate... is the same as if

(*a*) the fluid were absent, and
(*b*) the wave packet had momentum equal to its quasimomentum."

The quasimomentum is a wave property in the sense that it can be evaluated from linearized wave theory alone. In the simplest theoretical approximations (slow modulation

*This is because its conservation corresponds to a different translational symmetry (of the propagating medium rather than of the total physics). There is no widely agreed name for **q**. Other names used with different kinds of waves include 'wave-vector', 'Poynting's momentum', 'Minkowski momentum', 'radiation momentum', 'acoustic momentum', 'crystal momentum', 'phonon momentum', 'tensor momentum', 'field momentum', 'canonical momentum', 'wave momentum', and 'momentum'. This has led to some confusion. For the basics plus some history going back to the time of Rayleigh, Poynting, Abraham, and Minkowski, and for keys to the literature, the reader may consult my 1981 essay and its bibliography, also Loudon and Paige (1991, p.236), Peierls (1991 §§2.4–6), and Shepherd (1990, §5).

as well as small amplitude), we may take $\mathbf{q} \approx E\mathbf{k}/\omega$ where E is the intrinsic wave energy in the sense discussed e.g. by Bretherton and Garrett (1968), $\mathbf{k}$ is the wavenumber vector, and ω is the intrinsic frequency, or frequency Doppler shifted to a reference frame moving with the local mean flow. Like E, $\mathbf{q}$ is $O(a^2)$ in wave amplitude a.

The analogy summarizes a body of special and general results from theories in which a is considered small and in which wave induced momentum transport, and all the associated mean effects, are self-consistently described correct to $O(a^2)$. As is well known the theoretical calculations can be elaborate and tricky, equatorial Rossby–gravity waves being a case in point (Andrews and McIntyre 1978a §9, & refs.). Such theories are often qualitatively applicable, and may also be quantitatively applicable to important parts of the problem, such as the part concerned with wave-induced momentum transport between the sites of wave generation and dissipation. Amplitudes in the intervening wave field can in some cases be truly small (in the relevant sense, measured by wave slopes), as for instance in the case of surface gravity waves crossing the Pacific. Some aspects of the theory and the photon analogy extend to finite amplitude as well (Andrews and McIntyre 1978a,b). The analogy has relevance to all the situations and all the wave types mentioned in this lecture, including the equatorial Kelvin and Rossby gravity waves that are thought to contribute to the real QBO.

But wait, I hear someone say, why all this hair-splitting about an 'analogy' — why the 'as if' and the 'quasi', to say nothing of the 'pseudo'? Isn't this complicating things unnecessarily? Surely wave packets in a fluid really do have momentum $\mathbf{q}$, which they really do carry around with them, just like photons in a vacuum. Is it not well known that dissipating waves exchange 'their momentum' (meaning $\mathbf{q}$) with the mean flow? How could they do such a thing if they didn't really have momentum $\mathbf{q}$ to exchange? Besides, how else could those waves that propagate across the Pacific, after generation by storms in the far south, drive *longshore* currents on northern beaches? In that situation we do not have a steady wavemaker, and we do not have a steady 'radiation stress' spanning the whole ocean for weeks on end. So surely wave packets, and finite wavetrains, must just carry momentum with them. And what about the mesospheric circulation and the noctilucent clouds? Real internal gravity waves are highly intermittent; and isolated wave packets are again, arguably, a more relevant idealization than steady waves spanning the whole depth of the middle atmosphere.

Well, what really happens is interesting, and worth a brief digression. Take for instance an idealized version of the situation in mid-Pacific. Figure 4, from my 1981 essay, shows an isolated, non-dissipating, two-dimensional packet of surface gravity waves on deep water. More precisely, it shows semi-schematically the leading-order theoretical solution describing the disturbance and its accompanying $O(a^2)$ velocity field, derivable from the work of Longuet-Higgins and Stewart (1962). That solution can be used to compute the total momentum of the propagating disturbance. In order for the photon idea to be literally true, in the manner just envisaged, it would be necessary for that momentum to be well defined and equal to $\mathbf{q}$. It turns out that in this case the momentum is indeed well defined. But a careful computation shows what Longuet-Higgins and Stewart also found, namely that the momentum is *not* equal to $\mathbf{q}$ (see figure caption). To leading order, it is zero!

This is not a paradox. Rather, it is one of many counterexamples showing that, for waves in material media, the photon idea cannot be taken literally as a general principle. In order for the photon idea to make sense in general — in fact to make sense outside a very limited set of circumstances — it is indeed crucial to regard it as an analogy, i.e., to retain the words 'as if' preceding items (*a*) and (*b*) above, and to continue to recognize the distinction between momentum and quasimomentum. Thus stated, the analogy is both useful, and capable of general theoretical justification. One approach is

to use 'generalized Lagrangian means' in conjunction with Kelvin's circulation theorem (Andrews and McIntyre 1978a); the importance of the circulation theorem in this kind of problem was recognized by Rayleigh (1896), and its connection with the photon analogy was, I think, first recognized by Bretherton (1971). Additional considerations that have improved our general understanding, but have yet to be fully worked out, can be found in my paper with Norton (1990). But, to return to the idealized 'Pacific' problem, how then *does* the whole thing work fluid-mechanically?

The main point is this. As a wave packet propagates past any given fluid element between its generation and dissipation sites, it gives rise to an $O(a^2)$ mean forcing *whose time integral is zero* for that fluid element. Details are complicated but the most significant aspect of this forcing can be thought of, for present purposes, as coming from the divergence of a radiation stress spanning the region occupied by the wave packet (and satisfying Newton's third law of 'action and reaction'). This causes the given fluid element first to feel a mean push, and then a mean pull, against other fluid elements in the region. Consequently, the time-integrated force on the given fluid element is zero. Corresponding statements are true of the other aspects of the $O(a^2)$ mean forcing, such as apparent mass sources and sinks (e.g. Andrews and McIntyre 1978a).

Computing the $O(a^2)$ response to a given $O(a^2)$ mean forcing is a linear problem. Therefore, during a time interval in which one or more wave packets propagate from A to B, one can regard the $O(a^2)$ mean forcing as the sum of two contributions: first a steady forcing, corresponding to a steady wavetrain and its radiation stress spanning the entire region between A and B — and conforming to the photon analogy — and second an oscillatory forcing in the same region whose time integral vanishes everywhere.

It is only the first of these two contributions that is interesting from the present viewpoint, i.e. that corresponds to the notion of a cumulative, irreversible wave-induced transport of momentum from A to B. The second, oscillatory contribution evokes a response that is non-cumulative, because of the vanishing of its time integral. It tends moreover to be strongly dependent on circumstances such as conditions at remote boundaries, and how the waves were generated. Its details can be complicated. For instance if one were to include stable stratification and Coriolis effects in a less idealized model of the Pacific ocean, then the $O(a^2)$ response to the passage of a non-dissipating wave packet would be quite different from that shown in Figure 4. It would involve the excitation of very weak $O(a^2)$ internal Coriolis–gravity (inertio–gravity) waves over a large area of ocean. In fact something similar happens even in the special case of Figure 4 (see my 1981 essay for further discussion) since in general there are very distant, very weak, fast-propagating, ultra-long $O(a^2)$ surface gravity waves, which I have not attempted to depict in the figure but which embody significant amounts of momentum, and which depend on how the wave packet was generated.

In summary, then, what is complicated, and circumstance-dependent, is the detailed, unsteady $O(a^2)$ mean response of the fluid medium to the generation, propagation and dissipation of a wave packet, or of many wave packets. What is simple, and general, is the fact that the time integral of this $O(a^2)$ response is zero apart from the cumulative contribution given by the photon analogy. This is true, in a wide range of problems of this kind, whether or not any well-determined, well-localized $O(a^2)$ mean momentum appears temporarily within the fluid as wave packets propagate past, or whether for instance $O(a^2)$ mean momentum is temporarily taken up by distant boundaries, as is sometimes signalled by divergent momentum integrals in idealized versions of the problem — or whether $O(a^2)$ mean momentum is taken up by distant long-wave disturbances, as in the problem of Figure 4 and its variants.

Another thought-experiment of fundamental interest in this connection is to scatter the wave packet of Figure 4 from an immersed obstacle, and ask what the mean recoil

force is. One finds that it is given by the photon analogy, appropriately re-stated. This again is 'despite' the fact that, in this case, the wave packet has a well defined momentum equal to zero. What happens is that more very weak, fast propagating, ultra long wave $O(a^2)$ disturbances, of the kind already referred to, are radiated during the reflection process. One can derive a general result, comprising a non-trivial extension of Noether's theorem, that shows why the net effect of all these $O(a^2)$ phenomena must be given by the photon analogy in many such cases including this one. The result also shows why there are some 'exceptional' problems for which the analogy fails.*

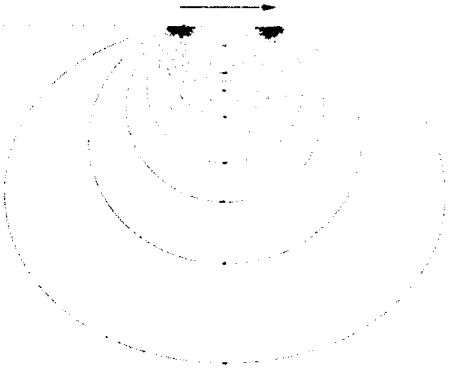


Figure 4. Packet of surface gravity waves propagating toward the right in deep water, and its accompanying $O(a^2)$ velocity field plotted quantitatively except that the Stokes drift (near the surface) is not depicted. The total momentum of the wave packet is well defined, and comprises the momentum of the Stokes drift, which, for these particular waves, equals $\mathbf{q}$, plus the momentum of the return flow underneath, which equals $-\mathbf{q}$ (because return-flow acceleration reactions feel the free surface as effectively rigid — for further discussion see McIntyre 1981).

6. ROSSBY WAVES, VORTICES AND SHEAR INSTABILITIES

Of all the examples of irreversible wave-induced momentum transport, some of those associated with Rossby waves are arguably the simplest as well as among the most important. Rossby waves and related phenomena are ubiquitous in the atmosphere and oceans, and are fundamental to almost every aspect of large-scale atmosphere ocean dynamics. For example, it is Rossby waves and related phenomena that drive most of the mean circulation illustrated in the lower part of Figure 2. The 'Rossby-wave elasticity' to which the wave propagation owes its existence is important also, for instance, in strongly inhibiting the turbulent transport of chemicals into the Antarctic ozone hole, or of drier air into the moist eye wall of a tropical cyclone. A still wider interpretation of 'related phenomena', meaning phenomena depending on 'Rossby elasticity', would

*The key step is to consider a certain translational symmetry operation giving a conservation law for the sum of the quasimomentum of the waves and the momentum of the immersed obstacle. There are 'exceptional' cases because this translational symmetry operation is singular. The singularity is strong enough, in some cases, to invalidate the extension of Noether's theorem. Details will be given in a forthcoming paper with S. D. Mobbs (1993). I discovered one of these exceptions by chance some time ago, via some very careful $O(a^2)$ calculations (McIntyre 1972, 1973). The existence of such exceptions is a telling confirmation that the photon analogy is only an analogy, highly useful when valid but not, in general, to be taken literally.

include many important types of *shear instabilities* all the way from ordinary small-scale shear instabilities to the large-scale, buoyancy-powered 'baroclinic instabilities' that can lead to the formation of common types of atmospheric and oceanic eddies and vortices, including extratropical weather cyclones (Hoskins *et al.* 1985, & refs.).

Rossby waves and related phenomena occur in dynamical systems of the generic form

$$DQ/Dt = 0\ , \qquad \mathbf{u} = \mathbf{I}(Q)\ , \tag{6.1a,b}$$

where $D/Dt = \partial_t + \mathbf{u}\cdot\nabla = \partial_t + u\partial_x + v\partial_y$, the two-dimensional material derivative, and where $\mathbf{I}(\cdot)$ is a time-independent functional of the materially conserved scalar field Q. The simplest case is the familiar case of two-dimensional inviscid, incompressible vortex dynamics, for which

$$\mathbf{I}(Q) = (-\partial_y,\ \partial_x)\nabla^{-2}Q_{\mathrm{r}} \tag{6.2}$$

where $Q_{\mathrm{r}} = Q = v_x - u_y$, the ordinary vorticity and, to make the inverse Laplacian unambiguous, suitable boundary conditions are understood such as evanescence of $|\mathbf{u}|$ at infinity. Note that (6.2) implies $\nabla\cdot\mathbf{u} = u_x + v_y = 0$. The next simplest case is the same thing on a rotating earth whose vertical component of absolute vorticity is Q_{e}, say, a prescribed function of horizontal position but not of time. Then, retaining the notation $\mathbf{u} = (u, v)$ for the relative velocity and $Q_{\mathrm{r}} = v_x - u_y$ for the relative vorticity, we may take $Q = Q_{\mathrm{e}} + Q_{\mathrm{r}}$ in (6.1) and retain (6.2) unchanged, remembering that Q_{e} is prescribed. The notation in (6.1) has been chosen to emphasize the fact that the single scalar field Q contains all the dynamical information. At every instant, the Q field can be 'inverted' to recover the velocity field $\mathbf{u}$. One may call $\mathbf{I}(\cdot)$ the 'inversion functional'.

In more realistic models of the rotating, stratification-constrained, vortical flows that occur in the real atmosphere and oceans, the same generic mathematical structure (6.1) applies, in many cases to remarkable accuracy. This is why simple two-dimensional vortex dynamics has always been such an important idealization in the context of atmosphere–ocean dynamics. The coordinates x and y now measure horizontally-projected distances along the (approximately horizontal) stratification surfaces. D/Dt and $\mathbf{u} = (u, v, 0)$ are still two-dimensional on each such surface, and $u_x + v_y$ is still 'zero' to some useful approximation (more precisely, has typical magnitudes $\ll$ typical magnitudes of $v_x - u_y$). What is new is that the inversion functional $\mathbf{I}(\cdot)$ is now three-dimensional. Away from the equator, it still has the qualitative character of (6.2) but with ∇^{-2} more like a three-dimensional inverse Laplacian, in coordinates vertically stretched by Prandtl's ratio, the ratio of the buoyancy to Coriolis frequencies. Distortions of the stratification surfaces are also determined as part of the inversion operation. One may generally characterize such stratified, rotating flows as approximately 'layerwise two-dimensional'. In the most accurate models, which include models whose validity extends into the tropics, Q is the Rossby–Ertel potential vorticity, and inversion is no longer a linear operation (e.g. Hoskins *et al.* 1985, & refs; Thorpe 1985; Davis 1992; Raymond 1992).

One has here, incidentally, a framework for the general characterization of coherent structures such as vortices and vortex pairs, in atmosphere–ocean dynamics – amounting to variations on a theme from Professor Roshko's Opening Lecture in this Proceedings. For instance a 'vortex', in any dynamical system of the form (6.1), is the coherent structure represented by a strong, isolated anomaly in the Q field together with the induced velocity and any other relevant fields in and around it, where 'induced' means given by whatever inversion operator $\mathbf{I}(\cdot)$ characterizes that dynamical system. Atmospheric cyclones and anticyclones, and oceanic Gulf Stream rings and 'Meddies' (e.g. Armi *et al.* 1988), are all cases in point. Figure 5 illustrates one such structure in a model atmosphere, somewhat idealized but instantly recognizable to any meteorologist

familiar with large-scale atmospheric behaviour. It is a cyclonic (earth-co-rotating) extratropical vortex induced by a strong, compact anomaly in the Rossby–Ertel potential vorticity field near the tropopause. Coherent structures such as these are often important for weather developments, in which fast advection of potential-vorticity anomalies near the tropopause is an important aspect of the dynamics, approximately satisfying eq. (6.1a) over timescales of several days.

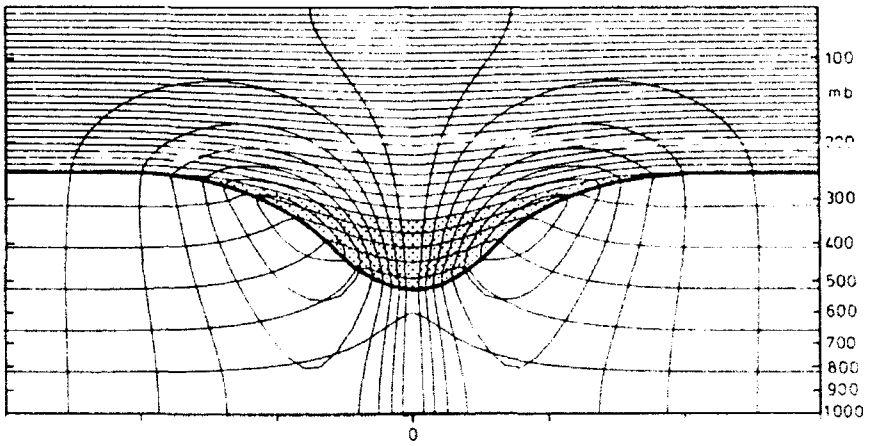


Figure 5. Section across the axisymmetric structure induced by an isolated, axisymmetric, cyclonic potential-vorticity anomaly (stippled region) near a model tropopause (heavy curve) across which the Rossby–Ertel potential vorticity has a strong discontinuity, by a factor of 6. The family of thin curves some of which are closed are isopleths of tangential velocity, at $3\,\mathrm{ms}^{-1}$ intervals — the greatest velocities $> 21\,\mathrm{ms}^{-1}$ being at the tropopause — and the other family of thin curves, more nearly horizontal, are the stratification surfaces. These are isopleths of potential temperature θ defined to coincide with actual temperature at pressure 1000 mb or 1000 hPa, plotted at 5 K intervals. 'Anomaly' means a potential-vorticity contrast on a θ surface (Hoskins *et al.* 1985, eq. 29). The induced surface-pressure minimum is 41 hPa below ambient. The structure is typical for middle latitudes; the Coriolis parameter is $10^{-4}\,\mathrm{s}^{-1}$ (as at latitude 43.3°N). The domain shown has a radius of 2500km. From the work of Thorpe (1985).

Now to 'Rossby-wave elasticity'. The simplest example occurs in the strictly two-dimensional dynamical system specified by (6.2) with $Q_\mathrm{r} = Q - Q_\mathrm{e} = v_x - u_y$, and Q_e a linear function of y so that when $\mathbf{u} \equiv 0$ we have

$$\partial Q/\partial y = \partial Q_\mathrm{e}/\partial y = \beta = \text{constant} \,. \tag{6.3}$$

This is Rossby's famous 'β-plane' or 'nearly flat earth' model, with (x, y) taken to be Cartesian. Then (6.1) is satisfied — in this case without linearization, as it happens — by expressions of the form $\mathbf{u} = (-\partial_y,\ \partial_x)\psi(x, y, t)$ $(Q_\mathrm{r} = \nabla^2\psi)$ with $\psi(x, y, t) \propto \cos ly \cos\{k(x - ct)\}$, provided that the phase speed c and the wavenumber components k, l satisfy the dispersion relation

$$c = -\beta/(k^2 + l^2) \,. \tag{6.4}$$

The one-signedness previously referred to shows up here: (6.4) is a formula for c, and not c^2 as in classical small-vibration problems in non-rotating reference frames. We can see the reasons for the one-signedness and appreciate its robustness as follows.

By (6.1), the contours of constant Q are also material contours. If a disturbance makes these contours undulate as suggested in Figure 6, then Q_r in (6.2) will be alternately positive and negative as indicated by the plus and minus signs in Figure 6. Then since $\psi = \nabla^{-2}Q_\mathrm{r}$ the contours of ψ can be pictured as the equipotentials of the electrostatic field due to a pattern of alternating positive and negative charges (with the

sign changed), or as the topographical contours giving the displacement of a stretched elastic membrane that is pulled up (−) and pushed down (+) alternately in the same pattern. Hence ψ will have hills and valleys centred respectively on the minus and the plus signs, implying that the strongest north-south velocities (at right angles to the electric field in the electrostatic analogy) will occur at intermediate positions, a quarter wavelength out of phase with the displacement, and in the sense shown by the heavy, dashed arrows in Figure 6. If one now makes a moving picture in one's mind's eye of what this induced velocity field will do to the material contours, one can see at once that the behaviour must be oscillatory, and also such that the undulations propagate from right to left only.

The same physical picture exhibiting one-way propagation applies when $Q_e = 0$ but Q has a background gradient $\partial Q_0/\partial y$ due to a mean shear flow $\mathbf{u} = \{u_0(y), 0\}$,

$$\partial Q_0/\partial y = -\partial^2 u_0/\partial y^2 = -u_{0yy} \tag{6.5}$$

so that $Q = Q_0(y) + Q'(x, y, t)$ and $\psi = -\int^y u_0(\tilde{y})\, d\tilde{y} + \psi'(x, y, t)$ and, for small disturbances,

$$(\partial_t + u_0\partial_x)Q' - u_{0yy}\psi'_x = 0\ , \qquad \psi' = \nabla^{-2}Q'\ , \tag{6.6}$$

essentially the Rayleigh equation. It is no surprise, therefore, to find that the simplest classical shear instabilities, such as the instability of the $u_0 = \tanh y$ shear layer, or its Rayleigh (piecewise linear) counterpart, can be understood (consistently with the Rayleigh, Fjørtoft and Arnol'd stability theorems) in terms of a coupled pair of Rossby waves, in a certain frame of reference, each of which propagates against the local mean flow and phase-locks with the other in such a way as to bring it to rest. Furthermore, each Rossby wave makes the other grow exponentially, via a reduced phase shift between disturbance velocities v' and the sideways displacements η' of the Q contours (Lighthill 1963, Bretherton 1966, Hoskins *et al.* 1985 §6b). The Miles (1957) wind-wave instability is in some ways fundamentally similar, except that one of the Rossby waves is replaced by the surface gravity wave, and there is a mismatch between the strengths of the gravity and Rossby elasticities leading among other things to relatively slow growth.

7. ROSSBY-WAVE BREAKING: A DEFINITIVE EXAMPLE

In Figure 6 the Q (material) contours are depicted as undulating reversibly, the situation described by linearized, dissipationless wave theory. Indeed, for linearized theory to be self-consistently and generally applicable, along with associated concepts such as the principle of superposition, the undulations must also be gentle. Strictly speaking, the sideways slopes $(\partial y/\partial x)_Q$ must be infinitesimal. The opposite extreme, that of infinite sideways slopes followed by sideways overturning, and rapid, irreversible deformation, lengthening and folding of the Q contours, is a commonplace occurrence and can be recognized as a Rossby-wave version of 'wave breaking'. An idealized example is shown in Figure 7; see also Figure 9 below. This phenomenon, in a variety of forms, is ubiquitous in the real atmosphere and oceans and plays an important role in the irreversible transport of momentum and angular momentum by Rossby waves.

The case of Figure 7, known as the Stewartson–Warn–Warn (SWW) Rossby-wave critical-layer solution, is now described in more detail. It is important out of all proportion to the restrictive idealizations used because it is an unequivocal example, described by an analytical solution, of irreversible wave-induced angular momentum transport due *solely* to wave breaking, with no other wave dissipation mechanisms involved.

We return to the dynamical system described by (6.1)–(6.3), but now introduce an undisturbed flow $u_0(y)$ having constant shear u_{0y}, say. Again the Q contours are made to

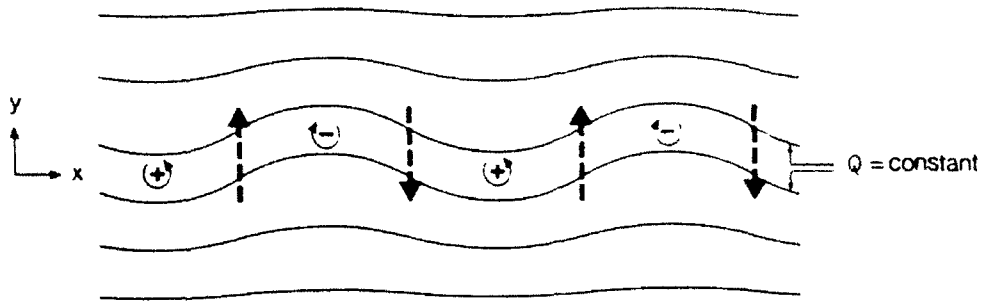


Figure 6. Sketch of the Q contours and the Q-anomaly (Q') pattern, and the induced velocity field giving rise to the sideways 'Rossby elasticity' in a simple, non-breaking Rossby wave.

undulate, this time by introducing a gently undulating boundary near some value y_0 of y, exciting Rossby waves of constant phase speed c and y-lengthscale u_{0y}/β. The boundary displacement amplitude is of order $\epsilon\, u_{0y}/\beta$ where ϵ is a small dimensionless parameter. The x-wavelength $2\pi/k$ of the undulation is assumed long enough, $\gg 2\pi\, u_{0y}/\beta$, for the inverse Laplacian ∇^{-2} in (6.2) to be approximated by $(\partial/\partial y)^{-2}$, simplifying the mathematics. (The lengthscale u_{0y}/β then corresponds to the lengthscale l^{-1} in (6.4) when $c \simeq l^{-1} u_{0y}$.) Under these restrictions and with suitable choices of y_0, this problem can be solved analytically (Stewartson 1978, Warn and Warn 1978). The Q contours behave in an approximately undular manner except in a narrow region surrounding the critical line, or y-location where $u_0(y) = c$. This narrow region, the 'critical layer', is separated from the undulating boundary by a distance $\gtrsim u_{0y}/\beta$ and has width of order

$$b = \epsilon^{1/2}\, u_{0y}/\beta\ . \tag{7.1}$$

The 'inner problem' for this region is mathematically the same as the nonlinear pendulum problem, incompressible fluid flow replacing incompressible phase-space flow. It is analytically soluble in terms of elliptic functions, and the solution confirms that the region is, indeed, a region of Rossby wave breaking in the sense envisaged. Figure 7 shows the solution at four successive times. Q contours are overturning sideways, and deforming in a manifestly irreversible way.

Let us suppose that the resulting rearrangement of the Q field has, in a coarse-grain view, something of the character of a mixing layer as shown in idealized form in Figure 8a. Then it is a trivial matter to see that there must be an associated irreversible transport of momentum. Let $\delta\bar{Q}(y)$ be the change in Q represented by the difference between the solid and dashed lines in Figure 8a, and $\delta\bar{u}(y)$ the corresponding change in u. Application of the inversion operator (6.2) gives (since $\delta\bar{Q} = \delta\bar{Q}_r = -\delta\bar{u}_y$ here) the parabolic profile

$$\delta\bar{u}(y) = -\int_{-\infty}^{y} \delta\bar{Q}(\tilde{y})\, d\tilde{y} = \begin{cases} \frac{1}{2}\beta(y^2 - \frac{1}{4}{b_m}^2) & (-\frac{1}{2}b_m < y < \frac{1}{2}b_m) \\ 0 & (y < -\frac{1}{2}b_m \text{ or } y > \frac{1}{2}b_m) \end{cases} \tag{7.2}$$

where b_m is the breadth of the mixing layer. Negative momentum

$$\delta M \propto \int_{-\infty}^{\infty} \delta\bar{u}(y)\, dy = -\tfrac{1}{12}\beta {b_m}^3 \tag{7.3}$$

has been transported irreversibly into the region where Q has been rearranged. This phenomenon is robust: any change $\delta\bar{Q}(y)$ qualitatively like that in Figure 8a will be associated with a momentum change of the order of magnitude, and sign, indicated by (7.3). For instance the SWW solution has a limiting value of δM as $t \to \infty$, which can be expressed in the form (7.3) with the value of b_m shown by the bar at the centre

of Figure 7d. This value of b_m may be thought of as an 'effective mixing width' for the SWW solution. In this case the momentum (7.3) has come from the wave source comprising the undulating boundary at large Y. When β, or more generally the initial gradient $\partial(Q_e + Q_0)/\partial y$, is positive, this momentum is necessarily negative — another manifestation of the one-signedness of Rossby dynamics.

The SWW solution provides us, incidentally, with a definitive counterexample to arguments saying that Q-mixing scenarios like that of Figure 8a are impossible because they violate momentum conservation. These arguments overlook the possibility of wave-induced momentum transport from outside the region. Further historical remarks and references, going back to various issues surrounding, for instance, G.I. Taylor's 'vorticity transfer theory', and its recent developments including links with the photon analogy, can be found in my 1992 review and in the paper with Norton (1990).

Another interesting point about Figure 7 is that the predicted Q configuration becomes shear-unstable after the contours first overturn. So if any noise is present initially, the actual evolution is quite different in detail. In typical cases the result is an apparently chaotic form of Rossby wave breaking, and an increase in the effective mixing width b_m by a modest factor such that δM ($\propto {b_m}^3$) increases in magnitude by a factor 2 or 3. Figures 8b,c, from the definitive study by Haynes (1989), show the Q and $\delta\bar{u}$ profiles, defined as Eulerian x-averages, in one such case. The $\delta\bar{u}$ profile has the approximately parabolic form suggested by the idealization (7.2).

The foregoing examples are conceptually important in another way, already hinted at. The inviscid, two-dimensional fluid-dynamical system under consideration is known to have mathematically regular behaviour, over arbitrarily long time intervals. The examples are therefore cases of wave dissipation and irreversible momentum transport that do not depend on overtly dissipative processes like viscosity. The irreversibility involved is a purely fluid-dynamical irreversibility, precisely that associated with the persistent lengthening of the Q contours as time goes on, and familiar from other fluid-dynamical paradigms such as 'random straining' and 'turbulence' (e.g. Batchelor 1952).

8. THE DEFINING PROPERTY OF WAVE BREAKING, THE STRATOSPHERIC 'SURF ZONE', AND OZONE CHEMISTRY

What should one mean by wave breaking for general, non-acoustic* waves in fluids? Even in the most familiar case, ordinary surface gravity waves, the phenomenon usually recognized as breaking has an extensive 'zoology' of shapes and time-evolutions. The same is true of internal gravity and Rossby waves. The question does not seem to have any natural answer from a 'zoological' or morphological viewpoint. However, a natural answer does suggest itself if one wants the concept of 'wave breaking' to be relevant to the general question of wave-induced momentum transport, or, more precisely, to the question of when wave-induced momentum transport becomes irreversible.

One can then use the rapid, irreversible material-contour deformation illustrated above as the defining property of wave breaking. 'Rapid' means that deformation rates are comparable, at least, to the local intrinsic wave frequency. Such a definition is entirely compatible with the accepted phenomenology, and relevance, of wave breaking in the case of surface gravity waves and longshore ocean-beach currents. The case for such a generalization is carefully argued in three papers with T. N. Palmer (1983–5). It avoids 'zoological' definitions, and requirements to decide whether 'turbulence' is involved, but does take account of the relevant general theorems, particularly Kelvin's

*Shock formation in acoustic waves seems best kept conceptually separate, if only because its essential dependence on the existence of overtly dissipative processes (microscopic irreversibility).

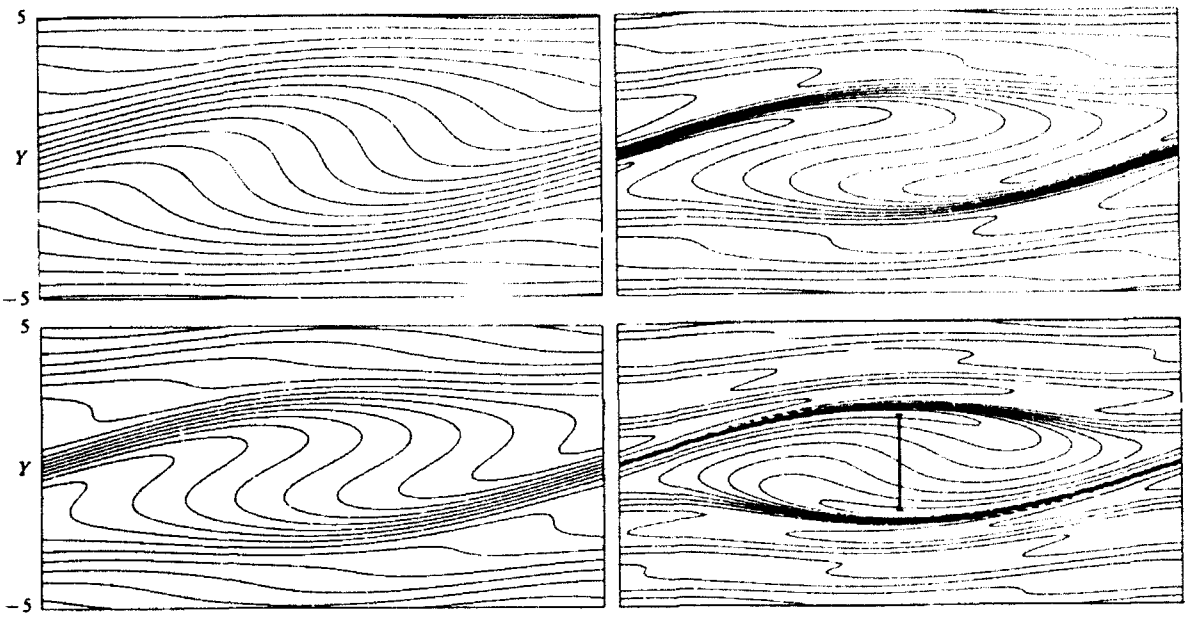


Figure 7. Contours of constant Q for the SWW solution, a special but clear-cut example of Rossby-wave breaking (contrast the purely undular, non-breaking Rossby wave in Figure 6), and the consequent irreversible wave-induced momentum transport. See also Figure 9. The y-scale has been expanded using the re-scaled coordinate $Y = y/b$ with $b = \epsilon^{1/2}\, u_{0y}/\beta$; the range $-5 \leq Y \leq 5$ is plotted. Four successive stages in the evolution are shown, at times 1, 1.5, 2, 3 in units of $2^{1/2}(k\, b\, u_{0y})^{-1}$, where k is the x-wavenumber. The vertical bar in panel (d) gives the effective mixing scale $b_{\rm m} \simeq b$; cf. (7.2), (7.3) and Figure 8a. From Killworth and McIntyre (1985), after Stewartson (1978) and Warn and Warn (1978).

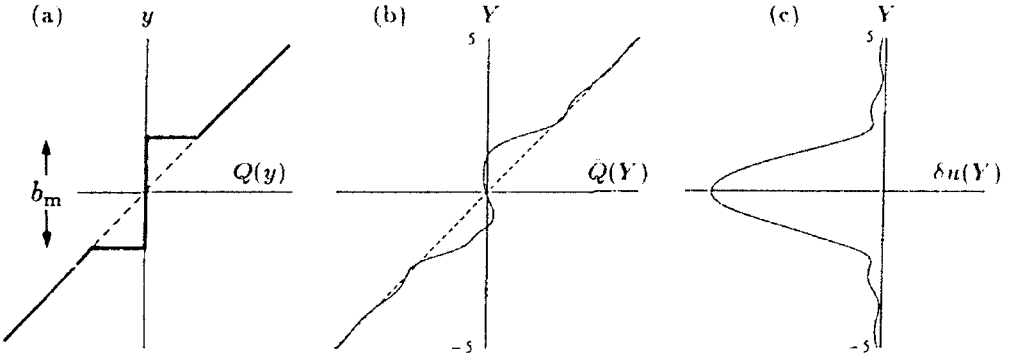


Figure 8. Q-mixing scenarios and associated momentum change; see (7.2), (7.3): (a) the simplistic, but qualitatively relevant, idealization that assumes perfect Q-mixing over width $b_{\rm m}$; (b) an actual Eulerian-mean $Q(y)$ profile from an accurate, quasi-chaotic Rossby-wave critical-layer solution (P. H. Haynes, personal communication); and (c) the approximately parabolic $\delta u(y)$ profile corresponding to (b), showing the momentum change due to the Rossby wave breaking (first of (7.2) and of (7.3)). As in Figure 7, the re-scaled coordinate $Y = y/b$ is used, over the same range -5 to 5 ($b = \epsilon^{1/2}\, u_{0y}/\beta$).

circulation theorem and the way it manifests itself in exact, formally complete theories of wave mean interaction. The relevant material contours are defined to be those that would otherwise undulate reversibly under the generalized elasticity, or restoring mechanism, that gives rise to the wave propagation.

Figure 9, of which an animated version was seen in the lecture, shows a less idealized example of Rossby wave breaking and its effects, taken from the work of Norton (1993). The parameter conditions are far closer to those in the real stratosphere than those assumed by critical-layer theory. Here the dynamical system comprises the shallow-

water equations on a sphere, solved numerically by a high-resolution pseudospectral method. The projection is polar stereographic, and the winter northern hemisphere is shown. An axisymmetric initial state is disturbed by smoothly distorting the lower boundary in a large-scale pattern so as to imitate the effect, on the real stratosphere, of planetary-scale Rossby waves propagating up from the much denser troposphere below (Charney and Drazin 1961). The result looks remarkably similar to what is seen in the real winter stratosphere at altitudes of the order of 25 to 50 km.

The left panel of Figure 9 shows Q, now the shallow-water potential vorticity, defined as absolute vorticity over local layer depth. The model problem still has the generic form (6.1) to excellent approximation, but with a shorter-range potential-vorticity inversion operator* $\mathbf{I}(\cdot)$. The central region is the model's 'stratospheric polar vortex', where the relevant material contours, which in this experiment lie initially along latitude circles, are almost coincident with the Q contours and undulate nearly reversibly. In this region there is little Rossby-wave breaking. This is illustrated by comparison with the right panel, which shows the behaviour of some material contours computed very accurately using a high-precision 'contour advection' technique adapted from the work of Dritschel (1988). Outside the polar vortex, in middle latitudes, is a region in which the waves are breaking vigorously. There, the initially-latitudinal material contours are deformed rapidly and irreversibly, and mixed into a broad, 2D-turbulent 'Rossby-wave surf zone'; e-folding times for contour lengthening were estimated to be about 4 days (Norton, *op. cit.*). This 'surf zone' is the real-stratospheric counterpart of the idealized Rossby-wave critical layer. It is far broader, and very different in detail, but it illustrates equally well the robustness and one-signedness of the angular momentum transport associated with irreversible Q-rearrangement. As already suggested, such Rossby-wave breaking is an important part of how the mean circulation illustrated in Figure 2 is driven. Furthermore, its recognition in models of the global-scale stratospheric transport of trace chemicals like CFCs is beginning to lead to improved realism in the predictions of such models, both via a more realistic mean circulation, and also via a more realistic representation of the quasi-horizontal turbulent transport (Garcia *et al.* 1992).

This latter aspect may be especially important for ozone photochemistry (and relevant to some current controversies about ozone depletion — see §10 of my 1992 review). This is because ozone photochemistry by its nature could be more sensitive than, for instance, CFC photochemistry, to the timing of a typical molecule's excursions across the 'surf zone'. Timescales for such excursions are comparable, at certain altitudes, to photochemical timescales. Together with the mean circulation itself, these complicated fluid motions control the rate at which ozone is produced photochemically, mainly in the high tropical stratosphere, and carried thence to the extratropical lower stratosphere where it accumulates (unless destroyed by 'ozone-hole chemistry') in far greater concentrations than can be produced by tropical photochemistry alone.

Model simulations like that of Figure 9 are also relevant to understanding the Antarctic ozone hole. The contours in the right-hand panel of Figure 9 can be regarded as isopleths of an advected passive tracer, the advection being very accurately simulated, with no artificial diffusion. The simulation shows that, at least in the model, chemical substances in the surf zone do not penetrate past the region of strong Rossby elasticity concentrated in the steep Q gradient near the vortex edge. This is believed to be important

*To rough approximation, this $\mathbf{I}(\cdot)$ is given by (6.2) with ∇^{-2} replaced by $(\nabla^2 - \kappa^2)^{-1}$, corresponding to an elastic membrane tethered by local springs, somewhat like a spring mattress, with a latitude-dependent e-folding scale κ^{-1} (the 'Rossby radius') of about 1400 km at the pole, 2000 km at 45°N, and 3000 km in the tropics. McIntyre and Norton (1990, 1993) give examples of much more accurate (but much more elaborate) potential-vorticity inversion operators $\mathbf{I}(\cdot)$ for shallow-water models.

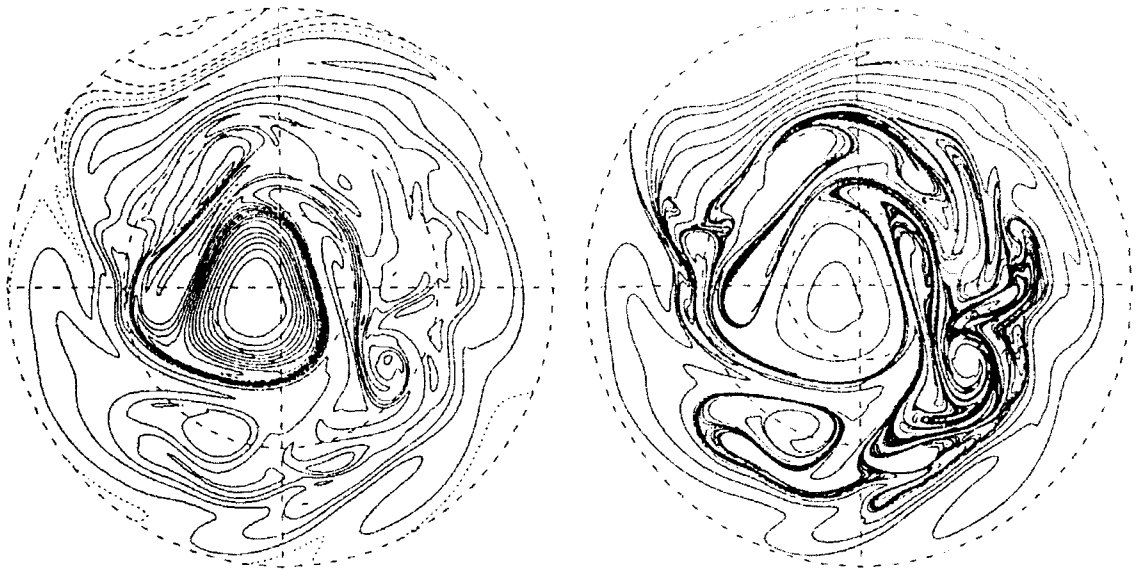


Figure 9. Winter hemisphere in a high-resolution, shallow-water numerical model of the stratosphere, from Norton (1993). The mean depth is 4 km, and the numerical resolution (triangular truncation at total wavenumber 127) corresponds to a mesh size roughly 1° latitude. *Left panel*: potential vorticity Q, contour interval $4 \times 10^{-9}\,\mathrm{m^{-1}s^{-1}}$, zero contour dotted. (Negative Q values would, in a three-dimensional model, signal the three-dimensional mode of Rossby-wave breaking pointed out by O'Sullivan and Hitchman 1992.) *Right panel*: isopleths of an advected tracer field, initially axisymmetric and coincident with the Q contours, showing the fluid-dynamical irreversibility characteristic of Rossby-wave breaking and two-dimensional 'turbulence'. This was computed with near-perfect accuracy using a high-precision 'contour advection' techique, introduced independently by Norton (1993) and by Waugh and Plumb (1993) using an algorithm developed in another context by Dritschel (1988).

for ozone-hole chemistry. The same phenomenon has been demonstrated in the laboratory by Sommeria *et al.* (1989, 1991). The contour-advection technique, conceived of as a benchmark numerical tracer advection algorithm, was introduced independently by Norton (*op. cit.*) and by Waugh and Plumb (1993).

Recently, O'Sullivan and Hitchman (1992) have shown that an entirely different mode of Rossby-wave breaking is possible near the equator, where the potential vorticity itself changes sign. It conforms to the general wave-breaking definition, with three-dimensional rather than layerwise-two-dimensional material contour deformations, arising from an asymmetric inertial (quasi-centrifugal) instability. Among other things, this may have new implications for the quantitative modelling of the QBO.

9. WIND-GENERATED WATER WAVES: TWO NEW MECHANISMS?

What does this old but elusive problem look like from the foregoing perspective? The first point is that Rossby-wave dynamics is involved, albeit on much faster timescales than before. The velocity profile in the air, whether or not approximately logarithmic, will usually have a strong curvature u_{0yy} near the water surface (y vertical). Therefore there is a vertical gradient of spanwise vorticity and hence, in the present language, a Rossby elasticity, in the airflow just above the water waves; recall (6.5) and Figure 6. That is why the Miles inviscid wind–wave instability, for instance, can be regarded as a coupled 'Rossby-wave, gravity-wave' instability. One can think in terms of a pair of phase-locked, counterpropagating waves, a backward-propagating Rossby wave in the air coupled to a forward-propagating gravity wave in the water. (Other examples of Rossby-wave, gravity-wave instabilities go back to G. I. Taylor's work in the 1930s on

the 'Taylor–Goldstein equation'; see also, e.g., Griffiths *et al.* (1982), Hayashi and Young (1987), and Sakai (1989).) In the Miles instability, the Rossby elasticity is relatively weak. This implies not only that the water wave largely determines the phase speed c, but also that the growth rate is slow and that the most significant Rossby effects occur near the critical line $u_0(y) = c$, where intrinsic phase speeds are slow. The mathematical underpinning for these statements is well known and is given in detail by Miles (1957) and Lighthill (1962).

Now the self-consistency of the linearized instability theory requires that conditions near the critical line resemble those in Figure 7a. In the present language, the Rossby wave is not only weak and slow, but is also in the earliest stages of wave breaking, in the generalized sense already referred to. The later stages can be expected to look more like those of Figure 7b–7d and beyond, with a small value of the effective mixing scale b_m, viewed in a suitably undulating coordinate system. This suggests that the idealized scenario represented by the inviscid Miles mechanism lacks robustness, when taken literally. The delicate phase-locking and synchronization of the normal mode's displacement fields will be disrupted by Rossby wave breaking before much vertical rearrangement of spanwise vorticity can take place (and long before the water waves break). The outcome will be net vorticity rearrangement, in the airflow, over only a small effective mixing depth b_m, of the same order as the scale b given by (7.1) except that the lengthscale u_{0y}/β is replaced by a typical value of $-u_{0y}/u_{0yy}$ near the critical line. There is a correspondingly small net wave-coherent momentum transport $\delta M \propto {b_\mathrm{m}}^3$ from air to water, having the order of magnitude implied by (7.3) with β again replaced by a typical value of $-u_{0yy}$ near the critical line.

However, there is one important feature of this idealized scenario that does, on the other hand, look robust. This is the *sign* of the net wave-coherent momentum transport, which tends to be such that the water waves are amplified when travelling in the same direction as the wind. More precisely, the sign is determined by the sign of $-u_{0yy}$ and the one-signedness of the associated Rossby-wave dynamics — the same one-signedness that so strongly controls the sense of the mean circulation throughout most of Figure 2. In the wind–wave problem the sign is determined quite independently of whether or not we have strictly x-periodic waves and delicately synchronized exponential, normal-mode growth.

This sign-robustness must imply a kind of 'ratchet effect'. Almost any spatio-temporal intermittency — whether it be any tendency of the water waves to arrange themselves in groups through, for instance, subharmonic instabilities, or any gustiness of the wind that might be modelled as an intermittency in quantities like $-u_{0yy}$ — will tend to favour intermittent wave-coherent momentum transport whose effects are cumulative. They may possibly also be such as to reinforce the intermittency. There is no longer any reason, moreover, why the effective depths b_m associated with any Rossby wave breaking should be especially small.

More generally, the sign-robustness points to the likely effectiveness of any process in which wave-coherent undulation of the airflow over the water waves causes Rossby wave activity in the air to increase in total amount, or to dissipate, or both. The relevant measure of 'activity' for this purpose is a suitably defined quasimomentum (Killworth and McIntyre 1985; Shepherd 1990, & refs.), since the photon analogy can be shown to apply here, in the required sense. Any such process will result in wave-coherent momentum transport in such a sense as to amplify the water waves. This could well be important for the real wind–wave problem, despite the added complexities of turbulence and other three-dimensional effects. There appear, in particular, to be two distinguishable mechanisms whose possible role deserves closer attention, both theoretically and experimentally.

One is simply local Rossby wave breaking characterized by relatively large effective depths b_m. If, for instance, a gust produces an effective $-u_{0yy}$ that then encounters a wave group, leading to rapid vertical rearrangement of spanwise vorticity (reducing $-u_{0yy}$), then the water waves will be correspondingly amplified for a short time. The transient nature of such phenomena would add to the well-known difficulties of observing the associated pressure phase shift experimentally. The other mechanism, which in reality might tend to occur at the same time, is the formation of Rossby lee waves in the airflow. This will give rise to 'wave resistance' in the usual way, not depending on wave breaking of any kind. Indeed the lee-wave mechanism should be at its most efficient when both Rossby wave breaking and water wave breaking are unimportant. Then one should get unseparated flow over the water waves (e.g. Banner 1990, & refs.) and, for at least some $-u_{0yy}$ profiles, an efficient shedding of Rossby-wave quasimomentum into the downstream airflow. It seems possible that the amplifying wave group in the experiments of Chu *et al.* reported by Professor Phillips, which they call the 'short wave energy front', might depend on the Rossby lee-wave mechanism — and the experiments might offer a chance to study it under controlled conditions.

These ideas might also help resolve some longstanding questions about the effects of three-dimensional turbulence on different timescales, from the relatively long, 'dissipative' timescales required to restore $-u_{0yy}$ profiles, to the relatively short, 'elastic' timescales on which rapid-distortion theory might be appropriate, merely modifying the Rossby elasticity in the air — as with internal gravity waves in the important observational study by Finnigan and Einaudi (1981), in which eddy-viscosity or other steady-state dissipative turbulent modelling was shown to be wrong by 'an order of magnitude'. In some wind–wave models, turbulent processes are modelled semi-empirically for instance by adding something like an eddy viscosity to the Miles theory. Such devices (drastically modifying the type of behaviour suggested by Figure 7) would seem unavoidable in models that insist on x-periodicity, or on horizontal statistical stationarity and linear superposability of wave-coherent processes. But a physically correct turbulence model might need to bring in a wider range of timescales. This suggests that it could be fruitful to try to model the spatio-temporal intermittency of the water waves and the airflow explicitly.

Acknowledgements: Early versions of the ideas developed here were part of an essay that shared the 1981 Adams Prize in the University of Cambridge. Their development has been influenced by many colleagues going back to early ideas from F. P. Bretherton, D. O. Gough, T. Matsuno, R. E. Peierls, E. A. Spiegel, and most recently P. H. Haynes, W. A. Norton, and other co-authors with whom it has been a pleasure to work. The experimental demonstrations in Figure 3 were stimulated by one of M. S. Longuet-Higgins' beautiful wave-tank demonstrations, and G. I. Barenblatt stimulated me to think harder about real turbulent airflow over water. Support from the Natural Environment Research Council (through the UK Universities' Global Atmospheric Modelling Project and through the British Antarctic Survey), the Science and Engineering Research Council, the UK Meteorological Office, and the US Office of Naval Research is gratefully acknowledged.

REFERENCES

Andrews, D. G., Holton, J. R., Leovy, C. B., 1987: *Middle Atmosphere Dynamics*. Academic Press, 489 pp.

Andrews, D. G., McIntyre, M. E., 1976: Planetary waves in horizontal and vertical shear: the generalized Eliassen–Palm relation and the mean zonal acceleration. J. Atmos. Sci., **33**, 2031–2048. [See pp. 2043–5]

Andrews, D. G., McIntyre, M. E., 1978a: An exact theory of nonlinear waves on a Lagrangian-mean flow. J. Fluid Mech., **89**, 609–646. [Note the caveats about Lagrangian means in §10 (regarding cases like that of Figure 7 above), and in McIntyre 1980.]

Andrews, D. G., McIntyre, M. E., 1978b: On wave-action and its relatives. J. Fluid Mech., **89**, 647–664 (Corrigendum **95**, 796).

Armi, L., Hebert, D., Oakey, N., Price, J., Richardson, P. L., Rossby, T., Ruddick, B., 1988: The history and decay of a Mediterranean salt lens. Nature, **333**, 649–651.

Banner, M. L., 1990: The influence of wave breaking on the surface pressure distribution in wind wave interactions. J. Fluid Mech., **211**, 463–495.

Barnett, J. J., Corney, M., 1985: Middle atmosphere reference model derived from satellite data. In: Handbook for MAP, vol. **16**, ed. K. Labitzke, J. J. Barnett, B. Edwards, pp. 47–143. Available from SCOSTEP secretariat, University of Illinois, 1406 W. Green St. Urbana, Ill. 61801, U.S.A.

Batchelor, G. K., 1952: The effect of homogeneous turbulence on material lines and surfaces. Proc. Roy. Soc. **A 213**, 349-366.

Booker, J. R., Bretherton, F. P., 1967: The critical layer for internal gravity waves in a shear flow. J. Fluid Mech., **27**, 513–539.

Bretherton, F. P., 1966: Baroclinic instability and the short wavelength cut-off in terms of potential vorticity. Q. J. Roy. Meteorol. Soc., **92**, 335–345.

Bretherton, F. P., 1971: The general linearised theory of wave propagation. Amer. Math. Soc. Lectures in Applied Mathematics, **13**, 61–102. [See §6.]

Bretherton, F. P., Garrett, C. J. R., 1968: Wavetrains in inhomogeneous moving media. Proc. Roy. Soc. Lond., **A 302**, 529–554.

Brillouin, L., 1925: On radiation stresses. Annales de Physique, **4**, 528–586. (In French.)

Charney, J. G., Drazin, P. G., 1961: Propagation of planetary-scale disturbances from the lower into the upper atmosphere. J. Geophys. Res., **66**, 83–109.

Chu, J. S., Long, S. R., Phillips, O. M., 1992: Measurements on the interaction of wave groups with shorter wind-generated waves. J. Fluid Mech., **245**, 191–210

Davis, C. A., 1992: Piecewise potential vorticity inversion. J. Atmos. Sci., **49**, 1397–1411.

Dickinson, R. E., 1968: Planetary Rossby waves propagating vertically through weak westerly wind waveguides. J. Atmos. Sci., **25**, 984–1002. [See p. 1001 for the suggestion about the QBO.]

Dickinson, R. E., 1969: Theory of planetary wave-zonal flow interaction. J. Atmos. Sci., **26**, 73–81.

Dritschel, D. G., 1988: Contour surgery: a topological reconnection scheme for extended integrations using contour dynamics. J. Comput. Phys., **77**, 240–266.

Dunkerton, T. J., 1983: Laterally-propagating Rossby waves in the easterly acceleration phase of the quasi-biennial oscillation. Atmos.-Ocean, **21**, 55–68.

Finnigan, J. J., Einaudi, F., 1981: The interaction between an internal gravity waves and the planetary boundary layer. Part II: effect of the wave on the turbulence structure. Q. J. Roy. Meteorol. Soc., **107**, 807–832.

Fritts, D. C., 1984: Gravity wave saturation in the middle atmosphere: a review of theory and observations. Revs. Geophys. Space Phys. **22**, 275–308.

Fritts, D. C., 1993: Gravity wave sources, source variability and lower and middle atmospheric effects. In: Coupling Processes in the Lower and Middle Atmosphere, ed. E. V. Thrane, T. A. Blix, D. C. Fritts; Dordrecht, Kluwer, 191–208.

Garcia, R. R., Stordal, F., Solomon, S., Kiehl, J. T., 1992: A new numerical model of the middle atmosphere. 1. Dynamics and transport of tropospheric trace source gases. J. Geophys. Res., **97**, 12967–12991.

Garrett, C., Smith, J., 1976: On the interaction between long and short surface waves. J. Phys. Oceanography, **6**, 925–930.

Gill, A. E., 1982: *Atmosphere-Ocean Dynamics*. Academic Press, 662 pp.

Gray, W. M., Sheaffer, J. D., Knaff, J. A., 1992: Influence of the stratospheric QBO on ENSO variability. J. Meteorol. Soc. Japan, **70**, 975–995.

Green, J. S. A., 1970: Transfer properties of the large-scale eddies and the general circulation of the atmosphere. Q. J. Roy. Meteorol. Soc., **96**, 157–185.

Griffiths, R. W., Killworth, P. D., Stern, M. E., 1982: Ageostrophic instability of ocean currents. J. Fluid Mech., **117**, 343–377.

Hamilton, K., Garcia, R. R., 1984: Long-period variations in the solar semidiurnal atmospheric tide. J. Geophys. Res., **89**, 11705–11710.

Hayashi, Y.-Y., Young, W. R., 1987: Stable and unstable shear modes of rotating parallel flows in shallow water. J. Fluid Mech., **184**, 477–504.

Haynes, P. H., 1989: The effect of barotropic instability on the nonlinear evolution of a Rossby wave critical layer. J. Fluid Mech., **207**, 231–266.

Haynes, P. H., Marks, C. J., McIntyre, M. E., Shepherd, T. G., Shine, K. P., 1991: On the "downward

control" of extratropical diabatic circulations by eddy-induced mean zonal forces. J. Atmos. Sci., **48**, 651–678.

Haynes, P. H., *et al.*, 1993: Notes on the UK National Summer School in Geophysical and Environmental Fluid Dynamics. In preparation.

Hoskins, B. J., McIntyre, M. E., Robertson, A. W., 1985: On the use and significance of isentropic potential-vorticity maps. Q. J. Roy. Meteorol. Soc., **111**, 877–946. Also **113**, 402–404.

Killworth, P. D., McIntyre, M. E., 1985: Do Rossby-wave critical layers absorb, reflect or over-reflect? J. Fluid Mech., **161**, 449–492.

Lighthill, M. J., 1962: Physical interpretation of the mathematical theory of wave generation by wind. J. Fluid Mech., **14**, 385–398.

Lighthill, M. J., 1963: Boundary layer theory. In: Laminar Boundary Layers, ed. L. Rosenhead; Oxford University Press, 46–113. [See p. 93.]

Lighthill, M. J., 1978a: *Waves in Fluids.* Cambridge University Press, 504 pp.

Lighthill, M. J., 1978b: Acoustic streaming. J. Sound Vib., **61**, 391–418.

Lindzen, R. S., Holton, J. R., 1968: A theory of the quasi-biennial oscillation. J. Atmos. Sci., **25**, 1095–1107.

Lindzen, R. S., Tsay, C. Y., 1975: Wave structure of the tropical stratosphere over the Marshall Islands area during 1 April – 1 July 1958. J. Atmos. Sci., **32**, 2008–2021.

Longuet-Higgins, M. S., 1969: Action of a variable stress at the surface of water waves. Phys. Fluids, **12**, 737–740.

Longuet-Higgins, M. S., Stewart, R. W., 1962: Radiation stress and mass transport in gravity waves, with application to 'surf beats'. J. Fluid Mech., **13**, 481–504.

Lorenz, E. N., 1967: *The Nature and Theory of the General Circulation of the Atmosphere.* Geneva, World Meteor. Org., 161 pp. [See especially pp. 85, 150.]

Loudon, R., Paige, E. G. S., 1991: Alan Frank Gibson. Biogr. Mem. F. Roy. Soc. London, **37**, 221–244.

McIntyre, M. E., 1972: On Long's hypothesis of no upstream influence in uniformly stratified or rotating flow. J. Fluid Mech., **52**, 209–243.

McIntyre, M. E., 1973: Mean motions and impulse of a guided internal gravity wave packet. J. Fluid Mech., **60**, 801–811.

McIntyre, M. E., 1980: Towards a Lagrangian-mean description of stratospheric circulations and chemical transports. Phil. Trans. Roy. Soc. Lond., **A 296**, 129–148.

McIntyre, M. E., 1981: On the 'wave momentum' myth. J. Fluid Mech., **106**, 331–347.*

McIntyre, M. E., 1992: Atmospheric dynamics: some fundamentals, with observational implications. In: Proc. Internat. School Phys. "Enrico Fermi", CXV Course, ed. J. C. Gille, G. Visconti; Amsterdam, Oxford, New York, Toronto, North-Holland, pp. 313–386. [A list of bibliographical updates and minor corrections is available from the author, on paper or by email (*mem2@phx.cam.ac.uk*). For caveats about Lagrangian means, see also the 1980 paper cited above.]

McIntyre, M. E., Mobbs, S. D., 1993: On the 'quasimomentum rule' for wave-induced mean forces on obstacles immersed in a fluid medium. To be submitted to Proc. Roy. Soc. Lond. [See McIntyre 1973 for an *exception* to the rule.]

McIntyre, M. E., Norton, W. A., 1990: Dissipative wave-mean interactions and the transport of vorticity or potential vorticity. J. Fluid Mech., **212**, 403–435 (G. K. Batchelor Festschrift Issue.); Corrigendum **220**, 693.

McIntyre, M. E., Norton, W. A., 1993: Potential-vorticity inversion on a hemisphere. J. Atmos. Sci., to appear.

McIntyre, M. E., Palmer, T. N., 1983: Breaking planetary waves in the stratosphere. Nature, **305**, 593–600.

McIntyre, M. E., Palmer, T. N., 1984: The "surf zone" in the stratosphere. J. Atm. Terr. Phys., **46**, 825–849.

McIntyre, M. E., Palmer, T. N., 1985: A note on the general concept of wave breaking for Rossby and gravity waves. Pure Appl. Geophys., **123**, 964–975.

Miles, J. W., 1957: On the generation of surface waves by shear flows. J. Fluid Mech., **3**, 185–204.

Munk, W. H., Miller, G. R., Snodgrass, F. E., Barber, N. F., 1963: Directional recording of swell from

*There is a slip on p. 339, sixth line from bottom: 'length of the wavetrain' should read 'wavelength'. Also, I now think that the statement on p. 338 about a horizontal distance of order $H \gg ct$ (with $c = (gH)^{1/2}$, tenth line from bottom) is wrong. The main points are unaffected.

distant storms. Phil. Trans. Roy. Soc. Lond., **A 255**, 505–584.

Norton, W. A., 1993: Breaking Rossby waves in a model stratosphere diagnosed by a vortex-following coordinate system and a contour advection technique. J. Atmos. Sci., submitted.

Okuda, K., 1982: Internal flow structure of short wind waves. Part 2. The streamline pattern. J. Oceanog. Soc. Japan, **38**, 313–322.

O'Sullivan, D. J., Hitchman, M. H., 1992: Inertial instability and Rossby wave breaking in a numerical model. J. Atmos. Sci., **49**, 991–1002.

Peierls, R., 1991: *More Surprises in Theoretical Physics*. Princeton, 106 pp.

Phillips, O. M., 1977: *The Dynamics of the Upper Ocean*, 2nd edition. Cambridge University Press, 336 pp.

Plumb, R. A., McEwan, A. D., 1978: The instability of a forced standing wave in a viscous stratified fluid: a laboratory analogue of the quasi-biennial oscillation. J. Atmos. Sci., **35**, 1827–1839.

Rayleigh, Lord, 1896: The theory of sound, volume 2. New York, Dover (reprinted 1945), 504 pp. [See §352.]

Raymond, D. J., 1992: Nonlinear balance and potential-vorticity thinking at large Rossby number. Q J. Roy. Meteorol. Soc., **118**, 987–1015.

Rhines, P. B., 1975: Waves and turbulence on a beta-plane. J. Fluid Mech., **69**, 417–443.

Sakai, S., 1989: Rossby-Kelvin instability: a new type of ageostrophic instability caused by a resonance between Rossby waves and gravity waves. J. Fluid Mech., **202**, 149–176. [See also Griffiths *et al.* 1982, Hayashi and Young 1987.]

Saravanan, R., 1993: Idealized modelling of tropospheric processes in the mechanism of the quasi-biennial oscillation. Manuscript in preparation. (The author may be contacted at the National Center for Atmospheric Research, Boulder, CO., U. S. A.)

Shepherd, T. G., 1990: Symmetries, conservation laws, and Hamiltonian structure in geophysical fluid dynamics. Adv. Geophys., **32**, 287–338.

Snodgrass, F. E., Groves, G. W., Hasselmann, K. F., Miller, G. R., Munk, W. H., Powers, W. H., 1966: Propagation of ocean swell across the Pacific. Phil. Trans. Roy. Soc. Lond., **A 259**, 431–497. [All the way to Alaska, via several recording stations including New Zealand.]

Solomon, S., Kiehl, J. T., Garcia, R. R., Grose, W., 1986: Tracer transport by the diabatic circulation deduced from satellite observations. J. Atmos. Sci., **43**, 1603–1617.

Sommeria, J., Meyers, S. D., Swinney, H. L., 1989: Laboratory model of a planetary eastward jet. Nature, **337**, 58–61.

Sommeria, J., Meyers, S. D., Swinney, H. L., 1991: Experiments on vortices and Rossby waves in eastward and westward jets. In: Nonlinear Topics in Ocean Physics, ed. A. R. Osborne; Amsterdam, North-Holland, 227–269.

Starr, V. P., 1968: *Physics of negative viscosity phenomena*. McGraw-Hill, 256 pp.

Stewartson, K., 1978. The evolution of the critical layer of a Rossby wave. Geophys. Astrophys. Fluid Dyn., **9**, 185-200.

Takahashi, M., Boville, B. A., 1992: A three-dimensional simulation of the equatorial quasi-biennial oscillation. J. Atmos. Sci., **49**, 1020–1035.

Teitelbaum, H., Vial, F., Bauer, P., 1993: The stratospheric quasi-biennal oscillation observed in the semidiurnal ground pressure data. J. Climate, submitted.

Thomas, G. E., Olivero, J. J., Jensen, E. J., Schröder, W., Toon, O. B., 1989: Relation between increasing methane and the presence of ice clouds at the mesopause. Nature, **338**, 490–492.

Thorncroft, C. D., Hoskins, B. J., McIntyre, M. E., 1993: Two paradigms of baroclinic-wave life-cycle behaviour. Q. J. Roy. Meteorol. Soc., **119**, 17–55.

Thorpe, A. J., 1985: Diagnosis of balanced vortex structure using potential vorticity. J. Atmos. Sci., **42**, 397–406.

Wallace, J. M., Holton, J. R., 1968: A diagnostic numerical model of the quasi-biennial oscillation. J. Atmos. Sci., **25**, 280–292.

Van Dyke, M., 1982: *An Album of Fluid Motion*. Stanford, Parabolic Press.

Warn, T. and Warn, H., 1978. The evolution of a nonlinear critical level. Stud. Appl. Math., **33**, 2021-2024.

Waugh, D. W., Plumb, R. A., 1993: Contour advection with surgery: a technique for investigating fine scale structure in tracer transport. J. Atmos. Sci., submitted.

Yoden, S., Holton, J. R., 1988: A new look at equatorial quasi-biennial oscillation models. J. Atmos. Sci , **45**, 2703–2717.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

COMPUTATIONAL ASPECTS OF INTEGRATION ALONG THE PATH OF LOADING IN ELASTIC-PLASTIC PROBLEMS

J B Martin

FRD/UCT Centre for Research in Computational and Applied Mechanics, University of Cape Town, 7700 Rondebosch, South Africa

Abstract

The determination of the mechanical response of a body subjected to some prescribed history of loading is the classical problem in the analysis of elastic-plastic bodies. Closed form solutions are not possible, except in a very limited number of problems, and numerical methods must be adopted. Spatial discretisation is most commonly carried out by the finite element method, and the process is well understood. Time discretisation, and the associated algorithms to advance from one time step to the next, however, have been the subject of considerable discussion in recent years.

Work contributing towards this problem has been carried out in the Centre for Research in Computational and Applied Mechanics at the University of Cape Town, resulting in a coherent overall framework. This paper is intended to summarise and consolidate the major features of this framework.

The process begins with the time discretisation of the continuous problem. The backward difference algorithm is adopted because of its relevance in plasticity: this step is equivalent to replacing the elastic-plastic material by a nonlinear elastic material for the time step. The minimisation of the potential energy of this nonlinear elastic problem leads to a nonlinear programming problem. If the elastic-plastic material is stable, the nonlinear programming problem will be convex.

The minimisation of the programming problem can be carried out by a two step iterative process which can be identified with the conventional Newton-Raphson algorithm. In the first (predictor) step, the nonlinear functional is approximated by a quadratic functional whose minimisation is a linear problem. The choice of quadratic functional affects the convergence of the problem; several choices are possible. The second (corrector) step involves the calculation of the properties of the nonlinear functional for the state predicted by means of the approximation; this requires a local calculation and can be identified with the return algorithm.

The spatial discretisation is now introduced into the predictor step - this is essentially the

discretisation of a linear problem. The final finite element formulation is thus achieved. Several other questions are amenable to analysis in this approach, and generalisation to trapezoidal and midpoint rules is possible. These issues will be discussed briefly.

1 Introduction

Our purpose in this paper is to summarise and review work carried out at the Centre for Research in Computational and Applied Mechanics at the University of Cape Town in recent years [1 - 21] on the topic of the time integration of the classical problem in plasticity. What is sought is insight into the mechanics of the integration process, as opposed to a purely mathematical approach to the integration of the differential equations.

The work referred to in [1-21] is framed largely in terms of an internal variable theory of plasticity: this theory has provided several insights which are not immediately clear in the conventional formulation. In this paper, however, we will present the concepts in a conventional formulation.

We shall consider an elastic-plastic body subject to some quasi-static loading history. We shall assume that the displacements are small, and that the material is stable. The basic incremental equations will be formulated, and time discretisation will be achieved by using the backward difference scheme. On this basis we will show that the incremental elastic-plastic problem reduces to a nonlinear elastic or holonomic problem, with which a convex nonlinear programming problem can be associated.

We will then discuss a two step iterative algorithm for the minimisation of the programming problem; this can be identified with the conventional Newton-Raphson method. Issues of convergence can also be clearly understood. Finally, the backward difference approximation can be relaxed, giving an equivalent trapezoidal and midpoint rule.

2 Formulation of the incremental problem

We consider a body of volume V and surface S. The coordinates of a point in the body are denoted by $x_i, \quad i = 1,2,3$, and time is denoted by t. The strain - displacement relations are

$$\epsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \tag{1}$$

where $u_i(x_i, t)$ is the displacement field and $\epsilon_{ij}(x_i, t)$ is the strain tensor.

The equilibrium equations are

$$\frac{\partial \sigma_{ij}}{\partial x_j} + F_i = 0 \quad \text{on} \quad V, \tag{2}$$

where $F_i(x_i,t)$ are body forces and $\sigma_{ij}(x_i,t)$ is the stress tensor. In addition

$$\sigma_{ij}\,\nu_j = p_i \quad \text{on} \quad S, \tag{3}$$

where $p_i(x_i,t)$ are surface tractions, and ν_i is the outward normal to the surface.

The loading history comprises specified body forces $F_i(x_i,t)$ on V, surface tractions $p_i(x_i,t)$ on part of the surface S_p, and prescribed displacements $u_i(x_i,t)$ on the remainder of the surface S_u. For simplicity we shall assume that $u_i(x_i,t) = 0$ on S_u.

The governing equations are completed by the constituitive relations. We assume that the strain ϵ_{ij} can be divided into elastic and plastic components, ϵ^e_{ij}, ϵ^p_{ij}:

$$\epsilon_{ij} = \epsilon^e_{ij} + \epsilon^p_{ij}. \tag{4}$$

The elastic strain is given by

$$\epsilon^e_{ij} = D_{ijkl}\,\sigma_{kl}, \tag{5}$$

where D_{ijkl} is the compliance tensor. We define a yield function

$$\phi = \phi(\sigma_{ij}, H_\alpha) \tag{6}$$

where H_α are internal variables which control hardening. Then

$$\dot{\epsilon}^p_{ij} = \lambda\,\frac{\partial\phi}{\partial\sigma_{ij}}, \tag{7}$$

with

$$\lambda = 0 \quad \text{if} \quad \phi < 0 \tag{8}$$

$$\text{or} \quad \phi = 0 \quad \text{and} \quad \frac{\partial\phi}{\partial\sigma_{ij}}\dot{\sigma}_{ij} \le 0. \tag{9}$$

These are the cases of elastic behaviour and unloading from a plastic state. For the case of plastic loading,

$$\lambda \ge 0 \quad \text{for} \quad \phi = 0 \quad \text{and} \quad \frac{\partial\phi}{\partial\sigma_{ij}}\dot{\sigma}_{ij} \ge 0. \tag{10}$$

We note that during plastic deformation

$$\dot{\phi} = \frac{\partial \phi}{\partial \sigma_{ij}} \dot{\sigma_{ij}} + \frac{\partial \phi}{\partial H_\alpha} \dot{H}_\alpha = 0. \tag{11}$$

Depending on the specific form of H_α, and ensuring that equation (11) is satisfied, we may show that

$$\lambda = G \frac{\partial \phi}{\partial \sigma_{kl}} \dot{\sigma_{kl}}, \quad G \geq 0. \tag{12}$$

In addition, we must specify evolution equations for the internal variables H_α. We note that $\dot{H}_\alpha = 0$ for $\phi < 0$ or $\phi = 0$ and $(\partial \phi / \partial \sigma_{ij}) \dot{\sigma_{ij}} \geq 0$.

The constitutive equations are invertible, and may be written in the form

$$\dot{\sigma_{ij}} = B_{ijkl}\, \dot{\epsilon_{kl}}, \tag{13}$$

where $B_{ijkl} = B_{ijkl}(\epsilon_{ij}, H_\alpha)$. For stability, we require that $\dot{\sigma_{ij}}\, \dot{\epsilon_{ij}} \geq 0$, and hence that B_{ijkl} is positive semi-definite. Equation (13) can be integrated with respect to time along any given strain path in order to determine the stress response.

3 Discretisation in time

We now introduce a time discretisation. We suppose that we know the solution at time t_{n-1} and we denote the fields by $F_i^{n-1}, p_i^{n-1}, u_i^{n-1}, \epsilon_{ij}^{n-1}, \sigma_{ij}^{n-1}$. At a subsequent instant t_n we know F_i^n, p_i^n on S_p, u_i^n on S_u, and we must find u_i^n on V and S_p, ϵ_{ij}^n, σ_{ij}^n.

Let

$$\Delta u_i^n = u_i^n - u_i^{n-1}, \quad \Delta \epsilon_{ij}^n = \epsilon_{ij}^n - \epsilon_{ij}^{n-1}. \tag{14}$$

The kinematic equations (1) are linear, and hence we can write for time t_n:

$$\frac{\partial \sigma_{ij}^n}{\partial x_j} + F_i^n = 0 \quad \text{on} \quad V, \tag{15}$$

$$\sigma_{ij}^n \nu_j = p_i^n \quad \text{on} \quad S, \tag{16}$$

$$\Delta \epsilon_{ij}^n = \frac{1}{2} \left(\frac{\partial \Delta u_i^n}{\partial x_j} + \frac{\partial \Delta u_i^n}{\partial x_i} \right). \tag{17}$$

To these equations we must add a set of discrete constitutive equations relating $\Delta \epsilon_{ij}^n$ and σ_{ij}^n in order to complete the governing equations for time t_n.

4 Discrete constitutive relations

As a result of the time discretisation, we must introduce an additional assumption in order to be able to integrate the constitutive equations (13) over the interval t_{n-1}, t_n. Physically, this can be interpreted as choosing a path in strain space along which the strain changes from the previous value ϵ_{ij}^{n-1} to the new value ϵ_{ij}^{n}, as shown diagrammatically in Figure 1.

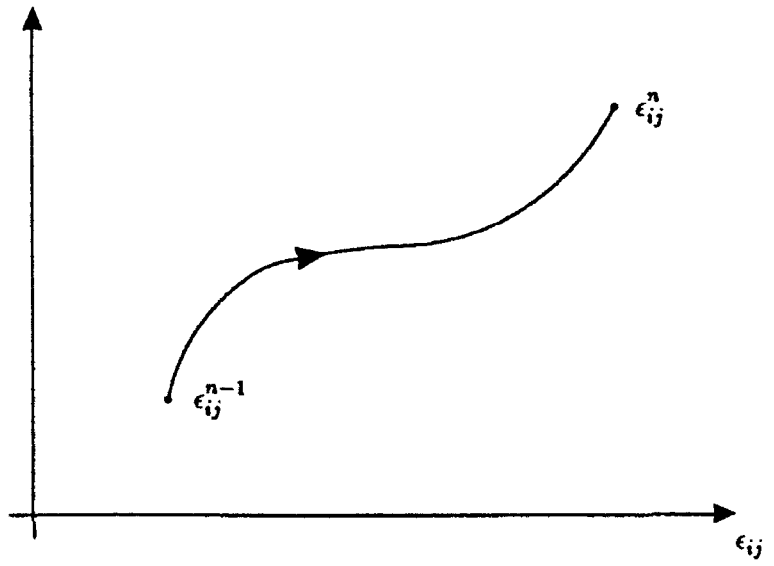


Figure 1: Assumed path in strain space

There are a variety of possible choices, but one which has particularly desirable properties is the *minimum work path* [22,12]. The minimum work path is defined as a path in strain space for which the integral

$$W_n = \int_{\epsilon_{ij}^{n-1}}^{\epsilon_{ij}^{n}} \sigma_{ij}\dot{\epsilon}_{ij}dt \tag{18}$$

takes its least value W_n^o for given ϵ_{ij}^{n-1} and arbitrary ϵ_{ij}^{n}. The integral W_n is evaluated by integrating the constitutive equation (13) along the path. The minimum work path is not unique, in the sense that there may exist a family of paths for which the work is the least value W_n^o.

The function W_n^o has several important properties. If the elastic-plastic material is stable, $W_n^o\{\epsilon_{ij}^{n}, \epsilon_{ij}^{n-1}, H_\alpha(t_{n-1})\} = W_n^o\{\Delta\epsilon_{ij}^{n}, \epsilon_{ij}^{n-1}, H_\alpha(t_{n-1})\}$ will be a convex function of $\Delta\epsilon_{ij}^{n}$. Further, it is a potential function which provides the stress associated with ϵ_{ij}^{n} at the end of the path. Thus

$$\sigma_{ij}^n = \left.\frac{\partial W_n^o}{\partial \Delta\epsilon_{ij}}\right|_{\Delta\epsilon_{ij}^n} . \tag{19}$$

For conventional plastic materials, the minimum work path [23] is characterised by a straight line path in internal variable space. Alternatively, it is a path such that the plastic strain direction is fixed. This implies that, for a stress change from σ_{ij}^{n-1} to σ_{ij}^n, the path in stress space , as shown in Figure 2, is divided into two parts: first an elastic segment, with the stress path remaining inside the yield surface, and second, a segment in which the stress remains constant at yield (for an elastic, perfectly plastic material) or moves out along a normal to the yield surface for a hardening material. Integration of the constitutive equations along this path is generally fairly straightforward, and in most cases equation (18) can be evaluated explicitly.

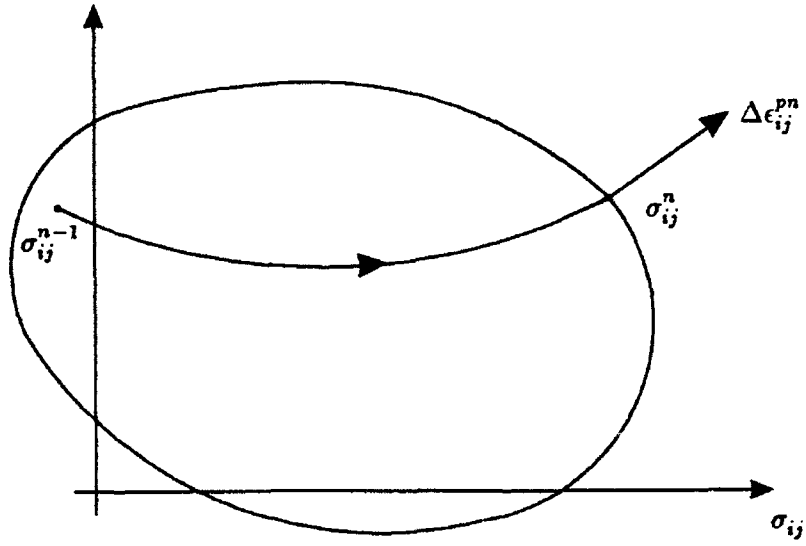


Figure 2: Minimum work path for an elastic, perfectly plastic material

In the case of an elastic, perfectly plastic material the path in stress space is particularly straightforward (see Figure 2): a stress change occurs first, accompanied by elastic strain changes, and the the stress remains constant while plastic strain changes take place. It is evident from Figure 2, for an elastic, perfectly plastic material, that if the yield surface is convex, we may write

$$(\sigma_{ij}^n - \sigma_{ij}^{n-1})\Delta\epsilon_{ij}^{pn} \geq 0. \tag{20}$$

This may be recognised as an incremental form of the *principle of maximum plastic work*; it is the existence of this inequality, and its appropriate generalisation for hardening materials, which leads to the convexity of W_n^o.

The minimum work path assumption in fact corresponds to a *backward difference assumption* for plastic strain rates: if we were to put

$$\dot{\epsilon}^p_{ij} = \{(1-\alpha)\dot{\epsilon}^{p(n-1)}_{ij} + \alpha\dot{\epsilon}^{pn}_{ij}\}, \tag{21}$$

and to integrate the constitutive equation (13) over the interval t_{n-1}, t_n, we would recover the minimum work path with $\alpha = 1$.

5 The incremental problem

The incremental problem is now defined in terms of the field equations (15-17) and the constitutive equation (19). We may recognise that the constitutive equation is essentially a stress-strain relation for a *nonlinear elastic* or *holonomic* material, with W^o_n playing the role of a strain energy function. Thus, as a result of the minimum work path or backward difference assumption, the incremental problem in plasticity is transformed into an *incremental holonomic or nonlinear elastic problem.*

The problem may be equivalently stated in the form of a programming problem, in which we are required to minimise the potential energy of the holonomic problem. The solution to the incremental problem is given by the fields Δu^n_i, $\Delta\epsilon^n_{ij}$ which give the least value of the functional

$$U^o_n = \int_V W^o_n(\Delta\epsilon_{ij})dV - \int_V F^n_i \Delta u_i dV - \int_{S_p} p^n_i \Delta u_i dS. \tag{22}$$

The functional U^o_n is convex if W^o_n is convex. When the incremental solution is obtained, equation (22) is updated, with $(n+1)$ replacing n.

6 Solution algorithm

We now consider an appropriate algorithm for finding the least value of the functional U^o_n. The simplest and most direct algorithm is a two step algorithm. In the first step, which we shall refer to as the predictor step, we approximate the nonlinear functional U^o_n by a quadratic functional: the minimisation of this quadratic approximation is then a linear problem, and leads to improved estimates of the kinematic variables. In the second step, which will be referred to as the corrector step, we recompute those properties of the functional U^o_n which are needed to proceed with the next iteration.

We shall consider each of these steps in turn for the $(i+1)th$ iteration: we have estimates Δu^{ni}_i, $\Delta\epsilon^{ni}_{ij}$, and we seek improved estimates $\Delta u^{n(i+1)}_i$, $\Delta\epsilon^{n(i+1)}_{ij}$. We put

$$\Delta\hat{u}_i = \Delta u^{n(i+1)}_i - \Delta u^{ni}_i, \quad \Delta\hat{\epsilon}_{ij} = \Delta\epsilon^{n(i+1)}_{ij} - \Delta\epsilon^{ni}_{ij}. \tag{23}$$

The functional U_n^o is then written as

$$U_n^o = \int_V W_n^o(\Delta\epsilon_{ij}^{n\,i} + \Delta\hat{\epsilon}_{ij})dV - \int_V F_i^n(\Delta u_i^{ni} + \Delta\hat{u}_i)dV - \int_{S_p} p_i^n(\Delta u_i^{ni} + \Delta\hat{u}_i)dS. \tag{24}$$

6.1 The predictor step

In the predictor step we simplify the nonlinear problem by introducing a quadratic approximation to the function W_n^o. Some flexibility is permissible in the choice of the approximation: we write

$$\begin{aligned}\hat{W}_n^o(\Delta\epsilon_{ij}^{n\,i} + \Delta\hat{\epsilon}_{ij}) &= W_n^o(\Delta\epsilon_{ij}^{n\,i}) + \left.\frac{\partial W_n^o}{\partial\Delta\epsilon_{ij}}\right|_{\Delta\epsilon_{ij}^{n\,i}} \Delta\hat{\epsilon}_{ij} + \frac{1}{2}\hat{C}_{ijkl}^i\Delta\hat{\epsilon}_{ij}\Delta\hat{\epsilon}_{kl} \\ &= W_n^o(\Delta\epsilon_{ij}^{n\,i}) + \sigma_{ij}^{ni}\Delta\hat{\epsilon}_{ij} + \frac{1}{2}\hat{C}_{ijkl}^i\Delta\hat{\epsilon}_{ij}\Delta\hat{\epsilon}_{kl}.\end{aligned} \tag{25}$$

Substituting $\hat{W}_n^o$ in equation(25) for W_n^o in equation (24) we may define a new functional U_n^{oi} given by

$$U_n^{oi} = U_n^o(\Delta u_i^{ni},\ \Delta\epsilon_{ij}^{n\,i}) + \hat{U}_n^o, \tag{26}$$

where

$$\hat{U}_n^o = \int_V \frac{1}{2}\hat{C}_{ijkl}^i\Delta\hat{\epsilon}_{ij}\Delta\hat{\epsilon}_{kl}\,dV - \left\{\int_V F_i^n\Delta\hat{u}_i\,dV + \int_{S_p} p_i^n\Delta\hat{u}_i\,dS - \int_V \sigma_{ij}^{ni}\Delta\hat{\epsilon}_{ij}\,dV\right\}. \tag{27}$$

The least value of $\hat{U}_n^o$ gives values $\Delta\hat{u}_i$, $\Delta\hat{\epsilon}_{ij}$ and hence improved estimates $\Delta u_i^{n(i+1)}$, $\Delta\epsilon_{ij}^{n(i+1)}$ from equation (19). What is significant is that $\hat{U}_n^o$ is a quadratic function, and hence the problem of finding its least value is a linear problem. A variety of choices of $\hat{C}_{ijkl}^i$ can be made: these will be discussed in detail when we refer to the rate of convergence of the algorithm.

6.2 The corrector step

We note that in the expression for $\hat{U}_n^o$, given in equation (27), we need to know F_i^n, p_i^n, which are part of the data, and σ_{ij}^{ni}, which is not given. Thus before advancing to the next iteration we must compute $\sigma_{ij}^{n(i+1)}$; this is the corrector step.

The stress $\sigma_{ij}^{n(i+1)}$ depends on the strain at the end of the n-th *time step*, and the strain at the end of the $(i+1)$-th iteration for the $(n+1)$-th time step, or the total accumulated increment $\Delta\epsilon_{ij}^{n(i+1)}$. The accumulated increment is assumed to take place along the path

designated in Section 4, and the constitutive equations are integrated along this path. We note thus that the computation of $\sigma_{ij}^{n(i+1)}$ is a *local* calculation at each point in the body.

If the backward difference assumption is adopted, integration of the constitutive equations reduces to the problem of determining $\Delta\epsilon_{ij}^{pn(i+1)}$, since we can put

$$\sigma_{ij}^{n(i+1)} = \sigma_{ij}^{(n-1)} + C_{ijkl}(\Delta\epsilon_{kl}^{n\,(i+1)} - \Delta\epsilon_{kl}^{pn(i+1)}), \tag{28}$$

where $C_{ijkl} = D_{ijkl}^{-1}$ are the elastic constants.

It is convenient to define an elastically predicted stress $\sigma_{ij}^{nE(i+1)}$, defined by

$$\sigma_{ij}^{nE(i+1)} = \sigma_{ij}^{(n-1)} + C_{ijkl}\Delta\epsilon_{kl}^{n\,(i+1)}. \tag{29}$$

Then, if H_α^{ni} are internal variables associated with the $i-th$ iteration, we note that

$$\sigma_{ij}^{n(i+1)} = \sigma_{ij}^{nE(i+1)}, \quad \Delta\epsilon_{ij}^{pn(i+1)} = 0, \quad \text{if} \quad \phi(\sigma_{ij}^{nE(i+1)}, H_\alpha^{ni}) \leq 0. \tag{30}$$

If $\phi(\sigma_{ij}^{nE(i+1)}, H_\alpha^{ni}) > 0$, then $\Delta\epsilon_{ij}^{pn(i+1)} \neq 0$. The computation of $\sigma_{ij}^{n(i+1)}$ is then generally represented graphically, and is referred to as the *return algorithm*. The simplest case, that of an elastic, perfectly plastic material, is shown in Figure 3. The minimum work path, as described earlier, is made up of an elastic change from $\sigma_{ij}^{(n-1)}$ to $\sigma_{ij}^{n(i+1)}$, followed by plastic strain changes at constant stress if $\sigma_{ij}^{n(i+1)}$ lies on the yield surface. In Figure 3, for the case where $\Delta\epsilon_{ij}^{pn(i+1)} \neq 0$, we show the vectors $(\sigma_{ij}^{n(i+1)} - \sigma_{ij}^{(n-1)})$ and $C_{ijkl}\Delta\epsilon_{ij}^{pn(i+1)}$, which make up $\sigma_{ij}^{nE(i+1)}$. It can clearly be seen that if $\sigma_{ij}^{nE(i+1)}$ is given, we can drop a perpendicular back onto the yield surface in order to find $\sigma_{ij}^{n(i+1)}$. Similar, slightly more elaborate rules of the same type can be constructed for hardening materials, for both continuously differentiable and piecewise continuously differentiable yield surfaces.

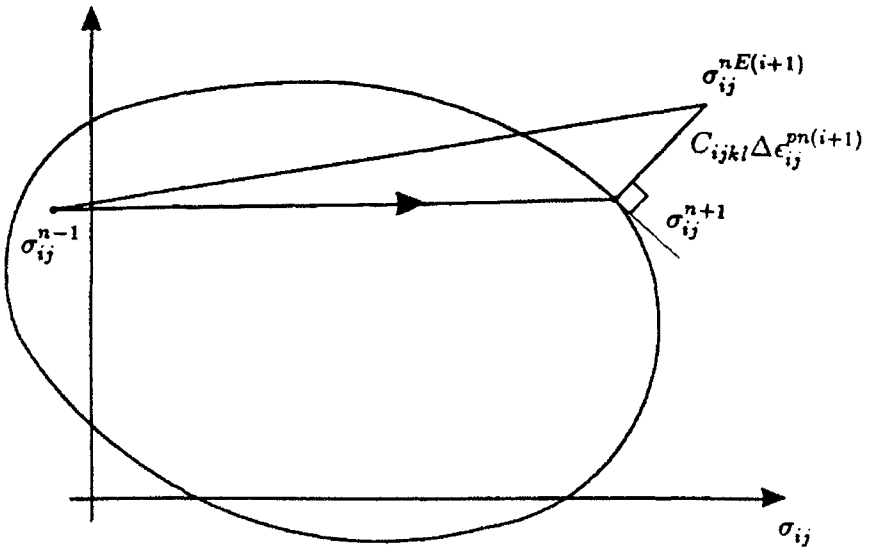


Figure 3: Return algorithm for an elastic perfectly plastic material

We emphasise, however, that the return algorithm is not a heuristic choice, but can be set up rigorously for any elastic-plastic material in terms of the minimum work path. Alternatively, in the internal variable formulation the return algorithm corresponds to a local minimum principle [20]. Piecewise linear yield surfaces are included in these formulations.

7 Convergence

The steps of the iterative algorithm are shown diagrammatically in Figure 4. The sketch shows the function U_n^o we set out to minimise, together with the $i-th$ estimate of the solution. The approximating quadratic function U_n^{oi} is also shown, leading to the $(i+1)-th$ estimate. The corrector step permits us to re-establish the new estimate on the original functional U_n^o.

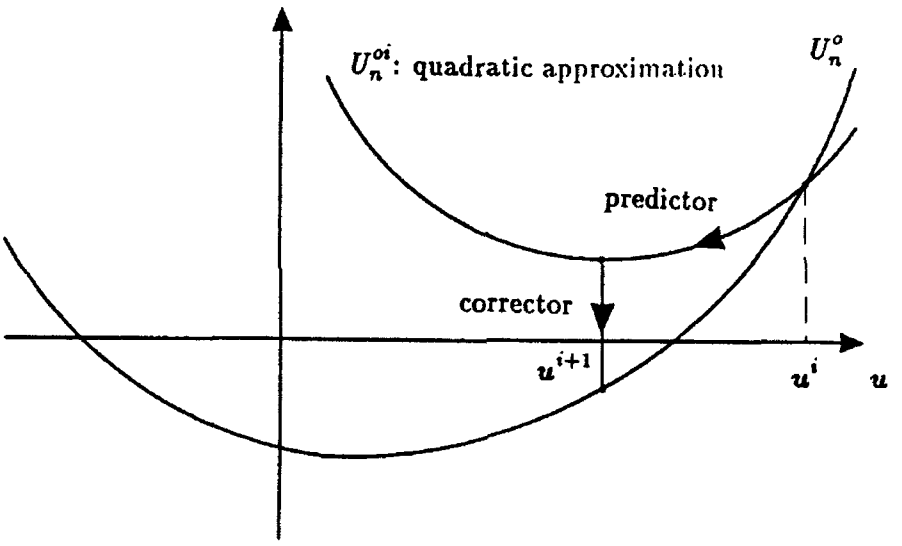


Figure 4: Steps in the iterative algorithm

Thus at the end of the two steps of the iteration we can compute $U_n^o(\Delta u_i^{n(i+1)}, \Delta\epsilon_{ij}^{n\,(i+1)})$. A comparison of $U_n^o(\Delta u_i^{ni}, \Delta\epsilon_{ij}^{n\,i})$ and $U_n^o(\Delta u_i^{n(i+1)}, \Delta\epsilon_{ij}^{n\,(i+1)})$ provides us with information on the convergence of the algorithm: it is clear that a sufficient condition for monotonic convergence is that

$$U_n^o(\Delta u_i^{n(i+1)}, \Delta\epsilon_{ij}^{n\,(i+1)}) \leq U_n^o(\Delta u_i^{ni}, \Delta\epsilon_{ij}^{n\,i}). \tag{31}$$

Further, since by the definition of the process,

$$U_n^{oi}(\Delta u_i^{n(i+1)}, \Delta\epsilon_{ij}^{n\,(i+1)}) \leq U_n^o(\Delta u_i^{ni}, \Delta\epsilon_{ij}^{n\,i}), \tag{32}$$

it follows that the sufficient condition for monotonic convergence is met if

$$U_n^o(\Delta u_i^{n(i+1)}, \Delta\epsilon_{ij}^{n(i+1)}) \leq U_n^{oi}(\Delta u_i^{n(i+1)}, \Delta\epsilon_{ij}^{n(i+1)}) \tag{33}$$

In geometrical terms, this implies that sufficient conditions for monotonic convergence can be met if the least value of the functional U_n^{oi} lies within the functional U_n^o. Whether or not this occurs depends on the choice of the predictor modulus $\hat{C}_{ijkl}^i$.

It can be shown that this condition is met if $\hat{C}_{ijkl}^i$ is chosen as the elastic constants C_{ijkl} [13,16]. However, the disadvantage of this choice is that the rate of convergence is very slow. Improvements can be made by choosing the secant modulus [13] or a softened elatic modulus γC_{ijkl}, where $0.5 \leq \gamma \leq 1$ [24]. The improvements, however, are not substantial, and the rate of convergence remains slow.

The rate of convergence can be increased substantially if the consistent tangent predictor [25] is chosen: here we put

$$\hat{C}_{ijkl}^i = \left.\frac{\partial^2 W_n^o}{\partial\Delta\epsilon_{ij}\partial\Delta\epsilon_{kl}}\right|_{\Delta\epsilon_{ij}^{n_i}} \tag{34}$$

We note that if $\Delta\epsilon_{ij}^{pni} = 0$,

$$\left.\frac{\partial^2 W_n^o}{\partial\Delta\epsilon_{ij}\partial\Delta\epsilon_{kl}}\right|_{\Delta\epsilon_{ij}^{n_i}} = C_{ijkl}, \tag{35}$$

i.e. the consistent tangent predictor is simply the elastic modulus. It is also important to note that in the first iteration in a load increment $\Delta\epsilon_{ij}^{pn1} = 0$ at all points in the body, and hence the elastic moduli are chosen as the elastic predictor for this iteration. This ensures convergence in the first step even when the load increment represents unloading.

In general, however, the consistent tangent predictor does not meet the sufficient condition for monotonic convergence of equation (33). This follows because the least value of the functional U_n^{oi} does not necessarily lie within the functional U_n^o, and therefore constraint (33) is not met in all circumstances.

8 Line search algorithm

Convergence of the algorithm with the consistent tangent predictor can be assured if a line search algorithm [25,26] is adopted. In our present context, this can be demonstrated by simple geometric arguments.

In the line search algorithm we compute $\Delta\hat{u}_i$, $\Delta\hat{\epsilon}_{ij}$ as before, but we consider improved estimates

$$\Delta u_i^{n(i+1)} = \Delta u_i^{ni} + a\Delta\hat{u}_i, \quad \Delta\epsilon_{ij}^{n(i+1)} = \Delta\epsilon_{ij}^{ni} + a\Delta\hat{\epsilon}_{ij}, \tag{36}$$

where $a > 0$ is a scalar multiplier.

Consider first the approximating quadratic function U_n^{oi}, given in equations (26) and (27). We may put

$$\begin{aligned} U_n^{oi}(\Delta u_i^{ni} + a\Delta\hat{u}_i,\ \Delta\epsilon_{ij}^{ni} + a\Delta\hat{\epsilon}_{ij}) = \\ U_n^o(\Delta u_i^{ni},\ \Delta\epsilon_{ij}^{ni}) + a^2 \int_V \frac{1}{2}\hat{C}_{ijkl}^i \Delta\hat{\epsilon}_{ij}\Delta\hat{\epsilon}_{kl}\, dV \\ -a\left\{\int_V F_i^n \Delta\hat{u}_i\, dV + \int_{S_p} p_i^n \Delta\hat{u}_i\, dS - \int_V \sigma_{ij}^{ni}\Delta\hat{\epsilon}_{ij}\, dV\right\}. \end{aligned} \tag{37}$$

As expected, this function has its least value for $a = 1$; note also that $U_n^{oi}(\Delta u_i^{ni} + a\Delta\hat{u}_i,\ \Delta\epsilon_{ij}^{ni} + a\Delta\hat{\epsilon}_{ij}) = U_n^o(\Delta u_i^{ni},\ \Delta\epsilon_{ij}^{ni})$ for $a = 0$ and $a = 2$. The function is shown diagrammatically in Figure 5.

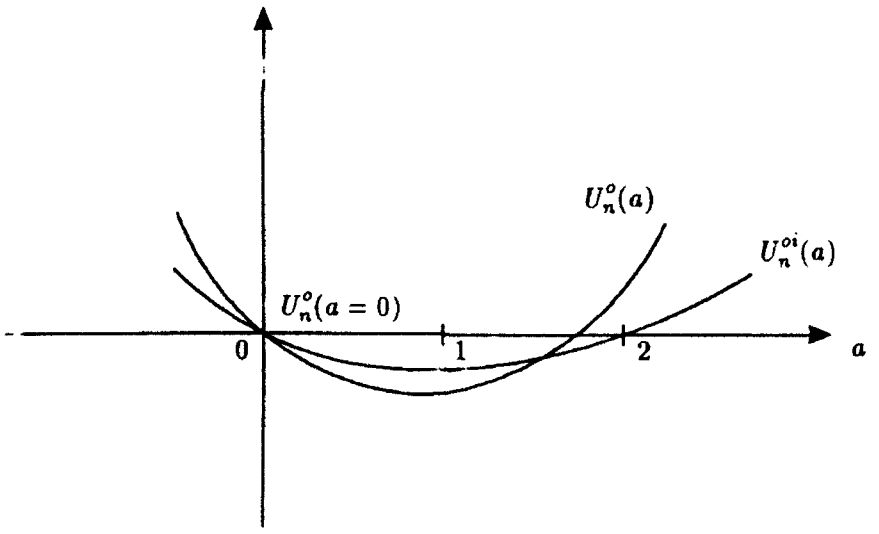


Figure 5: Convergence of the line search algorithm

Now consider the nonlinear functional $U_n^o(\Delta u_i^{ni} + a\Delta\hat{u}_i,\ \Delta\epsilon_{ij}^{ni} + a\Delta\hat{\epsilon}_{ij})$. First, we note that

$$U_n^o(a=0) = U_n^{oi}(a=0). \tag{38}$$

Further, we see that

$$\left.\frac{dU_n^o}{da}\right|_{a=0} = \left.\frac{dU_n^{oi}}{da}\right|_{a=0} . \tag{39}$$

The function $U_n^o(a)$ is also plotted diagrammatically in Figure 5. The geometry of this function permits us to reach two conclusions. First, the least value of U_n^0 will occur for $a > 0$, and will be less than $U_n^o(\Delta u_i^{n\,i}, \Delta\epsilon_{ij}^{n\,i})$, following from equations (38) and (39). Second, $U_n^o(a = 1)$ may be less than or greater than $U_n^o(a = 0)$. Thus, if the line search algorithm is not routinely used, the iteration may or may not provide an improved estimate in the sense of an improved value of U_n^o. However, if $U_n^o(a = 1) \geq U_n^o(a = 0)$, it is clear that there exists a value a', $0 < a' < 1$, for which $U_n^o(a = a') \leq U_n^o(a = 0)$.

In general, the application of the line search algorithm involves finding, at least approximately, the value of a which provides the least value of $U_n^o(a)$. This may occur for $a < 1$, or for $a > 1$; in either case monotonic convergence is assured and the optimum improvement for the iteration is found.

9 Spatial discretisation

With the framework described above in place, we may now consider the implementation of the finite element discretisation. This is applied to the weak form of equations (26) and (27), i.e. the minimisation of U_n^{oi} or $\hat{U}_n^o$. This is a quadratic form, and standard finite element discretisation procedures may be applied. The discrete form of equation (27) will be

$$\hat{U}_n^o = \frac{\mathrm{i}}{2}\Delta\hat{\boldsymbol{u}}^T\boldsymbol{K}^{\mathrm{i}}\Delta\hat{\boldsymbol{u}} - \Delta\hat{\boldsymbol{u}}^T\boldsymbol{R}^{\mathrm{i}}, \tag{40}$$

where $\Delta\hat{\boldsymbol{u}}$ is the vector of improvements in the nodal displacements, $\boldsymbol{K}^{\mathrm{i}}$ is the consistent tangent matrix and $\boldsymbol{R}^{\mathrm{i}}$ is the residual. The tangent matrix $\boldsymbol{K}^{\mathrm{i}}$ will be symmetric if $\hat{C}_{ijkl}^{\mathrm{i}}$ is symmetric, as is the case with the choices of $\hat{C}_{ijkl}^{\mathrm{i}}$ which have been discussed. It may be noted that the magnitude of the residual depends on the difference between the external nodal forces associated with F_i^n, p_i^n and the internal nodal forces associated with $\sigma_{ij}^{n\mathrm{i}}$. When the external and internal nodal forces are in equilibrium, $\boldsymbol{R}^{\mathrm{i}} = 0$ and equation (40) gives $\Delta\hat{\boldsymbol{u}} = 0$, indicating that the iteration is complete.

10 Generalised trapezoidal rule

The development of the solution algorithm has been presented in terms of the adoption of the minimum work path as the strain path for the increment. This is equivalent to a backward difference assumption in the approximation of the plastic strain over the interval (see equation 21).

Experience has shown that the backward difference assumption is robust and effective [27]. It also has the advantage of providing a symmetric consistent tangent stiffness matrix

and unconditional B - stability. Nevertheless, the backward difference approximation is associated with linear rather than quadratic convergence. First attempts to formulate a trapezoidal rule led to non-symmetric operators.

Ortiz and Popov [28] proposed a general class of integration algorithms, including both a generalised mid- point rule and a generalised trapezoidal rule. These rules do not provide symmetric operators, except for the special case of the backward difference rule in each case; this is a major disadvantage. Simo and Govindjee [29] have shown that a midpoint rule can be formulated which provides both quadratic convergence and a symmetric operator. It can be shown simply that this is equivalent to a generalised trapezoidal rule with a symmetric operator.

Full details of the development have been given by Rencontre, Caddemi and Martin [19] and Caddemi [21]. The assumed strain path over the interval remains the same: the initial stress is $\sigma_{ij}^{(n-1)}$, and the final stress is σ_{ij}^{n}. The strain path is made up of two parts, and elastic part and a part where plastic strain takes place with a fixed direction in strain space. In computing the strain increment, however, we include a part of the plastic strain at the end of the previous increment and a part of the plastic strain at end of the current increment. Thus, we put

$$\Delta\epsilon_{ij}^{n} = D_{ijkl}(\sigma_{ij}^{n} - \sigma_{ij}^{(n-1)}) + (1-\alpha)\Delta\epsilon_{ij}^{p(n-1)} + \alpha\Delta\epsilon_{ij}^{pn}, \tag{41}$$

where $0 \leq \alpha \leq 1$.

When $\alpha = 1$, we recover the backward difference formulation. The rule is shown diagrammatically in Figure 6 for a perfectly plastic material. Note that $(1-\alpha)\Delta\epsilon_{ij}^{p(n-1)}$ is *known* from the previous increment; if $\sigma_{ij}^{(n-1)}$ lies inside the yield surface $\Delta\epsilon_{ij}^{p(n-1)}$ will of course be zero. This is the essential difference from the trapezoidal rule of Ortiz and Popov [28], where the direction of the plastic strain associated with $\sigma_{ij}^{(n-1)}$ was assumed known, but not its magnitude.

We note also that $\Delta\epsilon_{ij}^{pn}$ will be zero if σ_{ij}^{n} lies within the yield surface. In the case of unloading, where $\Delta\epsilon_{ij}^{p(n-1)}$ is not zero, we see that $\Delta\epsilon_{ij}^{n}$ may thus involve plastic strain projected from the previous increment. This suggests that the generalised trapezoidal rule may not be optimally accurate for unloading steps, where the backward difference assumption is more effective.

The operator for the generalised trapezoidal rule so defined is symmetric, is unconditionally B - stable for $\alpha \geq \frac{1}{2}$, and provides second order accuracy for $\alpha = \frac{1}{2}$; it is indeed directly equivalent to the generalised midpoint rule of Simo and Govindjee.

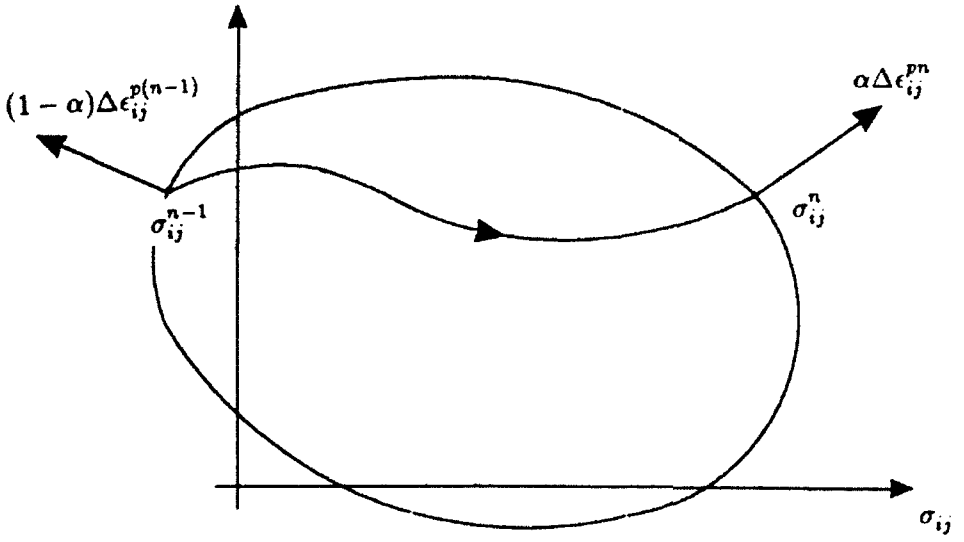


Figure 6: Generalised trapezoidal rule for an elastic perfectly plastic material

11 Conclusions

A coherent overall framework for the iterative solution of the incremental elastic-plastic problem which is based on the physical concepts of extremal paths and piecewise holonomic materials has been presented.

The underlying physical concepts allow several insights which are not accessible in a purely mathematical treatment, and work is continuing to attempt to exploit these insights.

Acknowledgement

The work reported in this paper has been supported by Foundation for Research Development. The author is indebted to Professor B D Reddy for his comment and criticism.

References

[1] J B Martin, 'A complementary work bounding principle for forward integration along the path of loading', J Appl Mech, **54**, 341 - 345, 1987.

[2] J B Martin, B D Reddy, T B Griffin and W W Bird, 'Applications of mathematical programming concepts to incremental elastic-plastic analysis', Engineering Structures, **9**, 171 - 176, 1987.

[3] B D Reddy, J B Martin and T B Griffin, 'Extremal paths and holonomic constitutive laws in elastoplasticity', Quart App Math, **45**, 487 - 502, 1987.

[4] J B Martin and B D Reddy, 'Variational principles and solution algorithms for internal variable formulations of problems in plasticity', in *Omaggio a Giulio Ceradini*, U Andreas *et al* (editors), Dipartimento di Ingegneria Strutturale e Geotecnica, Universita di Roma La Sapienza", 465 - 478, 1988.

[5] T B Griffin, B D Reddy and J B Martin, 'A numerical study of holonomic approximations to problems in plasticity', Int J Num Meth Engg,**26**, 1449 - 1466, 1988.

[6] J B Martin, 'Convergence and shakedown for discrete load steps is statically loaded elastic-plastic bodies', Mech Struct and Mach, **16**, 1 - 16, 1988.

[7] J B Martin and W W Bird, 'Integration along the path of loading in elastic-plastic problems', Engineering Computations, **5**, 217 - 223, 1988.

[8] U Perego, 'Explicit backward difference operators and consistent predictors for linear hardening elastic-plastic constitutive laws', Solid Mech Arch, **13**,65 - 102, 1988.

[9] B D Reddy and T B Griffin, 'Variational principles and convergence of finite element approximations of a holonomic elastic-plastic problem', Num Math, **52**, 101 - 117, 1988.

[10] R A Eve and B D Reddy, ' Variational principles and finite element approximations to problems in finite-strain plasticity', in *Computational Plasticity: Models, Software and Applications*, D R J Owen, E Hinton and E Oñate (editors), Pineridge Press, Volume 1, 71 - 84, 1989.

[11] G Maier and A Nappi, 'Backward difference time integration, nonlinear programming and extremum theorems in elastoplastic analysis', Solid Mech Arch, **14**, 17 - 36, 1989.

[12] M Ortiz and J B Martin, ' Symmetry-preserving return mapping algorithms and incrementally extremal paths', Int J Num Meth Engg, **28**, 1839 - 1853, 1989.

[13] W W Bird and J B Martin, ' Consistent predictors and the solution of the piecewise holonomic incremental problem in elasto-plasticity', Engineering Structures, **12**, 9 - 14, 1990.

[14] J B Martin and A Nappi, 'An internal variable formulation for perfectly plastic and linear hardenting relations in plasticity', Eur J Mech A:Solids, **9**, 107 - 131, 1990.

[15] R A Eve, B D Reddy and R T Rockafellar, ' An internasl variable theory of plasticity based on the maximum plastic work inequality', Quart App Math, **48**, 59 - 84, 1990.

[16] S Caddemi and J B Martin, 'Convergence of the Newton-Raphson algorithm in elastic-plastic incremental analysis', Int J Num Meth Engg, **31**, 177 - 191, 1991.

[17] J B Martin and S Caddemi, ' Sufficient conditions for convergence of the Newton-Raphson algorithm in elastic-plastic analysis', Eur J Mech A:Solids, to appear.

[18] B D Reddy and J B Martin, 'Finite element approximations and return mapping algorithms algorithms for internal variable formulations of problems in plasticity', Comp Meth Appl Mech Engg, **93**, 253 - 273, 1991.

[19] L J Rencontre, S Caddemi and J B Martin, 'The relationship between the generalised midpoint and trapezoidal rules in incremental elasto- plasticity', Comp Meth in Appl Mech Engg, **96**, 201-212, 1992.

[20] L J Rencontre, J B Martin and W W Bird, 'Internal variable formulation of a backward difference corrector algorithm for piecewise linear yield surfaces', Meccanica, **27**, 13-24, 1992.

[21] S Caddemi, ' Computational aspects of the integration of the von Mises hardening constitutive laws', Int J Plasticity, to appear.

[22] A R S Ponter and J B Martin, 'Some extremal properties and energy theorems for inelastic materials and their relationship to the deformation theory of plasticity', J Mech Phys Sol, **20**, 281 - 300, 1972.

[23] P Carter and J B Martin, 'Work bounding functions for plastic materials', J Appl Mech, **43**, 434 - 438, 1976.

[24] C Comi and G Maier, ' On the convergence of a backward difference iterative procedure in elastoplasticity with nonlinear kinematic and isotropic hardening', in *Computational Plasticity: Models, Software and Applications*, D R J Owen, E Hinton and E Oñate (editors), Pineridge Press, Volume 1, 323 - 334, 1989.

[25] J C Simo and R S Taylor, 'Consistent tangent operators for rate-independent elastoplastcity', Comp Meth Appl Mech Engg, **48**, 101 - 118, 1985.

[26] M A Crisfield, 'Accelerating and damping the modified Netwon-Raphson method', Comp Struct, **18**, 395 - 407, 1984.

[27] R D Krieg and D B Krieg, ' Accuracies of numerical solution methods for the elastic-perfectly plstic model', J Press Vessel Tech, ASME, **99**, 510 - 515, 1977.

[28] M Ortiz and E P Popov, ' Accuracy and stability of integration algorithms for elastoplastic constitutive relations', Int J Num Meth Engg, **21**, 1561 - 1576, 1985.

[29] J C Simo and S Govindjee, ' Nonlinear B- stability and symmetry preserving return mapping algorithms for plasticity and viscoplasticity, Int J Num Meth Engg, **31**, 151 - 176, 1991.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

CONSTITUTIVE MODELING AND ANALYSIS OF CREEP, DAMAGE AND CREEP CRACK GROWTH UNDER NEUTRON IRRADIATION

S. Murakami

Department of Mechanical Engineering, Nagoya University
Furo-cho, Chikusa-ku, Nagoya, 464-01, Japan

Abstract

Nuclear fusion technology is currently making remarkable progress, and the structural design of the experimental reactors is becoming one of the most crucial problems in this technology. The present paper is concerned with the constitutive modeling of inelastic behavior and damage evolution of materials under thermal and neutron-irradiation conditions applicable to the rational structural analysis of nuclear reactor components. After discussing the mechanisms and the physical theories of neutron irradiation reported so far, a constitutive model of creep, swelling and damage under irradiation was developed. It was assumed that the creep under irradiation can be decomposed into irradiation-induced creep and irradiation-affected thermal creep. The applicability of the proposed constitutive equations was demonstrated by analyzing the creep, swelling and creep damage of type 316 stainless steel under various conditions of irradiation. As an application of the proposed equations, creep crack growth under neutron irradiation was also analyzed.

1. INTRODUCTION

Polycrystalline metals exposed to the irradiation of fast neutron at elevated temperatures not only show salient intrinsic phenomena of irradiation growth, swelling, irradiation creep, irradiation damage, etc., but also are subjected to significant influence on their mechanical behaviors of deformation and fracture under stresses [1-4]. Fuel cladding tubes of FBR(Fast Breeder Reactors) [5-7], for example, are irradiated by fast neutron flux of the energy E>0.1MeV at the temperatures of 400-700°C, and their mechanical behaviors have been investigated extensively.

Another and more important objective of irradiation effects is the plasma facing components of fusion reactors. Hitherto, the central problem of the nuclear fusion has been concentrated to the realization of the fusion reaction. However, a plasma-burn experiment due to D-T(Deuterium-Tritium) fusion reaction was demonstrated in the JET(Joint European Torus) in November 1991 for the first time in the history, and the success of this experiment has confirmed the feasibility of the controllable nuclear fusion reaction. This success implies that the problems of nuclear fusion are coming to the engineering stage, and the intensive

international collaborations for the conceptual and engineering design of the experimental fusion reactors have been already effected.

The Tokamak type fusion reactor of ITER(International Thermonuclear Experimental Reactor) project, in particular, is supposed to have the burn time of 1000 sec and burn cycles of 10^5, and the neutron energy and the neutron wall loading are 14.1MeV and $1MW/m^2$, respectively. If Type 316 stainless steel is employed as the first wall material, it will be subject to the maximum temperature of 400°C and the displacement damage of 31.4 dpa during the whole life; the material will undergo swelling, irradiation creep, irradiation damage under considerable cyclic thermal stresses.

The recent development of the nuclear fusion technology is making the structural design of the experimental fusion reactors one of the most essential problems in that technology. Though a number of investigation have been performed on the creep and damage under neutron irradiation, they are mainly concerned either with the kinetics of the interstitial atoms, vacancies and dislocations, or with crystal physics and metallurgy, and thus continuum mechanics theories on the constitutive behavior of material under irradiation have been still undeveloped.

The present paper is concerned with the mechanical modeling of creep, swelling and damage of polycrystalline metals under thermal and neutron irradiation conditions applicable to the structural analyses of nuclear reactor components under multiaxial states of stress. A constitutive equation of creep, swelling and damage under irradiation will be formulated in the framework of continuum mechanics, and by taking account of physical mechanisms of neutron irradiation. Continuum damage mechanics will be employed to describe the irradiation damage and final fracture of materials. It will be assumed that the creep under irradiation is decomposed into irradiation-induced creep and irradiation-affected thermal creep. The proposed constitutive equation was applied to analyze the creep and damage of type 316 stainless steel under various conditions of irradiation. Finally, the utility of the proposed equations was demonstrated by analyzing the creep crack growth by means of local approach of fracture based on finite element method.

Since the irradiation-induced creep is weakly dependent on temperature [2, 8], the effect of temperature variation will be disregarded in this paper. Though the bubble formation at grain boundaries, driven by helium generation, may affect the mechanical behavior of austenitic stainless steels at high temperatures [3, 8], it will not be taken into account because of the lack of experimental data.

2. MECHANISMS AND PHYSICAL THEORIES OF IRRADIATION-RELATED CREEP

Preceding to the continuum mechanics modeling of creep and damage subject to the irradiation effects, let us first review the current understanding of the physical mechanisms and mechanical aspects of the relevant phenomena.

2.1. Mechanisms of irradiation related creep

When a fast neutron collides with an atom in a crystal, the atom is knocked on from the lattice cite, and gives a pair of an interstitial atom and a vacancy (*Frenkel pair*). If the *primary knock-on atom* has sufficiently high energy, the atom will hit another atom, initiates a sequence of collision and knocking-on of the constituent atoms, and thus induces a *collision cascade* in the crystal [1, 3]. Then, a part of the resulting interstitial atoms and vacancies may annihilate by recombining with each other or by clustering, or may be absorbed by sinks such as voids, dislocations and grain boundaries [1-4].

Interstitial atoms have larger mobility than vacancies. A part of interstitial atoms coalesce to form dislocation loops, or are absorbed by existing dislocations to facilitate their climb. The annihilation or absorption of interstitial atoms increases the number of lattice sites, and brings about dilatation (*swelling*). The excess vacancies, on the other hand, aggregate to form voids, and contribute to void swelling. Thus when polycrystalline metals are subjected to irradiation at elevated temperatures, in addition to the ordinary thermal creep, the coalescence and the absorption of interstitial atoms together with the aggregation of vacancies give rise to *irradiation-induced creep* and *swelling*.

Though the motion of dislocations under stress may be accelerated by their facilitated climb due to the absorption of interstitials, they may be decelerated at the same time by the formation of dislocation loops, dislocation networks, vacancies, clusters and precipitates brought about by the irradiation. The ordinary thermal creep caused by stress-controlled dislocation motion, therefore, is significantly affected by the neutron irradiation. Thus, the creep under irradiation may be divided into *irradiation-induced creep* and *irradiation-affected thermal creep*. The dilatational part of the irradiation-induced creep is identified as *swelling*. The term *irradiation creep* will represent the creep which is obtained by subtracting the stress-free swelling from the irradiation-induced creep.

2.2. Physical models of irradiation-induced creep

Though a number of physical models have been proposed so far to describe the creep under irradiation, they have been focused mainly on irradiation-induced creep, or irradiation creep and swelling [1-3]. One of the most successful models to describe the irradiation-induced creep is the *Climb-Enhanced Glide* (CEG) mechanism, originally due to Ansell and Weertman [9]. Absorption of interstitial atoms brought about by the irradiation facilitates the climb and glide of dislocations. By postulating the operation of the Frank-Read dislocation source and assuming that the creep rate is proportional to the climb velocity, we have the creep rate as follows [2,3]:

$$\dot{\varepsilon} = k(\sigma^2/G^2)\dot{S} \tag{1}$$

where $(\dot{})$ is the time derivative, and k, σ and $\dot{S}$ stands for a material constant, modulus of rigidity and the swelling rate. This mechanism postulates the flow of excess interstitials into dislocations, and the creep rate is proportional to the rate of point defect nucleation, and hence to the swelling rate $\dot{S}$. By assuming that the climbed dislocation segment can bow and glide

according to the Mott-Friedel equation for dislocation-bowing, Gittus [10, 11], furthermore, developed the following equations:

$$\dot{\varepsilon} = k(\sigma / G)\dot{S} \tag{2}$$

$$\dot{\varepsilon}_{ij} = k\Psi_D \sigma_{Dij} \tag{3}$$

where $\dot{\varepsilon}_{ij}$, σ_{Dij} and Ψ_D are the strain rate tensor, deviatoric stress tensor and the dislocation flux, respectively.

Heald and Speight [12], on the other hand, by taking account of the elastic interaction of a point defect with the strain field around an edge dislocation, proposed *Stress-Induced Preferential Absorption* (SIPA) mechanism. Namely, the magnitude of the long range force between an edge dislocation and an interstitial atom depends on the relative orientation between the directions of applied stress and the Burgers vector, and the dislocations of different Burgers vectors have different rates of interstitial flux. Therefore, dislocations of the preferential orientation may easily absorb interstitials, easily climb and give rise to the macroscopic creep rate in the corresponding direction. Since the anisotropic rate of dislocation climb is proportional to the rate of nucleation of point defects K_i,the creep rate is finally given as follow [1, 2, 12]:

$$\dot{\varepsilon} = kK_i\sigma \tag{4}$$

The SIPA mechanism has been ascertained to explain a number of important features of creep under irradiation, especially when the swelling is not significant. Since the CEG and SIPA mechanisms are not mutually exclusive but may operate concurrently, the total creep rate can be given by the combination of these two mechanisms [13]:

$$\dot{\varepsilon} = k_1\sigma\dot{S} + k_2K_i\sigma \tag{5}$$

Finally the nucleation of irradiation-induced interstitial loops also depends on the direction of the applied stress; they nucleate preferentially on crystallographic planes nearly perpendicular to the applied stress. Brailsford and Bullough [14], for example, assuming this SIPN (*Stress-Induced Loop Nucleation*) mechanism, derived the following equation:

$$\dot{\varepsilon} = k_1 b^2 \rho_D \sigma\dot{S} \tag{6}$$

where b and ρ_D are the Burgers vector and the dislocation density of the interstitial loops. The SIPN mechanism is valid for low dislocation density, but has a serious restriction. Namely, according to this model, once a prefered loop orientation is nucleated, the strain rate will be controlled by the irradiation dose rate, but is not affected further by the applied stress.

The swelling rates $\dot{S}$ in eqs. (1), (2), (5) and (6) have been given by the rate theory based on the diffusion rates and density of interstitials and vacancies [1, 3, 13], or by the empirical equations expressed in terms of stress, flux of irradiation and temperature [2, 7, 15].

3. MECHANICAL MODELING OF CREEP, SWELLING AND DAMAGE UNDER IRRADIATION

In nuclear reactor components operating at elevated temperatures, besides the irradiation-induced creep discussed above, the stress-controlled thermal creep under irradiation is also predominant. Thus for the engineering design of the reactor components, we need not only to reformulate the above irradiation-induced creep in terms of mechanics, but also to develop a rational constitutive equation which unifies the irradiation creep, swelling, thermal creep and damage and can be easily incorporated into the computer algorithms.

For this purpose, we will note that the creep under irradiation ε_{ij}^{c} can be divided into the irradiation-induced creep ε_{ij}^{IIC} and the irradiation-affected thermal creep ε_{ij}^{ITC}, the former is caused exclusively by the irradiation and may occur even under vanishing stress, while the latter is the ordinary thermal creep but accelerated or decelerated by irradiation.

3.1. Mechanical modeling of irradiation-induced creep [16-18]

For the mechanical modeling of irradiation-induced creep ε_{ij}^{IIC}, let us first postulate two important mechanisms of CEG and SIPA. According to these mechanisms, as described in the preceding sections, the rate of irradiation-induced creep $\dot{\varepsilon}_{ij}^{IIC}$ will be specified by the current state of stress σ_{ij} and the current rates of nucleation of interstitial atoms and vacancies, with respect to the current state of internal structure of the material developed by the irradiation. The rates of nucleation of interstitial atoms and vacancies will be proportional to the neutron flux ϕ, and the state of the internal structure of the material will be identified by the densities and configurations of interstitial atoms, vacancies and dislocations brought about by the irradiation. Since the dislocation density, above all, is identified by the neutron fluence $\Phi = \int \phi \, dt$ [2, 20], we will assume that the internal state of material can be described by Φ. The rate of irradiation-induced creep $\dot{\varepsilon}_{ij}^{IIC}$ may be represented as follows:

$$\dot{\varepsilon}_{ij}^{IIC} = F_{ij}(\sigma_{kl}, \phi, \Phi) \tag{7}$$

In contrast to the dislocation-controlled thermal creep, the distribution of dislocations and the internal structure of the material caused in the irradiation-induced creep will have not significant anisotropy. Thus, in view of the linear stress dependence of CEG and SIPA described by eqs. (2) and (4), the general form of eq. (7) may be expressed as an isotropic tensor function of stress σ_{ij} of the order one and zero.

$$\dot{\varepsilon}_{ij}^{IIC} = \eta(\phi,\Phi)\delta_{ij} + \xi(\phi,\Phi)\sigma_{kk}\delta_{ij} + \zeta(\phi,\Phi)\sigma_{ij} \tag{8}$$

where symbols η, ξ, and ζ are material functions of ϕ and Φ, and δ_{ij} denotes the Kronecker delta. If we decompose the stress and strain into isotropic and deviatoric parts, eq. (8) leads to

$$\dot{\varepsilon}_{ij}^{IIC} = (1/3)\dot{\varepsilon}_{kk}^{IIC}\delta_{ij} + \dot{\varepsilon}_{Dij}^{IIC} \tag{9a}$$

$$\dot{\varepsilon}_{kk}^{IIC} = 3\eta + (3\xi + \zeta)\sigma_{kk} \equiv \dot{S} \tag{9b}$$

$$\dot{\varepsilon}_{Dij}^{IIC} = \zeta\sigma_{Dij} \tag{9c}$$

where $\dot{\varepsilon}_{Dij}^{IIC}$ and σ_{Dij} are the deviatoric parts of $\dot{\varepsilon}_{ij}^{IIC}$ and σ_{ij}, respectively.

The symbol $\dot{\varepsilon}_{kk}^{IIC}$ of eq. (9) stands for the volumetric strain rate, and hence corresponds exactly to the swelling rate $\dot{S}$. Therefore, in view of the incubation period of swelling [2, 7, 21] and the linear dependence of creep and swelling on the neutron fluence Φ [2, 21], and employing the transient expression of Bates and Korenk [7], we finally have the constitutive equation of irradiation-induced creep and swelling [16, 18]:

$$\dot{\varepsilon}_{Dij}^{IIC} = (1/3)\dot{S}\delta_{ij} + (3/2)P\phi\sigma_{Dij} \tag{10a}$$

$$\dot{S} = C[1 - e^{R(\chi-\Phi)}/\{1 + e^{R(\chi-\Phi)}\}]\langle 1 + Q\sigma_{kk}\rangle\phi \tag{10b}$$

where C, P, Q and R are material functions of ϕ and Φ, and χ stands for the incubation fluence for swelling. Where symbol $\langle\ \rangle$, furthermore, represents the Macauley bracket.

3.2. Mechanical modeling of irradiation-affected thermal creep [16, 18]

Irradiation-affected thermal creep is brought about by the ordinary thermal creep affected by the irradiation effects. The thermal creep of polycrystalline metals at elevated temperatures, on the other hand, is induced by the diffusion controlled dislocation motion activated by stress, and is usually accompanied by creep damage; the creep damage occurs as a result of the nucleation and growth of grain boundary cavities brought about mainly by the grain boundary sliding [22].

By examining the results of creep tests under irradiation [2, 6, 23], and by comparing the results with those of the ordinary unirradiated tests, the effects of the concurrent neutron irradiation on creep may be

1) acceleration and deceleration of dislocation motion as a result of the nucleation of point defects, void clusters, dislocation networks and the precipitation of inclusions.

2) deceleration of nucleation and growth of grain boundary voids and the delay of the onset of the tertiary creep [6], probably due to the supply of interstitial atoms from rich source of point defects.

These features imply that the mechanisms of thermal creep and creep damage *per se* are not changed by the irradiation, and the rates of creep and damage will be subject to quantitative change depending on the neutron flux ϕ and the neutron fluence Φ.

Thus, if we postulate the Kachanov-Rabotnov creep-damage theory [24] together with the strain-hardening hypothesis of McVetty type creep law [25], and assume that the effects of irradiation on the constitutive and the evolution equations can be described through the change of their material constants, we have [16, 18]

$$\dot{\varepsilon}_{ij}^{IIC} = (3/2)A^{*}\sigma^{*}e^{-\alpha t^{*}}\sigma_{EQ}^{n^{*}-1}\sigma_{Dij} + (3/2)A(\phi,\Phi)[\sigma_{EQ}/(1-D)]^{n(\phi,\Phi)}\sigma_{Dij}/(1-D) \tag{11a}$$

$$\dot{D} = B(\phi,\Phi)[\sigma^{(1)}/(1-D)]^{k(\phi,\Phi)} \tag{11b}$$

where $\sigma^{(1)}$ and σ_{EQ} are the maximum principal stress and the equivalent stress. The symbols A, B, n and k are material functions of ϕ and Φ, while A^*, σ^* and n^* are material constants for the transient creep law [25]. The symbol D in eq.(11), furthermore, is an internal state variable describing damage state (damage variable), and defined as follows:

$D = 0$, at $t = 0$ (initial undamaged state)
$D = 1$, at $t = t_R$ (final ruptured state) (12)

According to Gilbert and Chin's experimental data for 20% cold worked type 316 stainless steel at 650°C [5] (see Fig. 3 below), no significant difference is observed in the stress-dependence of rupture times among unirradiated creep, creep under irradiation and post-irradiation creep; this implies

$$k(\phi,\Phi) = k_0 \tag{13a}$$

As regards the creep exponent, the results observed in Reference [2] allows us to assume

$$n(\phi,\Phi) = n_0 \tag{13b}$$

Since the material function $A(\phi, \Phi)$ is related to the creep rate, it depends on the behavior of irradiation-induced internal defects. The effects of fluence Φ on the microstructure of the material is cumulative, and develops dislocation structures. The dislocation density and configuration will tend to an asymptotic state as Φ increases [2, 20]. The effect of ϕ , on the other hand, is brought about by the interaction of the irradiation-induced defects with the existing defects [26]. Namely such interaction is governed by the rate of increase of point defects induced by the irradiation together with the rate of their annihilation due to their absorption by the existing sinks, such as point defect clusters, dislocation networks, etc. Thus, sufficiently large value of ϕ again leads to a stationary defect structures [26].

Furthermore, the function $B(\phi, \Phi)$ is related to the rate of creep damage. The creep damage in polycrystalline metals is brought about by the formation and the development of microcavities on the grain boundaries, which are governed mainly by the local diffusion of atoms and the matrix creep [22]. The effect of Φ on the material is embrittlement due to the formation of point defect clusters and the sessile dislocation networks. The effect of ϕ, on the other hand, supplies interstitial atoms, which in turn decelerate the formation of grain boundary cavities and retards the material damage. Increase in both of ϕ and Φ, therefore, will lead to an asymptotic state of damage rates. In view of these microstructural mechanisms, we will assume the following functions for A and B:

$$A(\phi,\Phi) = A_0[1 + a_1(1 - e^{-a_2\phi})][1 + a_3(1 - e^{-a_4\Phi})] \tag{13c}$$

$$B(\phi,\Phi) = B_0[1 + b_1(1 - e^{-b_2\phi})][1 + b_3(1 - e^{-b_4\Phi})] \tag{13d}$$

where a_1, a_2, a_3, b_1, b_2, b_3 and b_4 are constants that can be identified by comparing the creep curves of unirradiated creep, creep under irradiation and of post-irradiation creep under constant stress.

3.3. Constitutive equation of creep, swelling and damage under irradiation

Summarizing the results of the preceding discussions, the constitutive equations of irradiation creep and swelling are given by eqs. (10), (11) and (13) as follows [16, 18]:

$$\dot{\varepsilon}_{ij}^{C} = \dot{\varepsilon}_{ij}^{IIC} + \dot{\varepsilon}_{ij}^{IIC}$$
$$= (3/2)A^{*}\sigma^{*}e^{-at^{*}}\sigma_{EQ}^{n^{*}-1}\sigma_{Dij} \tag{14a}$$
$$+ (3/2)A_0[1+a_1(1-e^{-a_2\phi})][1+a_3(1-e^{-a_4\Phi})][\sigma_{EQ}/(1-D)]^{n_0-1}[\sigma_{Dij}/(1-D)]$$

$$\dot{D} = B_0[1+b_1(1-e^{-b_2\phi})][1+b_3(1-e^{-b_4\Phi})][\sigma^{(1)}/(1-D)]^{k_0} \tag{14b}$$

$$\dot{S} = C[1-e^{R(\chi-\Phi)}/(1+e^{R(\chi-\Phi)})]\langle 1+Q\sigma_{kk}\rangle\phi \tag{14c}$$

By taking $\phi = \Phi = 0$, eq. (14) is reduced to the thermal creep under unirradiated condition, while the cases of $\phi \neq 0$, $\Phi = \int \phi dt$ and $\phi = 0$, $\Phi \neq 0$ account for the creep under irradiation and the post-irradiation creep, respectively. Thus eq. (14) furnishes a unified constitutive equation for unirradiated creep, creep under irradiation and for post-irradiation creep.

3.4. Analysis of creep under irradiation of 20% cold-worked type 316 stainless steel

To evaluate the validity and the applicability of eq. (14), we will now perform the following analyses for 20% cold worked type 316 stainless steel at 650°C.

Material constants in eq. (14) can be identified by the procedure discussed in the preceding papers [17, 18], and can be determined by use of the experimental results of References [5, 6, 20, 21, 23]. The material constants of eq. (14) thus determined are as follow:

$$A_0 = 3.20\times10^{-13}\text{MPa}^{-n_0}\cdot\text{hr}^{-1},\ n_0 = 3.50,\ B_0 = 6.00\times10^{-10}\text{MPa}^{-k_0}\cdot\text{hr}^{-1},\ k_0 = 2.80$$
$$A^{*} = 1.30\times10^{-7}\text{MPa}^{-n^{*}},\ a^{*} = 8.50\times10^{-3}\text{hr}^{-1},\ n^{*} = 2.00 \tag{15a}$$

$$\chi = 5.00\times10^{22}\,\text{n}\cdot\text{cm}^{-2},\ C = 4.00\times10^{-25}(\text{n}\cdot\text{cm}^{-2}\cdot\text{hr})^{-1}\cdot\text{hr}^{-1},$$
$$Q = 4.75\times10^{-3}\text{MPa}^{-1},\ R = 1.25\times10^{-22}(\text{n}\cdot\text{cm}^{-2})^{-1} \tag{15b}$$

$$a_1 = 5.00\times10^{-2},\ a_2 = 2.60\times10^{-19}(\text{n}\cdot\text{cm}^{-2}\cdot\text{hr})^{-1},\ b_1 = -0.95,$$
$$b_2 = 4.5\times10^{-19}(\text{n}\cdot\text{cm}^{-2}\cdot\text{hr})^{-1},\ P = 2.50\times10^{-28}(\text{MPa}\cdot\text{n}\cdot\text{cm}^{-2}\cdot\text{hr})^{-1} \tag{15c}$$

$$a_3 = -9.00\times10^{-2},\ a_4 = 2.60\times10^{-21}(\text{n}\cdot\text{cm}^{-2})^{-1},$$
$$b_3 = 1.30,\ b_4 = 2.60\times10^{-21}(\text{n}\cdot\text{cm}^{-2})^{-1} \tag{15d}$$

where the material constants of eqs. (15a) through (15d) were identified by the test results of unirradiated creep, swelling, creep under irradiation and postirradiation creep, respectively.

Fig. 1 shows the results of prediction of swelling for 20% cold worked type 316 stainless steel at 400°C. The solid lines and the symbols are the predictions of eq. (14) and the corresponding experimental results of Porter, Takata and Wood [21]. The calculations were performed by use of material constants related to stainless steel at 650°C, and the material constants of eq.(15) were modified by use of temperature dependence of swelling identified by the experiments [5].

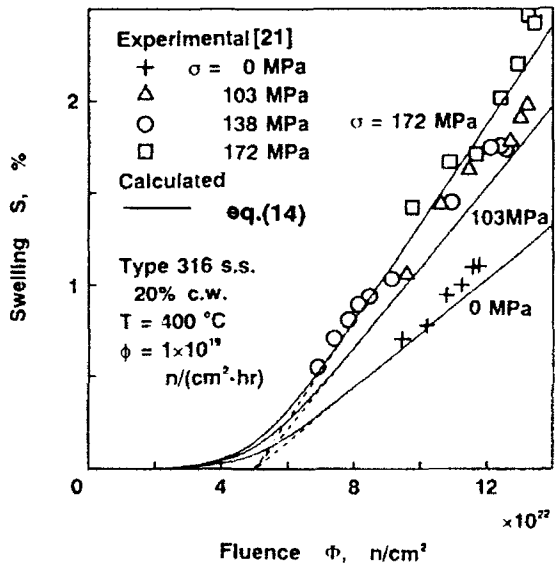


Fig. 1 Swelling of 20% cold worked type 316 stainless steel at 400°C

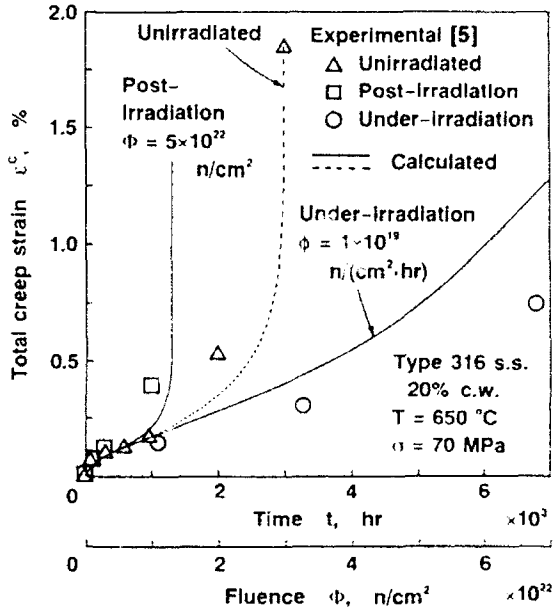


Fig.2 Creep curves of unirradiated, post-irradiation and under-irradiation creep for 20% cold worked type 316 stainless steel

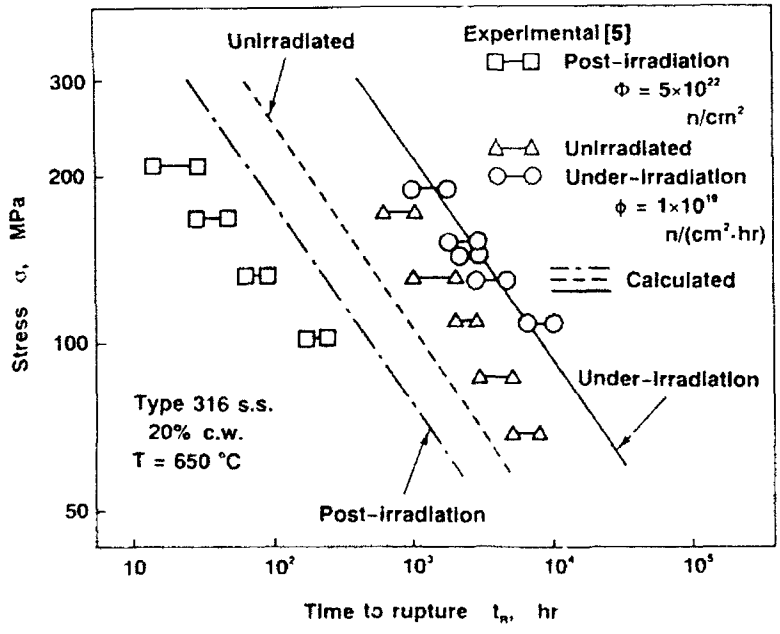


Fig. 3 Stress rupture properties of unirradiated, post-irradiation under-irradiation creep for 20% cold worked type 316 stainless steel

As shown in Fig.1, eq. (14) represents the continuous transition from the incubation to the steady-state period without changing the characteristic of incubation and steady-state periods. Fig.1 suggests the validity of eq. (14).

Figs. 2 and 3 show the comparison between the experimental results [5] for unirradiated creep, post-irradiation creep and creep under irradiation, and the corresponding predictions of eq. (14). The symbols in the figures stand for the results of experiment, and the temperature, stress and irradiation conditions are entered in the figures.

Fig. 2, to begin with, represents the creep curves for different conditions of irradiation. It will be observed that the constitutive equation (14) describes accurately the essential features of irradiation creep; i.e., creep curves under irradiation including transient creep stage, embrittlement due to prior irradiation, and the retardation of the start of the tertiary creep and increase of tensile ductility due to neutron flux. Fig. 3, on the other hand, shows the stress-rupture time relations for unirradiated, post-irradiation and under-irradiation creep, predicted by eq. (14). It will be observed from Fig.3 that eq. (14) describes the dependence of rupture times on the irradiation conditions observed in the corresponding experiments.

4. ELABORATION OF CONSTITUTIVE EQUATIONS FOR CONSIDERABLE VARIATION OF STRESS STATES

4.1 Elaboration based on creep-hardening surface model of variable stress creep [19]

Components of nuclear structures are often subject to large change of stress states, not only in their magnitude but also in their direction. The first wall of fusion reactors, for example, is subject to reversed thermal stress due to the intermittent burn of plasma. This change in stress induces significant transient increase of creep rates, and has a large influence on the initiation and the growth of creep cracks in the materials. The irradiation-affected thermal creep of eq.(11) is based on the strain-hardening theory of creep, and cannot be applied adequately to the problems of large stress variation [25]. Thus we will extend the creep-hardening surface model of variable stress creep [27] to include the effects of irradiation and material damage.

In the case of the classical strain-hardening theory, in particular, the thermal creep rate $\dot{\varepsilon}_{ij}^{TC}$ is expressed as a function of stress and the history of creep strain ε_{ij}^{TC} as follows [27]:

$$\dot{\varepsilon}_{ij}^{TC} = (3/2) f(\sigma_{EQ};\, q)(\sigma_{Dij}/\sigma_{EQ}), \qquad \dot{q} = (2/3)^{1/2} (\dot{\varepsilon}_{ij}^{TC} \dot{\varepsilon}_{ij}^{TC})^{1/2} \tag{16a,b}$$

where σ_{EQ} and σ_{Dij} denote the equivalent stress and the deviatoric stress, respectively. The symbol q in eq.(16a), furthermore, is a creep-hardening variable (i.e. an internal state variable) representing the state of creep-hardening of the material.

The conventional theory represented by eqs.(16a,b), however, cannot describe the transient increase of creep rate which is observed after stress reversals or change in stress direction [27], because the creep-hardening variable q cannot adequately describe the recovery of material hardening after the stress change. In order to overcome this difficulty, Murakami and Ohno [27] introduced a hyper-sphere (creep-hardening surface) in a creep strain space as follows:

$$g = (2/3)(\varepsilon_{ij}^{TC} - \alpha_{ij})(\varepsilon_{ij}^{TC} - \alpha_{ij}) - \rho^2 \le 0 \tag{17a}$$

where α_{ij} and ρ denote center and radius of the hyper-sphere, respectively. The evolution equations of these variables are given as follows:

$$\dot{\alpha}_{ij} = \begin{cases} (1-\lambda)\dot{\varepsilon}_{kl}^{TC} n_{kl} n_{ij} & g = 0 \text{ and } (\partial g / \partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} > 0 \\ 0 & g < 0 \text{ or } (\partial g / \partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} \le 0 \end{cases} \tag{17b}$$

$$\dot{\rho} = \begin{cases} (2/3)^{1/2}\lambda \dot{\varepsilon}_{ij}^{TC} n_{ij} & g = 0 \text{ and } (\partial g / \partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} > 0 \\ 0 & g < 0 \text{ or } (\partial g / \partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} \le 0 \end{cases} \tag{17c}$$

$$n_{ij} = (\varepsilon_{ij}^{TC} - \alpha_{ij}) / [(\varepsilon_{kl}^{TC} - \alpha_{kl})(\varepsilon_{kl}^{TC} - \alpha_{kl})]^{1/2} \tag{17d}$$

where λ is a material constant specifying the rate of development of ρ. Thus, the creep-hardening variable q representing the state of creep hardening under a given deformation history is specified by use of α_{ij} and ρ in eq.(17) as follows:

$$q = [1/(2\lambda)][\rho + (\varepsilon_{ij}^{TC} - \alpha_{ij})(\sigma_{Dij} / \sigma_{EQ})] \tag{18}$$

If we employ the Bailey-Norton creep equation of constant uniaxial stress

$$\varepsilon^{TC} = A\sigma^n t^m \tag{19a}$$

and postulate the strain-hardening hypothesis, we have the following equation of multiaxial creep [27]:

$$\dot{\varepsilon}_{ij}^{TC} = (3/2)mA^{1/m}\sigma_{EQ}^{n/m} q^{(m-1)/m}(\sigma_{Dij} / \sigma_{EQ}) \equiv (3/2) f(\sigma_{EQ};\ q)(\sigma_{Dij} / \sigma_{EQ}) \tag{19b}$$

where A, n and m denote material constants.

In view of the concept of continuum damage mechanics, the damage state of material under creep can be expressed by a damage variable D. If we employ the creep damage theory of Kachanov-Rabotnov, eq. (19b) can be extended to incorporate the effects of creep damage as follows:

$$\dot{\varepsilon}_{ij}^{TC} = (3/2)mA^{1/m} q^{(m-1)/m}[\sigma_{EQ}/(1-D)]^{(n-m)/m}[\sigma_{Dij}/(1-D)], \quad \dot{D} = B[\sigma^{(1)}/(1-D)]^k \tag{20a,b}$$

where $\sigma^{(1)}$ denotes the largest principal stress, and B and k are material constants. The hardening parameter in eq. (20a) can be given by eqs. (17) and (18).

Since the essential mechanisms of irradiation-affected thermal creep are unchanged from those of the creep and damage under unirradiated condition, the constitutive equation of irradiation-affected thermal creep ε_{ij}^{ITC} and the related evolution equation of creep damage can be provided from eqs. (17),(18),(20) by replacing the coefficients A, m, n B, k and λ with the corresponding material function of ϕ and Φ.

To recapitulate eqs. (14), (17)-(19) and the above argument, the constitutive equations of creep, swelling and damage under irradiation under multiaxial and variable states of stress are expressed as follows:

$$\dot{\varepsilon}_{ij}^{C} = \dot{\varepsilon}_{ij}^{IIC} + \dot{\varepsilon}_{ij}^{ITC}$$
$$= (1/3)\dot{S}\delta_{ij} + (3/2)P\phi\sigma_{Dij}$$
$$+ (3/2)m_0 A(\phi,\Phi)^{1/m_0} q^{(m_0-1)/m_0}[\sigma_{EQ}/(1-D)]^{(n_0-m_0)/m_0}[\sigma_{Dij}/(1-D)] \tag{21a}$$

$$q = [1/(2\lambda_0)][\rho + (\varepsilon_{ij}^{TC} - \alpha_{ij})(\sigma_{Dij}/\sigma_{EQ})] \tag{21b}$$

$$\dot{\alpha}_{ij} = \begin{cases} (1-\lambda_0)\dot{\varepsilon}_{kl}^{TC} n_k n_{ij} & g = 0 \text{ and } (\partial g/\partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} > 0 \\ 0 & g < 0 \text{ or } (\partial g/\partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} \le 0 \end{cases} \tag{21c}$$

$$\dot{\rho} = \begin{cases} (2/3)^{1/2}\lambda_0 \dot{\varepsilon}_{ij}^{TC} n_{ij} & g = 0 \text{ and } (\partial g/\partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} > 0 \\ 0 & g < 0 \text{ or } (\partial g/\partial \varepsilon_{ij}^{TC})\dot{\varepsilon}_{ij}^{TC} \le 0 \end{cases} \tag{21d}$$

$$n_{ij} = (\varepsilon_{ij}^{TC} - \alpha_{ij})/[(\varepsilon_{kl}^{TC} - \alpha_{kl})(\varepsilon_{kl}^{TC} - \alpha_{kl})]^{1/2} \tag{21e}$$

$$\dot{D} = B(\phi,\Phi)[\sigma^{(1)}/(1-D)]^{k_0} \tag{21f}$$

$$\dot{S} = C[1 - e^{R(\chi-\Phi)}/(1+e^{R(\chi-\Phi)})](1+Q\sigma_{kk})\phi \tag{21g}$$

where $A(\phi,\Phi)$ and $B(\phi,\Phi)$ are again given by eqs. (13c) and (13d).

4.2. Analysis of creep under irradiation subject to variable states of stress

The material constants of eqs. (21) can be determined by the procedures similar to that of Section 3.4 and Reference [27]. The material constants were identified for the 20% cold worked type 316 stainless steel at 650°C and were given in Reference [19].

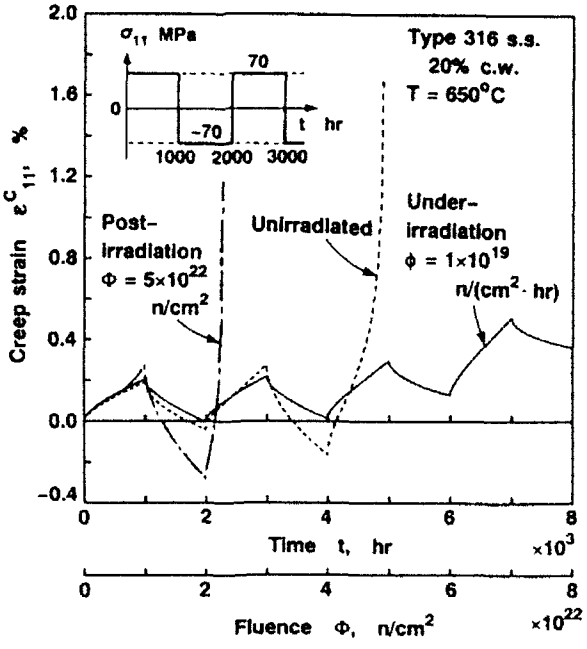


Fig. 4 Creep strain ε_{11}^{c} under uniaxial stress reversals between σ_{11} = 70 MPa and −70MPa with an interval of 1000 hours

Figure 4 shows the predictions of eq.(21) for the creep subjected to uniaxial stress reversals under three different conditions of irradiation. The stress changes between σ_{11}=70MPa and −70MPa with the intervals of 1000 hours as shown in the figure. It will be observed that eq. (21) describes well the transient increase of creep rate after stress reversals under each irradiation condition. Conventional strain-hardening theory expressed in eq. (16), in general, cannot describe this increase in creep rate, and accordingly underestimates the relaxation of stress. It means that conventional strain-hardening theory predicts smaller redistribution of stress in the materials.

Rupture time under stress reversal conditions in Fig.4, however, becomes about twice as long as the rupture time under the constant stress conditions of Fig.1. It is because the maximum principal stress, which governs the process of creep damage [see evolution equation of creep damage (21f)] is zero when stress is negative σ_{11}=−70MPa. Noting the behavior of creep under irradiation, the absolute value of the creep rate under positive stress σ_{11}=70MPa is found to be larger than that under negative stress σ_{11}=−70MPa for neutron fluence larger than $\Phi = 5\times10^{22}\mathrm{n\cdot cm^{-2}}$, and this can be accounted for by the swelling at a constant rate after the incubation period. Thus, creep under irradiation is found to increase with stress cycles in contrast to the creep under other irradiation conditions.

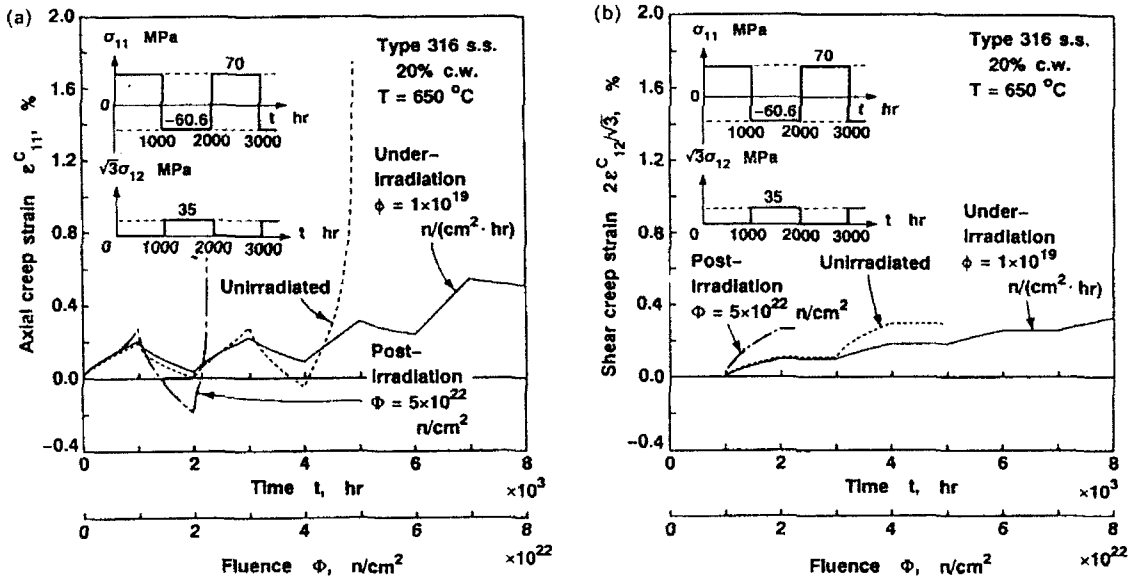


Fig. 5 Creep strain under change of stress direction between 0° and 150° at the interval of 1000 hours. (a) Axial creep strain ε^c_{11}. (b) Shear creep strain $2\varepsilon^c_{12}/(3)^{1/2}$.

Finally, Figs. 5 (a), (b) show the axial and shear creep strain predicted by eq. (21) under different irradiation conditions brought about by the multiaxial and non-proportional loading. The alternative stress change between 0° and 150° in $\sigma_{11}-(3)^{1/2}\sigma_{12}$ stress plane under constant equivalent stress of 70MPa at intervals of 1000 hours are shown in the figure. Unfortunately,

there are no experimental data corresponding to Figs.4 and 5 under post-irradiation and irradiation conditions to confirm the validity of eq. (21).

5. ANALYSIS OF CREEP CRACK GROWTH UNDER NEUTRON IRRADIATION

As an application of the constitutive equations developed so far, we will elucidate the effects of neutron irradiation on creep crack growth observed in nuclear reactor components. For this purpose, the constitutive equation (14) was incorporated into a computer code of finite element analysis (FEM), and the creep crack growth in a plate of 20% cold worked stainless steel at 650°C under unirradiation, post-irradiation and irradiation conditions was analyzed [28] by a local approach to fracture in continuum damage mechanics. The material was assumed to show elastic-creep behavior. When the damage variable D of an element calculated from eq. (14b) has attained to a critical value, the element is assumed to be ruptured. The aggregation of these ruptured elements is identified as a macroscopic crack.

The mesh division by simplex elements employed in the analysis is shown in Fig. 6. The plate is subject to the constant distributed stress σ_0 in the direction perpendicular to the initial crack, and was assumed to be in a state of plane stress. Number of elements and nodes in Fig. 6 are 485 and 286, respectively. The material constants of eq.(15) together with the following value of the modulus of rigidity were employed:

$$G=54.8\text{GPa} \tag{22}$$

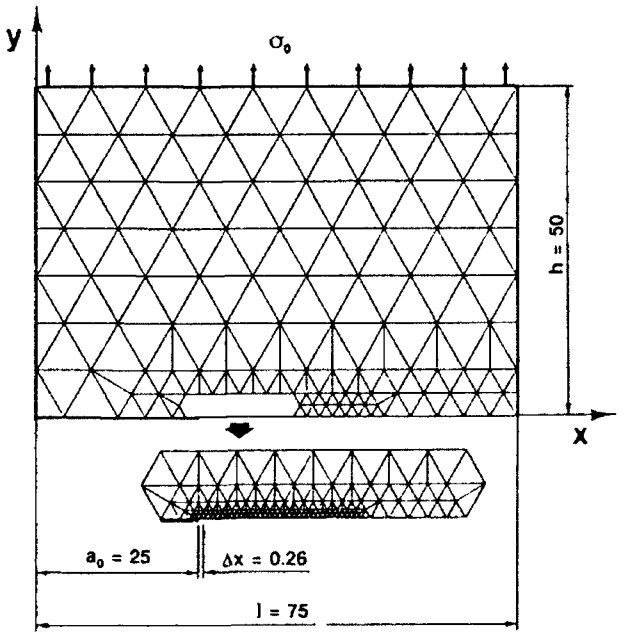


Fig. 6 Finite element discretization of a cracked plate subject to uniform tension under neutron irradiation

Fig.7 shows the relation between the incipient creep crack growth rate v and the tensile stress σ_0 under different conditions of irradiation entered in the figure. The conditions of irradiation correspond to a fuel cladding of fast breeder reactors. The corresponding results of a simplified analysis by a double cantilever beam (DCB) model [29] were also entered in the figure by dashed lines. As observed in the results of FEM analysis, the crack growth rate under neutron irradiation is about 1/20 of that of unirradiated condition. The crack growth rate of the post-irradiated material, on the other hand, is 5 times as large as the unirradiated condition. It should be noted that though the creep crack rate is suppressed by the irradiation and the crack growth under irradiation looks more stable, it will be accelerated by about 100 times once the neutron flux vanishes. The creep crack growth rates predicted by DCB model is 2-4 times larger than those of FEM analysis, but shows similar results.

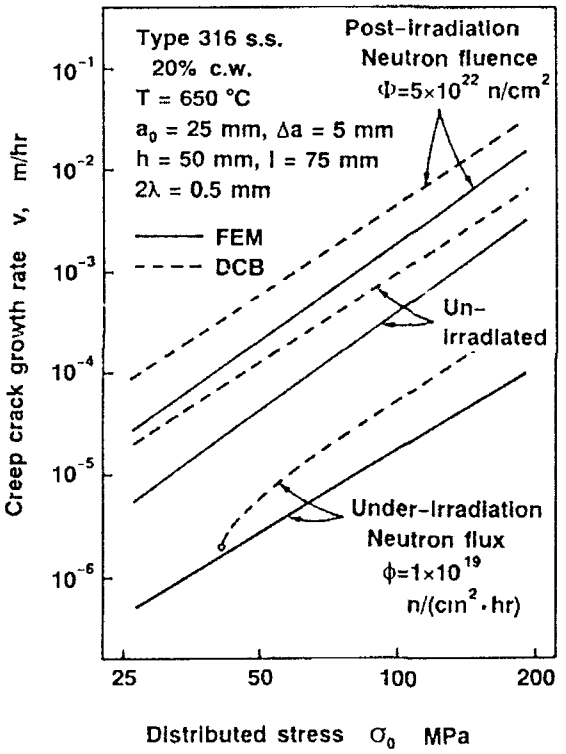


Fig. 7 Relation between tensile stress and crack growth velocity under different conditions of irradiation

6. CONCLUSIONS AND REMARKS

To facilitate engineering design of nuclear reactor components under intensive irradiation, especially of future fusion reactors, constitutive models unifying irradiation creep, swelling and thermal creep should be developed on the sound bases of physics and continuum mechanics. In the present paper, a constitutive equation for this purpose was developed by dividing the creep

rate into irradiation-induced creep and irradiation-affected thermal creep, with special emphasis on austenitic stainless steels. The resulting constitutive equation was applied to analyze the creep crack growth under neutron irradiation to elucidate the effects of concurrent and prior irradiation on creep crack growth rates.

Though a great deal of results have been accumulated so far in the physical and metallurgical investigation on the mechanical behavior of materials under neutron irradiation, they hardly have a relation with the current development of the theories of plasticity, viscoplasticity and damage. To narrow the gap between the present knowledge of irradiation creep and the current state of design technology developed so far for fusion reactors, the collaboration of continuum mechanics and metallurgical physics will be able to do a significant contribution.

7. REFERENCES

1. J. Gittus, Irradiation Effects in Crystalline Solids, Applied Science Publishers, London, 1978.
2. K. Ehrlich, J. Nucl. Mater., 100(1981)149.
3. J. R. Matthews and M. W. Finnis, J. Nucl. Mater., 159(1988)257.
4. F. A. Garner and D. S. Gelles, J. Nucl. Mater., 159(1988)286.
5. E. R. Gilbert and B. A. Chin, Nucl. Technol., 52(1981)273.
6. E. R. Gilbert, B. A. Chin and D. R. Duncan, Metall. Trans. A, 18A(1987)79.
7. J. Bates and M. K. Korenko, Nucl. Technol., 48(1980)303.
8. M. L. Grossbeck, K. Ehrlich and C. Wassilew, J. Nucl. Mater., 174(1990)264.
9. G. S. Ansell and J. Weertman, Trans. AIME, 215(1959)838.
10. J. H. Gittus, Philos. Mag., 25(1972)345.
11. J. H. Gittus, Philos. Mag., 28(1973)261.
12. P. T. Heald and M. V. Speight, Philos. Mag., 29(1974)1075.
13. P. T. Heald and J. E. Harbottle, J. Nucl. Mater., 67(1977)229.
14. A. D. Brailsford and R. Bullough, Philos. Mag., 27(1973)49.
15. A. Boltax, J. P. Foster, R. A. Weiner and A. Biancheria, J. Nucl. Mater., 65(1977)174.
16. S. Murakami and M. Mizuno, Creep in Structures, Proc. IUTAM Symposium on Creep in Structures(ed. by M. Zyczkowski), Springer-Verlag Berlin (1991)237.
17. S. Murakami, M. Mizuno and T. Okamoto, Nucl. Engng Design, 131(1991)147.
18. S. Murakami and M. Mizuno, Nucl. Technol., 95(1991)219.
19. S. Murakami and M. Mizuno, Int. J. Solids Struct., 29(1992)2319.
20. W. G. Wolfer, J. Nucl. Mater., 90(1980)175.
21. D. L. Porter, M. L. Takata and E. L. Wood, J. Nucl. Mater., 116(1983)272.
22. H. E. Evans, Mechanisms of Creep Fracture, Elsevier Applied Science Publishers, London, 1984.
23. E. E. Bloom and J. R. Weir, J. Nucl. Techol., 66(1972)45.
24. F. A. Leckie and D. R. Hayhurst, Acta Metall., 25(1977)1059.
25. H. Kraus, Creep Analysis, John-Wiley, New York, 1980.
26. M. Kiritani, J. Nucl. Mater., 169(1989)89.
27. S. Murakami and N. Ohno, Int. J. Solids Struct., 18(1982)597.
28. S. Murakami and M. Mizuno, Mechanical Behaviour of Materials-VI, Proc. the 6th Int. Conf. Mech. Behaviour of Mater., Vol.3, ed. by M. Jono and T. Inoue, Pergamon Press, Oxford, 1991, 853.
29. S. Murakami, K. Hayakawa and M. Mizuno, Eur. J. Mech., A/Solid, 11(1992) (in press).

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

STABILITY AND BIFURCATION IN DISSIPATIVE MEDIA

Nguyen Quoc Son

Laboratoire de Mécanique des Solides, Centre commun X - Mines - Ponts - CNRS
Ecole Polytechnique, 91128 - Palaiseau, France

Abstract

This lecture addresses energy-related aspects of anelasticity. Anelasticity also means irreversibility and dissipation. To study the behaviour of irreversible media (anelastic materials and structures), an interesting and important direction is to consider their description in the general framework of thermodynamics. Considerable progress has been achieved and various approaches are proposed in this direction. The most usual one is based on a systematic description of irreversible systems via the expressions of two potentials : the free energy and the dissipative potential. This approach is associated with the work of several authors in different domains such as metal plasticty, friction, fracture and damage mechanics. Our objective is to present this description in relation with common and practical models of anelasticity. Some general results of the literature concerning the quasi-static response of irreversible systems are presented. In particular, when the behaviour is time-independent (e.g. plasticity, contact with dry friction, brittle fracture or brittle damage), general results concerning stability and bifurcation analysis can be expressed in an unified framework. This framework will be presented and illustrated by recent examples.

1. INTRODUCTION

Thermodynamics provide the largest framework for the description of the irreversible behaviour of anelastic media (materials or structures). In this direction, considerable progress has been achieved and various approaches have been proposed. For example, the functional approach found in Coleman & Gurtin's work has been successfully applied to visco-elasticity in the fifties. In the last two decades, significant results have been however obtained within the so-called internal variable theory. In particular, the introduction of internal state variables can be complemented by a systematic description of the material behaviour via the expressions of two potentials : the free energy and the dissipation potential (cf. [1-3] for example). This approach is associated with the works of several authors, in particular Biot, Mandel [4-5] in visco-elasticity, Moreau, Mandel, Rice, Halphen & Nguyen .. [6-9] in plasticity and visco-plasticity, Lemaitre & Chaboche [10] in damage mechanics. ..etc. It provides a systematic and general framework for the description of usual laws and for the mechanical modeling of new phenomena. It is now widely accepted and becomes more and more popular for the description of anelasticity.

Our objective is to present this approach and to underline some general results concerning the quasi-static evolution of irreversible systems. In particular, when the behaviour is time-independent as it is in the context of plasticity, contact with dry friction, brittle fracture and brittle damage, general results concerning stability and bifurcation analysis can be expressed in an unified framework. These results are illustrated by some recent examples borrowed from the literature.

2. THERMODYNAMIC FRAMEWORK

The notions of state variables, state equations, free energy, complementary laws and dissipation potential D are here assumed for all dissipative systems. This approach leads to the definition of specific models of materials and structures denoted as standard materials and standard systems.

2.1. Standard materials

A material can be described by states variables $q = (\epsilon,\alpha)$ where ϵ is a deformation and α denotes internal parameters . By unit reference volume, the associated free energy is $W(\epsilon,\alpha)$. The dissipation is :

$$d = \sigma.\dot{\epsilon} - \dot{W} \tag{1}$$

where σ denote the dual stress associated to ϵ. Dissipation represents a product of fluxes and forces :

$$d = (\sigma - \sigma_R).\dot{\epsilon} + A.\dot{\alpha} \tag{2}$$

where σ_R and A denote the associated forces :

$$\sigma_R = W_{,\epsilon} \quad , \quad A = - W_{,\alpha} \tag{3}$$

To simplify the presentation, it will be accepted here that the deformation is not a dissipative mechanism, all dissipative mechanisms are taken into account by α. It follows then from (2) that :

$$\sigma = \sigma_R \tag{4}$$

To complete the constitutive equations, complementary relations between forces and fluxes must be introduced. These complementary laws of evolution can be conveniently expressed under the form :

$$A = D_{,\dot{\alpha}} \tag{5}$$

or under the dual form :

$$\dot{\alpha} \in D^*_{,A} \tag{6}$$

where $D(\dot{\alpha},q)$ denotes the dissipation potential , a function of $\dot{\alpha}$ depending eventually on the present state q, D^* is obtained from D by Legendre transform. If the dissipation potential is assumed to be **convex**, the notion of differentiation in (5) can be understood in the sense of sub-gradients, cf. Moreau [6], and this enables us to include in the same framework the important case of non-differentiable dissipation potentials associated with time-independent dissipative processes.

Finally, the constitutive equations of a standard material are described by the expressions of two potentials W and D.

This description is sufficiently large to include most models of anelasticity :

1 For example, in Maxwell's model of linear visco-elasticity, state variables are (ϵ,ϵ^v), free energy $W(\epsilon,\epsilon^v) = \frac{1}{2}(\epsilon-\epsilon^v).L.(\epsilon-\epsilon^v)$ and $D = \frac{1}{2}\dot{\epsilon}^v.N.\dot{\epsilon}^v$.

If ϵ^v is replaced by ϵ^p and $D = k\,|\dot{\epsilon}^p|$ while w is unchanged, relation (5) gives :

$$\sigma = k\,\frac{\dot{\epsilon}^p}{|\dot{\epsilon}^p|} \quad \text{if } \dot{\epsilon}^p \neq 0 \text{ and } |\sigma| \leq k \quad \text{if } \dot{\epsilon}^p = 0$$

The obtained model is the elastic-perfectly plastic material with Mises criterion and normality law :

$$f(\sigma) = |\sigma| - k \leq 0 \;,\; \dot{\epsilon}^p = \mu\, f_{,\sigma} \quad \text{with } \mu \geq 0,\; f \leq 0 \;,\; \mu f = 0 \,.$$

In the same spirit, the model can be extended into finite strain with the decomposition $F = \nabla u = E.P$ (cf. [7, 11]) :
Mandel's model of perfect plasticity consists in choosing $\epsilon = F$ and $\alpha = P$, $W = W(E) = W(F.P^{-1})$. In this case, $\sigma = W_{,F}$ is the unsymmetric Piola-Kirchoff stress tensor and $A = -W_{,P} = \sigma.E$. The plastic criterion is $f(\psi) \leq 0$, ψ denotes the stress tensor $\psi = \det P\; P.A$ and the flow rule is $\dot{P} = \mu\, f_{,A}$ with $f = f(\det P\; P.A)$.

Standard models of plasticity have been much discussed at small strain. Most models with associated flow rules are standard models. For example, Chaboche's model of cyclic plasticity [10] is an extension of Ziegler-Prager or Mroz's model of combined isotropic and kinematic hardening [12]. Camclay model is a popular standard model in soil mechanics. The notion of visco-plastic potential $D^*(\sigma)$ has been discussed in particular by Rice [8]. Standard models of damage mechanics are considered in [10] etc....

It is important to underline that the description of standard models is stable with respect to the choice of state variables and with the expression of forces and fluxes.

For example, if an arbitrary regular change of variable $\beta = \beta(\alpha)$ is introduced, the new associated force is $B = -w_{,\beta} = A\,.\,\alpha_{,\beta}$ while $\dot{\beta} = \beta_{,\alpha}\,.\,\dot{\alpha}$. It is clear that $A.\dot{\alpha} = B.\dot{\beta}$ and all the principal ingredients of the model (free energy, forces, dissipative potential, convexity) remain available.

Forces A and rates $\dot{\alpha}$ can also be transported by convection to a different configuration to new forces $B = L(q).A$ and new fluxes $b = L(q)^{T^{-1}}.\dot{\alpha}$ where L(q) denotes a state-dependent linear operator. The dissipation potential D can be then written in terms of new fluxes as it has been proposed in different models of plasticity at finite strain.

If $\epsilon = \lambda(t)$ is a given function of time, the following equations are obtained :

$$A = -W_{,\alpha}(\lambda,\alpha) = D_{,\dot{\alpha}}(\dot{\alpha},\lambda,\alpha). \tag{7}$$

Thus the evolution of internal parameters $\alpha(t)$ satisfies a differential equation with initial condition :

$$W_{,\alpha} + D_{,\dot{\alpha}} = 0 \;,\quad \alpha(0) = \alpha^0 \tag{8}$$

which can also be expressed in an equivalent way as :

$$\begin{aligned} A &= -W_{,\alpha}(\lambda(t),\alpha) \\ \dot{\alpha} &= D^*_{,A}(A,\lambda(t),\alpha) \\ \alpha(0) &= \alpha^0 \end{aligned} \tag{9}$$

2.2. Standard systems

Equations (8) also describes the quasi-static evolution of discrete or continuous structures, cf. Biot [4].

For a system defined by parameters $\mathbf{q} = (\mathbf{u},\alpha)$ where u are reversible parameters, α irreversible parameters, and $\lambda(t)$ are (displacement or force) control parameters, let $\mathbf{W}(\mathbf{q},\lambda)$ be the whole free energy of the system. This energy results from the contribution of reversible energy stored by all different elements.

If the system is dissipative, let us introduce as in the previous section the dissipation potential $\mathbf{D}(\dot{\alpha},\mathbf{q})$ in the sense that :

$$\mathbf{d} = \mathbf{A} \cdot \dot{\alpha} \quad \text{with} \quad \mathbf{A} = \mathbf{D}_{,\dot{\alpha}} \ . \tag{10}$$

The equation describing the dynamic evolution of the system has then the general form :

$$\begin{aligned} &\mathbf{W}_{,\mathbf{q}} + \mathbf{D}_{,\dot{\mathbf{q}}} + \mathbf{F}(\mathbf{q},\lambda) + \mathbf{J} = 0 \\ &\mathbf{q}(0) = \mathbf{q}^0 \ , \ \dot{\mathbf{q}}(0) = \mathbf{p}^0 \end{aligned} \tag{11}$$

where F denotes external forces and J inertia forces.

If the external forces F are conservative, by definition $\mathbf{F} = -\Psi_{,\mathbf{q}}$ where $\Psi(\mathbf{q},\lambda)$ denotes the force potential. It is convenient to include the force potential Ψ in the expression of W (which represents then the **total free energy of the system**) to write the quasi-static equation of evolution in the previous form (8) :

$$\begin{aligned} &\mathbf{W}_{,\mathbf{q}} + \mathbf{D}_{,\dot{\mathbf{q}}} = 0 \\ &\mathbf{q}(0) = \mathbf{q}^0 \end{aligned} \tag{12}$$

3. TIME-INDEPENDENT PROCESSES

3.1. Quasi-static evolution of time-independent processes

We are interested principally in the particular case of time-independent processes. This is an important case since it includes all usual descriptions of dry friction, plasticity, fracture and damage.

Such phenomena are associated with positive homogenous dissipation potential of degree one :

$$\mathbf{D}(m\dot{\mathbf{q}},\mathbf{q}) = m\,\mathbf{D}(\dot{\mathbf{q}},\mathbf{q}) \quad \text{if} \quad m > 0 \ . \tag{13}$$

This function is not differentiable when $\dot{q} = 0$, the set C of sub-gradients at this point is a convex set in the force space A and represents the domain of admissible forces. This domain is state-dependent $\mathbf{C} = \mathbf{C}(\mathbf{q})$ when D depends on the present state q. It can be convenient to represent this domain C by inequalities $f(A,q) \leq 0$.

Conversely, D is recovered from C by Hill's principle of maximum dissipation :

$$\mathbf{D}(\dot{\mathbf{q}},\mathbf{q}) = \max_{\mathbf{A}^* \in \mathbf{C}} \mathbf{A}^* \dot{\mathbf{q}} \tag{14}$$

In the spirit of the previous examples of plasticity, equations (14) can also be written following (11) or (13) as :

$$\begin{aligned} & W_{,u} = O \quad , \\ & A = - W_{,\alpha} \quad , \\ & \dot{\alpha} = N_{C(q)} \ (A) \ , \\ & \dot{q} \ (0) = q^0 \end{aligned} \tag{15}$$

where N_C denotes the operator of normality, N_C (A) is a normal to convex C at point A.

In (15), the first group of equations $W_{,u} = 0$ gives the possibility to eliminate completely the reversible variables u as $u = u(\alpha,\lambda)$ when W,uu is not singular. This yields a more compact equation of evolution for α :

$$\begin{aligned} & A = - W_{,\alpha} \quad , \\ & \dot{\alpha} = N_{C(\alpha,\lambda)} \ (A) \ , \\ & \alpha(0) = \alpha^0 \end{aligned} \tag{16}$$

where $W = W(\alpha,\lambda) = W(u(\alpha,\lambda),\alpha,\lambda)$ is the reduced expression of energy.

3.2. Incremental response

It is well-known that normality law represents an incremental law :

For example, in plasticity, $\dot{\varepsilon}^p = \mu \ f_{,\sigma}$ with $\mu \geq 0$, $f \leq 0$ and $\mu f = 0$. Plastic multiplier μ is an additional unknown but equality $\mu \ f = 0$ provides an additional equation. If r.h.s. derivability is assumed, it follows that $\dot{\mu} \ f + \mu \ \dot{f} = 0$. Thus if the plastic criterion is satisfied, i.e. $f = 0$, the plastic multiplier μ is then given by equations $\mu \geq 0$, $\dot{f} \leq 0$, $\mu \ \dot{f} = 0$. This is the usual way to address elastic-plastic response.

However, it is also possible to express the same result directly from the dissipation potential. The following equivalent equations are derived from normality law and from r.h.s. derivability [13] :

$$- \dot{A}.\dot{\alpha} + \dot{q} \ . \ D_{,q} \ (\dot{\alpha},q) = 0 \tag{17}$$

$$- \dot{A}.\delta\alpha + \dot{q} \ . \ D,q \ (\delta\alpha,q) \geq 0 \tag{18}$$

$\forall \ \delta q = (\delta u,\delta\alpha)$ such that $\delta\alpha$ is an admissible rate i.e. a vector of the normal cone $\delta\alpha \in N_C$ (A) .

If the admissible domain C does not depend on the present state q, equation (17) gives the orthogonal condition $\dot{A}.\dot{\alpha} = 0$ which is for example the well-known relation $\dot{\sigma} \ \dot{\varepsilon}^p = 0$ of perfect plasticity at small strains.

When $\dot{A}$ is computed from $A = - W_{,\alpha}$, equations (17) and (18) lead directly to the following characterization of the rate response :

The rate $\dot{q} = (\dot{u},\dot{\alpha})$ is a solution of the variational inequality :

$$\begin{aligned} & (W_{,uu} \ . \ \dot{u} + W_{,u\alpha} \ . \ \dot{\alpha} + W_{,u\lambda} \ . \dot{\lambda} \ ). \ \delta u = O \\ & (W_{,u\alpha} \ . \ \dot{u} + W_{,\alpha\alpha} \ . \ \dot{\alpha} + W_{,\alpha\lambda} \ . \ \dot{\lambda} \ ). \ (\delta\alpha - \dot{\alpha}) + \\ & \dot{q} \ . \ (\ D_{,q} \ (\delta\alpha,q) - D_{,q} \ (\dot{\alpha},q) \ ) \geq 0 \end{aligned} \tag{19}$$

for all admissible rates $\delta q = (\delta u, \delta\alpha)$.

3.3. Some general results on the incremental response

The incremental response has been studied in the context of plasticity [14-17]... or for contact problems with dry friction [18-19]. Variational inequality (19) can be studied by standard methods of convex analysis [20]. The following remarks are straightforward :

- This variational inequality is not symmetric when D depends on u. If D depends on α, it is symmetric if and only if the term $\delta\alpha \cdot D_{,\dot{\alpha}}(\dot{\alpha},\alpha)$ is symmetric. The dependence of the dissipation potential on the present state is thus a source of mathematical difficulty. It is useful to recall that a non-symmetric variational inequality admits no associated extremum principle.

- Existence of a rate response $\dot{\mathbf{q}}$ associated with an arbitrary implied control rate $\dot{\lambda}$ is ensured by the following positivity condition :

$$\delta\mathbf{q} \cdot \mathbf{W}_{,\mathbf{qq}} \cdot \delta\mathbf{q} + \delta\mathbf{q} \cdot \mathbf{D}_{,\mathbf{q}}(\delta\alpha, \mathbf{q}) > 0 \tag{20}$$

for all admissible rates $\delta\mathbf{q} = (\delta\mathbf{u},\delta\alpha)$.

- Uniqueness of the rate response is ensured if the following inequality is satisfied for any couple of admissible rates $\delta\mathbf{q}^1 \neq \delta\mathbf{q}^2$:

$$(\delta\mathbf{q}^1-\delta\mathbf{q}^2) \cdot \mathbf{W}_{,\mathbf{qq}} \cdot (\delta\mathbf{q}^1-\delta\mathbf{q}^2) + (\delta\mathbf{q}^1-\delta\mathbf{q}^2) \cdot (\mathbf{D}(\delta\alpha^1,\mathbf{q}) - \mathbf{D}(\delta\alpha^2,\mathbf{q})) > 0 \tag{21}$$

- Variational inequality (19) is a mixed formulation in which the principal unknown is $\dot{\mathbf{q}} = (\dot{\mathbf{u}},\dot{\alpha})$. It may be more convenient in some situations to eliminate $\dot{\mathbf{u}}$ or $\dot{\alpha}$ and to retain only $\dot{\alpha}$ or $\dot{\mathbf{u}}$ as principal unknown :

In plasticity for example, in the case of simple plastic potential, the plastic strain rate can be eliminated. The resulting rate displacement problem with unknown $\dot{\mathbf{u}}$ has been discussed in detail [14-17] in small and finite strains. The associated expressions of (20) and (21) for an elastic-plastic solid is, [15] :

$$\int_V \delta\sigma \cdot \nabla\delta\mathbf{u}\, dV + \delta\mathbf{u} \cdot \Psi_{,\mathbf{uu}} \cdot \delta\mathbf{u} > 0 \tag{22}$$

$$\int_V (\delta\sigma^1-\delta\sigma^2).(\nabla\delta\mathbf{u}^1-\nabla\delta\mathbf{u}^2)\, dV + (\delta\mathbf{u}^1-\delta\mathbf{u}^2) \cdot \Psi_{,\mathbf{uu}} \cdot (\delta\mathbf{u}^1-\delta\mathbf{u}^2) > 0 \tag{23}$$

for all admissible rates of displacement.

However, in the case of multiple plastic potential, the plastic criterion is represented by several inequalities and such elimination is not possible. It is thus necessary to retain as principal unknown the couple $\mathbf{q} = (\mathbf{u},\alpha)$.

Another possibility is to eliminate $\mathbf{u}$ and keep α as principal unknown following the reduced formulation (16). Such formulation has been discussed by Maier [21] in plasticity for the study of the quasi-static evolution and of the asymptotic behaviour of elastic-plastic structures such as elastic or plastic shake-down. In fracture mechanics, α represents the crack surfaces or crack lengths, the elimination of displacement $\mathbf{u}$ is also necessary. The expressions of (20) or (21) associated with the reduced formulation (16) are :

$$\delta\alpha \cdot \mathbf{W}_{,\alpha\alpha} \cdot \delta\alpha + \delta\alpha \cdot \mathbf{D}_{,\dot{\alpha}}(\delta\alpha,\alpha,\lambda) > 0 \tag{24}$$

$$(\delta\alpha^1-\delta\alpha^2) \cdot \mathbf{W}_{,\alpha\alpha} \cdot (\delta\alpha^1-\delta\alpha^2) + \tag{25}$$

$$(\delta\alpha^1-\delta\alpha^2) \cdot (D_{,\dot\alpha} (\delta\alpha^1,\alpha,\lambda) - D_{,\dot\alpha} (\delta\alpha^2,\alpha,\lambda)) > 0$$

3.4. Bifurcation and stability analysis

It has been established by Hill [15] that positivity condition (23) is a sufficient condition of non-bifurcation. Bifurcation analysis in plasticity has been extensively studied [22-24] and most analyses are based on Hill's result. It is relatively clear that if two different rates exist then bifurcation occurs eventually. However the possibility of smooth bifurcation, i.e. two responses with same tangent but different curvatures for example, can also be discussed in the light of Hill's criterion (23) by considering higher order terms.

In this spirit, conditions (21) or (23) or (25) represent the same criterion of non-bifurcation.

Condition (22) has been proposed by Hill [15] as a sufficient condition of stability of the present equilibrium of an elastic-plastic system. It appears as a direct generalization of the well-known criterion of second variation in elastic buckling. Much discussions have been devoted for a full justification of Hill's result, cf. [13, 25, 26] in connection with the notion of dynamic stability i.e. stability in the sense of Lyapunov.

An equilibrium is stable if small perturbations of the system imply only small perturbed motion near equilibrium. The energy E_p injected into the system by perturbation forces during time interval [0,t] is :

$$E_p (t) = K(t) + \Phi(t) - \Phi(0) \quad \text{with} \tag{26}$$

$$\Phi(t) = W(q,\lambda) + \int_{t^0}^{t} D(\dot\alpha,q)\ dt$$

where $K(t) \geq 0$ denotes the kinetic energy and $\Phi(t)$ the **total energy** which is the sum of total free energy and dissipated energy of the system at time t. In general, the total energy is a path-dependent functional.

It has been established that [26] :

$$\Phi(t) - \Phi(0) = \frac{1}{2}\ \delta^2\Phi\ \frac{t^2}{2} + \text{higher order terms...} \tag{27}$$

$$\text{with} \quad \delta^2\Phi = \int_V \dot\sigma\ \nabla\dot u\ dV + \dot u \cdot \Psi_{,uu} \cdot \dot u$$

Thus, if (22) is satisfied, then it will be necessary to inject energy to remove the system from equilibrium. The system will have no natural tendency to change itself without external actions. The considered equilibrium satisfies certain notions of stability qualified as static stability or directional stability [13, 19, 26].

To obtain a complete justification of dynamic stability, it is necessary to derive further estimates on the perturbed motions. References [13, 25] provide some discussions on this difficult subject. In particular, when dissipation potential D is state-dependent, the loss of symmetry in the rate response is an important difficulty and the problem is still open to further investigations. However, if D is state-independent, then dynamic stability is also ensured, at least for discrete systems.

In the same spirit, conditions (20) or (22) or (24) give the same criterion of (static or directional) stability. When D is state-independent, the considered equilibrium is also dynami-

cally stable, at least for discrete systems. Stability criterion (20) or (24) has been discussed for various problems of plasticity, friction, damage, fracture [27, 28].

Note that if $\mathbf{D}$ is state-independent, then stability criterion (24) is :

$$\delta\mathbf{A} \cdot \delta\boldsymbol{\alpha} > 0 \quad \forall \quad \delta\boldsymbol{\alpha} \text{ admissible, where } \delta\mathbf{A} = \mathbf{W}_{,\boldsymbol{\alpha}\boldsymbol{\alpha}} \cdot \delta\boldsymbol{\alpha}. \tag{28}$$

It is important to underline that bifurcation does not mean exchange of stability as in the classical context of elasticity, cf. [29, 30]. This result is well understood after Shanley's discussion on the buckling of elastic-plastic columns (Fig. 1) [13, 28]. Condition (20) for example is less restrictive then (21).

4. ILLUSTRATIONS

4.1 System of interacting linear cracks

The study of crack propagation and stability is a classical problem in Fracture Mechanics. In brittle fracture, when Griffith's criterion for crack propagation is assumed, a large body of research work has been dedicated to the two-dimensional problem of one linear crack in plane strain, plane stress or antiplane shear. Its generalization to a system of interacting linear cracks or a plane crack of arbitrary shape has been also discussed.

An elastic solid with Griffith cracks is a system undergoing irreversible transformation subsequent to crack propagation. Its quasi-static evolution fits perfectly in the framework of time-independent standard dissipative systems. For example, the quasi-static behaviour of a system of interacting linear cracks can be compared to the quasi-static behaviour of a descrete elastic plastic structure. The equations of evolution of these systems are of the same mathematical nature.

Consider the following crack propagation problem in brittle fracture, [28, 32-34]. An elastic solid V with m linear cracks of lengths ℓ_i is subjected to a loading path defined by load parameters λ. The quasi-static response $\boldsymbol{\ell}(t)$ associated with a given loading path $\boldsymbol{\lambda}(t)$ from a given initial configuration $\boldsymbol{\ell}_0$ is discussed. The cracks are assumed to propagate in their direction (no crack kinking). The response of the solid is reversible if there is no crack propagation. The crack lengths $\boldsymbol{\ell}(t)$ represent clearly a system of geometric parameters describing the irreversible evolution of this system. It is thus natural to take $\boldsymbol{\ell}$ as parameters $\boldsymbol{\alpha}$.

Total free energy of the system is :

$$\mathbf{W}(\mathbf{u},\boldsymbol{\ell},\boldsymbol{\lambda}) = \int_V W(\varepsilon(\mathbf{u}))\ dV - \int_{S_T} T_d\ \mathbf{u}\ dS$$

Displacement at equilibrium $\mathbf{u} = \mathbf{u}(\boldsymbol{\ell},\boldsymbol{\lambda})$ can be eliminated and leads to the definition of total free energy at equilibrium $\mathbf{W}(\boldsymbol{\ell},\boldsymbol{\lambda}) = \mathbf{W}(\mathbf{u}(\boldsymbol{\ell},\boldsymbol{\lambda}),\boldsymbol{\ell},\boldsymbol{\lambda})$.

Generalized force $A_i = -\dfrac{\partial \mathbf{W}}{\partial \ell_i}$ represents the **energy release rate** G_i associated to the i-th crack length.

In brittle fracture, the Griffith criterion consists in adopting the following crack propagation law :

If $G_i < G_c$ then $\dot{\ell}_i = 0$ (no propagation)
If $G_i = G_c$ then $\dot{\ell}_i \geq 0$ (possible propagation)

where G_c denotes a critical surface energy. Most often, G_c is a constant of the considered material. In some applications however, G_c may also be considered as a function of the effective length of propagation $\Delta\ell_i = \ell_i - \ell_i^0$ in order to take into account the stabilizing effect due to the near-tip plastic strain. The associated dissipation potential is :

$$\mathbf{D}(\dot{\boldsymbol{\ell}},\boldsymbol{\ell}) = \sum_{I} G_c(\Delta\ell_i)\,\dot{\ell}_i \quad \text{if } \dot{\ell}_i \geq 0 \ , \ = +\infty \text{ otherwise}$$

where I denotes the set of possible propagating cracks i.e. such that $A_i = G_c(\Delta\ell_i)$.

The rate problem is now described by variational inequality :

$$\dot{\ell}_i \geq 0 \ , \ i \in I \text{ and satisfy :}$$
$$(\Phi_{,ij} \cdot \dot{\ell}_j + W_{,i\lambda} \cdot \dot{\lambda})(\delta\ell_i - \dot{\ell}_i) \geq 0 \tag{29}$$
$$\text{where } \Phi_{,ij} = W_{,ij} + G'_c\,\delta_{ij} \ .$$

The present state $(\boldsymbol{\ell},\boldsymbol{\lambda})$ is stable in the dynamic sense if the following condition is satisfied :

$$\delta\ell_i \cdot \Phi_{,ij} \cdot \delta\ell_j > 0 \ \forall\ \delta\boldsymbol{\ell} \neq \mathbf{0} \ , \ \delta\ell_i \geq 0 \ , \ i \in I \tag{30}$$

The present state $(\boldsymbol{\ell},\boldsymbol{\lambda})$ is not an angular bifurcation state if :

$$\text{Matrix } \Phi_{,ij} \ , \ (i,j \in I) \text{ is positive-definite.} \tag{31}$$

If G_c is a constant, the stability criterion is :

$$-\delta G \cdot \delta\boldsymbol{\ell} > 0 \qquad \text{for any } \delta\boldsymbol{\ell} \neq O,\ \delta\ell_i \geq O \ , \ i \in I \tag{32}$$

where $\delta G = -\mathbf{W}_{,\boldsymbol{\ell}\boldsymbol{\ell}} \cdot \delta\boldsymbol{\ell}$.

For example, Fig. 2 represents an elementary system of two cracks simulating the delamination of laminated composite [35-38]. The analytical solution can be constructed and compared to experimental results [35]. It may be interesting to compare also this system to previous Shanley's model of elastic-plastic column of Fig. 1 .

4.2 Propagation of a plane crack or a damaged zone

The propagation of a plane crack of arbitrary shape Ω in a three-dimensional solid is an important problem in fracture mechanics. The study of interface cracks in thin films or laminate composites provide in the same spirit similar examples of plane cracks of arbitrary shape, [35-40]. In damage mechanics, some models of brittle damage also lead to the extension of a damaged zone, [41-46].

These examples are here considered in order to illustrate general results when the irreversible parameter $\boldsymbol{\alpha}$ is a plane surface Ω of boundary S , cf. Fig. 3. The solid is assumed to be elastic, the potential energy at equilibrium is $\mathbf{W}(\Omega,\lambda) = \mathbf{W}(\mathbf{u}(\Omega,\lambda),\Omega,\lambda)$ with :

$$\mathbf{W}(\mathbf{u},\Omega,\lambda) = \int_V W(\epsilon(u))\ dV - \int_{S_T} T(\lambda).u\ ds$$

To follow the general description, it is necessary to make the derivation $\mathbf{W}_{,\Omega}$ and $\mathbf{W}_{,\Omega\Omega}$ by techniques of derivation with respect to a domain. It is established that :

$$\mathbf{W}_{,\Omega} \cdot \delta\Omega = - \int_S G\, \delta\Omega \; ds \tag{33}$$

where G denotes the energy release rate at a point of the moving surface S and represents the local value of generalized force **A** associated with the motion of Ω.

For a plane crack in a three-dimensional solid, G is the limiting value of local Rice-Eshelby integrals :

$$G = J^0 \quad \text{with} \quad J^0 = \text{Lim}_{\Gamma \Rightarrow O} \int_\Gamma (W n_1 - n.\sigma.u_{,n}) \, d\Gamma \tag{34}$$

For a damage zone, the expression of G is :

$$G = [W - n.\sigma.u_{,n}] \tag{35}$$

For a delamination crack in a composite plate or a thin film :

$$G = [W - n.N.u_{,n} - n.M.\nabla\nabla w.n] \tag{36}$$

The second derivative of energy is :

$$\delta\Omega \cdot \mathbf{W}_{,\Omega\Omega} \cdot \delta\Omega = \int_S - \left(\delta G + \frac{G}{R} \right) . \delta\Omega \; ds \tag{37}$$

where δG denotes the variation of G following the motion $\delta\Omega$ of Ω and R is the curvature of S. A symmetric expression of the second derivative can be obtained from the expression of δG in terms of $\delta u(\delta\Omega)$, cf. [13]. The boundary perturbation problem to obtain $\delta u(\delta\Omega)$ is a familiar problem of shape optimization.

The propagation law :

$$\begin{aligned} &\text{If } G(s) \leq G_c \quad \text{then } \dot{\Omega}(s) = 0 \quad \text{(no propagation)} \\ &\text{If } G(s) = G_c \quad \text{then } \dot{\Omega}(s) \geq 0 \quad \text{(possible propagation)} \end{aligned} \tag{38}$$

is associated with dissipation potential :

$$\mathbf{D}(\dot{\Omega},\Omega) = \int_S G_c \; \dot{\Omega}(s) \quad ds \quad \text{for} \quad \dot{\Omega}(s) \geq 0 \tag{39}$$

When G_c is a constant, the total energy depends only on the present state :

$$\mathbf{\Phi}(\Omega,\lambda) = \mathbf{W}(\Omega,\lambda) + G_c \int_\Omega dV \tag{40}$$

The rate problem of propagation of the damaged zone Ω is given by variational inequality (19) which can be now written as :

$$\begin{aligned} &\dot{\Omega}(s) \geq 0 \quad \text{on } S_c \quad \text{and satisfies } \forall \; \delta\Omega(s) \geq O \quad \text{on } S_c : \\ &(\delta\Omega - \dot{\Omega}) \cdot (\mathbf{\Phi}_{,\Omega\Omega} \cdot \dot{\Omega} + \mathbf{W}_{,\Omega\lambda} \cdot \dot{\lambda}) \geq 0 \end{aligned} \tag{41}$$

where S_c denotes the portion of S such that $G(s) = G_c$.

Stability and bifurcation again can be discussed as in previous chapters. For example, the following proposition is obtained :

The present equilibrium is stable in the dynamic sense if the quadratic form $\delta\Omega \,.\, \Phi_{,\Omega\Omega} \,.\, \delta\Omega$ is positive definite on the set of admissible rates $\delta\Omega(s) \geq 0$ on S_c .

The stability criterion can also be written as :

$$\int_S - \delta G \,.\, \delta\Omega \; ds \; > \; 0 \quad \text{for any } \delta\Omega \neq O \text{ such that } \delta\Omega(s) \geq 0 \text{ on } S_c \,. \tag{42}$$

As an illustration, consider the torsion of a brittle elastic cylinder [46] Fig. 4. If ℓ is the length of the cylinder and $\lambda\ell$ the torsion angle between the two extreme sections, the displacement u is given by :

$$u_x = - \lambda\, yz \;\; , \; u_y = \lambda\, xz \;\; , \; u_z = \lambda\, \eta(x,y)$$

where $\eta(x,y)$ denotes the warping function. The expression of energy is :

$$W(\eta,\Omega,\lambda) \;=\; \lambda^2 \;\mu\ell \int_{V-\Omega} [(\eta_{,x} - y)^2 + (\eta_{,y} + x)^2 \,] \, dV$$

Equilibrium equation $W_{,\eta} = 0$ leads to :

$$\Delta\eta = 0 \text{ in } V-\Omega \;, \quad \frac{\partial\eta}{\partial n} = n_x\, y - n_y\, x \text{ on } S$$

For example, for a cylinder of circular section, these equations give the trivial equilibrium $\eta = 0$.

The dependence $d\eta(d\Omega)$ can be obtained directly from the derivation of the previous equations with respect to Ω. The following equations are obtained for a circular section :

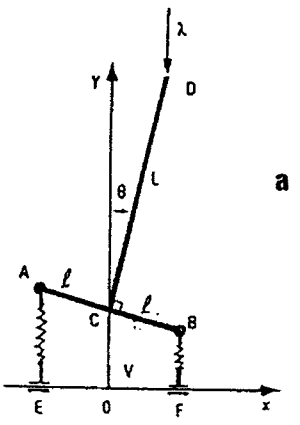
$$\Delta(d\eta) = 0 \text{ in } V-\Omega \;\; , \; \frac{\partial(d\eta)}{\partial n} = R \; d\Omega_{,s}$$

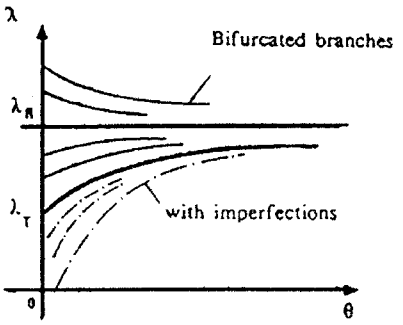
$$\text{Let} \quad d\Omega(\theta) \;=\; da_0 \;+\; \sum_{j=1}^{\infty} \; (\, da_j \cos j\theta + db_j \sin j\theta \,)$$

be the Fourier expansion of $d\Omega$. The following result is obtained :

$$d\Omega \,.\, \Phi_{,\Omega\Omega} \,.\, d\Omega \;=\; 2\, \lambda^2\, \mu\ell\, 2\pi R^2 \left\{ \; da_0^{\,2} \; - \sum_{j=1}^{\infty} \; (j-1) \;\; [\; da_j^{\,2} + db_j^{\,2} \;] \right\}$$

This form is not positive definite on the set $\{ \, da_j \;, \; j = 0,1,.. \, \}$ such that the rate $d\Omega(\theta)$ is non-negative. Thus, trivial equilibrium is not stable.





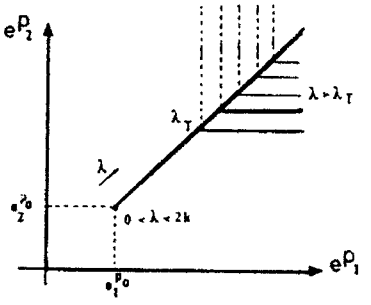
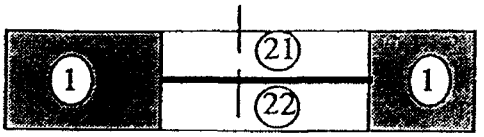
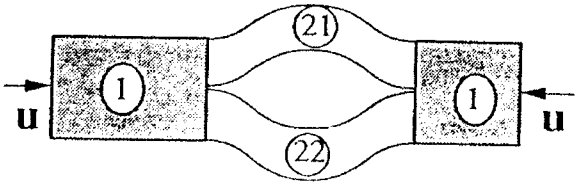


Figure 1. Plastic buckling of Shanley's column (a). Bifurcation diagrams are given by force-displacement curves (b) or by plastic strain curves (c).

Figure 2. An elementary system simulating delamination of composites.





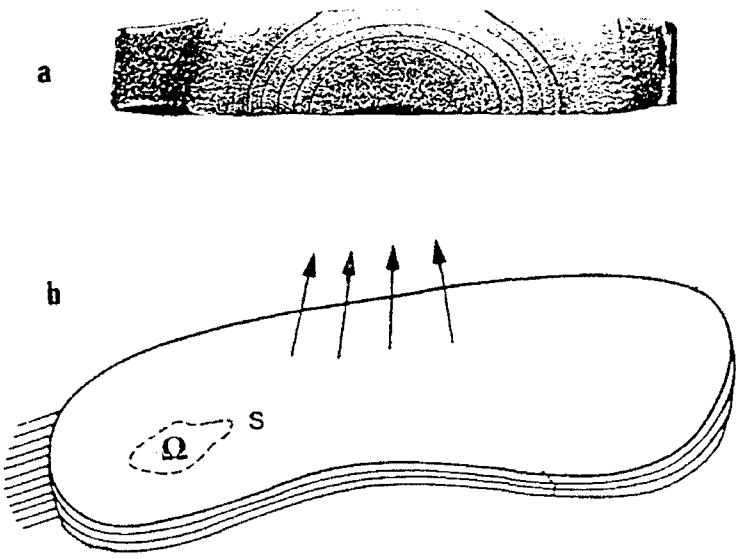


Figure 3. Propagation of a plane crack of arbitrary shape :
(a) in a three-dimensional solid,
(b) in the interfaces of multilayered composites.

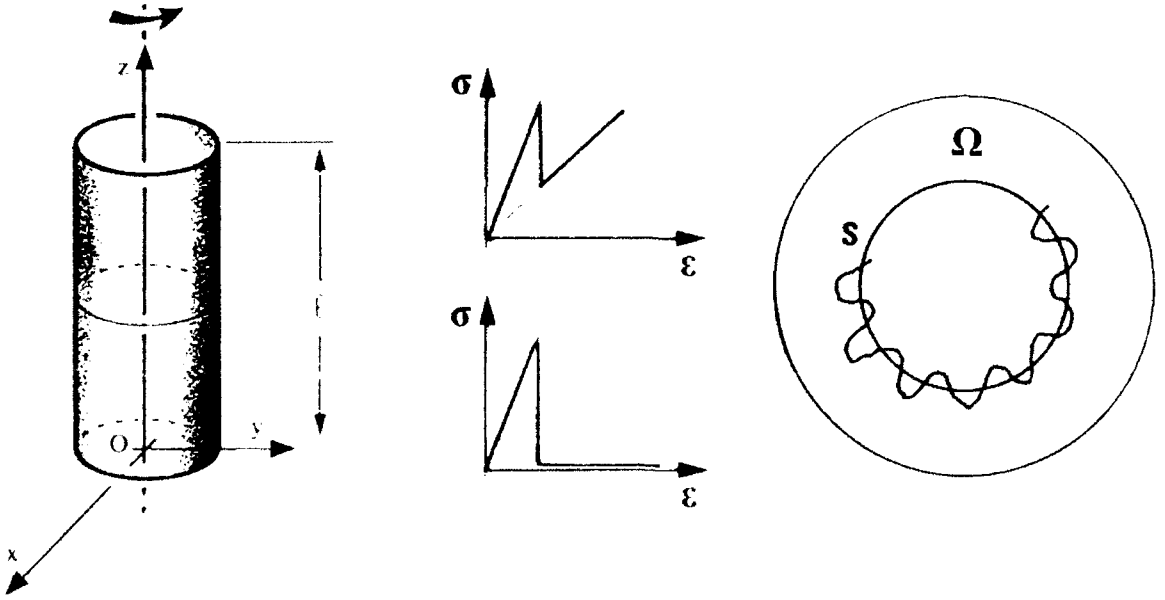


Figure 4. Torsion of a brittle elastic circular cylinder. Trivial response is instable under force or displacement control.

5. REFERENCES

1 Germain P., Mécanique des Milieux Continus.
Masson, Paris, 1973.

2 Gyarmati I. , Non equilibrium thermodynamics.
Springer Verlag, 1970.

3 Maugin G. , The thermomechanics of Plasticity and Fracture,
Cambridge University Press, 1992.

4 Biot M.A., Mechanics of incremental deformations
Wiley, 1965.

5 Mandel J. Plasticité classique et Viscoplasticité.
CISM course, 1971.

6 Moreau J.J., On unilateral constraints, friction and plasticity.
CIME Course, Bressanone 1973.

7 Mandel J. Cours de Mécanique des Milieux Continus.
Gauthier-Villars, 1961.

8 Rice J.R., Inelastic constitutive relations for solids : an internal variable theory
and its application in metal plasticity. J. Mech. Phys. Solids,19 (1971).

9 Halphen B. & Nguyen Q.S., Sur les matériaux standards généralisés.
J. de Mécanique, vol 1 (1975).

10 Lemaitre J. & Chaboche J.L., Mécanique des Matériaux solides.
Dunod, Paris, 1985.

11 Lee E.H. Elastic-plastic deformation at finite strain .
J. Applied Mech., 36 (1969).

12 Mroz Z. A description of work-hardening of metals with application to
variable loading. Foundations of plasticity, Varzawa, 1972.

13 Nguyen Q.S., Stability and bifurcation of standard dissipative systems.
CISM Course, Udine, 1991.

14 Halphen B. , Sur le champ des vitesses en thermoplasticité finie.
Int. J. Solids & Structures,11 (1975).

15 Hill R., A general theory of uniqueness and stability in elastic-plastic solids.
J. Mech. Phys. Solids, (1958) 236.

16 Martin J.B. , Plasticity : fundamentals and general results.
MIT Press, 1975.

17 Stolz C. , Anelasticity and Stability.
Doctorat thesis, Paris 1987.

18 Klarbring A. Derivation and analysis of rate boundary value problems
of frictional contact. Eur. J. Mech. A, 9 (1990).

19 Chateau X. & Nguyen Q.S. Buckling of elastic structures
in unilateral contact with or without friction.
Eur. J. Mech., A, 10 (1991).

20 Duvaut G. & Lions J.L., Les inéquations en mécanique et en physique.
Dunod, Paris, 1972.

21 Maier G. , A shake-down matrix theory allowing for workhardening
and second order geometric effect.
Symposium "Foundations of Plasticity", Varsovie, 1972.

22 Hutchinson J.W., Plastic buckling.
Advances in Applied Mechanics, (1973).

23 Sewell M.J. , A survey of plastic buckling.
Stability, Waterloo Press, 1972.

24 Triantafyllidis N., On the bifurcation and post-bifurcation analysis
of elastic plastic solids under general pre-bifurcation conditions.
J. Mech. Phys. Solids, 31 (1983).

25 Nguyen Q.S. & Radenkovic D., Stability of an equilibrium

in elastic plastic solids. IUTAM Symposium, Marseille, 1975.
26 Petryk H. , On energy criteria of plastic instability, Plastic buckling, Proc. Considere Memorial, ENPC, Paris 1985.
27 Nguyen Q.S., Bifurcation et stabilité des systèmes irréversibles obéissant au principe de dissipation maximale. J. de Mécanique, Paris (1984).
28 Bazant Z.P. , Cedolin L., Stability of structures. Elastic, inelastic, fracture and damage theories. Oxford University Press, 1991.
29 Koiter W.T., On the stability of elastic equilibrium. Ph.D. thesis, Delft, 1945.
30 Budiansky B., Elastic buckling. Advances in Applied Mechanics, vol. 14, 1974.
31 Potier-Ferry M., Towards a catastrophe theory for the mechanics of plasticity and fracture. Int. J. Engng & Sc. (1985).
32 Bazant Z.P., Ohtsubo H., Stability conditions for propagation of a system of cracks in brittle solids. Adv. Civil Engng, (1977).
33 Nemat-Nasser, Sumi Y. & Keer L.M., Unstable growth of tensile cracks in Brittle Solids. Int. J. Solids & Structures, (1980).
34 Nguyen Q.S., Stolz C., Energy methods in fracture mechanics. IUTAM Symposium, Masson, Paris (1987).
35 Pradeilles R. Délaminage des composites. Doctorat thesis, 1992
36 Storakers B., Non linear aspects of delamination in structural members. Theoretical and Applied Mechanics, ICTAM, Grenoble 1988.
37 Destuynder P. & Nevers T., Un modèle de calcul des forces de délaminage dans les plaques minces multicouches, J.M.T.A.,6 (1987).
38 Cochelin B , Flambage et délaminage dans les plaques composites stratifiées. Doctorat thesis, Metz 1989.
39 Hutchinson J.W., Thouless M.D., & Liniger E.G. Growth and configurational stability of circular buckling driven film delamination. Acta Metall. Mater. 40 (1992).
40 Jensen H.M., Thouless M.D. Effect of residual stresses in the blister test. Report 439, The Technical University of Denmark, 1992.
41 Bérest P. , Problèmes de Mécanique associés au stockage souterrain. Doctorat thesis, Paris, 1989.
42 Bui H.D. & Ehrlacher A., Propagation dynamique d'une zone endommagée. C.R. Acad. Sciences, t290, 1980.
43 Dems K. & Mroz Z., Stability conditions for brittle plastic structures with propagating damage surfaces. J. Struc. Mech., Vol 13, (1985) 95.
44 Ehrlacher A., Contribution à l'étude thermodynamique de la progression de fissure et à la mécanique de l'endommagement brutal. Doctorat thesis, Paris, 1985.
45 Marigo J.J., Modélisation et analyse qualitative de l'endommagement des matériaux et des structures. Cours EDF-INRIA-CEA. Paris, 1990.
46 Fedelich B. & Bérest P., Torsion d'un cylindre élasto-fragile, Arch. Mech. Stos., Vol 40, 5-6, (1988).

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Bubble mechanics: luminescence, noise, and two-phase flow

Andrea Prosperetti

Department of Mechanical Engineering, 122 Latrobe Hall, The Johns Hopkins University, Baltimore MD 21218

Abstract

A descriptive treatment of light emission from pulsating bubbles (sonoluminescence), sound generation by rain falling on water and breaking waves, and the propagation of shock waves in bubbly liquids is presented. The first section contains a brief survey of some occurrences of bubbles in literature and in the figurative arts. Considerations on the etymology of the word are presented in the appendix.

1. INTRODUCTION

Πομφόλυξ ο άνθροπος – Man is a bubble – is an ancient Greek proverb that enjoyed a singular favor in the western culture as an expression of the caducity and impermanence of human life. It is used by a number of Latin and Greek authors such as Varro who, apologizing in the Preface to his *De Re Rustica* (36 b.c.) that the work is not as polished as he would like, says that nevertheless, being 80 years old, he should go ahead with its publication since, "ut dicitur, si est homo bulla, eo magis senex" ("if, as they say, man is a bubble, all the more so is an old man"). Petronius (1st century a.d.), in a mocking passage of the *Satyricon* in which he compares man to inflated walking bags and flies, also says "nos non pluris sumus quam bullis" ("we are no more than bubbles," 42, 4). Lucian (117-180) elaborates: "I've thought of a simile to describe human life as a whole ... You know the bubbles that rise to the surface below a waterfall – those little pockets of air that combine to produce foam? ... Well, that's what human beings are like. They're more or less inflated pockets of air ... but sooner or later they're all bound to go pop" (*Charon* 19).

The advent of Christianity, with its message of hope and salvation, rendered the idea less relevant and probably the only pre-Renaissance textual reference to it is in the *Lexicon* by the 10th-11th century Byzantine scholar Suida ("as a bubble immediately disappears when it is broken, so does the memory of the splendid and powerful upon their death").

However, in the northern post-reformation cultural climate of the 16th and 17th century, the metaphor regained its appeal. Most succint – and first – is Erasmus: "Homo bulla" [man (is a) bubble, *Adagia*, 1508, n. 1990]. Taverner, in his *Proverbs or adagies ... gathered out of the Chiliades of Erasmus* (1539) echoes the concept, and so does sir Thomas Elyot (1545), Golding ("When man seemeth to bee at his best, he is al-

Figure 1: Engraving after a 1575 painting by Cornelis Ketel illustrating the Greek proverb "Man is a bubble." (Reproduced from Ref. 2)

togither nothing else but a bubble blowen togither of vanitie" *Psalms* xxxix 6, II 155, 1571), Samuell Rowlands ("Poore mortals in my hands are brittle ware, Like Vapor, Bubble, Flower, wither'd Hay," *A Terrible Battell betweene the two consumers of the whole World: Time, and Death*, p. 23, 1606) Sir William Browne ("What's he ... whose life's a bubble, and in length a span; ...? 'Tis a man," *Britannia's Pastorals*, Bk. i, song 2, l. 192, 1613), Simon Wastell (circa 1628), Jeremy Taylor (1651), and several others.

Sir Francis Bacon, in his *The World* (1629) expresses a more radical point of view: "The world's a bubble," which Francis Quarles quickly echoes "My soul, what's lighter than a feather? ... This bubble world. What then this bubble? Nought" (*Emblems*, Bk. i, No. 4, published in 1635), and also "The pleasure, honour, wealth of sea and land / Bring but trouble; / The world itself, and all the world's command / Is but a bubble" (ibid., Bk. i, No. 6).

Of the 9 times the word "bubble" recurs in Shakespeare's theatre (it is never used in his poetry), 3 times it similarly suggests volatility and insubstantiality such as "the bubble of reputation" (*As You Like It* II, vii, 139), or "A dream of what thou wast ... a breath, a bubble" (*Richard III* IV, iv, 83).

The same notion accompanies the word throughout the subsequent English language literature. William Cowper (*The Task*, published in 1785, Bk. III, l. 175) admonishes: "Eternity for bubbles proves, at last, / A senseless bargain." Samuel Johnson, in an ironic passage thus commiserates the fashion-conscious: "Hard is his lot, that here by fortune plac'd / must ... chase the new-blown bubbles of the day." Alexander Pope is most impartial: "Who sees with equal eye, as God of all, / a hero perish, or a sparrow fall, / atoms or systems into ruin hurled, / and now a bubble burst, and now a world" (*An Essay on Man*, 1734, Epis. i, l. 87) and also: "Like bubbles on the sea of matter borne,/ They rise, they break, and to the sea return" (ibid. Epis. iii, l. 19). Coleridge thus uses the metaphor to express a lover's disappointment "I listen for thy voice / Beloved! 't is not thyne; thou are not here! Then melts the bubble into idle air,/ and wishing without hope I restlessly despair" (*Blossoming of the Solitary Date-Tree*, st. 4). His contemporary Sir Walter Scott joins in the lover's desperation: "Like the dew on the mountain,/ Like

the foam on the river,/ Like the bubble on the fountain, / Thou art gone, and for ever!" (*The Lady of the Lake*, Canto iii, st. 16, 1810). The idea can be followed in Lord Byron: "The eternal surge of time and tide rolls on, and bears afar our bubbles" (*Don Juan*, 1824, canto xv, st. 99), and all the way down practically to modern times.

It does not seem that English language poets ever found any word other than *trouble* to rhyme with *bubble*. An early example by Quarles was cited before. Shakespeare's is probably the most well known one: "Double, double toil and trouble, / fire burn and cauldron bubble" (*Macbeth* IV, i, 10). One can also add Dryden: "War, he sung, is toil and trouble;/ honor but an empty bubble" (*Alexander's Feast, or the Power of Music* l. 97), the Australian poet Adam L. Gordon: "Life is mostly froth and bubble; / Two things stand like stone: / Kindness in another's trouble, / Courage in your own" (*Ye Weary Wayfarer*, Fytte viii), and even the peripheral Nathaniel Ward, who combines it with a blatant expression of anti-feminism: "The world is full of care, and much like unto a bubble;/ Women and care and care and women, and women and care and trouble" (*The Simple Cobbler of Aggawam*, p. 25).

While the metaphor of bubble as volatility is present in other more or less unrelated traditions, such as in the 12th-13th century Japanese writer Kamo no Chomei ("The flow of the river is ceaseless and its water never the same. The bubbles that float in the pools, now vanishing, now forming, are not of long duration: so in the world are man and his dwellings" *Hojoki – An Account of my Hut*, 1212), or the 11th century Persian poet Omar Khayyam ("A deathless Saki draws Khayyams in thousands/ like wine bubbles out of Creator's bowl," *The Rubaiyyat* st. 49, transl. Graves), it is curiously absent from other major literatures. The closest the Bible gets to the idea is in *Proverbs, 21: 6* with the words *hevel niddaf* – literally "driven vapor." The *New American Bible* translates: "He who makes a fortune by a lying tongue is chasing a bubble over deadly snares," but most other English versions use different constructs. For example, the *King James Bible* says: "The getting of treasures by a lying tongue is a vanity tossed to and fro of them that seek death." The Septuagint translation is far from literal, with the words *θησαυρίσματα μάταια*, "vain possessions."

Dante's *Divina Commedia* abounds with literal bubbles. The violent rulers are punished by boiling in blood: "lungo la proda del bollor vermiglio, / dove i bolliti facieno alte strida" [along the brink of the vermilion boiling, / wherein the boiled were uttering loud laments (trad. Longfellow), *Inferno* xii, l. 101] and the barrators (i.e., corrupt officials) by boiling in pitch: "bollia lá giuso una pegola spessa, / che 'nviscava la ripa d'ogni parte./ I' vedea lei, ma non vedea in essa / mai che le bolle che 'l bollor levava, / e gonfiarsi tutta e risieder compressa" (was boiling down below there a dense pitch / which upon every side the bank belimed. / I saw it, but I did not see within it / aught but the bubbles that the boiling raised, / and all swell up and resubside compressed, *Inferno* xxi, l. 17), but the *Homo bulla* idea is absent. Similarly, I have been unable to find any reference to this notion in other Italian poets or in the French or Spanish literature.

An interesting and direct offshoot of Erasmus's exhumation of the Greek proverb can however be found in the Netherlands. The Remonstrant minister Uytenbogaert admonished the Stadholder's court in the by now familiar terms: "Man is nothing other than a bubble that children blow in a mussel shell, that glistens in the round ... but which in an instant disappears and is gone." Rembrandt painted a portrait of Uytenbogaert (sold

in July 1992 at Sotheby's for 3.8×10^6 pounds), and one of Rembrandt's better pupils, Gerrit Dou, contributed to the tremendously successful iconography of a child blowing a soap bubble, "an image that reflected both levity and gravity, that floated like childish reveries and popped as childhood itself had to end" [1]. Many other Dutch painters, such as Cornelis Ketel (perhaps the earliest one, on the back of a 1575 portrait, see Fig. 1 [2]), Hendrik Goltzius, Jacques de Gheyn, Jan Miense Molenaer, and Frans Hals have left us a number of images of children blowing bubbles. It is interesting that, while the ancient authors had in mind the bubbles briefly floating on water surfaces, from Ketel onward the reference is always to soap bubbles. This fact is probably also due to the dramatic improvements in the soap technology intervened since antiquity.

In the mid 1730's, the French painter Jean-Baptiste Siméon Chardin used his five-year old son as a model for several pictures of a child blowing bubbles and later again an older boy. By this time, however, the original Calvinist moral embodied in the *Vanitas* genre had faded into the background leaving its place to the painter's delight with his subject and his own ability as we find in several others artists all the way down to the pre-Raffaellite Sir John Everett Millais, whose painting "Bubbles" was used in some of the Lever Brothers' commercial advertising for Pear's Soap [1].

Other than its career as a lofty metaphor, the word *bubble* has also enjoyed another – and disreputable – one. With reference to the *bulla*, a spherical gold ornament worn by the noble Roman children after an Etruscan custom, Juvenal uses the word to fashion an insult: "senior, bulla dignissime" ("Old man, worthy of a bubble," *Satirae* 13, l. 33), and in old Italian *bolla acquaiola* ("water bubble") was a disparaging designation. In Latin, *animam ebullire* ("to gurgle one's life out," similar to the bubbles that gurgle in the neck of a flask when the liquid is poured out) was a vulgar expression for "to die" used, among others, by Seneca and Persius.

In Shakespeare's *All's Well That Ends Well* the Second Lord Dumaine qualifies Paroles as "On my life, my lord, a bubble" (III, vi, 6) and goes on to explain: "He's a most notable coward, an infinite and endless liar, an hourly promise-breaker, the owner of no one good quality worthy of your lordship's entertainment" (ibid., 10-12). Later, in 17th and 18th century English, a *bubble* denoted a "dupe" or gullible person. In the underworld's slang, "to betray" was indicated as *to bubble*, with *bubble* also standing for a piece of information. *Bubble duster* was used for "handkerchief" (presumably after the Scottish usage of *bubble* for "mucus," whence also *bubbly* for "snotty, dirty"), and we have the plethora of American slang expressions *bubble brain, bubble head, bubble gummer* (for "teen-ager"), *bubble dancer* (for "dish washer," in addition to the original reference to a night-club performer), and others. As a symbol for financial bust one does not have to wait until the recent crash of the Tokyo stock exchange to encounter the word. The *South Seas Bubble* was a scheme devised in England in 1711 to eliminate the national debt – at the expense of the usual "bubbles" – and the French were quick to catch on with their similar 1717 *Mississippi Bubble*.

In French *faire des bulles* signifies something like "to make a fuss" and *coincer la bulle* or *buller* is a popular expression for "doing nothing." In Dutch, *blaasjes verkopon* and the corresponding English *to blow bubbles* similarly stand for "to devise baseless theories, to amuse oneself in a childish manner," and *bolla di vento* ("wind bubble") recurs in 18th-century Italian in a similar sense.

It will be my task in the rest of this paper to try to convince the reader that, in spite of all of the above, bubbles are "investigatione dignissimae" – most worthy of serious scientific study.

It should be stressed at the outset that the type of bubbles that will form the object of this paper differ in a major way from those featured in the previous quotations. Although the bursting of floating bubbles and the related problem of the stability of foams is a major scientific and technical problem, here I shall confine myself to gas bubbles more deeply submerged in a liquid.

The breadth of the subject prevents me from any attempt at exhaustiveness. I will limit myself to a few recent developments that, hopefully, will give an idea of the vast range and far-reaching implications of bubble phenomena.

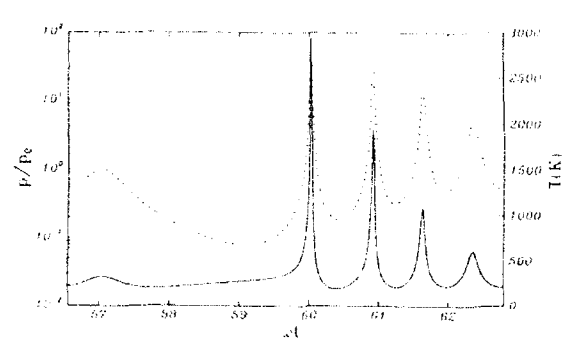
2. SONOLUMINESCENCE

From the mathematical point of view, the richness exhibited by bubble phenomena may be considered as a direct consequence of the $1/r$ singularity of the fundamental solution of Laplace's equation in three-dimensional space. Since bubble dimensions are often "small" (with respect to other, problem-dependent lengths), this singular behavior has the opportunity to manifest itself very strongly. With bubbles, one's life is enriched by the proximity to this singularity.

A very striking phenomenon due precisely to this fact is *sonoluminescence*, the light emitted by a sound-irradiated liquid. In an acoustic wave the energy density is of the order of $\mathcal{E} = p'^2/\rho c^2$, where p' is the pressure amplitude, ρ the fluid density, and c the speed of sound. Since the mole density per unit volume is ρ/M, where M is the molecular mass, the corresponding energy per molecule is $(M/\rho)\mathcal{E}/N_A$, where N_A is Avogadro's number. For a fairly intense pressure wave in water water, with $p' \sim 1$ bar, this amounts to approximately 10^{-9} eV per molecule, which is of the order of 10^{-9} times smaller than the energy of low-eV radiation. The appearance of such radiation therefore implies that a mechanism exists to amplify the energy density by a factor of 10^9.

First discovered in the late 20's, the effect prompted a number of more or less fanciful speculations as to its origin which subsequent research has put to rest – to some extent [3]. The current understanding is that sonoluminescence is due to the dissociation and recombination of gaseous molecules contained in bubbles that pulsate in a strongly non-linear fashion under the action of the sound field. It is the $1/r$ singularity of the potential for the associated radial motion of the liquid that is responsible for the effect.

A theory of radial bubble pulsations and sonoluminescence is given elsewhere [4]. For purposes of illustration I reproduce here in Fig. 2 a graph of the internal pressure (solid line) and center temperature (dashed line) for an argon bubble in water subject to a 0.93 bar sound field. The bubble radius is 26 μm and the sound frequency 21 kHz. Since the radius of a resonant bubble at this frequency would be 150 μm, the bubble is driven at about one fifth of its resonance frequency. A parallel chemical kinetics calculation [4], the results of which are shown in Fig. 3, indicates that a substantial number of hydroxyl radicals are produced in these conditions by collisional dissociation of the vapor molecules. This prediction is in agreement with the experimental fact that liquids saturated with



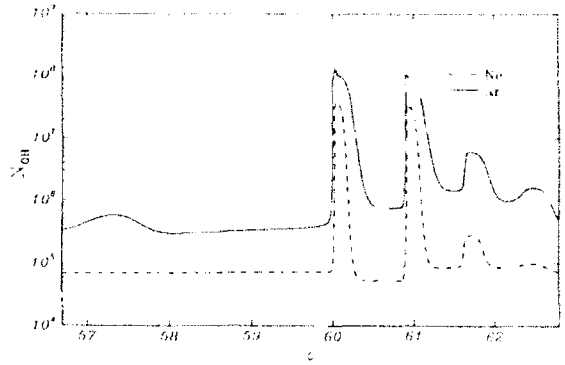


Figure 2: Pressure (dashed line) and center temperature (solid line) during steady oscillations of a 26 μm-radius argon bubble driven at 21 kHz and 0.93 bar acoustic pressure amplitude

Figure 3: Number of hydroxyl radicals as a function of time during the steady oscillations of the argon bubble of the previous figure (solid line), and an equal one containing neon (dashed line).

noble gases emit a strong 310 nm-line corresponding to a transition of the excited OH radical to its ground state [5]. Heat transport inside the bubble has a strong effect on this process. Neon has a thermal conductivity 3-4 times that of argon, and the corresponding OH number, shown in Fig. 3 by the dashed line, is much smaller than for the argon case. It should be realized that, during the period of peak temperature, the temperature gradient inside the bubble is of the order of 10^5 K/mm. Clearly, this is only possible due to the very short duration of the peak temperature, for this example of the order of 300 nsec.

However short, this time interval is many orders of magnitude longer than the duration of the sonoluminescence light flashes, which is currently estimated below 50 psec [6]. This extremely brief duration is perhaps the most baffling aspect of sonoluminescence. A possible explanation could be the following. It is well known since Wood's experiments in 1904 that gases are very opaque to radiation corresponding to their emission bands. It is for this reason, for example, that we can only observe the outermost layer of the sun as the light emitted by the inner regions is absorbed by the surrounding layers. As the bubble compresses, the temperature is highest at the center and, when a certain threshold is exceeded, H_2O molecules begin to dissociate there. A fraction of them is excited and radiates upon decaying to the ground level. Very soon afterward, the temperature in the next layer of gas exceeds the OH-production threshold, and these radicals are now available to absorb the radiation still coming from the center of the bubble. For the case considered previously our computations indicate that it takes about 100 psec for the temperature "wave" to travel to a distance of 0.01 radii from the origin center.

To explain why no light is emitted from the outer layer, one may note that, by symmetry, the temperature gradient must vanish at the bubble center (another consequence of the presence of a singularity !). Hence, the temperature in the central region of the bubble is much more homogeneous than in the surrounding layers. In particular, in the

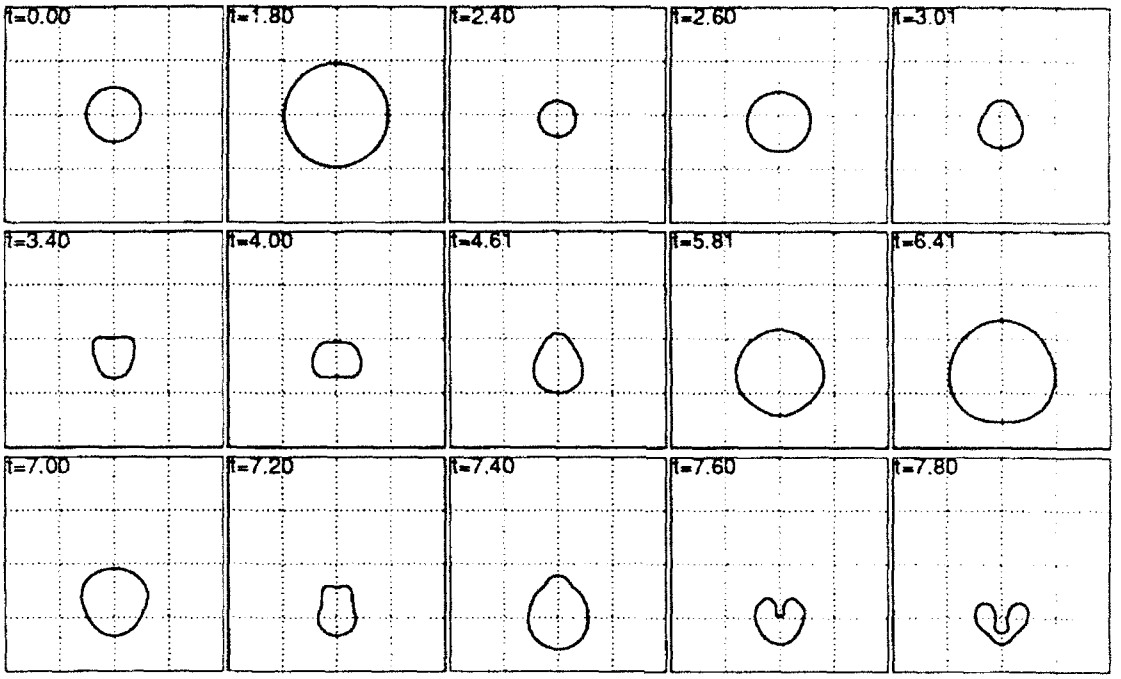


Figure 4: Successive shapes of a 20 μm gas bubble driven at 20 kHz by a 1 bar-amplitude sound field switched on at $t = 0$. The time unit is 10.5 μsec. Initially the bubble is located 1/4 wavelength above the antinode of the sound field.

outermost layer where the temperature barely exceeds the threshold, virtually all the OH's are produced in the ground state and therefore can absorb, but not emit, radiation. Clearly, radiationless transitions to the ground state must also occur, and these are rendered probable by the high number density of the molecules (of the order of $10^{18}/\text{cm}^3$ for the example of Figs. 2 and 3).

With numbers of this order of magnitude involved, one might wonder whether one could get "cold fusion" to occur in a liquid subjected to a sufficiently intense sound field and, indeed, the idea for such a reactor has been patented [7]. Unfortunately, there is a snag: the spherical shape of an oscillating bubble has a strong tendency to instability, so that any attempt to drive the bubble too strongly will lead to its shattering rather than to increased temperatures and pressures. This instability is similar to the well-known Rayleigh-Taylor one as, when the pressure in the bubble rises, the liquid is being accelerated upon by a lighter fluid, the gas.

We present the results of a calculation that illustrates this instability in Fig. 4. The bubble here has a radius of 20 μm and is forced by a 1-bar, 20-kHz plane standing acoustic wave. In spite of the fact that the wavelength is nearly 4,000 times larger than the bubble radius, the non-zero pressure gradient at the bubble location (1/4 wavelength above the antinode) induces a perturbation that, as can be seen, is strongly amplified during the compression phase of the oscillations. This calculation ignores various dissipative processes and therefore it probably overestimates somewhat the magnitude of the effect.

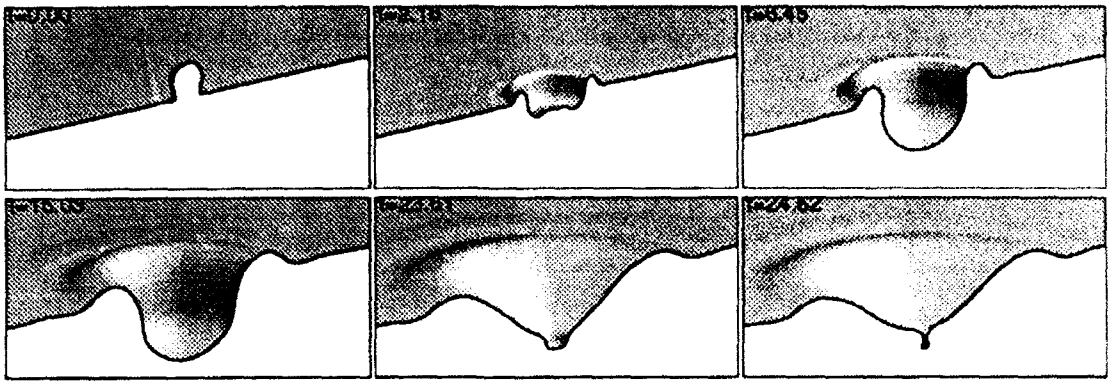


Figure 5: Bubble entrapment (last frame) by a 2.86 mm-diameter water drop impacting a water surface at 2 m/sec (first frame). The free surface of the liquid is shaded. Times are in msec.

3. SOUND GENERATION

A completely different range of phenomena in which bubbles are of major importance is sound production in liquids. The flushing of a toilet or simply the filling of a glass usually make a far louder noise than water flowing in the pipes of a house. The reason is that a homogeneous liquid emits sound due to turbulence, which is a quadrupole-type source, while bubbles behave as monopole sources. To illustrate these processes, I will briefly discuss two bubble-related mechanisms responsible for the production of oceanic underwater noise. Since sound propagates over great distances in the ocean, these mechanisms are of importance, among other reasons, as diagnostic tools for oceanic conditions and processes.

An interesting example is rain noise that, at low wind speeds, is experimentally found to possess a spectrum – strongly peaked at about 14 kHz - very nearly independent of rain drop size distribution and rain intensity. This seeming "universality" of the rain noise spectral shape has puzzled early investigators until the key was identified in the fluid dynamics of bubble production by drops striking a free surface. This work is discussed at length elsewhere [8]. Briefly, it is found that drops striking a liquid surface make very little noise unless a bubble is entrained in the liquid. The results of a calculation illustrating this process are presented in Fig. 5. Bubble entrapment occurs only in a narrow range of drop sizes and impact velocities which, for impacts at terminal velocity, corresponds to diameters between about 0.8 and 1.1 mm. Since drops in this size range are always present, and since other drops are acoustically irrelevant, all rains sound very much alike.

Anyone who has listened to a wave breaking on the shore is quite aware that a substantial amount of noise exists in the ocean also at frequencies much lower than the 14-20 kHz characteristic of rain noise. As a matter of fact, measurements indicate a substantial amount of wind-dependent (and therefore not man-made) oceanic noise at frequencies of a few hundred Hz and lower [9]. The linear natural frequency of oscillation of a bubble

having an equilibrium radius a is, approximately [10],

$$\omega_0 = \frac{1}{a}\sqrt{\frac{3\kappa p_0}{\rho}}, \tag{1}$$

where p_0 is the equilibrium internal pressure in the bubble and $1 \leq \kappa \leq \gamma$, with γ the ratio of the specific heats, is a polytropic index dependent in a complex way on the frequency. According to this formula, for a bubble to pulsate at 100 Hz, the diameter should be 6-7 cm. It is very unlikely that such large bubbles are produced in substantial numbers by waves that, in the open ocean, break in a spilling rather than the plunging mode characteristic of near-shore breaking. Instead, there are substantial reasons to believe that this low-frequency noise is originated by the *collective* oscillations of the bubble clouds produced by breaking waves [11,12]. A simple argument can be presented in the following terms. At low frequencies, the speed of sound c_m in a liquid containing gas bubbles is given approximately by

$$c_m^2 = \frac{\kappa p_0}{\beta \rho}, \tag{2}$$

where $\beta = V_G/V$ is the fraction of the mixture volume V occupied by the gas. The simplest way to derive this expression is to note that, if the liquid compressibility is neglected, a change in the mixture density ρ_m can only be due to a change in the volume V_G occupied by the gas sc that $\delta\rho_m/\rho_m = -\delta V_G/V$. On the other hand, if the bubbles compress polytropically, $\delta V_G/V_G = -\kappa\delta p/p_0$ and (2) readily follows upon identifying c_m^2 with $\delta p/\delta\rho_m$. [1]

The striking fact is that already at very small volume fractions c_m is much lower than c. For instance, for $\beta = 1\%$ and atmospheric pressure, one finds $c_m \simeq 100$ m/sec to be compared with $c = 1500$ m/sec in pure water. Thus, in view of the great disparity between c and c_m a bubble cloud can, to a first approximation, be regarded as a compressible medium in a rigid enclosure. The lowest normal mode of oscillation of the cloud has therefore a frequency ω_{min} of the order of c_m/L, where L is the linear scale of the cloud, so that, from (1) and (2), one finds

$$\frac{\omega_{min}}{\omega_0} \sim \frac{1}{\beta^{1/6}N^{1/3}}, \tag{3}$$

where N is the number of bubbles in the cloud. This result shows that 1,000 1-mm bubbles - which individually have a natural frequency of about 3 kHz - would give rise to a cloud oscillating at about 300 Hz. Since the liquid surrounding the cloud is not actually incompressible, the cloud eigenmodes' energy leaks out and is radiated to great distances.

An indirect quantitative confirmation of this process has been obtained by using an "artificial" cloud in the laboratory [13]. If air is bled out of a disk-shaped bank of needles at the bottom of a tank, a hydrophone picks up a signal similar to that shown in Fig. 6. The peaks correspond to eigenfrequencies of the system and can be calculated as a check of the theory. Some results of this comparison are shown in Fig. 7.

[1] The result (2) fails to converge to the correct limit as $\beta \to 0$ since the liquid compressibility has been neglected. It can be shown that this is justified at gas volume fractions above about $10^{-3}\%$.

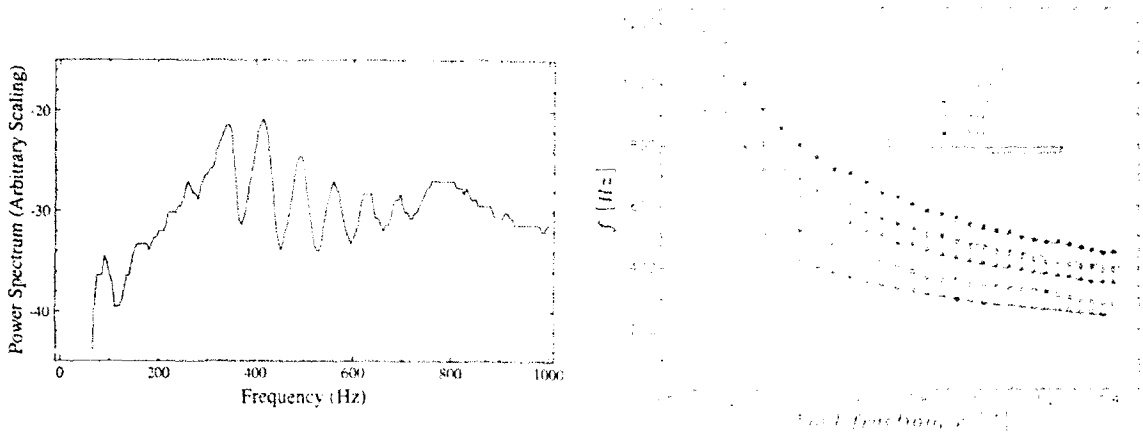


Figure 6: Frequency spectrum recorded in a laboratory tank in which a steady cylindrical column of bubbles is maintained by air injection at the bottom. Note the many resonances of the system. (From Nicholas *et al.*, Ref. 13.)

Figure 7: Comparison between theory and experiment for the frequencies of the normal modes shown in the previous figure. The horizontal scale is the gas volume fraction. The differences at the higher frequencies are due to the proximity of the resonance frequency of the individual bubbles. (From Nicholas *et al.*, Ref. 13.)

The agreement deteriorates at the lower volume fractions due to the proximity of the resonance frequency of the individual bubbles, approximately 1.9 kHz in this case [14].

It is remarkable that the only source of excitation for this system is the "pinch-off" noise emitted by the bubbles as they leave the needles. A computed example of this process is shown in Fig. 8, where the compression of the bubble as it detaches from the needle is quite clear. In a breaking wave, the processes giving rise to the formation of the bubbles are far more violent and a much higher noise emission level can be expected.

4. SHOCK WAVES IN BUBBLY LIQUIDS

The last example of the previous section provides an instance in which it is not possible nor desirable to account in detail for the behavior of the individual bubbles, but it is more efficient to treat the liquid-gas mixture as a continuum endowed with a complex internal structure. The area of gas-liquid multi-phase flows is one of great industrial relevance which is currently the object of intense research efforts.

After the case of linear pressure waves touched upon in the previous section, perhaps the simplest case to consider is that of shock waves propagating in a dilute mixture of bubbles and liquid. Such waves may originate e.g. in pipelines during operational transients. Experiment [15] shows that the steady profile of weak waves is either completely smooth (Fig. 9, left) or smooth at the wave front and "rippled" in the middle (Fig. 9, right). Our recent study [16] shows that this fact is due to the peculiar nature of the

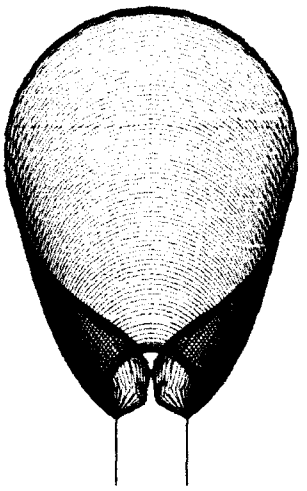


Figure 8: Successive shapes of a bubble blown out of a 1 mm-diameter needle.

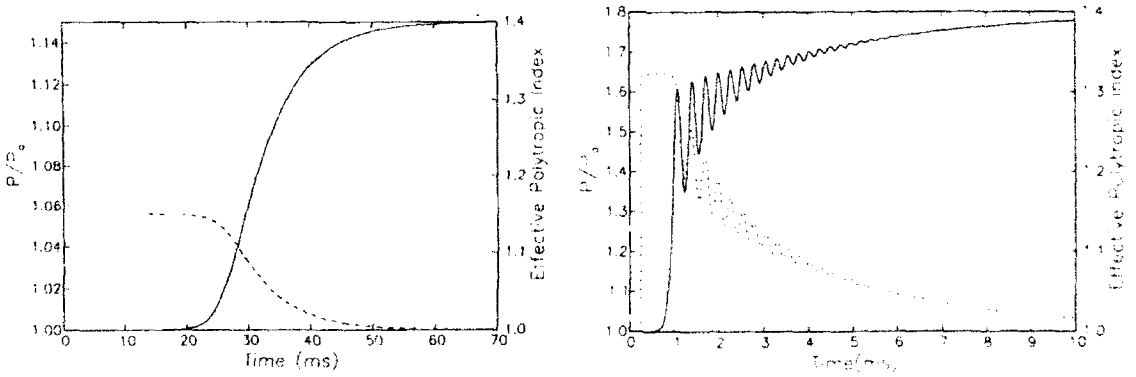


Figure 9: Examples of pressure profiles in steady shock waves in a bubbly liquid. At left is a completely smooth waveform, at right an oscillating one. The dashed line is the effective polytropic index along the wave. (From Ref. 16).

dispersion relation for pressure waves in a bubbly liquid. Consider first linear waves at very low frequency. As the bubbles are compressed, the gas temperature has time to equilibrate with that of the surrounding liquid, the process is isothermal and, in Eq. (2), $\kappa = 1$. As the frequency increases, complete temperature equilibration is not possible and κ becomes greater than 1 so that the speed of sound increases. At still higher frequencies the natural frequency of the bubbles is approached, it is easier and easier to compress them, and physical processes not accounted for in Eq. (2) greatly decrease the speed of sound. The phase velocity of the waves possesses therefore a mild maximum below the resonance frequency of the bubbles. This feature is clear in Fig. 10 where the solid lines show the dispersion relations corresponding to the region near the front and the back of the waves of Figs. 9. The dashed lines are the group velocity. The speed U of the shock itself, shown by the horizontal dash-and-dot lines, is given by

$$U^2 \simeq \frac{P_b}{\rho\beta_a} = \frac{P_a}{\rho\beta_b}, \tag{4}$$

where the subscripts a and b refer to conditions ahead and behind the wave respectively. The two panels of Fig. 10 exhibit a major qualitative difference. In the case of the smooth shock, there is a substantial range of frequencies where monochromatic waves near the front and the back propagate faster than the shock itself: any perturbation falling in these frequency ranges will therefore propagate away from the shock. The shock is only faster than very low-frequency, long-wavelength perturbations - which exceed the scale of the shock and therefore are not generated - or than high-frequency, highly damped ones. In this situation the integrity of the shock can only be maintained by nonlinear effects and the profile is smooth. In the case of the oscillating shock, however, the shock speed is faster than that of any wave near its front, which is therefore steep and smooth. In the intermediate shock region, however, waves having a speed comparable with that of the shock are possible, and it is these waves that remain "trapped" in the shock and confer to it the oscillatory nature seen in the Fig. 10.

5. CONCLUSIONS

Reasons of space have only allowed me to give the briefest indication of some current problems in the mechanics of bubbles and bubbly liquids. It has not been possible to even mention many important topics such as cavitation - both flow and acoustic - bubble nucleation, break-up, and coalescence, air entrainment in flowing liquids, biomedical applications, micro-gravity manufacturing, boiling heat transfer, and many others. Gas-liquid multiphase flows are of great relevance e.g. in the power generation and oil industries and very poorly understood. One can expect theoretical as well as experimental challanges in these areas for many years to come.

In conclusion I would like to return for a moment to the single pulsating bubble with an idea that is perhaps far-fetched but nevertheless intriguing. It is possible that here we shall find a "power plant" for the micro-robots of the future. It is not inconceivable that such machines could be used e.g. inside the human body where power from any source other than acoustic would be impractical.

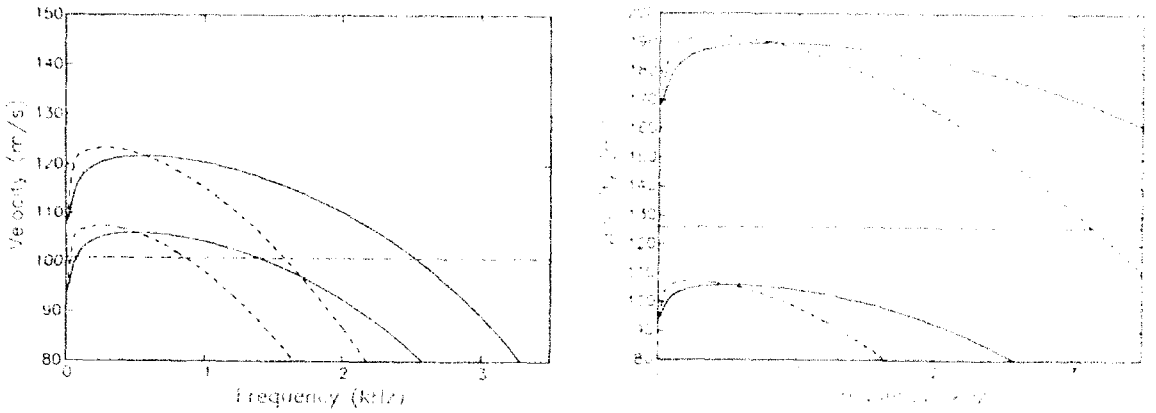


Figure 10: Dispersion relations for the two shock waves of Fig. 9. The horizontal line is the wave velocity, the solid lines the phase velocity and the dashed lines the group velocity. The upper pair of curves is for conditions at the back of the wave, and the lower pair for conditions near the front of the wave. (From Ref. 16.)

ACKNOWLEDGMENTS

The author wishes to express his gratitude to several persons who have contributed to the present paper: H.N. Oğuz, M. Nicholas, R.A. Roy, L.A. Crum, L.H. Prosperetti, X. Cheng, and M. Watanabe.

The work described here has been supported by the National Science Foundation (grant no. CBT-8918144), the Department of Energy (grant no. DE-FG02-89ER14043), and the Ocean Acoustics Program of the Office of Naval Research.

APPENDIX – AN ETYMOLOGICAL NOTE

According to Ref. [17], the Greek word for bubble, $\pi o \mu \phi \acute{o} \lambda \upsilon \xi$, is related through an intermediate variant $\pi o \mu \phi \acute{o} \varsigma$ to $\pi \acute{\epsilon} \mu \phi \iota \xi$ which has the basic meaning of "something that is created by inflating or blowing." Interestingly one of the etymologies offered for the Latin *bulla* is from the root $\phi \alpha \lambda -$ (whence e.g. *phallus*), indicating "any object swelling up" [18]. However, there does not seem to be any consensus on this point as other authors indicate a derivation from the sanskrit *belih*, which denotes the external female genitals [19]. Here the connection might be through the idea of "rotundity." Another Greek word sometimes juxtaposed might be $\beta o \lambda \beta \acute{o} \varsigma$, "bulb" [17].

Be that as it may, it seems very likely that onomatopoeic elements have also shaped the word, as we find the basic sound in unrelated languages such as Sumerian (*bubul*), Accadian (*bubutu*), Hindi (*bud-bud, bull-bula*), Bengali (*bud-bud*), Cebuano-Visayan (*bulâ*), and others.

As for the English *bubble* and its ancestors *boble, bobel, bubbul*, the onomatopeic element is probably dominant as also indicated by the closeness with *burble* [20]. Parallel

substantives in other Teutonic languages are e.g. the Swedish *bubbla*, the Danish *boble*, and the dialectal German words *bobbel*, *bubbel*. The current German word for "bubble" is *blase*, similar to the Dutch *blaas*, "bladder."

All the modern Romance languages have a word deriving from the Latin *bulla*, although sometimes the meaning of "bubble" has been lost. An example is Spanish where the word *bula* now primarily denotes the seal apposed to Papal documents, but was in the past used for *burbuja*, that is the modern word for "bubble."

REFERENCES

1. S. Schama, *The Embarrassment of Riches*, Knopf, New York, 1987, pp. 512-516.

2. W. Stechov, *The Art Bulletin* **20**, 227-228 (1938).

3. R.E. Verral and C.M. Sehgal, in K.S. Suslick (ed.), *Ultrasound: its Chemical, Physical, and Biological Effects*, VCH Publishers, New York, 1988, pp. 227-286.

4. V. Kamath, A. Prosperetti, and F. Egolfopoulos, *J. Acoust. Soc. Am.*, in press.

5. C. Sehgal, R.G. Sutherland, and R.E. Verrall, *J. Phys. Chem.* **84**, 388-395 (1980).

6. B.P. Barber and S.J. Putterman, *Nature* **352**, 318-319 (1991); B.P. Barber, R. Hiller, K. Arisaka, H. Fetterman, and S.J. Putterman, *J. Acoust. Soc. Am.* **91**, 3061-3063 (1992).

7. H.G. Flynn, Method of generating energy by acoustically induced cavitation fusion and reactor, U.S. Patent No. 4,333,796 (1982).

8. A. Prosperetti, L.A. Crum, and H.C. Pumphrey, *J. Geophys. Res.* **94**, 3255-3259 (1989); H.N. Oğuz and A. Prosperetti, *J. Fluid Mech.* **219**, 143-179 (1990); H.N. Oğuz and A. Prosperetti, *J. Fluid Mech.* **228**, 417-442 (1991); H.N. Oğuz and A. Prosperetti, *Ann. Rev. Fluid Mech.*, in press.

9. see e.g. B.R. Kerman, *J. Acoust. Soc. Am.* **75**, 149-165 (1984); P.C. Wille and D. Geyer, *J. Acoust. Soc. Am.* **75**, 173-185 (1984); D.J. Kewley, D.G. Browning, and W.M. Carey, *J. Acoust. Soc. Am.* **88**, 1894-1902 (1990).

10. M.S. Plesset and A. Prosperetti, *Ann. Rev. Fluid Mech.* **9**, 145-185 (1977).

11. W.M. Carey, *J. Acoust. Soc. Am.* **78**, S1-S2 (1985); W.M. Carey and D. Browning, in B.R. Kerman (ed.), *Sea Surface Sound*, Kluwer, Boston, 1988, pp. 361-376; W.M. Carey, J.W. Fitzgerald, and D.G. Browning, in B.R. Kerman (ed.), *Natural Sources of Underwater Sound*, Kluwer, Boston, in press.

12. A. Prosperetti, *J. Acoust. Soc. Am.* **78**, S2 (1985); *J. Acoust. Soc. Am.* **84**, 1042-1054 (1988); A. Prosperetti, N.Q. Lu, and H.S. Kim, *J. Acoust. Soc. Am.*, in press.

13. N.Q. Lu, A. Prosperetti, and S.W. Yoon, *IEEE J. Ocean Eng.* **15**, 275-281 (1990); S.W. Yoon, L.A. Crum, A. Prosperetti, and N.Q. Lu, *J. Acoust. Soc. Am.* **89**, 700-706 (1991); M. Nicholas, L.A. Crum, R.A. Roy, H.N. Oğuz, and A. Prosperetti, *J. Acoust. Soc. Am.*, submitted.

14. K.W. Commander and A. Prosperetti, *J. Acoust. Soc. Am.* **85**, 732-746 (1989).

15. L. Noordzij and L. van Wijngaarden, *J. Fluid Mech.* **66**, 115-143 (1974).

16. M. Watanabe and A. Prosperetti, *J. Fluid Mech.*, submitted.

17. P. Chantraine, *Dictionnaire Etymologique de la Langue Grecque*, Klincksieck, Paris, 1968.

18. C.T. Lewis and C. Short, *A Latin Dictionary*, Clarendon, Oxford, 1975 (first edition 1879).

19. A. Ernout and A. Meillet, *Dictionnaire Etymologique de la Langue Latine*, Klincksieck, Paris, 1959.

20. J.A. Simpson and E.S.C. Weiner eds., *Oxford English Dictionary*, Clarendon, Oxford, 1989.

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
1993 IUTAM.

Wave propagation in non-isotropic structures

Mahir Behar Sayir

Institute of Mechanics, ETH, Zurich, Switzerland

Abstract

In structural wave propagation the characteristic wavelengths of the propagating pulse are chosen to be sufficiently large with respect to the thickness of the structure. Narrow-band transient structural waves are particularly well-suited for example to study and quantitatively evaluate global ("homogenized") dynamical properties of sandwich or fiber-reinforced composite structures such as laminates, tubes etc. The waves are generated in various membrane or flexural modes by piezo-electric transducers excited at narrow frequency bands well below their resonant frequencies to avoid coupling and thus allow perfect control of the pulse shape. The theoretical models used to interpret and evaluate the experimental results obtained in sandwich and composite structures must be carefully developed in order to predict accurately the features following from the interaction of the various geometric, dynamic and material scales. A few experimental results obtained in recent years in our laboratory are discussed in conjunction with the corresponding theoretical models. Part of these results are also used to interpret dynamical phenomena observed during the propagation of fast-running cracks in cross-ply laminates.

1. TERMINOLOGY AND MOTIVATION

Consider a transient wave in the amplitude range of linear elasticity in a thin structure such as a beam, plate or shell . The time function is supposed to be of *narrow-band* type and can be generated by applying a suitable window (for example a translated cosine-window of relatively large duration) on a sine- or cosine-function of a given frequency. In the amplitude-frequency spectrum of this function the non-vanishing amplitudes will lie for all practical purposes between two frequency limits. The higher limit f_{max} will correspond in the structure to a wavelength λ which we call the *characteristic wavelength* of the travelling pulse. In a *structural wave* the ratio of the largest thickness H of the structure to the characteristic wavelength of the propagating pulse (called *"dynamic parameter"*) is supposed to be sufficiently small, i.e.

$$\varepsilon_\lambda := H / L << 1$$

The stress- and strain-distributions generated by such a wave across the thickness of the structure are stationary and usually quite simple mathematical functions of the thickness coordinate. The wave travels then in *membrane, torsional* or *flexural* modes, with possible coupling between them. Furthermore, in fiber-reinforced composite structures, due to the much lower scale of the fiber diameter or of the ply thickness, most of the important features of structural wave propagation can be interpreted by assuming homogeneous material behaviour.

Per definition in anisotropic materials with orthotropic symmetry three mutually perpendicular principal axes of material symmetry can be associated with each material point. With respect to structural waves structures may behave orthotropically either because the material itself is orthotropic or because a combination of geometric features and material distribution through the corss-section of the structure generate "globally" non-isotropic structural behaviour in spite of the fact that the material itself is isotropic. An interesting case of such "*structural non-isotropy*" has been studied in [1], where a flexural wave in an aluminum beam with nearly perfect circular cross-section (measured deviations from a circular boundary less then 0.05 %) "rotated" most probably because of an extremely slight inhomogeneity in the distribution of the axial modulus of elasticity E through the cross-section (largest difference in the weighted average of the modulus in two perpendicular directions of the order of 1 %). The material itself was shown to behave istoropically within the accuracy of the measurements.

In isotropic materials the ratio of the shear modulus G to the tensile modulus E

$$\varepsilon_M := G / E$$

which will be called "*material parameter*" in the following has to be larger than 0.33.for reasons of material stability. Of course in anisotropic structures such as fiber-reinforced laminates or filament-wound tubes this material parameter may reach much lower values (for example 0.01). Also in sandwich structures with a soft isotropic core and stiff isotropic facings this parameter (determined from equivalent stiffness values) is usually quite low. In our laboratory sandwich beams and plates with ε_M as low as 0.0004 have been tested [2]. Thus from the point of view of effects associated with low values of the material parameter ε_M (for example transverse shear coupling) sandwich structures may behave non-isotropically even though the material of the components is isotropic.

Our experimental research activity on wave propagation in non-isotropic structures which started about 15 years ago was first motivated by the need to reach a better understanding of coupling phenomena which are characteristic of non-isotropic behaviour. Thus to study mode-coupling in non-principal directions we investigated flexural waves in unidirectionally reinforced plates [3] or in-plane waves in cross-ply laminates [4, 17]. To study coupling phenomena related to low values of ε_M we generated flexural waves in unidirectionally reinforced beams [5], in sandwich-plates with extremely soft cores [2] or axisymmetric waves in filament-wound tubes [6, 4]. These experimental investigations were interpreted by developing for each case appropriate theoretical models based on a careful asymptotic analysis of the basic relations of three-dimensional linear elasticity. Approximations based on a priori assumptions were avoided since in all cases the presence of at least two competing small parameters, namely ε_λ, ε_M and of some other small geometric ratios (as in sandwich plates or filament-wound tubes) rendered the choice of physically justified simplifying assumptions without a straightforward ordering scheme rather difficult. With a more thorough knowledge of the above-mentioned coupling phenomena we were also able to determine structural properties such as flexural stiffnesses and various

moduli of elasticity with quite high degrees of accuracy [7]. Our experimental and theoretical methods of analysis can also be used to detect flaws or structural errors which affect local stiffness values to some extent, even though the characteristic wavelengths of the generated pulses may be much larger than the size of the flaws. Especially flexural waves which by nature are particularly sensitive to changes in thickness are well suited to detect flaws such as delaminations or to check for possible errors in ply combinations, filament-wounding etc. Details related to current research on applications to non destructive detection will be reported elsewhere. In the following I will present a review of the main features related to the experimental and theoretical work briefly mentioned above, together with some unpublished recent results.

2. SHEAR COUPLING IN FLEXURAL WAVES

2.1 Unidirectionally reinforced beams

In [8] a theory for flexural waves with large wavelengths in unidirectionally reinforced beams was elaborated. According to this theory which was derived by asymptotic analysis of the three-dimensional equations of linear elasticity, the order of magnitude of the parameter

$$p := \pi \frac{H}{\lambda} \sqrt{\frac{E}{G}} \tag{2.1}$$

determines decisively the propagation characteristics of transverse structural waves in beams. The following table summarizes the role of the parameter p:

Order of magnitude	Degree of anisotropy	Nature of the shear coupling
$p \ll 1$	Isotropy, Weak or moderate anisotropy	*Weak*: Becomes active at the *second* step of approximation
$p = O(1)$	"Quasi-strong" anisotropy (transitional case)	*Strong*: Becomes active at the *first* step of approximation
$p \gg 1$	Strong anisotropy	*Sandwich-behaviour*

In the transitional case of "quasi-strong" anisotropy with $p = O(1)$ the usual assumption that cross-sections remain plane after deformation does not hold since the transverse shear strains become so large that they reach the magnitude of the axial strains due to bending. Because of this strong shear coupling, the variations of the axial displacements along the axis of the beam are connected with their variations through the cross-section. The axial strains and stresses are then in general nonlinearly distributed through the cross-section. For sinusoidal waves of a given amplitude of transverse displacement the amplitudes of distributions of axial strain and stress are illustrated in Fig. 1 for a rectan-

gular cross-section. In [8] distributions for a circular cross-section were also given. As p takes larger values, the nonlinearity of the strain and stress distribution becomes also more pronounced. As one reaches values of p much larger than 1, axial strains and stresses concentrate on the parts of the cross-section which are near the lateral surfaces of the beam. A core of finite thickness frees itself almost completely of axial stresses and carries only a constantly distributed shear stress (fig. 2). Thus the unidirectionally reinforced beam behaves like a "natural" ideal sandwich structure.

In [8] the dispersion relation

$$c := (1 - \frac{\mathrm{Th}\, p}{p})^{1/2} \sqrt{\frac{G}{\rho}} \tag{2.2}$$

(c: phase velocity of flexural waves, ρ: specific mass per unit volume) was derived for the transitional case of "quasi-strong" anisotropy in a beam with rectangular cross-section. Since this case corresponds to a "distinguished limit" in the asymptotic sense, it can also be used in the case of moderate anistropy by expanding for small p. One obtains eventually

$$c := \frac{p}{\sqrt{3}} (1 + \frac{2}{5} p^2)^{-1/2} \sqrt{\frac{G}{\rho}} = f_S \frac{k H}{\sqrt{12}} \sqrt{\frac{E}{\rho}} \quad , \quad f_S = \frac{1}{\sqrt{1 + \frac{k^2 H^2}{10} \frac{E}{G}}} \tag{2.3}$$

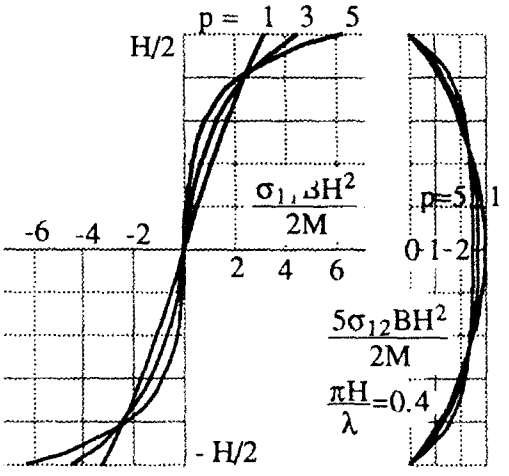


Fig. 1: Distribution of normal and shear stress through the cross-section for p = O (1)

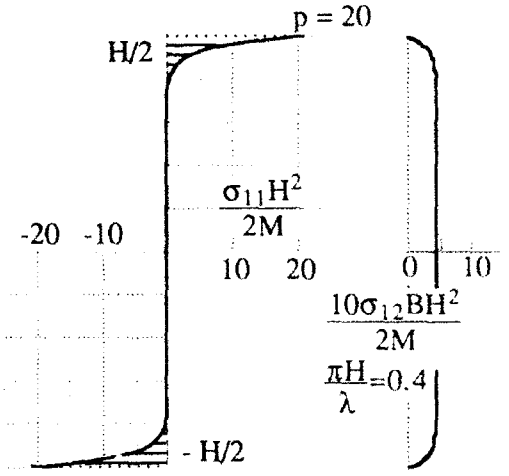


Fig. 2: Sandwich behaviour for strong anisotropy with p >> 1

(k = 2 π / λ = "wave number"). This relation has also been derived directly in [7] for a plane bending wave travelling in one of the principal directions of structural symmetry in an *orthotropic plate*. It includes with the factor f_S the correction for shear deformation but excludes the effect of rotary inertia which for large ratios E/G (for example larger than 10

as in most applications to fiber-reinforced structures) becomes indeed negligible. In [7] we showed how this relation could be conveniently used to measure "global" values of E and G in the principal directions of an orthotropic plate in bending mode (see also Section 4 of the present paper).

Experimental work to check the dispersion relations for phase velocities as a function of frequency or wave number calculated in [8] was performed by Kolsky and Mosquera [9] who used beams with a rubber matrix reinforced by steel wires. They covered the range of strong anisotropy, changed the value of the parameter p by varying the frequency and obtained excellent agreement with the theoretical predictions. Experimental work was also performed by Goodbread and myself in our laboratory using beams with an Araldite matrix reinforced by Kevlar 49 fibers at 60 volume %. Results were briefly presented at ICTAM, Toronto in 1980 and published much later in [5]. To vary the parameter p we investigated not only various frequencies but tried also to change the material parameter ε_M by heating the beam in a constant temperature oven. Comparing the phase spectra of the measured transient pulses at two different locations we obtained the flexural phase velocities both as a function of temperature and frequency.and compared them with theoretical predictions using values of tensile and shear moduli determined in independent resonance experiments. The results given in some detail in [5] showed excellent agreement between theory and experiment.

2.2 Sandwich plates

To study strong shear coupling further M. Koller built in our laboratory a sandwich plate exhibiting ideal sandwich structural behaviour in flexure. In this case only the facings should contribute to the bending moment and only the core should carry the shear force. Such behaviour is assumed a priori in most of the theoretical work on sandwich structures (see for example [10, 11]). Our study [2] was aimed at determining the limits of applicability of the idealized theoretical models. Aluminum or glass facings of thickness h = 0.5 mm were combined with a polyurethan or polyethylen core of thickness 2H = 3 mm to obtain a sandwich plate with the desired characteristics. The ratio of the shear modulus of the core G_c to the tensile modulus E_f of the facings was in the range 0.26 - 1.3 10^{-3}. Koller used the plates with glass facings to study impact by steel balls of various sizes (radius varying from 1 to 10 mm) in the elastic range [12]. The peculiarities of sandwich behaviour as compared to an isotropic structure can also be demonstrated under static loading. Fig. 3 illustrates two examples of the differences in static response of sandwich beams as compared to the isotropic counterparts. The original photographs of the actual experiments are published in [2]. Koller showed in his impact experiments that whereas in a thin isotropic plate the cross-sections in the impact zone are translated rapidly and remain stationary until waves are reflected from the edges of the plate, in a sandwich plate, due to the shear compliance of the core which acts like a spring, the cross-sections bounce back after reaching a maximum displacement value and remain stationary at a lower amplitude [12].

The theoretical part of our study in [3] was developed by asymptotic analysis of the three dimensional basic relations of linear elasticity for isotropic behaviour of both facings and core. Four competing small dimensionless parameters

$$\varepsilon_\lambda := \frac{H+h}{\lambda} \quad , \quad \varepsilon_h := \frac{h}{H} \quad , \quad \varepsilon_\rho := \frac{\rho_c}{\rho_f} \quad , \quad \varepsilon_M := \frac{G_c}{E_f}$$

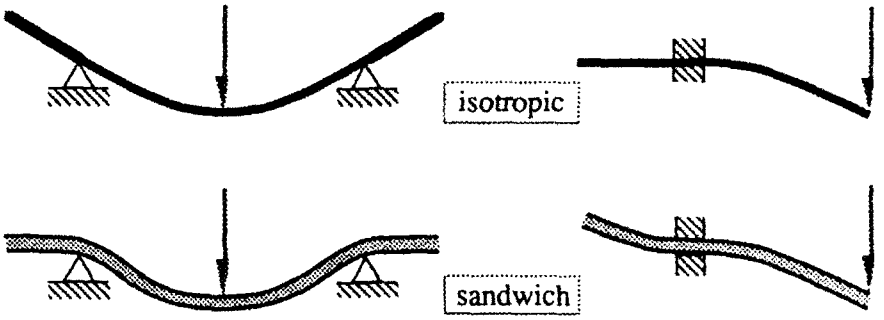


Fig. 3: Differences in the static behaviour of a sandwich beam and of its isotropic counterpart

and their relation to each other had to be considered (ρ_c, ρ_f : specific mass of the core or facings). Provided that ε_M is sufficiently small with respect to $(\varepsilon_h)^3$, three wavelengths of decreasing orders of magnitude in the range of structural wavelengths (large compared to the total thickness 2H + 2h of the plate) can be defined. The role of these wavelengths with respect to the characteristic wavelength of the propagating pulse is characterized in the table below.

Characteristic wavelength λ comparable with	Main dynamic contribution of the facings besides inertia	Main dynamic contribution of the core besides inertia
$\lambda_1 := 2\pi H\sqrt{\frac{E_f}{G_c}\frac{h}{H}}$	In-plane forces through in-plane deformation	Transverse shear force
$\lambda_2 := 2\pi H\sqrt{\frac{E_f}{G_c}\frac{h^2}{H^2}}$	Only second order values	Transverse shear force
$\lambda_3 := 2\pi H\sqrt{\frac{E_f}{G_c}\frac{h^3}{H^3}}$	Bending moments through flexural deformation	Transverse shear force

For wavelengths comparable with λ_1 the dispersion behaviour (phase velocity as a function of frequency or reciprocal wavelength) is similar to the corresponding behaviour of a fully isotropic homogeneous plate but is strongly influenced by shear deformation even at relatively low frequencies. For wavelengths comparable with λ_2 the phase velocity takes values in the vicinity of the shear wave velocity of the core corrected by the inertia of the facings. If the wavelength is comparable with λ_3 (which for ε_M sufficiently large

with respect to $(e_h)^3$ is still larger than the total thickness of the plate) the flexural behaviour of the sandwich plate is dictated by the flexural properties of the facings which behave like thin plates on an elastic foundation with appreciable shear stiffness and negligible compressive or tensile stiffness. As reported in [3] the theoretically predicted dispersion curve in the latter range of wavelengths comparable with λ_3 has been checked experimentally for sandwich plates designed to fulfill the requirements on the four small parameters mentioned above. Radially propagating flexural waves were generated with the help of a piezoelectric transducer glued on one of the facing surfaces.sufficiently far from the edges to avoid interferences with reflections during the recording of the transient pulse. The phase velocity was obtained as a function of frequency by comparing the phase spectra of the displacements at two or more radial locations on the plate. The transient signals were measured with laser-interferometry, fed into a transient recorder and Fourier-analyzed subsequently. The fit between theoretical and experimental results was remarkable, since all parameters used in the theoretical predictions (E_f, G_c, etc) were measured in independent experiments. The results are illustrated in Fig. 4. A less spectacular but still satisfactory agreement between theory and experiment was obtained in the elastic impact study carried out by Koller [12].

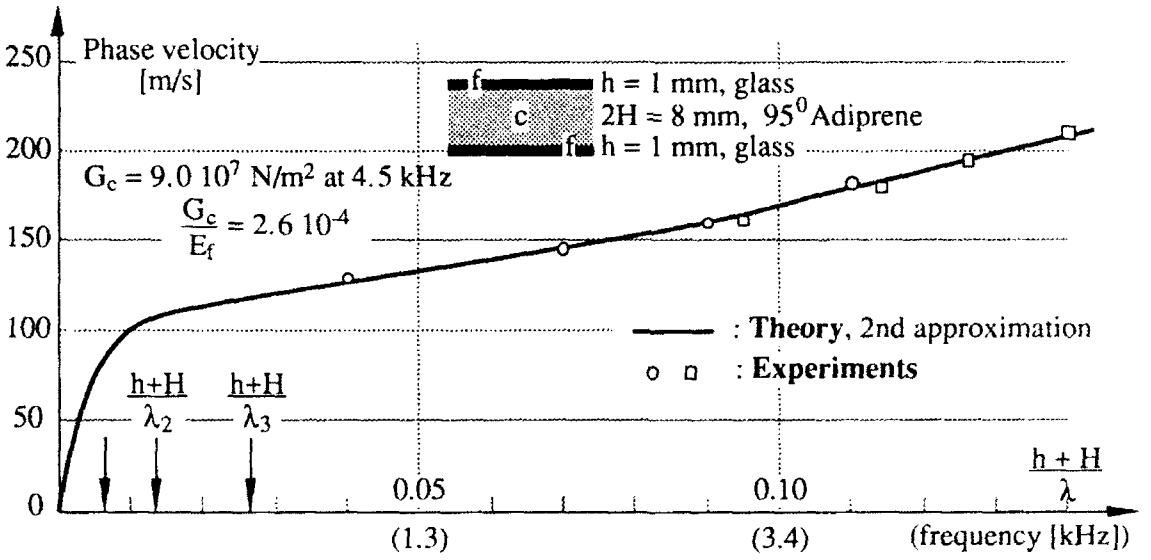


Fig. 4: Theoretical and experimental dispersion curves for sandwich plates

Current research work still in progress concentrates on certain aspects of damage caused on plates with both isotropic and fiber-reinforced facings. In another current research project possibilities of non-destructive detection of damage in sandwich plates are being investigated.

3. FLEXURAL WAVES IN UNIDIRECTIONALLY REINFORCED PLATES

In the sandwich plates mentioned in the previous section, since both the facings and the core were isotropic, the sinusoidal "components" of the flexural waves propagated in the radial direction generating phase cylinders of circular shape (*phase cylinder:* cylindrical surface of constant phase, its axis is perpendicular to the plate and its intersection with the plate surface is a *phase curve*). Thus the phase velocities were independent of the radial direction along which they were measured. In a unidirectionally fiber-reinforced composite plate loaded in the same manner with the help of a transducer mounted on one of its two outer surfaces the phase of a sinusoidal component is also expected to be constant along cylindrical surfaces but of course these have no circular symmetry (Fig. 5).

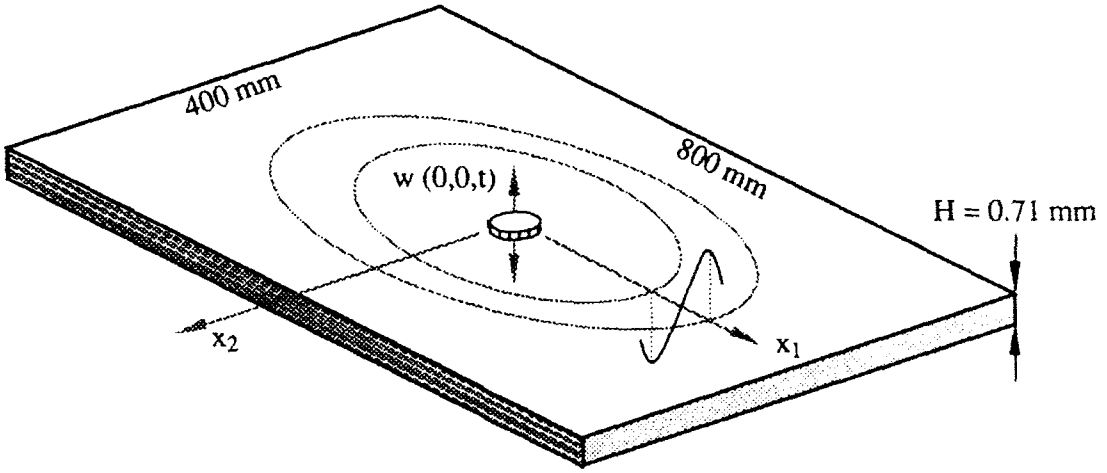


Fig. 5: Flexural waves generated by a transducer in a unidirectionally fiber-reinforced plate

The main purpose of the study described in [3] and [13] was to investigate the shape of the phase curves and to determine the dependence of the radial phase velocity on the angle φ with respect to the fiber direction x_1. A unidirectionally reinforced laminate made out of 5 plies of 0.14 mm thickness of carbon-reinforced epoxy was subjected to transverse loading with the help of a piezoelectric transducer. The transient signal was recorded at various radial directions and distances from the transducer. From the difference of the phase spectra of the measured signal at two radial locations with the same angle with respect to the fiber direction the wavelength and hence the radial phase velocity was determined as a function of frequency and angle of propagation. My former doctoral student A. Kreis who performed the experiments concentrated on the frequency value f = 29.3 kHz and measured for example in the fiber direction the wavelength λ_1 = 20.3 mm and in the direction perpendicular to the fibers λ_2 = 10.7 mm. Since λ_1 was about 30 times and λ_2 about 15 times larger than the total thickness H = 0.71 mm of the plate, shear deformation effects were judged to be too small to affect the behaviour at the mentioned frequency and a simple first order theory was used to derive the governing differential equation for the

transverse displacement w (t, x_1, x_2). The following equation was derived by asymptotic analysis of the three dimensional basic equations (derivatives are indicated as usual with a comma, the symbols for the various moduli of elasticity are self-explanatory):

$$E_1 w,_{1111} + 2 (\nu_{12} E_2 + 2 G_{12}) w,_{1122} + E_2 w,_{2222} + \frac{12 \rho}{H^2} w,_{tt} = 0 \quad .$$

For sinusoidal time functions of circular frequency ω and with the coordinate transformations

$$x := x_1 \left[\frac{12 \rho \omega^2}{E_1 H^2}\right]^{1/4} \quad , \quad y := x_2 \left[\frac{12 \rho \omega^2}{E_2 H^2}\right]^{1/4}$$

the differential equation above was simplified to obtain

$$\overline{w},_{xxxx} + \overline{w},_{yyyy} + 2 \delta \overline{w},_{xxyy} = \overline{w} \qquad (3.1)$$

for the amplitude function $\overline{w}$ (x,y) where

$$\delta := \frac{E_2 \nu_{12} + 2 G_{12}}{\sqrt{E_1 E_2}}$$

is supposed to be small and was evaluated a posteriori on the basis of the experimental results as 0.29.

Kreis tried to solve (3.1) for the case of concentrated load by Fourier-transform and contour-integration to perform the inverse transform. Some of the integrals resulting from this approach had to be evaluated numerically. The solution of (3.1) contains not only the cylindrical wave (without circular symmetry) generated by the transducer but also the stationary motion whose amplitude decays strongly with increasing distance from the transducer and whose role is to supply to the travelling wave the appropriate rhythm of exchange of potential and kinetic energy, thus allowing propagation of the energy introduced by the concentrated load. Since all measurements were performed at a sufficient distance from the transducer where for all practical purposes the stationary motion had fully decayed, only the part of the solution which describes the propagating wave is really of interest in the evaluation of the experimental data. To obtain this part A. Kreis used the method of stationary phase on the expressions for the inverse Fourier-transform [3]. In a different approach I tried recently to obtain a physically meaningful approximate solution for the differential equation (3.1) by direct asymptotic analysis, assuming that the coordinates x, y correspond to "large" radial distances from the transducer (distances more than three times larger than the diameter of the transducer). The following procedure proved to be quite efficient:

Set

$$\overline{w}\ (x,y) = \exp\ [u\ (x,y)]$$

in the differential equation (3.1) and obtain a differential equation of the form

$$D_1\ [u] + D_2\ [u] + D_3\ [u] + D_4\ [u] = 1 \qquad (3.2)$$

where D_k [.] are nonlinear differential operators involving derivatives of order k = 1, 2, 3 or 4. It can be shown that for large distances from the origin (0,0), in a first step of approximation, the terms with the higher order differential operators can be neglected and the differential equation is reduced to

$$D_1 [u] := (u,_x)^4 + (u,_y)^4 + 2\,\delta\,(u,_x)^2\,(u,_y)^2 = 1 \quad .$$

For $\delta = 0$ its exact solution corresponding to our problem is

$$u_0 (x,y) = - i\, f_0 (x,y) = - i\, [x^{4/3} + y^{4/3}]^{3/4} \quad .$$

For small $\delta > 0$ the approximate solution

$$u (x,y) = - i f (x,y) = - i\, [(1 + \frac{2}{9}\,\delta^2)\,(x^{4/3} + y^{4/3}) - \frac{2}{3}\,\delta\, x^{2/3}\, y^{2/3}]^{3/4} + O(\delta^3) \qquad (3.3)$$

can easily be derived by regular asymptotic expansion in terms of δ. Thus the phase curves of the propagating wave are given by f (x,y) = constant. The next steps of approximation in the asymptotic approach to (3.2) involve increasing orders of the operators D_k [.]. Eventually one obtains a solution of the form

$$\overline{w}\,(x,y) = \frac{g\,(y/x)}{\sqrt{x^2 + y^2}}\,\exp\,[-i\,f\,(x,y)] \quad . \qquad (3.4)$$

The factor $(x^2 + y^2)^{-1/2}$ is obtained by considering the operator D_2 [u] in (3.2). The function g (y/x) which determines the dependence of the amplitude from the angle of propagation for large distances from the transducer follows by considering the operator D_3 [u] in (3.2).

The asymptotic expansion (3.3) in terms of δ becomes singular near the axes. The word "near" means in this case either $y/x = O(\delta^{3/2})$ ("near" the x-axis) or $x/y = O(\delta^{3/2})$ ("near" the y-axis). Performing the usual singular perturbation analysis with the appropriate boundary layer variable one obtains for example near the x-axis a solution of the type (3.4) with

$$f\,(x,y) = x\,\{1 + \frac{1}{2\,\delta}\,(\frac{y}{x})^2 + O\,[\frac{1}{\delta^4}\,(\frac{y}{x})^4]\} \quad . \qquad (3.5)$$

Evaluating (3.5) to obtain the radial phase velocity at small angles with respect to the fiber direction in a unidirectionally reinforced plate, one can show that its value is quite sensitive to small changes in this angle. In the particular plate which was tested in our laboratory, the radial phase velocity at an angle of 7^0 was 10 % lower with respect to its value along the fiber direction. Thus to obtain accurate values of the tensile modulus near the fiber direction by measuring the corresponding phase velocity, particular care must be taken to remain in the proper radial direction while choosing the locations where the displacement is measured.

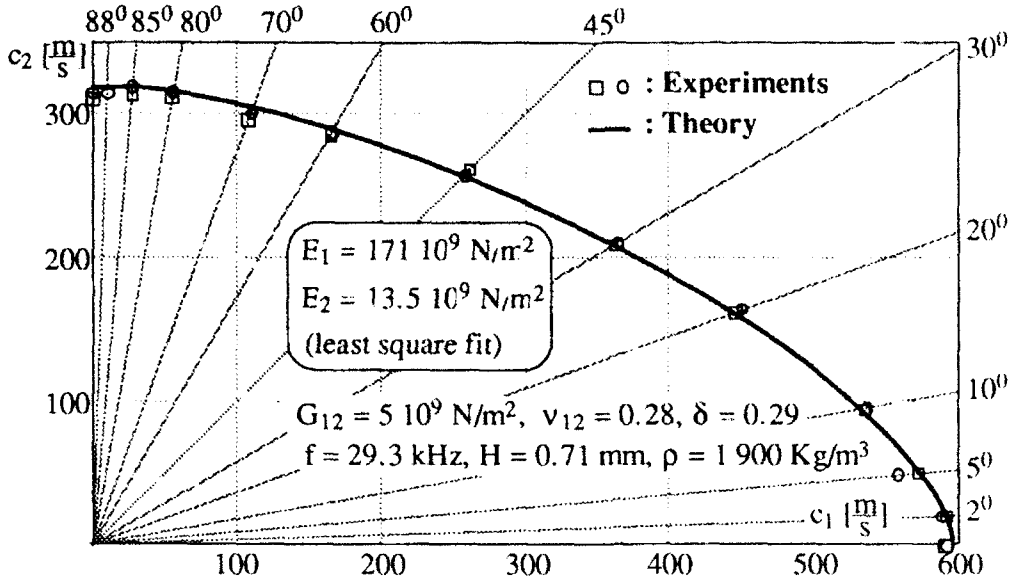


Fig. 6: Theoretical phase curve fitted to experimental results by choosing E_1, E_2

Fig. 6 shows phase velocities obtained from wavelengths measured in various radial directions. The theoretical phase curve corresponding to the developments described above has been fitted to the experimental results by proper choice of two parameters, i.e. the two tensile moduli E_1 and E_2 corresponding to the fiber direction and the direction perpendicular to it respectively. The dimensionless parameter δ which includes both Poisson's ratio ν_{12} and the shear modulus G_{12} was calculated using values of these two material constants determined independently. Since 2×13 experimental values have been fitted with only 2 parameters, the obtained results confirm quite successfully (within 5.5 m/s variance with respect to measured phase velocities) both the experimental investigation and the theoretical development.

4. EVALUATION OF FLEXURAL MODULI IN ORTHOTROPIC LAMINATES

The experimental methods developed to study the propagation flexural waves with (non circular) cylindrical phase surfaces in unidirectionally reinforced laminates have been subsequently refined to allow the systematic experimental evaluation of global flexural moduli in orthotropic laminates with arbitrary symmetric lay-up. In spite of the apparent success illustrated in Fig. 6 where, by appropriately choosing values of E_1 and E_2, an excellent fit with measured phase velocity values along different radial directions was obtained, the method described in Section 3 has mainly three weaknesses:

a) Although the wavelength of $\lambda_1 = 20.3$ mm measured in the fiber direction x_1 was about 30 times larger than the total plate thickness $H = 0.71$, the assumption that shear deformation effects can be neglected is not quite justified. This is immediately shown by evaluating the factor f_S in (2.3) with the values $k_1H = 2\pi H/\lambda_1 = 0.220$ and $E_1/G_{13} = 34.2$. One obtains $f_S = 0.926$. This is quite different from the ideal value of 1 for which shear deformation would be vanishingly small. In fact a more accurate analysis including shear deformation effects according to (2.3) reveals that the experimental data leads to a value of E_1 along the fibers of about 200 10^9 N/m^2 which is 17 % higher than the value obtained in Section 3.

b) Curve fitting with the phase curve delivers for all practical purposes only E_1 and E_2. The other moduli such as G_{12}, G_{13}, G_{23} and ν_{12} needed for an accurate fitting have to be determined from additional independent experiments.

c) To obtain the phase curve experimentally a series of $2n$, preferably $3n$ measurements must be performed where n is the number of different radial directions along which the phase velocity has to be determined. To measure along different radial directions requires a certain preparation which costs time.

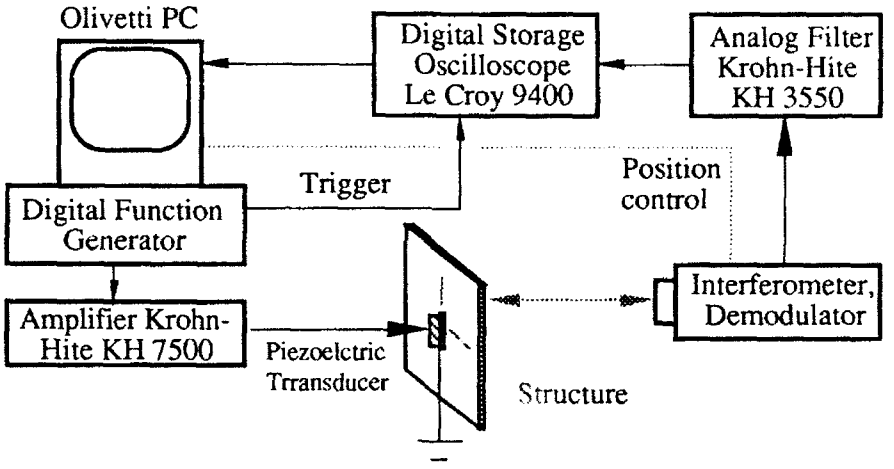


Fig. 7: Experimental set-up for a systematic evaluation of laminate stiffneses

More satisfactory results were obtained by measuring phase velocities only along the two principal axes of orthotropy (n = 2). To ensure that in spite of the various experimental inaccuracies to be usually expected in such a reduced programme, the accuracy of the global moduli resulting from the analysis remained sufficiently high, the phase velocity was measured as a function of *frequency* by comparing the phase spectra at two or more points along each of the two principal directions. Thus for each pair of measured signals a large number of phase velocity values for different frequencies were produced in the subsequent computer analysis and compared with theoretical predictions based on linear-elastic homogeneous behaviour including shear deformation effects. Theoretical results and experimental values were fitted (least squares) by appropriate choice of the global flexural tensile moduli E_1, E_2 and global transverse shear moduli G_{13}, G_{23} along the principal directions. However, for orthotropic laminates with arbitrary symmetric lay-up, one

has to bear in mind that the global moduli obtained from flexural modes are generally expected to be different from the ones which can be determined in membrane modes (see Section 6).

The set-up in our Laboratory corresponding to the systematic experimental evaluation mentioned above and main results for particular laminates which we produced at the Laboratory of Composite Design in our University thanks to the courtesy of Prof. M. Flemming have been described in some detail in [7]. Fig. 7 shows the main features of our experimental set-up. Narrow band flexural pulses with well-defined functions are introduced through a computer and transferred with the help of a function generator to a piezoelectric transducer properly glued at some central location of the surface of a laminate whose stiffnesses are to be determined. The flexural displacement of the plate is measured at two or more locations in the radial direction at some distance from the transducer by properly positioning (the position is computer-controlled) a *heterodyne laser interferometer* measuring displacements in the direction of the laser beam according to the Doppler-effect. The details of the interferometer which was originally built and further developed at our Laboratory by J. Goodbread have been described in [14]. The demodulated interferometer signal is fed into a digital storage transient recorder and subsequently transferred to the computer to be Fourier-analyzed and further processed. The main rules to be observed in such experiments have been listed in some detail in [4]. Here three essential features will be briefly mentioned:

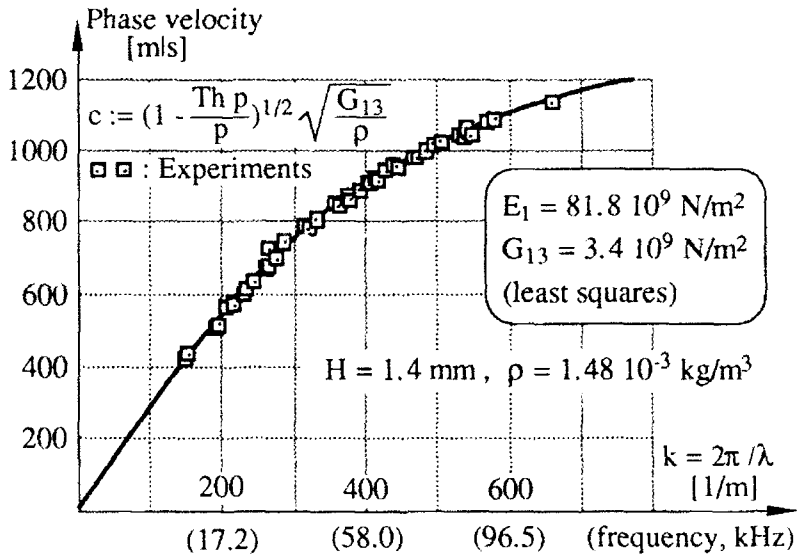


Fig. 8: Obtaining E_1 and G_{13} in a cross-ply laminate from data on phase velocities in flexure

(1) The frequency band of the signal should be chosen to lie well below the lowest resonance frequency of the transducer to avoid elastic coupling between the wave generator and the responding structure.

(2) The time duration of the signal and the frequency band of the signal should be coordinated with the location of the transducer to avoid reflections from the edges of the plate which could interfere with the measurements.

(3) Since the time function of the signal is well-defined, the same wave is generated a few hundred times at time intervals allowing waves generated by one signal to decay before the next signal is sent. By taking a simple additive average of the hundreds of responses measured at the same location (and generated with the same time function at the source) the noise to signal ratio is decreased considerably, since thanks to averaging stochastic noise due to various mechanic or electronic sources is either cancelled or reduced below levels of relevance.

M. Veidt measured at our Laboratory phase velocities as a function of frequency in carbon-fiber reinforced cross-ply laminates which were subsequently needed to study dynamic fracture phenomena (see Section 7). Since cylindrical flexural waves travel at some distance from the transducer along the principal directions of an orthotropic plate essentially like corresponding plane waves (see Section 3) and since in cross-ply laminates ν_{12} (corresponding to in-plane lateral contraction resilience) is very low (about 0.04) the result (2.2) for phase velocities can be directly used to analyze experimental data on phase velocities along the principal directions of orthotropy (in [7] the expanded form (2.3) was used). Fig 8 shows an example of fitting theoretical predictions to experimental data for various frequencies to obtain both the global tensile modulus E_1 and the transverse shear modulus G_{13} in flexural mode of a cross-ply laminate with a symmetric lay-up $[0, 90, 0, 90, 0]_S$.

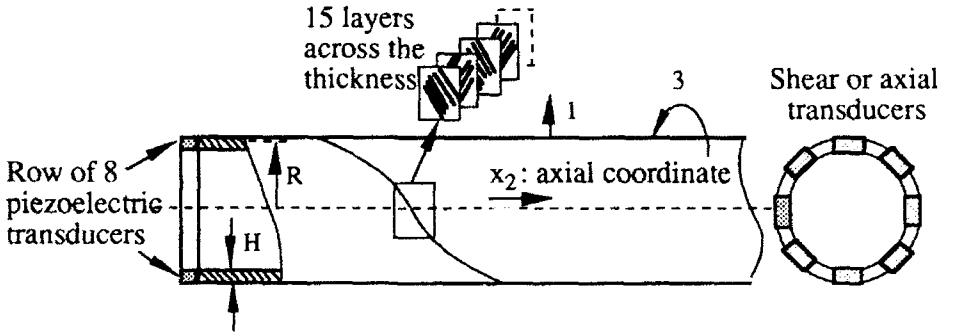


Fig. 9: Filament-wound tube axisymmetrically loaded at one end by piezoelectric transducers

5. AXISYMMETRIC WAVES IN FILAMENT-WOUND TUBES

Axisymmetric waves in isotropic tubes have been studied in great detail both experimentally and theoretically by a number of scientists (see for example [15]). In contrast, data on composite tubes are rather scarce. Thus, as part of his PhD work [6] J. Dual performed a series of systematic experiments on carbon-reinforced tubes produced by fila-

ment winding on a total length of 2 m with 12 000 filaments per roving and a total number of 15 layers (fig. 9). He considered 5 cases with winding angles 0^0, 22.5^0, 45^0, 67.5^0 and 82^0. The main purpose of the experiments was to measure the dispersion behaviour and to obtain the relevant moduli of the tube. The study should also serve as a basis for further investigations regarding non-destructive detection of possible flaws and structural errors. Dual used in his di sertation a theoretical framework based on numerical evaluations of power-series expansions of the full three-dimensional equations for homogeneous, orthotropic behaviour. Particularly efficient and accurate analytical expressions based on an asymptotic analysis of the three-dimensional equations for the range of structural wavelengths covered by the experiments have since been developed and will be published in [16]. The asymptotic analysis has also the merit of providing an extremely useful *physical insight* into the various patterns of deformation in the various ranges of wavelengths and into the corresponding propagation characteristics. A short report on main results of this analysis and of its confrontation with experimental data produced by Dual has been published in [4]. In the following a summary of the salient features will be presented.

The asymptotic analysis mentioned above is based on a careful exploitation of the three small parameters

$$\varepsilon_R := \frac{H}{R} \quad , \quad \varepsilon_\lambda := \frac{H}{\lambda} \quad , \quad \varepsilon_M := \frac{G_{12}}{E_2}$$

which are inherent to the problem of axisymmetric structural waves in a thin orthotropic tube. In the tubes tested the value of the geometric parameter ε_R was about 0.1. For the tubes with 0^0, 22.5^0 and 45^0 it was found a posteriori that $\varepsilon_M \approx 0.04$, 0.06 and 0.2 respectively. For the remaining two angles of winding the axial modulus E_2 in ε_M has to be replaced by the hoop modulus E_3. Thus in all tubes $\varepsilon_M = O(\varepsilon_R)$ could be assumed. This is the case which has been called in [8] "moderately strong anisotropy". Depending on the relation of ε_λ to ε_R the following three different ranges of wavelengths can be defined:

1) Wavelengths comparable with the perimeter of the tube, $\varepsilon_\lambda = O(\varepsilon_R)$

The waves propagate in a "membrane mode" with both radial (u_1) and axial (u_2) components of the displacement uniformly distributed across the thickness of the tube wall. The following explicit analytical expression for the phase velocity c as a function of frequency f follows from the simplified differential equations of the asymptotic analysis ($\nu \equiv \nu_{23}$):

$$C^2 = \frac{1 - \Omega^2}{1 - \Omega^2 (1 - e\nu^2)} \quad , \quad C := \frac{c}{\sqrt{E_2 / \rho}} \quad , \quad \Omega := \frac{2\pi R}{\sqrt{E_3 / \rho}} f \quad , \quad e := \frac{E_3}{E_2} \quad . \tag{5.1}$$

For values of $\Omega << 1$ corresponding to very large wavelengths (exceeding considerably the perimeter of the tube) the waves propagate nondispersively ($c = \sqrt{E_2/\rho}$, independently of the frequency as long as $\Omega << 1$) under uniaxial stress conditions. The main displacement component is u_2, radial displacements u_1 can be neglected. For increasing values of Ω, as long as $\Omega < 1$, the influence of the radial displacement component u_1 becomes stronger and causes more and more energy to be canalized to radial motion rather than axial wave propagation. Thus the value of phase velocity drops. For values or Ω near 1 the radial motion becomes so strong that energy transport in the axial direction becomes

almost impossible in the membrane mode ("radial resonance"). For increasing values $\Omega > 1$, the radial motion decreases and again axial displacements become more important. Since in this case the wavelengths are smaller than the perimeter (but still large with respect to the wall thickness H), the curvature of the tube ceases to be effective and the waves tend to propagate in a longitudinal nondispersive "plate mode" with negligible hoop strains and $c = \sqrt{E_2/\rho(1-e\nu^2)}$. Of course this discussion follows very closely arguments well-known from isotropic theory (see for example [9]).

2) Wavelengths comparable with the geometric average of radius and thickness. $\varepsilon_R = O[(\varepsilon\lambda)^2]$

Anisotropic behaviour shows in this case a feature which is essentially different from the isotropic case: Transverse shear deformation coupled with flexural stiffness in the axial direction becomes so important that simplifying a priori assumptions on the distribution of strains and stress across the thickness (for example linear or cubic) can no longer be justified even in a first step of approximation. The proper distribution must be determined from partial differential equations containing derivatives both with respect to x_2 and to the thickness coordinate x_1 by considering the boundary conditions on the stress-free surfaces of the tube wall. An exact analytic solution of the equations and of the boundary problem corresponding to the first asymptotic step of approximation can eventually be found, following ideas similar to the ones developed in [8] for beams Details of the analysis will be given in [16]. The phase velocity as a function of frequency or wavelength follows from the implicit dispersion relations

$$\Omega^2 = 1 + \frac{\varepsilon_M}{e}\left[1 - \frac{\mathrm{Th}(\alpha H/2)}{\alpha H/2}\right]\left(\frac{2\pi R}{\lambda}\right)^2 , \quad \alpha := \frac{2\pi}{\lambda\sqrt{\varepsilon_M(1-e\nu^2)}} ,$$

$$c = \lambda f = \frac{\lambda}{2\pi R}\Omega\sqrt{\frac{E_3}{\rho}} . \qquad (5.2)$$

3) Transition regime near frequencies $\Omega = 1$

Both solutions given above for moderately large and moderately short wavelengths and the corresponding dispersion relations (5.1) and (5.2) fail for values of Ω approaching 1 from the left side (membrane solution fails) or from the right side (flexural solution fails). This is a typical problem of singular perturbation schemes. It can be solved for sinusoidal waves by careful analysis leading to a simple differential equation of 6. order. The corresponding dispersion relation in terms of frequency and wavelength can be explicitly expressed as

$$\Omega^2 = 1 - e^2\nu^2\left(\frac{\lambda}{2\pi R}\right)^2 + \frac{1}{12\,e\,(1-e\nu^2)}\frac{(2\pi R H)^2}{\lambda^4} . \qquad (5.3)$$

Fig. 10 shows for the case of winding angles of $\pm 22.5^0$ experimentally determined phase velocities and the corresponding theoretical values based on (5.1), (5.2) and (5.3). To obtain the values of the moduli E_2, E_3 and ν needed in (5.1), the theoretical phase velocities were fitted (least squares) to 135 pairs of experimental values for various frequencies in the range of long and moderately long wavelengths corresponding to the mem-

brane behaviour both for $\Omega < 1$ and $\Omega > 1$. A relative variance as low as 0.8 % could be reached. With such a high degree of accuracy one can safely assume that the fitting not only delivers the proper values of the moduli (in the case at hand $E_2 = 75.2\ 10^9$ N/m^2, $E_3 = 11.55\ 10^9$ N/m^2 and $\nu \equiv \nu_{23} = 1.275$) but also a meaningful confirmation of the quality of the experiments and of the adequacy of the asymptotic theory leading to the simple relation (5.1). Once the moduli E_2, E_3 and ν were determined, the theoretical predictions based on (5.2) and (5.3) were fitted to the experimental phase velocities in the range of moderately short wavelengths (involving flexure of the tube wall and transverse shear) by proper choice of the transverse shear modulus G_{12}. Using a set of 153 pairs of experimental values for the tube with $\pm 22.5^0$ winding angle the value $G_{12} = 4.3\ 10^9$ N/m^2 could be determined with 0.9 % relative variance. If one excludes the first 3 of the 153 experimental pairs of values where measurements are quite difficult since they occur in the transitional range corresponding to the vicinity of the critical value $\Omega = 1$, the variance drops even to less than 0.2 %. Again this remarkable fit with just one parameter may be considered as a brilliant confirmation of both the experimental work of Dual and of the efficiency of the asymptotic approach used in the theoretical modelling leading to (5.2).

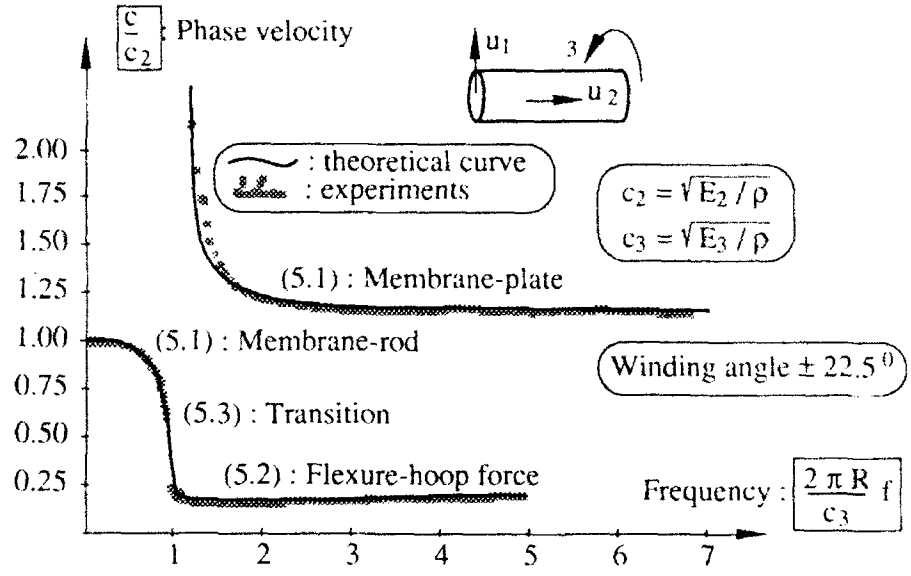


Fig. 10: Obtaining E_2, E_3, ν_{23} and G_{12} in a filament-wound tube from data on phase velocities

6. IN-PLANE WAVES IN A CROSS-PLY LAMINATE

Cross-ply laminates which were studied in flexural wave propagation modes according to the program described in Section 4 were also subjected to transient loading by torsional piezo-electric transducers generating in-plane waves (Fig. 11). A short account of the experimental and theoretical results has been given in [4]. A more detailed treatment will be

presented in [17]. In the following the main ideas leading to the theoretical prediction of the phase curves will be briefly reviewed. These theoretical phase curves have been compared with experimental data produced by A.Eisenhut at our Laboratory on a particular laminate produced out of 9 carbon-fiber-reinforced laminae arranged symmetrically as (0, 90, 0, 0, 90, 0, 0, 90, 0)-plies with respect to the x_1-axis.

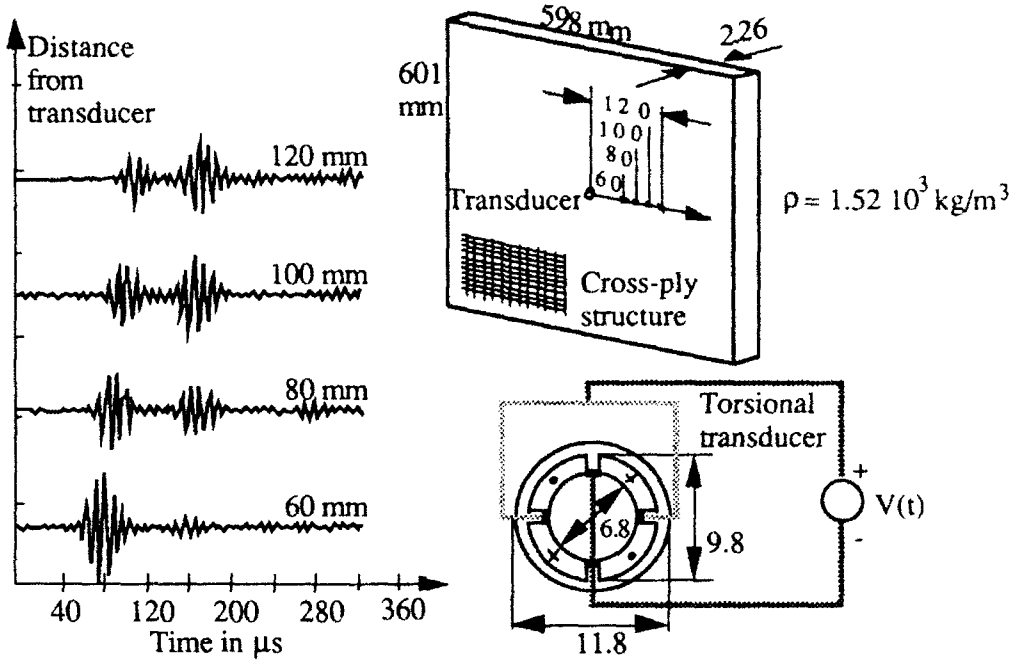


Fig. 11: In-plane waves in a cross-ply laminate generated by a torsional transducer

With x_α as coordinates along the principal directions of orthotropy $\alpha = 1, 2$ the basic equations of linear elastodynamics lead in the case of in-plane waves in an orthotropic plate to the differential equations

$$u_{1,11} + \delta^2 u_{2,12} + \varepsilon^2 u_{1,22} = (c_1)^{-2} u_{1,tt} \quad ,$$
$$\varepsilon^2 u_{2,11} + \delta^2 u_{1,21} + e^2 u_{2,22} = (c_1)^{-2} u_{2,tt} \qquad (6.1)$$

where

$$e^2 := \frac{E_2}{E_1} \quad , \quad \varepsilon^2 := \frac{G_{12}\,[1 - (e\,\nu_{12})^2]}{E_1} \quad , \quad \delta^2 := \varepsilon^2 + e^2\,\nu_{12} \quad , \quad (c_1)^2 := \frac{E_1}{\rho\,[1 - (e\,\nu_{12})^2]} \quad .$$

In a cross-ply laminate E_1, E_2 and ν_{12} are global in-plane moduli of elasticity corresponding to the moduli of the equivalent homogeneous orthotropic plate. Whereas e^2 can be expected to take a value of about 0.5 in the laminate mentioned above with twice as many

0^0-plies than 90^0-plies, ε^2 should be an order of magnitude smaller (for example 0.05) since the cross-ply laminate is quite soft in shear. Also the global coefficient of lateral contraction ν_{12} should be small (about 0.04), so that the value of δ^2 is not expected to exceed 0.07. To derive the shape of the phase curve at distances from the transducer where for all practical purposes a sinusoidal mode of propagation can be assumed (distances larger than 1 wavelength) set

$$u_\alpha = A_\alpha \exp [i \omega t - i f (x, y)] \quad , \quad (\alpha = 1, 2) \quad , \quad (x, y) := \frac{\omega}{c_1} (x_1, x_2) \tag{6.2}$$

in (6.1) to obtain a set of two linear homogeneous equations for A_1 and A_2 which only admit non-trivial solutions if the determinant of the coefficients of A_1 and A_2 vanishes. This leads to a differential equation for f (x, y) of the form

$$d_1 [f] + i\, d_2 [f] + d_3 [f] = 0 \tag{6.3}$$

where d_2 [.] is a nonlinear differential operator involving products of first and second derivatives of f (x, y) and d_3 [.] is a nonlinear differential operator containing only second derivatives of f (x, y). The explicit expression for the first differential operator is

$$d_1 [f] := [(f_{,x})^2 + \varepsilon^2 (f_{,y})^2 - 1] [\varepsilon^2 (f_{,x})^2 + e^2 (f_{,y})^2 - 1] - \delta^4 (f_{,x}\, f_{,y})^2 \quad . \tag{6.4}$$

Since at large distances (larger than at least one wavelength) from the transducer centered at the origin (0, 0) the terms d_2 [f] and d_3 [f] with the derivatives of second order can be neglected (see [17] for details, the arguments are similar to those used in connection with (3.2)), the expression for the phase curve can be obtained by setting d_1 [f] ≈ 0 according to (6.3). Using further that δ^4 in the expression (6.4) is a small quantity (less than 0.005) one obtains the following two possible expressions for f (x,y):

$$f^{(1)} = [x^2 (1 + \frac{\delta^4}{e^2}) + \frac{y^2}{\varepsilon^2}]^{1/2} + O(\delta^8) \quad , \quad f^{(2)} = [\frac{x^2}{\varepsilon^2} + \frac{y^2}{e^2} (1 + \frac{\delta^4}{e^2})]^{1/2} + O(\delta^8) \quad . \tag{6.5}$$

Whereas in isotropic structures two solutions corresponding to rotational (with displacements in the hoop direction) and dilatational (with displacements in the radial direction) motion and circular phase curves are expected, in the cross-ply laminate at hand the two possible phase curve shapes are obviously elliptic with large excentricities. One can further show that in a first step of approximation $f^{(1)}$ is associated only with the displacement component u_1 along the first principal direction of orthotropy and $f^{(2)}$ with the displacement component u_2 along the second principal direction of orthotropy. The low shear stiffness ($G_{12} << E_1, E_2$) is responsible for this decoupling phenomenon with respect to the principal directions. Thus along the x_1-axis $f^{(1)}$ corresponds to pure dilatational motion and along the x_2-axis to pure shear motion. Conversely $f^{(2)}$ corresponds to pure shear motion along the x_1-axis and to pure dilatational motion along the x_2-axis. Furthermore, the cylindrical waves with elliptic phase curves propagate obviously without dispersion, since the radial phase velocities do not depend on the frequency, only on the angle with respect to the first principal direction x_1.

In experimental work leading to his diploma thesis A. Eisenhut studied at our Laboratory in-plane waves produced by a torsional transducer and measured the arrival times of both waves carrying displacements u_1 and u_2 respectively at various radial loca-

tions as a function of the angle φ with respect to x_1. Thus he obtained the radial velocities $c_{r1}(\varphi)$ and $c_{r2}(\varphi)$ for both waves. For each value of φ two pairs of measured radial velocities were available. I fitted (least squares) the experimental results to the theoretical predictions by properly choosing the values of the moduli of elasticity E_1, E_2 and G_{12} (the influence of ν_{12} with expected values less than 0.05 is negligible) and obtained the results depicted in Fig. 12 which exhibit some problems in the dilatational modes with large velocities near the principal axes of orthotropy where the difference of the arrival times between two locations were too short (7-10 μs) to allow accurate measurements. Nevertheless, in the shear modes with low velocities near the principal axes and at larger angles with respect to those axes the accuracy of the fit is quite satisfactory (less than 3 % relative variance) and confirms both the quality of the experimental study and the validity of the corresponding theoretical development.

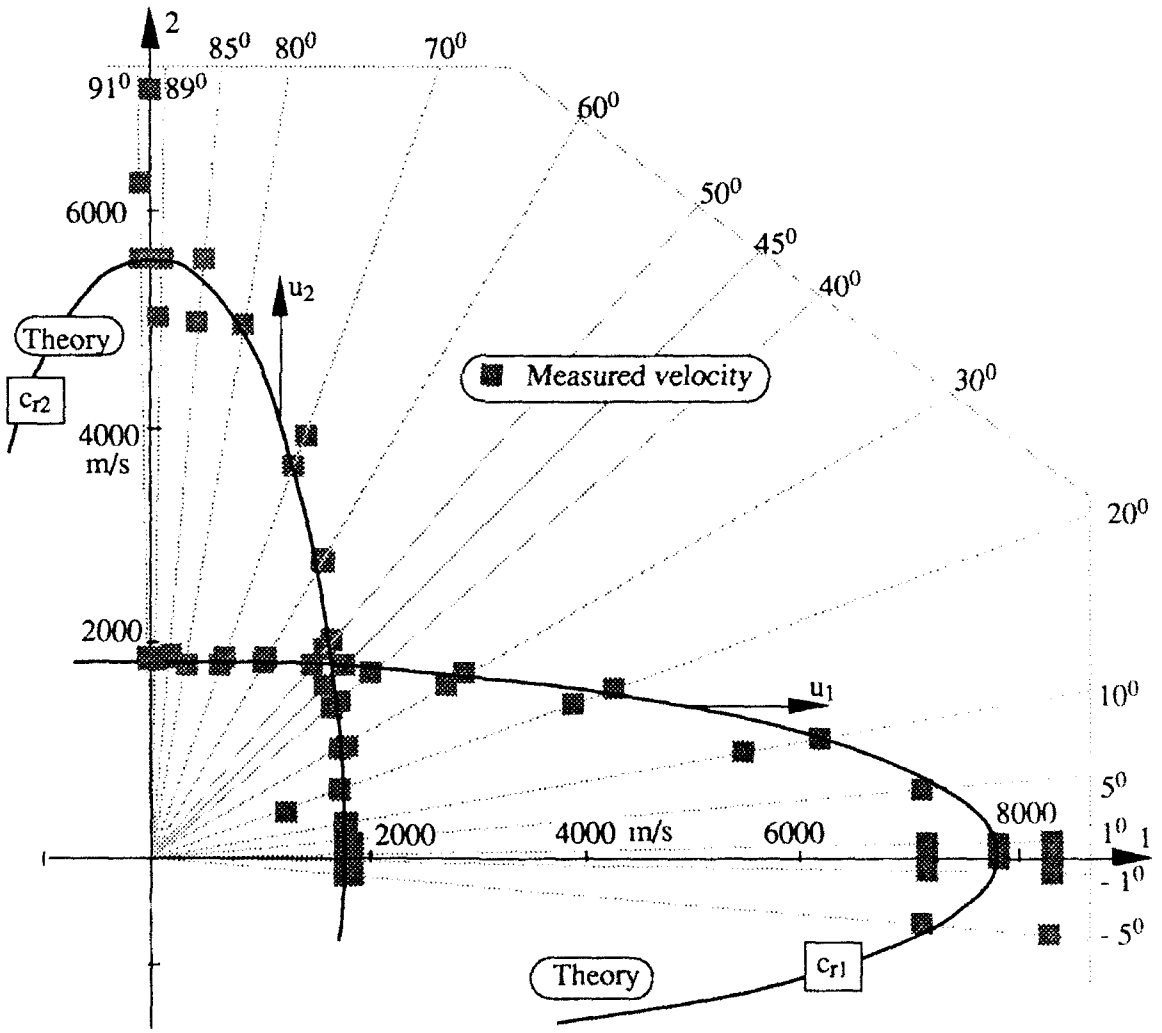


Fig. 12: Phase curves of cylindrical in-plane waves in a cross-ply laminate

It is interesting to compare results for the moduli obtained from in-plane waves as discussed in the present section with those generated by M. Veidt on the same laminate in the context of his doctoral work by using flexural modes as described in Section 4 (see [7] for more details).

Moduli: 10^9 N/m^2	E_1	E_2	G_{12}	G_{13}	G_{23}
From in-plane waves	93	47	4.8	-	-
From flexural waves	92	42	-	4.2	3.0

The appreciable difference in the values of E_2 for the two modes is most probably due to the fact that in the flexural mode the central 90^0-ply with its fibers parallel to the x_2-direction contributes almost nothing to the global value of E_2 whereas in the membrane (in-plane) mode it is as effective as the other two 90^0-plies of the laminate with its (0, 90, 0, 0, 90, 0, 0, 90, 0) lay-up. Thus E_2 (membrane) > E_2 (flexural). Besides, since there are twice as many 0^0-plies as 90^0-plies, the value of E_1 in the membrane mode is roughly twice as large as E_2.

7. APPLICATIONS TO FAST FRACTURE IN CROSS-PLY LAMINATES

In the context of experimental work leading to his doctoral thesis [18] E. Moor showed in our Laboratory that in thin carbon-fiber reinforced laminates submitted to tensile loading and conveniently notched as in classical tests of fracture mechanics, a crack perpendicular to the direction of the load propagates from the tip of the notch towards the other edge of the plate with an almost constant velocity reaching values in the range of 500 -1200 m/s for the various lay-ups tested. Moor used shadow-optical methods [19, 20] to provide experimental data on the stress field first near the tip of the crack both before and after it started running. Preliminary calculations by M. Veidt [21] showed that the results could be interpreted by assuming that the laminate behaved as a homogeneous orthotropic structure, provided that the global shear modulus near the crack tip was adjusted to account for the fact that damage near the crack tip substantially decreases locally the global shear stiffness characteristics of the structure.

To calculate the stress and strain field near the tip of a crack running with constant velocity v along one of the principal directions of orthotropy x_1 of a thin orthotropic plate we may use relations (6.1) and look for a stationary solution by assuming that the stress and strain fields remain stationary with respect to a frame moving with the crack tip (this is of course acceptable only as long as reflections from the edges of the structure can be neglected). Thus we set

$$x := x_1 - v\,t$$

and look for a solution of (6.1) of the form $u_\alpha(x, x_2)$. The transformed relations (6.1) are

$$\begin{aligned} &(1 - \gamma^2)\, u_{1,xx} + \delta^2\, u_{2,x2} + \varepsilon^2\, u_{1,22} = 0 \quad , \\ &(\varepsilon^2 - \gamma^2)\, u_{2,xx} + \delta^2\, u_{1,2x} + e^2\, u_{2,22} = 0 \quad , \end{aligned} \tag{7.1}$$

where $\gamma := v / c_1$. For the carbon-fiber cross-ply laminates tested in our laboratory the value of γ was in the range 0.10 - 0.15, so that $\gamma < \varepsilon$.

We now look for a particular solution of (7.1) which is derivable from a potential according to

$$u_1 = \Phi_{,x} \quad , \quad u_2 = q\, \Phi_{,2} \quad , \tag{7.2}$$

where q is a constant factor which can be adjusted to fulfill both equations (7.1) with the single function $\Phi\ (x, x_2)$. From (7.1) and (7.2) after integrating the first equation with respect to x and the second with respect to x_2 one obtains

$$\begin{aligned} &(1 - \gamma^2)\, \Phi_{,xx} + (q\, \delta^2 + \varepsilon^2)\, \Phi_{,22} = 0 \quad , \\ &[q\, (\varepsilon^2 - \gamma^2) + \delta^2]\, \Phi_{,xx} + q\, e^2\, \Phi_{,22} = 0 \quad , \end{aligned} \tag{7.3}$$

$\Phi\ (x, x_2)$ is a solution of both equations if these are identical, i.e. if the coefficients of $\Phi_{,xx}$ and $\Phi_{,22}$ in the first and second equation are proportional. This leads to the following two values for the parameter q:

$$q_1,\, q_2 = \frac{\beta \pm \sqrt{\beta^2 - 4\, (\varepsilon^2 - \gamma^2)\, \delta^4\, \varepsilon^2}}{2\, (\varepsilon^2 - \gamma^2)\, \delta^2} \quad , \quad \beta := (1 - \gamma^2)\, e^2 - \delta^4 - (\varepsilon^2 - \gamma^2)\, \varepsilon^2 \quad . \tag{7.4}$$

Thus the general solution for (7.1) can be written as a sum of two particular solutions

$$u_1 = \Phi_{1,x} + \Phi_{2,x} \quad , \quad u_2 = q_1\, \Phi_{1,2} + q_2\, \Phi_{2,2} \quad , \tag{7.5}$$

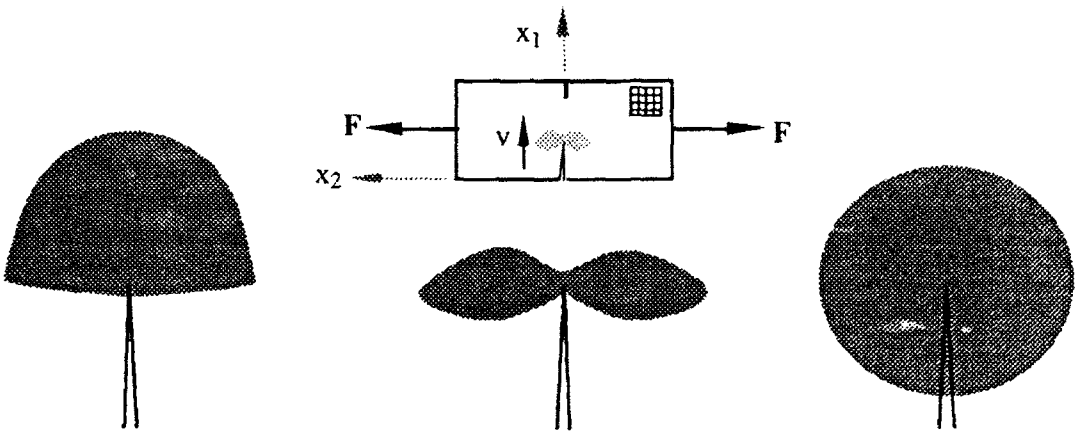
where according to (7.3) the functions $\Phi_1\ (x, x_2)$ and $\Phi_2\ (x, x_2)$ satisfy the uncoupled differential equations

$$\begin{aligned} &(1 - \gamma^2)\, \Phi_{1,xx} + (q_1\, \delta^2 + \varepsilon^2)\, \Phi_{1,22} = 0 \quad , \\ &(1 - \gamma^2)\, \Phi_{2,xx} + (q_2\, \delta^2 + \varepsilon^2)\, \Phi_{2,22} = 0 \quad , \end{aligned} \tag{7.6}$$

which can easily be expressed as Laplace differential equations by transforming x_2 in each equation and solved by classical methods of potential theory .

The shadow-optical data gathered by E. Moor near the tip of cracks in cross-ply laminates fall mainly into the two patterns (a) and (b) illustrated in Fig. 13. The pattern (a occurs at the tip of the notch before the crack starts running, actually at loads well belo (less than 20 %) the fracture load. Pattern (b) has been observed at the tip of a runnin crack. These patterns can be interpreted and simulated on the basis of the theoretical background presented above which allows the prediction of the displacement, strain and stress fields near the tip of a crack running at constant velocity, provided that reflecions from the edges of the structure can be neglected. M. Veidt has shown that the main differences in the shapes (a) for a cross-ply laminate and (c) for an isotropic structure is due to the much lower ratio ε_M (shear modulus / tensile modulus) in the former case. The shape (b) can be theoretically simulated by assuming that the shear modulus is much lower (by

a factor of about 3) than the one measured in the undamaged cross-ply laminate. These findings could also be substantiated by E. Moor with the help of direct strain measurements with strain gauges conveniently placed in the vicinity of the running crack. The details of these experimental studies and of the theoretical developments required for their interpretation and evaluation will be published elsewhere.



(a) Low load, stationary crack, cross-ply structure

(b) Fracture load, running crack, cross-ply structure

(c) Isotropic structure

Fig. 13: Measured patterns of shadows at the crack tips of cross-ply and isotropic structures

8. DEDICATION

Both this paper and the corresponding lecture presented at ICTAM 18 in Haifa are dedicated to the memory of *Professor **Harry Kolsky*** who motivated most of us so strongly in our research activities towards reaching a better understanding of the *physical aspects* of wave propagation and dynamic fracture.

9. REFERENCES

[1] M. Staudenmann and M.B. Sayir, *Apparent Rotation of a Flexural Wave in an Isotropic Beam with Circular Cross Section* (to be published)

[2] M.B.Sayir and M.G. Koller, *Dynamic Behaviour of Sandwich Plates*, ZAMP, 37 (1986) 78-103

[3] A. Kreis and M.B. Sayir, *Propagation of Flexural Waves in a Thin Trnsversely Isotropic Plate*, ZAMP, 34 (1983) 816-831

[4] M.B. Sayir, *Wave Propagation in Composite Structures: Experimental and Theoretical Studies*, Acta Mechanica, 3 (1992) 7-21
[5] M.B. Sayir, *Theoretical and Experimental Results on the Dynamic Behaviour of Composite Beams, Plates and Shells*, "Refined Dynamical Theories of Beams, Plates and Shells and their Applications", 72-88, Springer 1987
[6] J. Dual, *Experimental Methods in Wave Propagation in Solids and Dynamic Viscometry*, Diss. ETH 8659, Zurich 1989
[7] M. Veidt and M.B. Sayir, *Experimental Evaluation of Global Composite Laminate Stiffnesses by Structural Wave Propagation*, J. Composite. Materials 24 (1990) 688-706
[8] M.B. Sayir, *Flexural Vibrations of Strongly Anisotropic Beams, Ingenieur Archiv*, 49 (1980) 309-323
[9] H. Kolsky andJ.M. Mosquera, *Dynamic Loading of Fiber-reinforced Beams, Mechanics of Mat. Behaviour*, 201-218, Elsevier 1984
[10] E. Reissner, *On Bending of Elastic Plates*, Q. Appl. Math., 5 (1947) 55-68
[11] N.J. Hoff, *Bending and Buckling of Rectangular Sandwich Plates*, NACA TN 2225, Nov. 1950
[12] M.G. Koller, *Elastic Impact of Spheres on Sandwich Plates*, ZAMP 37 (1986) 256-269
[13] A. Kreis and M.B. Sayir, *On Wave Propagation and Energy Flux in Thin Transversely Isotropic Plates*, ZAMP 36 (1985) 549-567
[14] J.H. Goodbread, *Mechanical Properties of Spongy Bone at Low Ultrasonic Frequencies*, Diss. ETH 5856, Zurich 1976
[15] J. H. Heimann and H. Kolsky, *The Propagation of Elastic Waves in Thin Cylindrical Shells, J. Mech. Phys. Solids*, 14 (1966) 121-130
[16] M.B. Sayir, J. Dual and B. Gasser, *Structural Axisymmetric Waves in Anisotropic Cylindrical Shells: Asymptotic Theory and Experiments* (to be published)
[17] M.B. Sayir, M.Veidt and A.Eisenhut, *Centrally Induced In-Plane Structural Waves in Cross-Ply Laminates: Asymptoticd Theory and Experiments* (to be published)
[18] E. Moor, *Anwendung der Methode der Schattenoptik zur Untersuchung der Rissausbreitung in faserverstärkten Kunststoffen*, Diss. ETH , Zurich 1992
[19] P. Mannogg, *Anwendungen der Schattenoptik zur Untersuchung des Zerreissvorgangs von Platten*, Diss. Albert-Ludwigs-Universität zu Freiburg i. Br. 1964
[20] J.F. Kalthoff, *Shadow Optical Method of Caustics*. Handbook on Experimental Mechanics, 430-500 Prentice-Hall, 1987
[21] M. Veidt, *Studien zum bruchmechanischen und strukturdynamischen Verhalten von Faserverbundplatten*, Diss. ETH 9424, Zurich 1991

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Self-similar multiplier distributions and multiplicative models for energy dissipation in high-Reynolds-number turbulence

K.R. Sreenivasan and G. Stolovitzky

Mason Laboratory, Yale University
New Haven, CT 06520-2159, USA

Abstract

We begin with a brief description of the multiplier distribution for ϵ_r, the average over a linear interval r of the energy dissipation rate, ϵ. Using measured multiplier distributions obtained for atmospheric surface layer data on ϵ, we show that quasi-deterministic multiplicative models for bases 2 and 3 (that is, binary and tertiary breakdown processes) can be developed on a rational basis. For r in the inertial range, moments computed up to a fairly high order from these models are found to be in good agreement with experimental values. For bases larger than three, such quasi-deterministic approximations for multiplier distributions are not possible. Some applications of multiplier distributions are presented.

1. INTRODUCTION

A Gaussian process is completely described in a statistical sense by its mean and standard deviation. It is conceivable that a nearly Gaussian process can be described well by its first few moments – at least well enough for many purposes. This is the situation with respect to velocity or temperature traces obtained in high-Reynolds-number fully turbulent flows not too close to the wall. On the other hand, the situation is quite different for quantities such as the energy dissipation rate, ϵ, in high-Reynolds-number turbulence. Figure 1 is (effectively) a one-dimensional section through the field of ϵ in the atmospheric surface layer a few meters over land. In contrast to Gaussian or nearly Gaussian processes, information about the first few low-order moments does not describe the signal in any detail. Peaks which are hundreds of times the mean are not uncommon, and the signal is at other times of very low amplitude; this strongly intermittent character is a generic

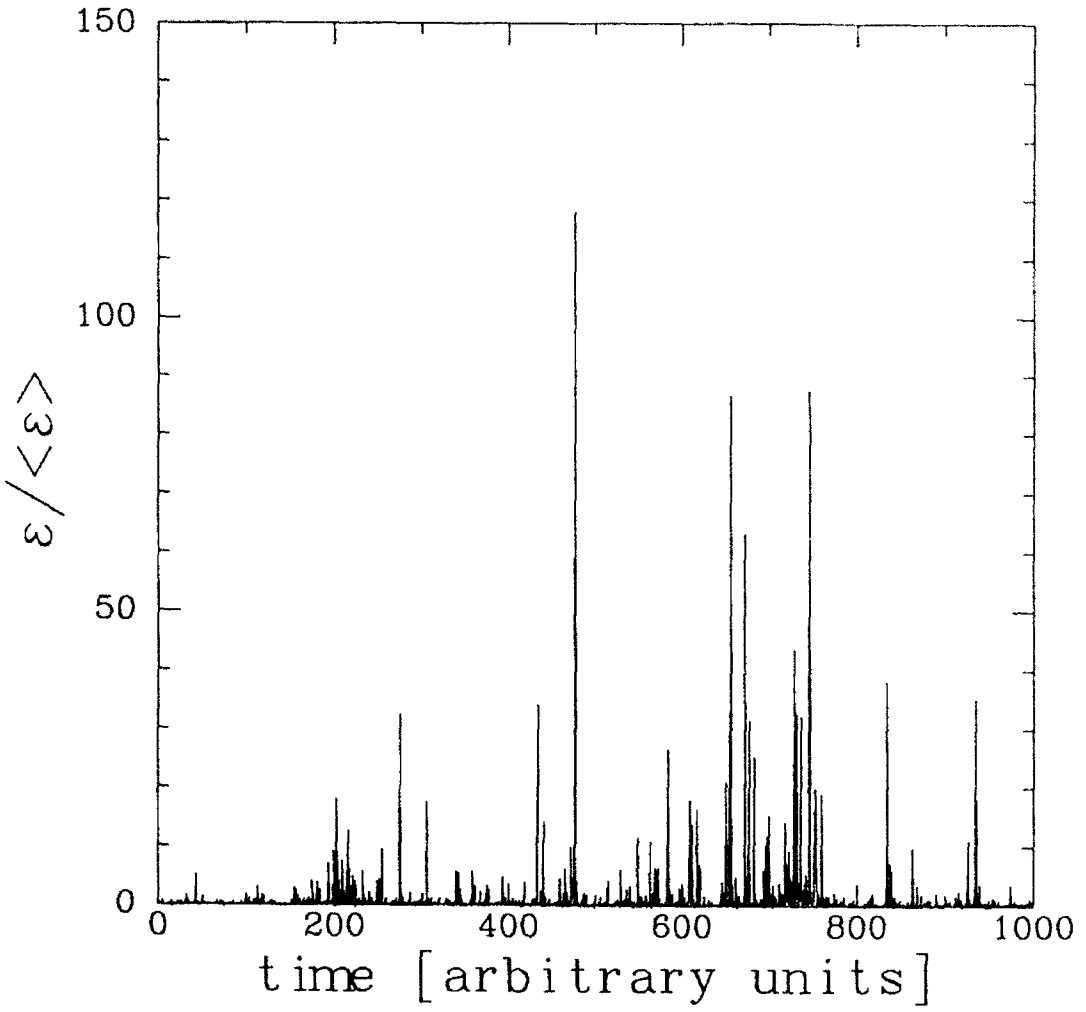


Figure 1: A typical signal of a representative component of ε, namely $\varepsilon' = (du/dx)^2 \sim (du/dt)^2$, normalized by its mean. Here, u is the velocity fluctuation in the direction x of the mean velocity U. In writing the last step of the above approximation, it has been assumed that Taylor's frozen flow hypothesis, namely that the spatial derivative can be approximated by the temporal derivative, holds. The velocity fluctuation u was obtained by a hot-wire mounted on a pole 6 m above the ground level over a wheat canopy. The microscale Reynolds number is of the order 2500.

property of ε in high-Reynolds-number turbulence. Its understanding and modeling is important to any practical scheme for computing turbulent flows. The intermittency has important implications also in contexts such as the structure of turbulent flames.

In the last few years, much work based on multifractals has occurred on the description and modeling of the intermittent character of energy dissipation rate (and other similar characteristics). For a summary, see [1]. In Refs. [2-4], several simple quasi-deterministic multifractal models were shown to describe the statistical properties of the energy dissipation rate quite accurately. Here, we provide an organized basis for developing such simplified intermittency models.

The energy dissipation rate is a positive definite quantity which is additive (in the sense that ε over two non-overlapping intervals equals the sum of ε values distributed over the sum of the two intervals). In this sense, it is convenient to think of ε as a measure distributed on an interval.

2. MULTIPLIER DISTRIBUTIONS

Consider a long data string of ε distributed over an interval which is N integral scales in extent, N being some large integer. Divide the interval into 'a' equal-sized sub-intervals, and obtain the ratios of the measures in each of the sub-intervals to that in the entire interval. These ratios, to be called multipliers, are clearly positive and lie between zero and unity. Subdivide each sub-intervals into 'a' pieces as before, and repeat the procedure. When we reach sub-intervals of the size of the integral scale of turbulence, L, there will be sufficiently large population of the ratios $M_i^{(L)}$, $1 \le i \le N$, and one can obtain a converged histogram of the multipliers $M_i^{(L)}$. Proceed with further subdivisions. At the n-th subsequent level, where each sub-interval is of size $r/L = a^{-n}$, there are $N.a^n$ multipliers $M_i^{(r)}$. Construct the histogram of the multipliers at each level. Repeat the procedure until the smallest sub-interval reached is of the order of the Kolmogorov scale.

The thought behind this hierarchical construction is that the nonlinear processes occurring in the inertial range of scales may be abstracted by a breakdown process in which each eddy subdivides into 'a' pieces, with the energy flux redistributed in some unequal fashion without loss among the sub-eddies; since the energy flux, as it cascades down to smaller scales, is ultimately converted into energy dissipation, the two quantities are equal on the average. It is further thought that this unequal distribution among sub-eddies is the heart of the observed intermittency. The reality is, of course, more complex. For instance, it is

not obvious what the appropriate value of 'a' (that is, the base for the cascade process) must be, or whether it remains the same from one step of the cascade to another. In spite of this ambiguity, it is clear that if there is a scale-similar breakdown occurring in the cascade process, the histograms of the multipliers should be identical at each step of the subdivision or cascade.

The probability density P(M) of the multipliers M - here and subsequently, we omit the indices on the $M_i^{(r)}$ and denote them simply by M - have been obtained for different stages of subdivision of the interval. Since the value of 'a' is not known *a priori*, Chhabra & Sreenivasan [4] obtained P(M) for various bases. Figure 2 shows the results for 'a' = 2, 3 and 5. The shape of each of the distributions is invariant over a certain range of scales, suggesting that some type of self-similarity occurs in this scale range, whatever the assumed base. This range of scales over which P(M) is self-similar agrees quite well with the inertial range of scales determined by the scaling range in spectra and structure functions. The larger symbols show an average over steps involving comparisons between boxes of size 'm' and those of size 'm*a', where m ranged from 50 to 1000 in units of sampling intervals. (For the very smallest scales, the distributions have a concave shape. This concavity is related to the divergence of moments [5] and will be discussed elsewhere. For very large box-sizes, multiplier distributions approach a delta function centered around 0.5, as would be the case for random measures.)

The scale-invariant multiplier distributions obtained in figure 2 are fundamental to the understanding of the observed multifractal scaling [2]. One can compute [4] from them not only the asymptotic scaling properties such as the multifractal spectrum (or the $f(\alpha)$ curve [6]) of a measure, but also finite-size fluctuations of scaling properties [4]. In addition, even in instances where high-order moments diverge, P(M) remains well-defined. Finally, the $f(\alpha)$ function may extend over $(-\infty,\infty)$ whereas P(M) is a compact function defined on $M \in [0,1]$.

A disadvantage of P(M) is that it is base-dependent. However, if the cascades giving rise to the observed intermittency are randomly multiplicative, then the multiplier distributions corresponding to different bases are related by convolution, and one can scale out this base-dependency [4]. If the multiplicative process is random (i.e., successive multipliers are uncorrelated) several base-independent functions can be constructed from these multiplier distributions. In particular, for any two bases 'a' and 'b', we have

$$\log\langle(M_a)^q\rangle/\log(a) = \log\langle(M_b)^q\rangle/\log(b) = -[\tau(q) + D_0] \qquad (1)$$

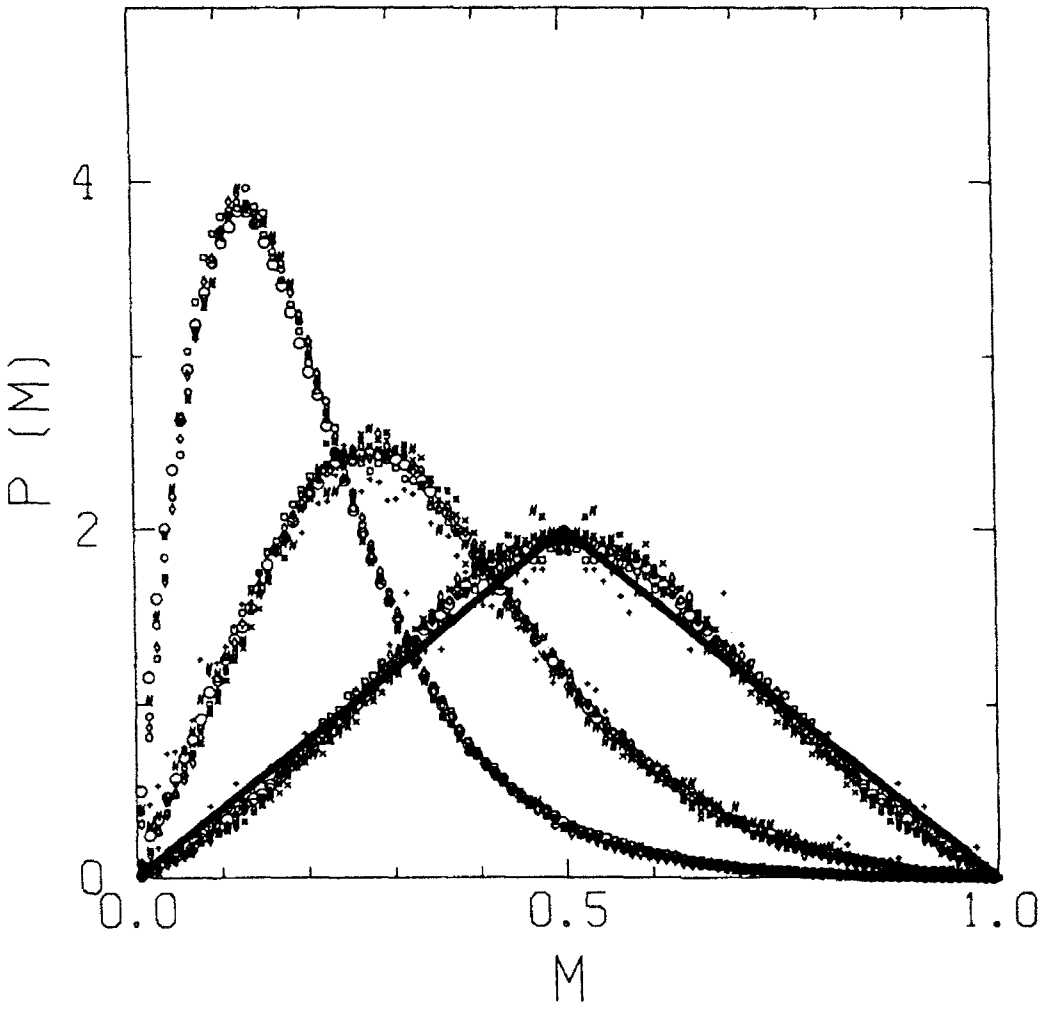


Figure 2: Multiplier distributions P(M) for bases (from right to left) 'a' = 2, 3 and 5. The larger symbols show averaged multiplier distribution, which are the mean of multiplier distributions obtained by comparing measure in boxes of size 'm' to those of 'm*a' , where m ranged from 50 to 1000 in units corresponding roughly to the Kolmogorov scale. The smaller symbols show the distributions obtained for m = 50, 80, 150, 200, 400 and 1000. The solid line is the triangular approximation to the binary case. The figure is adapted from [4].

where $\tau(q) = (q-1)D_q$, D_q being the so-called generalized dimensions [7] of q-th order, and $D_0 = D_{q=0}$ is the fractal dimension of the support of the measure. The $f(\alpha)$ function can also be easily derived from the multiplier distribution. In Ref. [4], it was shown that the $f(\alpha)$ functions computed from these different distributions were in good agreement with each other as well as with those obtained from direct methods such as box-counting. This agreement indicates the existence of a probablistic cascade where no single base is preferred.

Incidentally, a good approximation for P(M) in the binary case is the triangular distribution shown in figure 2. In Ref. [4] it was shown that the $f(\alpha)$ function for this model is in excellent agreement with that obtained directly [3]. Further, the triangular model displays the correct behavior with respect to sample-to-sample fluctuations in $f(\alpha)$. It also reproduces the stretched exponential tails, $P(\varepsilon) \sim \exp(-\beta(\varepsilon)^{1/2})$, observed in Refs. [3,8] for the probability distribution of ε.

3. SIMPLE MODELS

The multiplier distributions shown in figure 2 are extracted directly from the experiment and their analytical forms are yet to be found from the theory. The question meanwhile is a simple representation of these distributions in a way that permits one to evaluate most of the measured properties quite accurately. The goal is to seek models that are simple enough to be tractable mathematically and realistic enough to represent the spirit of the underlying physics. We already mentioned the triangular distribution as a good approximation. An even simpler possibility is the p-model [2], which is a model for a binary cascade ('a'=2). We first discuss the p-model and show how it can be obtained as a rational approximation to the measured multiplier distribution for the binary case. We will then discuss how models in the same spirit can be obtained for the tertiary case ('a'=3). The limitations of the procedure for high order subdivisions ('a'>3) will be highlighted.

From a physical point of view, the cascading process with 'a'=2 can be thought of as the break-up of a structure (the parent structure or eddy) into two sub-structures. For the one-dimensional case corresponding to figure 1, a pertinent question is the following: is there any difference between the left and right offsprings in terms of the energy flux they receive from the parent structure? One can determine experimentally that left and right are statistically indistinguishable. (This is not true for the velocity signal itself, as can be concluded from Kolmogorov's 4/5 law [9]). Now, for the sake of simplicity and modeling, let us assume that one of the two sub-eddies always receives a fixed fraction p of the

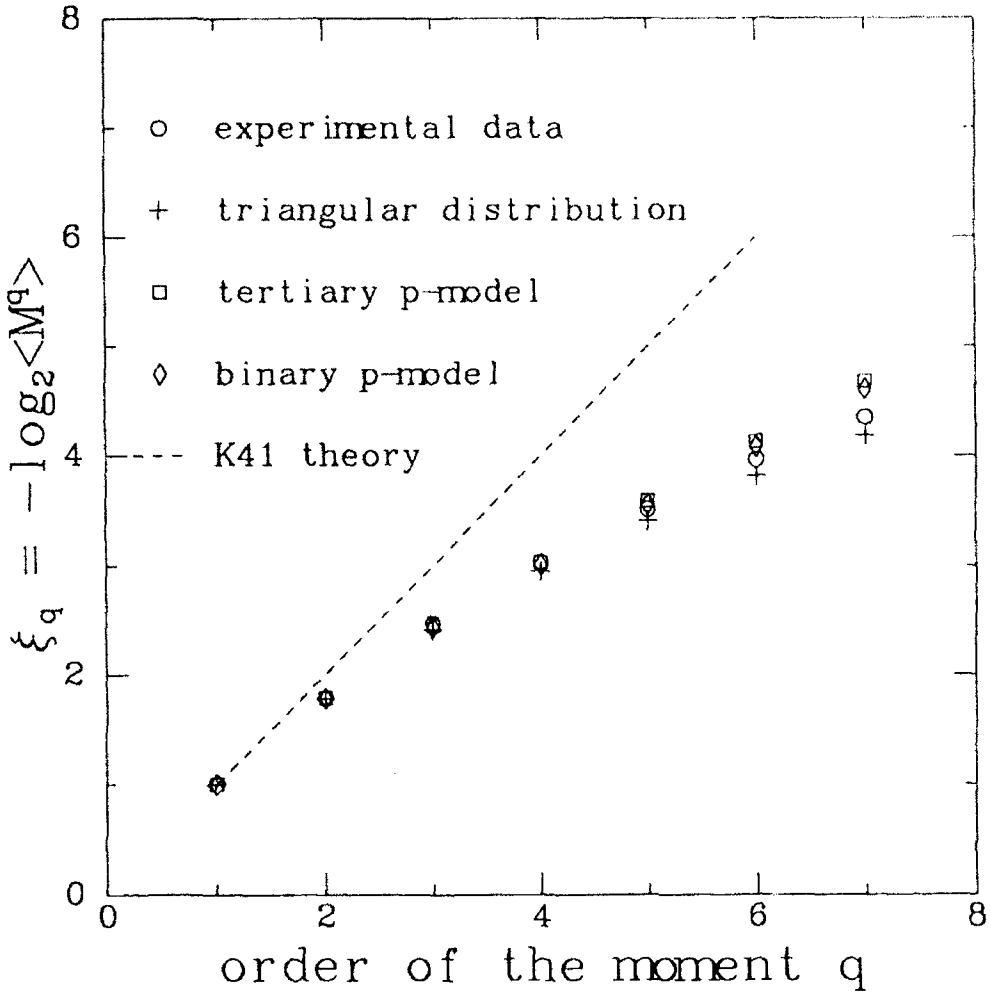


Figure 3: A comparison between moments computed from the measured multiplier distributions and those computed for the different models considered in the text. Experimental data were obtained from a record length of 810,000 data points. The convergence of moments was reasonable; for example, in the last half decade of the record length, the variations observed were smaller (in the log scale) than the symbol size.

energy contained in the parent eddy; naturally, the other will receive 1-p. In this sense, this model has some determinism. However, it is only quasi-deterministic in the sense that either one of the two eddies could receive the fraction p; because of the left-right symmetry mentioned above, a given piece will receive p as often as 1-p. Then, the multiplier distribution for the p-model becomes

$$p_{'a'=2}(M) = \frac{\delta(M-p) + \delta(M-(1-p))}{2}. \qquad (2)$$

If p=1/2, there is no intermittency and the physical situation corresponds to Kolmogorov's 1941 theory [10]. To obtain intermittency, we should have a value of p different from 1/2.

How can we choose p? A natural way is to match the moments of $p_{'a'=2}(M)$ with those of the real p(M). For both distributions, the zero-order moment (normalization) and the first-order moment (mean value) coincide, and are 1 and 1/2, respectively. The first non-trivial condition is to match the second order moment. When this is done, we obtain the value p=0.697, or 1-p=0.303, which can be rounded off to excellent accuracy by 0.7 and 0.3, respectively. This is the p-model of Ref. [2]. It turns out, purely by luck, that high-order moments computed for the p-model also agree with those computed for the real data (see figure 3). It had been shown in [2] that the $f(\alpha)$ spectrum for the binary p-model with p = 0.7 fit the experimental data quite accurately.

We now discuss a general scheme for developing for all 'a'≠2 quasi-deterministic models of the sort developed above for the p-model. Again, we attempt to do this by matching moments. The general multiplier distribution for any 'a' in the p-model scheme is

$$P_{'a'}(M) = \frac{\sum_{i=1}^{a} \delta(M-p_i)}{a} \qquad (3)$$

where

$$\sum_{i=1}^{a} p_i = 1, \quad (0 \le p_i \le 1) \qquad (4)$$

We may now equate the moments of $P_{'a'}(M)$ to the moments of the real multiplier distributions. Since multiplier distributions for any base yield

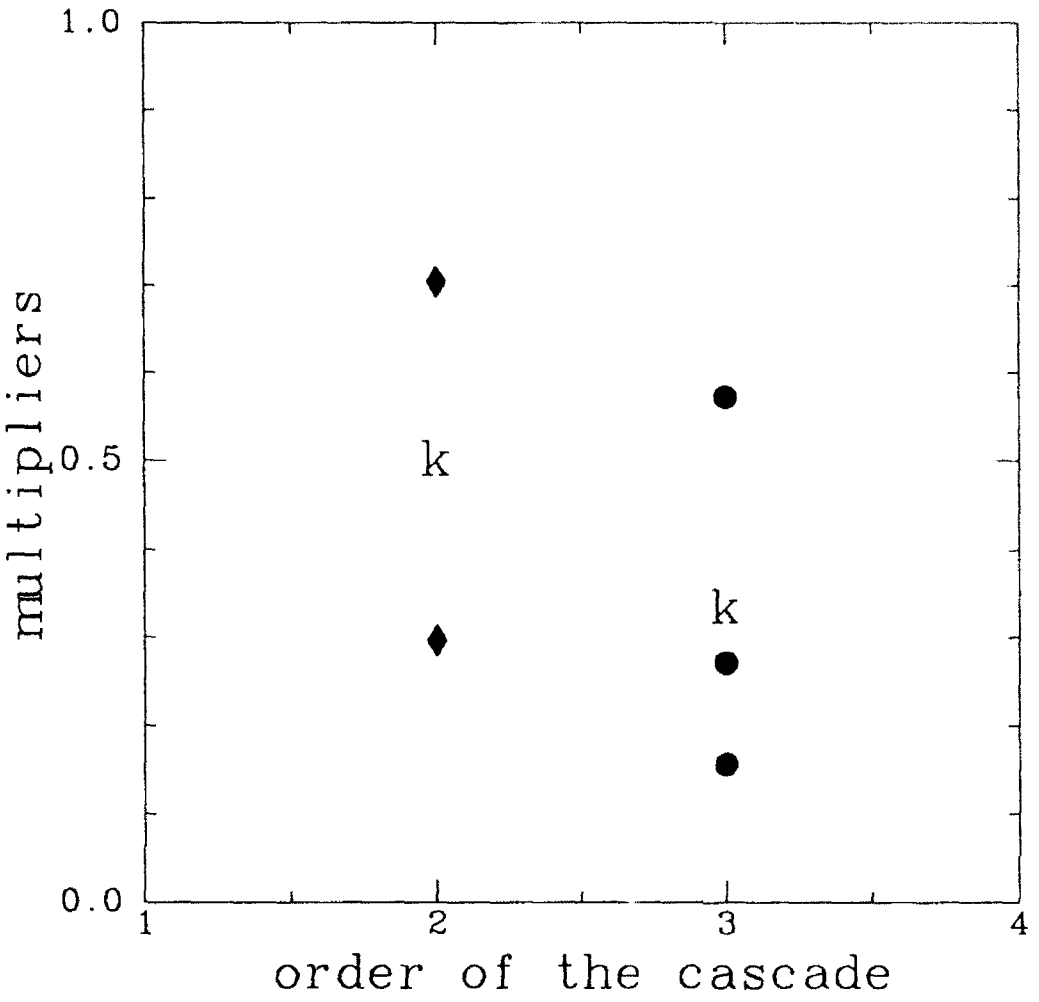


Figure 4. Values of the multipliers for the binary (diamonds) and tertiary (circles) p-models. The letter K indicates the location of the multipliers if there were no intermittency, consistent with Kolmogorov's 1941 theory [10].

the same results, we may take the distribution corresponding to the binary cascade of figure 2. Computing the moments $\langle M^q \rangle$ from it, we are led to the equations

$$\sum_{i=1}^{a} p_i^q = a \langle M^q \rangle^{\log(a)/\log(2)}, \quad q=1, 2, ..., a. \tag{5}$$

This is a system of 'a' equations with 'a' number of p_i's to be determined. Using Girard's rule [11], it is easy to find a polynomial of degree 'a' whose roots are the desired p_i's. The problem thus reduces to the determination of the roots of the polynomial. It turns out *a posteriori* that this problem has physical solutions only for 'a'=2 and 'a'=3: for larger values of 'a', some of the roots turn out to be complex, and have no physical meaning. The values of p_i for the tertiary cascade ('a'=3) are p_1=0.155, p_2=0.283 and p_3=0.562. We designate this as the tertiary p-model. In this scheme, the classical Kolmogorov theory would yield $p_1 = p_2 = p_3 = 1/3$.

The values of p_i for the binary and tertiary cascade are shown in figure 4; also marked by K are the classical non-intermittent values applicable to Kolmogorov's 1941 theory. Although the binary and tertiary p-models, respectively, are generated to possess the first three and four moments correctly, it is remarkable that the binary and tertiary p-models and the measured distributions have approximately the same high order moments up to, say, about 7 (see figure 3).

Other models have also been proposed. For example, Novikov [12] proposed a uniform distribution for P(M). At that time, however, the multiplier distribution had not been obtained experimentally. It is now clear, however, that a uniform distribution is not a good model for any of the curves in figure 2. For example, a good approximation to the binary cascade is the triangular distribution shown by a solid line in figure 2.

4. CONCLUSIONS

The multiplier distributions are a basic tool for understanding many of the scale-similar properties of energy dissipation in turbulence. In the absence of an *ab initio* theory that yields these distributions in a deductive way, analytical progress can be made only by modeling them with reasonable schemes. Here, we have summarized the attempts

made in the last few years. Below, we mention some applications of this work.

The first application is in the calculation of E_r, the energy dissipation over an interval of size r; $E_r = r\epsilon_r$. According to the definition of the multipliers,

$$E_r = E_L \prod_{i=1}^{n} m_i, \tag{6}$$

where $n(r)=\log_a(r/L)$, and L is the large-eddy (integral) scale. Within the p-model scheme for a binary cascade, the probability density function (PDF) for the ratio $x = E_r/E_L$ is

$$p(x)=\frac{A}{x}\frac{r}{L}\frac{\Gamma(\log_2(r/L)+1)}{\Gamma(k(x)+1)\,\Gamma(\log_2(r/L) - k(x)+1)} \tag{7}$$

where A is a normalization constant, Γ is the gamma function and

$$k(x) = \frac{\log(x\,(r/L)^{\log_2(1-p)})}{\log(p/(1-p))}. \tag{8}$$

A second application is in computing the exponents ξ_q, defined as

$$\langle E_r^q \rangle \sim (r/L)^{\xi_q}. \tag{9}$$

The result for the triangular distribution is

$$\xi_q = -\log_2\left\{4\left(1-\frac{1}{2q+1}\right)\frac{1}{(q+1)(q+2)}\right\}. \tag{10}$$

Corresponding results for the p-model are given in [2]. The PDFs of the velocity increments can also be computed. On using the second refined similarity hypothesis [13], the velocity increments can be written as

$$\Delta u(r) = V\,E_r^{1/3}. \tag{11}$$

In Ref. [14], we obtained the probability density of V. The PDF of Δu can be computed if some model is assumed for E_r. In Ref. [8], those PDFs were computed using the binary p-model.

The tertiary p-model has been used to generate a signal that shares many features of a real turbulent velocity trace. This issue will be addressed elsewhere.

5. REFERENCES

[1] K.R. Sreenivasan, Annu. Rev. Fluid Mech. 23, 539 (1992)
[2] C. Meneveau & K.R. Sreenivasan, Phys. Rev. Lett. 59, 1424 (1987)
[3] C. Meneveau & K.R. Sreenivasan, J. Fluid Mech. 224, 429 (1991)
[4] A.B. Chhabra & K.R. Sreenivasan, Phys. Rev. Lett. 62, 1327 (1992)
[5] B.B. Mandelbrot, J. Fluid Mech. 62, 331 (1974)
[6] U. Frisch & G. Parisi, in *Turbulence and predictability in geophysical fluid dynamics and climate dynamics*, eds. M. Ghil, R. Benzi, G. Parisi, pp. 84-88, North-Holland; T.C. Halsey, M.H. Jensen, L.P. Kadanoff, I. Procaccia & B.I. Shraiman, Phys. Rev. A 33, 1141 (1986)
[7] H.G.E. Hentchel & I. Procaccia, Physica D, 8, 435 (1983)
[8] P. Kailasnath, K.R. Sreenivasan & G. Stolovitzky, Phys. Rev. Lett. 68, 2766 (1992)
[9] A.N. Kolmogorov, Dokl. Akad. Nauk. SSSR 30, 301 (1941)
[10] A.N. Kolmogorov, Dokl. Akad. Nauk SSSR 32, 16 (1941)
[11] J. Rey Pastor, P. Pi Calleja & C. Trejo, Analisis Matematico (Kapelusz, Buenos Aires, 1957), vol I
[12] E.A. Novikov, Prikl. Mat. Mekh. 35, 266 (1971)
[13] A.N. Kolmogorov, J. Fluid Mech. 13, 82 (1962)
[14] G. Stolovitzky, P. Kailasnath & K.R. Sreenivasan, Phys. Rev. Lett. 69, 1178 (1992)

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Cardiovascular Fluid Mechanics

A.A. van Steenhoven, J.D. Janssen and D.H. van Campen

Department of Mechanical Engineering, Eindhoven University of Technology
P.O. Box 513, 5600 MB Eindhoven,The Netherlands

Abstract

Fluid flow in the larger arteries is essentially unsteady; under normal physiological circumstances it may be regarded as laminar. To obtain more insight in the complex blood flow patterns in the carotid artery bifurcation and around the aortic valve, finite element calculations have been carried out, validated by flow visualization studies and laser-Doppler velocity measurements. In the numerical models, the arterial wall motion and leaflet-fluid interaction are taken into account, using a quasi 1-D linear wave propagation model and a fully coupled iterative method, respectively.

Introduction

The analysis of blood flow around heart valves and in bifurcations is a major research topic in biomechanics. Due to the highly complicated geometry of the arterial system, the unsteadiness of the flow, the large displacements of the valve leaflets, the deformability of the vessel wall and the non-Newtonian behaviour of blood, the subject is a challenge for the engineering disciplines. The recent progress in this field [1] is mainly due to the fact that the experimental tools and computational methods are so advanced now that unsteady three-dimensional flow analyses in complex geometries have become common practice. Besides, the software and hardware in detection apparatus (like ultrasound and MRI) are so improved that non-invasive real-time blood flow analyses can be performed clinically in the near future and, hence, detailed fluid mechanical information is needed to interpret the clinical data [2]. The present state of research [1] is that in various bifurcations a similar steady flow behaviour is found, comprising flow separation, secondary flow and skewing of axial velocity profiles. Differences observed are mainly due to different geometries and Reynolds numbers. However, the influences of flow unsteadiness, wall distensibility, non-Newtonian effects and stenoses are not completely clear yet.

As examples of the interesting flow phenomena in the cardiovascular system and the tools available, we will focus in this paper on the physical-mathematical methods used to analyse the entrance flow in a rigid-walled curved tube, the fluid flow in a distensible model of the carotid artery bifurcation and the interaction between fluid flow and leaflet motion during the opening and closing of the aortic valve. The geometry of the carotid artery bifurcation is shown in figure 1. It consists of a main branch,the common carotid artery,which asymmetrically divides into two branches,the internal and external arteries. A characteristic feature of the carotid artery bifurcation is the widening in the proximal part of the internal carotid artery, the carotid sinus. As a second example the aortic valve is shown in the same figure. It is one of the four valves controling the blood flow through the heart and is situated between the aorta and the left ventricle. It consists of three flexible leaflets and behind each leaflet a cavity in the aorta wall is present; the so-called sinus of Valsalva.

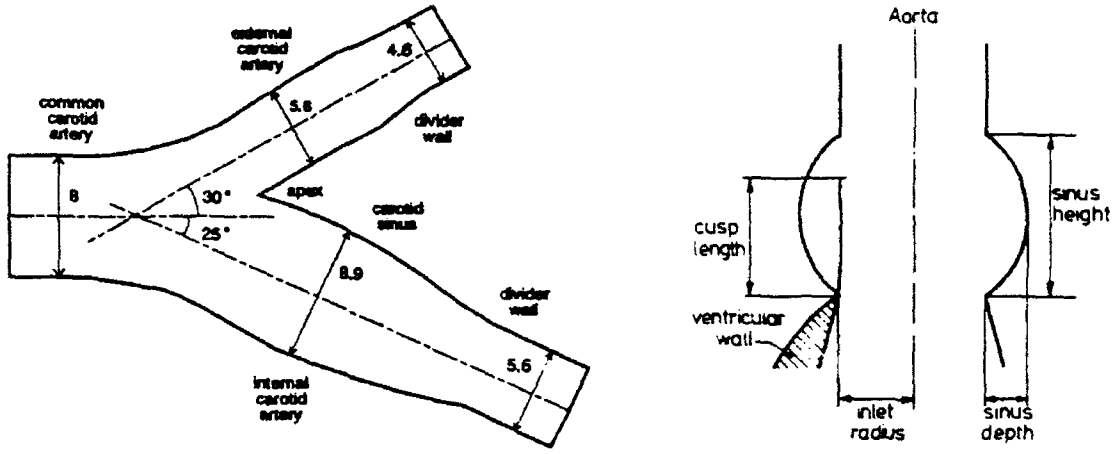


Figure 1. Sketches of the carotid artery bifurcation and the aortic valve.

Under normal conditions the human heart contracts and relaxes about once a second. The arterial volume flow curve consists of a period of relaxation (diastole) followed by a period of contraction (systole), see figure 6. Due to the contraction of the heart, blood flow is essentially unsteady. The Strouhal and Reynolds numbers are defined as:

$$Sr \equiv \frac{D\omega}{\bar{u}} \quad ; \quad Re \equiv \frac{\rho D\bar{u}}{\mu} \qquad \text{eq.1}$$

Here $\bar{u}$ is the mean velocity at the inlet of the artery, D is the diameter of the artery and ω is the angular frequency of flow variation. For the density ρ and viscosity μ of blood the constant values of 1050 kg/m^3 and 3-4 mNs/m^2 are commonly used, although in regions with flow

separation non-Newtonian effects may play a role [3]. In the present study the blood flow is assumed to be laminar, although in some arteries the peak Reynolds number is quite high. However, the unsteadiness of the flow strongly affects the critical Reynolds number [4]. Finally, the arterial walls are distensible. This results in the occurrence of wave phenomena; i.e. pressure and flow disturbances travel with finite velocity, experience damping, and will reflect at discontinuities [5]. The physiological values of the various parameters for the aorta and the carotid artery are summarized in Table 1, adapted after Caro et al. [6].

	Ascending aorta	Carotid artery
Diameter D [mm]	15	5.0
Peak blood velocity u_{max} [m/s]	1.2	0.5
Mean blood velocity $\bar{u}$ [m/s]	0.2	0.2
Measured wave speed c [m/s]	5.0	8.0
Reynolds number	750	275
Reynolds number (peak)	4500	700
Strouhal number	0.95	0.28
Speed-ratio u_{max}/c	0.24	0.06

Table 1. Normal values for canine cardiovascular parameters at a heart rate of 2 Hz [6].

Many researchers have contributed to a better insight into the flow phenomena around the aortic valve and in the carotid artery bifurcation. Significant contributions to the understanding of the fluid mechanical aspects in aortic valve functioning were delivered by Henderson and Johnson [7], Bellhouse and Talbot [8], Peskin [9] and Wipperman [10]. Extensive calculations on blood flow in straight arteries were performed by Womersley [11]. Olson [12] concluded from a steady flow study in a symmetrical three-dimensional bifurcation that the flow phenomena occurring in the daughter branches are highly determined by curvature effects. Yao and Berger [13] further investigated the steady entry flow in a curved pipe. Detailed experimental and numerical information about the flow field in the carotid bifurcation has been obtained by Bahradvaj et al. [14] and for the unsteady case by Ku et al. [15] and Perktold [16]. In our group, model studies of the flow phenomena around the aortic valve were performed [17,18] and measurement techniques were developed for oscillatory flow [19]. To gain insight

into secondary flow patterns, entrance flow in a 90-degree bend was studied [20-22]. Next, steady and unsteady flow in models of the carotid artery bifurcation were analysed [23-26]. Here, we will briefly describe the methods used and present some characteristic results. Experimental validation of the methods used is reported elsewhere [27-30]. Finally, some future research directions will be pointed out.

Entrance flow in a 90-degree curved tube

The flow of an incompressible isothermal Newtonian fluid is described by the Navier-Stokes and continuity equations. To solve these equations they are discretized using a standard Galerkin finite element method which is implemented in the finite element package SEPRAN [31]. The spatial discretization results in the following set of non-linear first-order differential equations [32]:

$$M\dot{\underline{u}} + [S + N(\underline{u})]\,\underline{u} + L^T \underline{p} = \underline{f} + \underline{b} \qquad \text{eq.2a}$$

$$L\underline{u} = 0 \qquad \text{eq.2b}$$

where $M\dot{u}$ represents the local acceleration term, $S\underline{u}$ the viscous term, $N(\underline{u})\underline{u}$ the convective acceleration term, $L^T\underline{p}$ the pressure gradient term and $L\underline{u}$ the velocity divergence term. The body and boundary forces are represented by $\underline{f}$ and $\underline{b}$ respectively, while $\underline{u}$ contains the velocity and $\underline{p}$ the pressure unknowns in the nodal points. The local time derivative in eq.2a is approximated by :

$$\dot{\underline{u}}^{n+\Theta} = \frac{\underline{u}^{n+1} - \underline{u}^{n}}{\Delta t} \qquad \text{eq.3}$$

in which $\underline{u}^n$ is an abbreviation for $u(n\Delta t)$ with Δt the time step. The non-linear convective term is linearized using one step of a Newton-Raphson iteration scheme. In order to reduce the number of unknowns, a penalty function method is applied [33]. For the solution of the matrix equation a direct technique was used .

The elements used belong to the group of so-called modified Crouzeix-Raviart elements. For 2D situations, like the aortic valve model, a 7-noded element with 3 pressure and 14 velocity unknowns is used. 3D situations, like the curved tube and the carotid artery bifurcation, require the application of a 27-noded element with 4 pressure and 81 velocity unknowns. (See figure 2). The accuracy is $0(\Delta x^3)$ for the velocity and $0(\Delta x^2)$ for the pressure with Δx being a characteristic dimension of the spatial discretization. The resolution in time, using $\theta = 1/2$, is shown to be $0(\Delta t^2)$.

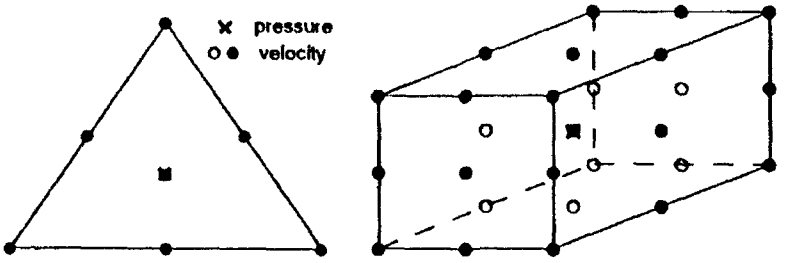


Figure 2. The 7-noded and 27-noded elements used in solving the 2D and 3D fluid motion equations, respectively.

The flow at the inlet is supposed to be fully developed. The velocities at the wall are presumed to be zero. At the outlet the normal and both tangential stresses are set to zero, while in the plane of symmetry both tangential stresses and the normal velocity component are put to zero.

As an example some results of unsteady entrance flow in a 90-degree curved tube will be shown. The flow rate consisted of a sinusoidally varying unsteady flow component superimposed on a steady flow component. With a curvature ratio of $\delta = 1/6$, the Dean number $\left(\equiv \sqrt{\delta}\,\mathrm{Re}\right)$ varied between 82 and 327. The Strouhal number was 0.46. In figure 3 the results of axial and secondary flow at $\theta = 22.5°$ from the straight inlet tube are shown. Axial flow is presented by axial isovelocity lines and secondary flow is visualized by means of velocity vectors. Contour level 0 corresponds to zero axial velocity. As a consequence of centrifugal forces, the secondary velocities near the plane of symmetry are directed towards the outer bend (O), whereas near the side wall of the curved tube these secondary velocities are directed towards the inner bend (I), resulting in a Dean-type secondary flow field. For the total period of time the center of this secondary vortex is situated near the center line of the cross-sectional plane. The secondary velocities at t=1/4T and 1/2T are about equal, as well as the secondary velocities at t=3/4T and t=0. At minimal flow rate a region with low secondary velocities is found near the inner bend. As a consequence of secondary flow a shift of the maximum of axial velocity towards the outer bend is observed for the whole period of time. This shift is maximal at minimal flow rate (t=3/4T). At this time interval a region with negative axial velocities is observed at the inner bend. Similar results were obtained at other axial positions [22]. There it also appeared that axial flow is highly determined by secondary flow. The influence of the frequency parameter and the flow wave form appeared to be of minor importance.

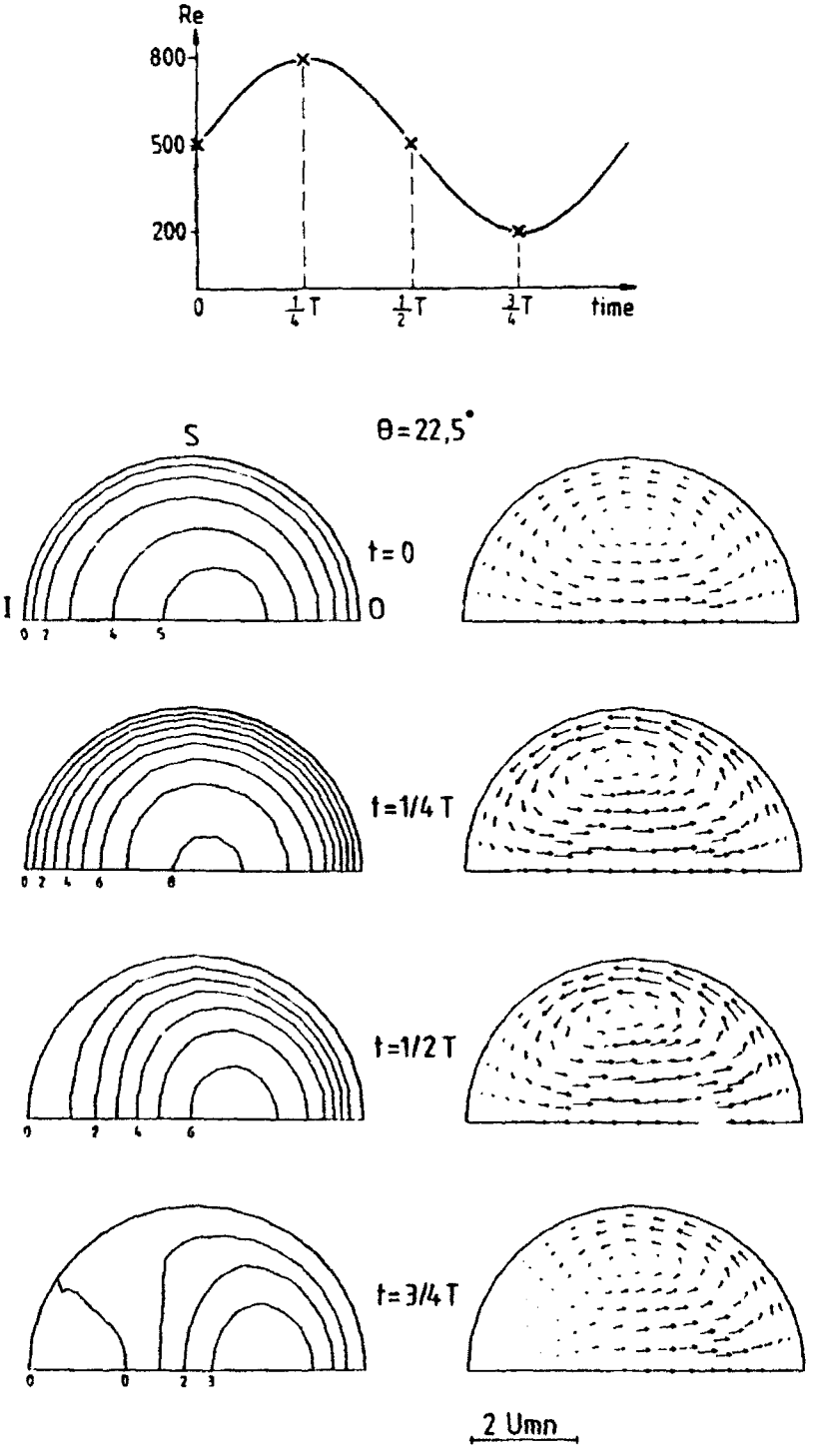


Figure 3. Axial and secondary unsteady entrance flow in a 90-degree curved tube [28].

Fluid flow in a distensible model of the carotid artery bifurcation

In the numerical analysis of the flow phenomena in distensible geometries not only the fluid motion equations but also the equations governing wall motion should be taken into account. Motivated by the small ratio of fluid velocity over wave velocity ($\simeq 0.06$ for the carotid artery) an uncoupled approach is followed. First, the time dependent pressure distribution due to wave propagation is determined. Next, the wall motion due to this time dependent pressure distribution is calculated. Finally, the wall motion is prescribed as a boundary condition for the solution of the fluid motion. We will briefly describe the three steps in this procedure and finally an illustrative result will be given.

The time dependent pressure distribution is calculated using a quasi-1D linear model describing wave propagation. The linearized one-dimensional mass and momentum equations for unsteady viscous flow through a visco-elastic tube are used, which read [34]:

$$\frac{\partial A}{\partial t} + A_o \frac{\partial u}{\partial x} = 0 \quad ; \quad \rho \frac{\partial u}{\partial t} + \frac{\partial p}{\partial x} = f\,u \qquad \text{eq.4a,b}$$

In these equations $A = \pi R^2$ is the cross-sectional area of the tube, A_0 is the cross-sectional area in the reference state, u(x,t) and p(x,t) are the axial velocity and transmural pressure, both averaged over the cross-sectional area, and f is the so-called friction function. The assumption of one-dimensionality holds if the wave length is large compared to the diameter of the tube; upto the fifteenth harmonic in the carotid artery tnis ratio is larger than 40. Due to the assumed linearity the equations can easily be solved using harmonic functions. To that end, two additional assumptions have to be made; one for the friction function and the other for the constitutive relationship between A and p. For the friction term a frequency dependent expression is used, derived from the two-dimensional theory of unsteady fully-developed flow in uniform visco-elastic tubes [11]. Visco-elastic wall behaviour is taken into account by an experimentally determined constitutive relationship between the complex amplitudes of pressure and cross-sectional area variations. The influence of wave reflections is incorporated in the model by treating the carotid artery bifurcation as a bifurcation of uniform tubes. Due to the very large wave length/diameter-ratio the contribution of local inhomogeneities, like the carotid sinus, to the reflection and transmission coefficients appeared to be negligible. Reflections from the outflow ends of the internal and external arteries were incorporated in a standard manner [29].

Next, the wall motion of the carotid bifurcation due to the propagating pressure pulse is analysed. Because of the magnitude of the deformation (relative diameter changes up to 10 % are found) the system to be analysed is geometrically non-linear. For practical reasons the mechanical behaviour of the vessel wall is described with a Hookean constitutive relation. The

material parameters are selected in such a way that diameter changes similar to the ones observed in vivo are obtained. Visco-elastic and inertia effects are neglected, reducing the problem to a quasi-static one. Although the wall of the artery has a thickness which is about 20 times smaller than its diameter, experiments show that its bending stiffness is considerable [35]. These considerations lead to the use of shell elements; compared with the application of three-dimensional brick elements this results in a considerably smaller system of equations. The finite element package MARC is used with an 8-noded thick shell element with global displacements and rotations as degrees of freedom and four integration points [36]. Boundary conditions for the wall motion are applied in a rather pragmatical way [29].

The final step in the numerical approach is the solution of the fluid motion equations using the finite element method described before. However, in this case for each time step the mesh is adapted to account for the wall motion. As inflow condition the fully developed axial and radial velocity profiles as given by Womersley [11] were prescribed. The boundary conditions at the wall were obtained by taking the time derivative of the wall displacement. At both outflow ends boundary conditions of the stress-type $\sigma_n = -p$ and $\vec{\sigma}_t = \vec{0}$ were prescribed. The pressures needed for this purpose were calculated using the quasi-1D model for wave propagation. To investigate the effect of the wall motion, the flow field calculated for the distensible carotid artery bifurcation model was compared with the flow field calculated for a corresponding rigid model. The geometry of the rigid model was chosen to be equal to the end-diastolic geometry of the distensible model. In both cases an identical flow rate at the entrance was prescribed.

In figure 4 the flow rates at the outflow of the rigid model are compared with those in the distensible model. Due to the increase in its volume, part of the volume flow entering the distensible model is accumulated during acceleration. During deceleration this part is released. This effect accounts for the difference between the outflow rates in the rigid and the distensible model. As an illustration of the results, the axial flow field in the internal carotid artery at minimal flow rate (end-systole) is shown in the same figure. Here, contour level 1 corresponds to zero axial velocity; a reversed flow area is observed at the non-divider wall (N). At the entrance of the internal carotid artery (I0) and halfway the sinus (I1) the reversed flow area occupies about 40 % of the local cross-sectional area. For the distensible case both the size of the reversed flow area and the magnitude of the negative velocities are found to be smaller. In both cases Dean-type secondary vortices are observed resulting in C-shaped curvatures of the corresponding axial isovelocity profiles. In the rigid model the curvature is more pronounced. Most likely, the release of "flow accumulated during acceleration" in the distensible model is the dominant factor causing the differences with the flow field observed in the rigid model [29].

The above described flow field in the carotid artery changes considerably as a function of time. The recirculation area is found to develop during deceleration and to disappear during acceleration. Also the secondary flow field and the curvature of the axial isovelocity lines are

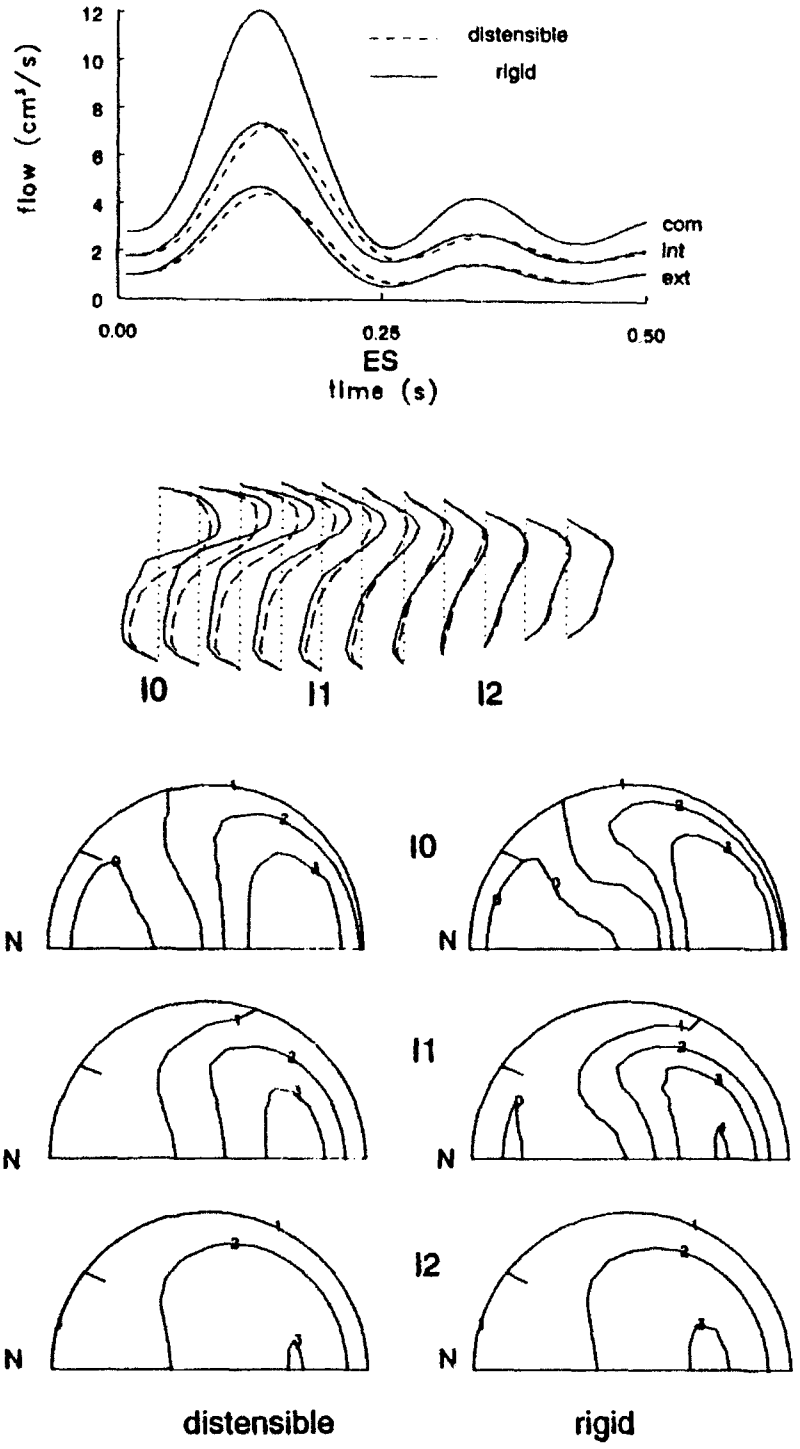


Figure 4. Comparison between axial velocity profiles in the symmetry plane and at three-cross-sections in the internal carotid artery at end-systole (ES), calculated for the distensible and rigid models [29].

strongly time-dependent. In general, the distensibility of the wall is found to reduce some of these effects, although the global structure of the flow field remains unchanged [29]. However, this conclusion should be handled with care as no comparison has been made between the 3D pressure field with the pressure distribution calculated using the quasi-1D model for wave propagation. Since the latter was applied to calculate the wall motion, it should not deviate too much from the 3D pressure field if a consistent solution of the fluid motion is to be obtained. This will be a future research topic.

Fluid-leaflet interaction in the aortic valve

Due to the strong interaction between structure motion and fluid flow, in the numerical model for the dynamic behaviour of the aortic valve fluid and structure are iteratively coupled. The equations of motion of the subsystems are evaluated separately and an iteration is performed until convergence is achieved. Because of computational reasons, the fluid flow and the structure are restricted to be two-dimensional. For the fluid the full unsteady two-dimensional incompressible Navier-Stokes equations are solved by means of the standard Galerkin finite element method described before. For stability reasons, the time integration is performed with a complete Euler implicit scheme (Θ=1 in eq.3). As initial condition, the steady state solution for a fully opened valve is taken. As a boundary condition at the entrance, a time-varying parabolic axial velocity profile is prescribed and the radial flow is set to zero. The entrance channel is chosen to be long enough to guarantee a full development of the unsteady velocity profile. Due to the lack of experimental data this fully developed entrance condition is used, although it certainly does not correspond to the physiological one. As contact condition on the valve, the fluid velocity is set equal to the local valve velocity, determined from the actual valve position and the valve position at the previous time level. At the outlet a stress-free flow condition is prescribed. At the channel walls a no-slip condition is used.

To incorporate the moving structure a segmented valve is considered, consisting of several rigid segments connected to each other. The physiological relevance of this model valve is that with an increasing number (n) of links, the segmented valve will behave more and more like a fully flexible valve. In the special case when n=1, the valve reduces to the so-called rigid valve [37]. In figure 5 some segment quantities are defined. Each segment has one degree of freedom: its angle of rotation φ_i. The inertia of the valve is neglected. At each point of time a segment must satisfy the equilibrium conditions:

$$f^i_{hx1} + f^i_{hx2} + f^i_{fx} = 0 \quad ; \quad f^i_{hy1} + f^i_{hy2} + f^i_{fy} = 0 \qquad \text{eq.5a,b}$$

$$f^i_{hx2}\, l_i \sin\varphi_i + f^i_{hy2}\, l_i \cos\varphi_i + m^i_f = 0 \qquad \text{eq.5c}$$

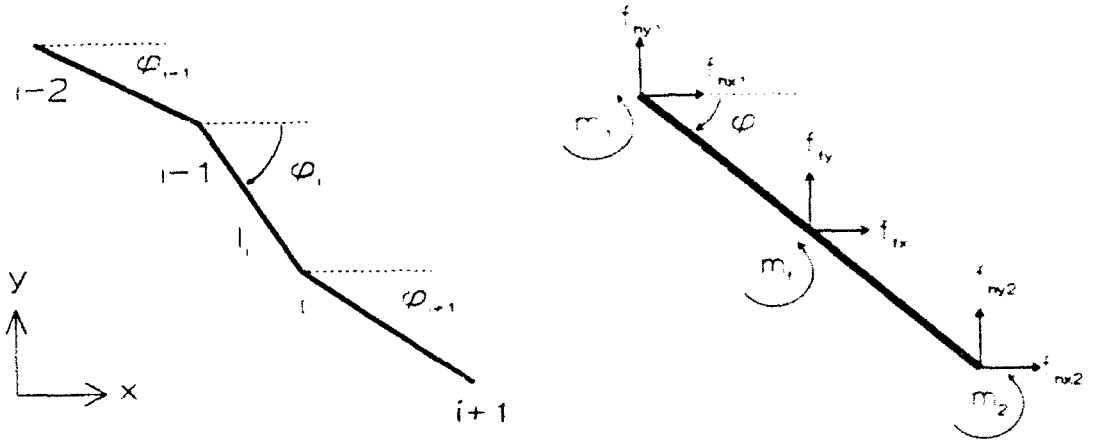


Figure 5. Definition of the geometry and the forces, acting on a segmented valve.

with f_f the fluid force and f_h the hinge force respectively. l is the length of the segment. Index i refers to segment i, subscripts x and y refer to the horizontal and vertical directions, whereas 1 and 2 refer to the left and right edge, respectively m^i_f denotes the moment due to the fluid flow about the center of the hinge with segment i-1. The fluid forces and moments must be calculated by the finite element method. On the segment edges the force continuity conditions:

$$f^i_{hx2} = -f^{i+1}_{hx1} \quad ; \quad f^i_{hy2} = -f^{i+1}_{hy1} \qquad \text{eq.6a,b}$$

are imposed. The boundary conditions for segment n:

$$f^n_{hx2} = 0 \quad ; \quad f^n_{hy2} = 0 \quad ; \quad m^n_2 = 0 \qquad \text{eq.7a,b,c}$$

complete the set of equations. For an arbitrary set φ_i (i=1,n) the system is not in equilibrium. For those states a residual moment m^i_{res} per segment is defined as:

$$m^i_{res} = f^i_{hx2}\, l_i \sin\varphi_i + f^i_{hy2}\, l_i \cos\varphi_i + m^i_f \qquad \text{eq.8}$$

The unknown hinge forces can be calculated in a recursive way:

$$f^i_{hx1} = f^{i+1}_{hx1} - f^i_{fx} \quad ; \quad f^i_{hy1} = f^{i+1}_{hy1} - f^i_{fy} \qquad \text{eq.9a,b}$$

Starting from the valve tip, where boundary conditions (eq.7) can be applied, the hinge forces f^i_{hx} and f^i_{hy} can be calculated for every segment with eq.9, which in turn can be substituted in eq.8. This leads, together with eqs. 5 and 6 and expressions for the fluid moment and forces,

to n equations with unknowns φ_i (i=1,n). This set of equations is solved by reformulating the problem as a non-linear least squares problem. To this end a residual moment is defined:

$$r\left(\underset{\sim}{\varphi}\right) = \sum_{i=1}^{n} m_i \qquad \text{eq.10}$$

where r is the total residual moment to be minimized and m_i the moment acting on segment i, which is a function of a position set $\underset{\sim}{\varphi}$. Once a minimum of r is found which equals zero, $\underset{\sim}{\varphi}$ contains the equilibrium position of the valve segments and therefore of the valve itself. The segment moments m_i depend on all φ_0 ..... φ_n. In general no explicit relation for this dependence can be given, especially not when the finite element fluid model is used. A consequence is that derivatives cannot be calculated. Therefore, a method must be used which does not require the evaluation of derivatives. A very suitable non-linear least squares method for this study is Powell's hybrid method [38]. It combines a robust convergence far from the equilibrium position, with a fast, superlinear convergence close to it. Notwithstanding the robustness of Powell's method, it occasionally fails to find the equilibrium position. In these cases, a new initial estimate is generated and the iteration process is restarted. Using this method, both the numbers of segments (n=1 and n=4) and the physical parameters, like the Strouhal and Reynolds numbers, are varied.

In the numerical calculations a pulsatile flow rate is used. Since the flow rate decleration time is an important parameter in the valve closure mechanism [17] ,this time is chosen as the characteristic time τ. Figure 6 shows the velocity field for n=1, Re=313 and Sr=3.60. At t=0.0 the velocities are small and a vortex is present in the sinus. At t=1.0 the fluid is accelerated and the valve starts to open. At t=1.5 the valve has reached its maximum velocity and pushes out the contents of the sinus. On t=2.0 the flow rate is maximal, the valve is slowing down but still moving towards the fully opened position. Just at t=2.5, when the flow has its maximum deceleration, the valve reaches its maximum position and its velocity becomes zero. At t=2.75 the valve is moving towards the closed position. A vortex in the sinus is being formed. At t=3.0, when the flow rate is nearly zero, the valve has its maximal closing velocity. At t=4.0 the valve has nearly reached its pseudo-equilibrium position and its velocity becomes zero. It is interesting to note that the valve moves already towards its closed position during flow deceleration, just like the natural aortic valve does. This observation is consistent with quasi-1D model predictions [17].

For lower Strouhal numbers the motion of the valve is found to be nearly in phase with the flow rate and its amplitude is larger [30]. The influence of a variation of the Reynolds number between 300 and 1100 appeared to be only marginal. Higher Reynolds number values could not be reached as the fluid solver does not allow computations for Re >1100. Also, similar

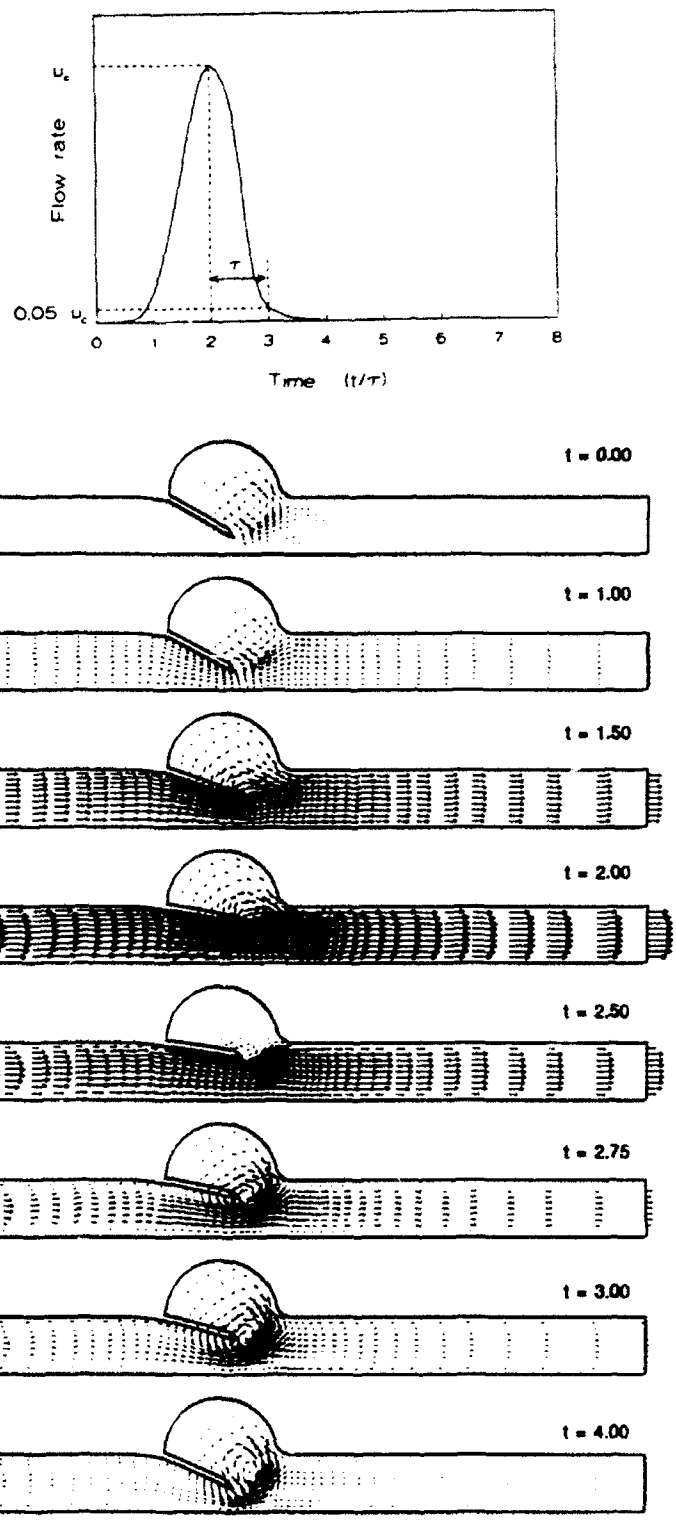


Figure 6. Velocity field around a moving valve in a pulsatile flow at several time levels [30].

results were obtained for a segmented valve with n=4; the curvature changes of the leaflet are only modest. Only at a much higher Strouhal number or when a harmonically varying flow rate was used, large deformations of the valve leaflet were observed. Finally, it need to be mentioned that if the simulation of a full physiological flow cycle is required, including back flow and full valve closure, the model must be extended with a closure algorithm, which is not included in the present model. This and the extension to the three-dimensional case will be a future research topic.

Concluding discussion

From the present study it is concluded that the finite element method can be used for detailed analyses of unsteady blood flow in complex three-dimensional geometries. The local geometry of the bifurcation highly affects the axial and secondary flow fields, but the wall distensibility has only a marginal influence. During flow acceleration and deceleration the valve leaflet shows large displacements, which induce vortex like velocity patterns in the cavity behind the leaflet. As presented elsewhere, a good agreement is found between the numerical and experimental results [27-30].

Some future research topics in the analysis of fluid flow in large blood vessels are [1]:

--- Improvements are needed of the models for fluid-structure interaction. In case of the analysis of the influence of wall distensibility on fluid flow, the improvements need to include a proper account for the effects of reflections originating from the peripheral vascular bed, the application of physiologically relevant boundary conditions in the calculation of wall motion, the use of a model describing the complex mechanical behaviour of the wall and an experimental validation of the uncoupled approach to calculate the flow field in such a distensible geometry. For a complete analysis of the dynamic heart valve behaviour, a three-dimensional fluid model must be applied and in some cases valve inertia or visco-elasticity of the leaflet material has to be incorporated. Furthermore, attention should be given to incorporate the possibility of full valve closure. Then the models can be used to improve the design of existing or newly developed one- or bi-leaflet disc type or flexible leaflet type heart valve prostheses.

--- The influences of non-Newtonian effects and minor stenoses on the local flow field are not completely clear yet. Especially, the visco- elastic properties of the blood may require a considerable effort. For the analysis of the influence of minor stenoses on the flow field spectral element methods and PIV-methods may be favourable, as detailed analyses have to be performed of eddy generation. Besides, the influence of biological variabilities (like bifurcation angle and divider geometry in case of a bifurcation) need research attention.

--- Atherosclerotic plaque formation is found at positions where low wall shear rates occur [39]. Here focal regions of elevated permeability to macromolecules, temporary gaps and

a thin endothelial glycocalyx are found. Modelling of the arterial wall is worthwhile, focussing on maximal stress concentration in the intima and the mass transport of atherogenic proteins between plasma and vessel wall [40]. Then probably an improvement can be reached with regard to the understanding of the fluid mechanical contribution to the process of atherogenesis.

Acknowledgements

The work presented here is mainly based on the PhD-theses of F.N. van de Vosse, C.C.M. Rindt, P.J. Reuderink and J.B.A.M. Horsten. The research was supported by the Dutch Organization for Scientific Research and the Dutch Technology Foundation.

References

[1] AA van Steenhoven, APG Hoeks, JD Janssen and RS Reneman (eds), Three-dimensional blood flow in bifurcations, Abstractbook Euromech 286, Eindhoven, 1991.

[2] RS Reneman, T van Merode and APG Hoeks, News in Physiolog. Sciences, 4 (1989) 185

[3] D Liepsch and S Moravec, Biorheology, 21 (1984) 571

[4] TJ Pedley, The fluid mechanics of large blood vessels, Cambridge, Cambridge, 1980

[5] MJ Lighthill, Waves in fluids, Cambridge, Cambrige, 1978

[6] CG Cáro, TJ Pedley, RC Schroter and WA Seed, The mechanics of the circulation, Oxford, Oxford, 1978.

[7] Y Henderson and FE Johnson, Heart, 4 (1912) 69

[8] BJ Bellhouse and L Talbot, J. Fluid Mech., 35 (1969) 721

[9] CS Peskin, Annu. Rev. Fluid Mech., 14 (1982) 235

[10] FK Wipperman, J. Fluid Mech., 159 (1985) 487.

[11] JR Womersley, J. Physiol., 127 (1955) 553

[12] DE Olson, Fluid Mechanics relevant to respiration, PhD-thesis, University of London, 1971

[13] LS Yao and SA Berger, J. Fluid Mech., 67 (1975) 177

[14] BK Bharadvaj, RF Mabon and DP Giddens, J. Biomech., 15 (1982) 349

[15] DN Ku, DP Giddens, CK Zarins and S Glagov, Atherosclerosis, 5 (1985) 293

[16] K Perktold, M Resch and OP Reinfried, J. Biomech., 24 (1991) 409

[17] AA van Steenhoven and MEH van Dongen, J. Fluid Mech., 90 (1979) 21

[18] AA van Steenhoven and MEH van Dongen, J. Fluid Mech., 166 (1986) 93

[19] AA van Steenhoven and FJHM van de Beucken, J. Fluid Mech., 231 (1991) 599

[20] PH Bovendeerd, AA van Steenhoven, FN van de Vosse and G Vossers, J. Fluid Mech., 177 (1987) 233

[21] FN van de Vosse, AA van Steenhoven, A Segal and JD Janssen, Int. J. Num. Meth. Fluids, 9 (1989) 27

[22] CCM Rindt, AA van Steenhoven, JD Janssen and G Vossers, J. Fluid Mech., 226 (1991) 445

[23] CCM Rindt, FN van de Vosse, AA van Steenhoven, JD Janssen and RS Reneman, J. Biomech., 20 (1987) 499

[24] FN van de Vosse, AA van Steenhoven, JD Janssen and RS Reneman, Biorheology, 27 (1990) 163

[25] CCM Rindt, AA van Steenhoven and RS Reneman, J. Biomech., 21 (1988) 985

[26] CCM Rindt, AA van Steenhoven, JD Janssen, RS Reneman and A Segal, J. Biomech., 23 (1990) 461

[27] FN van de Vosse, Nummerical analysis of carotid artery flow, PhD-thesis, Eindhoven University of Technology, 1987

[28] CCM Rindt, Analysis of the 3D flow field in the carotid artery bifurcation, PhD-thesis, Eindhoven University of Technology, 1989

[29] PJ Reuderink, Analysis of the flow in a 3D distensible model of the carotid artery bifurcation, PhD-thesis, Eindhoven University of Technology, 1991

[30] JBAM Horsten, On the analysis of moving heart valves, PhD-thesis, Eindhoven University of Technology. 1990

[31] A Segal, Sepran Manual, Sepra, Delft, 1984

[32] FN van de Vosse, A Segal, AA van Steenhoven and JD Janssen, Int . J. Num. Meth. Fluids, 6 (1986) 427

[33] C Cuvelier, A Segal and AA van Steenhoven, Finite element methods and Navier-Stokes equations, Reidel, Dordrecht, 1986

[34] JBAM Horsten, AA van Steenhoven and MEH van Dongen, J. Biomech., 22 (1989) 477

[35] WR Milnor, Hemodynamics, Williams and Wilkins, Baltimore, 1989

[36] Marc K.3, Manuals A-E, Marc Analysis, Palo Alto, 1988

[37] JBAM Horsten, AA van Steenhoven and DH van Campen, in: P. Wesseling (ed), Notes on Numerical Fluid Mechanics 29, Vieweg, Braunschweig, 1990

[38] MJD Powell, in: P Rabinowitz (ed), Numerical methods for algebraic equations, Gordon and Breach, London, 1970

[39] CK Zarins, DP Giddens, BK Bharadvaj, VS Sorriurai, RF Mabon and S Glagov, Circulation Res., 53 (1983) 502

[40] TJ Pedley, IUTAM Proceedings, 15 (1980) 151

Theoretical and Applied Mechanics 1992
S.R. Bodner, J. Singer, A. Solan & Z. Hashin (Editors)
Elsevier Science Publishers B.V.
© 1993 IUTAM. All rights reserved.

Trends in transonic research

Jürgen Zierep

Institut für Strömungslehre und Strömungsmaschinen, Universität (TH) Karlsruhe, Kaiserstrasse 12, D-W7500 Karlsruhe 1, FRG

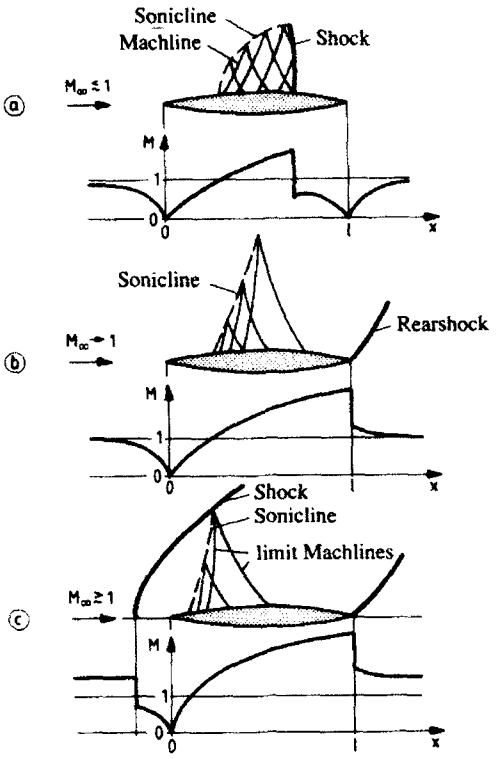
Abstract

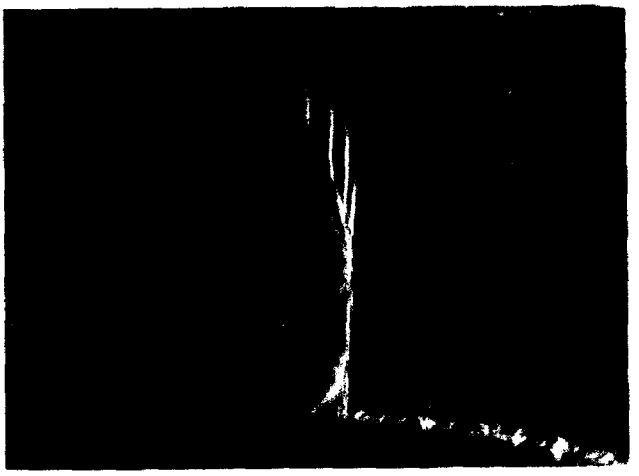
Transonic flows are now as before fascinating and very important in respect to theory, experiment and applications. In the early days analytical methods prevailed that led to fundamental results. Now a days due to increasing power of computers numerical methods dominate. Validation problems of Euler- and Navier-Stokes Codes are of primary interest. This led to simple but fundamental experiments (DFVLR-F5 wing, vortex flow, shock boundary layer interaction, condensation phenomena, ...) and the comparison with theory. The status quo of validation is today not sufficient in all respects. There are complex dependences on the block structures, the meshes, the laminar-turbulent transition and the turbulence models. Some future trends: Search for reliable codes for 3D-boundary value problems to develop airfoils of low drag - especially laminar transonic airfoils; study of diabatic flows of phase changing media with several applications (lift and drag of airfoils, steam turbines, cryogenic tunnel, Organic Rankine Cycle, ...); transonic flows in retrograde media (e.g. BZT fluids).

1. INTRODUCTION: WHAT IS A TRANSONIC FLOW?

In transonic flows the magnitude of the fluid velocity is comparable with the sonic speed (= signal velocity). The body moves nearly with the same velocity with which disturbances propagate. The theoretical and experimental investigation of these flows is now and as has been before, of great interest and has become of considerable technical significance. Now a days there are many applications in aerodynamics of transonic flight for instance in the field of civil aviation and thermal fluid machinery which underline even the economical importance of these investigations. The declared aim is: reliability, high capacity, less drag, less noise, less pollution.

The typical qualities of these flow fields, especially the mathematical and physical problems involved, can already be understood by studying slender body flows (Fig. 1). In transonic slightly subsonic flow $M_\infty \leq 1$ a *local supersonic region* develops in the neighbourhood of maximum thickness which as a rule is limited downstream by a normal shock (Fig. 2). The local Mach number distribution on the airfoil explains the flow. In front of the body the flow decelerates, behind the stagnation point acceleration to supersonic





⇑ Figure 2. Schlieren picture. Rough surface. Mach lines and shock. Supersonic region.

⇐ Figure 1. Fluid fields and Mach number distributions on the profiles.
a) $M_\infty \lesssim 1$, b) $M_\infty \to 1$, c) $M_\infty \gtrsim 1$.

velocity exists. In the shock the velocity jumps to subsonic velocity with following postexpansion (Figs. 3,4). After this deceleration to backward stagnation point with following approach to conditions at infinity. In transonic slightly supersonic flow $M_\infty \gtrsim 1$ the shock-wave detaches as a rule. Between shock and body a *local subsonic region* builds up. The displacement effect of the body leads to an acceleration to supersonic

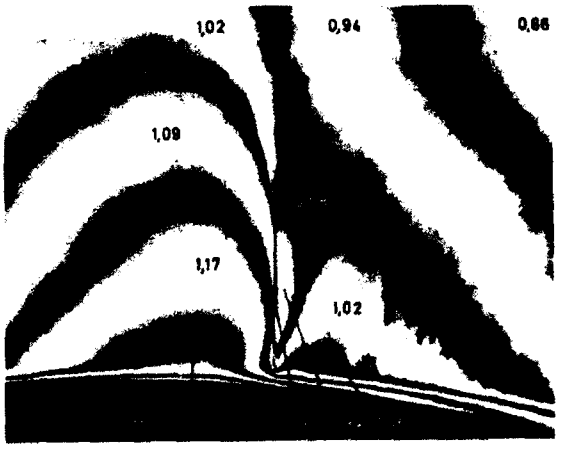


Figure 3. Black-white Mach-Zehnder interferogram of transonic flow at a curved surface. M = local Mach number.

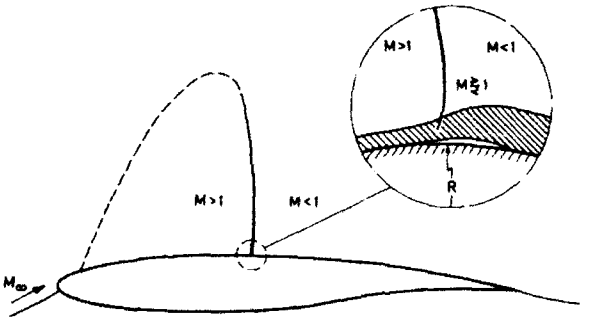


Figure 4. Shock boundary layer interaction at an airfoil.

velocity till the trailing edge shock is reached. The limit Machline is the last characteristic from the airfoil that reaches the subsonic region, while the last characteristic running down from the subsonic domain limits the range of influence downstream. $M_\infty \to 1$ leads to the limiting case of sonic velocity at infinity. The sonic line extends to infinity. At the body the flow accelerates until the rear shock is reached. Comparing the Mach number distributions on the airfoil we see, that only slight changements exist; the *flow freezes.* A simple explanation is as follows: if M_∞ is only slightly above one the shock wave stands far upstream with Mach number behind the shock $\hat{M}_\infty \leq 1$. That means the profile don't realize whether is stands in a transonic supersonic or subsonic flow. With other words the local dates at the profile change from $M_\infty \leq 1$ to $M_\infty \geq 1$ only slightly.

The mathematical description leads even for slender bodies with small disturbance theory to quasilinear partial differential equations. On the sonic line (sonic surface) the type changes, across the shock curve (shock surface) the shock relations have to be fulfilled. Both curves (surfaces) are not known a priori and have to be determined with the solution.

2. TRENDS IN THE PAST

The great technical importance of transonic flows on the one hand and the interesting mathematical-physical problems on the other hand had led to an extensive theoretical and experimental treatment of relevant problems in the past 50 years. The Proceedings of IUTAM Symposium Transsonicum I 1962 (Aachen) [1], II 1975 (Göttingen) [2], III 1988 (Göttingen) [3] give an interesting, unique survey of the worldwide efforts until today. Concerning theory, 1962 the analytical methods prevailed. Among others we have the hodograph method, the parabolic (local linearization) method, the characteristics method and the integral equation method. A lot of general results came out like similarity laws, equivalence rule, area rule, asymptotic expansions, ... Beside this some exact solutions for special profiles had been discovered. The small power of computers at that time didn't allow the numerical treatment of the direct problem. Already 1975 the interest changed remarkable to Computational Fluid Dynamics (CFD). The increasing power of computers allowed for the first time the direct numerical treatment of the nonlinear differential equations in the fluid field. Questions of stability, convergence and estimation of errors were more or less open this time. This changes drastically in the following years.

3. TRENDS TODAY ILLUSTRATED BY DIFFERENT CODE VALIDATIONS

The Proceedings of the Symposium Transsonicum III [3] describe this development. The concluding remarks by Hornung give a (subjective) pointed description worth reading of the status quo with the progress achieved but with open questions too. Concerning the CFD-methods the validation of Euler- and Navier-Stokes Codes are of primary interest. This led to some simple but fundamental experiments, e.g. the DFVLR-F5 wing experiment (Fig. 5) and the vortex flow experiment (Fig. 6) and the comparison

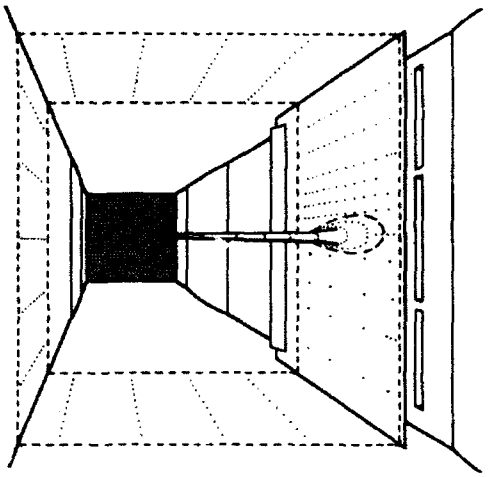


Figure 5. Test airfoil DFVLR-F5 in the transonic 1m x 1m test section in Göttingen.

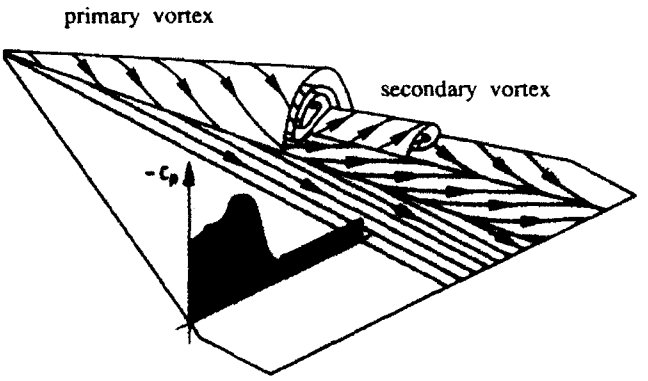


Figure 6. Streamlines of the primary and secondary vortex ($M_\infty = 0.85, \alpha = 10°$).

with numerical calculations for these cases. Some other experiments and validations will be discussed later on.

In principle there are two concepts for the numerical calculation of the viscous flow around wings. By using the *zonal numerical approach* in different domains of the flow field different differential equations according to the physical effects will be used (Fig. 7). Outside the boundary layer in front of the shock we use the potential equation, behind the shock the Euler equations, near the body the boundary layer equations, near the trailing edge and in the wake region the Navier-Stokes equations. The shock-boundary layer interference is given by a separate local solution which has to be integrated iteratively in the global flow field. The second method solves the *full Navier-Stokes equations* in the total flow field. The main point is to choose a suitable block-structured grid which secures a high resolution in the viscous interaction region and at the shock and the trailing edge. Within the inviscid flow field zones a reduced resolution is possible. This concept needs an optimization in the choise of computational grids and blocks.

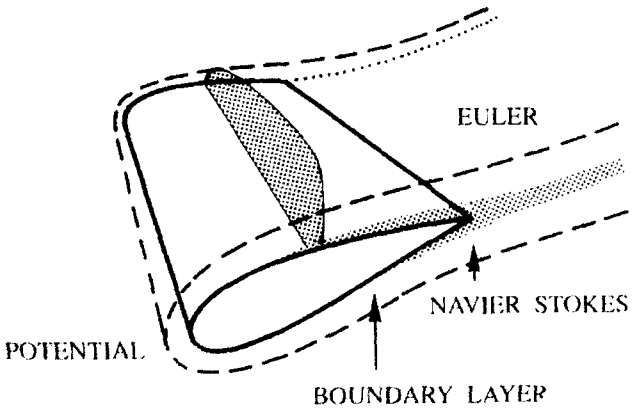


Figure 7. Phenomena and zonal model equations for transonic flow at an airfoil.

3.1

A workshop was organized by Kordulla and Sobieczky [4] before the Symposium Transsonicum III. A well-defined boundary value problem, the flow over a wing (DFVLR-F5) in a wind tunnel was formulated (Figs. 8, 9). Boundary conditions were given at the walls and in the entrance and exit area. All groups and institutions that had developed suitable Navier-Stokes programs, were invited to participate in solving this flow problem numerically. Active participants came from: NASA Ames, NASA Langley, NAL Japan, FFA Sweden, MBB Munich, DFVLR Göttingen. All theoretical results were collected, compared with each other and with the experiments performed in Göttingen. This seems to be a milestone and should be continued in future.

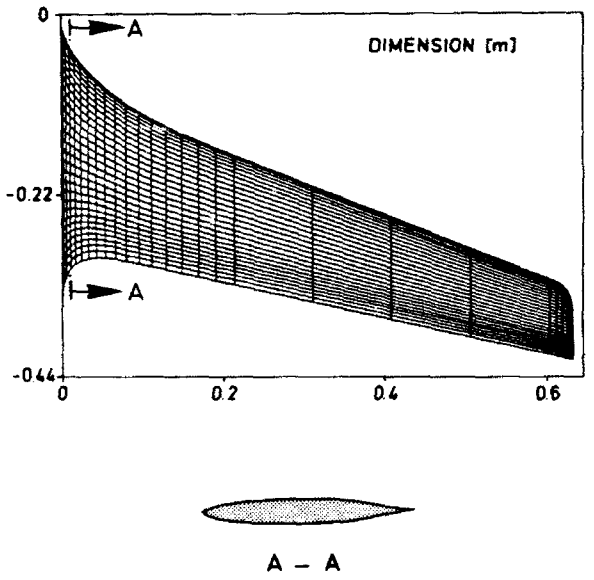


Figure 8. 3D - test airfoil [4].

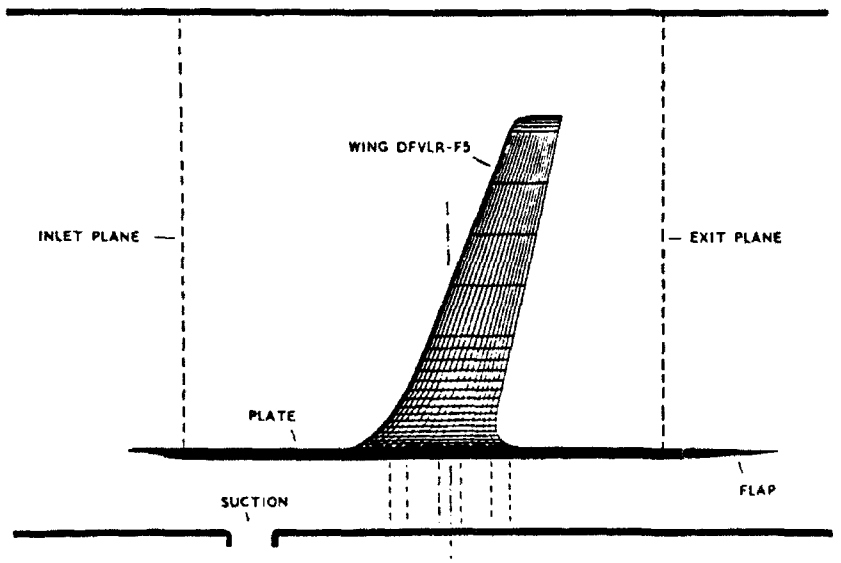


Figure 9. Wing-plate configuration in 1m x 1m transonic wind tunnel [4].

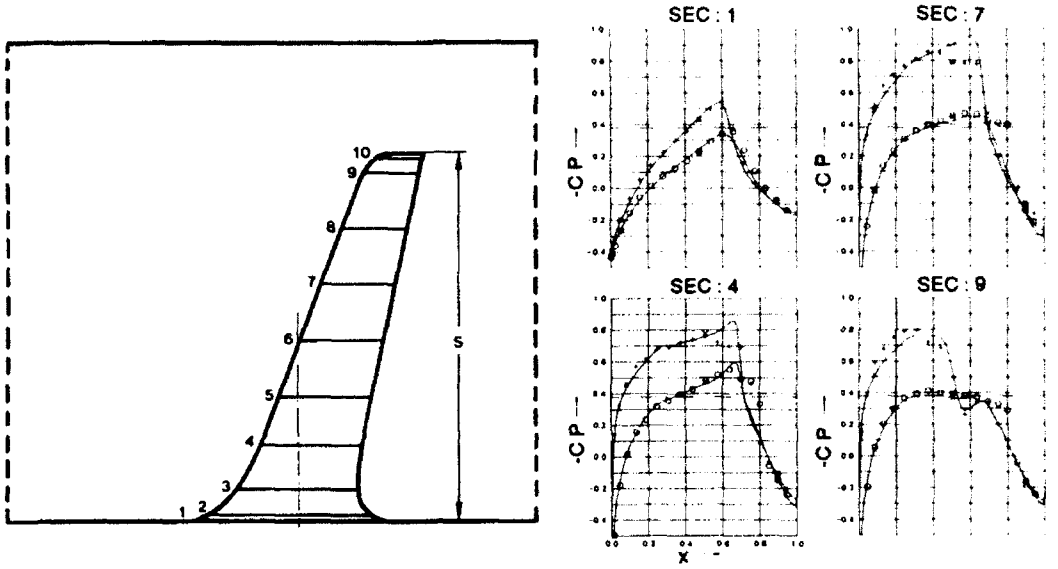


Figure 10. Chordwise c_p - distributions on the wing surface in different sections (1,4,7,9). Experimental data given by symbols, solid lines theoretical results(Schwamborn). Flow conditions: Free-stream Mach number $M_\infty = 0.82$, $\alpha = 2°$, $Re/m = 10^7$ [4].

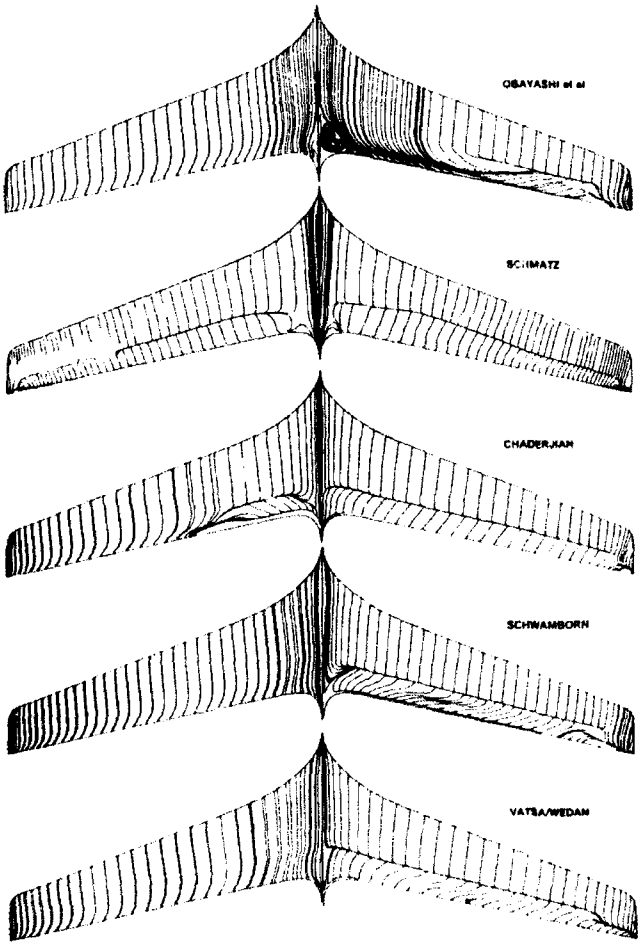


Figure 11. Comparison of predicted skin friction line patterns for the wind tunnel test. $\alpha = 2°$. Left: lower wing surface. Right: upper wing surface. [4].

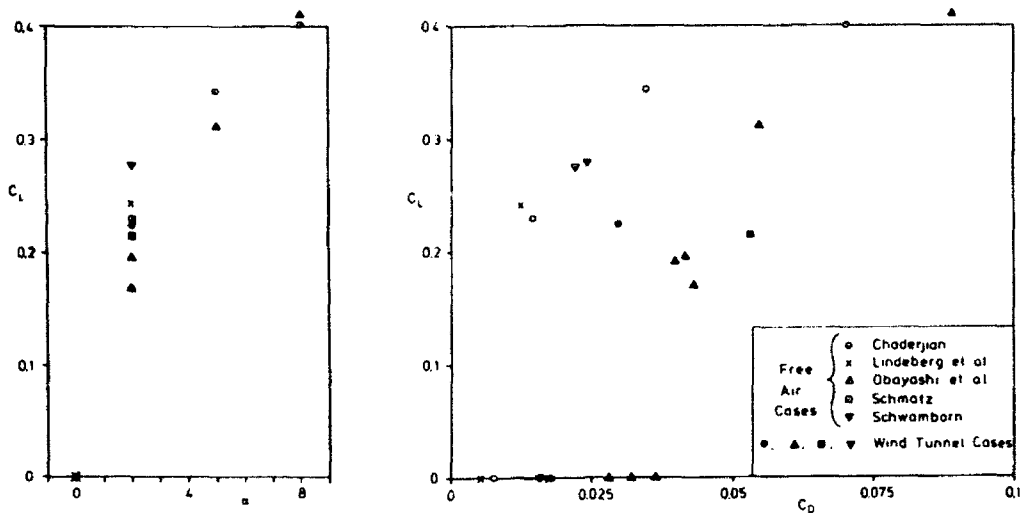


Figure 12. Lift and drag coefficients of the DFVLR-F5 Navier-Stokes solutions [4].

The pressure distributions on the wing and in the field agree quite well among the participants and with the experiment (approximately 10%). The results are not very sensitive (Fig. 10). The skin friction lines, especially if separation exists, are much more sensitive (Fig. 11). Disappointing is the scatter of the lift (c_L) and the drag coefficients (c_D) (Fig. 12). There is a remarkable and complex dependence on the nature of the used meshes and block structures in streamwise and chordwise direction (C-O, C-H, O-O, H-H-type) and the discretization (FVM, FDM). Beyond this the laminar-turbulent transition (switch on) and the used turbulence model (Baldwin-Lomax) may be of considerable influence. This is doubtless a disillusion and not recognized with pleasure but it is important for all further investigations in that field. Without strict validation of the numerical codes no reliability can be guaranted!

3.2

The next validation complex is experiment and computation of transonic viscous channel flow with ventilation. Many efforts at different places had been done [5, 6, 7]. The practical idea is the reduction of shock strength by passive control in the domain of turbulent shock-boundary layer interaction. Figure 13 describes on the one hand the scheme of the test section in a laboratory and on the other hand the different zonal methods for the numerical approach. Outside the boundary layer we have the nonviscous flow part. We solve the potential equation after suitable grid generation by well known difference methods. At the wall we start with laminar boundary layer, after transition we continue with turbulent boundary layer. In the shock boundary interaction area we fit in an analytical model [8]. In the cavity closed with a punched sheet we prescribe a constant pressure and fulfil the conservation of mass. Downstream we have a turbulent boundary layer with or without separation. An inverse integral method has to be applied [9]. The

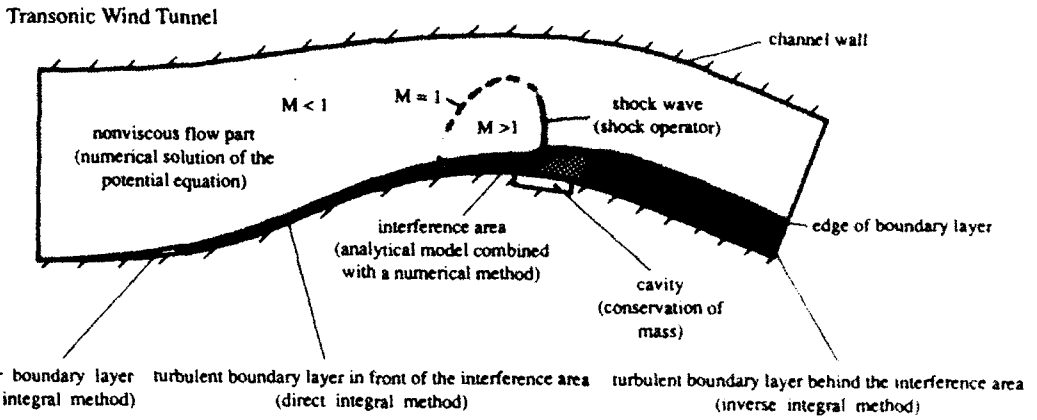
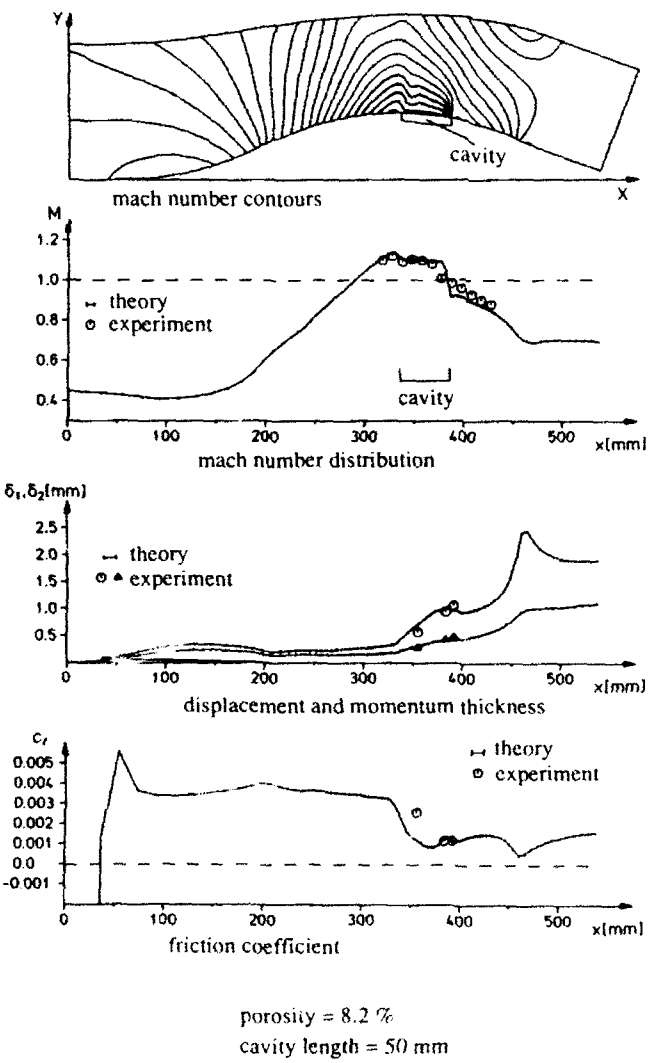


Figure 13. Zonal methods for computation of transonic viscous channel flow with ventilation.



porosity = 8.2 %
cavity length = 50 mm

Figure 14. Comparison of theory and experiment for ventilation. Porosity = 8.2%, cavity length = 50 mm. The mass flux corresponds to $M_{max} = 1.2$ in the case without ventilation [6].

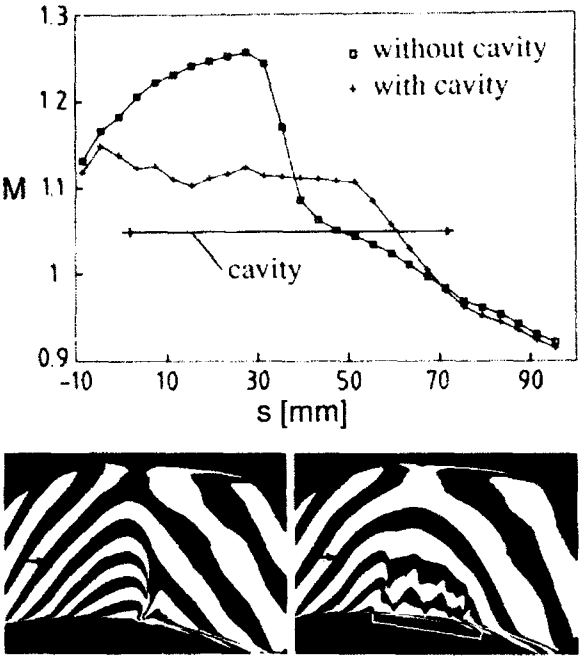


Figure 15. Interferogram without and with ventilation and the corresponding Mach number at the wall [7].

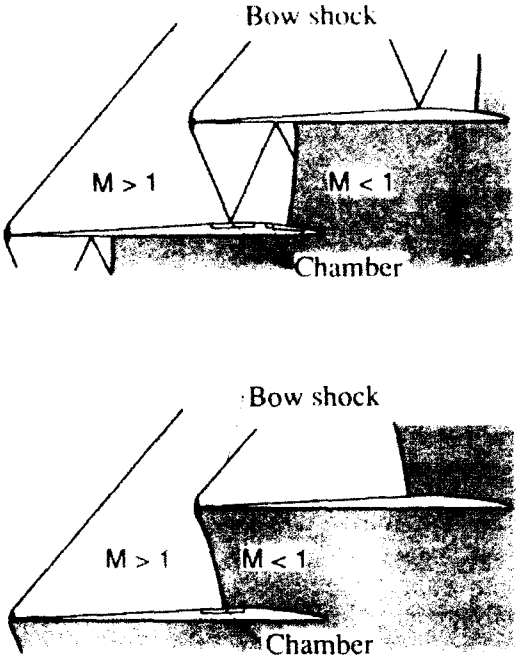


Figure 16. Cascade in supersonic flow. Top: Reflexion of bow shock wave at the profile. Bottom: Bow shock wave closes the supersonic region.

calculation and the measurements for Mach number and skin friction at the lower wall and the displacement - and momentum thickness agree well (Fig. 14). The comparison with the nonventilated case is convincing. The shock strength is considerably reduced (Fig. 15). For a single profile the wave resistance decreases - until 50% - and the shock position is fixed. It is remarkable that the effect of passive control is already reached with a perforation of about 8%. Further on the inclination of the perforation-channels is of considerable influence (oblique perforation).

The method described here may be of interest for several applications. For compressor cascades in supersonic flow for instance it is possible to reduce the loss by passive control on the suction side of the blades (Fig. 16). Experiments in a cascade tunnel [10] and in channel test section obviously show this effect. Calculation with zonal methods are possible in a way we have described above.

In the last time another idea of passive control has been propagated [11]. The cavity is closed by an elastic membrane. Depending upon the elasticity and the inner pressure the membrane builds up some kind of adaptive wall that may have a favourable influence on the shock-boundary layer interaction.

3.3

A very interesting new experiment for validation of modern transonic numerical methods is the condensation, or more general the study of flows with phase transition, in Laval nozzles and around profiles. This is important for several reasons. On the one hand this is a diabatic flow with interesting new physical effects on a microscopic and a macroscopic scale (nucleation theory, droplet growth), on the other hand there are several actual applications in steam turbines and in cryogenic tunnels, not to forget the

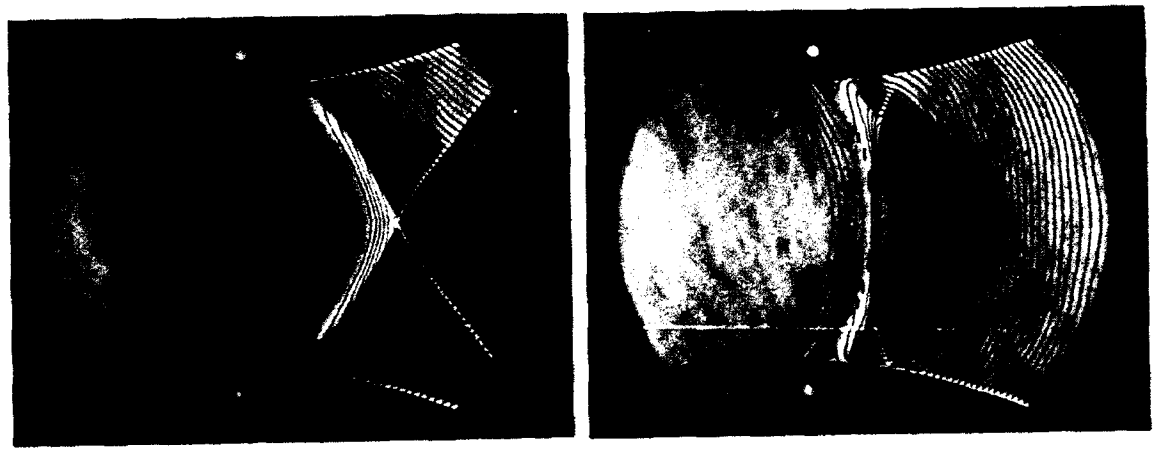


Figure 17a,b. Schlierenpictures and calculated iso Mach lines. Left: $\phi_0 = 29\%$, subcritical heat addition; Right: $\phi_0 = 60\%$, supercritical heat addition.

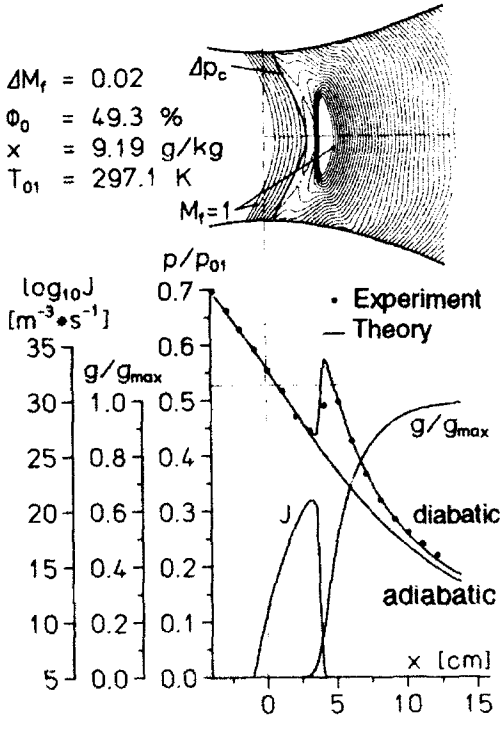
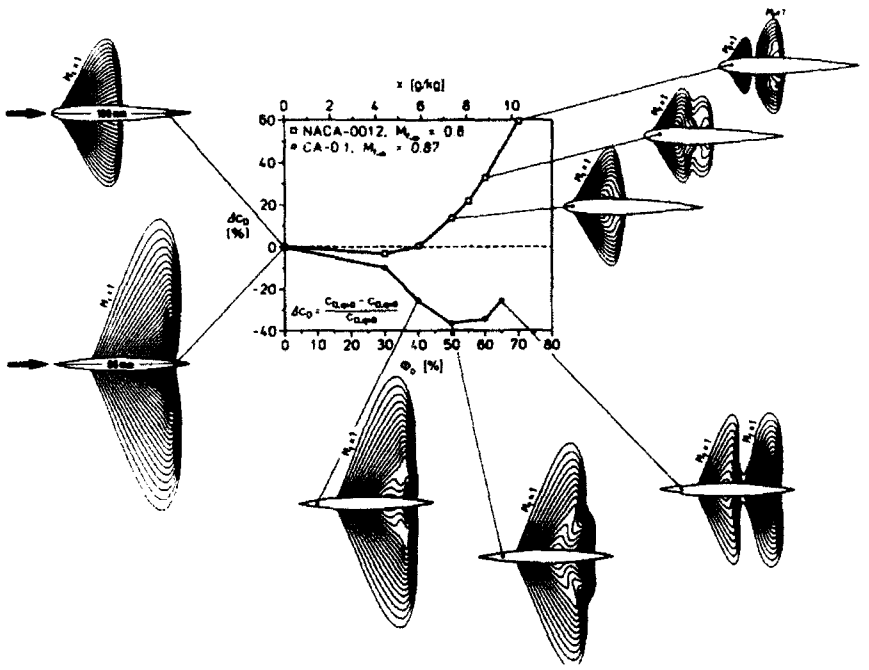


Figure 18. Diabatic flow in a hyperbola nozzle. On the axis: Nucleation rate J, Condensing mass g/g_{max}, Pressure ratio p/p_{01}.



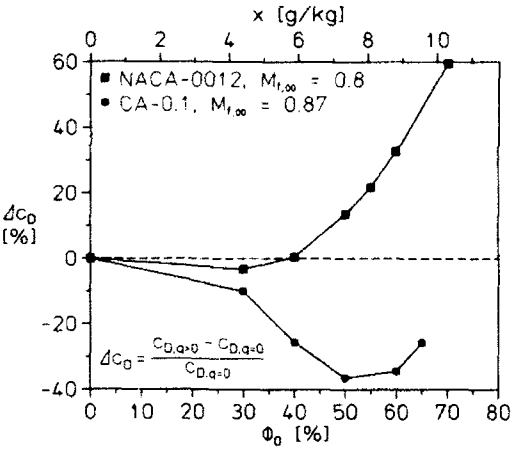
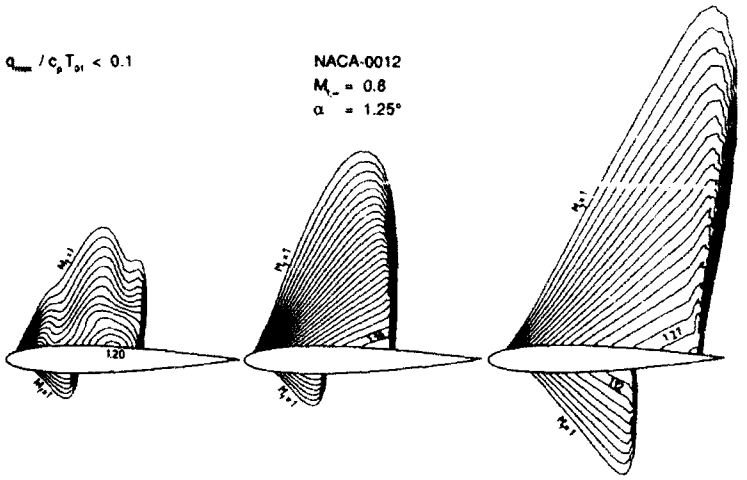


Figure 19a,b. Changement of the pressure drag for circular arc profile (CA-0.1) chordlength c = 80 mm and NACA-0012 profile (c = 100 mm) by heat addition due to homogeneous condensation, without incidence [12].

fligh of an aircraft in humid air. Condensation gives rise to heat release that influences the pressure distribution considerably. Strong changements of lift and wave drag coefficients are possible in moist atmospheric air [12]. The numerical method solves the unsteady 2D-Euler equations. Iteration is finished if the stationary final solution is obtained. All equations are used in conservation form and the condensation is modelled by two successive processes. For homogeneous condensation we start with the nucleation theory by Volmer [13] and continue with the droplet growth theory according to the Hertz-Knudsen model. The calculation uses a finite-volume method that generalizes a corresponding procedure for adiabatic flows [14]. Depending upon the stagnation humidity ϕ_0 we find different structures of condensation phenomena in the supersonic part of a nozzle - subcritical and supercritical heat addition (Fig. 17a,b) [15]. The



Nonequilibrium		Equilibrium
Δc_D = - 12%	adiabatic	Δc_D = 207%
Δc_L = - 30%		Δc_L = 30%

Figure 20. Non-equilibrium, adiabatic and equilibrium flow for NACA-0012 profile, $\alpha = 1.25°$, $M_f = 0.8$, $\Delta M_f = 0.015$ [16].

comparison between calculation and measurements is convincing. This is valid in respect to the Schlierenpictures with iso-Machlines and in respect tc the static pressure measurements along the axis. Figure 18 gives as microscopic quantity the nucleation rate J and as macroscopic values the condensed mass g/g_{max} and the static pressure p/p_{01}. Concerning profile flow figure 19 shows for circular arc (CA-0.10) and NACA-0012 profile the results. Depending upon the stagnation humidity ϕ_0 the pressure drag may be considerably *reduced* (CA-0.10) or *increased* (NACA-0012) [12]. A simple explanation is as follows: For circular arc profile and adiabatic flow the shock position is in the neighbourhood of the rear, for $M_{f,\infty} = 0.87$. Condensation will not move this shock. The changement of the pressure due to heat addition leads to *de*creasing drag with increasing humidity. For NACA profile on the contrary the shock position in adiabatic flow is in the middle of the profile, for $M_{f,\infty} = 0.8$. Here due to condensation the shock moves. The changement of the pressure due to heat addition leads to an *in*creasing drag with increasing humidity. Remarkable is the typical double shock structure for high humidity! The first is due to super-critical heat addition, the second is the usual gasdynamic shock. If unsymmetric flows with incidence are considered things change once more [12]. Now for NACA profile the shock is for adiabatic flow in the neighbourhood of the rear and drag and lift *de*crease with increasing humidity. Figure 20 shows the different effects between non-equilibrium condensation and equilibrium condensation on the one hand and adiabatic flow on the other hand. The case of equilibrium condensation corresponds to atmospheric flight conditions with chordlength c = 1500 mm. A drastic increase of the drag is the consequence.

4. FUTURE TRENDS

The last examples lead already to efforts in future. In the field of CFD-methods now as before reliable Euler- and Navier-Stokes codes for actual problems of modern aerodynamics are wanted. Validation is necessary in any case. Of special interest are airfoils of low drag. This aim may be reached by design of shockless profiles [17] or by active or passive control of the shock boundary layer interaction.

The transonic laminar airfoil is simply the optimum. If it is possible to suppress the instabilities in the boundary layer and keep it laminar over the main part of the airfoil this would be a great success.

Phase transition effects are of increasing importance as we have seen. The situation in cryogenic tunnels belongs to these subjects and the study of condensation of N_2 is of great interest [18,19]. A possible application of phase changing fluid flow lies in a new type of fluid flow machines, using a so called Organic Rankine Cycle (ORC). Some new investigations show that remarkable features exist which lead to large efficiencies of these systems [20].

Last not least the transonic flow in retrograde (dense) fluids is in many respects surprising [21]. Some well known classical gasdynamic effects invert [22,23,24]. For instance, *expansion shocks* may be possible with a favourable influence on boundary layer separation; *isentropic compression* of these media may lead to condensation and so on.

At Symposium Transsonicum III Earl Murman gave a delightful comment: "What we are doing is fun, it might even be important". After working in the fascinating field of transonic flow for over 35 years I am convinced that it *is* important.

5. REFERENCES

1 Proceedings IUTAM Symposium Transsonicum I, ed: Oswatitsch, K., Springer-Verlag, 490 pages, 1964

2 Proceedings IUTAM Symposium Transsonicum II, eds: Oswatitsch, K. and Rues, D., Springer-Verlag, 574 pages, 1976

3 Proceedings IUTAM Symposium Transsonicum III, eds: Zierep, J. and Oertel, H., Springer-Verlag, 504 pages, 1989

4 Numerical Simulation of the Transonic DFVLR-F5 Wing Experiment. Ed: Kordulla, W., Notes on numerical fluid mechanics, Vol. 22, Vieweg, 305 pages, 1988

5 Délery, J.: Recent Basic Studies in Transonic Shock-Wave/Turbulent Boundary-Layer-interactions. See [3], pp 357-377

6 Breitling, Th.: Berechnung transsonischer, reibungsbehafteter Kanal- und Profilströmungen mit passiver Beeinflussung. Dissertation Universität Karlsruhe 1989, 90 pages

7 Braun, W.: Experimentelle Untersuchung der turbulenten Stoß - Grenzschicht-Wechselwirkung mit passiver Beeinflussung. Dissertation Universität Karlsruhe 1990, 98 pages

8 Bohning, R. und Zierep, J.: Der senkrechte Verdichtungsstoß an der gekrümmten Wand unter Einfluß der Reibung. ZAMP 27, 225-240 (1976)

9 Bohning, R. and Jungbluth, H.: Turbulent Shock-Boundary-Layer Interaction with Control. Theory and Experiment. See [3], 389-398.

10 Yu, Shen; Li, Jianyi; Yu, Bin and Fan, Chungfu: Some preliminary experimental results on passive control of shock wave boundary layer interaction in compressor cascades. In: Proceedings of the 1st International Symposium 'Experimental and Computational Aerothermodynamics of Internal Flows', Beijing 1990, World Publishing Corporation, pp. 217-219

11 Breitling, R. and Zierep, J.: Computation of transonic viscous flow over airfoils with control by an elastic membrane and comparison with control by passive ventilation. Acta Mechanica 87, 23-36 (1991)

12 Schnerr, G.H. and Dohrmann, U.: Drag and Lift in Nonadiabatic Transonic Flow. AIAA 91-1716, AIAA 22nd Fluid Dynamics, Plasma Dynamics and Lasers Conference, June 24-26, 1991, Honolulu

13 Volmer, M.: Kinetik der Phasenbildung. Steinkopf, Leipzig, 1939

14 Eberle, A.: A new flux extrapolation scheme solving the Euler equations for arbitrary 3-D geometry and speed. MBB/LKE 122/PUB/140, 1984

15 Delale, C.F.; Schnerr, G.H. and Zierep, J.: The Mathematical Theory of Thermal Choking in Nozzle Flows. J. Fluid Mech., submitted for publication (1992)

16 Schnerr, G.H.: Transonic Aerodynamics including strong Effects from Heat Addition. 4th Int. Symposium on Compressible Fluid Dynamics, University of Davis, California. Computers & Fluids (1992)

17 Sobieczky, H.: Verfahren für die Entwurfsaerodynamik moderner Transportflugzeuge. DFVLR-FB 85-05 (1985)

18 Wegener, P.P.: Cryogenic Wind tunnels and the Condensation of Nitrogen. Exp. in Fluids 11, 333-338 (1989)

19 Schnerr, G.H.: Gasdynamische Effekte in transsonischen Strömungen durch Energiezu- und -abfuhr nach partiellem Phasenwechsel des Fluids. ZAMM 71, T 401-408 (1991)

20 Zörner, W. and Blumenberg, J.: The Organic Rankine Cycle for solar-dynamic space power generation. Zeitschr. f. Flugwiss. und Weltraumforschung 13, 260-270 (1989)

21 Cramer, M.S.: Transonic Flows of BZT fluids. Proceedings of 13th IMACS Congress on Appll. Math. and Scientific Computations, Dublin 1991

22 Thompson, P.A.: A Fundamental Derivative in Gasdynamics. Physics of Fluids 14, 1843-1849 (1971)

23 Cramer, M.S.: Nonclassical Dynamics of Classical Gases. In: Waves in Real Fluids (ed. Kluwick, A.), Springer-Verlag, Heidelberg, New York, Tokyo, in press

24 Cramer, M.S. and Kluwick, A.: On the Propagation of Waves Exhibiting Positive and Negative Nonlinearity. J. Fluid Mech. 142, 9-37 (1984).

LIST OF CONTRIBUTED PAPERS PRESENTED AT THE CONGRESS

Abbott, J.R., Graham, A.L., Mondy, L.A., Brenner, H.
DISPERSION ACCOMPANYING THE SETTLING OF A BALL THROUGH A CONCENTRATED NEUTRALLY BUOYANT SUSPENSION

Abeyaratne, R., Knowles, J.K.
REFLECTION AND TRANSMISSION OF WAVES FROM AN INTERFACE WITH A PHASE-TRANSFORMING SOLID

Acrivos, A., Fan, X., Mauri, R., Nir, A.
VISCOUS RESUSPENSION IN COUETTE FLOWS

Acrivos, A.: See Kapoor, B.

Acrivos, A.: See Wu, W.-Y.

Adan, M., Sheinman, I., Altus, E.
POST-BUCKLING BEHAVIOR OF BEAMS UNDER CONTACT CONSTRAINTS

Adziev, T., Wnuk, M.P.
INFLUENCE OF PRESTRAINING AND RESIDUAL STRESSES ON MATERIAL RESISTANCE TO CRACK PROPAGATION

Afeti, G.M., Resch, F.J.
FROM BREAKING WAVES TO BREAKING BUBBLES: PRODUCTION CHARACTERISTICS OF THE MARINE DROPLET AEROSOL

Agarwal, R.K.
A CFD-BASED APPROACH FOR THE SOLUTION OF AEROACOUSTICS, MAXWELL, SCHROEDINGER, AND BOLTZMANN EQUATIONS

Agnon, Y., Glozman, M., Stiassnie, M.
HIGH-ORDER HAMILTONIAN AND CHAOS IN SURFACE WAVES

Aifantis, E.C.
PATTERN-FORMING INSTABILITIES DURING PLASTIC FLOW AND FRACTURE

Ait el Ferrane, M.I., Bremand, F., Lagarde, A.
NONLINEAR VISCOELASTIC AND PHOTOVISCOELASTIC CONSTITUTIVE LAWS FOR POLYCARBONATE MATERIAL IN QUASISTATIC LOAD

Akylas, T.R., Grimshaw, R.H.J.
SOLITARY INTERNAL WAVES WITH OSCILLATORY TAILS

Altenbach, H., Zolochevsky, A.
A UNIFIED MODEL OF ISOTROPIC MATERIALS INCLUDING SECOND-ORDER-EFFECTS

Altus, E.: See Adan, M.

Amadio, C., Rajgelj, S., Nappi, A.
AN EXTENDED VARIATIONAL FORMULATION APPLIED TO THE ANALYSIS OF CREEP PROBLEMS

Amari, T., Asakura, T.
CREEP BEHAVIOR OF MAGNETIC SUSPENSION AT THE STRESS LEVEL OF YIELD VALUE

Anh, N.D.
AN APPROXIMATE TECHNIQUE FOR STATIONARY RANDOM VIBRATIONS

Anselmet, F., Djeridi, H., Fulachier, L.
PASSIVE SCALAR DISSIPATION IN FULLY TURBULENT FLOWS

Arbocz, J., de Vries, J., Rebel, G.
KOITER'S STABILITY THEORY EXTENDED TO THE BUCKLING OF ANISOTROPIC SHELLS WITH (NEARLY) SIMULTANEOUS BUCKLING MODES

Arbocz, J.: See Guo, Q.Z.

Arcisz, M., Kosinski, W.
HEAT CONDUCTION LAWS AND HUGONIOT RELATIONS IN FINITE ELASTICITY

Ariaratnam, S.T.: See Asokanthan, S.F.

Arts, T.: See Janssen, J.D.

Asakura, T.: See Amari, T.

Ashida, F., Tauchert, T.R., Noda, N.
A GENERAL SOLUTION TECHNIQUE FOR THREE-DIMENSIONAL PROBLEMS OF PIEZOTHERMOELASTICITY IN HEXAGONAL SOLIDS OF CLASS 6MM

Ashlani, F.: See Hanagud, S.

Asokanthan, S.F., Ariaratnam, S.T.
DYNAMIC STABILITY OF AXIALLY MOVING BANDS UNDER RANDOM TENSION FLUCTUATIONS

Austermann, R., Kurnik, W., Popp, K.
CROSS FLOW-INDUCED INSTABILITIES IN A TUBE ARRAY - EXPERIMENTS, MODELLING, ANALYSIS

Aydin, E.M.: See Leutheusser, H.J.

Bai, Y.L., Lu, C.S., Ling, Z., Xia, M.F., Ke, F.J.
COALESCENCE OF MICROCRACKS AS A CASCADE TO CRITICAL FAILURE

Baker, G.R.: See Cowley, S.J.

Bakker, P.G.
A MATHEMATICAL MODEL FOR 'OPEN' SEPARATION

Bammann, D.J.
MODELING THE FINITE INELASTIC DEFORMATION AND FAILURE OF METALS

Bar-Cohen, Y.: See Mal, A.K.

Barthes-Biesel, D.: See Drochon, A.

Batra, R.C., Zhu, Z.G., Wang, Y.
CONSIDERATION OF PHASE TRANSFORMATIONS AND KINEMATIC HARDENING IN THE STUDY OF SHEAR BANDS IN A DYNAMICALLY LOADED STEEL BLOCK

Bauer, J.: See Gutkowski, W.

Bazant, Z.P., Cedolin, L., Ozbolt, J.
INSTABILITIES AND BIFURCATIONS DUE TO CRACK INTERACTIONS AND DAMAGE

Bazhlekov, I.: See Zapryanov, Z.

Belsky, V.G.
A SOLUTION SCHEME FOR CONTACT PROBLEMS AND THE ANALYSIS OF BODIES WITH CRACKS USING MULTI-GRID APPROACH

Ben-David, D.: See Tirosh, J.

Ben-Haim, Y.
CONVEX MODELS FOR UNCERTAINTY IN RADIAL PULSE BUCKLING OF SHELLS

Benallal, A.
ILL-POSEDNESS AND LOCALIZATION IN SOLIDS AND STRUCTURAL MECHANICS

Benveniste, Y.
UNIVERSAL RELATIONS IN PIEZOELECTRIC COMPOSITES WITH EIGENSTRESS AND POLARIZATION FIELDS

Berdichevsky, V., Fridlyand, A.
VARIATIONAL APPROACH TO FLUID DYNAMICS

Berest, P.
STABILITY OF AN IRREVERSIBLE SYSTEM: DETACHMENT OF A MEMBRANE

Berkooz, G., Elezgaray, J., Holmes, P.
THE LOCAL NATURE OF COHERENT STRUCTURE INTERACTION

Bershadskii, A.: See Branover, H.

Bert, C.W.:See Libai, A.

Berthaud, Y., Billardon, R., Comi, C.
IDENTIFICATION OF DAMAGEABLE ELASTIC-PLASTIC MATERIALS FROM LOCALIZATION MEASUREMENTS

Beylinson, Ya.Z., Goldberg, A.Z., Goldberg Z.A.
NONLINEAR WAVES AND FERMI-PASTA-ULAM PROBLEM

Bigoni, D., Zaccaria, D.
ON FLUTTER INSTABILITY IN CONTINUUM ELASTOPLASTICITY

Billardon, R.: See Berthaud, Y.

Birker, T.: See Zhou, M.

Blajer, W., Schiehlen, W., Schirm, W.
SOME NEW DEVELOPMENTS IN THE COORDINATE PARTITIONING METHOD

Blanze, C., Danwe, R., Ladeveze, P., Maurel, Ph.
A NEW SIMPLIFIED METHOD FOR THE ANALYSIS OF MASSIVE-BODY STRUCTURES

Blau, M., Moshaiov, A.
THREE DIMENSIONAL MODELING AND ANALYSIS OF A SINGLE LAP JOINT

Blekhman, I.I.
THE VIBRATIONAL MECHANICS AND VIBRATIONAL RHEOLOGY: MAIN THEOREMS, BODY OF MATHEMATICS, NONLINEAR EFFECTS, APPLICATIONS

Bleuler, H., Blöchlinger, M., Dual, J., Gähler, C.
MEASUREMENT OF MATERIAL DAMPING IN FERROMAGNETIC METALS USING ACTIVE MAGNETIC BEARINGS

Blöchlinger, M.: See Bleuler, H.

Bodner, S.R.: See Ravid, M.

Boehler, J.-P.: See Millien, A.

Bonnecaze, R.T.: See Kraynik, A.M.

Bontcheva, N.
PLASTIC LOCALIZATION IN ANISOTROPICALLY HARDENING DAMAGED MATERIALS

Borgas, M.S.
STOCHASTIC MODELS OF LAGRANGIAN STATISTICS IN TURBULENT FLOW

Borodich, F.M.
SELF-SIMILARITY PROPERTIES OF STATIC AND DYNAMIC CONTACT BETWEEN A PUNCH AND ANISOTROPIC MEDIA

Boucher, S.: See Verlinden, O.

Boulahbal, D.: See Crandall, S.H.

Boulanger, Ph., Hayes, M.
SHEAR, SHEAR STRESS AND SHEARING

Bourassa, P.
DYNAMICS OF A MULTIBODY SYSTEM WITH CONTROLS APPLICATION TO BIOMECHANICS MOTION

Bovendeerd, P.H.M.: See Janssen, J.D.

Braat, G.F.M.
MECHANICS OF CONTACT BETWEEN TWO ROLLERS COATED WITH MULTI-LAYERED, VISCOELASTIC RUBBER

Brady, J.F.: See Kraynik, A.M.

Brandt, R.
RELAMINARISED BOUNDARY LAYERS IN ADVERSE PRESSURE GRADIENTS

Branover, H., Bershadskii, A., Eidelman, A., Nagorny, M.
THE LABORATORY SIMULATION OF TRANSFER PROCESSES AND FRACTAL VARIABILITY OF STRATOSPHERIC TURBULENCE

Bratukhin, Y.K., Makarov, S.O.
ON THE THERMOCAPILLARY MOVEMENTS OF THE SOLITONIC TYPE

Bremand, F.: See Ait-el-Ferrane, M.I.

Brenner, H.: See Abbott, J.R.

Brenner, H.: See Frankel, I.

Brenner, H.: See Mauri, R.

Brewer, D.: See Calomino, A.

Brock, L.M.
THERMAL EFFECTS IN TRANSIENT MODELS OF DYNAMIC BRITTLE FRACTURE

Broday, D., Kimmel, E.
ANALYSIS OF THE ENHANCED AXIAL DISPERSION IN A VIBRATING TUBE

Brodsky, N.S., Fossum, A.F, Chan, K.S., Munson, D.E.
EXPERIMENTAL EVALUATION OF A CONSTITUTIVE MODEL FOR INELASTIC FLOW AND DAMAGE EVOLUTION IN SOLIDS SUBJECTED TO TRIAXIAL COMPRESSION LOADING

Browne, L.W.B.: See Dinckelacker, A.

Budiansky, B., Fleck, N.A.
COMPRESSIVE FAILURE OF FIBRE COMPOSITES

Bühler, K.
VORTEX BREAKDOWN PHENOMENA IN ROTATING FLUIDS

Bui, H.D.: See Rittel, D.

Burde, G.I.
SOME EXACT NON-SIMILARITY SOLUTIONS OF THE BOUNDARY LAYER EQUATIONS

Byskov, E.
FINITE ELEMENT PRINCIPLES, ASYMPTOTIC POSTBUCKLING PROBLEMS AND NONLINEAR MEMBRANE LOCKING

Calomino, A., Brewer, D., Chudnovsky, A.
SUBCRITICAL AND CRITICAL CRACK GROWTH IN NC-132 MONOLITHIC SILICON NITRIDE

Camotim, D., Roorda, J.
PLASTIC BUCKLING WITH RESIDUAL STRESSES - IMPERFECTION SENSITIVITY

Carpenter, P.W.: See Lucey, A.D.

Caulliez, G.: See Ricci, N.

Cederbaum, G., Mond, M., Kahn, P.B., Zarmi, Y.
STABILITY ANALYSIS OF THE NON-LINEAR MATHIEU EQUATION

Cedolin, L.: See Bazant, Z.

Cercignani, C.
ON THE INVERTED TEMPERATURE GRADIENT IN A VAPOR

Chan, K.S.: See Brodsky, N.S.

Chang, C.-C.
ON THE SOURCES OF AERODYNAMIC FORCES

Chang, M.H.: See Chen, F.

Chapuliot, S., Combescure, A.
PLASTIC BUCKLING ANALYSIS OF AXIALLY STIFFENED SHELLS

Charru, F.
BIFURCATION ANALYSIS OF INTER-FACIAL WAVES IN A TWO-LAYER COUETTE-POISEUILLE FLOW

Chashechkin, J.D.
INTERNAL WAVES AND VORTICES IN THE WAKES IN A CONTINUOUSLY STRATIFIED LIQUID

Chaskalovic, J.
A NEW MATHEMATICAL MODEL FOR THE TORNADOES GENESIS BY BIFURCATION OF THE SOLUTIONS OF NAVIER-STOKES EQUATIONS

Chau, L.-K.: See Stumpf, H.

Chaudhuri, R.A., Garala, H.J.
ON DEFORMATION AND FAILURE BEHAVIOR OF THICK ADVANCED COMPOSITES UNDER COMPRESSION

Chen, C., Han, M., Liu, Z., Su, X., Wang, R.
STRESS WAVE EFFECT ON THE DYNAMIC PLASTIC BUCKLING OF CYLINDRICAL SHELL UNDER AXIAL IMPACT

Chen, C.-T.: See Su, F.-C.

Chen, F., Chang, M.H.
STABILITY OF TAYLOR-DEAN FLOW IN A SMALL GAP BETWEEN ROTATING CYLINDERS

Chen, X.-N., Sharma, S.D.
SLENDER SHIP MOVING AT NEAR-CRITICAL SPEED IN SHALLOW CHANNEL

Chen,. X-N., Wei, R-J.
DYNAMICAL BEHAVIOUR OF A NONPROPAGATING SOLITON UNDER A PERIODICALLY MODULATED OSCILLATION

Cherepanov, G.P.
DYNAMICS OF AN ELEMENTARY ACT OF A CRACK GROWTH AND DISLOCATION EMISSION AT THE INSTABILITY POINT

Cherkaev, A.
A VARIATIONAL METHOD FOR ESTIMATING EFFECTIVE PROPERTIES OF COMPOSITES WITH ARBITRARY MICROSTRUCTURES

Chien, N.: See Honein, T.

Chinilov, A. Y.
ON A BOUNDARY VALUE PROBLEM IN THE STABILITY THEORY OF A SHOCK WAVE

Chiskis, A.: See Freidin, A.

Chiu, C.-H.: See Gao, H.

Christensen, R.M., DeTeresa, S.J.
ELIMINATION/MINIMIZATION OF EDGE-INDUCED STRESS SINGULARITIES IN FIBER COMPOSITE LAMINATES

Chu, C.-C., Wang, C.-T., Hsieh, C.-S.
AN EXPERIMENTAL INVESTIGATION OF VORTEX MOTIONS NEAR SURFACES

Chudnova, R.S.: See Smirnov, B.I.

Chudnovsky, A.: See Calomino, A.

Chudnovsky, A.: See Kunin, B.

Chudnovsky, A.: See Wu, S.

Chung, J.: See Hulbert, G.M.

Churilov, S.M., Shukhman, I.G.
ON A WEAKLY-NONLINEAR INTERPRETATION OF MODE ALTERNATION IN A CIRCULAR SHEAR FLOW

Churilov, S.M., Shukhman, I.G.
CRITICAL LAYER AND NONLINEAR EVOLUTION OF DISTURBANCES IN WEAKLY SUPERCRITICAL SHEAR FLOWS

Clifton, R.J., Prakash, V.
TIME-RESOLVED FRICTION IN PRESSURE-SHEAR IMPACT

Cochelin, B., Damil, N., Potier-Ferry, M.
ASYMPTOTIC-NUMERICAL METHODS FOR NONLINEAR STRUCTURAL MECHANICS

Coene, R.
THE DRIFT PAST AN AIRFOIL WITH CIRCULATION

Cohen, B., Lai, W.M., Mow, V.C.
THE ANISOTROPY INDUCED NON-LINEARITY OF THE GROWTH PLATE BEHAVIOR - A FINITE DEFORMATION ANALYSIS

Cohen, H., Mac Sithigh, R.G.
ON THE STABILITY OF SPINNING ELASTIC PSEUDO-RIGID BODIES

Cohen, H., Mac Sithigh, G.P.
THE PSEUDO-RIGID MODEL: APPLICATIONS TO IMPACT PROBLEMS

Collins, I.F.
INSTABILITIES IN CRITICAL STATE SOIL MECHANICS

Combescure, A.: See Chapuliot, S.

Comi, C.: See Berthaud, Y.

Conti, C.: See Verlinden, O.

Cowley, S.J., Baker, G.R., Tanveer, S., Page, M.
SINGULARITY FORMATION AND ROLL-UP ON VORTEX SHEETS

Crandall, S.H., Boulahbal, D.
DYNAMIC INSTABILITIES OF ROTATING DISK INFORMATION STORAGE DEVICES

Crisfield, M.: See Shi, J.

Cristescu, N.
A PROCEDURE TO DETERMINE NONASSOCIATED CONSTITUTIVE EQUATIONS FOR GEOMATERIALS

Crocombe, A.D.: See Dabnichki, P.

Cuche, D.: See Mayer, M.

Curry, F.-R.: See Tsay, R.-V.

Cveticanin, L.
THE INFLUENCE OF THE REACTIVE FORCE ON THE VIBRATION PROPERTIES OF THE ROTOR WITH TIME VARIABLE MASS

Dabnichki, P., Crocombe, A.D.
MODELLING AND FINITE ELEMENT ANALYSES OF BUTTOCK-CUSHION INTERACTION

Damil, N.: See Cochelin, B.

Danwe, R.: See Blanze, C.

Davies, H.G., Pan, R.
PERIODIC MOTION THROUGH A BIFURCATION

de Boer, R.
COMPRESSIBLE POROUS MEDIA - TOWARDS A GENERAL THEORY

de Borst, R.: See Mülhaus, H.-B.

de Borst, R.: See Sluys, L.J.

DeBotton, G., Ponte Castaneda, P.
THE ELASTIC/PLASTIC BEHAVIOR OF FIBER COMPOSITES

de Bremaecker, J.C.: See Wei, K.

de Kraker, A.: See Van Campen, D.H.

Dehombreux, P.: See Verlinden, O.

Demetriades, A.: See Liang, F.P.

De Teresa, S.J.: See Christensen, R.M.

de Vries, J.: See Arbocz, J.

De-Wachter, D.: See Verdonck, P.

Dias, F., Iooss, G.
CAPILLARY-GRAVITY SOLITARY WAVES WITH DAMPED OSCILLATIONS

Dickstein, P.A., Ward, C.A.
PREDICTION OF THE SHAPE OF ROTATING DROPLETS IN TERMS OF EXPERIMENTALLY CONTROLLABLE VARIABLES

Dimas, A.A., Triantafyllou, G.G.
NON-LINEAR INTERACTION OF SHEAR FLOW WITH A FREE SURFACE

Dinkelacker, A., Kowalski, T., Browne, L.W.B.
ON THE CORRELATION BETWEEN TURBULENT VELOCITY AND WALL PRESSURE FLUCTUATIONS

Djeridi, H.: See Anselmet, F.

Dolinski, K.
ON STOCHASTIC MODELING OF FATIGUE CRACK GROWTH

Dost, S., Erbay, S., Erbay, H.A.
EFFECT OF GRAVITY ON THE DISSOLUTION AND LIQUID PHASE EPITAXIAL GROWTH PROCESS OF SILICON IN A HORIZONTAL SANDWICH SYSTEM

Drochon, A., Leyrat-Maurin, A., Barthes-Biesel, D.
MOTION OF A CAPSULE THROUGH A CONSTRICTION: A MODEL FOR RED BLOOD CELL FILTRATION

Drozdov, A.D., Kolmanovskii, V.B.
STABILITY OF NONHOMOGENEOUS AGING ELASTIC AND VISCOELASTIC BODIES

Drucker, D.C., Li, M.
EXPLORING INSTABILITY RESULTING FROM A NON-ASSOCIATED FLOW RULE

Dual, J.: See Bleuler, H.

Dundurs, J.: See Markenscoff, X.

Dunne, F.P.E., Hayhurst, D.R.
VISCOPLASTIC DAMAGE CONSTITUTIVE EQUATIONS FOR CREEP AND CYCLIC PLASTICITY

Dupre, J.C., Lagarde, A.
OPTICAL METHOD OF THE MEASUREMENT AT DISTANCE OF LOCAL STRAINS

Durban, D., Stronge, W.J.
DIFFUSION OF INSTANTANEOUS LOADS IN PRESTRAINED BARS

Duszek-Perzyna, M.K.
ADIABATIC SHEAR BAND LOCALIZATION IN DYNAMIC LOADING PROCESSES OF ELASTIC-PLASTIC CRYSTALS

Dvorak, G.J.
TRANSFORMATION FIELD ANALYSIS OF INELASTIC COMPOSITE MATERIALS

Dvorkin, J., Mavko, G., Nur, A.
THE THEORY OF CEMENTED GRANULAR MATERIALS

Dziedzic, M., Leutheusser, H.J.
EVOLUTION OF VORTEX RINGS GENERATED AT AN ORIFICE

Dzodzo, M.B.
LAMINAR NATURAL CONVECTION IN ROMB SHAPED ENCLOSURES

Earles, S.W.E.: See Seneviratne, L.D.

Eckelmann, H.: See König, M.

Eckelmann, H.: See Noack, B.R.

Eckelmann, H.: See Ohle, F.

Eidelman, A.: See Branover, H.

Einav, S.: See Ginzberg, I.

Elad, D.: See Ginzberg, I.

Elezgaray, J.: See Berkooz, G.

Elishakoff, I.
OPTIMIZATION AND ANTI-OPTIMIZATION OF STRUCTURES DUE TO UNCERTAINTY: PROBABILITY OR CONVEXITY OR BOTH?

Erbay, H.A.: See Dost, S.

Erbay, S.: See Dost, S.

Eschenauer, H.A., Kobelev, V.V., Schumacher, A.
BUBBLE METHOD OF TOPOLOGY STRUCTURAL OPTIMIZATION

Essén, H.
PROJECTING NEWTON'S EQUATIONS ONTO NON-COORDINATE TANGENT VECTORS OF THE CONFIGURATION SPACE; A NEW LOOK AT LAGRANGE'S EQUATIONS IN TERMS OF QUASI-COORDINATES

Etay, J.
STABILISATION OF A SURFACE GRAVITY WAVE BY A VERTICAL UNIFORM MAGNETIC FIELD: ANALYSIS AND EXPERIMENTS

Etay, J., Fautrelle, Y.
DANCE OF SPHERES

Fabrikant, A.L., Stepanyants, Y..A.
RADIATIVE FORCES ACTING ON MOVING SOURCES IN ANISOTROPIC MEDIA

Faciu, C.
INITIATION AND GROWTH OF STRAIN BANDS IN RATE-TYPE VISCOELASTIC MATERIALS

Falkovich. A.I.: See Khain, A.P.

Fan, X.: See Acrivos, A.

Faria, L.O.: See Martins, J.A.C.

Fautrelle, Y.: See Etay, J.

Fenner, D.N.: See Seneviratne, L.P.

Fernando, H.J.S, McGrath, J., Perera, M.A.P.
TURBULENT MIXING ACROSS DENSITY INTERFACES: SOME NEW CONCEPTS AND RESULTS

Ferron, G., Gardin, C., Grandidier, J-C., Potier-Ferry, M.
COMPRESSIVE BEHAVIOR OF COMPOSITE MATERIALS: A THEORETICAL AND EXPERIMENTAL STUDY

Fey, R.H.B.: See Van Campen, D.H.

Fiebig, M.: See Güntermann, T.

Fish, J.
MICROMECHANICAL STUDIES OF COMPOSITE MATERIALS WITH THE S-VERSION OF THE FINITE ELEMENT METHOD

Fiszdon, W.: See Stamm, G.

Fleck, N.A.: See Budiansky, B.

Fleck, N.A.: See Sutcliffe, M.P.F.

Floryan, J.M., Inculet, I.I.
INDUCED POLARIZATION EFFECTS IN LIQUID DROPLETS

Foerster, A.: See Kuhn, G.

Fossum, A.B.: See Brodsky, N.S.

Frankel, I., Brenner, H.
TRANSPORT OF ORIENTABLE BROWNIAN PARTICLES IN HOMOGENEOUS SHEAR FLOWS - A *NEGATIVE* DISPERSION?

Franzese, P.: See Zannetti, L.

Freidin, A., Chiskis, A.
THE PHASE TRANSITION ZONE IN NONLINEAR ELASTIC ISOTROPIC MEDIA

Freidin, A., Vaynblat, D.
HETEROGENEOUS DEFORMATION OF SOLIDS DUE TO APPEARANCE OF NEW PHASE NUCLEI AND LAYERS

Fressengeas, C.: See Jeanclaude, V.

Fridlyand, A.: See Berdichevsky, V.

Frolov, I.V., Vakulenko, A.A.
ON THE CONNECTION BETWEEN MICRO- AND MACRO-PROPERTIES IN THE ELASTO-PLASTIC MEDIA

Fukuyu, A., Ruzi, T.
THE RESPONSE OF HILL'S VORTEX TO A THREE-DIMENSIONAL DISTURBANCE

Fulachier, L.: See Anselmet, F.

Furta, S.D.: See Kozlov, V.V.

Gähler, C.: See Bleuler, H.

Galper, A., Miloh, T.
HYDRODYNAMICAL SELF-PROPULSION OF OSCILLATING DEFORMABLE SMOOTH SURFACE

Gandelsman, M.I.
REFLECTION OF PLANE ELASTIC WAVE FROM THE BOUNDARY OF FIBER-REINFORCED COMPOSITE MATERIAL

Gao, H., Chiu, C.H.
STRESS DRIVEN INTERFACIAL INSTABILITIES

Gao, Q.: See Zeng, P.

Garala, H.J.: See Chaudhuri, R.A.

Gardin, C.: See Ferron, G.

Gawinecki, J.A.
GLOBAL SOLUTIONS TO THE INITIAL VALUE PROBLEMS IN NONLINEAR HYPERBOLIC THERMOELASTICITY

Geller, A.S.: See Mondy, L.A.

Gendelman, O.V., Ginzburg, V.V., Manevitch, L.I.
INSTABILITIES CAUSED BY THE PRESENCE OF STRUCTURAL DEFECTS IN GLASSES AND POLYMER CRYSTALS

Georgantopoulos, G.A.: See Lekas, T.I.

George, J., Grisenti, M., Minel, F.
INTERFACIAL AGITATION PARAMETERS INFLUENCING G[illegible]LIQUID MASS TRANSFER

Georgiou, E., Papageorgiou, D., Maldarelli, C., Rumschitzki, D.
THE ASYMPTOTIC LINEAR AND NONLINEAR STABILITY OF CORE-ANNULAR FILM FLOWS IN THE THIN ANNULAR LIMIT

Gershtein, M.
NONLINEAR DYNAMICS OF MULTI-LAYERED SHELLS

Giat, Y.: See Mizrachi, J.

Ginis, I.D.: See Khain, A.P.

Ginzberg, I., Elad, D., Einav, S.
ANALOG MODEL OF THE BRONCHIAL TREE

Ginzburg, V.V. See Gendelman, O.V.

Givoli, D., Rivkin, L., Keller, J.B.
A NEW FINITE ELEMENT METHOD FOR FRACTURE PROBLEMS

Glenn, L.A.
ENERGY-DENSITY EFFECTS ON SEISMIC DECOUPLING

Glozman, M.: See Agnon, Y.

Goddard, J.D.
NEW THEORETICAL METHODS FOR REYNOLDS DILATANCY IN GRANULAR MATERIALS

Goldberg, A.Z. See Beylinson, Ya.Z.

Goldberg, Z.A.: See Beylinson, Ya.Z.

Golden, J.M., Graham, G.A.C.
THE BEHAVIOUR OF CRACKS IN VISCOELASTIC MEDIA SUBJECT TO A BENDING MOMENT

Goldenveizer, A.L., Kaplunov, J.D., Nolde, E.V.
ASYMPTOTIC ANALYSIS OF SHEAR DEFORMATION SHELL THEORIES

Goldshtein, A., Shapiro, M.
A MODEL FOR COLLISIONAL MOTION OF GRANULAR MATERIALS

Goldshtik, M., Husain, H.S., Hussain, F.
PARADOXICAL SPECIES SEPARATION PHENOMENON

Goldstein, R., Jitnikov, J., Spector, A.
EVOLUTION PROBLEMS OF CONTACT INTERACTION MECHANICS WITH FRICTION

Golos, K., Osinski, Z.
FATIGUE CRACK GROWTH RATE MODEL IN TERMS OF ENERGY

Golovan, A.A., Matasov, A.I.
THE LEAST-SQUARES METHOD FOR GUARANTEED ESTIMATION IN MECHANICAL SYSTEMS

Gorelik, M.: See Kunin, B.

Grabacki, J.
CONSTITUTIVE EQUATIONS FOR SOME DAMAGING MATERIALS

Graham, A.L.: See Abbott, J.R.

Graham, G.A.C.: See Golden, J.M.

Grandidier, J.C.: See Ferron, G.

Greenberg, J.B., Stavsky, Y., Sabag, M.
BUCKLING OF EDGE-DAMAGED COMPOSITE CYLINDRICAL SHELLS SUBJECTED TO RADIAL PRESSURE

Grimshaw, R.H.J.: See Akylas, T.R.

Gringauz, M.
THE FRICTION CONTACT PROBLEMS OF FRACTURE MECHANICS

Grisenti, M.: See George, J.

Gross, D., Zhang, C.
A COHESIVE DAMAGE ZONE MODEL FOR STABLE CRACK GROWTH IN DUCTILE MATERIALS

Gross, T.S.: See Mendelsohn, D.A.

Guest, S.D., Pellegrino, S.
PROPAGATION OF DESTABILISING WAVES IN THE FOLDING OF FACETED TUBES

Guimaraes, J.: See Martins, J.A.C.

Gulic, I., Ostrach, S., Kamotani, Y.
THERMAL CONVECTION IN A SHALLOW ROTATING ANNULUS

Güntermann, T., Fiebig, M., Mitra, N.K.
OSCILLATING LONGITUDINAL VORTEX GENERATION IN LAMINAR CHANNEL FLOW

Guo, Q.Z., Arbocz, J.
BUCKLING OF ANISOTROPIC CONICAL SHELLS

Guo, Z-H.
π-METHOD IN FINITE DEFORMATIONS

Gurevich, B., Lopatnikov, S.L., Sadovnichaja, A.P.
ELASTIC WAVES IN NON-HOMOGENEOUS POROUS MATERIALS THEORY AND NUMERICAL EXPERIMENTS

Gürtl, R.: See Perktold, K.

Gutkowski, W., Bauer, J.
STRUCTURAL OPTIMIZATION WITH SENSITIVITY CONSTRAINTS

Guz, A.N.
ELASTIC WAVES IN COMPRESSIBLE MATERIALS WITH INITIAL STRESSES AND NON-DESTRUCTIVE ULTRASONIC METHOD OF DETERMINATION OF TWO-AXIAL RESIDUAL STRESSES

Haddow, J.B., Janeie, P., Mioduchowski, A.
SOLUTION OF SOME NONLINEAR ELASTODYNAMIC PROBLEMS

Hadouaj, H.: See Maugin, G.A.

Hagedorn, P., Kazimierczyk, P.
ON THE OPTIMAL PLACEMENT OF SPACERS IN BUNDLED CONDUCTORS

Hagedorn, P.: See Meirovitch, L.

Han, M.: See Chen, C.

Han, R.P.S.: See Luo, C.

Hanagud, S., Ashlani, F.
ROUTES TO CHAOS IN STRUCTURAL DYNAMIC SYSTEMS

Handler, R.A., Levich, E. Sirovich, L.
DRAG REDUCTION IN TURBULENT CHANNEL FLOW BY PHASE RANDOMIZATION

Hansson, T.: See Neumeister, J.M.

Hardy, S.J., Malik, N.H.
MODEL STUDIES OF NUMERICAL DERIVATIVES FOR GRADIENT-BASED OPTIMIZATION TECHNIQUES

Hashin, Z.
EFFECT OF MATRIX NON-LINEARITY AND IMPERFECT INTERFACE ON COMPRESSIVE STRENGTH OF FIBER COMPOSITES

Hasimoto, H.
CREEPING SHEAR FLOW ALONG A HOLE IN A WALL

Haughton, D.M.
CIRCULAR SHEARING OF COMPRESSIBLE ELASTIC CYLINDERS

Haung, W.: See Wu, S.

Hayes, M.: See Boulanger, Ph.

Hayhurst, D.R.: See Dunne, F.P.E.

Haythornthwaite, R.M.
ISOTROPIC HARDENING BASED ON SIMPLE SHEAR

Hazanov, S.
NEW MINIMUM THEOREMS FOR HEREDITARY ELASTICITY

Healey, J.J.
NONLINEARITY IN BOUNDARY LAYER TRANSITION

Hedrih-Stefanovic, K.
ON SOME INTERPRETATIONS OF THE RIGID BODIES KINETIC PARAMETERS

Herrmann, G.: See Honein, T.

Herve, E.: See Thebaud, F.

Hetnarski, R.B.: See Tylikowski, A.

Hiller, W.J.: See Kowalewski, T.A.

Hishinuma, E.: See Kondo, K.

Holcman, I.: See Ravid, M.

Holmes, P., Swart, P.
THE DYNAMICS OF PHASE TRANSFORMATION IN AN ANTI-PLANE SHARE PROBLEM

Holmes, P.: See Berkooz, G.

Honein, T., Chien, N., Herrmann, G.
A NOVEL METHOD TO ESTABLISH CONSERVATION LAWS FOR DISSIPATIVE SYSTEMS

Hong, J.: See Liu, Y.

Horacek, J.
ANNULAR-FLOW-INDUCED INSTABILITY OF CYLINDRICAL SHELLS

Horii, H., Okui, Y.
A MICROMECHANICS-BASED CONTINUUM THEORY FOR LOCALIZATION PHENOMENA

Hsieh, C.-S.: See Chung, C.-C.

Hui, W.H., Lepage, C.Y.
DISCONTINUITY-ADAPTIVE GODUNOV SCHEME

Hulbert, G.M., Chung, J.
A TIME INTEGRATION ALGORITHM FOR STRUCTURAL DYNAMICS WITH IMPROVED NUMERICAL DISSIPATION: THE GENERALIZED-α METHOD

Husain, H.S.: See Goldshtik, M.

Hussain, F.: See Goldshtik, M.

Hussain, F.: See Shtern, V.

Huyghe, J.M., van Gemert, R., Janssen, J.D.
VISCOELASTIC REMODELLING OF PASSIVE MYOCARDIAL TISSUE: A REAPPRAISAL OF ISOTROPY

Huyghe, J.M.: See Janssen, J.D.

Hwang, K-Ch., Zhang, L.
HIGHER-ORDER ASYMPTOTIC SOLUTION FOR MODE-I CRACK GROWING IN ELASTIC PERFECTLY-PLASTIC MATERIAL

Ido, Y.: See Kiya, M.

Inculet, I.I: See Floryan, J.M.

Ingber, M.S.: See Mondy, L.A.

Iooss, G.: See Dias, F.

Iskenderov, I.A.
CIRCULAR LAW OF FULL MECHANICAL ENERGY CONSERVATION

Ivanov, V.V., Lopatnikov, S.L., Rok, V.E.
ON POSSIBLE USE OF HYDROACOUSTIC FIELD FROM AN UNDERWATER EARTHQUAKE FOR TSUNAMI WARNING

Ivanov, V.
KORTEWEG-DE VRIES EQUATION AND INTERNAL WAVES IN THE MEDITERRANEAN SEA

Janele, P.: See Haddow, J.B.

Janssen, J.D., van Campen, D.H., Bovendeerd, P.H.M., Arts, T., Huyghe, J.M.
THE INFLUENCE OF MUSCLE FIBER ORIENTATION ON LOCAL LEFT VENTRICULAR WALL MECHANICS

Janssen, J.D.: See Huyghe, J.M.

Jasiuk, I.: See Jun, S.

Jeanclaude, V., Fressengeas, C.
SPATIAL ORGANIZATION IN THE PORTEVIN-LE CHATELIER RELAXATION OSCILLATIONS

Jekot, T.
THE ANISOTROPIC NONLINEAR THERMOELASTIC MODEL OF SI-SIO STRUCTURE

Jensen, H.M.
DELAMINATION SHAPE INSTABILITIES IN COATED MATERIALS

Jeong, J.-T.: See Moffatt, H.K.

Jitnikov, J.: See Goldstein, R.

Jullien, J.-F., Limam, A., Sanchez-Sanchez, H.
TOWARDS A BETTER ESTIMATE OF INSTABILITY OF CYLINDRICAL SHELL STRUCTURES UNDER A COMPLEX LOADING

Jun, S., Jasiuk, I.
ELASTIC MODULI OF 2-D COMPOSITES WITH SLIDING INCLUSIONS - A CRITICAL EVALUATION OF THE EFFECTIVE MEDIUM THEORIES

Kachanov, M.
SOLIDS WITH MANY CRACKS: STATISTICAL STABILITY OF EFFECTIVE MODULI AND INSTABILITY OF FRACTURE-RELATED PARAMETERS

Kahn, P.B.: See Cederbaum, G.

Kalamkarov, A.L., Kudryavtsev, B.A., Laurent-Gengoux, P.
GENERALIZED INTEGRAL TRANSFORMS IN MECHANICS OF PERIODIC COMPOSITES

Kamotani, Y.: See Gulic, I.

Kanaun, S.K., Kudrjavtseva, L.T., Levin, V.M.
EFFECTIVE FIELD METHOD IN MECHANICS OF COMPOSITE MATERIALS

Kaplunov, J.D.: See Goldenveizer, A.L.

Kapoor, B., Acrivos, A.
FLOW OF A SEDIMENT LAYER ON AN INCLINED PLANE

Karagiozov, V., Karagiozova, D., Manoach, E.
CHAOS-LIKE PHENOMENA IN THE DYNAMIC STABILITY OF ELASTIC-PLASTIC BEAMS UNDER PULSE LOADING

Karagiozova, D.: See Karagiozov, V.

Kaya, M.O.: See Kaykayoglu, C.R.

Kaykayoglu, C.R., Kaya, M.O.
SIMULATION OF SECONDARY VORTEX FORMATION DURING KARMAN VORTEX STREET-LEADING EDGE INTERACTION: ASSESSMENT OF THE LEADING EDGE KUTTA CONDITION

Kazimierczyk, P.: See Hagedorn, P.

Ke, F.J.: See Bai, Y.L.

Keer, L.M.: See Yu, M.

Keller, J.B.: See Givoli, D.

Khain, A.P., Falkovich, A.I., Ginis, I.D.
NUMERICAL MODELING OF THE INFLUENCE OF AIR-SEA INTERACTION ON THE DEVELOPMENT AND MOTION OF TROPICAL CYCLONES

Kimmel, E.: See Broday, D.

Kit, E.: See Tsitverblit, N.

Kitagawa, H., Nakatani, A.
CRACK TIP FIELD IN CRYSTALLINE MATERIAL-COUNTERPART IN ATOMISTIC SIMULATION OF SINGULARITY OBTAINED BY CRYSTAL PLASTICITY MODEL

Kiya, M., Mochizuki, O., Shimizu, M., Ogura, Y., Ido, Y.
ACTIVE CONTROL OF A TURBULENT SEPARATED FLOW ZONE

Klarbring, A.: See Sjödin, B.

Kleiber, M., Kowalczyk, P.
STRESSES IN HUMAN LUNG ACCOUNTING FOR TISSUE-GAS INTERACTION: MODELING AND NUMERICAL SIMULATION

Klepaczko, J.R.
ADIABATIC SHEAR BANDS - NEW EXPERIMENTAL TECHNIQUES AND RESULTS

Klimes, F., Korenar, J., Metelka, T.
LDA MEASUREMENTS OF PULSATILE VELOCITY FIELD BEHIND PROSTHETIC HEART VALVE

Klimov, D.M.
LIE GROUPS AND COMPUTER ALGEBRA IN PROBLEMS OF MECHANICS

Kluwick, A.
TRANSONIC NOZZLE FLOW OF DENSE GASES

Knauss, W.G., Minahen, T.M.
TIME DEPENDENT BUCKLING OF VISCOELASTIC STRUCTURES

Knowles, J.K.: See Abeyaratne, R.

Kobayashi, S., Shizawa, K., Takahashi, K.
CONTINUUM MECHANICS UNDER NONEQUILIBRIUM PROCESSES BASED ON A CONCEPT OF "THERMOPOLAR MATERIALS"

Kobelev, V.V.: See Eschenauer, H.A.

Kocks, U.F.
KINETIC PLASTIC INSTABILITIES

Kohn, R.V.
THE INFLUENCE OF SURFACE ENERGY ON THE MICROSTRUCTURE OF AN EXTREMAL ELASTIC COMPOSITE

Kolmanovskii, V.B.: See Drozdov, A.D.

Kondo, K., Hishinuma, E.
TRAVELLING HINGE SOLUTION FOR LARGE DEFORMATIONS OF RIGID-PLASTIC CIRCULAR ARCHES

König, M., Eckelmann, H.
AN EXPERIMENTAL STUDY OF THE THREE-DIMENSIONAL STRUCTURE OF THE WAKE OF WALL-BOUNDED CIRCULAR CYLINDERS

Koras, A.: See Lekas, T.I.

Korenar, J.: See Klimes, F.

Kosinski, W.: See Arcisz, M.

Kouhia, R.J.: See Menken, C.M.

Kowalczyk, P.: See Kleiber, M.

Kowalewski, T.A., Hiller, W.J.
TRANSIENT EFFECTS BY EVAPORATION FROM OSCILLATING DROPS

Kozlov, V.V., Furta, S.D.
ON ASYMPTOTIC SOLUTIONS OF EQUATIONS OF MOTION IN MECHANICAL SYSTEMS

Krajcinovic, D. , Mallick, K.
MATERIAL INSTABILITIES DURING PHASE TRANSITION IN THERMOSETS

Krasovskij, V.: See Manevich, A.

Kraynik, A.M., Bonnecaze, R.T., Brady, J.F.
THE EFFECT OF SUSPENSION STRUCTURE ON THE STRESS IN ELECTRORHEOLOGICAL FLUIDS

Kudrjavtseva, L.T.: See Kanaun, S.K.

Kudryavtsev, B.A.: See Kalamkarov, A.L.

Kuhn, G. Foerster, A.
A BOUNDARY-ELEMENT FORMULATION FOR FINITE STRAIN ELASTO-PLASTICITY

Kumagai, T., Muraoko, M.
NEWLY-RECOGNIZED MOTIONS OF SEVERAL NUMBERS OF EQUAL SPHERES IN FLUID AT LOW REYNOLDS NUMBERS

Kunin, B., Gorelik M., Chudnovsky, A.
STATISTICAL ORIGIN OF THE DIFFERENCE BETWEEN GRIFFITH ENERGIES OF CRACK INITIATION AND ARREST

Kurnik, W.: See Austermann, R.

Labisch, F.K.
ON SHEARABLE NONLINEARLY ELASTIC RINGS

Ladeveze, P., Reynier, M.
CONTROL OF COMPLEX STRUCTURE MODELS USING VIBRATION TESTS

Ladeveze, P.: See Blanze, C.

Lagarde, A.L: See Ait-el-Ferrane, M.I.

Lagarde, A.: See Dupre, J.C.

Lai, W.M.: See Cohen, B.

Lanir, Y.: See Nevo, E.

Laptev, A., Vassiliev, D.
MAGNETOELASTICITY OF THIN SUPERCONDUCTING SHELLS

Larsen, P.S.: See Nielsen, N.F.

Laurent-Gengoux, P.: See Kalamkarov, A.L.

Le Baron, L., Pedley, T.J.
BI-FLAGELLATE LOCOMOTION IN A SHEAR FLOW

Lekas, T.I., Koras, A., Georgantopoulos, G.A.
HYDROMAGNETIC FLOW THROUGH A THREE-DIMENSIONAL DUCT

Lenz, J.: See Thies, M.

Lepage, C.Y.: See Hui, W.H.

Leroy, Y.M., Molinari, A.
SPATIAL PATTERNING IN SHEAR BANDS

Leutheusser, H.J., Aydin, E.M., McLaren, D.G.
EXPERIMENTS ON COUETTE FLOW BETWEEN ROUGH SURFACES

Leutheusser, H.J.: See Dziedzic, M.

Levich, E.: See Handler, R.A.

Levich, E.: See Tur, A.V.

Levin, V.M.: See Kanaun, S.K.

Levitas, V.I.
THE POSTULATE OF REALIZABILITY: FORMULATION AND APPLICATION TO THERMOMECHANICAL SYSTEMS

Levy, M.: See Mizrahi, J.

Leyrat-Maurin, A.: See Drochon, A.

Li, M.: See Drucker, D.

Li, Y.N.
ON A STABILITY THEORY OF FRACTURE MECHANICS

Liang, F.-P., Reshotko, E., Demetriades, A.
STABILITY OF DEVELOPING FREE SHEAR FLOWS FORMED BY TWO SUPERSONIC LAMINAR STREAMS

Libai, A., Bert, C.W.
AN INTRINSIC VARIATIONAL FORMULATION OF NONLINEAR SHELL THEORY AND APPLICATION TO THE BEAM BENDING PROBLEM OF ORTHOTROPIC TUBES

Lighthill, J., Sir
NONLINEAR EDGE WAVES

Limam, A.: See Jullien, J.-F.

Ling, F.H.
MIXING WINDOWS IN CAVITY FLOWS FOUND BY BIFURCATION ANALYSIS

Ling, Z.: See Bai, Y.L.

Litovsky, E., Shapiro, M.
RELATIONSHIP BETWEEN MECHANICAL AND THERMOPHYSICAL PROPERTIES OF POROUS CERAMIC MATERIALS

Liu, Y., Hong, J.
MATHEMATICAL SIMULATION AND EXPLANATION OF THE MOTION OF THE HUMAN BODY

Liu, Z.: See Chen, C.

Lopatnikov, S.L.: See Gurevich, B.

Lopatnikov, S.L.: See Ivanov, V.V.

Lu, C.S.: See Bai, Y.L.

Lucey, A.D., Carpenter, P.W.
THE DYNAMICS OF AN EXCITED FLEXIBLE PLATE IN THE PRESENCE OF A FLUID FLOW

Lund, E.: See Olhoff, N.

Lundberg, R.: See Neumeister, J.M.

Luo, C. and Han, R.P.S.
DYNAMIC STABILITY ANALYSIS OF AN IRRATIONALLY NONLINEAR ROD

Luo, Z.
ARCHITECTURE AND DENSITY DISTRIBUTION OF CANCELLOUS BONE STUDIED BY OPTIMIZATION METHOD

Lurie, K.A.
ON A GENERAL CONCEPT IN OPTIMAL MATERIAL LAYOUT

Lyubimov, G.A.
THE CHOICE OF THE MODEL TO DESCRIBE THE ELASTIC PROPERTIES OF HUMAN LUNG

Mac Sithigh, G.P.: See Cohen, H.

Mac Sithigh, R.G.: See Cohen, H.

Mace, B.R.: See Shoavi, E.

Magnaudet, J.: See Thais, L.

Maier, G., Perego, U.
ON ELASTIC-PLASTIC DYNAMIC ANALYSIS IN THE PRESENCES OF SOFTENING

Maigre, H.: See Rittel, D.

Maity, A.K.: See Samanta, S.K.

Makarov, S.O.: See Bratukhin, Y.K.

Mal, A.K., Bar-Cohen, Y.
A THEORETICAL AND EXPERIMENTAL STUDY OF WAVE PROPAGATION IN COMPOSITE LAMINATES

Maldarelli, C.: See Georgiou, E.

Malik, N.H.: See Hardy, S.J.

Mallick, K.: See Krajcinovic, D.

Malomed, B.A.: See Maugin, G.A.

Manevich, A., Krasovskij, V.
AN EXPERIMENTAL TESTING OF COUPLING BUCKLING THEORY OF STIFFENED CYLINDRICAL SHELLS

Manevitch, L.I.: See Gendelman, O.V.

Manoach, E.: See Karagiozov, V.

Markenscoff, X., Dundurs, J.
ELASTIC SINGULARITIES AT INTERFACES

Markov, K.Z., Zvyatkov, K.D.
FUNCTIONAL SERIES AND HASHIN-SHTRIKMAN'S TYPE BOUNDS ON THE EFFECTIVE PROPERTIES OF RANDOM MEDIA

Markus, S., Stavsky, Y.
ONE-DIMENSIONAL COUPLED THERMAL SHOCK PROBLEM

Martins, J.A.C., Faria, L.O., Guimaraes, J.
FRICTION-INDUCED SURFACE FLUTTER INSTABILITIES IN LINEAR ELASTICITY

Mataga, P.A., Ponte-Castaneda, P.
ELASTIC-PLASTIC INTERFACIAL CRACK GROWTH

Matasov, A.I.: See Golovan, A.A.

Maugin, G.A., Hadouaj, H., Malomed, B.A.
SURFACE ELASTIC WAVES OF THE ENVELOPE SOLITON TYPE

Maugin, G.A., Trimarco, C.
THE ROLE OF FULLY MATERIAL BALANCE LAWS IN CONTINUUM MECHANICS

Maurel, P.: See Blanze, C.

Mauri, R., Brenner, H.
RHEOLOGICAL PROPERTIES OF NEUTRALLY BUOYANT SUSPENSIONS DETERMINED BY MOMENTUM TRACER METHODS

Mauri, R.: See Acrivos, A.

Mavko, G.: See Dvorkin, J.

Meijaard, J.P.
MECHANICAL DYNAMICAL SYSTEMS WITH DISCONTINUITIES

McGrath, J.: See Fernando, H.J.S.

McLaren, D.G.: See Leutheusser, H.J.

Meier, G.: See Wenskus, R.

Meirovitch, L., Hagedorn, P.
A NEW APPROACH TO THE MODELING OF DISTRIBUTED NON-SELF-ADJOINT SYSTEMS

Mendelsohn, D.A., Gross, T.S., Young, L.-J.J.
FRACTURE SURFACE INTERFERENCE IN SHEAR OR MIXED-MODE PLANE STRAIN LOADING

Menken, C.M., Kouhia, R.J.
USING THE ASYMPTOTIC APPROACH CORRECTLY AS A PREDICTOR TO THE POST-BUCKLING ANALYSIS

Mersov, G.A., Nezlina, Y.M., Moiseev, S.S.
THE INVESTIGATION METHOD OF EVOLUTION OF THE NONLINEAR SYSTEM DISTURBED BY MULTIPLICATIVE NOISE

Metelka, T.: See Klimes, F.

Meyer, L.W.
ADIABATIC SHEAR FAILURE UNDER BIAXIAL DYNAMIC COMPRESSION/ SHEAR LOADING

Meyer, M., Cuche, D.
THE ELASTO-PLASTIC PLATE WITH A HOLE: COMPARISON OF AN ANALYTICAL MODEL WITH HOLOGRAPHIC MEASUREMENTS

Middendorf, P.: See Wenskus, R.

Miehe, C.: See Stein, E.

Mikkelsen, L.
ON THE ANALYSIS OF VISCOPLASTIC BUCKLING

Millien, A., Boehler, J.-P.
ANALYSIS OF SHEAR BANDS IN ANISOTROPIC ROCKS

Miloh, T., Zilman, G.,
WAVE PATTERNS EXERTED BY A SHIP MOVING IN HEAVY STRATIFIED SEAS

Miloh, T.: See Galper, A.

Minahen, T.M.: See Knauss, W.G.

Minel, F.: See George, J.

Mioduchowski, A.: See Haddow, J.B.

Mirkin, M.A., Vakulenko, A.A., Zhuravlev, V.P., Zimin, V.A.
THE DIFFUSION FRACTURE MECHANISM IN THE HIGH-MOLECULAR SOLID BODIES

Mitra, N.K.: See Güntermann, T.

Mizrahi, J., Giat, Y., Levy, M.
A DYNAMICALLY DETERMINATE MODEL OF FATIGUE AND RECOVERY OF ELECTRICALLY STIMULATED PARALYSED MUSCLES

Mochel, A.N., Salganik, R.L.
MECHANICAL-MATHEMATICAL MODELS OF MATERIALS BEHAVIOR ACCOUNTING FOR DAMAGE ACCUMULATION

Mochizuki, O.: See Kiya, M.

Moffatt, H.K., Jeong, J.-T.
FREE SURFACE CUSPS IN FLOWS AT LOW REYNOLDS NUMBERS

Moiseev, S.S.: See Mersov, G.A.

Molinari, A.: See Leroy, Y.M.

Mond, M.: See Cederbaum, G.

Mondy, L.A., Geller, A.S., Rader, D.J., Ingber, M.S.
APPLICATION OF BOUNDARY ELEMENT TECHNIQUES TO THE MOTION OF NONSPHERICAL PARTICLES IN STOKES FLOW

Mondy, L.A.: See Abbott, J.R.

Moran, B.: See Yu, M.

Moshaiov, A.: See Blau, M.

Mountian, I. Nielens, H., Raucent, B., Samin J.Cl., Willems, P. Willems, P.Y.
IDENTIFICATION MODEL FOR NORMAL AND PATHOLOGICAL GAIT ANALYSIS

Mow, V.C.: See Cohen, B.

Mroz, Z.
SENSITIVITY OF NON-LINEAR STRUCTURES IN REGULAR AND CRITICAL STATES TO DESIGN AND LOAD VARIATIONS

Mülhaus, H.-B., de Borst, R., Sluys, L.J.
THE ROLE OF HIGHER ORDER GRADIENTS IN STRAIN SOFTENING MATERIALS: STATIC AND DYNAMIC ANALYSES

Mülhaus, H.-B: See Vardoulakis, I.

Munson, D.E.: See Brodsky, N.S.

Muraoka, M.: See Kumagai, T.

Musafir, R.E.
ON THE RELATIONSHIP BETWEEN THE SOUND AND FLOW FIELDS OF COHERENT STRUCTURES IN JET FLOWS

Nagorny, M.: See Branover, H.

Nakatani, A.: See Kitagawa, H.

Nappi, A.: See Amadio, C.

Needleman, A., Ortiz, M.
STABILITY AND FAILURE AT INTERFACES

Neumeister, J.M., Hansson, T., Lundberg, R.
MECHANISMS OF FAILURE IN CONTINUOUS FIBER REINFORCED BRITTLE MATRIX COMPOSITES

Nevo, E., Lanir, Y.
RESIDUAL STRAIN AND LV DIASTOLIC PERFORMANCE - A MORPHOLOGICALLY BASED STRUCTURAL ANALYSIS

Nevo, E., Lanir, Y.
COMPARISON OF STRUCTURAL AND PHENOMENOLOGICAL CHARACTERIZATION

Nezlin, M.V.
COMMON MECHANISM FOR VORTEX STRUCTURE DRIVE IN PLANETARY ATMOSPHERES AND GALAXIES

Nezlina, Y.M.: See Mersov, G.A.

Nielens, H.: See Mountian, I.

Nielsen, N.F., Larsen, P.S.
MODELS OF THE MUSSEL GILL PUMP COMPARED TO IN-VITRO EXPERIMENTS

Nikolaev, V.I.: See Smirnov, B.I.

Nir, A.: See Acrivos, A.

Noack, B.R., Eckelmann, H.
THEORETICAL INVESTIGATION OF THE CYLINDER WAKE WITH A GALERKIN METHOD

Noda, N.: See Ashida, F.

Nolde, E.V.: See Goldenveizer, A.L.

Nolte, L.P.: See Pingel, T.H.

Nouailhas, D.
CYCLIC BEHAVIOR OF SINGLE-CRYSTAL SUPERALLOYS: EXPERIMENTS AND MODELING

Nowak, Z.
MODELLING OF THERMAL, VISCO-PLASTIC RESPONSE IN PLATE IMPACT EXPERIMENTS

Nuller, B.: See Ryvkin, M.

Nur, A.: See Dvorkin, J.

Obermeier, F.: See Wenskus, R.

Ogden, R.W., Roxburgh, D.G.
STABILITY AND VIBRATION OF PRE-STRESSED ELASTIC BLOCKS

Ogura, Y.: See Kiya, M.

Ohle, F., Eckelmann, H.
LOW-DIMENSIONAL MODELING OF HYDRODYNAMICAL SYSTEMS

Okui, Y.: See Horii, H.

Olhoff, N., Rasmussen, J., Lund, E.
METHOD OF ERROR ELIMINATION IN FINITE ELEMENT BASED SEMI-ANALYTICAL SHAPE SENSITIVITY ANALYSES

Olszok, T.: See Stamm, G.

Ortiz, M.: See Needleman, A.

Osinski, Z.: See Golos, K.

Ostrach, S.: See Gulic, I.

Ozbolt, J.: See Bazant, Z.

Page, M.: See Cowley, S.J.

Pan, R.: See Davies, H.G.

Papageorgiou, D.: See Georgiou, E.

Papamichos, E.: See Vardoulakis, I.

Pecherski, R.B.
ON THE EFFECTS OF MICRO-SHEAR BANDS ON YIELD SURFACE VERTICES AND PLASTIC FLOW INSTABILITY

Pedersen, P.T.
SHIP COLLISION DYNAMICS

Pedley, T.J.: See Le Baron, L.

Pellegrini, O.
EXPERIMENTAL STUDIES ON THE INFLUENCE OF MICROCRACKS ON MATERIAL PROPERTIES

Pellegrino, S.: See Guest, S.D.

Peng, C.: See Reid, S.R.

Peradzynski, Z.
ON THE INITIAL VALUE PROBLEM FOR HVBK EQUATIONS OF SUPERFLUID He^4

Perego, U.: See Maier, G.

Perera, M.A.P. See Fernando, H.J.S.

Perktold, K., Peter, R.O., Gürtl, R.
COMPUTER SIMULATION OF PULSATILE NON-NEWTONIAN THREE-DIMENSIONAL FLOW CHARACTERISTICS IN COMPLIANT CAROTID ARTERY BIFURCATION MODELS

Perzyna, P.
INSTABILITY PHENOMENA AND ADIABATIC SHEAR BAND LOCALIZATION IN THERMOPLASTIC FLOW PROCESSES

Peter, R.O.: See Perktold, K.

Petryk, H.
INSTABILITY AND BIFURCATIONS IN INCREMENTALLY NONLINEAR SOLIDS

Pfeiffer, F.
SOME FUNDAMENTAL PROBLEMS IN STICK-SLIP VIBRATIONS

Piechna, J.: See Stamm, G.

Pierre, C.: See Shaw, S.W.

Pingel, T.H., Nolte, L.P., Stumpf, H.
CONTRIBUTION TO THE DERIVATION AND NUMERICAL REALIZATION OF A MATHEMATICAL MODEL FOR THE HUMAN SPINE

Ponte Castaneda, P., Zaidman, M.
THE EFFECT OF TEXTURE DEVELOPMENT ON THE LOCALIZATION OF POROUS METALS

Ponte Castaneda, P.: See DeBotton, G.

Ponte Castaneda, P.: See Mataga, P.A.

Popp, K.: See Austerman, R.

Potier-Ferry, M.: See Cochelin, B.

Potier-Ferry, M.: See Ferron, G.

Prakash, V.: See Clifton, R.J.

Rader, D.J.: See Mondy, L.A.

Rajgelj, S.: See Amadio, C.

Rasmussen, J.: See Olhoff, N.

Rassokha, A.
SOLITON LOCALIZATION OF PLASTIC DEFORMATION

Raucent, B.: See Mountian, I.

Ravid, M., Bodner, S.R., Holcman, I.
A TWO-DIMENSIONAL ANALYSIS OF PENETRATION BY AN ERODING PROJECTILE

Rebel, G.: See Arbocz, J.

Reid, S.R., Peng, C.
DYNAMIC ENHANCEMENT OF THE CRUSHING STRENGTH OF WOOD

Resch, F.J.: See Afeti, G.M.

Reshotko, E.: See Liang, F.-P.

Reynier, M.: See Ladeveze, P.

Ricca, R.L.
LOCAL DYNAMICS OF A TWISTED VORTEX FILAMENT

Ricci, N., Caulliez, G.
GROWTH STAGES AND PROPERTIES OF THE INITIAL GRAVITY-CAPILLARY WIND-WAVES

Rimrott, F.P.J.: See Yeh, K.Y.

Rittel, D., Maigre, H., Bui, H.D.
EXPERIMENTAL EVALUATION OF THE DYNAMIC STRESS INTENSITY FACTORS: A RECENT DEVELOPMENT

Rivkin, L.: See Givoli, D.

Rizzi, N., Ruta, G., Tatone, A.
THE INFLUENCE OF WARPING ON THE POSTBUCKLING BEHAVIOR OF THIN-WALLED BEAMS WITH OPEN CROSS-SECTION

Rogers, T.G.: See Watson, P.

Rok, V.E.: See Ivanov, V.V.

Rong, T.-Y.
THEORY AND METHOD OF STRUCTURAL VARIATIONS OF FINITE ELEMENT SYSTEMS IN SOLID MECHANICS

Roorda, J.: See Camotim, D.

Rotem, A.
EFFECT OF THE FREQUENCY OF THE ALTERNATING LOAD ON FATIGUE DELAMINATION OF LAMINATES

Rougee, P.
THE MANIFOLD OF LOCAL METRIC STATES OF CONTINUUM

Roxburgh, R.G.: See Ogden, R.W.

Rozvany, G.I.N.: See Zhou, M.

Rubin, M.B.
AN ALTERNATIVE FORMULATION OF CONSTITUTIVE EQUATIONS FOR AN ELASTICALLY ISOTROPIC ELASTIC-PLASTIC MATERIAL

Rubinstein, A.A.
STRENGTH AND TOUGHNESS ANALYSIS OF THE FIBER REINFORCED BRITTLE MATRIX COMPOSITE

Rumschitzki, D.: See Georgiou, E.

Rumyantsev, V.V.
STABILITY OF STEADY MOTIONS OF RIGID BODIES WITH CAVITIES CONTAINING LIQUID

Ruta, G.: See Rizzi, N.

Ruzi, T.: See Fukuyu, A.

Ryaboy, V.M.
LIMITING PERFORMANCE OF CONTROLLED AND NON-CONTROLLED MULTI-BODY VIBRATION ISOLATING SYSTEMS

Ryvkin, M., Nuller, B.
CUTTING OF A CELLULAR ELASTIC MATERIAL BY A RIGID TRIANGULAR CUTTER

Sabag, M.: See Greenberg, J.B.

Sabina, F.J., Willis, J.R.
SELF-CONSISTENT ANALYSIS OF WAVES IN A POLYCRYSTALLINE MEDIUM

Sadovnichaja, A.P.: See Gurevich, B.

Sakao, F., Sato, H.
COMPOSITION AND CALIBRATION OF NEW-TYPE LUNG-SOUND SENSORS BASED ON MICROPHONES

Salganik, R.L.: See Mochel, A.N.

Samanta, S.K., Maity, A.K.
A YIELD CRITERION AND AN ASSOCIATED FLOW RULE FOR CAVITATED TWO-PHASE MATERIALS

Samin, J.Cl.: See Mountian, I.

Sanchez Sanchez, H.: See Jullien, J.-F.

Sato, H.: See Sakao, F.

Sawyers, D.R., Sen, M., Chang, H.-C.
CHAOTIC MIXING WITH CHEMICAL REACTION

Sayir, M.: See Spirig, T.H.

Sayir, M.: See Staudenmann, M.

Schiehlen, W.: See Blajer, W.

Schirm, W.: See Blajer, W.

Schumacher, A.: See Eschenaner, H.A.

Sen, M.: See Sawyers, D.R.

Seneviratne, L.D., Earles, S.W.E., Fenner, D.N.
ANALYSIS OF MECHANISMS WITH CLEARANCE JOINTS USING FINITE ELEMENT RESULTS

Sepe, V., Sinopoli, A.
COUPLED MOTION IN THE DYNAMIC ANALYSIS OF A THREE-BLOCK STRUCTURE

Shapiro, M.: See Goldshtein, A.

Shapiro, M.: See Litovsky, E.

Sharf, I., Tabarrok, B.
NONLINEAR STIFFNESS TENSOR OF A BEAM ELEMENT FOR DYNAMICS SIMULATION OF MULTIBODY SYSTEMS

Sharma, S.D.: See Chen, X.N.

Shaw, S.W., Pierre, C.
MODAL ANALYSIS FOR NONLINEAR DYNAMICAL SYSTEMS

Sheinman, I.: See Adan, M.

Sheludko, G.A., Shupikov, A.N., Sotrichin, S.Y.
NON-STATIONARY VIBRATIONS OF MULTILAYER PLATES UNDER INTEN-SIVE ACTION OF SHORT DURATION AND THEIR OPTIMIZATION. THEORY AND EXPERIMENT

Shesterikov, S.A., Yumasheva, M.A., Yumashev, M.W.
INSTABILITY OF THERMOFRACTURE AT RESTRICTED DEFORMATION

Shi, J., Crisfield, M.
A SIMPLE INDICATOR AND BRANCH SWITCHING TECHNIQUE FOR HIDDEN UNSTABLE EQUILIBRIUM PATHS

Shilkrut, D.
STABILITY OF EQUILIBRIUM STATES OF NONLINEAR STRUCTURES AND CHAOS PHENOMENON. CHAOS IN STATICS?

Shillor, M.
ON THERMOELASTIC CONTACT

Shimizu, M.: See Kiya, M.

Shizawa, K.: See Kobayashi, S.

Shoavi, E., Mace, B.R.
VIBRATION ISOLATION BY PASSIVE DESTRUCTIVE INTERFERENCE

Shopov, P.: See Zapryanov, Z.

Shpeizman, V.V.: See Smirnov, B.I.

Shtern, V., Hussain, F.
INERTIAL INSTABILITY OF DIVERGENT FLOWS

Shtilman, L., Spector, M., Tsinober, A.
ALIGNMENTS IN HOMOGENEOUS TURBULENCE - KINEMATICS VERSUS DYNAMICS

Shukhman, I.G.: See Churilov, S.M.

Shupikov, A.N.: See Sheludko, G.A.

Sinopoli, A.: See Sepe, V.

Sirovich, L.: See Handler, R.A.

Sjödin, B., Klarbring, A.
NUMERICAL METHODS FOR ELASTO-PLASTIC MATERIALS AT FINITE STRAINS

Skoczen, B., Skrzypek, J., Zyczkowski, M.
INSTABILITY AND OPTIMAL DESIGN OF BELLOWS AGAINST BUCKLING BASED UPON THE EQUIVALENT COLUMN CONCEPT

Skrzypek, J.: See Skoczen, B.

Slepyan, L.I.
MICROMECHANICAL ASPECTS OF CRACK DYNAMICS AND ENERGY CRITERIA

Sluys, L.J., de Borst, R.
COMPUTATIONAL MODELLING OF SHEAR BANDING IN AN IMPACT BIAXIAL TEST

Sluys, L.J.: See Mülhaus, H.-B.

Smirnov, B.I., Chudnova, V.I., Nikolaev, V.I., Shpeizman, V.V.
PLASTIC INSTABILITY IN CRYSTALS UNDER CONSTANT STRESS AT ELEVATED TEMPERATURES

Smith, S.H.
THE DEVELOPMENT OF SOURCE-SINK FLOW IN A ROTATING FLUID

Soldatos, K.P.
A GENERAL LAMINATED PLATE THEORY ACCOUNTING FOR INTERLAMINAR STRESS CONTINUITY

Söller, C.: See Wenskus, R.

Song, W.P.: See Yeh, K.Y.

Sotiropoulos, D.A.
INSTABILITIES OF ELASTIC WAVES IN PERIODIC STRUCTURES

Sotrichin, S.Y.: See Sheludko, G.A.

Spector, A.: See Goldstein, R.

Spector, A.: See Shtilman, L.

Spence, D.A.
BUOYANCY-DRIVEN CRACK PROPAGATION: A NEW CLASS OF STEADY SOLUTIONS

Spencer, A.J.M.: See Watson, P.

Spirig, T.H., Sayir, M.B.
CRACK PROPAGATION INSTABILITIES IN A BRITTLE PIPE

Stamm, G., Fiszdon, W., Olszok, T. Piechna, J.
CYLINDRICAL CONVERGING MODERATE AMPLITUDE SECOND-SOUND WAVES IN SUPERFLUID HELIUM

Staudenmann, M., Sayir, M.B.
FLEXURAL WAVE PULSES IN A ROD WITH SLIGHTLY DIFFERENT STIFFNESSES IN THE PRINCIPAL DIRECTIONS

Stavsky, Y.: See Greenberg, J.B.

Stavsky, Y.: See Markus, S.

Steele, C.R.
INSTABILITIES IN BIOLOGICAL TISSUE: (1) ENHANCEMENT OF ELECTRO-MOTILITY IN COCHLEAR RECEPTOR CELLS AND (2) MORPHOGENESIS OF PLANTS

Steif, P.S.
EFFECT OF INTERFACE PROPERTIES ON THE STRENGTH OF FIBER-REINFORCED CERAMICS

Stein, A.A.
MECHANICAL ASPECTS OF BIOLOGICAL MORPHOGENESIS

Stein, E., Miehe, C.
A MODEL OF FINITE STRAIN THERMOVISCOPLASTICITY: ASPECTS OF THE FORMULATION AND NUMERICAL TIME INTEGRATION

Stepanyants, Y.A.: See Fabrikant, A.L.

Stiassnie, M.: See Agnon, Y.

Stronge, W.J.
NONLINEAR DYNAMICS OF SWERVE IN FRICTIONAL COLLISIONS

Stronge, W.J.: See Durban, D.

Stumpf, H., Chau, L.-K.
NONLINEAR CRACK PROBLEM: CLASSIFICATION OF SINGULARITIES FOR RUBBERLIKE MATERIALS

Stumpf, H.: See Pingel, T.H.

Su, F.-C., Chen, C.-T.
A MATHEMATICAL MODEL OF BOTH TIBIOFERMORAL AND PATELLO-FERMORAL JOINTS

Su, X.: See Chen, C.

Sulem, J.: See Vardoulakis, I.

Sumi, Y.
A SECOND ORDER PERTURBATION METHOD FOR THE STRESS ANALYSIS OF SOLIDS WITH SLIGHTLY WAVY OR IRREGULAR SURFACES

Sun, X.-F.: See Zeng, P.

Suquet, P.: See Thebaud, F.

Sutcliffe, M.P.F., Fleck, N.A.
EFFECT OF GEOMETRY ON COMPRESSIVE FAILURE OF NOTCHED COMPOSITES

Swart, P.: See Holmes, P.

Szyszkowski, W.
CONTROLLING STRUCTURAL INSTABILITIES VIA EIGENVALUES OPTIMIZATION IN BUCKLING ANALYSIS

Tabarrok, B., Tong, L., Tezer, C.
DUALITY IN MINIMAL SURFACES

Tabarrok, B.: See Sharf, I.

Takahashi, K.: See Kobayashi, S.

Tanveer, S.: See Cowley, S.J.

Tarnai, T.
APPLICATION OF MECHANICS TO DESIGN OF STRUCTURES OF MOLECULAR BIOLOGY: VIRUSES AND COATED VESICLES

Tatone, A.: See Rizzi, N.

Tatsumi, T., Yoshimura, T.
NONLINEAR INSTABILITY OF THE LAMINAR FLOW IN A RECTANGULAR DUCT AND GENERATION OF THE "TURBULENT SECONDARY FLOW"

Tauchert, T.R.: See Ashida, F.

Tezer, C.: See Tabarrok, B.

Thais, L., Magnaudet, J.
INTERACTIONS OF WIND WAVES WITH TURBULENCE: AN ATTEMPT TO CONNECT THEORY AND EXPERIMENT

Thébaud, F., Hervé, E., Suquet, P., Zaoui, A.
MODELLING NONLINEAR BEHAVIOR OF COMPOSITES WITH IMPERFECT INTERFACES

Thies, M., Lenz, J.
TRANSIENT AND RESIDUAL THERMAL STRESSES IN CERAMO-METALLIC CROWNS

Thomas, P.: See Wierzbicki, T.

Thomsen, O.T.
ANALYSIS OF LOCAL BENDING EFFECTS IN FRP-SANDWICH PANELS BY USE OF A TWO-PARAMETER ELASTIC FOUNDATION MODEL

Tirosh, J., Ben-David, D.
FLUID-PRESSURE-ASSISTED DEEP-DRAWING OF HARDENING AND RATE-SENSITIVE MATERIALS AT HIGH SPEEDS

Tong, L.: See Tabarrok, B.

Trabucho, L.
SOME RESULTS FOR A CLASS ON CONTACT PROBLEMS WITH FRICTION IN ELASTODYNAMICS

Triantafyllou, G.G.: See Dimas, A.A.

Trimarco, C.: See Maugin, G.A.

Truskinovsky, L.
KINKS VERSUS SHOCKS

Tsay, R.-Y., Weinbaum, S., Curry, F.-R.
A COMBINED JUNCTION-PORE-FIBER ENTRANCE LAYER MODEL FOR CAPILLARY PERMEABILITY

Tsinober, A.: See Shtilman, L.

Tsitverblit, N., Kit, E.
A NUMERICAL BIFURCATION STUDY OF DOUBLE-DIFFUSIVE CONVECTION IN A RECTANGULAR ENCLOSURE

Tur, A.V., Levich, E., Yanovsky, V.
ON FLUID MOTION IN PRESENCE OF DILUTE POLYMER ADDITIVES

Tvergaard, V.
EFFECT OF DUCTILE PARTICLE DEBONDING DURING CRACK-BRIDGING IN CERAMICS

Tylikowski, A., Hetnarski, R.B.
THERMALLY INDUCED DYNAMIC INSTABILITY OF BEAMS AND PLATES

Tyvand, P.A.
A FLOW MODEL FOR WOOD FIBERS: SLENDER-BODY THEORY

Uetani, K.
SYMMETRY LIMIT THEORY AND INELASTIC BIFURCATION THEORY (ON THE CRITICAL BEHAVIOR OF SIMPLY-SUPPORTED BEAM-COLUMNS SUBJECTED TO CYCLIC CENTRAL DEFLECTION)

Ursell, F.
TRAPPED MODES IN THE LINEAR THEORY OF WATER WAVES AND SOUND WAVES

Vakulenko, A.A.: See Frolov, I.V.

Vakulenko, A.A.: See Mirkin, M.A.

van Beek, P.
A SOLUTION FOR THE TWO-SPHERE PROBLEM IN POTENTIAL FLOW IN TERMS OF CHEBYSHEV POLYNOMIALS

van Campen, D.H., Fey, R.H.B., de Kraker, A.
LONG TERM STRUCTURAL DYNAMICS OF MECHANICAL SYSTEMS WITH LOCAL NONLINEARITIES

van Campen, D.H.: See Janssen, J.D.

van der Giessen, E.
INTERACTION EFFECTS DURING GRAIN BOUNDARY CAVITATION AND MICROCRACKING IN CREEPING POLYCRYSTALS

van Gemert, R.: See Huyghe, J.M.

van Wijngaarden, L.
HYDRODYNAMIC NOISE FROM CAVITATION AND OTHER BUBBLY FLOWS

Vakakis, A.F.
NONLINEAR INTERACTION OF STRUCTURAL WAVES AND ELASTIC JOINTS

Vardoulakis, I., Papamichos, E.
SURFACE SPALLING IN BRITTLE ROCK UNDER COMPRESSION

Vardoulakis, I., Sulem, J., Mülhaus, H.-B.
INTERNAL BUCKLING INSTABILITIES IN A FLUID SATURATED LAYERED ROCK

Vassiliev, D.: See Laptev, A.

Vaynblat, D.: See Freidin, A.

Verdonck, P., De Wachter, D., Verhoeven, R.
HEMODYNAMICAL ANALYSIS OF ARTIFICIAL ORGANS

Verhoeven, R.: See Verdonck, P.

Verlinden, O., Dehombreux, P., Conti, C., Boucher, S.
A NEW FORMULATION FOR THE DIRECT DYNAMIC SIMULATION OF FLEXIBLE MECHANISMS BASED ON THE NEWTON-EULER INVERSE METHOD

Vigdergauz, S.
TWO-DIMENSIONAL GRAINED COMPOSITES OF EXTREMAL RIGIDITY

Voloshin, A.S.
EXPERIMENTAL INVESTIGATION OF THE STRESS SINGULARITIES BY ENHANCED MOIRE INTERFEROMETRY

Vujicic, V.A.
A GENERAL THEOREM OF OPTIMAL CONTROL OF MOTION OF MECHANICAL NON-AUTONOMOUS SYSTEMS

Wang, C.-T.: See Chu, C.-C.

Wang, G.T.
DYNAMIC PLASTIC BUCKLING OF CYLINDRICAL SHELL UNDER AXIAL IMPACT - AN EXPERIMENTAL STUDY

Wang, R.: See Batra, R.C.

Wang, C.A.: See Dickstein, P.A.

Watson, P., Rogers, T.G., Spencer, A.J.M.
EXACT THREE DIMENSIONAL ELASTICITY SOLUTIONS FOR BENDING OF MODERATELY THICK INHOMOGENEOUS AND LAMINATED STRIPS UNDER NORMAL PRESSURE

Wegner, J.L.
ON THE PROPAGATION OF WAVES FROM A SPHERICAL CAVITY IN AN UNBOUNDED LINEAR VISCOELASTIC SOLID

Wei, K. and De Bremaecker, J.-C.
THE DIRECTION OF FRACTURE

Wei, R.-J.: See Chen, X.-N.

Weinbaum, S.: See Tsay, R.-Y.

Weinbaum, S.: See Wu, W.-Y.

Weitsman, Y. and Zhu, H.
MULTI-FRACTURE OF CERAMIC COMPOSITES

Weng, G.J.
SELF-CONSISTENT DETERMINATION OF THE TIME-DEPENDENT CREEP OF POLYCRYSTALS

Wenskus, R., Söller, C., Middendorf, P., Meier, G.E.A., Obermeier, F.
TOMOGRAPHIC RECONSTRUCTION OF THREE-DIMENSIONAL DENSITY FIELDS FROM INTERFEROMETRIC DATA

Wiener, Z.
INSTABILITY OF DEGENERATE EQUILIBRIUM

Wierzbicki, T., Thomas, P.
CLOSED-FORM SOLUTION FOR WEDGE CUTTING FORCE THROUGH A THIN METAL SHEET

Willems, P.: See Mountian, I.

Willems, P.Y.: See Mountian, I.

Willis, J.R.
ASYMPTOTIC ANALYSIS OF CRACK BRIDGING BY DUCTILE FIBERS

Willis, J.R.: See Sabina, F.J.

Wnuk, M.: See Adziev, T.

Wu, E., Yen, C.-S.
DETECTION OF IMPACT FORCE HISTORY AND LOCATION ON A LAMINATED PLATE

Wu, S., Haung, W., Chudnovsky, A.
ENERGY ANALYSIS OF PROCESS ZONE EVOLUTION

Wu, W.-Y., Yan, Z.-Y., Acrivos, A., Weinbaum, S.
FLUID SKIMMING AND PARTICLE SCREENING AT MICROVASCULAR BIFURCATONS

Xia, M.F.: See Bai, Y.L.

Yan, Z.-Y.: See Wu, W.-Y.

Yang, Z.
ON THE LINEAR STABILITY OF A TRAILING LINE VORTEX

Yanovsky, V.: See Tur, A.V.

Yarin, A.L.
STRONG FLOWS OF POLYMERIC LIQUIDS IN JETS AND FILMS: RHEOLOGY AND HYDRODYNAMICS

Yeh, K.Y., Song, W.P., Rimrott, F.P.J.
ON THE INSTABILITIES OF TRANSVERSELY LOADED CORRUGATED CIRCULAR DIAPHRAGMS

Yen, C.-S. See Wu, E.

Yoshida, H.
IMPACT PERFORATION OF URETHANE-FOAM SPONGE

Yoshimura, T.: See Tatsumi, T.

Young, L.J.: See Mendelsohn, D.A.

Yu, M., Moran, B., Keer, L.M.
A DIRECT ANALYSIS OF SHAKEDOWN IN ELASTIC-PLASTIC THREE-DIMENSIONAL ROLLING CONTACT

Yu, Yi.-Y.
EQUATIONS FOR LARGE DEFLECTIONS OF ELASTIC AND PIEZOELECTRIC PLATES AND SHELLS, INCLUDING SANDWICHES AND LAMINATED COMPOSITES, WITH APPLICATIONS TO VIBRATIONS, CHAOS AND ACOUSTIC RADIATION

Yumashev, M.W.: See Shesterikov, S.A.
Yumasheva, M.A.: See Shesterikov, S.A.

Zaccaria, D.: See Bigoni, D.

Zachariadis, D.C.
OIL WHIP LIKE INSTABILITY OF ROTORS ON SYMMETRIC LINEAR BEARINGS
Zaidman, M.: See Ponte Castaneda, P.

Zannetti, L., Franzese, P.
ADVECTION BY A POINT VORTEX IN CLOSED DOMAINS

Zaoui, A.: See Thebaud, A.

Zapryanov, Z., Bazhlekov, I., Shopov, P.
THE TIME-DEPENDENT DEFORMATION OF COMPOUND MULTIPHASE DROPS AT MODERATE REYNOLDS NUMBERS

Zarmi, Y.: See Cederbaum, G.

Zeng, P., Sun, X.-F., Gao, Q.
COMPUTATION DAMAGE MECHANICS: CONCEPT AND PRINCIPLE

Zhang, C.: See Gross, D.

Zhang, L.: See Hwang, K.-Ch.

Zhang, R.J., Zhang, W.
BENDING THEORY OF TOROIDAL SHELLS UNDER ARBITRARY LOADINGS

Zhang, W.: See Zhang, R.J.

Zhou, M., Rozvany, G.I.N., Birker, T.
TOPOLOGY OPTIMIZATION FOR ALTERNATIVE LOAD CONDITIONS AND MULTIPLE DESIGN CONSTRAINTS

Zhou, X., Cui, H.
NON-PROPAGATING SOLITARY WAVES AND SURFACE TENSION

Zhu, H.: See Weitsman, Y.

Zhu, Z.G.: See Batra, R.C.

Zhuravlev, V.P.: See Mirkin, M.A.

Zilman, G.: See Miloh, T.

Zimin, V.A.: See Mirkin, M.A.

Zolochevsky, A.: See Altenbach, H.

Zvyatkov, K.D.: See Markov, K.Z.

Zyczkowski, M.: See Skoczen, B.